Residual Versus Suppressed-Carrier Coherent Communications

M. K. Simon and S. Million
Communications and Systems Research Section

This article addresses the issue of when to suppress or not to suppress the transmitted carrier in designing a coherent communication system employing a carrier tracking loop for carrier synchronization. Assuming that a phase-locked loop (PLL) is used whenever there exists a residual carrier and a Costas loop is used whenever the carrier is suppressed, the regions of system parameters that delineate these two options are presented based on the desire to minimize the average probability of error of the system.

I. Introduction

In the design of coherent communication systems, engineers are forever faced with the decision of whether or not to suppress the transmitted carrier. In general, for a given total power available, P_t, a fraction of the power, P_c, would be allocated to the carrier signal (which determines the accuracy of the carrier synchronization process, e.g., the loop phase jitter) and the remainder of the power, P_d, would be allocated to the data bearing signal (which determines the accuracy of the data detection process, e.g., the error probability, in the presence of perfect carrier synchronization). Since it is clear that the more power is allocated to the carrier the better will be the synchronization accuracy while at the same time the more power is allocated to the data-bearing signal the better will be the ideal (perfect synchronization) error probability, the issue at stake is how to trade off between these two conflicting power hungry requirements so as to minimize the *average* error probability of the system. In fact, more than three decades ago, it was shown [1] that for coherent communication systems that use a phase-locked loop (PLL) to track the residual carrier, there always exists an optimum (in the sense of minimum average error probability) split of the total power into the two components mentioned above. Stated another way, for sinusoidal carriers that are phase modulated by a binary data stream, there always exists an optimum phase modulation index and a corresponding minimum average error probability performance. When one examines the numerical results in [1], one observes that as the ratio of data rate ($\mathcal{R} = 1/T$) to loop bandwidth (B_L) increases, the optimum fractional allocation of power to the carrier, i.e., $m^2 \triangleq P_c/P_t$, diminishes. In fact, defining this ratio by $\delta \triangleq \mathcal{R}/B_L = 1/B_L T$, then for values of δ on the order of a few hundred (which is typical of most system designs), the fraction of total power allocated to the carrier that yields the minimum average error probability is on the order of $m^2 = 0.1$ or less over a large range of total signal-to-noise ratio (SNR), $R_t = P_t T/N_0$. This trend suggests the possibility of using a suppressed-carrier system, i.e., $m^2 = 0$, which itself requires replacing the PLL with a loop capable of tracking a fully suppressed carrier, e.g., a Costas loop [2].

Since suppressed-carrier tracking loops inherently require a significantly larger loop SNR to track the carrier with a given accuracy (this comes about because of the nonlinear operation, e.g., multiplication of the in-phase (I) and quadrature-phase (Q) arms of the Costas loop, needed to reestablish the carrier), the ultimate question is whether or not there is an *overall* gain relative to the optimum residual carrier system described above. In particular, is the average error probability achieved with a suppressed carrier–Costas loop system smaller than the average error probability achieved with the optimally designed residual carrier–PLL system? Indeed we shall show that for a wide range of total SNRs (-10 dB $\leq R_t \leq 10$ dB) and a wide range of data rate-to-loop bandwidth ratios ($10 \leq \delta \leq 400$), the suppressed-carrier system achieves the smaller of the two average error probabilities. Thus, if it were not for other considerations, one would conclude that for a given total available power, loop bandwidth (determined by the system dynamics), and data rate, one should *always use a suppressed carrier–Costas loop system*! While, in principle, this statement is true, one must condition it on the ability of each loop (PLL for residual carrier and Costas loop for suppressed carrier) to stay in lock. Indeed, since as mentioned above, the Costas loop requires a larger loop SNR than does the PLL to yield a given tracking accuracy, the same is true in terms of the loop SNR required to maintain lock (herein referred to as the threshold value of loop SNR.) Thus, depending on the threshold SNR values decided upon for the two loops (to be discussed later on), there will exist a region of system parameters where one would be forced to employ a residual rather than a suppressed-carrier system, since in this region the loop SNR of the latter is below its threshold value whereas the loop SNR of the former is still above its threshold value. The purpose of this article is to define these regions, which will then clearly spell out for the system designer when to choose the suppressed-carrier option over the residual carrier one or vice versa. In the next section, we present the theoretical background necessary to establish these regions.

