Bandwidth-Efficient Constant-Energy Trellis-Coded Modulation Schemes with Prescribed Decoding Delay

Quinn Li[1], Bixio Rimoldi[2]* and Marvin Simon[3]

[1]Broadcom Corporation
115 Route 46, Suite F-1000
Mountain Lakes, NJ 07046
email: quinnli@broadcom.com

[2]Swiss Federal Institute of Technology
Communication Systems Department
CH-1015 Lausanne, Switzerland
email: bixio.rimoldi@epfl.ch

[3]Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, California 91109-8099
email: marvin.k.simon@jpl.nasa.gov

March 26, 2001

Submitted for publication to IEEE Transactions on Information Theory

Abstract

For a general class of constant-energy trellis-coded modulation schemes with 2^ν states, necessary and sufficient conditions to guarantee that a maximum-likelihood sequence estimator can decode each symbol with a fixed delay of ν symbols are derived. Additive white Gaussian noise is assumed. MSK is a special case that belongs to the family of modulation schemes with $\nu = 1$. It is shown that when these conditions are met, the minimum squared Euclidean distance is upper bounded by $4E_s$, where E_s is the signal's energy per interval. Necessary and sufficient conditions to achieve the upper bound are given and it it shown that these conditions are met if and only if the trellis-coded modulation scheme can be implemented as pulse amplitude modulation using a pulse that extends over $\nu + 1$ symbols. Signals that achieve this upper bound and maximize the power within a given bandwidth are found. The bandwidth efficiency of such schemes is significantly higher than that of MSK.

Key words: MSK, Viterbi decoder, modulation, coding, spectral shaping, PAM.

I Introduction

Minimum shift keying (MSK) is a coded modulation scheme that can be described as consisting of an encoder followed by a memoryless modulator [1]. The encoder is a finite state machine which maintains the phase continuity of the transmitted signal. As a result, a maximum-likelihood sequence estimator for MSK may be realized using the *Viterbi algorithm* (VA) [2].

In general, the Viterbi decoder can start to produce the maximum-likelihood estimate of the transmitted signal only after it observes the channel output corresponding to the entire transmitted sequence, which results in a large decoding delay. The delay may be reduced using a suboptimal modification of the VA. Remarkably, for MSK the Viterbi decoder can optimally decode the information symbol transmitted during the nth interval upon observation of the $(n + 1)$th interval or, equivalently, with a delay of one symbol interval [3, 1]. Among other properties of MSK is the fact that MSK and antipodal modulation have the same minimum squared Euclidean distance, but the former has higher bandwidth efficiency than the latter. This paper was primarily motivated by the desire of understanding what is behind the finite decoding-delay property of MSK, how to construct more general schemes that have similar properties, and the implication of fixed decoding delay on the minimum Euclidean distance.

We are interested in such a generalization since we consider MSK as a TCM scheme in which the purpose of the trellis is to shape the spectrum of the transmitted signal rather than to achieve a large minimum distance. This and the fact that it is optimally decodable with a fixed delay, makes MSK a very special kind of TCM scheme. Unlike conventional TCM, MSK is not the result of maximizing the minimum Euclidean distance and regarding decoding delay and spectral shaping (resulting from the memory introduced by the trellis) as incidental. Ultimately, one would like to understand the possible tradeoffs between minimum distance (or more generally distance profile), bandwidth (or more generally spectral characteristic), and decoding delay. While such a goal is far beyond the scope of this paper, from our generalization of MSK we do learn that for the class of TCM schemes considered here, a finite decoding delay implies that the minimum squared Euclidean distance cannot exceed that of MSK. Hence we can only trade the amount of decoding delay with bandwidth efficiency.

