Carrier Synchronization for Low Signal-to-Noise Ratio Binary Phase-Shift-Keyed Modulated Signals

V. Vilnrotter,[1] A. Gray,[1] and C. Lee[1]

A novel information-reduced carrier-synchronization system has been evaluated by means of both analysis and simulation. The idea is to use preliminary estimates of the received binary symbols to reduce the number of transitions before attempting to track the carrier phase, in order to minimize the effects of squaring loss on loop performance. It is shown that this technique indeed yields significantly improved phase estimates with coded input data, even when simple orthogonal block codes are employed. Reductions in squaring loss amounting to as much as 5 dB and more have been demonstrated with this technique in the low signal-to-noise ratio (SNR) regions corresponding to symbol SNR of −8 to −5 dB.

I. Introduction

Optimum closed-loop phase estimators for modulated and unmodulated carriers have been derived in the past based on maximum a posteriori (MAP) criteria [1,2] and have been shown to take the form of feedback loops that attempt to continuously null the difference between the phase of the received signal and its estimate, produced by a controllable oscillator within the loop. It is well-known that at a high loop signal-to-noise ratio (SNR) the variance of the phase estimate varies inversely as the loop SNR for unmodulated carriers, but that for modulated carriers such as binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK), etc., the variance of the phase error increases due to the inherent multiplication operations needed to generate the error signal; this additional degradation is generally referred to as "squaring loss." Here we shall concentrate on simple binary-modulated carriers (BPSK), but the results can be directly extended to QPSK or even higher-order modulations, where squaring losses are more severe.

It has been shown that phase estimation loops for BPSK symbols perform better when the data transitions are unequal [3,4], so that on the average one of the binary data values occurs more often than the other, reducing the number of times data transitions occur at the input. In fact, as the data transitions become less frequent, the structure and performance of the MAP phase estimator approaches that of the phase-locked loop, which operates without any squaring loss. However, reducing the average number of data transitions also reduces the information throughput of the system, which is incompatible with communications requirements that typically require maximum information throughput. An idea for reducing or even eliminating squaring losses in a large class of phase estimators without reducing the

[1] Communications Systems and Research Section.

information throughput of the communications system has been proposed in [4], where it was shown that data transitions at the input to the phase estimator could indeed be reduced (in order to combat squaring losses) without compromising communications performance.

The idea is to estimate the data modulation and attempt to remove it from the carrier before it reaches the phase estimator. Even with imperfect data estimation, some of the original data modulation will be removed, leaving only a residual error sequence modulation on the carrier, which necessarily has a lower transition rate than the original signal. With good estimation, transitions in the error sequence can be greatly reduced, resulting in virtual elimination of squaring loss. These concepts are illustrated in the system block diagram of Fig. 1, which shows the main components needed to accomplish this concept, namely a MAP phase estimator, a symbol-sequence estimator, a delay block (to account for delays introduced by the symbol-sequence estimator), and a multiplier to remove the data from the carrier. In the simulation, estimates of the residual error rate, $\hat{p}$, are obtained by comparing the input symbols with their estimates. An exact mathematical description of the waveforms at various points in the block diagram follows.

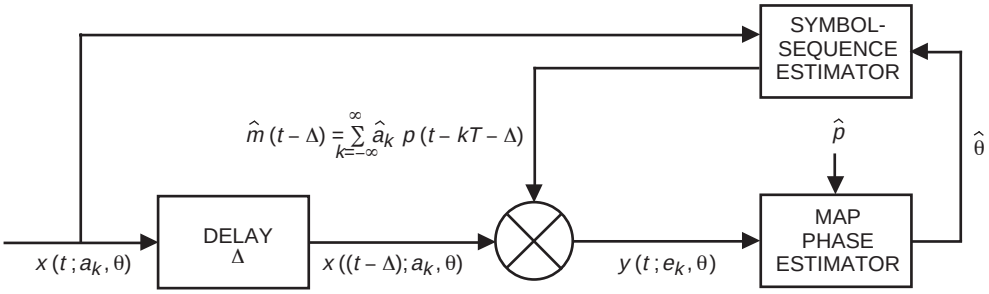

Fig. 1. Block diagram of the information-reduced carrier-phase estimation concept.

II. Mathematical Model of the Information-Reduced MAP Phase Estimator

Let the received signal be of the form

$$s(t;\theta) = \sqrt{2P}m(t)\sin\left(\omega_c t + \theta(t)\right) \tag{1}$$

as in [4], and let the additive noise be described by its narrowband representation,

$$n(t) = \sqrt{2}\left[N_c(t)\cos\left(\omega_c t + \theta(t)\right) - N_s(t)\sin\left(\omega_c t + \theta(t)\right)\right] \tag{2}$$

where $N_c(t)$ and $N_s(t)$ are lowpass Gaussian processes, each with single-sided power spectral density N_0 W/Hz; $\theta(t)$ is a slowly varying phase process; ω_c is the carrier's radian frequency; and P is the carrier power. The signal is modulated by the random pulse train

$$m(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT) \tag{3}$$

where $p(t)$ is a unit power pulse and $\{a_k\}$ are equiprobable binary random variables, $a_k = \pm 1$. The received signal-plus-noise waveform is denoted $x(t; a_k, \theta) = s(t;\theta) + n(t)$; a delayed version of this waveform is multiplied by the sequence of symbol estimates $\{\hat{a}_k\}$, $\hat{a}_k = \pm 1$, to obtain the error sequence $\{e_k\}$,

where $e_k = a_k \hat{a}_k$. If we denote the probability of an incorrect symbol estimate by p, it is clear that $\Pr\{e_k = -1\} = p$ and, hence, $\Pr\{e_k = 1\} = 1 - p$. Thus, if the symbols are estimated accurately, the error sequence consists almost entirely of the number +1.

After the multiplication operation, the resulting waveform, $y(t; e_k, \theta)$, is applied to the input of the phase estimator:

$$y(t; e_k, \theta) = \sqrt{2P} e(t) \sin(\omega_c t + \theta(t)) + N(t) \tag{4}$$

where

$$e(t) = \sum_{k=-\infty}^{\infty} e_k p(t - kT) \tag{5}$$

and $N(t) = \hat{m}(t) n(t)$ with $\hat{m}(t) = \sum_{k=-\infty}^{\infty} \hat{a}_k p(t - kT), \hat{a}_k = \pm 1$. The important point here is that, by attempting to estimate the data modulation and subsequently multiplying the received waveform by the symbol estimates, a new waveform has been created that has advantages in terms of synchronization performance, namely much fewer transitions.

