# Phase Smoothing Functions for Continuous Phase Modulation

David K. Asano, *Student Member, IEEE*, Harry Leib, *Member, IEEE* and Subbarayan Pasupathy, *Fellow, IEEE*

*Abstract*— **In this paper, the problem of signal design for continuous phase modulation is examined. Using the performance measures of effective bandwidth and minimum distance, "optimal" signal shapes are calculated for various receiver observation intervals for full and partial response signalling. For the partial response case, the optimization could only be done for a receiver observation interval of one symbol due to the difficulty of the optimization. Therefore, a family of signals based on previously calculated optimal signals is introduced. The tradeoffs that result between bandwidth and bit error rate are then compared as a function of the receiver observation interval and as a function of the memory length of the signalling scheme. Then, the signals introduced in this paper are compared to well known signals in terms of their effective bandwidth–minimum distance tradeoffs, power spectra and percent–in–band–power. The signals obtained by the optimization process are found to have good spectral properties in relation to well known signals.**

## I. Introduction

Continuous Phase Modulation (CPM) has become a popular modulation scheme in recent years due to its desirable properties of bandwidth efficiency and constant envelope [1], [2], [3]. In most previous work, the shape of the CPM signal was chosen to be "smooth" to give good bandwidth characteristics. While this resulted in a bandwidth reduction, the bit error rate was not as low as possible.

Some work has, however, tried to look at both requirements simultaneously, e.g., [4]. This paper showed the optimal tradeoff for given bandwidth and bit error rate criteria but unfortunately the receiver that was considered was one that observes for an infinite amount of time before making a decision. This gives no insight into the effects of receiver complexity on the choice of signal shape. Another drawback was the restriction to only full response CPM schemes. Partial response signals are known to be much more bandwidth efficient [5] and therefore have much more potential to provide a good tradeoff between bandwidth and bit error rate.

Paper approved by P.H. Wittke, the Editor for Communication Theory of the IEEE Communications Society. Manuscript received February 22, 1991; revised September 3, 1991 and January 28, 1992. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada. This paper was presented in part at the 1989 IEEE Pacific Rim Conference in Victoria, B.C., June 2, 1989.

D. Asano and S. Pasupathy are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, M5S 1A4, Canada.

H. Leib is with the Department of Electrical Engineering, McGill University, Montreal, P.Q., H3A 2A7, Canada.

IEEE Log Number 9401001

This paper attempts to provide some deeper insight into the tradeoffs between the bit error rate and bandwidth of CPM by studying the performance of various CPM schemes as a function of the signal shape. To this end, we calculate some "optimal" CPM signal shapes and compare them to previously suggested CPM signals.

In Section II the CPM signal and performance measures are introduced. Sections III and IV present the full and partial response cases respectively. Section V shows some comparisons between the various results. Finally, section VI contains the conclusions.

## II. Preliminaries

### A. System Model

The source of the communication system considered in this paper is assumed to emit one symbol from the alphabet $\{1,-1\}$ every symbol interval or $T$ seconds. The sequence of symbols is then $a(n) \in \{1,-1\}, \quad n \geq 0$, starting at $t=0$ with symbol $a(0)$.

The channel under consideration is the Additive White Gaussian Noise (AWGN) channel with power spectral density $N_0/2$. We also assume that the channel is free of distortion resulting from inter-symbol interference.

As for the receiver, we consider a minimum distance receiver [6]. This kind of receiver is optimal for the maximum likelihood decision criterion. The receiver operates by observing the received signal for $N$ symbol intervals, say from $t=0$ to $t=NT$. Then it estimates $a(0)$, which was transmitted in the interval $0 \leq t \leq T$, by dividing the possible transmitted signals into two sets $S_1$ and $S_{-1}$ defined as $S_1 = \{s(t) : a(0) = 1\}$ and $S_{-1} = \{s(t) : a(0) = -1\}$. The receiver chooses its estimate of $a(0)$, $\hat{a}(0)$, to be 1 if the received signal is "closer" to a signal in $S_1$ and $\hat{a}(1) = -1$ if it is "closer" to one in $S_{-1}$. Since the channel is an AWGN one, the concept of "closeness" can be interpreted in terms of Euclidean distance in a Euclidean space [7,pp.517-519] such as $C^2_{[a,b]}$ [8]. Also, the receiver is assumed to be perfectly coherent and to not suffer from timing errors.

### B. CPM Signal Description

A CPM signal can be defined by

$$s(t) = A\cos[2\pi f_c t + \phi(t)]. \quad (1)$$

The carrier frequency of the signal is denoted by $f_c$ while the energy of the signal, $E$, is proportional to $A^2$. The phase, $\phi(t)$, is defined by

$$\phi(t) = \phi(0) + 2\pi h \sum_{n=0}^{\infty} a(n) q(t - nT)$$

for $t \geq 0$. and contains all the information. The initial phase, $\phi(0)$, is assumed known because coherent systems are being studied. The parameter $h$ is the modulation index and represents the effective phase change that one symbol causes. The function $q(t)$ is the phase smoothing function. It is $q(t)$ that largely determines the signal shape and hence the performance of a CPM scheme. As the name CPM implies, the phase smoothing function is continuous and satisfies $q(t) = 0$ for $t \leq 0$ and $q(t) = 0.5$ for $t \geq LT$. $L$ can be thought of as the memory length of the modulation scheme because each symbol affects the signal shape for a maximum of $L$ symbol intervals. The case when $L = 1$ is called full response signalling. Partial response signalling results when $L \geq 2$.

It will be convenient in the following analysis to use the complex envelope representation of CPM. The complex envelope of a signal, $\tilde{s}(t)$, is related to the signal, $s(t)$, itself through the relation $s(t) = \Re\{\tilde{s}(t)e^{j2\pi f_c t}\}$, where $\Re\{z\}$ denotes the real part of the complex quantity $z$. Thus, from (1) the complex envelope of the CPM signal is $\tilde{s}(t) = Ae^{j\phi(t)}$.

### C. Performance Measures

As our measure of bit error rate, we choose $d_{min}$, which is the smallest distance between any signal in $S_{-1}$ and any signal in $S_1$, i.e.,

$$d_{min} = \min\{d(s_1, s_2) : s_1 \in S_1, s_2 \in S_{-1}\}.$$

Although $d_{min}$ does not give the bit error rate exactly, for large values of $E/N_0$ the probability of error can be approximated by

$$P_e \approx \frac{1}{2}\text{erfc}\left[\sqrt{\frac{E}{N_0}k d_{min}^2}\right]$$

for some constant $k$.