The TCM schemes considered in this paper have the following characteristics: (i) the trellis is generated by a ν-length tapped delay line; (ii) transitions are associated with constant-energy waveforms which are nonzero only over the duration of the transition. Under these assumptions, we derive the necessary and sufficient conditions such that optimal decoding with a delay of ν symbol intervals is possible for the AWGN channel. We show that the requirement of a fixed decoding delay upper bounds the minimum Euclidean distance to that of MSK or, equivalently, of antipodal modulation. Necessary and sufficient conditions to achieve the upper bound are found. Finally, we consider the problem of minimizing the bandwidth among all possible choices of signals that satisfy all of the above conditions. Numerical solutions for cases where the decoding delay is constrained to one and two symbol intervals ($\nu = 1, 2$) are found.

The paper is organized as follows. The class of TCM schemes considered in this paper is described in Section II. Necessary and sufficient conditions that ensures optimal decoding with delay $(\nu + 1)T$ are obtained in Section III, where T is the symbol interval. In Section IV we show that constant-energy pulse amplitude modulation (PAM) using a pulse of duration $(\nu + 1)T$ always satisfies the conditions for decoding delay νT. We also give necessary and sufficient conditions to guarantee that the class of modulation schemes considered in this paper are implementable as pulse amplitude modulation (PAM). In Section V, we study implications on the minimum squared Euclidean distance. Finally, for schemes with $\nu = 1, 2$, signals that maximize the power within a given bandwidth while satisfying the conditions for finite-delay decoding and maximum d_{min}^2 are found numerically in Section VI.

II   System Model

In this section, we consider all TCM schemes that can be modeled as shown in Figure 1, and derive the necessary and sufficient conditions the waveform set $\mathcal{S}$ has to fulfill so that

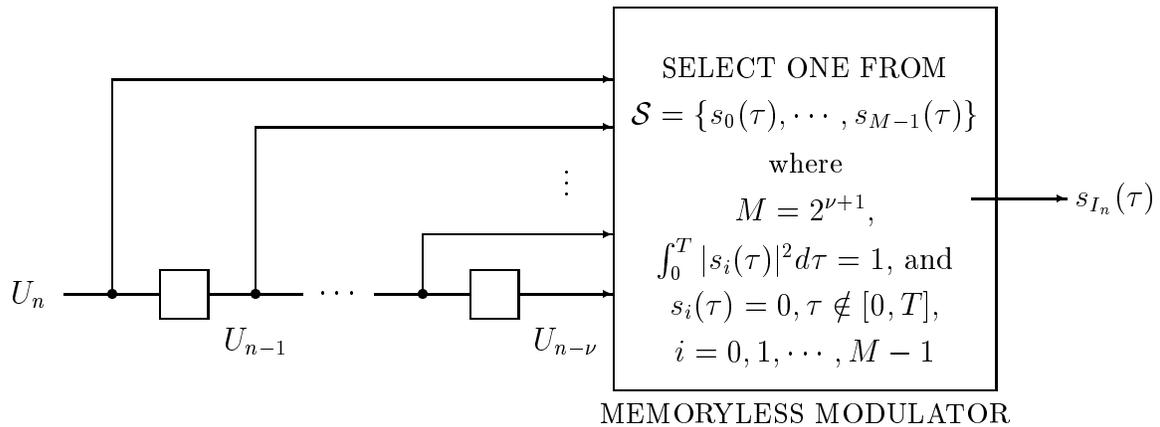optimal decoding with a fixed delay is possible using the Viterbi algorithm.

Figure 1: Trellis-coded Modulation Scheme under Consideration.

The input $\{U_n\}_{n=1}^{\infty}$ to the encoder, $U_n \in \{0, 1\}$, is a sequence of independent and identically distributed random variables referred to as the binary data sequence, and the corresponding encoder state sequence is $\{\sigma_n\}_{n=1}^{\infty}$, where $\sigma_n = U_{n-1}2^{\nu-1} + \cdots + U_{n-\nu} \in \{0, 1, \cdots, 2^{\nu} - 1\}$ is the random variable which represents the content of the shift registers in binary-coded decimal form. In the nth interval $nT \leq t < (n+1)T$, $n = 0, 1, \ldots$, the memoryless modulator outputs a signal taking value in a set $\mathcal{S} = \{s_0(\tau), \cdots, s_{M-1}(\tau)\}$ of finite-energy complex-valued baseband waveforms, where $\tau = t - nT$. For all $i \in \{0, 1, \cdots, 2^{\nu+1} - 1\}$, $s_i(\tau)$ is assumed to be zero outside the interval $[0, T]$ and normalized to have unit energy, i.e.,

