A ROBUST PHASE DETECTION STRUCTURE FOR M-PSK: THEORETICAL DERIVATIONS, SIMULATION RESULTS, AND SYSTEM IDENTIFICATION ANALYSIS

Yair Linn

University of British Columbia

2111 Lower Mall, Vancouver, BC, Canada, V6T-1Z4

e-mail: ylinn@ece.ubc.ca

Abstract

In this paper we shall present a new Non Data Aided (NDA) adaptive phase detection structure for carrier synchronization Phase Locked Loops (PLLs) in M-ary Phase Shift Keying (M-PSK) receivers. The structure's principal novelty is in the fact that it dynamically adapts the phase detector's S-curve to allow the PLL to perform optimally at virtually any input Signal-to-Noise Ratio (SNR) at which it can sustain lock.

Investigation of the detector will commence with theoretical derivations of its S-curve, self-noise, and phase-error variance performance. These theoretical predictions will then be verified by computer simulations. As a third method of analysis, simulated PLLs employing the new detector will be investigated using the Steiglitz-McBride system identification algorithm, which will prove that those PLLs indeed perform optimally at every SNR. The theoretical, simulation, and system identification investigations of the proposed structure will be contrasted with results obtained using other NDA and Decision Directed (DD) phase detectors. As those comparisons will show, the proposed detection scheme offers greatly improved phase-error variance performance as well as superior resistance to fading and to Automatic Gain Control (AGC) imperfections. Moreover, the new detector has a very compact hardware implementation.

Keywords - M-PSK, phase, detector, PLL, adaptive

1. Introduction

Carrier synchronization PLLs in coherent M-PSK receivers are tasked with canceling the carrier phase error, an estimate of which is provided by a carrier Phase Detector (PD). There are two general categories of PDs: Non Data Aided (NDA) and Decision Directed (DD). The M^{th} -order nonlinearity detector ([3],[6],[13]) and the multiphase NDA Costas loop ([3],[5],[13]) are examples

of NDA phase detection. Examples of DD detectors can be found in [3],[6],[10]-[13], and [16].

An inherent problem of DD detectors is that at low SNRs (and, as well, during acquisition) they suffer from considerable self-noise due to erroneous decisions, something which also has an effect on their S-curves (see [6] Fig. 6-2, [10]-[12]). NDA detectors, while not susceptible to such a phenomenon, are nonetheless seldom used for higher order modulations. This is because at higher M s NDA detectors experience high self-noise (due to the high-order nonlinearities which they include) and their implementations are significantly more complicated than their Decision Directed counterparts (see for example [5] p. 74, [13] fig. 5.54).

An additional problem which afflicts the DD and NDA detectors just cited is that their gain is often strongly linked to the AGC circuit's operating point and performance. As we shall see in this paper, the fact that the gain of the phase detector is not constant implies that - unless this change of gain is compensated for in some manner - the carrier PLL's characteristics will change accordingly.

This paper's objective is the investigation of a robust NDA adaptive phase detector structure for M-PSK carrier synchronization. This structure will be shown to produce a constant-gain detector, which allows the carrier PLL to maintain optimality at virtually any SNR at which it can lock, throughout the whole gamut of E_s/N_0 values¹, AGC behaviors, and signal amplitude levels². The self-noise and phase-error variance performance of the proposed structure will be shown to be superior to that of other NDA and DD detectors. Moreover, unlike other NDA phase detectors, the proposed structure has a

¹ Throughout this paper, the terms SNR and E_s/N_0 refer to the same thing and are used interchangeably.

² The terms "signal level" or "signal amplitude level" must not be confused with the term E_s/N_0 ratio. The formers refer to the total signal+noise power at the inputs of the samplers in Fig. 1, while the latter term refers to the signal-to-noise ratio of that signal.

compact implementation for all M which is quite suitable for use within an FPGA or ASIC.

2. Signal and Receiver Models

We define the baseband signal as $m(t) \triangleq \sum_{n=-\infty}^{\infty} a_n p(t-nT)$, with $p(t)$ being the baseband

data pulse and $a_n = \exp(j\phi_n)$, $\phi_n = 2\pi \cdot \frac{m_n}{M}$, with $m_n \in \{0, 1, \dots, M-1\}$ being the actual data. The modulated signal is $s_m(t) \triangleq \text{Re}[m(t) \exp(j\omega_c t + j\theta_i)]$ and that signal is corrupted by Additive White Gaussian Noise (AWGN). Fig. 1 shows a simplified diagram of the M-PSK receiver under discussion. In Fig. 1:

- 1) $1/T$ is the both the symbol rate and the sample rate.
- 2) $n(t) \sim N(0, N_0 W)$ where W is the width of the bandpass IF filter (not shown).
- 3) K represents the physical gain associated³ with the circuit. In general, K is a slow function of time controlled by the AGC to achieve a desired signal level at the sampler inputs. A detailed discussion of K is presented in Appendix A.
- 4) When the carrier loop is locked around a stable equilibrium point, we have $\Delta\omega = 0$ and (since M-PSK carrier synchronization has an inherent M-fold phase ambiguity ([3],[6])) $\theta_o \in \{\theta_i + 2\pi k/M - \theta_e | k = 0, 1, \dots, M-1\}$, where $|\theta_e| < \pi/M$ is the residual phase error.
- 5) The matched filter $h(t)$ is assumed ideal and the sampling at the outputs of matched filters is considered to be at the ideal time (i.e. the symbol synchronization loop is assumed locked).

6) From [3] we have $E_s = \int_{-\infty}^{\infty} [p(t) \cos(\omega_c t + \theta_i)]^2 dt$
 $\approx \frac{1}{2} \int_{-\infty}^{\infty} p^2(t) dt = \frac{1}{2} E_p$. Without loss of generality,

we assume for convenience $E_p = 1$.

Fig. 1. Simplified diagram of M-PSK Receiver.

7) We assume a narrowband bandpass signal (i.e. $\omega_i \gg 1/T$) and that the Nyquist criterion for zero-ISI [3] is obeyed regarding the output of the matched filters.

8) Throughout this paper we make the standard assumption made in synchronization texts, i.e. that the carrier PLL is a high-loop-gain second-order system, with a fixed loop filter. Hence, the linearized-model Laplace transfer function of the PLL is

$$H(s) \triangleq \theta_o(s)/\theta_i(s) = \frac{2\zeta\omega_n \cdot s + \omega_n^2}{s^2 + 2\zeta\omega_n \cdot s + \omega_n^2} \text{ where } \zeta \text{ is the}$$

damping ratio and ω_n is the natural frequency.

3. Building Block #1 - a Lock Detector

To arrive at the proposed adaptive detector, we shall adopt a “bottom-up” approach. There are two complementing and related structures - an M-PSK lock detector and an M-PSK phase detector - which form the main building blocks. Here we treat the first of those, namely the lock detector.

In this section we shall review the derivations presented in [1]. As this is only a brief summary of the derivations presented there, the reader is urged to review reference [1] in order to attain a more complete understanding of the formulas presented here.