II. Residual Carrier System Model

As discussed in the introduction, we consider a residual carrier system that transmits a sinusoidal carrier of total power P_t phase modulated by a binary data stream $d(t)$ of rate $\mathcal{R} = 1/T$ symbols/s. As such, the transmitted signal is given by

$$s(t) = \sqrt{2P_t}\sin\left(\omega_0 t + \theta_m d(t)\right) \tag{1}$$

where θ_m is the phase modulation angle and ω_0 is the carrier frequency in rad/s. Using simple trigonometry, $s(t)$ can be rewritten in terms of its carrier and data components as

$$\begin{aligned} s(t) &= \sqrt{2P_t\cos^2\theta_m}\sin\omega_0 t + \sqrt{2P_t\sin^2\theta_m}d(t)\cos\omega_0 t \\ &= \sqrt{2m^2P_t}\sin\omega_0 t + \sqrt{2\left(1-m^2\right)P_t}d(t)\cos\omega_0 t \\ &\triangleq \sqrt{2P_c}\sin\omega_0 t + \sqrt{2P_d}d(t)\cos\omega_0 t \end{aligned} \tag{2}$$

where

$$\left.\begin{aligned} m^2 &= \cos^2\theta_m = \frac{P_c}{P_t} \\ 1-m^2 &= \sin^2\theta_m = \frac{P_d}{P_t} \end{aligned}\right\} \tag{3}$$

are the fractional carrier and data power components, respectively. The signal $s(t)$ is transmitted over an additive white Gaussian noise (AWGN) channel with single-sided power spectral density N_0 W/Hz.

At the receiver, a PLL tracks the discrete (residual) carrier component in Eq. (2). Assuming uncoded data and matched-filter detection, the error probability of the receiver, conditioned on a phase error ϕ in the PLL, is well-known and is given as [2]

$$P\left(E;\phi\right) = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{R_d}\cos\phi\right) \tag{4}$$

where $R_d \triangleq P_d T/N_0$ is the data detection SNR and erfc x denotes the complementary error function with argument x. The probability density function (pdf) of the loop phase error is well-known to be a Tikhonov distribution and is given by [2]

$$p\left(\phi\right) = \frac{\exp\left(\rho_{PLL}\cos\phi\right)}{2\pi I_0\left(\rho_{PLL}\right)}, \quad |\phi| \le \pi \tag{5}$$

where $I_0(x)$ is the modified Bessel function of the first kind with argument x,

$$\rho_{PLL} = \frac{P_c}{N_0 B_L} = \frac{m^2 P_t T}{N_0}\frac{1}{B_L T} = m^2 R_t \delta \tag{6}$$

is the loop SNR, B_L is the single-sided loop noise bandwidth, and the total SNR and data rate-to-loop bandwidth ratio are defined as

$$\left.\begin{aligned} R_t &\triangleq \frac{P_t T}{N_0} \\ \delta &= \frac{1}{B_L T} \end{aligned}\right\} \tag{7}$$

The average error probability performance of the system is obtained by averaging the conditional error probability of Eq. (4) over the pdf in Eq. (5), which yields

$$\begin{aligned} P\left(E\right) &= \int_{-\pi}^{\pi} P\left(E;\phi\right) p\left(\phi\right) d\phi \\ &= \int_{-\pi}^{\pi} \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\left(1-m^2\right)R_t}\cos\phi\right)\frac{\exp\left(m^2 R_t \delta\cos\phi\right)}{2\pi I_0\left(m^2 R_t\delta\right)} d\phi \\ &= 2\int_{0}^{\pi} \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\left(1-m^2\right)R_t}\cos\phi\right)\frac{\exp\left(m^2 R_t \delta\cos\phi\right)}{2\pi I_0\left(m^2 R_t\delta\right)} d\phi \end{aligned} \tag{8}$$

As stated in the introduction, for given values of R_t and δ, there exists an optimum value of m^2 (or equivalently an optimum value of θ_m) in the sense of minimizing $P\left(E\right)$ of Eq. (8).

