Iterative Information-Reduced Carrier Synchronization Using Decision Feedback for Low SNR Applications

M. K. Simon and V. A. Vilnrotter
Communications and Systems Research Section

Traditional methods for carrier synchronization of uncoded binary phase-shift-keyed (BPSK) signals, e.g, conventional and polarity-type Costas loops, data-aided loops, demodulation–remodulation loops, etc., are obtained from approximations made to a closed-loop structure motivated by the maximum a posteriori (MAP) estimation of carrier phase. Inherent in all of these loops is the fact that their input data are assumed to be equiprobable (balanced) independent identically distributed (i.i.d.) binary sequences, and the MAP estimation loop from which these various structures are derived is predicated on this fact. The tracking performance limitation of such loops is the so-called squaring loss associated with the mean-squared phase error, which is the result of signal and noise products created by the necessity to remove the data modulation from the loop error signal. By reducing the amount of randomness (information) in the data that are input to the carrier synchronizer (hence the term information-reduced carrier synchronization), yet maintaining their i.i.d (but not necessarily balanced) property and also their independence of the additive noise, and then suitably modifying the synchronizer structure in accordance with this data reduction, one can obtain a significant squaring-loss improvement relative to what is achievable with the above-mentioned structures. One method for accomplishing this reduction in data randomness is through the use of decision feedback at the input *of the loop structure, which should be contrasted with conventional structures such as the data-aided or polarity-type Costas loops that use decision feedback* within *the loop structure itself, and hence do not modify the structure based on the amount of information reduction achieved. While applying such a method to uncoded modulations will not satisfy the noise independence constraint on the modified data sequence (for reasons explained in the article), for coded modulations, we show that it is indeed possible to approximately satisfy both the i.i.d and noise independence conditions and, thus, to achieve the above-mentioned gains. Although the article focuses on binary modulation, the concepts and results are easily extended to multiple phase-shift-keyed (*M*-PSK) modulations with* $M \geq 4$.

I. Introduction

Optimum closed-loop structures for tracking the phase of binary phase-shift-keyed (BPSK) signals as well as unmodulated carriers have been derived in the past based on maximum a posteriori (MAP) estimation criteria [1–3]. These structures have been shown to take the form of feedback loops that attempt to null the difference between the true input phase and its estimate produced by an oscillator in the loop. It has been customary in the past to distinguish between unmodulated and modulated carriers before deriving the optimum closed-loop structures, the former leading to a phase-locked loop (PLL) and the latter leading to an in-phase–quadrature (I-Q) loop with a hyperbolic tangent nonlinearity in its in-phase arm.

It is well known that PLLs attain a phase error variance that, in the limit of large loop signal-to-noise ratio (SNR), varies as the inverse of this loop SNR, whereas I-Q loops suffer an additional degradation because of the inherent multiplication required to produce an error signal in the I-Q structure. This multiplication of the I and Q signals results in the generation of signal and noise cross products, which leads to a loop SNR penalty (i.e., multiplication of the loop SNR by a factor less than unity) generically referred to as squaring loss. At low symbol SNR,[1] the squaring-loss degradation associated with I-Q loops can be severe, often prohibiting the ability to track. This suggests that substantial improvements may be possible if somehow one could convert the received modulated carrier to a pure (unmodulated) tone *before* applying it to the phase-tracking loop, since this would then allow use of a PLL which, as mentioned above, has no associated squaring-loss penalty.

In principle, an uncoded BPSK signal could indeed be converted to a pure tone if the data sequence were completely known, simply by multiplying the BPSK signal by the data waveform. Short of complete knowledge of the data waveform and in the presence of noise, the next best attempt at arriving at a "pure" input tone would be to feed back decisions (estimates) made on the data symbols. Referring to Fig. 1, the above statements can be put into mathematical terms as follows. Denoting the received signal plus noise waveform by $x\,(t; a_k, \theta)$, where a_k is the kth data symbol taking on values ± 1 with equal probability and θ is the unknown carrier phase to be estimated, it follows that if a perfect estimate, $\hat{a}_k$, of the data were available, so that $\hat{a}_k = a_k$, then the product of a_k and $\hat{a}_k$ would always be unity and the data modulation would in effect be removed from the signal input to the carrier synchronizer. Implicit in the previous statement is the assumption that accurate symbol synchronization has already been established and that data estimation delays have been accounted for by means of the signal delay, Δ, prior to the multiplication operation. In the absence of this ideal situation, the product of a_k and $\hat{a}_k$ forms an error sequence, $e_k \stackrel{\triangle}{=} a_k \hat{a}_k$, with statistics based on the probability, p, assigned to the error event, $\hat{a}_k \neq a_k$. Thus, ignoring for the moment the effect of this data estimate multiplication on the noise

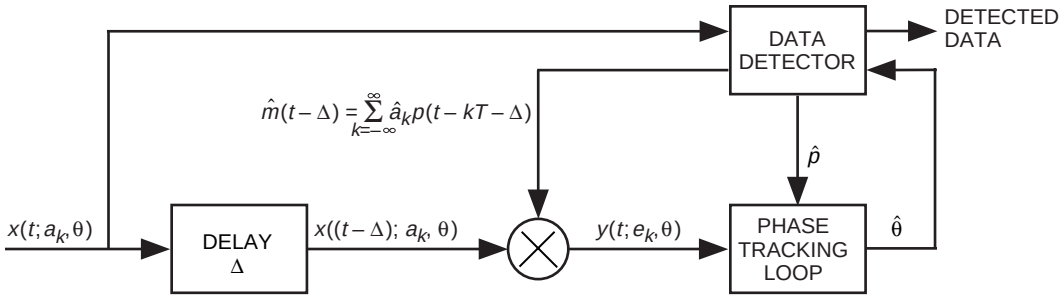

Fig. 1. Receiver with information-reduced carrier phase estimation.

[1] We use the term "symbol" rather than "bit" to include the case of coded data that will be our main focus of interest. Furthermore, it should be noted that assumption of low *symbol* SNR does not necessarily invalidate the assumption of high *loop* SNR.

component of $x(t; a_k, \theta)$ (which, as we shall see shortly, is an important consideration and is, in fact, what prevents this scheme from being successful for uncoded modulations), the input to the loop has been transformed into $y(t; e_k, \theta)$, whose signal component now represents the carrier modulated by the error sequence instead of the original data sequence. Assuming that the data estimates are independent and identically distributed (i.i.d.) (as would be the case for uncoded data), then the error sequence $\{e_k\}$ is an unbalanced i.i.d. sequence with the data estimate error probability, p, now also representing the probability that e_k takes on a value equal to -1. Hence, a value of $p = 0.5$ results in a balanced i.i.d. sequence and, thus, the feedback provides no advantage, whereas a reduction in the value of p to $p = 0$ would result in an all-ones input sequence, i.e., an unmodulated carrier. The degree to which p can be reduced from its nominal value of $p = 0.5$ depends on the nature of the data estimator (detector), which for uncoded data would be a matched filter followed by a hard decision threshold device.

If it were not for the effect of the decision feedback on the noise component of the received signal that is input to the loop, then modification of the loop structure in accordance with MAP theory based on the changed data sequence statistics discussed above should yield an improvement in performance. Again, this would occur because any unbalance (tilt) in the received data statistics, i.e., p other than 0.5, brings the signal component of the received signal closer to a "pure" tone, which intuitively should allow phase estimation with a smaller squaring loss. Unfortunately, for bit-by-bit detection of uncoded BPSK with a matched filter, each data decision that is output from the hard decision threshold device is based on the same symbol interval of noise as the corresponding interval of noise in the received signal that gets multiplied by this decision. Stated another way, the error statistic, $e_k \triangleq a_k \hat{a}_k$ (which of course contains the decision $\hat{a}_k$), corresponding to the kth transmission interval is totally correlated with the noise component of $x(t; a_k, \theta)$ during that same time interval and, thus, the assumption of a modified data sequence input to the loop that is independent of the additive noise is invalid. Indeed, we shall show later on (in the Appendix) that, for the configuration of Fig. 1 when implemented for uncoded BPSK,[2] because of the above-mentioned data sequence and noise correlation, the squaring loss performance cannot be improved relative to the well-known conventional techniques. We have decided to include the analysis that establishes this fact, even though our interest is focused on applying the information-reduction scheme to coded data, in order to inform the reader of the lack of utility of this scheme in the uncoded data application.