The squared Euclidean distance between two signals can be shown to be equal to [9]

$$d^2(s_1, s_2) = 2NT\{1 - \Re[\tilde{\rho}_{12}]\}. \tag{2}$$

The cross-correlation between $s_1(t)$ and $s_2(t)$, $\tilde{\rho}_{12}$, is given by

$$\tilde{\rho}_{12} = \frac{1}{NT}\int_0^{NT} \exp\left\{j2\pi h \sum_{n=0}^{\infty} b(n) q(t - nT)\right\} dt. \tag{3}$$

In this equation, $b(n) \in \{-2, 0, 2\}$ is the difference sequence $a_1(n) - a_2(n)$ where $a_1(n)$ and $a_2(n)$ are the symbol sequences corresponding to $\phi_1(t)$ and $\phi_2(t)$ respectively. From the Schwarz inequality [10], the magnitude of $\tilde{\rho}_{12}$ is bounded by $|\tilde{\rho}_{12}| \leq 1$. Therefore, from (2) the squared Euclidean distance between any two signals can be no more than $4NT$. Using (2), $d_{min}^2$ can be equivalently written as

$$d_{min}^2 = 2NT\{1 - \max \Re[\tilde{\rho}_{12}]\}. \tag{4}$$

The maximization is taken over all pairs of signals $s_1$ and $s_2$ such that $s_1 \in S_1, s_2 \in S_{-1}$. It will be convenient to define a normalized squared minimum distance as

$$D_{min}^2 = d_{min}^2/T.$$

The bandwidth measure used in this paper is the effective bandwidth defined by

$$B_e^2 = \frac{4\int_0^{\infty}(f - f_0)^2 S_+(f)\,df}{\int_0^{\infty} S_+(f)\,df},$$

where $S_+(f)$ is the positive part of the signal's spectrum and $f_0$ is the center frequency. This measure of bandwidth is similar to the variance measure of statistics and like its counterpart gives an indication of the spread of the function from its mean value, $f_0$. $B_e$ can be expressed in terms of the phase smoothing function, $q(t)$, as [4]

$$B_e^2 = \frac{4h^2}{T}\int_0^{LT} \dot{q}^2(t)\,dt. \tag{5}$$

Although the $B_e$ of a signal has no upper bound, it has a lower bound given by (see Appendix A) $(B_eT)_{min} = h/\sqrt{L}$, which is reached for the straight line phase smoothing function known as L-REC [2]. Thus, the minimum possible $B_e$ increases as $h$ increases while it decreases as $L$ increases. Therefore, to get small values of $B_e$, a small $h$ and/or a large value of $L$ are necessary.

### D. Sense of Optimality

In this paper, some optimal phase smoothing functions will be calculated. These functions will be optimal in the sense that they minimize $B_e$ for a given $d_{min}$.

## III. Full Response

### A. One Symbol Interval Demodulation ($N = 1$)

1) *Performance Measures*: For a receiver that observes only one symbol interval, there is only one distance since there are only two possible signals. Thus, $d_{min}$ is obtained from (2) with $\Re[\tilde{\rho}_{12}]$ given by

$$\rho \triangleq \Re[\tilde{\rho}_{12}] = \frac{1}{T}\int_0^T \cos[4\pi h q(t)]\,dt. \tag{6}$$

The minimum distance can then be completely described in terms of $\rho$ by $d_{min}^2 = 2T\{1 - \rho\}$.

The effective bandwidth in this case is given simply by (5) with $L = 1$. Since the bandwidth of a modulation scheme does not depend on the receiver, $B_e$ will be given by the same expression for all full response schemes.

2) *Heuristic Phase Smoothing Functions*: Some insight can be gained into the shape of the phase smoothing function that gives the best tradeoff between $B_e$ and $d_{min}$ by

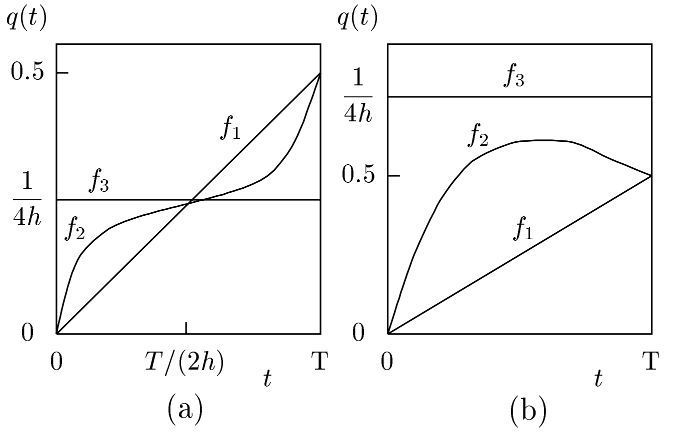


Fig. 1. Some heuristic phase smoothing functions. (a) $h \geq 0.5$. (b) $h \leq 0.5$.

looking at the functions that minimize either of these performance measures. As is shown in Appendix A, the smallest $B_e$ is obtained by the L-REC phase smoothing function shown as $f_1$ in Fig. 1. The phase smoothing function $f_3$ in Fig. 1 yields the largest $d_{min}$. This can be seen from (6). Since $|\rho|$ is bounded by 1, the best performance will be obtained when $\rho = -1$. This is possible only if $\cos[4\pi hq(t)] = -1$ for $0 \leq t \leq T$, i.e., when $q(t) = (4h)^{-1}$. This results in the two cases: $h > 0.5$ and $h < 0.5$ because when $h > 0.5$, $(4h)^{-1} < 0.5$ and vice–versa. As a sort of compromise, the function $f_2$ is drawn between the two. This compromise is the form of the phase smoothing function that would intuitively be expected to give the best tradeoff. Throughout the rest of this paper, whenever the term "heuristic phase smoothing functions" is used it will refer to the functions of this section.

3) *Optimal Phase Smoothing Functions*: In this case, the optimal phase smoothing functions are rather easy to calculate. The problem is to minimize $B_e^2$ subject to $\rho = constant$. The optimal solution then must satisfy the differential equation (see Appendix B)

$$\ddot{q}(t) = -\frac{\pi\lambda}{2h}\sin[4\pi hq(t)] \tag{7}$$

Here, $\lambda$ is chosen to so that the solution satisfies the subsidiary condition: $\rho = constant$.

Unfortunately, this is a non-linear differential equation whose solution was not analytically obtainable. Therefore, the solution of the above differential equation was found numerically by using the fourth order Runge-Kutta algorithm [11]. Some solutions of (7) are shown in Fig. 2(a) for $D_{min}^2 = 3$. It is interesting to note that these functions have a shape similar to the heuristic phase smoothing functions.

The $B_e$-distance tradeoffs that result are shown in Fig. 2(b). For each value of $h$, the leftmost points of the curves are the minimum bandwidths obtainable. If the distance is decreased further, the bandwidth will again increase. The maximum value of $D_{min}^2$ is 4 when $B_e = \infty$, which is the performance of antipodal PSK. Lowering the value of $D_{min}$ improves the bandwidth performance at the expense