We define the phase of the received complex sample as $\phi_n \triangleq \tan^{-1}(Q(n)/I(n))$. From [1], we have

$$I(n) = K(2E_s \cos(-\Delta\omega \cdot nT + \theta_i - \theta_o + \phi_n) + n_I(nT))$$

and

$$Q(n) = K(2E_s \sin(-\Delta\omega \cdot nT + \theta_i - \theta_o + \phi_n) + n_Q(nT)),$$

with $n_I(nT), n_Q(nT) \sim N(0, 2N_0 E_s)$. When locked

ϕ_n thus reduces to:

³ It is assumed that K has the same value in both the I and Q arms (i.e. the arms are “matched” to each other).

$$\varphi_n = \tan^{-1} \left(\frac{\sin(\theta_e + 2\pi(m_n - k)/M) + n_Q(nT)/(2E_S)}{\cos(\theta_e + 2\pi(m_n - k)/M) + n_I(nT)/(2E_S)} \right). \quad (1)$$

Furthermore, in [1] Section IV it was found that the random process $\Delta\phi_n \triangleq \varphi_n - \theta_e$ has a Rician phase distribution, as follows:

$$p(\Delta\phi) \triangleq p(\Delta\phi_n = \Delta\phi) = \frac{1}{2\pi} \exp\left(\frac{-E_S}{N_0}\right) \times \left[1 + \sqrt{\frac{2E_S}{N_0}} \cos(\Delta\phi) \exp\left(\frac{E_S}{N_0} \cdot \cos^2(\Delta\phi)\right) \cdot \int_{-\infty}^{\cos(\Delta\phi)\sqrt{\frac{2E_S}{N_0}}} e^{-x^2/2} dx \right] \quad (2)$$

where $|\Delta\phi| \leq \pi$. It was found[1] that at high SNR we have

$$p(\Delta\phi) \stackrel{\text{high SNR}}{\approx} \sqrt{\frac{E_S}{N_0\pi}} \exp\left(-\frac{E_S}{N_0}(\Delta\phi)^2\right) \quad (3)$$

and that therefore:

$$\Delta\phi_n \stackrel{\text{high SNR}}{\sim} N(0, N_0/(2E_S)). \quad (4)$$

In [1] an auxiliary random process was introduced via

$$x_M(n) \triangleq \text{Re}[(I(n) + j \cdot Q(n))^M] / (I^2(n) + Q^2(n))^{\frac{M}{2}} \\ = \sum_{k=0}^{M/2} \binom{M}{2k} (-1)^k I^{M-2k}(n) Q^{2k}(n) / (I^2(n) + Q^2(n))^{\frac{M}{2}}.$$

Elementary rectangular-to-polar manipulations and the use of De Moivre's theorem[8] yield[1]:

$$x_M(n) = \frac{\text{Re}[(I(n) + jQ(n))^M]}{(I^2(n) + Q^2(n))^{\frac{M}{2}}} = \cos(M\varphi_n). \quad (5)$$

As clearly seen in (5) and as elaborated upon in [1], the value of $x_M(n)$ is independent of K and, hence, of the AGC.

Using $x_M(n)$, a new lock detector was defined as

$$\hat{l}_{M,N} \triangleq \frac{1}{2N} \sum_{n=-N+1}^N x_M(n). \quad \text{It was found in [1] that when}$$

the carrier loop is unlocked $E[\hat{l}_{M,N}] = 0$. Conversely, when locked, a given $E_S/N_0 = \chi$ we have ([1] Section IV):

$$\mu_L(\chi) \triangleq E[\hat{l}_{M,N}] = E[\cos(M\Delta\phi_n)] E[\cos(M\theta_e)] \\ = \left(\int_{-\pi}^{\pi} \cos(M\Delta\phi) p(\Delta\phi) d\Delta\phi \right) E[\cos(M\theta_e)]. \quad (6)$$

Due to $x_M(n)$'s independence vis-à-vis the AGC, the same is true for $\hat{l}_{M,N}$. Regarding the distribution of $\hat{l}_{M,N}$, a good approximation is (see [1] Section III):

$$\hat{l}_{M,N} | \text{locked} \sim N(\mu_L(E_S/N_0), 1/(2N)) \\ \hat{l}_{M,N} | \text{unlocked} \sim N(0, 1/(2N)) \quad (7)$$

Further analysis of $\hat{l}_{M,N}$ can be found in [1].

4. Building Block # 2 - a Phase Detector

4.1 Definition

The second building block of our adaptive scheme is a (non-adaptive) M-PSK phase detector. This detector can be thought of as a modification of the M^{th} -order nonlinearity detectors ([3],[6],[13]), which are $c_M(n) \triangleq \text{Im}[(I(n) + j \cdot Q(n))^M]$. The modified detector is defined as follows:

$$d_M(n) \triangleq \frac{\text{Im}[(I(n) + jQ(n))^M]}{(I^2(n) + Q^2(n))^{\frac{M}{2}}} \\ = \frac{\sum_{k=0}^{\frac{M-1}{2}} \binom{M}{2k+1} (-1)^k I^{M-2k-1}(n) Q^{2k+1}(n)}{(I^2(n) + Q^2(n))^{\frac{M}{2}}}. \quad (8)$$

The denominator term in (8) performs adaptive normalization on the numerator, and - as we shall see - this normalization makes $d_M(n)$ behave quite differently from $c_M(n)$.

4.2 Stochastic Analysis

It is easy to show, using De Moivre's theorem[8], that:

$$d_M(n) = \frac{(I^2(n) + Q^2(n))^{\frac{M}{2}} \sin(M\varphi_n)}{(I^2(n) + Q^2(n))^{\frac{M}{2}}} = \sin(M\varphi_n) \quad (9)$$

which can be expanded to:

$$d_M(n) = \sin(M\varphi_n) = \sin(M(\Delta\phi_n + \theta_e)) \\ = \cos(M\Delta\phi_n) \sin(M\theta_e) + \sin(M\Delta\phi_n) \cos(M\theta_e). \quad (10)$$

To find the expected detector output as a function of the phase error θ_e (i.e., the S-curve) we have:

$$\begin{aligned}
S_d(\theta_e) &\triangleq E[d_M(n)|\theta_e] = E[\sin(M(\Delta\phi_n + \theta_e))|\theta_e] \\
&= E[\cos(M\Delta\phi_n)]\sin(M\theta_e) + E[\sin(M\Delta\phi_n)]\cos(M\theta_e) \quad (11) \\
&= \left(\int_{-\pi}^{\pi} \cos(M\Delta\phi) p(\Delta\phi) d\Delta\phi \right) \sin(M\theta_e).
\end{aligned}$$

We thus find that the S-curve of $d_M(n)$ is sinusoidal with M stable equilibrium points⁴ and a maximal amplitude of $\int_{-\pi}^{\pi} \cos(M\Delta\phi) p(\Delta\phi) d\Delta\phi$. Note that since $p(\Delta\phi)$ is independent of K , this is also true of $S_d(\theta_e)$. To arrive at a “quick-and-dirty” approximation formula for the S-curve we assume high SNR and use (4) in conjunction with integral 15.73 of [8] to yield:

$$S_d(\theta_e) \approx \exp(-M^2 N_0 / (4E_s)) \sin(M\theta_e). \quad (12)$$

An assessment of the correspondence $\int_{-\pi}^{\pi} \cos(M\Delta\phi) p(\Delta\phi) d\Delta\phi \approx \exp(-M^2 N_0 / (4E_s))$ can be found in [1] Fig. 3.

4.3 Hardware Realization

Since $|d_M(n)| = |\sin(M\phi_n)| \leq 1$, $d_M(n)$ has an efficient fixed-point hardware implementation in the form of a lookup table. This is due to exactly the same reason [1] that enabled such an implementation for $x_M(n)$, namely the small dynamic range that is needed to express $d_M(n)$. Lookup table #2 in Fig. 2 is an example of this implementation.

To see why the existence of such an implementation is significant, we make note of the fact that other phase detectors suffer from a large dynamic range that often renders a similar implementation unfeasible. To highlight this point, consider the M^{th} -order nonlinearity $c_M(n)$. It is easily seen that $c_M(n) \propto K^M$, which means that a phase detector lookup table and the ensuing datapath (in particular, the loop filter) must all be able to handle the dynamic range of K^M . This is often prohibitive to implement in fixed-point hardware (see Appendix A). Moreover, the dependence on K^M implies a nonlinear dependence upon the AGC, a dependence that $d_M(n)$ does not exhibit (so long as the AGC ensures that the samplers are not overdriven nor severely underdriven). A similar conclusion can be reached with regards to Decision Directed detectors, which are defined by ([10]-

⁴ This is consistent with the M -fold ambiguity present in all M-PSK phase synchronization circuits ([3],[6],[13]).

Fig. 2. Hardware implementation of $V_{M,N}(n)$.

[12]) $DD_M(n) \triangleq I(n) \cdot \hat{Q}(n) - Q(n) \cdot \hat{I}(n)$ (where $\hat{I}(n)$ and $\hat{Q}(n)$ are decisions on the I and Q channels). Simple substitution shows that $DD_M(n) \propto K$, so use of $DD_M(n)$ means that a dependence upon the dynamic range of K and the AGC would still exist.