$$||s_i||^2 = < s_i, s_i > = 1 \; \forall i \in \{0, 1, \cdots, 2^{\nu+1} - 1\}, \tag{1}$$

where

$$< s_i, s_j > = \int_0^T s_i(\tau) s_j^*(\tau) d\tau$$

denotes inner product.

Without loss of generality, the index I_n of the transmitted signal in the nth interval is the random variable given by the input and the state at time n in the following way,

$$I_n = U_n 2^{\nu} + \sigma_n. \tag{2}$$

Notice that the current input bit U_n is the most significant digit of I_n. For example, in the case where the encoder memory $\nu = 1$ (see Figure 2), the index I_n is related to (U_n, σ_n) according to Table 1. The last column of Table 1 specifies the signals for MSK, where it is assumed that $f_0 T$ is integer and that $f_1 T = f_0 T + 1/2$ (see e.g., [4]).

Table 1: The mapping performed by the memoryless modulator for $\nu = 1$.

U_n	σ_n	$s_{I_n}(\tau)$	MSK
0	0	$s_0(\tau)$	$\sqrt{\frac{2}{T}} \cos 2\pi f_0 t$
0	1	$s_1(\tau)$	$-\sqrt{\frac{2}{T}} \cos 2\pi f_1 t$
1	0	$s_2(\tau)$	$\sqrt{\frac{2}{T}} \cos 2\pi f_1 t$
1	1	$s_3(\tau)$	$-\sqrt{\frac{2}{T}} \cos 2\pi f_0 t$

Throughout the paper, U, I and σ represent random variables, whereas u and i represent elements in the alphabet of U and I, respectively.

The baseband transmitted signal of the coded modulation scheme can be expressed as

$$\sum_{n=0}^{\infty} s_{I_n}(t - nT), \ I_n \in \{0, 1, \cdots, 2^{\nu+1} - 1\}, \ t > 0. \tag{3}$$

The corresponding received signal is

$$R(t) = \sum_{n=0}^{\infty} s_{I_n}(t - nT) + Z(t), \ t > 0, \tag{4}$$

where $Z(t)$ is a zero-mean complex-valued white Gaussian process with power spectral density N_0.

III   Decoding Delay

For a TCM scheme, a *maximum-likelihood sequence estimator* (MLSE) minimizes the probability of error at the receiver. The Viterbi algorithm is an efficient method for implementing the MLSE[5]. The branch metric in the nth interval is given according to[1]

$$\lambda_n(s_{i_n}) = \Re\{\int_0^T R^*(\tau + nT)s_{i_n}(\tau) \, d\tau\}, \ i_n \in \{0, 1, \cdots, 2^{\nu+1} - 1\}, \tag{5}$$

where $\Re\{.\}$ denotes the real part of the enclosed complex number, and $R^*(t)$ is the complex conjugate of $R(t)$. A maximum-likelihood estimate of the transmitted sequence is a sequence

---

[1]We have used the fact that the signals have equal energy.

through the trellis with the largest sum metric. For optimality, the Viterbi decoder in general needs to observe the entire transmitted sequence before finding a maximum-likelihood estimate. As a result, the receiver experiences a large decoding delay. Suboptimal algorithms such as the *truncated Viterbi algorithm* [6] can be used to alleviate the delay problem.