The structure of the MAP phase estimator derived in [4] was based on the model of the input signal as described in Eq. (4), where the binary random variables defining the error sequence were independent and took on the favorable value +1 more often than the unfavorable value −1, although the equal probability case also was included as one of the endpoints for completeness. This model leads to the MAP estimator structure shown in Fig. 2, which differs from the MAP estimator for equiprobable data only in the form of the nonlinearity in the in-phase arm:

$$f_p(x) = \tanh\left(x - \frac{1}{2} \ln \frac{p}{1-p}\right) \tag{6}$$

Note that as p approaches 0.5, this nonlinearity reduces to $\tanh(x)$, as expected. An unavoidable consequence of unequal transition probabilities is that such probabilities are not unique, hence must be estimated and supplied to the phase estimator in real time in order to adjust the nonlinearity for optimal operation.

III. MAP Phase-Estimator Performance with Unbalanced Data

The MAP phase estimator for unbalanced data shown in Fig. 2 has been implemented in software (i.e., simulated) using the Signal Processing Workstation (SPW). Since this is a digital simulation that processes samples of the waveforms at discrete times, the analog model developed above does not strictly apply; however, the highest-frequency components in the sampling operation were sufficiently oversampled to ensure a close correspondence between the analog and digital models. In the simulation, there were 5 complete cycles of the carrier frequency in each symbol interval, and each cycle was sampled 5 times, so that each binary symbol was actually sampled 25 times. That means the carrier waveform was *oversampled* by a factor of 2.5 (since only two samples per cycle are required to establish Nyquist sampling), which was deemed sufficient to allow the use of the simpler analog notation to represent the digital model.

The MAP estimator for phase takes the form of an in-phase and quadrature-phase loop with a nonlinearity in the in-phase arm, as in Fig. 2. The phase estimator operates on the preprocessed signal-plus-noise

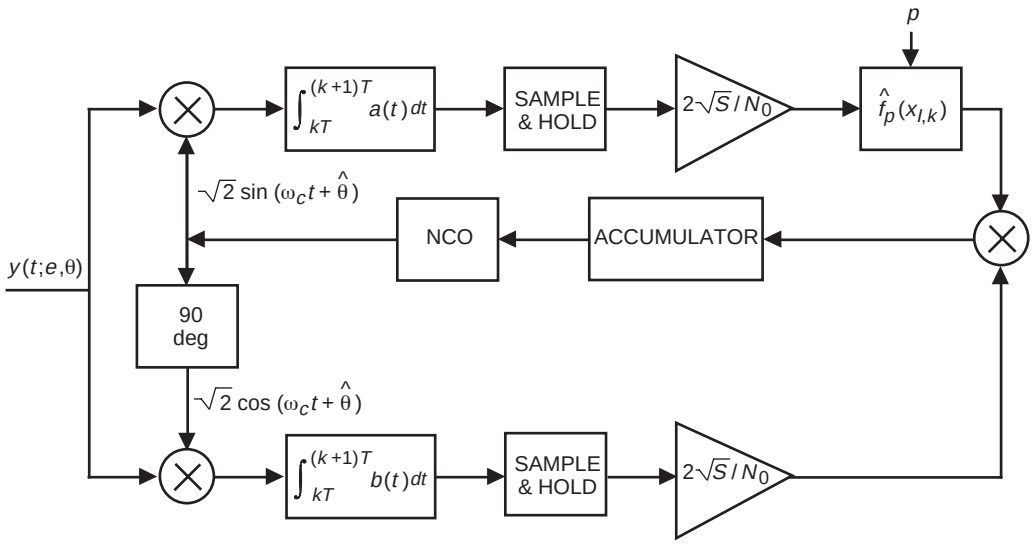

Fig. 2. The information-reduced carrier-synchronization loop.

waveform $y(t; e_k, \theta)$; it performs complex baseband downconversion using its internal numerically controlled oscillator (NCO), whose phase is continuously adjusted to the loop's estimate of the received phase, $\hat{\theta}$. It is assumed that symbol synchronization has been established, so that integration boundaries correspond to the true boundaries of the data symbols. The downconverted and noise-corrupted waveforms are integrated over each symbol interval to produce a single random variable whose value is held fixed throughout the next symbol interval. The loop samples in each arm are scaled by a factor proportional to the ratio of signal magnitude and noise power; the scaled in-phase samples, $x_{I,k}$, are applied to the nonlinear function to produce the value $f_p(x_{I,k})$, which then is multiplied by the corresponding quadrature samples, $x_{Q,k}$. This multiplication produces an error signal proportional to the phase difference between the local and received signals, thus providing a means to adjust the phase of the local oscillator (or NCO) to better correspond with that of the received signal. However, the multiplication operation also creates signal and noise cross-product terms, giving rise to squaring loss that degrades the performance of the loop at low SNRs. The product samples are further averaged by the accumulator to reduce random variations before this error signal is applied to the NCO, thus closing the loop.

After locking up to the phase of the received signal, the static performance of a phase-locked loop typically is described in terms of the variance of the phase error it produces, $\varphi(t) = \theta(t) - \hat{\theta}(t)$, at a fixed-loop bandwidth. The variance of the phase error, σ_ϕ^2, can be expressed in terms of the symbol SNR, S/N_0, and the squaring loss, $\mathcal{S}_L$, as

$$\sigma_\phi^2 = \frac{N_0 B_L}{S\,\mathcal{S}_L} \tag{7}$$

where, by convention, the squaring loss is a number between zero and one, taking on its maximum value only when there is no squaring loss within the loop (as is the case for a conventional phase-locked loop (PLL) tracking an unmodulated tone).

The performance of the simulated MAP phase estimator was tested by comparing it with the theoretical performance curves presented in [4], which show the reduction in squaring loss with decreasing transition probability. Since there is very little squaring loss above 0 dB, most of the simulations were carried out at much lower values, namely between −10 and −3 dB. Although not implicit in the calculations, special care had to be exercised to ensure that the loop bandwidth remained constant as the symbol SNR and

transition probabilities were changed; all simulations were carried out with a loop bandwidth of 5 Hz. The close agreement between simulation and calculated results is apparent in Fig. 3, verifying correct operation of the simulated system.