Evaluation of Eq. (8) can be performed by direct numerical integration (e.g., Riemann sum, Simpson's rule, etc.) or by employing a form of Gauss-quadrature integration, namely, Gauss–Chebyshev integration [3]. In the case of the latter, making the change of variables $x = \cos\phi$ in Eq. (8), we obtain

$$P(E) = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} \overbrace{\operatorname{erfc}\left(\sqrt{(1-m^2)R_t}x\right) \frac{\exp\left(m^2 R_t \delta\, x\right)}{2\pi I_0\left(m^2 R_t \delta\right)}}^{f(x)} dx \tag{9}$$

Using the Gauss–Chebyshev formula,

$$\left.\begin{aligned} \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\, dx &\cong \frac{\pi}{N} \sum_{k=1}^{N} f(x_k) \\ x_k &\triangleq \cos\left(\frac{(2k-1)\pi}{2N}\right) \end{aligned}\right\} \tag{10}$$

where N is a number typically much smaller than the number of points needed for numerical integration with uniform increments, we can compute Eq. (9) from

$$P(E) \cong \frac{1}{2N I_0\left(m^2 R_t \delta\right)} \sum_{k=1}^{N} \operatorname{erfc}\left(\sqrt{(1-m^2)R_t}x_k\right) \exp\left(m^2 R_t \delta\, x_k\right) \tag{11}$$

Figure 1 numerically illustrates this result by plotting $P(E)$ versus m^2 for several values of δ with R_t (in dB) as a parameter. (These curves are similar to those presented in [1].) Figure 2 illustrates the corresponding optimized values of m^2.

III. Suppressed-Carrier System Model

For a suppressed-carrier system, the transmitted signal is still given by Eq. (1) but with $\theta_m = 90$ deg. Thus, from Eq. (2), $m^2 = 0$ and hence the signal simplifies to

$$s(t) = \sqrt{2P_d}d(t)\cos\omega_0 t = \sqrt{2P_t}d(t)\cos\omega_0 t \tag{12}$$

At the receiver, a Costas loop tracks the suppressed-carrier signal. Again assuming uncoded data and matched-filter detection, the error probability of the receiver, conditioned on a phase error ϕ in the Costas loop, is still given by Eq. (4); however, the pdf of the loop phase error is now given by [2]

$$p(\phi) = \frac{\exp\left(\frac{\rho_{\text{Costas}}}{4}\cos 2\phi\right)}{2\pi I_0\left(\frac{\rho_{\text{Costas}}}{4}\right)}, \quad |\phi| \le \pi \tag{13}$$