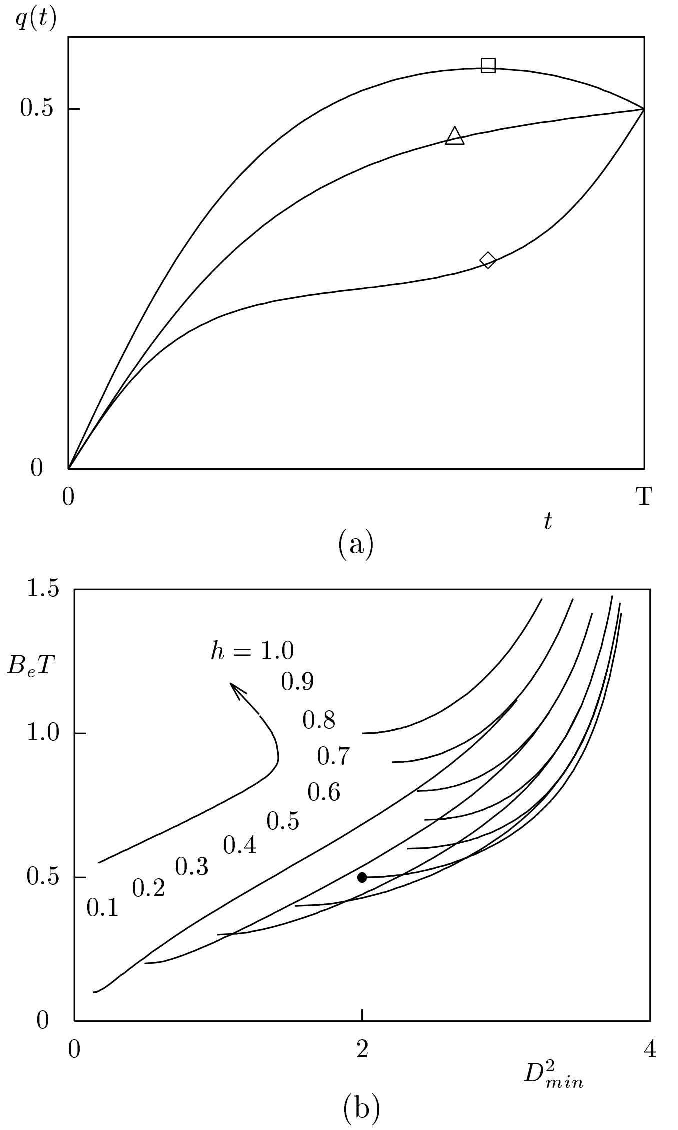


Fig. 2. Full response, $N = 1$. (a) Optimal phase smoothing functions for $D_{min}^2 = 3$. □:$h = 0.4$, $B_eT = 0.657$; △:$h = 0.5$, $B_eT = 0.638$; ◇:$h = 1.0$, $B_eT = 1.275$. (b) Optimal bandwidth-distance tradeoffs. • = MSK

of bit error rate performance. The ultimate in performance would be achieved if $B_e = 0$ and $D_{min}^2 = 4$. As can be seen from Fig. 2(b), the curves which come closest to this point are the ones for $0 < h \leq 0.5$.

### *B. Two or More Symbol Interval Demodulation ($N \geq 2$)*

1) *Minimum Distance*: For $N \geq 2$, the distance between two signals is given by (2) as before but the cross correlation, $\tilde{\rho}_{12}$, can be reduced further because the phase smoothing function only affects the signal for one symbol interval. As shown in Appendix C, (3) can be reduced to

$$\tilde{\rho}_{12} = \frac{1}{N}\sum_{n=0}^{N-1}\left[\tilde{\rho}_{b(n)}\exp\{j\pi h\sum_{m=0}^{n-1}b(m)\}\right], \tag{8}$$

where

$$\tilde{\rho}_{b(n)} = \begin{cases} \tilde{\rho} & b(n) = 2 \\ 1 & b(n) = 0 \\ \tilde{\rho}^* & b(n) = -2 \end{cases}$$

and

$$\tilde{\rho} = \frac{1}{T}\int_0^T e^{j4\pi hq(t)}\, dt \triangleq \rho_T + j\rho_T'.$$

Thus, the distance between two signals can be expressed in terms of $\tilde{\rho}$.

2) *Optimal Phase Smoothing Functions*: In this case the optimization cannot be done directly as in the $N = 1$ case due to the fact that the $d_{min}$ function is not continuously differentiable. Since $d_{min}$ can be written in terms of the maximum cross correlation and the cross correlation can be written in terms of $\tilde{\rho}$, the minimum $B_e$ for a given $d_{min}$ may be calculated in the following way. First, the minimum $B_e$ for a given $\tilde{\rho}$ is calculated. Second, $d_{min}$ is calculated from $\tilde{\rho}$. Then if all possible $\tilde{\rho}$ resulting in that $d_{min}$ are considered, the minimum $B_e$ for that distance can be found.

Minimizing $B_e$ subject to $\tilde{\rho} = constant$ can be written in the following way. Minimize

$$\frac{4h^2}{T}\int_0^T \dot{q}^2(t)\, dt$$

subject to

$$\rho_T = \frac{1}{T}\int_0^T \cos[4\pi hq(t)]\, dt = C_1$$

and

$$\rho_T' = \frac{1}{T}\int_0^T \sin[4\pi hq(t)]\, dt = C_2.$$

Using the calculus of variations, the problem can be reduced to solving the differential equation (see Appendix B)

$$\ddot{q}(t) = -\frac{\pi\lambda_1}{2h}\sin[4\pi hq(t)] + \frac{\pi\lambda_2}{2h}\cos[4\pi hq(t)]. \quad (9)$$

Again, $\lambda_1$ and $\lambda_2$ are chosen so that the solution satisfies the subsidiary conditions.

Due to space limitations, results are not presented here but can be found in [9]. The optimal phase smoothing functions have a shape similar to the heuristic ones.

### C. Infinite Observation Demodulation ($N = \infty$)

As the number of observation intervals gets very large, the receiver performance saturates to that of the infinite observation interval receiver. The tradeoff for $N = \infty$ will be the best since the receiver has an infinite amount of time to observe before making a decision. To find this best possible tradeoff, $d_{min}$ is examined in the limit as $N \to \infty$.

1) *Minimum Distance*: If the terms in the sum in (8) are examined closely, it is apparent that the correlation resulting from the difference sequence $\{2, -2, 0, 0, 0, \cdots\}$ will eventually become the largest and thus result in the minimum distance. This can be seen by examining the sum term by term. For a typical difference sequence, the sum would look something like $\tilde{\rho} + e^{j\pi 2h} + \tilde{\rho}\cdot e^{j\pi 2h} + \tilde{\rho}^* \cdot e^{j\pi 4h} + \cdots$ The first term is the same for all difference sequences because $b(0) = 2$. For the difference sequence $\{2, -2, 0, 0, 0, \cdots\}$, the sum is $\tilde{\rho}+\tilde{\rho}^*\cdot e^{j\pi 2h}+1+1+\cdots$. Since the magnitude of a term such as $\tilde{\rho}\cdot e^{j\pi 2h}$ is bounded by 1, its real part must be less than or equal to 1. It can only be equal to one for all such terms when $h = 1$ for the difference sequence $\{2, 0, 0, 0, \cdots\}$. However, when $h = 1$, $d_{min}$ does not increase with the receiver observation length because the phase at $t = kT$ is independent of the transmitted symbols. So, this case can be ignored. Therefore, the sum $\tilde{\rho} + \tilde{\rho}^* \cdot e^{j\pi 2h} + 1 + 1 + \cdots$ will eventually become the largest. It is not necessarily the largest for small values of $N$ because the second term is not necessarily larger than the corresponding second term of other sequences.

For sufficiently large $N$, the maximum correlation can then be written simply as,

$$\tilde{\rho}_{max} = \frac{1}{N}[\tilde{\rho} + \tilde{\rho}^* \cdot e^{j\pi 2h} + (N-2)].$$

Substituting this into (4), the normalized distance is

$$D_{min}^2 = 2\Re[2 - \tilde{\rho}_T - \tilde{\rho}_T^* \cdot e^{j\pi 2h}]$$

which can also be written as

$$D_{min}^2 = 2[2 - \rho_T' \sin(2\pi h) - \rho_T(1+\cos(2\pi h))].$$

The largest $D_{min}^2$ that is possible for any value of $N$ is the maximum of these equations. Remembering that the maximum magnitude of terms like $\tilde{\rho}^* \cdot e^{j\pi 2h}$ is 1, the maximum value of $D_{min}^2$ for any $N$ is 8.