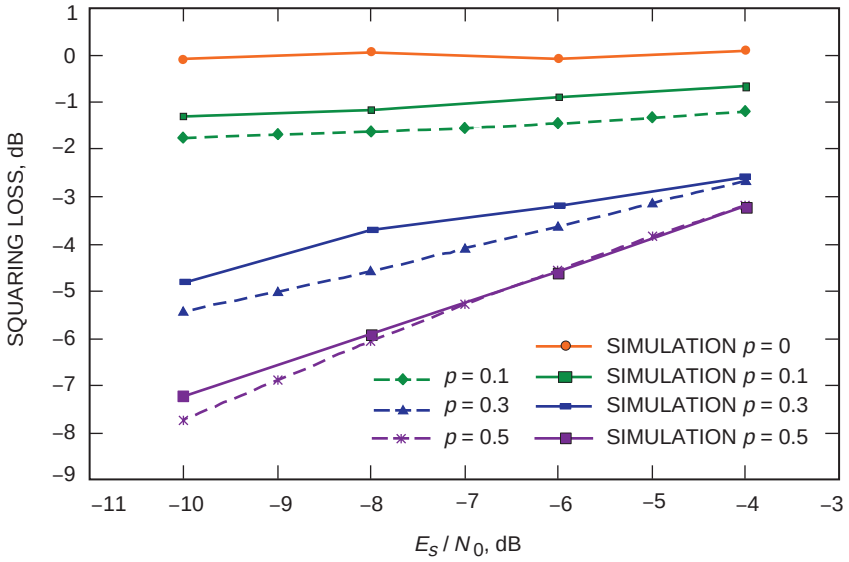

Fig. 3. Comparison of calculated and simulated squaring-loss results.

A. The Symbol-Sequence Estimator

The data modulation used for testing the MAP phase estimator in the previous section consisted of independent binary data symbols selected with predetermined a priori probabilities within the simulation. This allowed for testing the phase estimator at various transition probabilities ranging from $p = 0.5$ (the equilikely case) to $p = 0$ (the unmodulated carrier case). The fact that the data were independent from symbol to symbol is important because this corresponds to the condition under which the MAP structure was derived in [4]. However, in realistic communications scenarios, the phase estimator is not likely to encounter binary modulation with unequal a priori probabilities, because that typically would result in inefficient communication; whether the data are coded or uncoded, the resulting binary symbols observed by the receiver tend to be equilikely symbols, with transition probabilities close to one-half. For the purposes of phase estimation, this corresponds to the worst-case scenario, as derived analytically in [4] and verified with simulation here.

Since the received data symbols tend to be equiprobable in most cases of interest, we shall henceforth assume them to be equiprobable; that means that, in order to reduce the number of transitions due to modulation in the received waveform, some form of data estimation will be required in order to generate initial estimates of the data sequence, followed by a multiplication operation to partially remove the data from the received waveform, thus reducing the average number of transitions observed by the phase estimator. If this can be accomplished, then significant improvements in the phase estimates should be possible even though the symbol estimates may not be good enough to be of any real value in communications applications.

Preliminary estimates to be used for removing data modulation are obtained from the symbol-sequence estimator block shown in Fig. 1. We can immediately distinguish between two distinct cases likely to be encountered in applications, namely uncoded and coded BPSK. The structure of the symbol-sequence estimator block, as well as its performance, generally will depend on the type of coding employed, although robust structures that may work well for several different codes also can be envisioned. First, we shall consider uncoded BPSK data, followed by examples of symbol-sequence estimation with orthogonal block codes.

B. Single-Symbol Detection; Uncoded BPSK Data Modulation

For the case of uncoded binary symbols, the received signal is assumed to be a BPSK modulated carrier (or IF signal) with power P, observed in the presence of additive white Gaussian noise as in Eq. (1). Assuming that we have an estimate of the carrier phase, $\hat{\theta}$, we can obtain an estimate of the data by first demodulating the received waveform, followed by integration and polarity detection, as shown in Fig. 4.

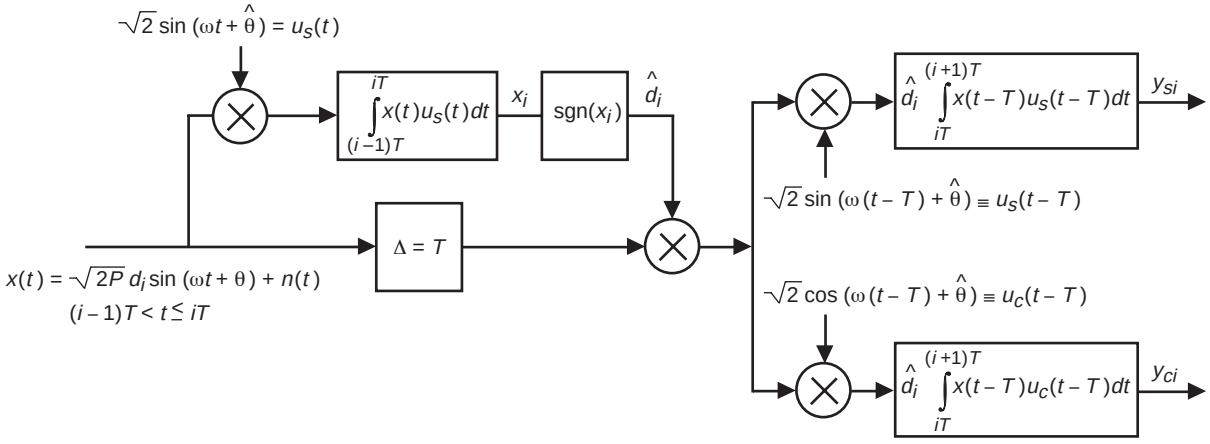

Fig. 4. The symbol-sequence estimator for uncoded BPSK: single-bit coding.