Suppose now that the BPSK symbols correspond to encoded data and the decisions result from, say, a convolutional (conventional or iterative) decoder. Then a memoryless data model for the error sequence at the phase tracking loop input is strictly speaking inappropriate. If the joint statistics of the input symbols were known as well as the actual decisions statistics, then, in principle, the structure of the MAP estimator could be derived. However, the mathematics would become intractable and, thus, it behooves us to find a way to justify the i.i.d. and noise independence assumptions for the error sequence. One simple way of validating the i.i.d assumption is to preprocess (e.g., interleave) the data input to the carrier synchronizer and at the same time perform the same preprocessing of the data being fed back from the decoder in such a way as to randomize the error sequence. We point out that such processing of the data and data feedback used as input to the carrier synchronizer does not affect the actual data-decoding process since this is performed external to the carrier synchronization operation. To validate the assumption that the components of the error sequence are approximately independent of the additive noise in the corresponding transmission intervals (a situation that, as mentioned above, is not possible for the uncoded data case), we argue as follows: In a coded system, the decisions on the data information bits (or equivalently on the encoded data symbols, as are needed for the carrier synchronizer), are made based on an observation interval of the received signal much longer than a single transmission interval. For example, in a convolutionally encoded data system, the output decision of the maximum-likelihood (ML) decoder, e.g, the Viterbi decoder, is obtained from the first bit of the current ML sequence whose length corresponds to a number of code constraint lengths sufficient to have all surviving paths converge

[2] Note that in this case the delay Δ becomes equal to the bit time T.

at their origins.[3] Thus, the noise that determines this ML sequence decision and, hence, the decision on the first bit (or equivalently the first encoded symbols) in this sequence is virtually uncorrelated with the noise in the same first transmission interval.

Assuming the above unbalanced i.i.d models, the next section derives the open-loop MAP estimator of carrier phase and the closed-loop carrier synchronizer motivated by this approach.

II. MAP Estimation of Carrier Phase for BPSK With Unbalanced Data

Consider the estimation of carrier phase for uncoded BPSK signals with unbalanced data in the presence of additive white Gaussian noise (AWGN). To begin, the received signal is of the form

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

while the additive noise has the narrowband representation

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

The carrier in Eq. (1) has power S, radian frequency ω_c, slowly varying phase $\theta(t)$, and is modulated by a random pulse train

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

where $p\left(t\right)$ is a unit power rectangular pulse of duration T seconds and $\{a_k\}$ is a binary (± 1) equiprobable data sequence assumed to be i.i.d. The noise process in Eq. (2) is modeled in terms of a pair of independent, low-pass Gaussian processes, $N_c(t)$ and $N_s(t)$, each with single-sided power spectral density (PSD) N_0 W/Hz and bandwidth $B_H < \omega_c/2\pi$.

In accordance with Fig. 1, the received signal plus noise, $x\left(t;a_k,\theta\right)$, is delayed by an amount Δ and then multiplied by the data estimator waveform

$$\hat{m}(t-\Delta) = \sum_{k=-\infty}^{\infty} \hat{a}_k p\left(t - kT - \Delta\right) \tag{4}$$

where $\{\hat{a}_k\}$ denotes the sequence of binary (± 1) data estimates. Assuming, as previously discussed, that this sequence of estimates is i.i.d. and has the probability statistics

$$\left.\begin{aligned} &\Pr\left\{\hat{a}_k \neq a_k\right\} = p \\ &\Pr\left\{\hat{a}_k = a_k\right\} = 1-p \end{aligned}\right\} \tag{5}$$

[3] It is envisioned for such applications that the symbol decision feedback could be obtained by storing the encoded symbol history of the survivor paths associated with the decoding algorithm rather than the input history normally associated with these paths. This soft symbol decision information would then be fed back synonymously with the time at which one would normally start making bit decisions with the decoder, in which case the delay Δ in Fig. 1 would be set equal to the decoder delay.

where p represents the error probability of the data estimator, then the result of the product of $x(t-\Delta)$ and $\hat{m}(t-\Delta)$ can be modeled as[4]

$$y(t) = \sqrt{2S}e(t)\sin(\omega_c t + \theta(t)) + N(t) \tag{6}$$

where

$$e(t) = \sum_{k=-\infty}^{\infty} e_k p(t - kT) \tag{7}$$

is a modified data modulation that reflects the errors in the detection of the input data sequence, i.e., $\{e_k\}$ is a binary (± 1) i.i.d. sequence with probability statistics[5]

$$\left.\begin{aligned} \Pr\{e_k = -1\} &= p \\ \Pr\{e_k = 1\} &= 1 - p \end{aligned}\right\} \tag{8}$$

and $N(t) \triangleq \hat{m}(t)n(t)$ is an additive noise process that is Gaussian in each T-second interval with statistics identical to $n(t)$ and, based on the arguments previously given for the coded application, is independent of the data.

The combined signal plus noise in Eq. (6) is observed for K data intervals, i.e., over the interval $0 \le t \le KT$. Assuming that the carrier phase $\theta(t)$ remains constant over this observation interval, then based on this observation and knowledge of $S, p(t)$ and ω_c, the MAP estimate of phase is that value $\hat{\theta}_{MAP}$ that maximizes the conditional probability density function $p(\theta|y(t))$, or equivalently $p(y(t)|\theta)$, since $\theta(t) = \theta$ is assumed to be uniformly distributed. The solution to this problem when $p = 1/2$ is well known and is given by

$$\hat{\theta}_{MAP} = \max_{\theta}{}^{-1} \ln \Lambda(\theta) \tag{9}$$

where $\ln \Lambda(\theta)$ is the logarithm of the likelihood function derived from $p(y(t)|\theta)$ and is given by

$$\ln \Lambda(\theta) = \ln \prod_{k=0}^{K-1} \cosh y_{sk}(\theta) = \sum_{k=0}^{K-1} \ln \cosh y_{sk}(\theta) \tag{10}$$

where

$$y_{sk}(\theta) \triangleq \frac{2\sqrt{2S}}{N_0} \int_{kT}^{(k+1)T} y(t)\sin(\omega_c t + \theta)\, p(t)dt \tag{11}$$

[4] In what follows, we ignore the delay Δ (i.e., set it equal to zero) since its value has no bearing on the derivation that follows. Also, for simplicity of notation, we herein use $y(t)$ instead of $y(t; e_k, \theta)$.

[5] In reality, the error probability of the decisions being fed back will be dependent on the tracking performance of the carrier synchronization loop itself since indeed the phase of the in-phase demodulation reference signal (the output of the phase tracking loop in Fig. 1) is an input to the data detector. For the moment, we ignore this effect since, as discussed later on, for large loop SNR, which is the case of interest, this is a second-order effect.

The "cosh x" form in Eq. (10) is obtained by averaging the conditional probability $p\left(x(t)|\theta, e_k\right)$, which from the AWGN assumption has an exponential form, over the equiprobable ($p = 1/2$) i.i.d. data sequence $\{e_k\}$. When this sequence has the unbalanced statistical properties of Eq. (8), then the analogous expression to Eq. (10) becomes

$$\ln \Lambda\left(\theta\right) = \sum_{k=0}^{K-1} \ln\left[(1-p)\exp\left(y_k(\theta)\right) + p\exp\left(-y_k(\theta)\right)\right]$$

$$= \sum_{k=0}^{K-1} \ln\left[\exp\left(y_k(\theta)\right) - 2p\sinh y_k(\theta)\right] \tag{12}$$

Since, from Eq. (9), $\hat{\theta}_{MAP}$ is the value of θ that maximizes the log likelihood function of Eq. (12), an equivalent statement is that $\hat{\theta}_{MAP}$ is the value of θ at which the derivative of Eq. (12) has zero value (and the second derivative is negative). For estimates $\hat{\theta}$ of θ in the neighborhood of $\hat{\theta}_{MAP}$, the derivative of the log likelihood function will be positive or negative in accordance with the sign of $\hat{\theta} - \hat{\theta}_{MAP}$ and, thus, this derivative can be used as an error signal in a closed-loop synchronizer to steer the loop in the direction of a locked condition corresponding to $\hat{\theta} = \hat{\theta}_{MAP}$.