It has been shown that MSK can be modeled as shown in Figure 1 with $\nu = 1$, and the receiver for MSK can decode optimally observing two symbol intervals using the VA [3, 1]. We are interested in generalizing this property of MSK to the class of TCM schemes shown in Figure 1. Specifically, we are looking for a set of necessary and sufficient conditions that the set of waveforms $\{s_i(\tau), i \in \{0, 1, \cdots, 2^{\nu+1} - 1\}\}$ has to fulfill so that the Viterbi decoder can optimally decode with a fixed delay of ν symbol intervals, i.e., the decoder is able to produce the estimate of the nth information symbol U_n upon observation of the received signal for $t \le (n + \nu + 1)T$.

Theorem 1 For any coded modulation scheme that can be modeled as shown in Figure 1 and where, without loss of generality, we assume the indexing described by (2), the necessary and sufficient conditions the signals $\{s_i(\tau), i \in \{0, 1, \cdots, 2^{v+1} - 1\}\}$ have to fulfill so that the Viterbi decoder can make an optimal decision with a delay of ν symbol intervals are

$$s_0(\tau) - s_{2^m}(\tau) = s_b(\tau) - s_{b+2^m}(\tau)$$
$$m \in \{0, 1, \cdots, \nu - 1\}, \qquad b \in \{2^{m+1}l : l = 1, \cdots, 2^{\nu-m} - 1\}. \tag{6}$$

Furthermore, (6) implies that the decision rule for the nth information symbol is given by

$$\sum_{i=n}^{n+\nu} \lambda_i\left(s_0 - s_{2^{n+\nu-i}}\right) \quad \begin{matrix} \hat{U}_n = 0 \\ \uparrow \\ > \\ < \\ \downarrow \\ \hat{U}_n = 1 \end{matrix} \quad 0. \tag{7}$$

Proof of Theorem 1. Let q and p be two arbitrary but fixed elements of the state alphabet $\{0, 1, \cdots, 2^{\nu} - 1\}$. There are exactly two paths which diverge at state p at time n and merge at state q at time $n + \nu + 1$. This is true since the paths have length $\nu + 1$ and the final state is in one-to-one correspondence with the last ν inputs. More precisely, if the binary representation of q is $(u_{n+\nu}, \cdots, u_{n+1})$, then these two paths correspond to input sequences $(0, u_{n+1}, \cdots, u_{n+\nu})$ and $(1, u_{n+1}, \cdots, u_{n+\nu})$, where u_{n+j} is the input at time $n + j$, $j = 1, 2, \cdots, \nu$.

Assume first that a genie informs the decoder that the maximum-likelihood estimate of the transmitted sequence passes through state p at time n and state q at time $n + \nu + 1$. Under these conditions, the maximum likelihood estimate for U_n is given by $\hat{U}_n = 0$ if

$$\lambda_n(s_{i_n}) + \lambda_{n+1}(s_{i_{n+1}}) + \cdots + \lambda_{n+\nu}(s_{i_{n+\nu}}) > \lambda_n(s_{i'_n}) + \lambda_{n+1}(s_{i'_{n+1}}) + \cdots + \lambda_{n+\nu}(s_{i'_{n+\nu}}), \quad (8)$$

where $s_{i_{n+j}}(\tau)$ and $s_{i'_{n+j}}(\tau)$, $j = 0, 1, \cdots, \nu$ are the transmitted signals in interval i_{n+j} corresponding to input sequences $(0, u_{n+1}, \cdots, u_{n+\nu})$ and $(1, u_{n+1}, \cdots, u_{n+\nu})$, respectively. Since $\lambda_n(.)$ is linear in its argument, the decision rule can be rewritten as: $\hat{U}_n = 0$ if

$$\lambda_n(s_{i_n} - s_{i'_n}) + \lambda_{n+1}(s_{i_{n+1}} - s_{i'_{n+1}}) + \cdots + \lambda_{n+\nu}(s_{i_{n+\nu}} - s_{i'_{n+\nu}}) > 0. \tag{9}$$