The output of the integrator, x_i, can expressed as

$$x_i = 2\sqrt{P}d_i\int_{(i-1)T}^{iT}\sin(\omega t+\theta)\sin\left(\omega t+\hat{\theta}\right)dt+\int_{(i-1)T}^{iT} n(t)\sin\left(\omega t+\hat{\theta}\right)dt \tag{8}$$

which can be rewritten more compactly as

$$x_i = T\sqrt{P}d_i\cos(\phi)+N_{si};\quad \phi\equiv\theta-\hat{\theta} \tag{9}$$

(where we assumed that integration effectively removes the double frequency terms). The output of the polarity detector is $+1$ if x_i is positive and -1 if x_i is negative; hence, we can write

$$\hat{d}_i = \text{sgn}(x_i) \tag{10}$$

In the in-phase arm, the received waveform is delayed by T seconds, demodulated using a delayed version of the local oscillator signal, multiplied by the data estimate, and integrated to obtain

$$y_{si} = 2\sqrt{P}d_i\hat{d}_i\int_{iT}^{(i+1)T}\sin(\omega(t-T)+\theta)\sin(\omega(t-T)+\hat{\theta})dt+\hat{d}_i\int_{iT}^{(i+1)T} n(t-T)\sin(\omega(t-T)+\hat{\theta})dt$$

$$=\hat{d}_i\left[T\sqrt{P}d_i\cos(\phi)+N_{si}\right]=\hat{d}_i x_i = x_i\ \text{sgn}(x_i)\equiv\alpha_i>0 \tag{11}$$

Note that α_i is always greater than zero, an observation that follows from the fact that the integral in the preprocessor and the integral of the delayed waveform in the in-phase arm are identical. This

implies that, with uncoded BPSK modulation and matched-filter detection, the in-phase arm cannot provide any useful information to help resolve the polarity of the signal, or the sign or magnitude of the phase error, particularly for small phase errors, since $\cos(\phi) \underset{\phi \to 0}{\longrightarrow} 1$. This is not inconsistent with the low SNR assumption, since, in the absence of phase dynamics, arbitrarily small phase-error variance can be obtained by integrating long enough or, equivalently, by operating with a narrow enough loop bandwidth.

Proceeding as above, the delayed waveform in the quadrature arm is demodulated using a delayed quadrature local signal to obtain

$$y_{ci} = 2\sqrt{P} d_i \hat{d}_i \int_{iT}^{(i+1)T} \sin(\omega(t-T)+\theta) \cos\left(\omega(t-T)+\hat{\theta}\right) dt + \hat{d}_i \int_{iT}^{(i+1)T} n(t-T) \cos(\omega(t-T)+\hat{\theta}) dt$$

$$= \hat{d}_i \left[T\sqrt{P} d_i \sin(\phi) + N_{ci}^* \right] = \hat{d}_i d_i T\sqrt{P} \sin(\phi) + N_{ci} \tag{12}$$

where $N_{ci} = \hat{d}_i N_{ci}^*$. For small ϕ, we can use the approximation $\sin(\phi) \cong \phi$; hence, from Eq. (12), it follows that the quadrature-arm samples, y_{ci}, provide the necessary error signal for closing the loop. Thus, with uncoded BPSK modulation and matched-filter detection, the in-phase arm can be ignored altogether when estimating phase error, since it provides no additional information.

The output of the quadrature-arm integrator turns out to be a scaled estimate of phase error, even after normalization to cancel the signal-power and symbol-duration terms. This can be explained using the following argument: define the phase estimate for small phase errors as

$$\hat{\phi} \cong \frac{y_{ci}}{T\sqrt{P}} \tag{13}$$

Clearly, for constant input phase and correctly decoded data, $E(\hat{\phi}) = \phi$ and $\mathrm{var}(\phi) = \sigma^2/T^2P$, where σ^2 is the variance of the noise samples, N_{ci}. As long as the estimate of the symbol polarity is correct, that is, as long as $d_i\hat{d}_i = 1$, the amplitude of the signal term in Eq. (12) remains positive; hence, the mean of the integrator's output, scaled as in Eq. (13), is a good estimate of the instantaneous phase error.

However, if the polarity of the estimate is incorrect, then the amplitude of the signal term changes sign and effectively cancels one of the correctly decoded symbols. Therefore, two symbols are annihilated whenever $d_i\hat{d}_i = -1$, subtracting two symbols from the total number available for phase estimation in a given time interval; this, in turn, introduces a scale factor into the phase estimate that must be taken into account if constant loop bandwidth is to be maintained. Recalling that the probability of symbol error for binary antipodal signals is $p = Q\left(\sqrt{2E_s/N_0}\right)$, the expected value of the time-averaged signal in the quadrature arm with imperfect data estimates becomes

$$E(y_{ci}) = (1-2p)T\sqrt{P}\sin(\phi) \tag{14}$$

so that, for small ϕ, the phase-error estimate is effectively scaled due to the cancellation:

$$E\left(\hat{\phi}\right) = \frac{E(y_{ci})}{T\sqrt{P}} \underset{\phi \to 0}{\longrightarrow} (1-2p)\phi \tag{15}$$

This means that a scaling factor must be applied to the quadrature-arm samples in order to obtain an accurate estimate of the phase error. Denoting this rescaled phase-error estimate as $\hat{\phi}^*$, we obtain

$$\hat{\phi}^* = \frac{\hat{\phi}}{(1-2p)} = \frac{y_{ci}}{T\sqrt{P}(1-2p)} \tag{16}$$

which yields $E(\hat{\phi}^*) = \phi$. Motivated by the discussion leading up to Eq. (16), a suboptimal loop structure for unbalanced BPSK symbols is shown in Fig. 5. This structure is very similar to the quadrature arm of the MAP phase estimator, except for the scaling factor of $1/(1-2\hat{p})$, which is necessary to keep the loop bandwidth constant in the presence of changing transition probability. Note that the suboptimal structure requires real-time estimates of transition probability; however, unlike the MAP phase estimator, it does not require real-time estimates of SNR.

We observe that, since $(1-2p) < 1$, this implies that the variance of the phase-error estimate in the suboptimal loop increases by a factor of $(1-2p)^2$ whenever the transition probability is greater than zero, which is exactly the squaring loss associated with a polarity-type Costas loop [4]. Thus, for the case of a matched-filter estimator for individual BPSK symbols, no advantage can be gained over the polarity-type Costas loop by attempting to strip the data from the received waveform prior to phase estimation.

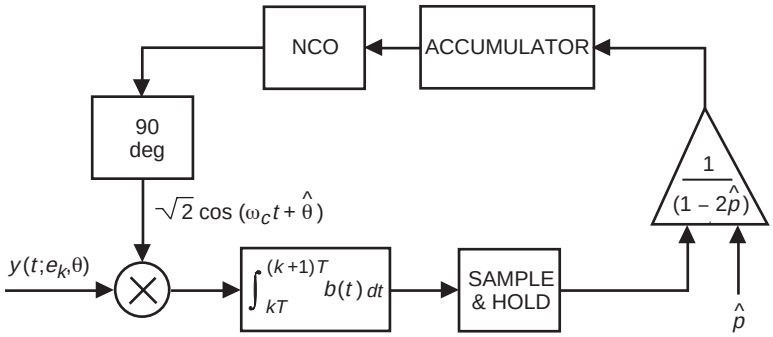

Fig. 5. The suboptimum loop for BPSK symbols with reduced transition probabilities.