Taking the derivative of Eq. (12) with respect to θ gives

$$\frac{\partial}{\partial\theta}\ln\Lambda\left(\theta\right) = \sum_{k=0}^{K-1} \frac{(1-p)\exp\left(y_{sk}(\theta)\right) - p\exp\left(-y_{sk}(\theta)\right)}{(1-p)\exp\left(y_{sk}(\theta)\right) + p\exp\left(-y_{sk}(\theta)\right)} y_{ck}(\theta)$$

$$= \sum_{k=0}^{K-1} \frac{\exp\left(y_{sk}(\theta)\right) - 2p\cosh y_{sk}(\theta)}{\exp\left(y_{sk}(\theta)\right) - 2p\sinh y_{sk}(\theta)} y_{ck}(\theta) \tag{13}$$

where, by analogy with Eq. (11),

$$y_{ck}\left(\theta\right) \triangleq \frac{2\sqrt{2S}}{N_0}\int_{kT}^{(k+1)T} y(t)\cos\left(\omega_c t + \theta\right)p(t)dt \tag{14}$$

Defining the zero memory nonlinearity

$$f_p\left(x\right) \triangleq \frac{(1-p)\exp\left(x\right) - p\exp\left(-x\right)}{(1-p)\exp\left(x\right) + p\exp\left(-x\right)} = \frac{\exp\left(x\right) - 2p\cosh x}{\exp\left(x\right) - 2p\sinh x} \tag{15}$$

then the closed-loop carrier synchronizer motivated by the above MAP estimation procedure is as illustrated in Fig. 2.[6] We herein refer to this loop as the *MAP estimation loop for BPSK with unbalanced data*. A similar result was obtained in [4], where it was also shown that the nonlinearity of Eq. (15) can be expressed in the form of a biased hyperbolic tangent function, namely,

[6] In Fig. 2, the nonlinearity is shown as $f_{\hat{p}}(x)$ since, in an actual implementation, $\hat{p}$ is an estimate of the true value of p as determined from measurements made on the receiver. Furthermore, in reality, $\hat{p}$ depends on the loop phase error and the detection SNR, R_d. Later on in this article, we discuss these dependencies as well as the effect of a mismatch between $\hat{p}$ and p. For the moment, we set $\hat{p} = p$.

$$f_p(x) \triangleq \tanh\left(x - \frac{1}{2}\ln\frac{p}{1-p}\right) \tag{16}$$

In the next section, we analyze the tracking performance of this loop in terms of its mean-squared phase error.

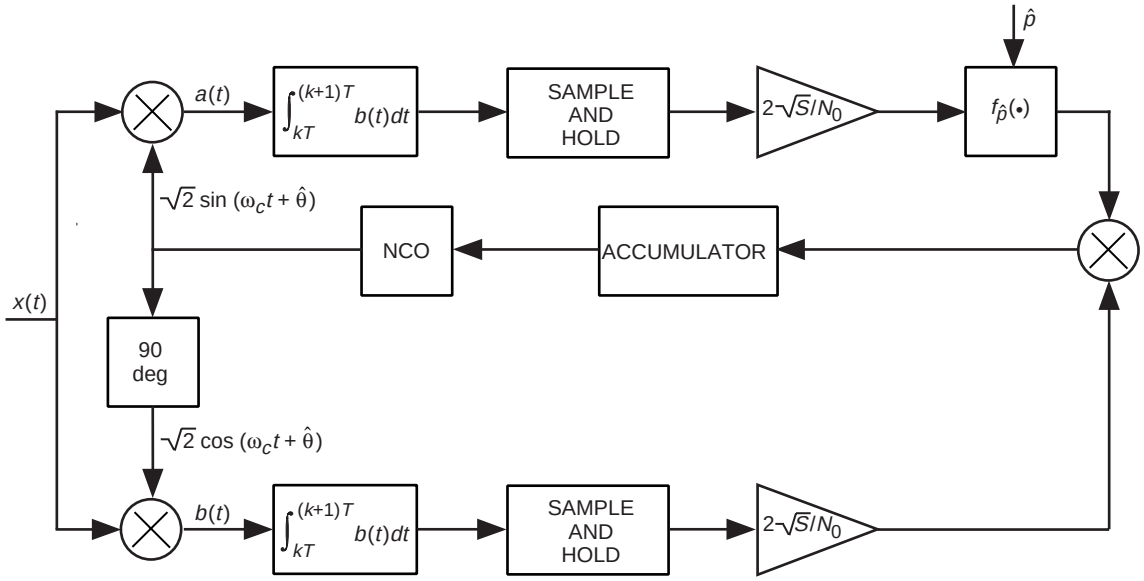

Fig. 2. Information-reduced carrier synchronization loop.

III. Tracking Performance of the MAP Estimation Loop for BPSK With Unbalanced Data

Consider the MAP estimation loop for phase synchronization of BPSK with unbalanced data illustrated in Fig. 2. Since this loop was derived based on the assumption of an input i.i.d. "data" sequence $\{e_k\}$ with probability statistics $\Pr\{e_k = -1\} = p, \quad \Pr\{e_k = 1\} = 1 - p$, we continue to use this signal model in assessing its performance. The analysis of the tracking performance of the loop in Fig. 2 parallels the approach taken in [1] for a similar loop with a hyperbolic tangent nonlinearity. We note, however, that there are some significant differences between the two analyses since several of the steps carried out in [1] depend upon the nonlinearity (e.g., hyperbolic tangent) being an odd function of its argument, which is not the case for the nonlinearity $f_p(x)$ (except when $p = 1/2$, in which case $f_p(x) = \tanh x$).

Denoting the input I and Q phase detector (multiplier) gains by K_m, then the sample-and-hold outputs $z_c(t)$ and $z_s(t)$ of Fig. 2 are given by

$$\left.\begin{aligned} z_c(t) &\triangleq \int_{kT}^{(k+1)T} y(\tau) r_c(\tau) d\tau = K_m \frac{2\sqrt{S}}{N_0}\left[T\sqrt{S}e_k \sin\phi - N_2 \sin\phi + N_1 \cos\phi\right] \\ z_s(t) &\triangleq \int_{kT}^{(k+1)T} y(\tau) r_s(\tau) d\tau = K_m \frac{2\sqrt{S}}{N_0}\left[T\sqrt{S}e_k \cos\phi - N_1 \sin\phi - N_2 \cos\phi\right]; \\ &\qquad\qquad (k+1)\,T \le t \le (k+2)\,T \end{aligned}\right\} \tag{17}$$

where N_1 and N_2 are zero-mean, independent Gaussian random variables[7] with variance $\sigma_{N_1}^2 = \sigma_{N_2}^2 = N_0T/2 \triangleq \sigma^2$, and $\phi(t) \triangleq \theta(t) - \hat{\theta}(t)$ is the loop phase error[8].

Multiplying $z_c(t)$ by the nonlinearly processed $z_s(t)$ gives the dynamic error signal

$$\begin{aligned} z_o(t) &= z_c(t) f_p\left(z_s(t)\right) \\ &= K_m\left[(2R_d)\, e_k \sin\phi - \sqrt{2R_d}X_2 \sin\phi + \sqrt{2R_d}X_1 \cos\phi\right] \\ &\quad \times f_p\left(K_m\left[(2R_d)\, e_k \cos\phi - \sqrt{2R_d}X_1 \sin\phi - \sqrt{2R_d}X_2 \cos\phi\right]\right); \\ &\qquad (k+1)\,T \le t \le (k+2)\,T \end{aligned} \tag{18}$$

where $f_p(x)$ is the nonlinearity defined in Eq. (15) that depends on the probability statistics of the input data sequence, i.e., the parameter p. Also, in Eq. (18), we have defined the detection signal-to-noise ratio (symbol energy-to-noise ratio for a coded data system) by $R_d \triangleq ST/N_0$ and the normalized (unit variance) Gaussian noise variables by $X_i \triangleq N_i/\sigma;\ i = 1, 2$.

It has been shown [1,2] that the tracking performance of a loop such as that in Fig. 2 can, in its linear region of operation (small phase error), be determined by examining the equivalent signal and noise components of the $z_o(t)$ process, more specifically, the slope of the equivalent S-curve at $\phi = 0$ and the variance of the equivalent additive noise. This makes the usual assumption that the loop bandwidth is much less than the data bandwidth.

Since X_1 and X_2 are zero-mean random variables, then from Eq. (18), the signal component $\eta(t)$ of $z_o(t)$ is[9]

$$\eta(\phi) = \overline{(2R_d e_k \sin\phi)\, f_p\left(2R_d e_k \cos\phi - \sqrt{2R_d}X_1 \sin\phi - \sqrt{2R_d}X_2 \cos\phi\right)}^{X_1, X_2, e_k} \tag{19}$$

where the overbar denotes statistical averaging over the distributions of the random variables indicated. The slope K_η of this S-curve at the origin is obtained from

$$\begin{aligned} K_\eta = \frac{d\eta(\phi)}{d\phi}\Big|_{\phi=0} &= \overline{(2R_d e_k)\, f_p\left(2R_d e_k - \sqrt{2R_d}X_2\right)}^{X_2, e_k} \\ &= 2R_d\left[(1-p)\overline{f_p\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} - p\overline{f_p\left(-2R_d - \sqrt{2R_d}X_2\right)}^{X_2}\right] \\ &= 2R_d\left[(1-p)\overline{f_p\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} - p\overline{f_p\left(-2R_d + \sqrt{2R_d}X_2\right)}^{X_2}\right] \end{aligned} \tag{20}$$

[7] Note that, as previously mentioned, multiplication of a Gaussian process $n(t)$ by a binary (± 1) waveform of rate $1/T$ does not change the Gaussian properties of the original process within any T-second time interval. Thus, for any input bit, the corresponding sample-and-hold outputs of the integrate-and-dump (I&D) filters are still Gaussian random variables.

[8] For notational convenience, we herein delete the dependence of $\phi(t)$ on t.