where

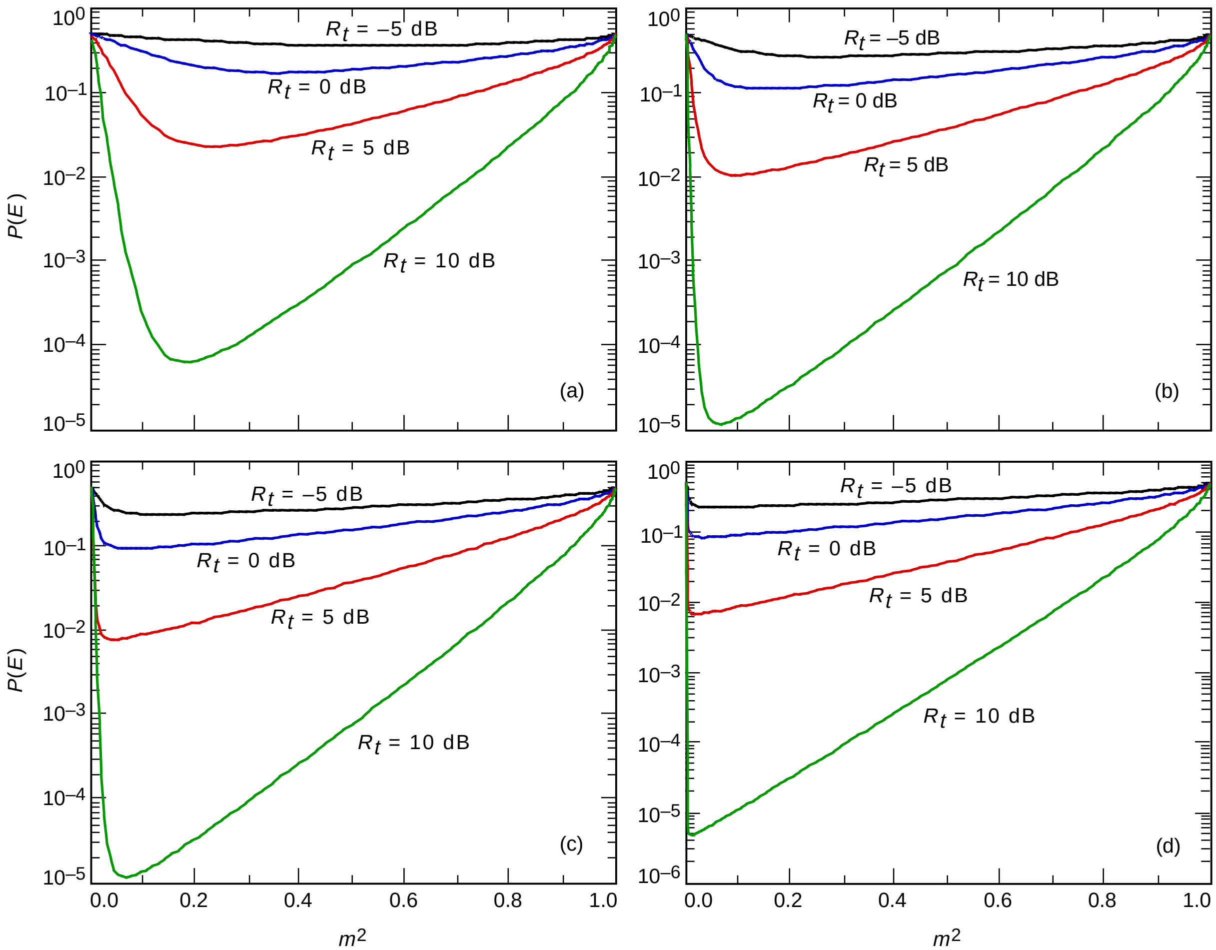

Fig. 1. Probability of error versus m^2 for (a) δ = 10, (b) δ = 50, (c) δ = 200, and (d) δ = 1000.

$$\rho_{\text{Costas}} = \frac{P_t}{N_0 B_L} S_L = \frac{P_t T}{N_0} \frac{1}{B_L T} S_L = R_t \delta S_L \tag{14}$$

is the loop SNR and S_L is the squaring loss, which for a Costas loop with integrate-and-dump (I&D) arm filters is given by

$$S_L = \frac{2R_t}{1 + 2R_t} \tag{15}$$

From Eq. (14) we observe, as is well-known, that the Costas loop exhibits a 180-deg phase ambiguity in that it is as likely to lock at $\phi = 0$ deg as it is at $\phi = 180$ deg. Assuming that this ambiguity can be perfectly resolved, then the average error probability performance of the system is obtained by averaging the conditional error probability of Eq. (4) over the pdf in Eq. (13) folded into the interval $-\pi/2 \leq \phi \leq \pi/2$, which yields