Indeed, we see that d_M has many merits (more of which will be outlined in Section 6.1.1), and it is an excellent M-PSK phase detector that can be used instead of c_M or DD_M . However, even better performance can be achieved by using d_M within a simple adaptive phase detection structure that is introduced in the next section.

5. A Robust Adaptive Phase Detector

Using $\hat{l}_{M,N}$ and $d_M(n)$ we now define the adaptive phase detection structure whose analysis is the primary goal of this paper. In this section, we confine ourselves to derivation of the S-curve of the adaptive detector. Performance analysis is relegated to subsequent sections.

5.1 Definition and Structure

We define the adaptive phase detector as:

$$V_{M,N}(n) \triangleq \begin{cases} (1/M) \cdot (d_M(n) / \hat{l}_{M,N}) & \text{Carrier is locked} \\ (1/M) \cdot (d_M(n) / \tilde{\alpha}_{SNR}) & \text{Carrier is unlocked} \end{cases} \quad (13)$$

where $\tilde{\alpha}_{SNR}$ is a constant that is discussed in Section 7.3. A block diagram of a fixed-point hardware implementation structure for $V_{M,N}(n)$ is presented in Fig. 2. Observe how the division by $2N$ is avoided, where it is assumed that $2N$ is a power of 2.

5.2 Comments Regarding the Hardware Implementation

Regarding Fig. 2, the implementation of lookup table #1 was discussed in Section 3 and in [1], and the implementation of lookup table #2 was discussed in Section 4. There it was established that lookup tables #1 and #2 can be efficiently realized in fixed-point hardware. With regards to lookup table #3, we observe that the lowest value of $\hat{l}_{M,N}$ that need be handled is the lock decision threshold value, since below this value $\tilde{\alpha}_{SNR}$ is used. Hence the largest absolute value that need be accommodated by lookup table #3 is $\max(|d_M(n)|)/(M \cdot \min(\tilde{\alpha}_{SNR}, \Gamma)) = 1/(M \cdot \min(\tilde{\alpha}_{SNR}, \Gamma))$, with Γ being the lock threshold. Typically, $\tilde{\alpha}_{SNR} \geq \Gamma$ (see Section 7.3) and neither parameter would be less than about 0.04, since below that value the SNR is so low (see [1] Fig. 3) that there is scarcely a hope of the carrier loop locking without first applying a coarse (=open loop) frequency correction. This means that $1/(M \cdot \min(\tilde{\alpha}_{SNR}, \Gamma)) \leq 25/M$, which implies that dynamic range of the data in lookup table #3 can be sufficiently limited to allow for its compact fixed-point hardware implementation. Even these requirements upon lookup table #3 can be relaxed because the phase detector operates within a closed loop, so if the phase detector is clamped to within reasonable values this will have only a marginal effect on the closed-loop system performance. This is because successive phase detector outputs will progressively compensate for any undercorrection caused by clamping. Quantitative investigation of such clamping effects is beyond the scope of this paper.

5.3 S-Curve Determination

First, we assume that the loop is locked (unlocked behaviour is treated in Section 7.3). We then have from (10) and (13):

$$\begin{aligned} V_{M,N}(n) &= \frac{d_M(n)}{M \cdot \hat{l}_{M,N}} = \frac{\sin(M\phi_n)}{M \cdot \hat{l}_{M,N}} \\ &= \frac{\cos(M\Delta\phi_n)}{M \cdot \hat{l}_{M,N}} \sin(M\theta_e) + \frac{\sin(M\Delta\phi_n)}{M \cdot \hat{l}_{M,N}} \cos(M\theta_e). \end{aligned} \quad (14)$$

To find the S-curve of $V_{M,N}(n)$ when locked, we choose to treat $\hat{l}_{M,N}$ as a constant. This is justified by noting that the rate of changes in the value of $\hat{l}_{M,N}$ is

significantly lower than the rate of change of $d_M(n)$ (it is lower by a factor of $2N$, which typically [1] would be at least in the order of 100). With that assumption and since $E[\sin(M\Delta\phi_n)] = 0$, we have from (14):

$$\begin{aligned} S_V(\theta_e) &\triangleq E[V_{M,N}(n)|\theta_e] \\ &= E[\cos(M\Delta\phi_n)] \sin(M\theta_e) / (M \cdot \hat{l}_{M,N}). \end{aligned} \quad (15)$$

6. Closed Loop Locked-State Analysis

It is easy to see, even from a cursory inspection of (13), that the adaptive detector $V_{M,N}$ will exhibit behaviour that is tightly linked to that of d_M . In this section, therefore, we shall first provide a theoretical analysis of d_M , and then use those results to develop conclusions regarding $V_{M,N}$. These theoretical predictions will then be compared to theoretical calculations that pertain to other phase detectors. To conclude this section, closed-loop nonlinear model simulations will be used to verify the theoretical analysis. In all the derivations of this section we assume a locked loop (i.e., acquisition-related transient effects are not treated).

6.1 Analysis of the Linearized PLL Model

PLLs are inherently nonlinear beings [9], an issue which presents great obstacles when theoretical analysis is attempted. In order to nonetheless arrive at theoretical predictions, a standard approach adopted by synchronization texts has been to assume that the PLL is locked and then analyze the linearized model of the loop. This will be the approach adopted in this subsection.

First, we define the *linearized gain* or *tracking-mode gain* of a detector P as $g_{L,P} \triangleq (\partial S_P(\theta_e) / \partial \theta_e) \big|_{\theta_e=0}$

where $S_P(\theta_e)$ is the detector's S-curve. For a given phase detector P we have [12]:

$$\text{Var}(N_{e,P}) = (N_0 / 2E_S) \cdot \xi_P \quad (16)$$

$$\begin{aligned} \text{Var}(\theta_e) &= \text{Var}(N_{e,P}) / \alpha_{SNR,P}^2 \cdot 2B_L \cdot T \\ &= (N_0 / 2E_S) \cdot \xi_P / \alpha_{SNR,P}^2 \cdot 2B_L \cdot T \end{aligned} \quad (17)$$

where:

- $B_L = \frac{1}{2} \omega_n (\zeta + 1 / (4\zeta))$ is the loop's *noise bandwidth*⁵

⁵ This definition contains a factor of $\frac{1}{2}$ w.r.t. the definition in [12]. This factor has been inserted in order to make the definition of B_L compatible with the (arguably) more widely used definition, as in [6]

- ξ_p is the phase detector's *self-noise* [12]
- $N_{e,p}(n)$ is the *equivalent loop noise* ([11],[12])
- $\alpha_{SNR,P}$ is the *amplitude suppression factor* ([10]-[12]). This factor is *the multiplicative factor by which the expected linearized gain of the detector is reduced due to the presence of additive noise at its inputs, as compared to the phase detector's expected linearized gain for noiseless inputs*. From this definition it is clear that at $E_S / N_0 = \chi$:

$$\alpha_{\chi,P} = (g_{L,P} | E_S / N_0 = \chi) / (g_{L,P} | E_S / N_0 = \infty) \quad (18)$$

The discussion in [12] as well as (implicitly) eq. (18) assume that K is constant (in [12] the assumption made is that $K=1$ identically). As seen in Appendix A, this is an unrealistic assumption that does not take into account the AGC's operation. Here we shall model the AGC's effects by defining an *effective amplitude suppression factor*, which we name $\beta_{SNR,P}$. This factor for a given phase detector P is the *multiplicative factor by which the expected linearized gain of the phase detector is reduced due to the presence of additive noise and variations of K at its inputs, as compared to the phase detector's expected linearized gain for noiseless inputs and $K=1$* . Let us assume that K is a function of the SNR, i.e. that at $E_S / N_0 = \chi$ we have $K = \Upsilon_{AGC}(\chi)$ where $\Upsilon_{AGC}(\chi)$ is (as the subscript suggests) determined by the AGC. This definition implies that at $E_S / N_0 = \chi$:

$$\beta_{\chi,P} = \frac{g_{L,P} | E_S / N_0 = \chi \text{ and } K = \Upsilon_{AGC}(\chi)}{g_{L,P} | E_S / N_0 = \infty \text{ and } K = 1} \quad (19)$$

When we wish to model AGC effects, we simply substitute $\beta_{SNR,P}$ instead of $\alpha_{SNR,P}$ in (17), yielding:

$$\text{Var}(\theta_e) = (N_0 / 2E_S) \cdot \xi_P / \beta_{SNR,P}^2 \cdot 2B_L \cdot T. \quad (20)$$

The *M-phase power loss* of a given phase detector P refers to the *multiplicative factor that quantifies the additional phase error jitter that is incurred because of noise adjoining to the signal that enters the nonlinearity constituting the phase detector* (see [3] eq. 6.2-59, [10]-[12]). In many texts this quantity is called the "squaring loss", a convention that, for succinctness, will also be adopted in this paper. In terms of previous definitions, the squaring loss is $\Omega_P \triangleq \xi_P / \alpha_{SNR,P}^2$, or $\Omega_P \triangleq \xi_P / \beta_{SNR,P}^2$ if modeling of the AGC's effects is required.

and [9]. However, note that (17) incorporates a factor of 2 (w.r.t. the corresponding equation in [12]), so the aforementioned factor of $\frac{1}{2}$ does not influence the phase-error variance results.