C. Vector Symbol Estimation; Uncoded BPSK Modulation

Having shown that no advantage can be gained over the polarity-type Costas loop by making individual decisions on uncoded BPSK symbols, the next question is: Can there be an improvement in phase-estimation performance over a polarity-type Costas loop if entire vectors of BPSK data are decoded at the same time instead of individual symbols? We now proceed to address this question.

Consider a sequence of K consecutive uncoded BPSK symbols with an information content of K bits, decoded all at once by correlating the received sequence with all 2^K possible binary realizations and selecting the vector corresponding to the greatest correlation value; this is the maximum-likelihood strategy for decoding vectors in the presence of additive Gaussian noise.

The structure of the vector symbol estimator is shown in Fig. 6; it is a bank of correlators, which also could be implemented equivalently as a bank of matched filters. Note that, when correlating against vectors of length K, the delay must be increased to KT, and the received sequence must be multiplied by each of the test vectors, $v_i(t)$ (representing all possible realizations of K binary symbols), before performing the integration operation.

In this application, we are trying to select a sequence that is close to the transmitted sequence in Hamming distance, but it need not be exactly the same; in a large vector, a few symbols could be decoded incorrectly without seriously impacting the effectiveness of the data-removal operation. Thus, unlike in communications applications, there is no great penalty for being in error by one or more symbols

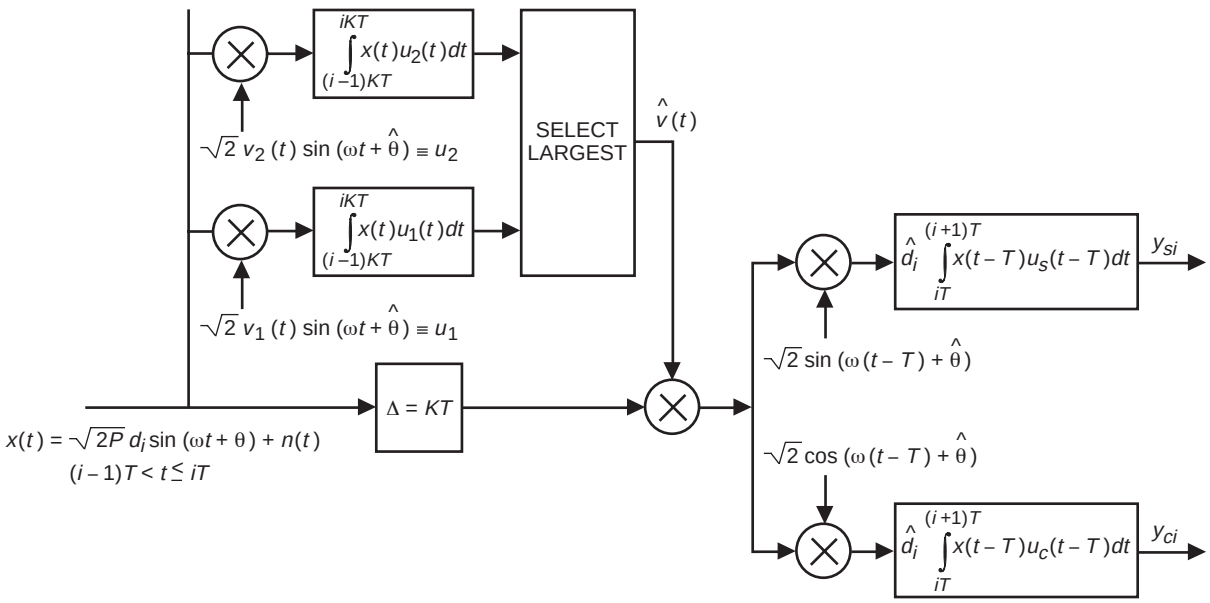

Fig. 6. The symbol-sequence estimator for BPSK vectors.

per codeword, provided the total number of errors remains much smaller than the length of the codeword K. With d_H denoting the Hamming distance, and designating the conditional probability that the selected test vector is Hamming distance l from the received vector given that the kth source vector was transmitted as $\Pr(d_H = l|\mathbf{s}_k)$, we can define the probability of a good sequence as the probability that the selected vector was within Hamming distance $m << K$ from the received vector, $d_H \leq m$, where $m/K = \alpha$ is a predetermined constant. With $P(gs|\mathbf{s}_k)$ denoting the probability that a good sequence was selected given the kth received vector, it follows that

$$P(gs|\mathbf{s}_k) = \sum_{l=0}^{m} \Pr(d_H = l|\mathbf{s}_k) \tag{17}$$

Consider the K-dimensional vector $\mathbf{s}_k = (s_1, s_2, \cdots, s_K)$ selected from the set of all possible binary vectors $\{\mathbf{s}_k\}$, where each component is an equally probable binary random variable with received magnitude $\pm\sqrt{E_s}$. Following the derivation in [5], we note that there are 2^K distinct signal vectors in this set, with each signal occupying one corner of a K-dimensional hypercube. If we assume that each vector is equally likely to be transmitted, then the probability of receiving a given vector is 2^{-K}. Let us assume that the all-positive vector was transmitted, designated as $\mathbf{s}_1 = (\sqrt{E_s}, \sqrt{E_s}, \cdots, \sqrt{E_s})$. Regardless of which vector was actually transmitted, there are exactly K vectors within a Hamming distance of one from the true received vector, K-taken-two-at-a-time vectors within a Hamming distance of two, and so on.