2) *Optimal Phase Smoothing Functions*: The optimal phase smoothing functions can be easily found by applying the calculus of variations once again. This time the problem to be solved is the following. Minimize

$$\frac{4h^2}{T}\int_0^T \dot{q}^2(t)\, dt$$

subject to

$$2[2 - \rho_T' \sin(2\pi h) - \rho_T(1+\cos(2\pi h))] = C$$

or equivalently

$$\rho_T' \sin(2\pi h) + \rho_T(1+\cos(2\pi h)) = C'$$

where $\rho_T'$ and $\rho_T$ are given by

$$\rho_T = \frac{1}{T}\int_0^T \cos[4\pi hq(t)]\, dt$$

and

$$\rho_T' = \frac{1}{T}\int_0^T \sin[4\pi hq(t)]\, dt.$$

This leads to the differential equation (see Appendix B)

$$\ddot{q}(t) = \frac{\pi\lambda}{2h}\{-\sin(2\pi h)\cos[4\pi hq(t)] + [1+\cos(2\pi h)]\sin[4\pi hq(t)]\}. \quad (10)$$

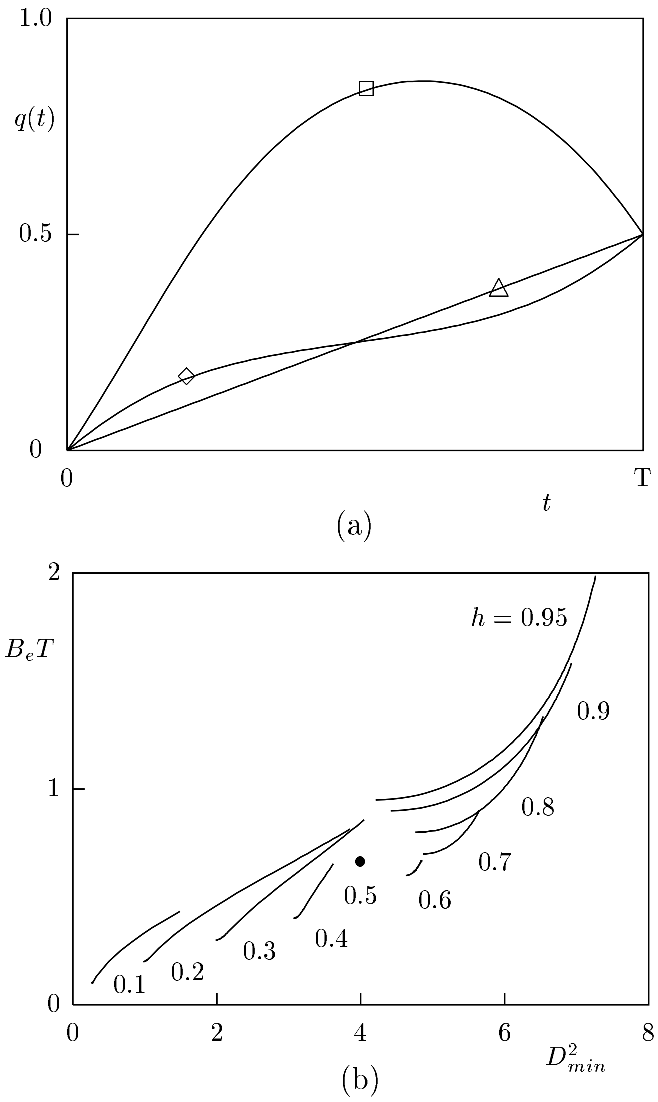


Fig. 3. Full response, $N = \infty$. (a) Optimal phase smoothing functions. $\Box$:$h = 0.3, B_eT = 0.844, D^2_{min} = 4$; $\triangle$:$h = 0.5, B_eT = 0.5, D^2_{min} = 4$; $\Diamond$:$h = 0.8, B_eT = 0.923, D^2_{min} = 5.74$. (b) Optimal bandwidth-distance tradeoffs

Numerical solution of (10) gives phase smoothing functions such as those shown in Fig. 3(a). Note that the shapes of the phase smoothing functions are again similar to the heuristic phase smoothing functions. The tradeoffs are shown in Fig. 3(b). The larger distances are obtained for values of $h$ close to 1.

## IV. Partial Response

### *A. One Symbol Interval Demodulation ($N = 1$)*

The optimal phase smoothing functions for $2 \leq L < \infty$ can be calculated but this kind of system would not be used because the receiver does not take advantage of the memory in the signal. For completeness, this case is mentioned here but no results will be presented. Interested readers are referred to [9].

### *B. Two or More Symbol Interval Demodulation ($N \geq 2$)*

1) *Performance Measures*: For $N = 2$, $d^2_{min} = 4T(1 - \max[\rho])$ where

$$\rho = \frac{1}{2T}\left\{\int_0^T \cos[4\pi h q(t)]\,dt + \int_T^{2T} \cos[2\pi h(2q(t) + b(1)q(t-T))]\,dt\right\}.$$

The maximization above is carried out by substituting the allowable values of $b_2$, which are $\{2, 0, -2\}$, into the equation. $B_e$ is given by (5).

Some difficulties arise in the minimization of $B_e$ for a given $d_{min}$. The calculus of variations cannot be used because the minimum distance function is not continuously differentiable and a time delay term appears in the integral. The correlation cannot be written in terms of the correlation over one symbol interval as in the full response case either. Therefore, it seems that the optimal phase smoothing functions cannot be found.

2) *Polynomial Phase Smoothing Functions*: Here, it is necessary to guess at the form of the optimal functions in order to proceed any further. As was seen in the previous section, the heuristic family of phase smoothing functions were of the same shape as the optimal ones found for the full response case. It is reasonable to suppose that functions of this form will lead to good tradeoffs.

A family of functions with the shape in question will be called POLY because its shape is determined by polynomials as the equations below show.

$$h \geq 0.5 : q(t) = \begin{cases} \frac{1}{4h} - k_1(\frac{LT}{2h} - t)^a, & 0 \leq t \leq \frac{LT}{2h} \\ k_2(t - \frac{LT}{2h})^a + \frac{1}{4h}, & \frac{LT}{2h} < t \leq LT \end{cases}$$

$$h < 0.5 : q(t) = \begin{cases} d - k(b - t)^a, & 0 \leq t \leq b \\ d - k(t - b)^a, & b < t \leq LT \end{cases}$$

The function is separated into two cases because of the different shape of the phase smoothing functions for $h > 0.5$ and $h < 0.5$. When $h = 0.5$, the two definitions are equivalent.

In the above equations, $a$ is the parameter describing the family. It is limited by $a \geq 1$. As $a$ increases, the function looks more and more rectangular. When the minimum value of $a$ is used, i.e. $a = 1$, the POLY phase smoothing function is a straight line and therefore gives the minimum $B_e$ possible.

The other variables, $b$, $k_1$, $k_2$ and $k$, are determined from $a$ so that $q(0) = 0$ and $q(LT) = 0.5$. This results in

$$k_1 = \frac{1}{4h}\left(\frac{2h}{LT}\right)^a, \; k_2 = \frac{\left(\frac{1}{2} - \frac{1}{4h}\right)}{\left[LT\left(1 - \frac{1}{2h}\right)\right]^a}, \; k = \frac{d}{b^a},$$

$$b = \frac{LT}{\left[1 + \left(1 - \frac{1}{2d}\right)^{\frac{1}{a}}\right]} \text{ and } d = \left(1 - \frac{1}{a}\right)\left[\frac{1}{4h} - \frac{1}{2}\right] + \frac{1}{2}.$$

Some examples of the polynomial phase smoothing functions are shown in Fig. 4. Notice the similarity to the

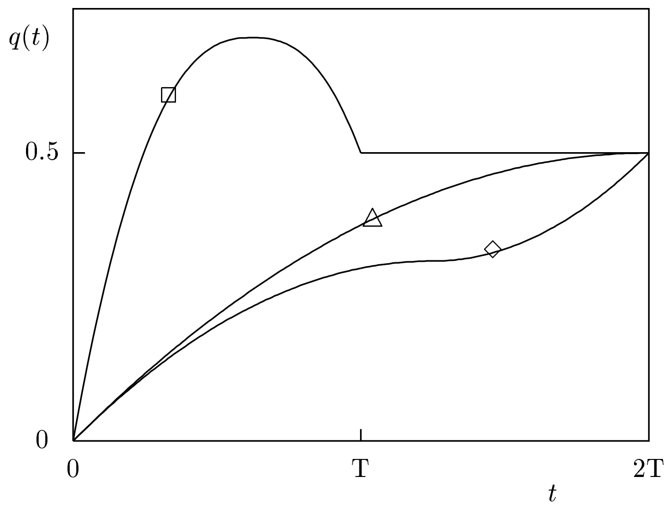


Fig. 4. Some polynomial phase smoothing functions. □:$h = 0.3, L = 1$; △:$h = 0.5, L = 2$; ◇:$h = 0.8, L = 2$.