6.1.1 Linear Modeling of $d_M(n)$.

Let us now resume our investigation $d_M(n)$ by computing $g_{L,d}$, $N_{e,d}(n)$, ξ_d , $\alpha_{SNR,d}$ and $\beta_{SNR,d}$. First, using (11), we have that the linearized gain of $d_M(n)$ is

$$g_{L,d} \triangleq (\partial S_d(\theta_e) / \partial \theta_e) |_{\theta_e=0} = M \cdot E[\cos(M\Delta\phi_n)] \quad (21)$$

From inspection of (21) we see that:

$$\begin{aligned} \beta_{SNR,d} &= \alpha_{SNR,d} = E[\cos(M\Delta\phi_n)] \\ &= \int_{-\pi}^{\pi} \cos(M\Delta\phi) p(\Delta\phi) \cdot d(\Delta\phi). \end{aligned} \quad (22)$$

Note that $0 \leq \alpha_{SNR,d} \leq 1$ and $0 \leq \beta_{SNR,d} \leq 1$. The fact that $\beta_{SNR,d} = \alpha_{SNR,d}$ is unsurprising, considering that we have already established that $d_M(n)$ is independent of K .

To find $N_{e,d}(n)$ we assume, for convenience, that the loop is operating around the equilibrium point corresponding to $M\theta_e \approx 0$, whereupon from (10):

$$\begin{aligned} d_M(n) &= \sin(M\phi_n) \\ &\approx \cos(M\Delta\phi_n) \cdot M\theta_e + \sin(M\Delta\phi_n) \\ &= M \left(\cos(M\Delta\phi_n) \theta_e + \frac{1}{M} \sin(M\Delta\phi_n) \right). \end{aligned} \quad (23)$$

This corresponds to the output of a linear, unity gain phase detector degraded by a factor of $\alpha_{SNR,d} = E[\cos(M\Delta\phi_n)]$, which is corrupted by additive noise $N_{e,d}(n) = \frac{1}{M} \sin(M\Delta\phi_n)$, and this sum is multiplied by M . This model is included in Fig. 4 (note: the fixed multiplicative factor of M can be compensated by design in the loop filter's gain. Hence, this factor is incorporated into K_a in the loop filter in Fig. 4). Observe that $N_{e,d}(n)$ is *not* Gaussian, though it may be approximated by Gaussian noise at high E_S / N_0 , since (using (4)) we have

$$N_{e,d}(n) \xrightarrow{E_S / N_0 \rightarrow \infty} \Delta\phi_n \underset{\text{high } E_S / N_0}{\sim} N(0, N_0 / (2E_S)).$$

The self-noise of $d_M(n)$ is then (from eq. (16)):

$$\begin{aligned} \xi_d &= 2E_S / N_0 \cdot \text{Var}(N_{e,d}) \\ &= 2E_S / N_0 \cdot \int_{-\pi}^{\pi} \left(\frac{1}{M} \sin(M\Delta\phi) \right)^2 p(\Delta\phi) \cdot d(\Delta\phi). \end{aligned} \quad (24)$$

Quantitative judgments regarding the squaring loss of d_M can be reached by comparing that loss to that which is incurred when c_M or DD_M is used. Such a comparison is shown in Fig. 3. Linear-model predictions

of closed-loop phase-error variance of d_M can be made via (17). Results are shown in Fig. 4, with $\zeta=0.95$ and $2B_L \cdot T$ fixed⁶ at all SNRs at $2B_L \cdot T=0.01$, and it is seen that d_M regularly provides superior performance, in particular compared to c_M . Also plotted in Fig. 4 is the Cramér-Rao bound[6] $CRB=(N_0/2E_S) \cdot 2B_L \cdot T$. The necessary data required to plot the results for c_M and DD_M was obtained from Table 1.

6.1.2 Linear Modeling of $V_{M,N}(n)$.

From (15) it follows that:

$$g_{L,V} \triangleq \partial S_V(\theta_e) / \partial \theta_e \big|_{\theta_e=0} = \left[\frac{\alpha_{SNR,d} M \cos(M\theta_e)}{M \cdot \hat{l}_{M,N}} \right]_{\theta_e=0} = \frac{\alpha_{SNR,d}}{\hat{l}_{M,N}}. \quad (25)$$

Inspection of (25) reveals that:

$$\beta_{SNR,V} = \alpha_{SNR,V} = \alpha_{SNR,d} / \hat{l}_{M,N}. \quad (26)$$

It is unsurprising that $\beta_{SNR,V} = \alpha_{SNR,V}$, since both the building blocks of $V_{M,N}(n)$, namely $d_M(n)$ and $\hat{l}_{M,N}$, are independent of K . When in lock, it is shown in [1] Section IV that at practical loop bandwidths we have to a good approximation $E[\cos(M\theta_e)] \approx 1$. Therefore, we have from (6) that $\hat{l}_{M,N} \approx E[\cos(M\Delta\phi_n)] = \alpha_{SNR,d}$, and hence it follows from (26) that $g_{L,V} \approx 1$ and $\beta_{SNR,V} = \alpha_{SNR,V} \approx 1$. It should be stressed that, to a good approximation, we have that $g_{L,V}$, $\alpha_{SNR,V}$, and $\beta_{SNR,V}$ are unity independent of (a) the value of K and the AGC's performance, and (b) the SNR and M (to be verified by simulations in Section 8).

Once again treating $\hat{l}_{M,N}$ as a constant (see Section 5.3), we see that the only random process in (13) is $d_M(n)$. Thus, the squaring loss of $V_{M,N}(n)$ should be identical to that of $d_M(n)$ (as seen in Fig. 3). Furthermore, with regards to phase-error variance performance, $V_{M,N}(n)$ should perform just like $d_M(n)$ (as predicted in Fig. 4), provided (of course) that the loops employing each detector have the same ω_n and ζ when phase-error variance comparisons are made. The

Fig. 3. Squaring loss as a function of E_s/N_0 .

Fig. 4. Calculated phase-error variance, using linearized baseband model. The loop bandwidth is held fixed at $2B_L \cdot T=0.01$. AGC effects are ignored. $\zeta=0.95$.

heuristic arguments just given are easy to formally verify, and hence those derivations are omitted.

The preceding discussion regarding the performance of $V_{M,N}$ is verified by nonlinear model simulations undertaken in Section 7.2. In Table 1 we have summarized the linearized-model parameters for the phase detectors discussed thus far. Fig. 5 shows plots of α_{SNR} for the phase detectors discussed in this paper. Plots of β_{SNR} , assuming the example AGC parameters of Appendix A, are given in Fig. 6. When we compare Fig. 5

⁶ See Section 6.2 for a discussion of how and why a fixed noise bandwidth is maintained.

to Fig. 6, we can see that the AGC's effect upon DD_M and c_M is quite pronounced.