Assuming that the average phase error is negligibly small, the transmitted signal vectors received in the presence of additive zero-mean Gaussian noise can be represented by the K-component vector $\mathbf{n} = (n_1, n_2, \cdots, n_K)$ with uncorrelated components. The probability density of this vector is Gaussian,

$$p_{\mathbf{n}}(n_1, n_2, \cdots, n_K) = \left(2\pi\sigma^2\right)^{-K/2} \exp\left[-(1/2\sigma^2)\sum\nolimits_{k=1}^{K} n_k^2\right] \tag{18}$$

while the density of the noise-corrupted signal vector $\mathbf{r} = \mathbf{s} + \mathbf{n}$ can be expressed as

$$p_{\mathbf{r}}(r_1, r_2, \cdots, r_K) = p_{\mathbf{n}}(r_1 - s_1, r_2 - s_2, \cdots, r_K - s_K) = \left(2\pi\sigma^2\right)^{-K/2} \exp\left[-(1/2\sigma^2)\sum_{k=1}^{K}(r_k - s_k)^2\right] \tag{19}$$

The decision regions are the generalized quadrants of the K-dimensional space. Since every coordinate of the all-positive vector $\mathbf{s}_1$ is the same, the point in K-dimensional space defined by this vector in the absence of noise is in the first generalized quadrant; there are K nearest neighbors, each with a single negative coordinate (that is, a Hamming distance of one from $\mathbf{s}_1$), etc. Thus, we can determine the probability of selecting a sequence within any given distance of the true received vector by integrating the probability density of Eq. (18) over the appropriate generalized quadrant. Proceeding in this manner, the probability of correct detection can be expressed as

$$\begin{aligned}
\Pr(d_H = 0|\mathbf{s}_1) &= \int_0^\infty dr_1 \int_0^\infty dr_2 \cdots \int_0^\infty dr_K \, p_{\mathbf{n}}(r_1 - s_1, r_2 - s_2, \cdots, r_K - s_K) \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty dr\, e^{-(r-\sqrt{E_s})^2/2\sigma^2}\right)^K \\
&= (1-p)^K
\end{aligned} \tag{20}$$

where p is defined as $p = Q(\sqrt{E_s/\sigma^2})$. A signal at Hamming distance one from the all-plus vector will have a negative sign for exactly one of its coordinates and, hence, will occupy a nearest corner of the hypercube. Since there are K nearest corners, and the integration is always over disjoint regions, it follows that the probability of selecting a vector at Hamming distance one from the received vector is

$$\begin{aligned}
\Pr(d_H = 1|\mathbf{s}_1) &= K\frac{1}{\sqrt{2\pi}\sigma^2}\int_{-\infty}^0 dx\, e^{-(x-\sqrt{E_s})^2/2\sigma^2}\left(\frac{1}{\sqrt{2\pi}\sigma^2}\int_0^\infty dy\, e^{-(y-\sqrt{E_s})^2/2\sigma^2}\right)^{K-1} \\
&= Kp(1-p)^{K-1}
\end{aligned} \tag{21}$$

Proceeding in this manner, the probability of selecting a vector at a Hamming distance of two from $\mathbf{s}_1$ becomes

$$\Pr(d_H = 2|\mathbf{s}_1) = \binom{K}{2} p^2(1-p)^{K-2} \tag{22}$$

while the general term is

$$\Pr(d_H = l|\mathbf{s}_1) = \binom{K}{l} p^l(1-p)^{K-l} \tag{23}$$

The expression in Eq. (23) is recognized as a binomial distribution in Hamming distance, equivalent to the probability of making l errors in K independent binary decisions. Indeed, since each coordinate of the hypercube is equally likely to be positive or negative independent of all other coordinates, and the noise component affecting each coordinate is also independent of all other noise components, it follows that a vector decision is equivalent to a sequence of K independent binary decisions when independent

identically distributed (i.i.d.) BPSK symbols are observed. Therefore, the vector correlator structure of Fig. 6 can be replaced by the much simpler binary sequential estimator for uncoded BPSK symbols shown in Fig. 4.

A direct implication of this equivalence is that the analysis of the single-symbol estimator also applies to the vector estimator when uncoded binary symbols are observed. It follows analogously that the in-phase arm is again irrelevant; hence, the performance of the MAP phase estimator reverts to that of the polarity-type Costas loop.

D. Vector Symbol Estimation; Orthogonal Block-Coded Modulation

Having shown that vector detection of uncoded BPSK symbols is equivalent to symbol-by-symbol decisions and, hence, cannot lead to further improvement over the performance of the simpler structure, the next question is whether the performance of the MAP phase estimator can be improved by using coded sequences of BPSK symbols. A convenient and well-understood modulation format that can be used for resolving this question is orthogonal block-coded modulation, where $\log_2 K$ bits of information are encoded onto orthogonal vectors of length K. Since there are K possible realizations, that means that only K correlators are required in the symbol-sequence estimator. Orthogonal vectors of length K can be generated conveniently from the rows of a Hadamard matrix of order K, designated as $\mathbf{H}_K$. A Hadamard matrix of order K can be constructed from a Hadamard matrix of order $K/2$ using Sylvester's construction, according to which

$$\mathbf{H}_K = \begin{bmatrix} \mathbf{H}_{K/2} & \mathbf{H}_{K/2} \\ \mathbf{H}_{K/2} & -\mathbf{H}_{K/2} \end{bmatrix} \tag{24}$$

Starting with a Hadamard matrix of order 2, Hadamard matrices of order 4, 8, 16, or any multiple of 4 can be generated:

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\mathbf{H}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

$$\mathbf{H}_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

Designating the ith row of a Hadamard matrix as $\mathbf{h}_i$, the ith signal vector can be written

$$\mathbf{v}_i(t) = \mathbf{s}(t)^* \mathbf{h}_i = (v_1(t), v_2(t), \cdots, v_K(t))$$

$$\mathbf{s}(t) = (s_1(t), s_2(t), \cdots, s_K(t))$$

$$\mathbf{h}_i = (h_{i,1}, h_{i,2}, \cdots, h_{i,K}); \quad h_{i,m} = \pm 1$$

$$s_m(t) = T\sqrt{2P}p_m(t)\sin(\omega(t-(m-1)T)+\theta)$$

and $p_m(t) = 1$ if $(m-1)T < t \leq mT$ and zero elsewhere.