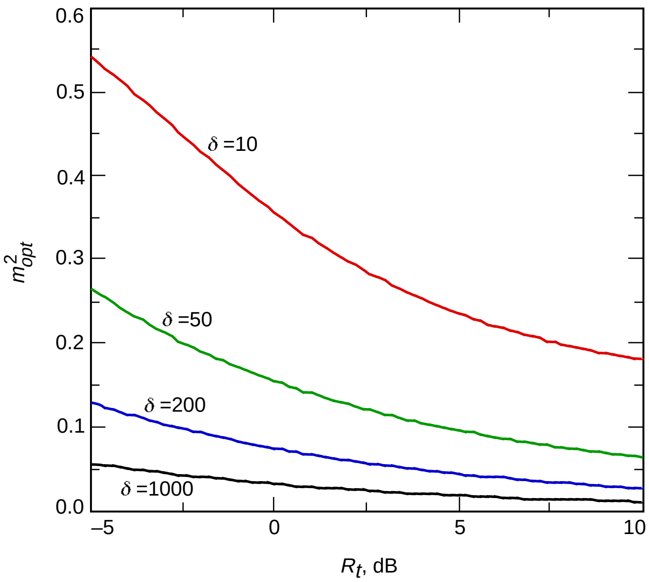

Fig. 2. Optimized values of m^2 for various δ.

$$P(E) = \int_{-\pi/2}^{\pi/2} P(E;\phi)\, p(\phi)\, d\phi$$

$$= \int_{-\pi/2}^{\pi/2} \operatorname{erfc}\left(\sqrt{R_t}\cos\phi\right)\left[2\frac{\exp\left(\frac{R_t\delta}{4}\cos 2\phi\right)}{2\pi I_0\left(\frac{R_t\delta}{4}\right)}\right] d\phi$$

$$= 4\int_0^{\pi/2} \frac{1}{2}\operatorname{erfc}\left(\sqrt{R_t}\cos\phi\right)\left[2\frac{\exp\left(\frac{R_t\delta}{4}\cos 2\phi\right)}{2\pi I_0\left(\frac{R_t\delta}{4}\right)}\right] d\phi \tag{16}$$

Once again, Eq. (16) can be evaluated by direct numerical integration or by employing Gauss–Chebyshev integration. For the latter, we again make the change of variables $x = \cos\phi$ in Eq. (16) to obtain

$$P(E) = 2\int_0^1 \frac{1}{\sqrt{1-x^2}} \overbrace{\operatorname{erfc}\left(\sqrt{R_t}x\right)\frac{\exp\left(\frac{R_t\delta}{4}\left(2x^2-1\right)\right)}{2\pi I_0\left(\frac{R_t\delta}{4}\right)}}^{g(x)} dx \tag{17}$$

Defining the even function

$$f(x) = \begin{cases} g(x) & 0 \le x \le 1 \\ g(-x) & -1 \le x \le 0 \end{cases} \tag{18}$$

then Eq. (17) can be rewritten as

$$P(E) = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} \overbrace{\operatorname{erfc}\left(\sqrt{R_t}\,|x|\right) \frac{\exp\left(\dfrac{R_t\delta}{4}\,\left(2x^2-1\right)\right)}{2\pi I_0\left(\dfrac{R_t\delta}{4}\right)}}^{f(x)} dx \tag{19}$$

Using the Gauss–Chebyshev formula of Eq. (10), we obtain

$$P(E) \cong \frac{1}{2NI_0\left(\dfrac{R_t\delta}{4}\right)} \sum_{k=1}^{N} \operatorname{erfc}\left(\sqrt{R_t}\,|x_k|\right) \exp\left(\frac{R_t\delta}{4}\,\left(2x_k^2-1\right)\right) \tag{20}$$

where x_k is defined as in Eq. (10).

IV. Performance Comparison

What is interesting is to compare the minimum average error probability of the residual carrier system (i.e., the best design) as obtained by minimizing Eq. (8) [or Eq. (11)] with respect to m^2 with the error probability of the suppressed-carrier system obtained from Eq. (17) [or Eq. (20)] for the same values of R_t and δ. This implies a comparison of the two systems for equal total available power, loop bandwidth, and data rate. Figure 3 is a plot of $P(E)$ versus R_t in dB for two values of δ, where for the residual carrier case the minimum value of $P(E)$ as discussed above is used. We observe from this figure that over the wide range of values of R_t and δ considered, *the Costas loop always results in a smaller average error probability*! Figure 4 is a plot of the corresponding optimum values of m^2 (those that produce the minimum average error probability for the residual carrier case) versus R_t in dB. We observe that as R_t and δ increase, the optimum value of m^2 approaches zero, suggesting the use of a suppressed-carrier system.