[9] For simplicity, we further set $K_m = 1$.

where the latter equality stems from the fact that averaging over a zero-mean Gaussian random variable is equal to averaging over the negative of that variable. We should also point out the following in Eq. (20): In an actual implementation, p would itself depend on the loop phase error (since the data decisions being fed back to the input of the loop depend on the phase of the in-phase carrier demodulation reference derived from the loop). However, since the slope of the S-curve is evaluated at $\phi = 0$, then the value of p needed in Eq. (20) is the value at $\phi = 0$, i.e., the ideal performance of the data detector. Thus, the dependence of p on ϕ is irrelevant to the evaluation of the S-curve slope. While we could denote the value of p at $\phi = 0$ as p_0 in Eq. (20), for simplicity we shall maintain the notation p, keeping in mind the above interpretation. We do remind the reader, however, that for an actual implementation, p is a function of R_d.

The noise component of $z_o(t)$ (evaluated at $\phi = 0$) is

$$N_e(t) = \sqrt{2R_d}X_1 f_p\left(2R_d e_k - \sqrt{2R_d}X_2\right) \tag{21}$$

Since X_1 and X_2 are independent and have unit variance, then the variance $\sigma^2_{N_e}$ of $N_e(t)$ is from Eq. (21):

$$\begin{aligned}\sigma^2_{N_e} &= 2R_d\left[(1-p)\overline{f_p^2\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} + p\overline{f_p^2\left(-2R_d - \sqrt{2R_d}X_2\right)}^{X_2}\right] \\ &= 2R_d\left[(1-p)\overline{f_p^2\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} + p\overline{f_p^2\left(-2R_d + \sqrt{2R_d}X_2\right)}^{X_2}\right]\end{aligned} \tag{22}$$

Here again the value of p needed in Eq. (22) is the value at $\phi = 0$, which is consistent with the previous usage of the same parameter in Eq. (20) for the slope of the S-curve.

Because of the sample-and-hold circuits in Fig. 2, the noise process $N_e(t)$ of Eq. (21) is piecewise constant over intervals of T-second duration, in particular, $T \le t \le 2T$. Thus, as long as the loop bandwidth is much less than the data bandwidth, this process can be approximated, as had been done in [1,2], by a delta-correlated process with the correlation function given by

$$R_{N_e}(\tau) \triangleq \overline{N_e(t)N_e(t+\tau)} = \begin{cases} \sigma^2_{N_e}\left[1 - \dfrac{|\tau|}{T}\right]; & |\tau| \le T \\ 0; & |\tau| > T \end{cases} \tag{23}$$

and equivalent single-sided noise spectral density

$$\begin{aligned}N_0' &\triangleq 2\int_{-\infty}^{\infty} R_{N_e}(\tau)d\tau = 2\sigma^2_{N_e}T \\ &= 4R_dT\left[(1-p)\overline{f_p^2\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} + p\overline{f_p^2\left(-2R_d - \sqrt{2R_d}X_2\right)}^{X_2}\right] \\ &= 4R_dT\left[(1-p)\overline{f_p^2\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} + p\overline{f_p^2\left(-2R_d + \sqrt{2R_d}X_2\right)}^{X_2}\right]\end{aligned} \tag{24}$$

The linearized phase error variance is given by

$$\sigma_{\phi}^{2} = \frac{N_0^{'} B_L}{K_{\eta}^{2}} \triangleq \frac{1}{\rho S_L} \tag{25}$$

where $\rho \triangleq S/N_0 B_L$ is the linear loop (PLL) signal-to-noise ratio and, analogous to conventional Costas loop terminology, S_L is the "squaring loss," which reflects the penalty paid due to the signal and noise cross products present in $z_o(t)$. Substituting Eqs. (20) and (22) in Eq. (25), the squaring loss can be identified as

$$S_L = \frac{\left[(1-p)\overline{f_p\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} - p\overline{f_p\left(-2R_d + \sqrt{2R_d}X_2\right)}^{X_2}\right]^2}{(1-p)\overline{f_p^2\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} + p\overline{f_p^2\left(-2R_d + \sqrt{2R_d}X_2\right)}^{X_2}} \tag{26}$$

Note that for $p = 1/2$, Eq. (25) reduces to

$$S_L = \frac{\left[\overline{\tanh\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2}\right]^2}{\overline{\tanh^2\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2}} \tag{27}$$

which is in agreement with Eq. (18) of [2]. Although the statistical averages required in Eq. (26) cannot be performed analytically, they are easily evaluated numerically using Gauss–quadrature techniques.

IV. Approximate Behavior at Low SNR

It is of interest to examine the behavior of the data-reduced carrier synchronization scheme at low SNR (small symbol energy-to-noise ratio) since this is the region where the squaring loss associated with conventional carrier synchronization schemes is large and thus limits their performances. Since the outputs of the I&D filters in Fig. 2 are proportional to R_d, then for small values of this parameter, one can approximate the in-phase arm nonlinearity of Eq. (15) by its expression for small values of its argument, namely,

$$f_p(x) \cong 1 - 2p + \left[1 - (1-2p)^2\right] x = 1 - 2p + 4p(1-p)x \tag{28}$$

which is a *linear* function with a bias and slope that depend on the input data probabilities and is consistent with Eq. (9) of [4]. Corresponding to this approximation of the nonlinearity, the slope of the S-curve at the origin, as defined in Eq. (20), can be evaluated analytically and becomes

$$K_{\eta} \cong (1-2p)^2 2R_d + \left[1 - (1-2p)^2\right] 4R_d^2 \tag{29}$$

Similarly, the equivalent noise spectral density of Eq. (24) becomes, for low SNR,

$$N_0^{'} \cong 4R_d T \left\{(1-2p)^2 + 4R_d (1-2p)^2 \left[1 - (1-2p)^2\right] + 2R_d (2R_d + 1)\left[1 - (1-2p)^2\right]\right\} \tag{30}$$

Substituting Eqs. (29) and (30) in Eq. (25), we obtain, after considerable simplification, the closed-form result

$$S_L = \frac{2R_d + \dfrac{(1-2p)^2\left[4R_d - (1-2p)^2(4R_d-1)\right]}{2R_d\left[1-(1-2p)^2\right]^2}}{2R_d + 1 + \dfrac{(1-2p)^2\left[4R_d + 1 - (1-2p)^2(4R_d)\right]}{2R_d\left[1-(1-2p)^2\right]^2}} \tag{31}$$

The limiting values of Eq. (31) for $p = 1/2$ and $p = 0$ are

$$S_L = \begin{cases} \dfrac{2R_d}{2R_d+1}; & p = \dfrac{1}{2} \\ 1; & p = 0 \end{cases} \tag{32}$$

which agree with the well-known results for the conventional Costas loop with I&D arm filters when tracking an uncoded BPSK modulation and the PLL, respectively.

The limiting value of Eq. (31) for $R_d \to 0$ will also be of interest later on in relation to a more traditional decision feedback scheme derived from ML considerations applied to a received signal with a correlated data sequence. In particular, it is straightforward to show from Eq. (31) that, for the information-reduced scheme of Fig. 2,

$$\lim_{R_d \to 0} S_L = (1-2p)^2 \tag{33}$$

Also, since S_L is a monotonically increasing function of R_d, then Eq. (33) represents a lower bound on the performance of the synchronizer at any SNR.

Figure 3 is a plot of the squaring loss as determined from the exact expression of Eq. (26) versus R_d with p as a parameter. While we recognize that for any given form of data detector p will be a function of R_d, we have chosen to maintain p as a constant parameter in this figure. The reason for this is that the results presented there can then be independent of the specific form of data detector or, equivalently, the specific form of error correction coding employed, provided that the conditions for an i.i.d and noise-dependent error sequence are still met. For any given coding application, only a single point on each curve of constant p would be applicable, namely, the one corresponding to the given error rate behavior of the data detector (decoder) and, thus, the performance for that particular application would be illustrated by a single curve passing through this series of points. With this interpretation in mind, we note that the curve labeled $p = 0.5$ corresponds to the in-phase arm nonlinearity $f_p(x) = \tanh x$ and, as such, agrees with the squaring-loss behavior performance of the MAP carrier synchronization structure of uncoded BPSK modulation. For purpose of illustration only, we also include in Fig. 3 curves corresponding to the conventional Costas loop and polarity-type Costas loop for uncoded BPSK with squaring losses given by Eq. (32) for the former and

$$S_L = \operatorname{erf}^2 \sqrt{R_d} \tag{34}$$

for the latter. Finally, we note that for improvement to take place in the carrier synchronizer, the data estimator is not required to operate with symbol error probabilities as small as those typically needed