Here $*$ refers to vector multiplication, defined as multiplying corresponding components of the two vectors to form a new vector. The correlator output, viewed as a vector, can be obtained by multiplying each received signal with a locally generated vector whose phase corresponds to the latest estimate, integrating to filter out the double frequency term, and forming the inner product with the same Hadamard matrix used by the encoder. Thus, the correlator output vector $\mathbf{c}$ can be expressed as

$$\mathbf{c} = (\mathbf{w}_i + \mathbf{n})\mathbf{H}^{\mathrm{T}}$$

where

$$\mathbf{w}_i = (w_{i,1}, w_{i,2}, \cdots, w_{i,K})$$

$$\mathbf{n} = (n_1, n_2, \cdots, n_K)$$

$$w_{i,m} = \int_{(m-1)T}^{mT} v_i(t)\sqrt{2}\sin\left(\omega(t-(m-1)T+\hat{\theta}\right)dt = T\sqrt{P}\cos(\phi)h_{i,m}$$

$$n_m = \int_{(m-1)T}^{mT} n(t)\sqrt{2}\sin\left(\omega(t-(m-1)T+\hat{\theta}\right)dt$$

The correlator output vector can be decomposed into Hadamard transforms of the signal and noise vectors. If the ith row represents the modulation of the signal, then because of orthogonality, the signal part of the ith component of $\mathbf{c}$ will be equal to $KT\sqrt{P}\cos(\phi)$, while all other components will be exactly zero. Each component of $\mathbf{c}$ will have a zero-mean noise term added to it, independent of the others, with common variance KN_0T. Thus, if the transmitted signal corresponds to the third row of a Hadamard matrix of order 8, the signal component of the correlation vector becomes

$$\mathbf{w}_i\mathbf{H}^{\mathrm{T}} = \left(0, 0, 8T\sqrt{P}\cos(\phi), 0, 0, 0, 0, 0\right)$$

Defining signal-to-noise ratio (or SNR) as the expected value of the square of the signal term divided by the variance of the noise term, it is clear that the SNR of the non-zero signal component is given by

$$\mathrm{SNR} = \frac{K^2T^2P\cos^2(\phi)}{KN_0T} = \frac{KTP\cos^2(\phi)}{N_0} \equiv \frac{KE_s\cos^2(\phi)}{N_0} \tag{25}$$

where signal energy is the time integral of the signal power over a symbol duration, that is, $E_s = TP$. Here we view the phase-error process as slowly varying over a great many codewords, so that its instantaneous value tends to dominate decoder performance. The symbol-sequence estimator determines the location of the largest component of the correlator output and declares the corresponding codeword as its estimate of

the received sequence. Thus, if the third component of **c** is greatest, then the third row of the Hadamard matrix will be selected as the transmitted codeword.

The probability of selecting the correct codeword can be determined using the signal and noise models developed in Eqs. (18) and (19). Conditioned on the phase error, the probability of selecting the correct codeword based on the correlation results, P_c, is simply the probability that a Gaussian random variable with mean value $KE_s\cos^2(\phi)$ and variance N_0 exceeds the remaining $M-1$ zero-mean Gaussian random variables:

$$P_c = \int_{-\infty}^{\infty} \frac{dy}{\sqrt{\pi N_0}} e^{-(y-KE_s\cos^2(\phi))^2/N_0} \left[\int_{-\infty}^{y} \frac{dx}{\sqrt{\pi N_0}} e^{-x^2/N_0} \right]^{(M-1)} \tag{26}$$

For block-coded M-dimensional orthogonal signaling, the probability of a symbol error, P_{se}, can be related to the probability of a correct decision as

$$P_{se} = \frac{M}{2(M-1)}(1-P_c) \tag{27}$$

Note that, since all possible realizations of binary vectors are not included in this code but only an orthogonal subset is used, the codeword estimate made by the symbol-sequence estimator does not reduce to individual binary symbol decisions, as was the case with vector estimates of uncoded BPSK symbols. Therefore, the integral in the in-phase arm is not equivalent to the integral in the preprocessor; hence, Eq. (11) does not apply. In other words, when coded vectors are observed, the in-phase arm of the MAP phase estimator provides additional information that must be used to attain best performance.

Symbol-error probabilities have been computed for the low-SNR region assuming negligible phase error, $\cos^2(\phi) \cong 1$, and the results of the computation are shown in Fig. 7. The theoretical curves for $M = 16, 32, 64$, and 128 indicate that better symbol estimates can be obtained by increasing the length of the codewords, yielding a direct improvement in the data removal—hence, a direct reduction in transition rate.

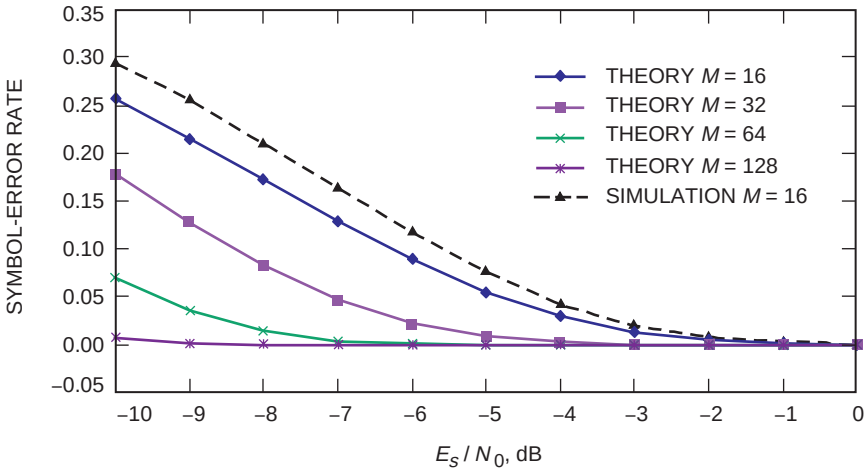

Fig. 7. Symbol-error probabilities for orthogonal block-coded modulation.

Note that transition rates of less than 0.1 can be obtained at an SNR of approximately −6 dB with $M = 16$, −8 dB with $M = 32$, and −10 dB with $M = 64$. Hence, squaring loss can be effectively eliminated in these regions by using a simple orthogonal block code of sufficiently high dimension.

IV. Simulation of a Closed-Loop Information-Reduced Phase-Estimation System

The entire information-reduced phase-estimation system described above, including the symbol-sequence estimator and the delay block, has been simulated in software using the SPW. Care was taken to keep critical system parameters (such as loop bandwidth) constant as other parameters were varied. Phase estimates supplied by the MAP estimator were used to demodulate the binary symbols in the symbol-sequence estimator; decisions were made on the encoded data; and the result was used to multiply a delayed version of the modulated signal in order to reduce the number of transitions.