The decoder can produce an estimate of U_n upon observing the received signal up to and including interval $n + \nu$ if and only if (9) results in the same decision regardless of q. This is true since by varying the noise process in the intervals $n + \nu + 1, n + \nu + 2, \cdots$, we can force q to take on any desired value of the state alphabet. But (9) is true (or false) independently of q if and only if $\forall j \in \{0, 1, \cdots, \nu\}$

$$s_{i_{n+j}}(\tau) - s_{i'_{n+j}}(\tau) \text{ is independent of the value of } q. \tag{10}$$

This can be seen as follows. Fix $j \in \{0, 1, \cdots, \nu\}$, let e and f be two elements of the state alphabet, and let $u_e(\tau) = s_{i_{n+j}}(\tau) - s_{i'_{n+j}}(\tau)$ when $q = e$ and let $u_f(\tau) = s_{i_{n+j}}(\tau) - s_{i'_{n+j}}(\tau)$ when $q = f$. If $u_e(\tau) \neq u_f(\tau)$, then one can choose the noise in the $(n+j)$th interval so that $\lambda_{n+j}(u_e)$ is, say, positive whereas $\lambda_{n+j}(u_f)$ is negative. By choosing the noise appropriately in each interval $i = n, n+1, \ldots, n+\nu$, one can make the left side of (9) have a sign that depends on the value of q.

The binary representation of i_{n+j} and i'_{n+j} are

$$u_{n+j}, \cdots, u_{n+1}, 0, u_{n-1}, \cdots, u_{n+j-\nu} \text{ and } u_{n+j}, \cdots, u_{n+1}, 1, u_{n-1}, \cdots, u_{n+j-\nu}, \tag{11}$$

respectively. Moreover, the first j digits in (11) are the j least significant bits in the binary representation of q, whereas the last $\nu - j$ digits are the $\nu - j$ most significant bits in the binary representation of p. Hence

$$\begin{aligned} i_{n+j} &= q_j + p_j \\ i'_{n+j} &= q_j + 2^{\nu-j} + p_j, \end{aligned}$$

where $q_j \in \{2^{\nu-j+1}l : l = 0, 1, \cdots, 2^j - 1\}$ is determined by q, and $p_j \in \{0, 1, \cdots, 2^{\nu-j} - 1\}$ is determined by p. With this notation (10) becomes

$$s_{p_j}(\tau) - s_{p_j+2^{\nu-j}}(\tau) = s_{q_j+p_j}(\tau) - s_{q_j+p_j+2^{\nu-j}}(\tau),$$

$$j \in \{0, 1, \cdots, \nu\}, \qquad q_j \in \{2^{\nu-j+1}l : l = 0, 1, \cdots, 2^j - 1\}, \qquad p_j \in \{0, 1, \cdots, 2^{\nu-j} - 1\}.$$

Notice that when $q_j = 0$ the above set of equations is trivially satisfied. Hence we may remove 0 from the range of q_j. We may also remove 0 from the range of j since $j = 0$ implies

$q_j = 0$. Finally, letting $m = \nu - j$ and renaming the (dummy) variable q_j and p_j with a and b, respectively, we obtain an equivalent set of equations:

$$s_a(\tau) - s_{a+2^m}(\tau) = s_{a+b}(\tau) - s_{a+b+2^m}(\tau), \tag{12}$$

$$m \in \{0, 1, \cdots, \nu - 1\}, \qquad a \in \{0, 1, \cdots, 2^m - 1\}, \qquad b \in \{2^{m+1}l : l = 1, \cdots, 2^{\nu-m} - 1\}.$$

It is useful to remember that a and b refer to the initial and final state, respectively.