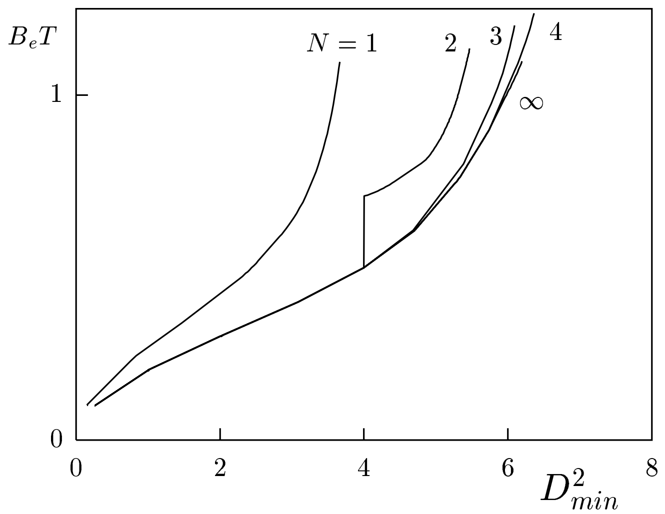


Fig. 5. Optimal bandwidth-minimum distance tradeoff envelopes for full response.

optimal phase smoothing functions found earlier. From the equations for the polynomial family, it is possible to calculate $B_e$ directly. For $h \geq 0.5$,

$$B_eT = ha/(\sqrt{L(2a-1)})$$

and for $0 \leq h \leq 0.5$,

$$B_eT = 2hak\sqrt{[b^{2a-1} + (L-b)^{2a-1}]/(2a-1)}.$$

From these equations it can be seen that $B_e$ increases when $h$ or $a$ increases.

## V. Comparisons

### A. Optimal Tradeoffs as a Function of N

First, the optimal tradeoffs between $B_e$ and $d_{min}$ as a function of the number of receiver observation intervals, $N$, will be examined. If the envelope of the tradeoffs calculated previously is considered, the dependence on $N$ can be seen. The envelopes were found by plotting the tradeoffs using a small step size for $h$. Fig. 5 shows the relationship between the tradeoff and $N$ for full response signalling. The area to the lower right of the curves represents values of $d_{min}$ and $B_e$ that are not attainable by CPM signals. It is interesting to note that the curve for $N = 2$ has a discontinuity at $D^2_{min} = 4$, which is exactly the distance given by MSK. The minimum $B_e$ for this distance is also that of MSK. Thus, MSK is optimal in terms of the $B_e$ criterion for $D^2_{min} = 4$.

As $N$ is increased from one to two, there is a relatively large increase in the $d_{min}$ obtainable for a given $B_e$. As $N$ is increased further, the increase becomes less and less. When $N = 4$, the tradeoff is very close to that of $N = \infty$ for the range of bandwidths that are shown. Therefore, the receiver is nearly saturated for $N = 4$. Another point to note is that for $B_e$ less than approximately 0.5, the tradeoff for $N \geq 2$ is the same as for $N = \infty$. The saturation of the receiver appears to happen more quickly for lower values of $B_e$.

A simple lower bound for the value of $N$ at which the tradeoff saturates can be easily derived. The difference sequence $\{2, -2, 0, \cdots\}$ corresponds to $d_{min}$ when $N = \infty$. The tradeoff will then be saturated when this difference sequence first leads to $d_{min}$, i.e., when this difference sequence leads to a smaller distance than all other difference sequences. Thus, it is clearly necessary that the distance corresponding to a sequence such as $\{2, 0, 0, \cdots\}$ must become larger than that corresponding to the sequence $\{2, -2, 0, \cdots\}$. The correlation corresponding to $\{2, 0, 0, \cdots\}$ for a given value of $N$ is

$$\tilde{\rho} = \frac{1}{N}[\rho_T + (N-1)\mathrm{e}^{j2\pi h}].$$

The minimum distance is then

$$D^2_{min} = 2N\{1 - \frac{1}{N}[\rho_T + (N-1)\cos(2\pi h)]\}.$$

This will be smaller than the distance corresponding to the difference sequence $\{2, -2, 0, \cdots\}$ until $N$ satisfies

$$N \geq \left\lceil \frac{3}{1-\cos(2\pi h)} \right\rceil,$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to $x$.

For partial response schemes, similar results were obtained [9]. Even though the optimal tradeoffs were not obtained, the results confirm the expectation that $L$ is more important than $N$ as far as the tradeoff between bandwidth and $d_{min}$ is concerned.

### B. Comparison to Well Known Functions

In this section, the performance tradeoffs, power spectra and percent–in–band–power for the phase smoothing functions found in this paper are compared to those for the commonly known phase smoothing functions defined in [3], [4]: Piecewise Linear (PCWL), Plateau Cosine (PLC),

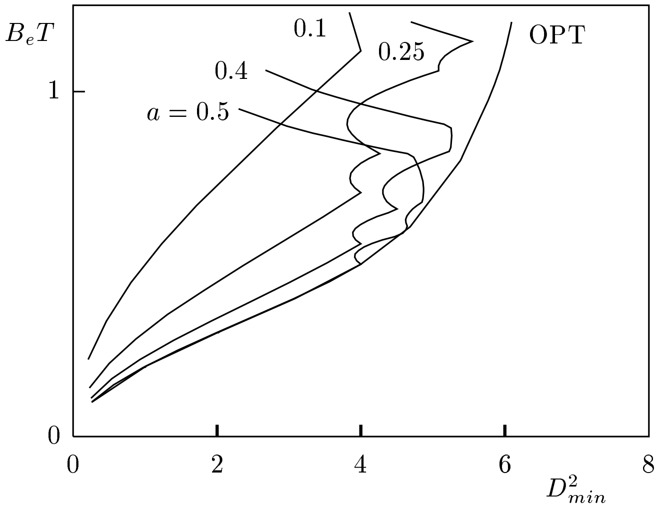


Fig. 6. Optimal tradeoffs compared to PCWL for $N = 3$.

TABLE I
POWER BANDWIDTH COMPARISONS FOR $L = 1, N = 3, D^2_{min} = 5.0$

| Family | $B_eT$ | 90% | 99% | 99.9% |
|---|---|---|---|---|
| OPT | 0.712 | 0.476 | 0.917 | 1.981 |
| PCWL | 0.751 | 0.481 | 0.986 | 2.179 |
| PLC | 0.841 | 0.491 | 1.521 | 2.378 |
| PLS | 0.992 | 0.503 | 1.653 | 3.427 |

Plateau Sine (PLS), Gaussian MSK (GMSK) and Tamed FM (TFM). The phase smoothing functions found by optimization will be referred to as the OPT family.