The reduction in the symbol transition rate at the input to the MAP phase estimator was measured directly using the integrator output in the in-phase arm of the loop (Fig. 2). Typically, several simulations of 10^5 or more samples were averaged for each measurement, after the loop converged to steady-state tracking. The average of several measurements is shown in Fig. 7 for the case of $M = 16$ orthogonal modulation, in addition to the theoretical results. The simulation results compare well with theory, although performance is uniformly worse than predicted due to unmodeled effects.

In order to establish a performance baseline, the MAP phase estimator first was evaluated with an unmodulated tone corresponding to $p = 0$, in which case the nonlinearity in the in-phase arm defined in Eq. (16) reduces to a constant; therefore, the estimator structure reduces to a conventional phase-locked loop. Since, by definition, a phase-locked loop has no squaring loss, it represents the upper limit on phase-estimation performance that can be achieved by the MAP estimator observing BPSK modulated signals. The performance of the conventional polarity-type Costas loop represents the other extreme of symbol-by-symbol decoding of uncoded BPSK modulation, which is known to suffer increasingly great squaring loss at low symbol SNRs.

The improved performance of the MAP phase estimator with orthogonal block-coded modulation and data removal is shown in Fig. 8 for $M = 16$ (optimum); it essentially matches the ideal performance of the phase-locked loop at SNRs above −5 dB by virtually eliminating squaring losses in this region and continues to combat squaring losses effectively at lower SNRs. In fact, the improvement over the polarity-type Costas loop remains greater than 5 dB in the entire region between −8 and −5 dB.

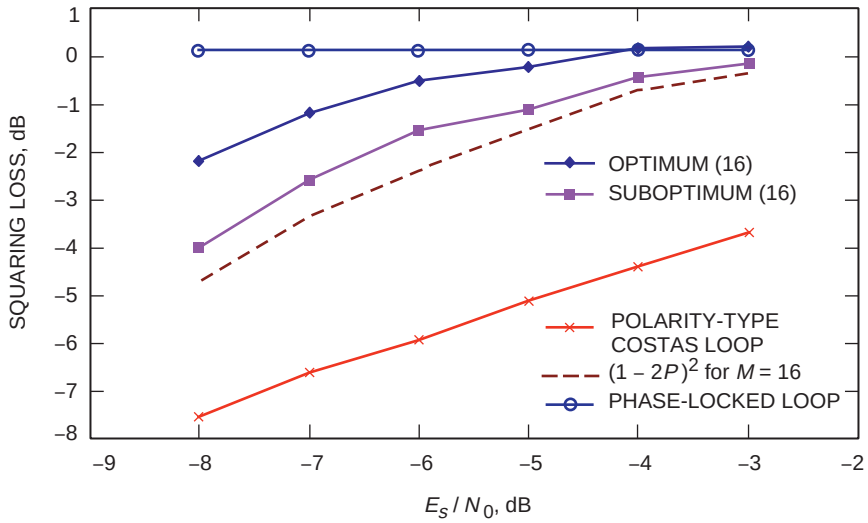

Fig. 8. Squaring-loss reduction for block-coded BPSK modulated signals.

A simplified estimator structure that uses only the quadrature arm of the MAP estimator but rescales the phase estimate according to Eq. (16) in order to preserve loop bandwidth also has been evaluated by means of analysis and simulations and is shown in Fig. 8, labeled "suboptimum." Since it is a suboptimal estimator, it cannot achieve the performance of the MAP estimator; however, it is much simpler to implement because it does not require nonlinear processing. As discussed previously, the squaring loss associated with this estimator is approximately equal to $(1-2p)^2$, represented by the dashed line in Fig. 8 for comparison.

These results can be extended to predict estimator performance with higher dimensional orthogonal signals by making use of the information in Fig. 7. Since the reduction in squaring loss is directly related to reduced transitions, it is clear that squaring losses can be kept to less than 1 dB down to a symbol SNR of −8 dB with 32 dimensions, and to less than −10 dB by resorting to 64-dimensional block codes.

V. Summary and Conclusions

The iterative information-reduced MAP phase estimator for BPSK symbols described in [4] has been simulated in SPW and evaluated at very low symbol SNRs, where squaring losses for conventional phase estimators start to become prohibitive. The approach is to try to obtain preliminary estimates of the data symbols and use these estimates to remove the data modulation from the carrier before attempting to estimate the phase. It was shown that, for the case of uncoded BPSK modulation, both symbol-by-symbol and block estimation were equivalent; hence, no further improvement in squaring loss could be obtained over the conventional polarity-type Costas loop, as described in [4]. However, it also was demonstrated that even simple coding schemes (such as block orthogonal codes) can lead to very dramatic squaring-loss reduction; the entire system, consisting of symbol-sequence estimator and MAP phase estimator operating in a closed-loop arrangement, was evaluated and shown to yield more than 5-dB improvement over a polarity-type Costas loop at extremely low symbol SNRs (between −8 and −3 dB). This improvement could be used to extend the telecommunications range of future spacecraft, operating with fixed power and antenna gain, by nearly a factor of two.

A simple suboptimal estimator structure also was derived and evaluated both analytically and by simulation and was shown to perform nearly as well as the optimal structure in the same range. It was demonstrated analytically that higher dimensional block codes employed in the same system could virtually eliminate squaring loss in the region of interest, effectively increasing the useful operating range of future spacecraft by more than a factor of three. Even greater improvements are expected from more sophisticated codes and symbol-sequence estimators, although these gains remain to be demonstrated.

References

[1] M. K. Simon and W. C. Lindsey, "Optimum Performance of Suppressed Carrier Receivers with Costas Loop Tracking," *IEEE Transactions on Communications*, vol. COM-25, no. 2, pp. 215–227, February 1977.

[2] J. J. Stiffler, *Theory of Synchronous Communications*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971.

[3] J. H. Chiu and L. S. Lee, "Maximum Likelihood Synchronizers with Arbitrary Prior Symbol Probabilities," *IEEE Proceedings*, vol. 138, no. 1, pp. 50–52, February 1991.

[4] M. K. Simon and V. A. Vilnrotter, "Iterative Information-Reduced Carrier Synchronization Using Decision Feedback for Low SNR Applications," *The Telecommunications and Data Acquisition Progress Report 42-130, April–June 1997*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–21, August 15, 1997. http://tmo.jpl.nasa.gov/tmo/progress_report/42-130/130A.pdf

[5] J. M. Wozencraft and I. M. Jacobs, *Principles of Communications Engineering*, New York: John Wiley and Sons, Inc., 1965.