Lemma 8 in the Appendix shows that the above system of equalities is satisfied if and only if it is satisfied for $a = 0$. In words, it is satisfied for all initial states if and only if it is satisfied for one (e.g. the zero) initial state. This proves that (6) is necessary and sufficient to decode with delay νT. Moreover, when (6) holds neither the initial nor the final state affects the arguments of λ in (9). Hence, in computing the left side of (9) we may assume that $p = q = 0$. Thus (9) is equivalent to (7), completing the proof. $\qquad\square$

Example 1 For $\nu = 1$ the transmitter is shown in Figure 2. The state transition diagram for the nth and $(n+1)$th intervals is shown in Figure 3, where the state σ_n corresponds to the state of the encoder at time n, and the transition between states is labeled with the input symbol and the output signal corresponding to that transition. Figure 4 shows the state transition diagram with branches labeled with the corresponding metric. From Theorem 1, the necessary and sufficient condition for the signal set $\{s_i(\tau), i = 0, 1, 2, 3\}$ to ensure optimal decoding with one symbol interval delay is

$$s_0(\tau) - s_1(\tau) = s_2(\tau) - s_3(\tau), \tag{13}$$

and the maximum-likelihood decision rule for the nth information symbol is given by

$$\lambda_n(s_0 - s_2) + \lambda_{n+1}(s_0 - s_1) \quad \begin{matrix} \hat{U}_n = 0 \\ \uparrow \\ \gtrless \\ \downarrow \\ \hat{U}_n = 1 \end{matrix} \quad 0.$$

Equation (13) may be rewritten as

$$s_0(\tau) - s_2(\tau) = s_1(\tau) - s_3(\tau). \tag{14}$$

Equation (14) says that the ordered difference between the signals of two transitions that diverge from any state is the same, whereas equation (13) states that the ordered difference between the signals of two transitions that merge to any state is the same. Hence, the

necessary and sufficient condition which ensures optimal decoding with a delay of one symbol interval also guarantees that the Euclidean distance between any pair of paths which diverge at time n and merge at time $n + 2$ is the same. We will see that this is not a coincidence.

Using Table 1 we verify that MSK fulfills equation (13) (hence equation (14)). Thus, as it is well known, MSK may be decoded with 1 symbol delay (see e.g. [4]).
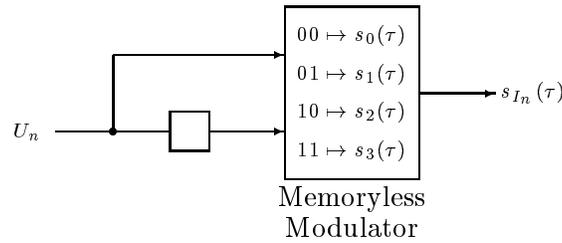

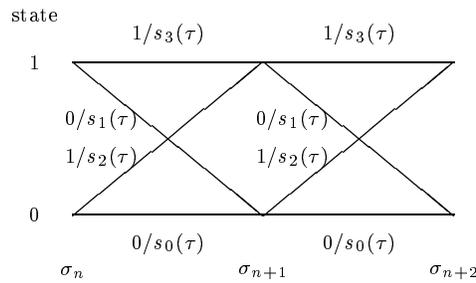Figure 2: Transmitter of the Proposed TCM Scheme for $\nu = 1$.

Figure 3: State Transition Diagram for $\nu = 1$.

Example 2 For $\nu = 2$ the transmitter is depicted in Figure 5. The state transition diagram for the nth, $(n + 1)$th, and $(n + 2)$th intervals is shown in Figure 6. From Theorem 1, we have that the necessary and sufficient conditions for the signal set $\{s_i(\tau), i = 0, 1, \cdots, 7\}$ to ensure optimal decoding with two-symbol interval delay are

$$
\begin{aligned}
s_0(\tau) - s_1(\tau) &= s_2(\tau) - s_3(\tau) = s_4(\tau) - s_5(\tau) = s_6(\tau) - s_7(\tau) \\
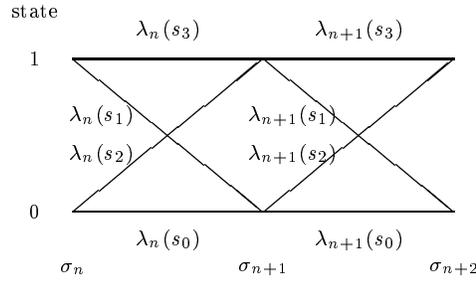s_0(\tau) - s_2(\tau) &= s_4(\tau) - s_6(\tau),
\end{aligned}
\tag{15}
$$