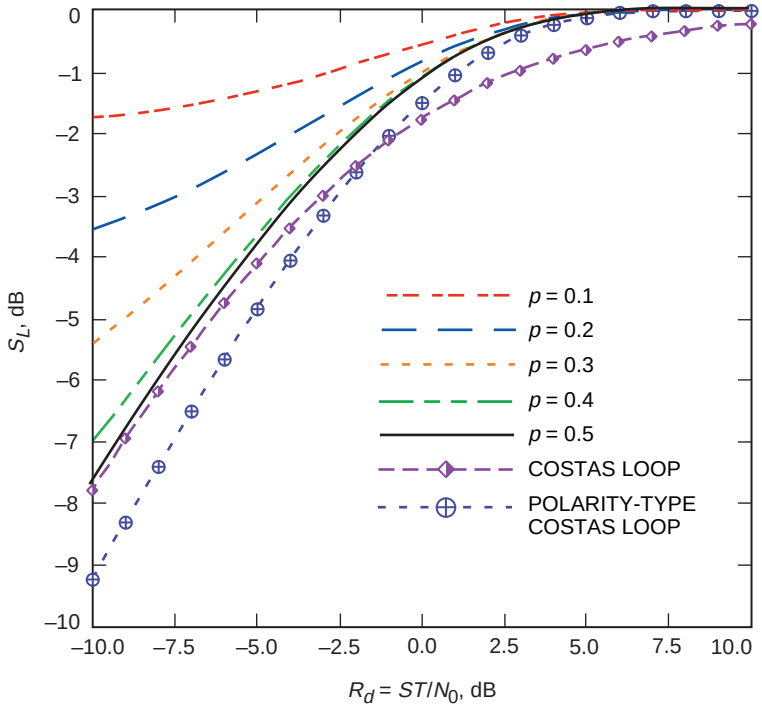

Fig. 3. Squaring-loss performance of an information-reduced carrier synchronization loop; perfect knowledge of *p*.

for reliable communication. In fact, a value of $p = 0.1$, which might be unacceptable for telemetry applications, can result in dramatic improvement in carrier synchronization performance when used in the manner illustrated in Fig. 1.

V. Sensitivity to Mismatch

Having improved the phase tracking by the use of data feedback, the new improved carrier phase estimate will help the data detector obtain better symbol estimates, which suggests an iterative procedure. That is, the phase tracking loop is initially designed assuming a value of p corresponding to no feedback. The initial data estimates obtained using this particular structure as a carrier synchronizer are fed back, and the carrier synchronizer structure is then modified based on the new estimated value of p. This updated carrier synchronizer provides improved phase tracking, which in turn provides improved data detection. These new data estimates, having a smaller value of p, are then fed back to the input of the carrier synchronizer, and the iteration continues. Thus, in practice, at any given time instant there will be a mismatch between the true value of p associated with the input error sequence $\{e_k\}$ and its estimate, $\hat{p}$, used to implement the nonlinearity of Eq. (15) in the loop. As such, we wish to investigate the sensitivity of the loop performance as measured by the squaring loss to mismatch between p and $\hat{p}$. That is, how good must our estimate of p be for the purpose of loop implementation and yet still achieve an improvement in performance relative to not implementing decision feedback at all?

We begin this discussion by quickly recognizing that when p and $\hat{p}$ are unequal, the squaring loss expression of Eq. (26) is modified to

$$S_L = \frac{\left[(1-p)\overline{f_{\hat{p}}\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} - p\overline{f_{\hat{p}}\left(-2R_d + \sqrt{2R_d}X_2\right)}^{X_2}\right]^2}{(1-p)\overline{f_{\hat{p}}^2\left(2R_d - \sqrt{2R_d}X_2\right)}^{X_2} + p\overline{f_{\hat{p}}^2\left(-2R_d + \sqrt{2R_d}X_2\right)}^{X_2}} \tag{35}$$

where $f_{\hat{p}}(x)$ is given by Eq. (15) with p replaced by $\hat{p}$. The first point to observe is that for $\hat{p} = 0.5$, $f_{\hat{p}}(x) = \tanh x$, which is an odd function of x; hence, the second average in the numerator of Eq. (35) is the negative of the first average, and the two averages in the denominator of Eq. (35) are equal. As a result of these observations, for this case the squaring loss becomes independent of p. That is, if we assume $\hat{p} = 0.5$ for the loop design, then the performance is independent of the true input probabilities. Next, suppose that we are optimistic in our estimate of p used for the design of the loop, i.e., $\hat{p} \leq p$. Figures 4(a) and (b) are plots of the squaring loss as determined from Eq. (35) versus R_d with $\hat{p}$ as a parameter and $p = 0.01$ and $p = 0.1$, respectively. The deviation between p and $\hat{p}$ is as much as a factor of 5. We see that for this case the performance of the loop is quite insensitive to the mismatch and yet we still obtain a significant improvement relative to the case of no feedback at all. Figures 5(a) and (b) show the analogous results for the case where we are pessimistic in our estimate of p used for the design of the loop, i.e., $\hat{p} \geq p$. Here, we see a good deal more sensitivity to the mismatch of the loop performance to the mismatch between p and $\hat{p}$, but always doing as well as or better than the case of no feedback at all.

VI. Relation to Other Structures Motivated by ML Considerations

Suppose that at the outset we had postulated a received signal $y(t)$ consisting of the sum of AWGN and a random binary data source with dependent data symbols, such as that output from a convolutional encoder, biphase modulated on a carrier. Based on an observation of $y(t)$ over K symbols and the assumption of a uniformly distributed unknown carrier phase θ, the MAP estimate of this phase is still that value $\hat{\theta}_{MAP}$ that maximizes $p(y(t)|\theta)$. In computing $p(y(t)|\theta)$ for the i.i.d. data source case, as in Section II, we first obtained the conditional probability $p(y(t)|\theta, \mathbf{a})$, where $\mathbf{a}$ is the vector of K transmitted data symbols and then averaged over the probability density function (pdf) of $\mathbf{a}$. For an i.i.d. data source as assumed in Section II (there the data source actually corresponds to the error sequence), this averaging over the data sequence, i.e., computing the average-likelihood ratio (ALR), can be performed symbol by symbol which, for binary data, results in a likelihood ratio (LR) in the form of a K-fold product of hyperbolic cosine functions [see Eq. (10)]. Taking the natural logarithm of this LR (which turns the product into a sum of "ln cosh" functions), and then differentiating with respect to the unknown parameter to find its maximum, results in the well-known MAP estimation loop with hyperbolic tangent nonlinearity (derivative of the "ln cosh" function) in its in-phase arm and a K-symbol accumulator following the I-Q multiplier.

When the data source is not i.i.d., then computation of the ALR by averaging $p(y(t)|\theta, \mathbf{a})$ over $\mathbf{a}$ cannot be performed symbol by symbol. Instead, the average must be computed over all 2^K possible data sequences, which for large K results in an estimation structure with high complexity. Furthermore, since this average cannot be written in the form of a K-fold product, then the advantage obtained by taking the natural logarithm of the ALR before differentiating is now lost [5]. Another possibility for obtaining a carrier phase estimator is first to maximize $p(y(t)|\theta, \mathbf{a})$ over $\mathbf{a}$, resulting in the ML estimate of the data sequence $\hat{\mathbf{a}}_{ML}$, and then to substitute this estimate into $p(y(t)|\theta, \mathbf{a})$ before maximizing over the carrier phase θ. Although this maximum-likelihood ratio (MLR) approach is suboptimum (since it does not result in maximization of the true LR $p(y(t)|\theta)$ with respect to θ), it nevertheless often produces an estimator whose performance is comparable with that obtained from the ALR approach. A closed-loop structure resulting from the MLR approach to carrier phase estimation is suggested in [5] and repeated here in Fig. 6 for clarity of this presentation. We observe that, in principle, the structure requires a maximum-likelihood sequence estimator (MLSE) in its in-phase arm prior to multiplication by the appropriate quadrature arm signal. An approximation of this vector decision feedback type of implementation that leads to a simplification in structure can be obtained in, for example, a convolutionally coded system wherein the MLSE is implemented as a Viterbi decoder that outputs symbol-by-symbol decisions after a suitable delay. Using these decisions to multiply the corresponding delayed (by the amount of the decoder delay) symbol-by-symbol output of the quadrature arm, one obtains a structure that resembles a polarity-type I-Q Costas loop, wherein the matched filter/hard limiter data detector combination in the in-phase arm is replaced by a Viterbi decoder. An analysis of the tracking performance of such a