6.2 Nonlinear Model Simulation Results

In order to verify the linear-model phase-error variance predictions of Fig. 4, nonlinear simulations were conducted using the model presented in Fig. 7. Note that the noise bandwidth B_L in Fig. 4 and Fig. 7 is maintained fixed by selecting the loop filter DC gain so that $2B_L T = 0.01$. Consider for example the PLLs which employ DD_M : if the loop gain needed to achieve $2B_L T = 0.01$ in noiseless conditions is $K_{a,DD}(E_S/N_0 = \infty)$, then (in order to maintain $2B_L T = 0.01$) the loop filter DC gain at a given $E_S/N_0 = \chi$ needs to be $K_{a,DD}(E_S/N_0 = \chi) = K_{a,DD}(E_S/N_0 = \infty) / \alpha_{\chi,DD}$ (see [10]-[12]). Maintaining a constant loop bandwidth is the practice adopted in most synchronization texts (e.g. [6],[13]), because it facilitates a meaningful comparison of the phase-error variance achievable by the compared detectors when employed in PLLs which have identical parameters. Yet, when inspecting Fig. 4 and Fig. 7, it is important to realize that since the loop filter's gain is different at each SNR, the results for a given phase detector cannot be obtained using a *single* PLL with a *fixed* loop filter. The correct way to interpret Fig. 4 and Fig. 7 (and similar graphs in texts such as [6] and [13]) is, rather, by considering the results at each SNR as if they were obtained by measurements on PLLs *unique to that* E_S/N_0 that were *optimized for operation at that* E_S/N_0 to yield $\zeta = 0.95$ and $2B_L \cdot T = 0.01$ at *that* E_S/N_0 . Indeed, a single PLL using $c_M(n)$, $DD_M(n)$ or $d_M(n)$ and a fixed loop filter cannot maintain a constant loop bandwidth and damping ratio over the entire SNR range, an issue that shall be discussed further in Section 7. Finally, recall that we have already noted in the previous subsection that $V_{M,N}$ should have the same phase-error variance characteristics of d_M , as displayed in Fig. 4 and Fig. 7. But, as a crucial difference of $V_{M,N}$ vis-à-vis d_M (and, indeed, vis-à-vis c_M and DD_M), Section 7 will show that $V_{M,N}$ can achieve the results predicted in those figures, at every SNR, using a single PLL with a fixed loop filter.

Fig. 5. α_{SNR} for various detectors.

Fig. 6. β_{SNR} for various detectors, assuming the AGC that is described in Appendix A. Also plotted is the AGC curve (i.e. $K = Y_{AGC}(E_S/N_0)$).

7. What is the Novelty of $V_{M,N}(n)$?

7.1 Locked State Analysis

We have shown in Section 6 that, during lock, we have for $V_{M,N}$ that $\beta_{SNR,V} = \alpha_{SNR,V} \approx 1$. To understand why this is significant, all one has to do is observe how the gain of a phase detector affects the PLL's loop-gain and, consequently, the PLL's behaviour. The most straightforward method with which one can arrive at a quantitative appreciation of this phenomenon is by studying the effect of the phase detector gain upon the PLL's damping ratio ζ and its natural frequency ω_n . Changes in ζ and ω_n will bring about corresponding changes in the PLL's noise bandwidth

Table 1. Comparison of important tracking-mode phase detector characteristics

PD	$Var(N_e) = (N_0/2E_S) \cdot \xi$	Linearized Gain	α_{SNR}	β_{SNR}
$c_M(n)$	$\frac{N_0}{2E_S} \cdot \frac{\left((M-1)! \sum_{i=0}^{M-1} \binom{M}{i} \frac{(E_S/N_0)^i}{i!} \right)}{\left(M (E_S/N_0)^{M-1} \right)}$ Source: ξ from eq. (17) in [15]	$g_{L,c} = M \cdot K^M$ Source: [6] eq. 6-116	$\alpha_{SNR,c} = 1$	$\beta_{SNR,c} = K^M$
$DD_M(n)$	See eq. (3) in [12]	$g_{L,DD} = K \cdot \alpha_{SNR,DD}$ Source: [11]	$\alpha_{SNR,DD}$ from eq. (10),(11), (32) in [11]	$\beta_{SNR,DD} = K \cdot \alpha_{SNR,DD}$
$d_M(n)$	$\int_{-\pi}^{\pi} \left(\frac{1}{M} \sin(M\Delta\phi) \right)^2 p(\Delta\phi) d(\Delta\phi)$	$g_{L,d} = M \cdot \alpha_{SNR,d}$ Source: eq. (21)	$\alpha_{SNR,d} = E[\cos(M\Delta\phi_n)]$	$\beta_{SNR,d} = \alpha_{SNR,d}$
$V_{M,N}(n)$	$\frac{\int_{-\pi}^{\pi} \left(\frac{1}{M} \sin(M\Delta\phi) \right)^2 p(\Delta\phi) d(\Delta\phi)}{\left[\int_{-\pi}^{\pi} \cos(M\Delta\phi) p(\Delta\phi) d(\Delta\phi) \right]^2}$	$g_{L,V} = 1$ Source: eq. (25)	$\alpha_{SNR,V} = 1$	$\beta_{SNR,V} = 1$

Notes: (a) Results for $V_{M,N}(n)$ assume $\hat{I}_{M,N} \approx \alpha_{SNR,d}$ (b) While $Var(N_{e,V}) \geq Var(N_{e,d})$, this does not result in an increase in phase-error variance, as substitution of the appropriate variables into (17) and/or (20) shows.

$B_L = \frac{1}{2} \omega_n (\zeta + 1/(4\zeta))$, its settling time $T_{set} \approx 2\pi/\omega_n$, its lock range $\Delta\omega_L \approx 2\zeta\omega_n$, pull-out range $\Delta\omega_{PO} \approx 1.8\omega_n (\zeta + 1)$, and its cycle-slip statistics (see [4], [6],[9],[13]-[14]). Hence, by quantifying the effects of the phase detector gain upon ζ and ω_n , we can easily quantify the effects upon the rest of the PLL's parameters.

Suppose that the PLL uses a given phase detector $P(n)$, and that in the absence of noise the natural frequency is $\bar{\omega}_n$ rads/sec and the damping ratio is $\bar{\zeta}$. Without loss of generality we assume that $K=1$ at noiseless conditions. At a given E_S/N_0 of χ the amplitude suppression factor of the phase detector (accounting for AGC effects) is $\beta_{\chi,P}$, and it can be shown (for example see eq. (15) in [11]) that at $E_S/N_0 = \chi$ the natural frequency $\omega_{n\chi}$ and damping ratio ζ_χ will be

$$\omega_{n\chi} = \bar{\omega}_n \sqrt{\beta_{\chi,P}} \text{ and } \zeta_\chi = \bar{\zeta} \sqrt{\beta_{\chi,P}}. \quad (27)$$

Hence (unless $\beta_{SNR,P}$ is constant) we see that ω_n and ζ change as a function of the SNR. To address the above problem, an engineering approximation is often adopted, which is to optimize the loop for operation at a certain

$E_S/N_0 = \lambda$ with the optimal parameters $\omega_{n\lambda}$ and ζ_λ . For example, λ might be the lowest SNR for which the error correction decoder provides an acceptable coding gain. It then immediately follows from (27) that

$$\omega_{n\chi} = \omega_{n\lambda} \sqrt{\beta_{\chi,P}/\beta_{\lambda,P}} \text{ and } \zeta_\chi = \zeta_\lambda \sqrt{\beta_{\chi,P}/\beta_{\lambda,P}}. \quad (28)$$

Thus (if we assume $\beta_{SNR,P}$ increases monotonically vs. E_S/N_0) we conclude from (28) that for $\chi > \lambda$ we have higher-than-optimal ω_n and ζ , and for $\chi < \lambda$ we have lower-than-optimal ω_n and ζ . Only at $\chi = \lambda$ does the PLL perform as desired. It thus follows that to achieve the same ω_n and ζ at all SNR and for all K , the phase detector must have an effective amplitude suppression factor which is identically unity. This is just the case for $V_{M,N}$, since $\beta_{SNR,V} \approx 1$. Hence, using $V_{M,N}$, the same carrier synchronization loop can be used to demodulate any M-PSK signal and achieve optimal performance at all E_S/N_0 ratios, signal-levels (i.e. all K), AGC behaviours, and M s. The only change in the loop structure when changing the modulation index M would be in the contents of the lookup tables which compute $d_M(n)$, $x_M(n)$, and $V_{M,N}(n)$.