The PCWL, PLC and PLS families were extended to partial response by replacing $T$ with $LT$ in their definitions. However, the GMSK and TFM families are essentially partial response schemes and therefore will not be used in full response comparisons. The power spectra of these and all succeeding graphs were calculated using the autocorrelation method in [3].

1) *Full Response Comparisons*: The tradeoffs between $B_e$ and $D^2_{min}$ for PCWL are shown in Fig. 6 for $N = 3$. The optimal curve is of course better than the other curves but give the best gain for larger values of $D^2_{min}$. The PCWL curves are closest to the OPT curve while similar results for PLC and PLS show that their curves are respectively farther away [9]. The power spectra of the signalling schemes are plotted for $D^2_{min} = 5$ and $N = 3$ in Fig. 7. In this figure is can be seen that the OPT power spectrum has a main lobe that is close in width to those of the other families but has lower side lobes. The percent–in–band–power comparisons for the $N = 3$ case are shown in Table I. The table shows that the OPT family is better than the others in every case.

2) *Partial Response Comparisons ($L = 2$)*: Some representative bandwidth-minimum distance comparisons are shown in Fig. 8. Even though it is not optimal, the POLY family has a better tradeoff than the other families for almost all values of the $D^2_{min}$. In Fig. 9 comparisons be-

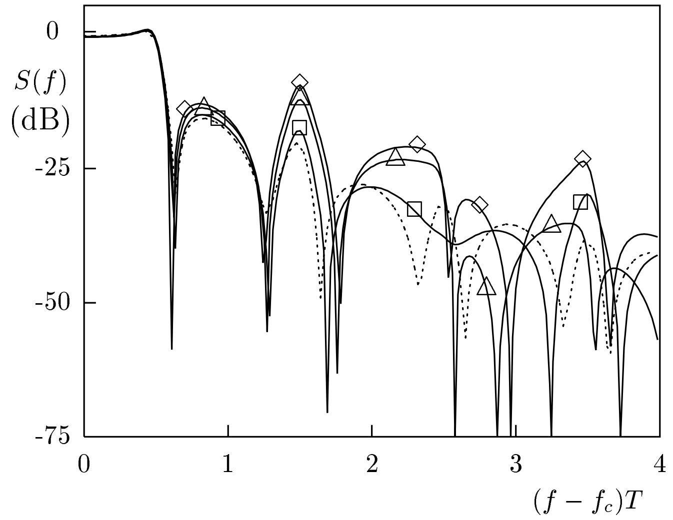


Fig. 7. Full response power spectrum comparison for $N = 3$ and $D^2_{min} = 5.0$. ···:OPT; □:PCWL; △:PLC; ◇:PLS.

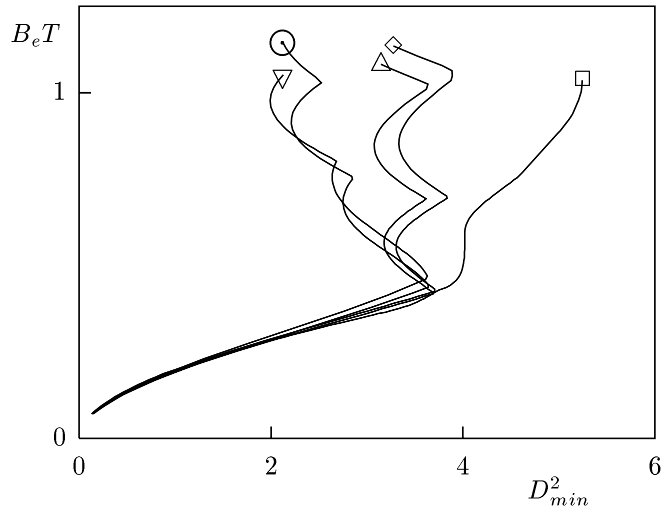


Fig. 8. Partial response POLY tradeoffs compared to GMSK and TFM for $L = 2$ and $N = 2$. □:POLY; △:TFM; ◇:GMSK, $B_b = 0.1$; ⊙:GMSK, $B_b = 0.5$; ▽:GMSK, $B_b = 1.0$.

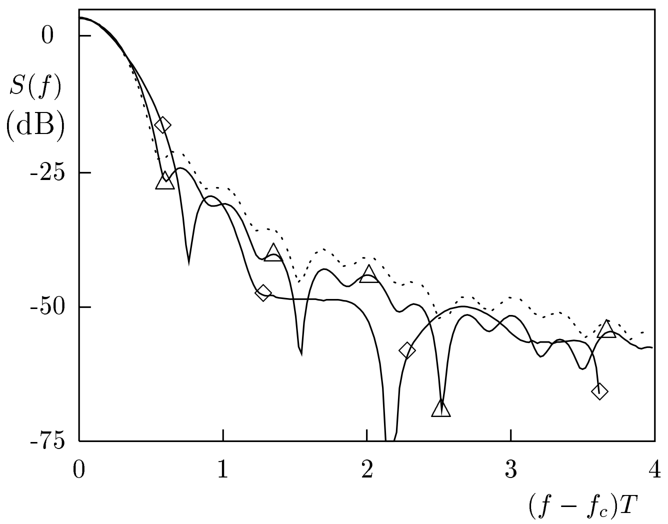


Fig. 9. Partial response power spectrum comparison for $L = 2$, $N = 2$ and $D^2_{min} = 3.0$. ···:POLY; △:TFM; ◇:GMSK.

TABLE II
POWER BANDWIDTH COMPARISONS FOR $L = 2, N = 2, D^2_{min} = 3.0$

| Family | $B_eT$ | 90% | 99% | 99.9% |
|---|---|---|---|---|
| POLY | 0.358 | 0.283 | 0.437 | 0.890 |
| GMSK | 0.375 | 0.310 | 0.497 | 0.639 |
| TFM | 0.363 | 0.293 | 0.441 | 0.747 |

tween power spectra are shown. In this case, the POLY spectrum has a narrower main lobe but slightly higher side lobes. The percent–in–band–power bandwidths are shown in Table II. The POLY spectrum is better than the other families at 90% and 99% for but is worse is terms of the 99.9% bandwidth.

## VI. CONCLUSIONS

In this paper, the problem of Continuous Phase Modulation signal design was examined for full and partial response. Using effective bandwidth and minimum distance as performance criteria, the signal shape was designed by choosing the shape that has the minimum possible effective bandwidth for a given minimum distance.

For full response signalling, the tradeoffs between performance criteria were shown to get better as the receiver observation time increased. This agrees with the intuition that increasing the length of observation leads to better performance. It was also found that values of $h$ closer to 0 or 1 led to slower performance saturation while a value of $h$ of 0.5 led to the fastest saturation. This can be explained by noting that for $h = 0.5$, the resulting signals are as dissimilar as possible. Values of $h$ near 0 or 1 lead to signals that are more similar and hence longer observations are required to distinguish them.

For the partial response case, the optimization was only possible for receiver observation times of one symbol interval. Therefore, a polynomial signal shape that is similar to previously found optimal ones was introduced. The tradeoffs for partial response signals were found to quickly overtake the performance of all full response signals. This fact shows that the memory of the signal is more important than the observation time of the receiver.

The signal shapes that were designed by using the previous optimization technique were compared against some well known signals. Of course the optimized signals gave the best tradeoff but the amount by which they were better gave an indication of the difference when other comparisons were made. A large difference in the tradeoffs resulted in a large difference in power spectra and percent–in–band–power.

Next, the power spectra were computed for a fixed value of the minimum distance. The optimal phase smoothing functions were found to generally have a narrower main lobe than the other signal shapes. This shows that the effective bandwidth measure puts emphasis on the main lobe. As a result, the optimal signals' spectra tended to have side lobes that were larger than the other signals' spectra. This was seen again in the percent–in–band–power comparisons.