The curves in Fig. 3 are an illustration of the results in Eq. (8) [or Eq. (11)] for the residual carrier system and Eq. (17) [or Eq. (20)] for the suppressed-carrier system without regard to whether the loop SNR in each case is sufficiently high for the loop (PLL or Costas) to remain in lock. Since, in practice, a residual versus suppressed carrier comparison only has significance in the range of system parameters where the loops can indeed remain in lock (i.e., at or above a threshold value of loop SNR), we must now restrict the above results to the regions where this constraint is satisfied. For a PLL, it is reasonable to assume that the loop is locked (i.e., the cycle slip rate is sufficiently small) when $\rho_{PLL} \geq 7$ dB. Since, as seen above, the Costas loop tracks a doubled phase error process, the loop SNR needed to maintain lock must be, at the least, 6-dB higher than the PLL [compare Eqs. (5) and (13)]. In addition, since the loop S-curve for the Costas loop is of the form $\sin 2\phi$, whereas that for the PLL is of the form $\sin\phi$, then the effective linear region of loop operation is half as wide for the former as it is for the latter. Thus, an additional 3 dB should be required of the Costas loop threshold SNR. Taking these facts into account, it is reasonable to assume that a Costas loop is locked when $\rho_{\text{Costas}} \geq 16$ dB. It is important to note that these threshold values of loop SNR for the PLL and Costas loop are not hard and fast numbers in that the failure of a loop to remain in lock is somewhat of a soft phenomenon. However, by choosing 7 dB and 16 dB as typical of the loop thresholds, we are able to portray a simple graphical illustration of the system parameter regions that are useful to the system engineer in deciding between residual and suppressed-carrier designs.

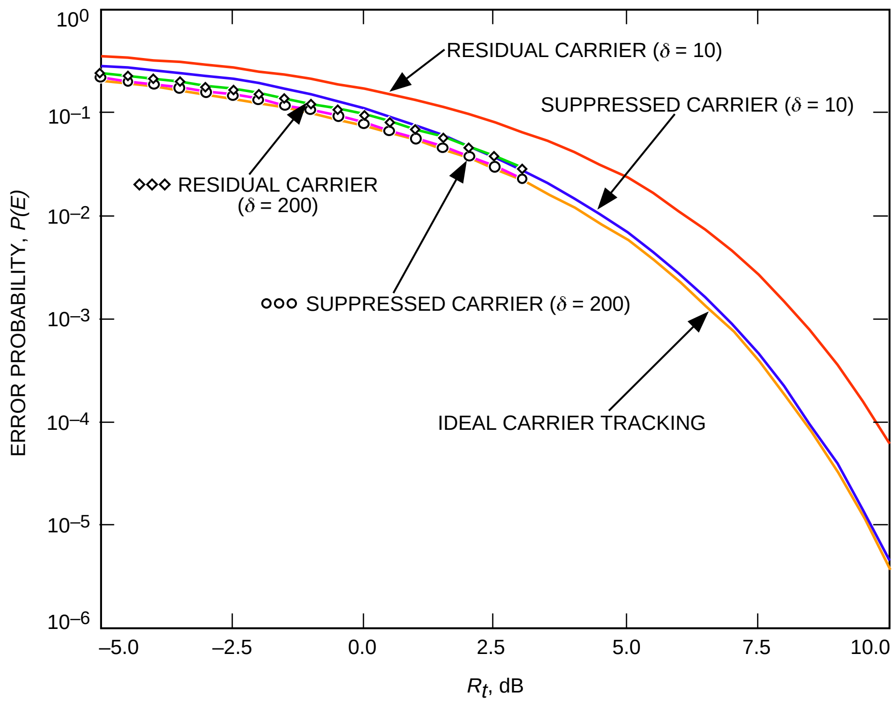

Fig. 3. Error probability performance of residual and suppressed-carrier systems.