Nonetheless, we must ensure that the delay $2N \cdot T$ incurred during the computation of $\hat{I}_{M,N}$ does not substantially impact the validity of the approximation $\hat{I}_{M,N} \approx \alpha_{SNR,d}$ when that lock detector value is used to compute $V_{M,N}(n)$. In Section 9, it is shown that a relatively small N is required to achieve good performance over most practical SNR ratios, so the delay may well be inconsequential for most systems (in particular those which have a high symbol rate as compared to the fading rate). One should also not forget that a further constraint exists for N , namely that it be determined so that the desired lock and false-alarm probabilities for $\hat{I}_{M,N}$ are attained (see [1]). If this constraint conflicts with those of Section 9, the Integrate-and-Dump module in Fig. 2 can be duplicated, with a different N being used in each module. One module would be used to generate $\hat{I}_{M,N}$ for lock detection (and to drive the “sel” input to the MUX), while the other would be used to drive the “1” input to the MUX.

7.2 Phase-Error Variance Comparison

To validate the previously made predictions regarding the phase-error variance performance of $V_{M,N}(n)$ (see end of Section 6.2), computer simulations were conducted using the nonlinear equivalent baseband model, assuming the AGC curve of Appendix A. The results are shown in Fig. 8. As is evident in that figure, the AGC has a profound effect upon c_M and DD_M and causes the phase-error variance to visibly increase at high-SNR. This is due to the changes in ω_n and ζ as described in the previous subsection, changes which also manifest themselves at low SNR (see Section 8). For the same reasons, the CRBs for c_M and DD_M become curved (for clarity those CRB curves are omitted from Fig. 8).

7.3 Unlocked State Operation

Since $\hat{I}_{M,N} \approx 0$ in the unlocked state (see [1]), an appropriately valued constant $\tilde{\alpha}_{SNR}$ must be substituted for $\hat{I}_{M,N}$ in (13) in order to facilitate carrier acquisition. $V_{M,N}(n)$ will then exhibit behavior identical to that of $d_M(n)$ during acquisition, as it is simply $d_M(n)$ multiplied by the constant $1/(M \cdot \tilde{\alpha}_{SNR})$. The effective phase detector gain is then $g_{L,V} = g_{L,d} / (M \cdot \tilde{\alpha}_{SNR}) = \alpha_{SNR,d} / \tilde{\alpha}_{SNR}$. In order to maintain the optimal loop parameters during acquisition,

Fig. 7. Simulated phase-error variance, using equivalent baseband nonlinear model. Loop bandwidth is held fixed at $2B_L \cdot T = 0.01$. AGC effects are ignored. $\zeta = 0.95$.

Fig. 8. $Var(\theta_e)$, obtained via nonlinear-model simulations including AGC effects. The loops for DD_M and c_M are optimized for input SNRs of 1, 4, 15, and 17 dB, for $M = 2, 4, 8$, and 16, respectively, where at those SNRs we desire to have $\zeta = 0.8$ and $\omega_n = 0.011/T$. For $V_{M,N}$, $\zeta = 0.8$ and $\omega_n = 0.011/T$ throughout. The arrows aid in finding the result for $V_{M,N}$ at the optimization SNRs (those data points are also circled). For $V_{M,N}$, $N = 256$. Note that DD_M appears to outperform $V_{M,N}$ at low SNR; however this is a fallacy, since this apparent advantage is due to the reduction of ω_n , ζ , and B_L in the PLL using DD_M . See Section 8.

we strive to have $g_{L,V}=1$, which implies that we should aspire to have $\tilde{\alpha}_{SNR}=\alpha_{SNR,d}$. Thus, as a general rule, an algorithm for deciding upon an appropriate $\tilde{\alpha}_{SNR}$ would try, for example, to determine the latter either by (in order of increasing complexity): (a) using a worst case value (i.e. the value of $\alpha_{SNR,d}$ for the lowest SNR for which operation is desired), (b) using the last measured value of $\hat{l}_{M,N}$ while locked (because $E[\hat{l}_{M,N}] \approx \alpha_{SNR,d}$), or (c) using some E_S/N_0 estimation technique (such as measurements on an auxiliary pilot signal) to estimate $\alpha_{SNR,d}$. Methods (a) and (b) are quite simple to implement, though it is important to note that performance during acquisition is only partially addressed by using a constant $\tilde{\alpha}_{SNR}$, due to the fact that, since $\alpha_{SNR,d}$ varies with the SNR yet $\tilde{\alpha}_{SNR}$ is constant, $g_{L,V}=\alpha_{SNR,d}/\tilde{\alpha}_{SNR}$ will vary vs. the SNR.

8. System Identification Analysis

To get a qualitative feel of the operation of $V_{M,N}$, Fig. 9 compares the step response for carrier PLLs which use DD_M , c_M , and $V_{M,N}$. The upper subplot of that figure shows the output phase trajectory for a single input data set for each SNR. Because of the input noise, it may be difficult to adequately distinguish the system response from a single data set, especially⁷ for the lower SNRs. This difficulty is overcome by using several data sets to drive the systems at each SNR, and for each SNR the measured responses are then averaged. It then becomes easy to discern the systems' responses. This is shown in the bottom subplot of Fig. 9, where it is seen that the system responses obtained by using $V_{M,N}$ are virtually identical at all SNRs, while, in contrast, the responses obtained by using c_M and DD_M are strongly dependent upon the SNR.

We can also arrive at quantitative system identification results by using the Steiglitz-McBride[18] system identification algorithm. To do this, at each SNR we average the PLLs' responses for a sufficient number of input data sets, that is, until the averaged response curves are sufficiently noise-free (much like we did in order to arrive at the bottom plot of Fig. 9). Then we can use the

⁷ Note that in the top subplot of Fig. 9 it appears that the response for $V_{M,N}(n)$ at $E_S/N_0=0$ dB is much noisier than that of the other detectors; but this is because the loop bandwidth for c_M and DD_M decreases considerably at low SNR (see Fig. 10).

Fig. 9. Simulated responses of carrier PLLs to a phase step of $\theta_i(n)=\Delta\theta_i \cdot u(nT-t_0)$, where $u(t)$ is the unit step function and $t_0=4100 \cdot T$. Modulation index is $M=4$ (QPSK). The loops are optimized for $E_S/N_0=10$ dB, where at that SNR we desire $\zeta=0.8$ and $\omega_n=9 \cdot 10^{-4}/T$. The upper subplot is the response obtained from a single input data set. The bottom subplot is the average of the responses obtained from 100 input data sets. For $V_{M,N}$, $N=2048$ was used. We assume that K behaves according to the example AGC described in Appendix A.

Fig. 10. Predicted and measured performance of DD_M , c_M , and $V_{M,N}$ over different E_S/N_0 ratios. Modulation index is $M=4$ (QPSK). For $V_{M,N}$, $N=2048$ was used. PLLs were designed to give $\zeta=0.8$ and $\omega_n=9 \cdot 10^{-4}/T$ at an E_S/N_0 ratio of 10 dB. We assume that K behaves according to the example AGC described in Appendix A.

Steiglitz-McBride algorithm upon the smoothed response in order to estimate ω_n and ζ . The results attained by following such a procedure are shown in Fig. 10. Also shown in Fig. 10 are curves of the expected ζ and ω_n as predicted in (28) for DD_M and c_M . As was predicted in Section 7, $V_{M,N}$ provides the desired damping ratio and natural frequency over the entire input SNR range; DD_M and c_M do not, and the parameters of their PLLs change according to (28).