In conclusion, it can be said that even though the optimization carried out in this paper does produce better signal shapes for the performance criteria used, the commonly known signal shapes are not too far from optimal performance.

## APPENDIX A
## MINIMIZATION OF $B_e$

The problem is to minimize

$$B_e^2 = \frac{4h^2}{T}\int_0^{LT} \dot{q}^2(t)\,dt$$

subject to

$$\int_0^{LT} \dot{q}(t)\,dt = 0.5.$$

Using the calculus of variations as outlined in [12,p.43] with the correspondences $y \sim q$, $x \sim t$, $a \sim 0$, $b \sim LT$, $F \sim \dot{q}^2$ and $G \sim \dot{q}$ leads to the differential equation

$$\frac{\partial}{\partial q}\left(\dot{q}^2 + \lambda\dot{q}\right) - \frac{d}{dt}\left\{\frac{\partial}{\partial \dot{q}}\left(\dot{q}^2 + \lambda\dot{q}\right)\right\} = 0.$$

Carrying out the partial derivatives, this becomes

$$0 - \frac{d}{dt}\left\{2\dot{q} + \lambda\right\} = 0.$$

After differentiating with respect to $t$, $\ddot{q} = 0$ results, which is the equation of a straight line. Applying the boundary conditions $q(0) = 0$ and $q(LT) = 0.5$ results in

$$q(t) = \begin{cases} 0 & t < 0 \\ t/(2LT) & 0 \leq t \leq LT \\ 0.5 & t > LT \end{cases}.$$

Substituting the optimal function q(t) into (5), the minimum bandwidth possible for a CPM scheme is $B_eT = \frac{h}{\sqrt{L}}$.

## APPENDIX B
## DERIVATION OF DIFFERENTIAL EQUATIONS

1) *Full Response ($N = 1$)*: Problem: minimize

$$B_e^2 = \frac{4h^2}{T}\int_0^{T} \dot{q}^2(t)\,dt$$

subject to

$$\frac{1}{T}\int_0^{T} \cos[4\pi h q(t)]\,dt = C.$$

Using the correspondences, $y \sim q$, $x \sim t$, $a \sim 0$, $b \sim T$, $F \sim \frac{4h^2}{T}\dot{q}^2$ and $G \sim \frac{1}{T}\cos[4\pi hq]$ in the the theorem of [12,p.43] results in

$$\frac{\partial}{\partial q}\left(\frac{4h^2}{T}\dot{q}^2 + \frac{\lambda}{T}\cos[4\pi hq]\right) - \frac{d}{dt}\left\{\frac{\partial}{\partial \dot{q}}\left(\frac{4h^2}{T}\dot{q}^2 + \frac{\lambda}{T}\cos[4\pi hq]\right)\right\} = 0.$$

Cancelling the $T$ and computing the partial derivatives gives

$$-4\pi h\lambda \sin[4\pi hq] = \frac{d}{dt}\left(8h^2\dot{q}\right).$$

Differentiating with respect to $t$ and rearranging gives the desired differential equation

$$\ddot{q}(t) = -\frac{\pi\lambda}{2h}\sin[4\pi hq(t)].$$

2) *Full Response ($N \geq 2$)*: Problem: minimize

$$B_e^2 = \frac{4h^2}{T}\int_0^T \dot{q}^2(t)\,dt$$

subject to

$$\frac{1}{T}\int_0^T \cos[4\pi hq(t)]\,dt = C_1$$

and

$$\frac{1}{T}\int_0^T \sin[4\pi hq(t)]\,dt = C_2.$$

Using $y \sim q$, $x \sim t$, $a \sim 0$, $b \sim T$, $F \sim \frac{4h^2}{T}\dot{q}^2$, $G_1 \sim \frac{1}{T}\cos[4\pi hq]$ and $G_2 \sim \frac{1}{T}\sin[4\pi hq]$ gives

$$\frac{\partial}{\partial q}\left(\frac{4h^2}{T}\dot{q}^2 + \frac{\lambda_1}{T}\cos[4\pi hq] + \frac{\lambda_2}{T}\sin[4\pi hq]\right) - \frac{d}{dt}\left\{\frac{\partial}{\partial \dot{q}}\left(\frac{4h^2}{T}\dot{q}^2 + \frac{\lambda_1}{T}\cos[4\pi hq] + \frac{\lambda_2}{T}\sin[4\pi hq]\right)\right\} = 0.$$

Cancelling the $T$ and carrying out the partial derivatives gives

$$-4\pi h\lambda_1 \sin[4\pi hq] + 4\pi h\lambda_2\cos[4\pi hq] = \frac{d}{dt}\left(8h^2\dot{q}\right).$$

Differentiating by $t$ and rearranging leads to the desired result,

$$\ddot{q}(t) = -\frac{\pi\lambda_1}{2h}\sin[4\pi hq(t)] + \frac{\pi\lambda_2}{2h}\cos[4\pi hq(t)].$$

3) *Full Response ($N = \infty$)* Problem: minimize

$$B_e^2 = \frac{4h^2}{T}\int_0^T \dot{q}^2(t)\,dt$$

subject to

$$\frac{1}{T}\int_0^T \{\sin(2\pi h)\sin[4\pi hq(t)] + [1+\cos(2\pi h)]\cos[4\pi hq(t)]\}\,dt = C.$$

The correspondences $y \sim q$, $x \sim t$, $a \sim 0$, $b \sim T$, $F \sim \frac{4h^2}{T}\dot{q}^2$ and $G \sim \frac{1}{T}\{\sin(2\pi h)\sin[4\pi hq(t)] + [1+\cos(2\pi h)]\cdot \cos[4\pi hq(t)]\}$ result in

$$\frac{\partial}{\partial q}\left(\frac{4h^2}{T}\dot{q}^2 + \frac{\lambda}{T}[\sin(2\pi h)\sin[4\pi hq(t)] + [1+\cos(2\pi h)]\cos[4\pi hq(t)]]\right) - \frac{d}{dt}\left\{\frac{\partial}{\partial \dot{q}}\left(\frac{4h^2}{T}\dot{q}^2 + \frac{\lambda}{T}[\sin(2\pi h)\sin[4\pi hq(t)] + [1+\cos(2\pi h)]\cos[4\pi hq(t)]]\right)\right\} = 0.$$

Cancelling the $T$ and computing the partial derivatives gives

$$4\pi h\lambda\{-\sin(2\pi h)\cos[4\pi hq(t)] + [1+\cos(2\pi h)]\sin[4\pi hq(t)]\} = \frac{d}{dt}\left(8h^2\dot{q}\right).$$

Differentiating with respect to $t$ and rearranging gives the desired differential equation

$$\ddot{q}(t) = \frac{\pi\lambda}{2h}\{-\sin(2\pi h)\cos[4\pi hq(t)] + [1+\cos(2\pi h)]\sin[4\pi hq(t)]\}.$$

## Appendix C
## Derivation of (8)

Dividing (3) into $N$ parts results in

$$\tilde{\rho}_{12} = \frac{1}{NT}\left\{\int_0^T \exp\{j2\pi h\sum_{n=0}^{\infty} b(n)q(t-nT)\}\,dt + \int_T^{2T}\exp\{j2\pi h\sum_{n=0}^{\infty} b(n)q(t-nT)\}\,dt + \cdots + \int_{(N-1)T}^{NT}\exp\{j2\pi h\sum_{n=0}^{\infty} b(n)q(t-nT)\}\,dt\right\}.$$