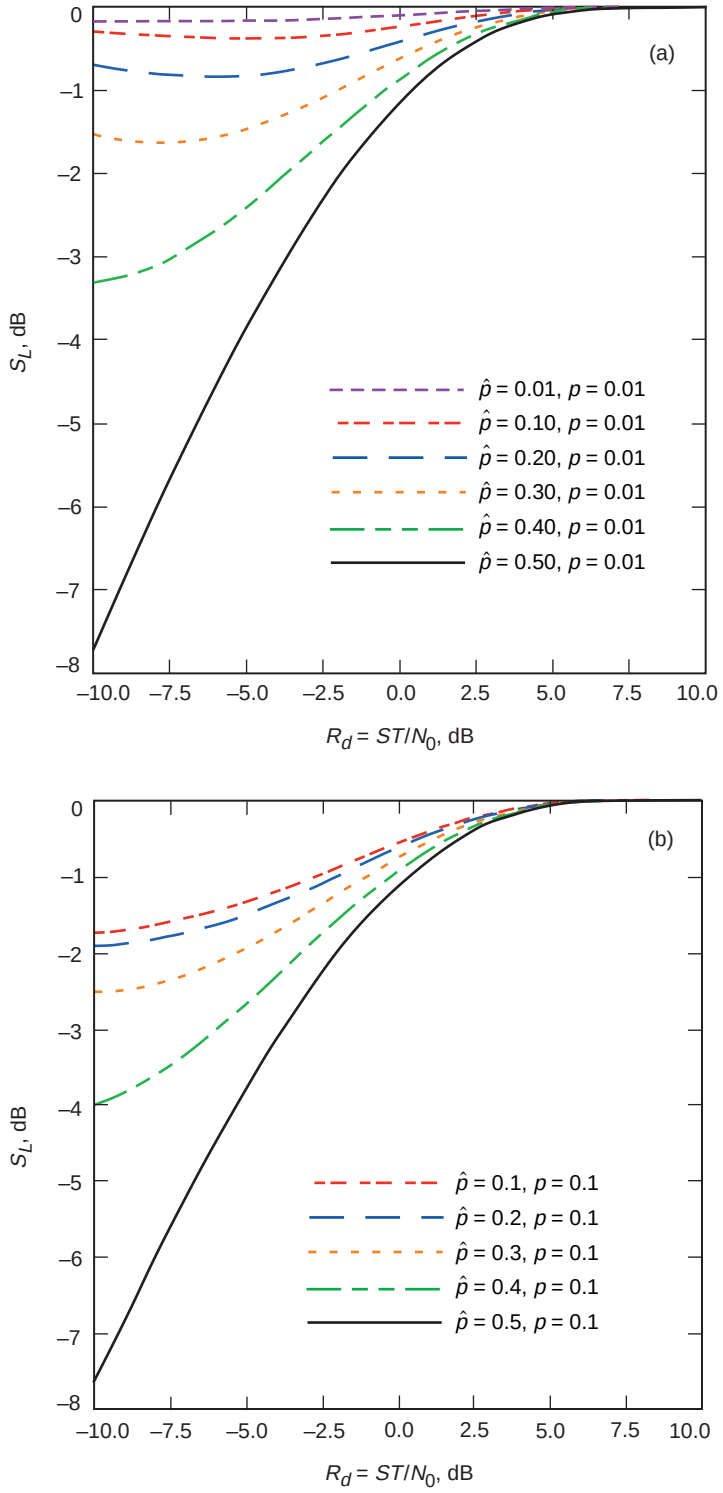

Fig. 4. The squaring-loss performance of an information-reduced carrier synchronization loop in the presence of mismatch: (a) $p = 0.01$, $\hat{p} \geq p$, and (b) $p = 0.1$, $\hat{p} \geq p$.

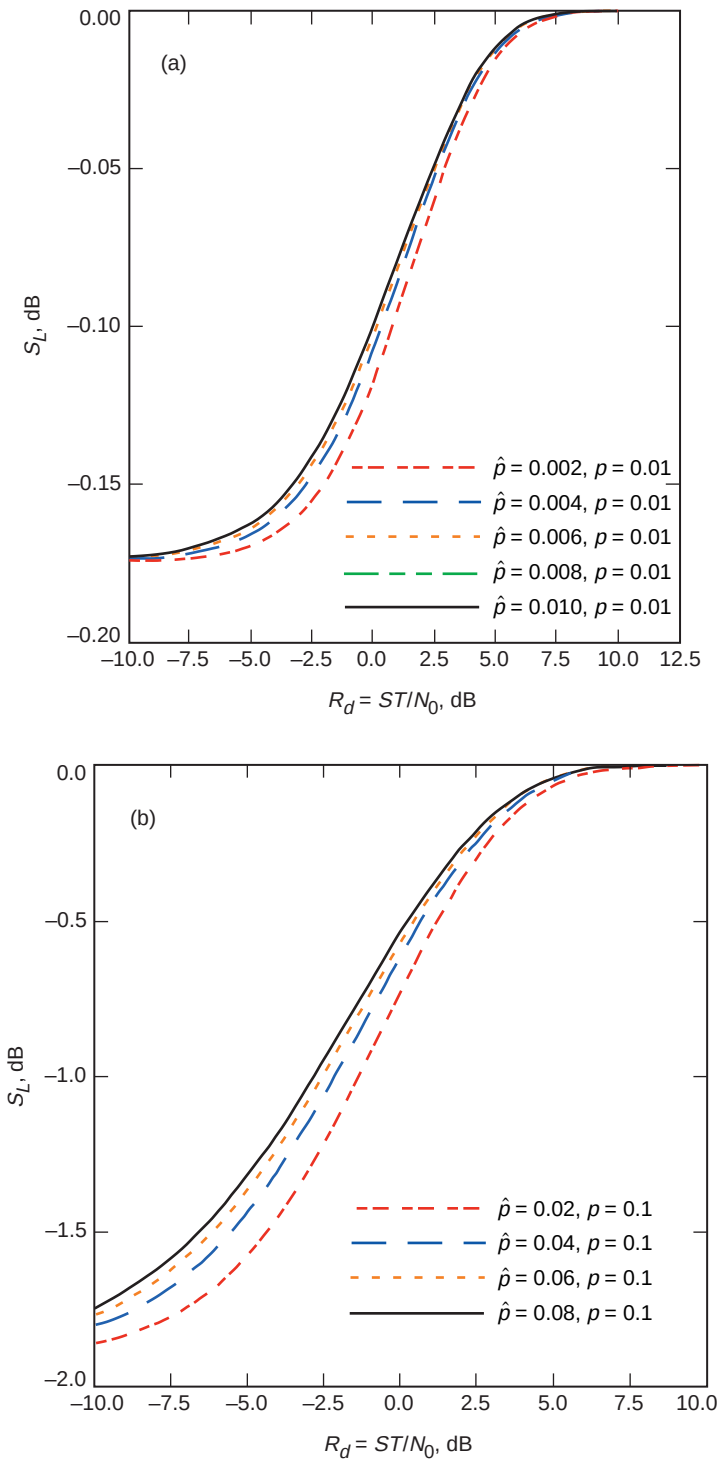

Fig. 5. The squaring-loss performance of an information-reduced carrier synchronization loop in the presence of mismatch: (a) $p = 0.01$, $\hat{p} \leq p$, and (b) $p = 0.1$, $\hat{p} < p$.

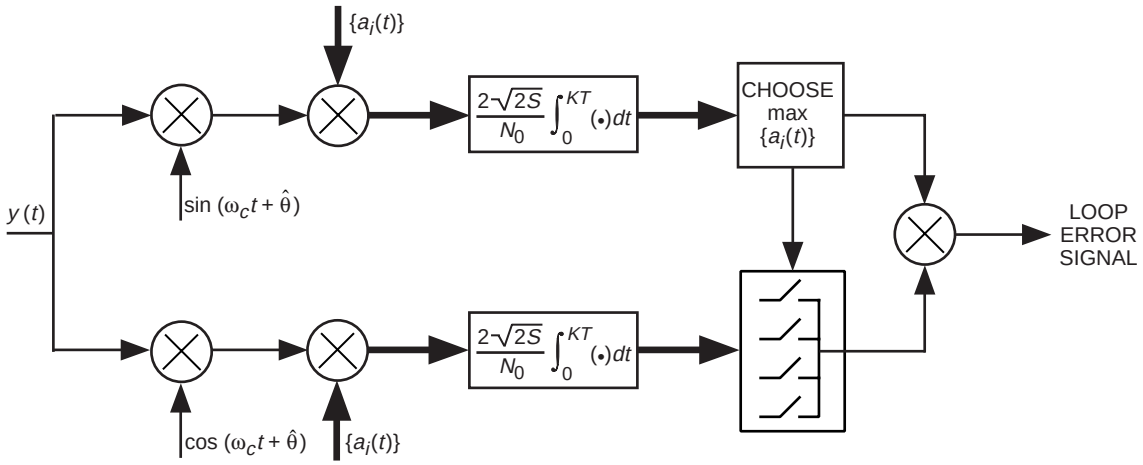

Fig. 6. Closed-loop carrier synchronizer motivated by MLR theory for coded BPSK, where $\{a_i(t)\}$ is a data waveform corresponding to data vector $\mathbf{a_i}$.

structure is straightforward and follows along the lines of that for an uncoded BPSK I-Q loop with a hard-limited in-phase arm whose squaring loss is given by Eq. (34). Suffice it to say that the analogous result to Eq. (34) for the above-mentioned I-Q structure with a Viterbi decoder in its in-phase arm is easily shown to be

$$S_L = (1-2p)^2 \tag{36}$$

where p is the symbol-error probability performance of the Viterbi decoder under ideal conditions, i.e., with zero phase error. Comparing Eq. (36) with Eq. (33) and keeping in mind the monotonically increasing nature of S_L with R_d, we conclude that for any $R_d \geq 0$, the information-reduced carrier synchronizer performs at least as well as (and often much better than) the scheme motivated by MLR considerations. The advantage of the MLR scheme, however, is that it can be implemented without the receiver having specific knowledge of p. We hasten to add that all schemes involving coded decision feedback of one form or another suffer from the potential instability brought about by the decoder delay introduced into the loop.