9. Bounds on N to Ensure Satisfactory Tolerances in PLL Parameters

In Sec. 7 we determined that in order to maintain optimal PLL parameters we desire $\beta_{SNR,V}=1$ identically, which in turn means that we strive to maintain $\hat{l}_{M,N}=\alpha_{SNR,d}$. Since (see Section 6.1.2) $E[\hat{l}_{M,N}]=\alpha_{SNR,d}$ and given (7), the way to increase the accuracy of the approximation $\hat{l}_{M,N}\approx\alpha_{SNR,d}$ is by increasing N . To obtain a quantitative measure of the value of N that is needed in order to achieve acceptable performance, let us denote the natural frequency and damping factor we are trying to achieve as $\bar{\omega}_n$ and $\bar{\zeta}$. We want to achieve them at all $E_S/N_0=\chi$ in the range $\chi\in[\Gamma,\infty]$ where Γ is some reasonable lower bound (for example, the PLL's lock threshold). The question we pose in this section is: can we find a lower bound on the N necessary to ensure, at each $E_S/N_0=\chi\in[\Gamma,\infty]$, that:

$$P(|\omega_{n\chi}/\bar{\omega}_n - 1| < tol) > C, P(|\zeta_\chi/\bar{\zeta} - 1| < tol) > C \quad (29)$$

where tol is the acceptable tolerance for ω_n and ζ , and C is the confidence. It can be easily shown from (26) and (28) that if we define $\rho \triangleq \sqrt{\alpha_{\chi,d}/\hat{l}_{M,N}}$ then we have $\omega_{n\chi} = \rho\bar{\omega}_n$ and $\zeta_\chi = \rho\bar{\zeta}$. Straightforward manipulations then show that an equivalent constraint to (29) is:

$$P\left(\alpha_{\chi,d} \left((1+tol)^{-2} - 1\right) < \hat{l}_{M,N} - \alpha_{\chi,d} < \alpha_{\chi,d} \left((1-tol)^{-2} - 1\right)\right) > C. \quad (30)$$

Since⁸ $E[\hat{l}_{M,N}]=\alpha_{\chi,d}$, to guarantee (30) it suffices that:

Fig. 11. Lower bound on N needed to achieve ω_n and ζ to a desired tolerance, at a given confidence.

$$P\left(|\hat{l}_{M,N} - E[\hat{l}_{M,N}]| < \alpha_{\chi,d} \cdot y\right) > C \quad (31)$$

where $y \triangleq \min\left\{\left((1-tol)^{-2} - 1\right), \left((1+tol)^{-2} - 1\right)\right\}$. Since $\hat{l}_{M,N}$ is Gaussian (31) is equivalent to the constraint

$$\text{erf}\left(\alpha_{\chi,d} \cdot y / \left(\sqrt{2} \sqrt{\text{Var}(\hat{l}_{M,N})}\right)\right) > C \quad (32)$$

with $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$. Now, since (see (7)) we

have $\text{Var}(\hat{l}_{M,N}) \leq 1/(2N)$, we can solve (32) for N , whereupon we find that for all $E_S/N_0=\chi \geq \Gamma$ a suitable lower bound on N would be $N > \left(1/\alpha_{\chi,d}^2\right) \cdot \left(\text{erf}^{-1}(C)/y\right)^2$. Graphs for N , computed in this manner, are shown in Fig. 11. As seen there, only a relatively small N is required in order to ensure good performance above reasonable SNRs for the respective modulations. In a practical implementation, the designer would probably choose N to be the lowest power of 2 that fulfills the minimum requirements, so as to ensure that the detector can be implemented as in Fig. 2 (i.e., to avoid the need to directly implement the division by $2N$).

10. Conclusions

In this paper we presented and investigated an adaptive phase detection structure for M-PSK. The detector was characterized via theoretical derivations, simulation

⁸ We made the assumption $E[\cos(M\theta_e)] = 1$. See Section 6.1.2.

results, and system identification analysis. It was shown that the proposed structure allows for optimal PLL parameters to be maintained at any SNR at which the PLL can attain lock. Moreover, the detector has superior phase-error variance performance and has a compact implementation that is suitable for use within an FPGA or ASIC. Hence, the proposed structure has many immediate applications in M-PSK receivers.

Acknowledgments

The author gratefully acknowledges the financial support provided to him by the National Sciences and Engineering Research Council of Canada (NSERC) through its Postgraduate Scholarship.

Appendix A: The AGC's Operation

A.1 The Parameter K

Obtaining a physical insight into the meaning of K is straightforward and should perhaps even be intuitively apparent to persons who have designed and built an M-PSK wireless receiver. To explicitly spell out this meaning, we assume that we are ideally locked (i.e. $\Delta\omega = 0$ and $\theta_e = 0$) and recall that throughout the paper we assumed for convenience and without loss of generality that $E_p = 1$. We then have $I(n) = K(\cos\phi_n + n_I(nT))$ and $Q(n) = K(\sin\phi_n + n_Q(nT))$.

Now, define the time-average operator $\langle \bullet \rangle$ as

$$\langle x(n) \rangle \triangleq \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{n=-N+1}^N x(n). \text{ It is then easy to see}$$

that:

$$K \xrightarrow{E_S/N_0 \rightarrow \infty} \sqrt{\langle I^2(n) + Q^2(n) \rangle} \quad (33)$$

i.e. at high SNR we have that K is roughly the RMS (Root-Mean-Square) of the M-PSK signal. Now, to inject a little more real-world issues into the model, we know that samplers have a finite number of bits and the AGC's job is, as already noted, to ensure that the samplers are not overdriven or underdriven. Consider, for example, a system which has 8-bit samplers, which give a range for $I(n)$ and $Q(n)$ of ± 127 . Let's also assume that the AGC controls the input signal so that, to avoid sporadic overdriving of the samplers due to noise, the input signal's RMS value is controlled to about 80% of the samplers' dynamic range. Note that 8-bit samplers and 80% driving of the samplers are certainly real-world parameters. In terms of the signal model used in the paper (see Fig. 1) where unity gain samplers are assumed, it can

be seen that at high SNR the model of Fig. 1 applies with about $K = 100$ (we rounded to $K = 100$. The exact number is $K = 80\% \times 127 = 101.6$).

Let us now look at the situation at low SNR. Since we assumed for our model a constant E_S , it follows that

$$\text{Var}(n_I(nT)), \text{Var}(n_Q(nT)) \xrightarrow{E_S/N_0 \rightarrow 0} \infty. \text{ The AGC}$$

still needs to control K so that the samplers are not overdriven, and hence we must have $K \xrightarrow{E_S/N_0 \rightarrow 0} 0$.

In summary then, if b is the number of bits in the samplers (including sign bit) and the AGC attempts to control the RMS of the input signal to $100 \cdot r$ percent of

the samplers' range, we have $K \xrightarrow{E_S/N_0 \rightarrow 0} 0$ and

$$K \xrightarrow{E_S/N_0 \rightarrow \infty} r \cdot (2^{b-1} - 1). \text{ For example, for the 8-bit samplers with 80\% driving we discussed above, we have } b=8 \text{ and } r=0.8, \text{ and the dynamic range of } K \text{ is about } 0 \leq K \leq 100.$$

We can now easily see the problem with the implementation of the M^{th} -order nonlinearity NDA phase detector, i.e. $c_M \triangleq \text{Im}[(I(n) + j \cdot Q(n))^M]$, for which the phase detector implementation must accommodate the dynamic range of K^M . Consider for example a 16-PSK receiver, i.e. a 16^{th} order nonlinearity. With such an implementation and $0 \leq K \leq 100$ as detailed above, the PLL for c_M would have to accommodate at least the dynamic range of magnitudes of between 0 and $(\max\{K\})^M = 100^{16} = 10^{32}$. This would require the output of a fixed-point lookup table to be $\log_2(10^{32}) + 1 \approx 107$ bits wide (!). Not only would this be in itself unfeasible, but this means that the ensuing loop filter would also have to handle data that is 107 bits wide. Moreover, since the registers in the loop filter have typically many more bits than the input data width, this implies an entirely unrealizable loop filter.

For the M^{th} power synchronizer, the only way to combat this explosion in logic resource usage is to artificially reduce K by discarding lower order bits of the samplers (thereby significantly worsening the effective sampler quantization), and to similarly discard the lower order bits of the output of the lookup table (thereby significantly limiting the dynamic range of the input I and Q signals that can be handled by the phase detector, and thus tightening the constraints on the AGC even more). Alternatively, floating-point implementations can be used; however, implementation of a floating-point

system in hardware (i.e. in an ASIC or FPGA) is always significantly more complex than implementation of a fixed-point system. Very similar problems afflict the implementation of the M^{th} -order nonlinearity NDA lock detector (see [1]).