Since $q(t) = 0$ for $t < 0$, the sums only extend up to $n$ where $nT$ is the upper limit of integration. For $t > T$, $q(t) = 0.5$ so the above expression for $\tilde{\rho}_{12}$ can be rewritten as

$$\tilde{\rho}_{12} = \frac{1}{NT}\left\{\int_0^T \exp\{j2\pi hb(0)q(t)\}\,dt + \int_T^{2T}\exp\{j2\pi h[0.5b(0) + b(1)q(t-T)]\}\,dt + \cdots + \int_{(N-1)T}^{NT}\exp\left\{j2\pi h\left[\sum_{m=0}^{N-2}0.5b(m) + b(N-1)q[t-(N-1)T]\right]\right\}\,dt\right\}.$$

This can be reduced by taking the constant terms of the sums outside the integrals giving

$$\tilde{\rho}_{12} = \frac{1}{NT}\left\{\int_0^T \exp\{j2\pi hb(1)q(t)\}\,dt + \exp\{j\pi hb(0)\}\int_T^{2T}\exp\{j2\pi hb(1)q(t-T)\}\,dt + \cdots + \exp\{j\pi h\sum_{m=0}^{N-2}b(m)\}\cdot \int_{(N-1)T}^{NT}\exp\{j2\pi hb(N-1)q[t-(N-1)T]\}\,dt\right\}.$$

Each of the integrals may be reduced to an integral from 0 to $T$ by a substitution of the form $s = t - nT$. This results in

$$\tilde{\rho}_{12} = \frac{1}{NT}\left\{\int_0^T \exp\{j2\pi hb(0)q(t)\}\,dt + \exp\{j\pi hb(0)\}\int_0^T \exp\{j2\pi hb(1)q(t)\}\,dt + \cdots + \exp\{j\pi h\sum_{m=0}^{N-2} b(m)\}\int_0^T \exp\{j2\pi hb(N-1)q(t)\}\,dt\right\}.$$

This can now be written as a sum of integrals as

$$\tilde{\rho}_{12} = \frac{1}{N}\sum_{n=0}^{N-1}\left\{\left[\frac{1}{T}\int_0^T \exp\{j2\pi hb(n)q(t)\}\,dt\right]\cdot \exp\{j\pi h\sum_{m=0}^{n-1} b(m)\}\right\}.$$

In this equation, the term inside the square brackets looks like the cross correlation for $N = 1$. Therefore, defining $\tilde{\rho}_{b(n)}$ as before results in

$$\tilde{\rho}_{12} = \frac{1}{N}\sum_{n=0}^{N-1}\left[\tilde{\rho}_{b(n)}\exp\{j\pi h\sum_{m=0}^{n-1} b(m)\}\right].$$

## References


[1] T. Aulin and C. Sundberg, "Continuous Phase Modulation - part I: full response signalling," *IEEE Trans. Commun.*, vol. COM-29, pp. 196–209, March 1981.

[2] T. Aulin, N. Rydbeck, and C. Sundberg, "Continuous Phase Modulation - part II: partial response signalling," *IEEE Trans. Commun.*, vol. COM-29, pp. 210–225, March 1981.

[3] J. Anderson, T. Aulin, and C. Sundberg, *Digital Phase Modulation.* New York: Plenum Press, 1986.

[4] M. Campanella, U. L. Faso, and G. Mamola, "Optimum bandwidth-distance performance in full response CPM systems," *IEEE Trans. Commun.*, vol. COM-36, pp. 1110–1118, Oct. 1988.

[5] P. Kabal and S. Pasupathy, "Partial-response signalling," *IEEE Trans. Commun.*, vol. COM-23, pp. 921–934, Sept. 1975.

[6] W. Osborne and M. Luntz, "Coherent and noncoherent detection of CPFSK," *IEEE Trans. Commun.*, vol. COM-22, pp. 1023–1036, August 1974.

[7] S. Haykin, *Digital Communications.* New York: John Wiley and Sons, 1988.

[8] A. Kolmogorov and S. Fomin, *Introductory Real Analysis.* New York: Dover Publications Inc., 1975.

[9] D. Asano, "Phase smoothing functions for Continuous Phase Modulation," M.A.Sc. Thesis, Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada, 1990.

[10] L. Franks, *Signal Theory.* Englewood Cliffs: Prentice Hall, 1969.

[11] R. Hornbeck, *Numerical Methods.* New York: Quantum Publishers Inc., 1975.

[12] I. Gelfand and S. Fomin, *Calculus of Variations.* Englewood Cliffs, New Jersey: Prentice-Hall, 1963.


**David K. Asano** received the B.A.Sc. degree (with honours) from the University of British Columbia, Vancouver, Canada in 1985, the M.A.Sc. degree from the University of Toronto, Toronto, Canada in 1990 and is currently working towards the Ph.D. degree at the University of Toronto, all in Electrical Engineering. He was involved in Northern Telecom's Network Engineering Program as an Assistant Coordinator in 1990 and 1991 and as a lecturer from 1990 to 1993. From 1990 to 1992 he managed the Communications Group computer system at the University of Toronto. He has been a teaching assistant at the University of Toronto since 1988.

He is presently working on coding and modulation for fading channels.

**Harry Leib** was born in 1953. He received the B.Sc. (cum laude) and M.Sc. degrees in Electrical Engineering from the Technion–Israel Institute of Technology, Haifa, Israel in 1977 and 1984 respectively. In 1987 he received the Ph.D. degree in Electrical Engineering from the University of Toronto, Toronto, Ontario, Canada.

From 1977 to 1984 he was with the Israel Ministry of Defense, working in the area of Communications. During his Ph.D. studies at the University of Toronto (1984-87) he was a teaching and research assistant in the Department of Electrical Engineering, doing research in Digital Communications and Signal Processing. He was with the University of Toronto as a Post-doctoral Research Associate in Telecommunications (September 1987 to December 1988) and as an Assistant Professor (January 1989 -August 1989). Since September 1989 he has been an Assistant Professor in the Department of Electrical Engineering at McGill University in Montreal, where he teaches courses in communications. His current research interests are in the areas of Digital Communications, Wireless telecommunication Systems, Signal Processing, Statistical Communication and Information Theory.

At the Technion, Dr. Leib received the Gutwirth Foundation M.Sc. award in 1984. During 1985-1987, he was a recipient of a University of Toronto Doctoral Fellowship, and a Post-doctoral scholarship in 1987.

**Subbarayan Pasupathy** was born in Madras, Tamilnadu, India, on September 21, 1940. He received the B.E. degree in telecommunications from the University of Madras in 1963, the M.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 1966, and the M.Phil. and Ph.D. degree in engineering and applied science from Yale University in 1970 and 1972, respectively.

He joined the faculty of the University of Toronto in 1973 and became a Professor of Electrical Engineering in 1983. He has served as the Chairman of the Communications Group and as the Associate Chairman of the Department of Electrical Engineering at the University of Toronto. His research interests lie in the areas of communication theory, digital communications, and statistical signal processing. He is a registered Professional Engineer in the province of Ontario. During 1982-1989 he was an Editor for *Data Communications and Modulation* for the IEEE TRANSACTIONS ON COMMUNICATIONS. He has also served as a Technical Associate Editor for the IEEE COMMUNICATIONS MAGAZINE (1979- 1982) and as an Associate Editor for the *Canadian Electrical Engineering Journal*(1980-1983). Since 1984, he has been writing a regular column entitled "Light Traffic" for the IEEE COMMUNICATIONS MAGAZINE.He was elected as a Fellow of the IEEE in 1991 "for contributions to bandwidth efficient coding and modulation schemes in digital communication."