If in Eq. (14) we set $\rho_{\text{Costas}} = 10^{1.6} = 39.81$ (corresponding to 16 dB), then using Eq. (15), we arrive at the following:

$$R_t \delta \frac{2R_t}{1 + 2R_t} = 39.81 \tag{21}$$

which is quadratic in R_t and has the solution

$$R_t = \frac{39.81}{2\delta} + \sqrt{\left(\frac{39.81}{2\delta}\right)^2 + \frac{39.81}{2\delta}} \tag{22}$$

Figure 4 is a plot of R_t versus δ in accordance with Eq. (22). The region above the top curve represents the range of system parameters where the Costas loop SNR is sufficient to track, and thus in this region, one would always employ a suppressed-carrier system. Note that this critical curve can be obtained without resorting to any error probability calculations.

For the residual carrier system, we obtain an analogous critical curve by taking the locus of points of minimum error probability obtained from Fig. 1 with the further constraint that $\rho_{PLL} = m_{opt}^2 R_t \delta$ $= 10^{0.7} = 5.01$ (corresponding to 7 dB). The curve so obtained is also illustrated in Fig. 4. The region between the two curves represents the range of system parameters where the Costas loop is below threshold, i.e., it will not maintain lock, but the PLL is above threshold, i.e., it will maintain lock. Thus, in this region, one would employ a residual carrier system. The region below the lower curve represents the range of system parameters where both the PLL and the Costas loops are below threshold and, thus, one should design the system to operate in this region.

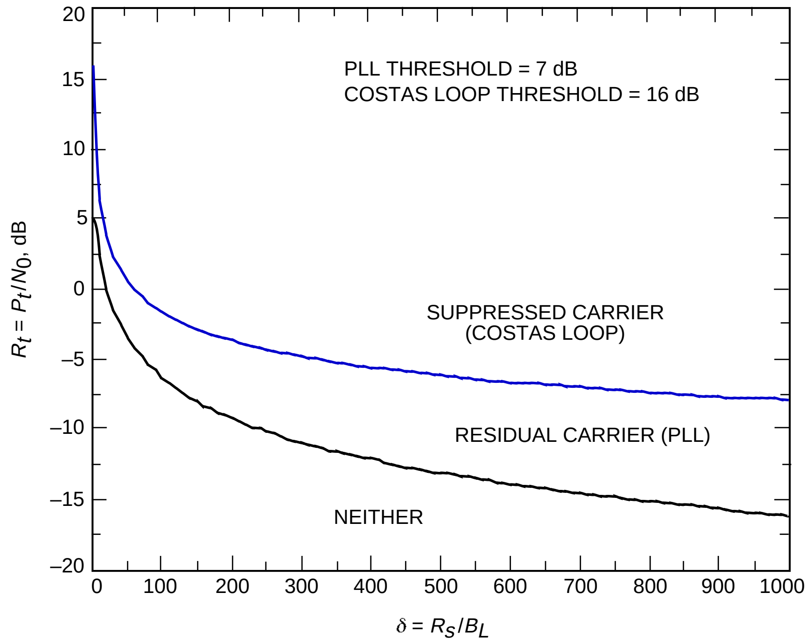

Fig. 4. Preferred regions of operation for residual and suppressed-carrier systems.

V. Conclusion

It has been shown that, for coherent communication systems that employ either a phase-locked loop (PLL) or a Costas loop for carrier synchronization, whenever the loop can indeed be locked, the system should be designed as a suppressed-carrier system. This conclusion is reached based on the desire to minimize the average error probability of the system.

References

[1] W. C. Lindsey, "Optimum Design of One-Way and Two-Way Coherent Communication Links," *IEEE Transactions on Communications*, vol. COM-14, no. 4, pp. 418–431, August 1966.

[2] W. C. Lindsey and M. K. Simon, "Optimum Performance of Suppressed Carrier Receivers With Costas Loop Tracking," *IEEE Transactions on Communications*, vol. COM-25, no. 2, pp. 2l5–227, February l977.

[3] A. H. Stroud and D. Secrest, *Gaussian Quadrature Formulas*, Englewood Cliffs, New Jersey: Prentice-Hall, 1966.