In contrast, for $d_M(n)$ and $x_M(n)$, the output of the lookup tables is always in the interval $[-1,1]$, regardless of K or M . Thus, with $d_M(n)$ and $x_M(n)$, a fixed-point lookup table with just an 8-bit output (quantization to 256 levels of the interval $[-1,1]$) will usually be more than sufficient, for any K and any M . As noted in Section 5.2, it then follows that a small dynamic range is needed for the lookup tables in Fig. 2. Again, as detailed in the previous paragraphs, this reduction in the necessary dynamic range has profound implications regarding the lookup table's size, the size of the ensuing datapath and, in particular, the complexity of the loop filter.

Finally, we could have just as easily adopted the convention that the binary point at the output of the samplers immediately follows the sign bit. For example, if the output of the sampler is 00000011_b , this could signify the value 3 or, alternatively, $\frac{3}{128}$ (if we think of the binary point as being at the right of the sign bit, i.e. 0.0000011_b). This decision is purely arbitrary and has no bearing upon the dynamic range considerations outlined above. Under the mathematically convenient assumption of the binary point being to the right of the sign bit of the sampler data, we have that $0 \leq K \leq r \leq 1$ regardless of the number of bits in the sampler. This is the assumption made in this paper.

A.2 Understanding the AGC's Operation

General analysis of AGC-induced effects is hindered by the fact that the constraints and parameters of AGC circuits are strongly dependent upon the specific communications system. It is thus, perhaps, less of a surprise that most contemporary synchronization texts ignore these effects (by assuming a constant $K=1$). This is the case, for example, in [6] and [13], which are some of the most comprehensive modern works on synchronization in wireless communications⁹.

Here, we strive to attain an intuitive (rather than mathematically rigorous) understanding of the AGC's operation. To that end, it is instructive to take a look at

⁹ Nonetheless, some treatment of AGC effects does exist. See for example [4] chap. 9 and [14] chap. 7, though the discussions there pertain to unmodulated carrier-wave synchronization.

Fig. 12. Pre-AGC and post-AGC waveforms.

the waveforms before and after the AGC. We assume as an example that our AGC attempts to control the waveform amplitudes at the input of the samplers so that the RMS value of the waveforms is 80% of the dynamic range of the samplers. We also assume, for simplicity, that the samplers' full-scale input range is ± 1 volt, which means that our AGC tries to ensure that the pre-sampler waveforms have an RMS of 0.8 volt. We also make, for the purposes of the following demonstration, the same convenient assumptions as in the previous subsection, in particular, that the receiver is locked. Now, let us look at the pre-AGC (=post-matched-filter) and the post-AGC (=pre-sampler) waveforms at various SNRs. We shall look at the I -channel of a BPSK signal that has rectangular baseband pulses (remember that the signal has passed through a matched filter, so the signal waveform is now composed of the triangular pulses $p(t) \otimes h(t)$).

Fig. 12 shows the effects of the AGC upon the waveforms. In the bottom subplots, the dashed horizontal lines represent the samplers' full-scale voltage. Of course, the AGC "sees" only the pre-AGC signal+noise waveform, and the samplers "see" only the post-AGC signal+noise waveform; the separation into separate signal and noise waveforms is only presented, courtesy of the computer simulation, for the benefit of the reader. As can be clearly seen in the graphs, the AGC insures that the signal+noise waveforms at the input of the samplers are such that the samplers are saturated infrequently, yet the 80% RMS driving level ensures that most of the dynamic range of the samplers is used at all SNRs. Clearly, had K not been reduced to accommodate for the noise power, the samplers would be saturated almost all the time, especially at low SNRs, as can be seen in the signal+noise waveforms before the AGC.

Now, notice that the assumption made here is of a constant E_S and a changing noise power. Put another way, we assume that the E_S/N_0 changes because N_0 changes. This is contrary to what happens in practice. In practice, the noise power is generally constant (we have $\text{NoisePowerPerHz}_{\text{dBm/Hz}} = \text{ThermalNoisePowerPerHz}_{\text{dBm/Hz}} + \text{ReceiverNoiseFigure}_{\text{dB}}$) and the E_S/N_0 changes because E_S changes. However, the adoption of the convention of a constant E_S does not impact the analysis, and in fact simplifies it. The analysis is simplified because we can assume that true matched filters (with energy $E_p = 2 \cdot E_S$) are present in the receiver. The receiver model adopted here is also the one used in most communications and synchronization texts (for example, see [6] and [13]). More importantly, the pre-sampler waveforms presented in Fig. 12 will be those that will indeed be encountered in practice, and, furthermore, the effective amplitude suppression factors in Fig. 6 are also those that will be encountered in practice.

It is of paramount importance to realize that we are still assuming an *ideal* AGC, i.e. the example AGC circuit discussed above is assumed to be devoid of lag time and is assumed to control the RMS of the pre-sampler waveforms to precisely 80% of the samplers' dynamic range. The assumption of a constant $K=1$, though undertaken in the vast majority of synchronization texts, does not really describe an *ideal* AGC, but rather an *atrophied* AGC which operates within a system whose samplers have an infinite dynamic range and an infinite number of quantization bits. Clearly, the AGC discussed

here (though still "ideal") is a much closer approximation of reality, as opposed to the assumption of a constant $K=1$. The curve of K as a function of the SNR for our example AGC is given in Fig. 6.

References

- [1] Y. Linn and N. Peleg, "A family of self-normalizing carrier lock detectors and E_S/N_0 estimators for M-PSK and other phase modulation schemes," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp.1659-1668, Sept. 2004.
- [2] R. N. McDonough and A. D. Whalen, *Detection of Signals in Noise*, 2nd ed. San Diego, CA: Academic Press, 1995.
- [3] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
- [4] Blanchard, *Phase-Locked Loops Application to Coherent Receiver Design*. New York: John Wiley & Sons, 1976.
- [5] W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*. New Jersey: Prentice-Hall, 1973.
- [6] H. Meyr, M. Moeneclaey, S. A. Fechtel, *Digital Communication Receivers*. New York: John Wiley & Sons, 1997.
- [7] M. K. Simon, S. M. Hinedi, W. C. Lindsey, *Digital Communication Techniques - Signal Design and Detection*. Englewood Cliffs, NJ: Prentice Hall, 1994.
- [8] M. R. Spiegel, *Mathematical Handbook of Formulas and Tables*. Singapore: McGraw-Hill, 1990.
- [9] F. M. Gardner, *Phaselock Techniques*, 2nd ed., New York: John Wiley & Sons, 1979.
- [10] W. P. Osborne and B. T. Kopp, "Synchronization in M-PSK modems", *Proc. ICC '92*, vol. 3, pp. 1436-1440
- [11] W. P. Osborne and B. T. Kopp, "An analysis of carrier phase jitter in an M-PSK receiver utilizing MAP estimation", in *Proc. MILCOM '93*, vol. 2, pp. 465-470.
- [12] B. T. Kopp and W. P. Osborne, "Phase jitter in MPSK carrier tracking loops: Analytical, simulation, and laboratory results," *IEEE Trans. Commun.*, vol. 45, no. 11, pp. 1385-1388, Nov. 1997.
- [13] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*. New York: Plenum Press, 1997.
- [14] H. Meyr and G. Ascheid, *Synchronization in Digital Communications - Volume 1 Phase-, Frequency-Locked Loops, and Amplitude Control*. New York: John Wiley & Sons, 1990.
- [15] S. A. Butman and J. R. Lesh, "The effects of bandpass limiters on n-phase tracking systems," *IEEE Trans. Commun.*, vol. 25, no. 6, pp. 569 - 576, Jun. 1977.
- [16] R. De Gaudenzi, T. Garde, V. Vanghi, "Performance analysis of decision-directed maximum-likelihood phase estimators for M-PSK modulated signals," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 3090-3100, Dec. 1995.
- [17] R. E. Best, *Phase-Locked Loops: Theory, Design, and Applications*, 2nd ed. New York: McGraw-Hill, 1993.
- [18] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automatic Control*, vol. AC-10 (1965), p. 461-464.