VII. Conclusions

A new type of iterative carrier-synchronization loop, designed for coded BPSK applications operating at low symbol SNR, has been described and analyzed. Using symbol estimates in a feedback configuration to reduce data transitions at the input of the tracking loop, the new structure achieves significant improvements in tracking threshold over conventional I-Q loops. It was shown that very accurate symbol estimates were not required to achieve significant improvements in low SNR tracking threshold or, equivalently, operation with greatly reduced phase error: a data detector operating at a symbol-error probability of 0.1 can, at low SNR, affect as much as a 7-dB reduction in squaring loss for BPSK signals. This means that the new carrier synchronization loop can operate effectively at signal levels that might be too low for a conventional I-Q loop to even lock. Another desirable feature of this new structure is that once phase locked, the improved phase estimates can be transferred to the data detector, yielding improved symbol estimates for feedback, thus achieving even better phase tracking; this kind of iterative implementation eventually could lead to a virtual elimination of squaring loss, hence approaching the tracking performance of a phase-locked loop operating on an unmodulated tone.

We conclude by noting several other scenarios where the use of information reduction can be used to improve tracking-loop performance. First, we point out that even though we have explicitly addressed

the *carrier* synchronization problem, the technique is equally applicable at the subcarrier level in residual carrier systems that employ suppressed subcarriers. Second, in systems that employ nonsuppressed carriers, i.e., the transmitted signal is a combination of a residual and suppressed-carrier signals, the MAP estimation procedure suggests a so-called hybrid loop [6], which is a weighted superposition of a PLL and a Costas loop. Here again, information reduction can be used to improve the performance of the suppressed-carrier portion (Costas loop) of the overall loop. Finally, the extension of the information-reduction concept to quadriphase-shift-keyed (QPSK) modulation is straightforward and results in even larger performance gains because of the inherently larger "squaring" (actually, fourth-order) losses associated with such a modulation.

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] M. K. Simon, "On the Optimality of the MAP Estimation Loop for Carrier Phase Tracking BPSK and QPSK Signals," *IEEE Transactions on Communications*, vol. COM-27, no. 1, pp. 158–165, January 1979.

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

[4] 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.

[5] H. Tsou, S. Hinedi, and M. K. Simon, "Closed-Loop Carrier Phase Synchronization Techniques Motivated by Likelihood Functions," *The Telecommunications and Data Acquisition Progress Report 42-119, July–September 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 83–104, November 15, 1994. http://tda.jpl.nasa.gov/tda/progress_report/42-119/119S.pdf

[6] M. K. Simon, H. Tsou, S. Hinedi, and K. Hamdan, "The Performance of a Coherent Residual Carrier PSK System Using Hybrid Carrier Phase Synchronization," *ICC '96 Conference Record*, Dallas, Texas, pp. 1275–1280, June 18–22, 1996.

Appendix
Iterative Information-Reduced Carrier Synchronization of Uncoded BPSK

In this appendix, we apply the notion of iterative information reduction to the problem of carrier synchronization of uncoded BPSK. In particular, for a rectangular pulse stream i.i.d. data-modulation biphase modulated on a carrier (i.e., uncoded BPSK), we consider the specific implementation of Fig. 1 illustrated in Fig. A-1. Here the data detector of Fig. 1 is a matched (integrate-and-dump) filter followed by a hard decision device and sample-and-hold circuit. Rather than model the product of the data decision and input data sequences as an error sequence with statistics as in Eq. (8), we instead allow these two sequences to maintain their separate identities so as to expose the all-important correlation of the former with the input noise in the corresponding transmission intervals.

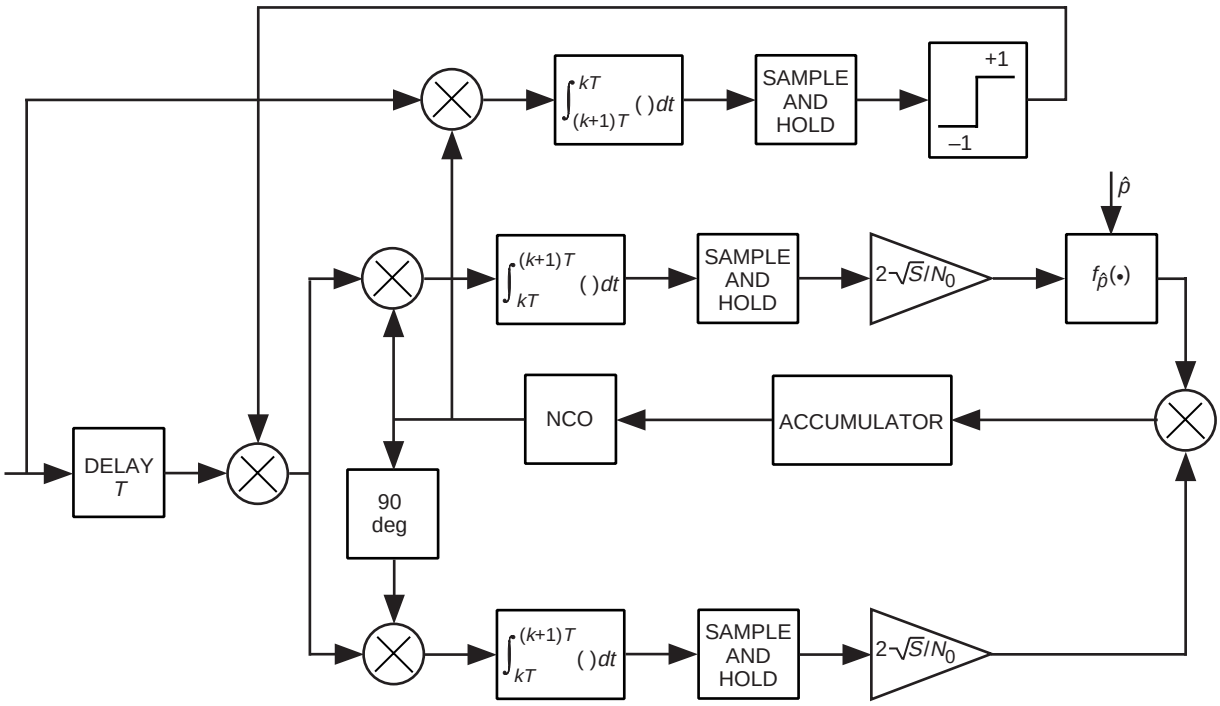

Fig. A-1. Information-reduced carrier synchronization loop for uncoded modulation.

Letting $y(t)$ denote the total received signal [i.e., the sum of Eqs. (1) and (2)], then analogous to the beginning of Section III, we can write the I and Q sample-and-hold outputs (again assuming $K_m = 1$ for convenience) as

$$\left.\begin{aligned} z_{ck} &\triangleq \int_{(k+1)T}^{(k+2)T} y(t-T) r_c(t)\,dt = \frac{2\sqrt{S}}{N_0}\left[T\sqrt{S} d_k \hat{d}_k \sin\phi - \hat{d}_k N_{sk}\sin\phi + \hat{d}_k N_{ck}\cos\phi\right] \\ z_{sk} &\triangleq \int_{(k+1)T}^{(k+2)T} y(t-T) r_s(t)\,dt = \frac{2\sqrt{S}}{N_0}\left[T\sqrt{S} d_k \hat{d}_k \cos\phi - \hat{d}_k N_{ck}\sin\phi - \hat{d}_k N_{sk}\cos\phi\right]; \\ & \qquad\qquad (k+2)\,T \le t \le (k+3)\,T \end{aligned}\right\} \quad \text{(A-1)}$$

where, similar to before, N_{ck} and N_{sk} are zero-mean, independent Gaussian random variables (r.v.'s) with variance $\sigma^2_{N_{ck}} = \sigma^2_{N_{sk}} = N_0T/2 \triangleq \sigma^2$ and now

$$\hat{d}_k = \text{sgn}\left[T\sqrt{S}d_k\cos\phi - N_{ck}\sin\phi - N_{sk}\cos\phi\right] \tag{A-2}$$

Note that $\hat{d}_k$ depends on N_{ck} and N_{sk}, which are the identical noise variables by which $\hat{d}_k$ is multiplied in Eq. (A-1). Thus, unlike the main text, it is no longer appropriate to characterize the r.v.'s $\hat{d}_kN_{sk}$ and $\hat{d}_kN_{ck}$ as Gaussian. Nevertheless, it is convenient to assign a notation to these r.v.'s, namely,

$$\left.\begin{aligned} M_{sk} &\triangleq \hat{d}_kN_{sk} = N_{sk}\,\text{sgn}\left[T\sqrt{S}d_k\cos\phi - N_{ck}\sin\phi - N_{sk}\cos\phi\right] \\ M_{ck} &\triangleq \hat{d}_kN_{ck} = N_{ck}\,\text{sgn}\left[T\sqrt{S}d_k\cos\phi - N_{ck}\sin\phi - N_{sk}\cos\phi\right] \end{aligned}\right\} \tag{A-3}$$

The output of the I-Q multiplier in Fig. 2, i.e., the error signal in the same transmission interval, is given by

$$\begin{aligned} z_{0k} \triangleq z_{ck}f_p\left(z_{sk}\right) = &\left[(2R_d)\,d_k\hat{d}_k\sin\phi - \frac{2\sqrt{S}}{N_0}M_{sk}\sin\phi + \frac{2\sqrt{S}}{N_0}M_{ck}\cos\phi\right] \\ &\times f_p\left((2R_d)\,d_k\hat{d}_k\cos\phi - \frac{2\sqrt{S}}{N_0}M_{ck}\sin\phi - \frac{2\sqrt{S}}{N_0}M_{sk}\cos\phi\right) \end{aligned} \tag{A-4}$$

Renormalizing as in Section II, we rewrite Eq. (A-4) as

$$\begin{aligned} z_{0k} = &\left[(2R_d)\,d_k\hat{d}_k\sin\phi - \sqrt{2R_d}Y_{sk}\sin\phi + \sqrt{2R_d}Y_{ck}\cos\phi\right] \\ &\times f_p\left((2R_d)\,d_k\hat{d}_k\cos\phi - \sqrt{2R_d}Y_{ck}\sin\phi - \sqrt{2R_d}Y_{sk}\cos\phi\right) \end{aligned} \tag{A-5}$$

where

$$\hat{d}_k = \text{sgn}\left[\sqrt{2R_d}d_k\cos\phi - X_{ck}\sin\phi - X_{sk}\cos\phi\right] \tag{A-6}$$

with normalized (unit variance) independent Gaussian noise r.v.'s $X_{sk} \triangleq N_{sk}/\sigma$, $X_{ck} \triangleq N_{ck}/\sigma$, and

$$\left.\begin{aligned} Y_{sk} &\triangleq \hat{d}_kX_{sk} = X_{sk}\,\text{sgn}\left[\sqrt{2R_d}d_k\cos\phi - X_{ck}\sin\phi - X_{sk}\cos\phi\right] \\ Y_{ck} &\triangleq \hat{d}_kX_{ck} = X_{ck}\,\text{sgn}\left[\sqrt{2R_d}d_k\cos\phi - X_{ck}\sin\phi - X_{sk}\cos\phi\right] \end{aligned}\right\} \tag{A-7}$$

The normalized S-curve is defined, analogous to Eq. (19), by

$$\eta(\phi) = \overline{\left(2R_d d_k \hat{d}_k \sin\phi\right) f_p\left(2R_d d_k \hat{d}_k \cos\phi - \sqrt{2R_d} Y_{ck} \sin\phi - \sqrt{2R_d} Y_{sk} \cos\phi\right)}^{X_{sk}, X_{ck}, d_k} \tag{A-8}$$

with slope at the origin

$$K_\eta = \frac{d\eta(\phi)}{d\phi}\,|_{\phi=0} = \overline{\left(2R_d d_k \hat{d}_k^{'}\right) f_p\left(2R_d d_k \hat{d}_k - \sqrt{2R_d} Y_{sk}^{'}\right)}^{Y_{sk}, Y_{ck}, d_k} \tag{A-9}$$

where

$$\left.\begin{aligned} \hat{d}_k^{'} &\triangleq \left.\hat{d}_k\right|_{\phi=0} = \operatorname{sgn}\left[\sqrt{2R_d} d_k - X_{sk}\right] = \operatorname{sgn}\left[2R_d d_k - \sqrt{2R_d} X_{sk}\right] \\ Y_{sk}^{'} &\triangleq \left. Y_{sk}\right|_{\phi=0} = X_{sk} \operatorname{sgn}\left[\sqrt{2R_d} d_k - X_{sk}\right] = X_{sk} \operatorname{sgn}\left[2R_d d_k - \sqrt{2R_d} X_{sk}\right] \end{aligned}\right\} \tag{A-10}$$

Using Eq. (A-10) in Eq. (A-9), we obtain, after some simplification,

$$K_\eta = 2R_d \overline{\operatorname{sgn}\left[2R_d - \sqrt{2R_d} X\right] f_p\left(\left|2R_d - \sqrt{2R_d} X\right|\right)}^{X} \tag{A-11}$$

where X is a zero-mean, unit variance Gaussian r.v.

The noise component of z_{0k} evaluated at $\phi = 0$ is

$$N_e(t) = \sqrt{2R_d} Y_{ck}^{'} f_p\left(2R_d d_k \hat{d}_k^{'} - \sqrt{2R_d} Y_{sk}^{'}\right) \tag{A-12}$$

where, analogous to the second equation in Eq. (A-10),

$$Y_{ck}^{'} \triangleq \left. Y_{ck}\right|_{\phi=0} = X_{ck} \operatorname{sgn}\left[\sqrt{2R_d} d_k - X_{sk}\right] = X_{ck} \operatorname{sgn}\left[2R_d d_k - \sqrt{2R_d} X_{sk}\right] \tag{A-13}$$

Note that $Y_{ck}^{'}$ is zero mean since X_{ck} and X_{sk} are independent zero-mean Gaussian r.v.'s. Also,

$$\left.\begin{aligned} E\left\{Y_{sk}^{'} Y_{ck}^{'}\right\} &= E\left\{X_{sk} X_{ck}\right\} = 0 \\ E\left\{\left(Y_{sk}^{'}\right)^2\right\} &= E\left\{\left(Y_{ck}^{'}\right)^2\right\} = 1 \end{aligned}\right\} \tag{A-14}$$

Thus, even though $Y_{sk}^{'}$ and $Y_{ck}^{'}$ are not Gaussian, they are nevertheless uncorrelated (but not necessarily independent). Fortunately, for our purposes in calculating squaring loss, the uncorrelated property is sufficient to establish what we wish. Furthermore, in view of the above, $N_e(t)$ is zero mean. Thus, as in the main text, we can compute the variance of $N_e(t)$ as

$$\sigma_{N_e}^2 = 2R_d E\left\{\left(Y_{ck}^{'}\right)^2 f_p^2\left(2R_d d_k \hat{d}_k^{'} - \sqrt{2R_d}Y_{sk}^{'}\right)\right\}$$

$$= 2R_d E\left\{\left(Y_{ck}^{'}\right)^2\right\}\overline{f_p^2\left(2R_d d_k \hat{d}_k^{'} - \sqrt{2R_d}Y_{sk}^{'}\right)}^{X_{sk},d_k}$$

$$= 2R_d \overline{f_p^2\left(2R_d d_k \hat{d}_k^{'} - \sqrt{2R_d}Y_{sk}^{'}\right)}^{X_{sk},d_k}$$

$$= 2R_d \overline{f_p^2\left(\left|2R_d - \sqrt{2R_d}X\right|\right)}^{X} \tag{A-15}$$

Finally, the squaring loss is obtained by substituting Eqs. (A-11) and (A-15) in Eq. (25) of the main text, which results in

$$S_L = \frac{\left(\overline{\operatorname{sgn}\left[2R_d - \sqrt{2R_d}X\right] f_p\left(\left|2R_d - \sqrt{2R_d}X\right|\right)}^{X}\right)^2}{\overline{f_p^2\left(\left|2R_d - \sqrt{2R_d}X\right|\right)}^{X}} \tag{A-16}$$

Evaluating Eq. (A-16) at the two extremes of $p = 1/2 (f_p(x) = \tanh x)$ and $p = 0 (f_p(x) = 1)$ results in Eqs. (27) and (33) [or equivalently Eq. (34)] of the main text, respectively. What is important to note is that here, for $p = 0$, Fig. 2 does not reduce to a PLL but rather to a polarity-type Costas loop! This occurs because of the full correlation of the data decisions with the additive input noise as discussed in connection with the definitions of M_{sk} and M_{ck} in Eq. (A-3). For any value of p between the above two extremes, the squaring-loss performance will lie between the curves corresponding to Eqs. (34) and (27), which are illustrated numerically in Fig. 3. Thus, in conclusion, the application of the information-reduction concept to an uncoded BPSK transmission results in no gain relative to conventional carrier synchronization techniques motivated by MAP theory.