Figure 4: State Transition Diagram Labeled with Path Metric.

and the maximum-likelihood decision rule for the nth information symbol is given by

$$
\lambda_n(s_0 - s_4) + \lambda_{n+1}(s_0 - s_2) + \lambda_{n+2}(s_0 - s_1) \quad
\begin{array}{c}
\hat{U}_n = 0 \\
\uparrow \\
\gtrless \\
\downarrow \\
\hat{U}_n = 1
\end{array}
\quad 0.
$$

It is straightforward to verify that (15) implies

$$
\begin{aligned}
s_0(\tau) - s_1(\tau) &= s_2(\tau) - s_3(\tau) = s_4(\tau) - s_5(\tau) = s_6(\tau) - s_7(\tau) \\
s_0(\tau) - s_2(\tau) &= s_1(\tau) - s_3(\tau) = s_4(\tau) - s_6(\tau) = s_5(\tau) - s_7(\tau) \\
s_0(\tau) - s_4(\tau) &= s_1(\tau) - s_5(\tau) = s_2(\tau) - s_6(\tau) = s_3(\tau) - s_7(\tau).
\end{aligned}
\tag{16}
$$

It can be seen from Figure 6 that the above set of equations ensure that the Euclidean distance between any pair of paths that diverge at time n and merge at time $n + 3$ is the same.
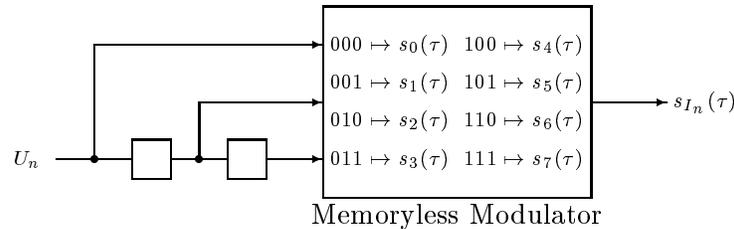

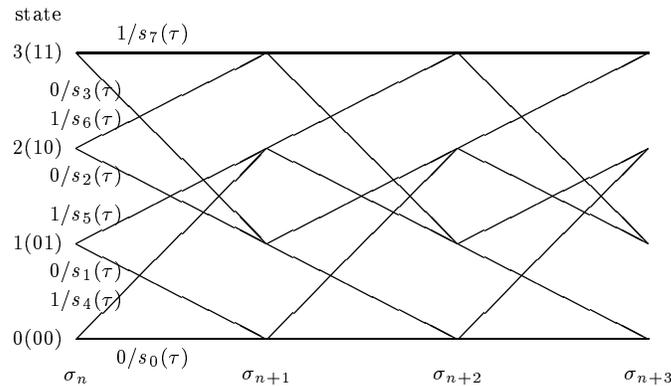Figure 5: Transmitter of the Proposed TCM Scheme for $\nu = 2$.

Figure 6: State Transition Diagram for $\nu = 2$.

Notice from the maximum-likelihood decision rule (7) that the required number of matched filters or correlators at the receiver is $\nu + 1$. When no conditions are imposed on the decoding delay, the required number of matched filters at the receiver may be as large as $2^{\nu+1}$ (i.e., the cardinality of the signal set), even when the truncated VA is used.

We now derive some conditions implied by (6) of Theorem 1.

Lemma 1 Condition (6) of Theorem 1 holds if and only if for all $I = \sum_{i=0}^{\nu} U_{-i} 2^{\nu-i}$, where $U_{-i} \in \{0, 1\}$,

$$s_I - s_0 = \sum_{i=0}^{\nu} U_{-i}(s_{2^{\nu-i}} - s_0). \tag{17}$$

Proof First of all we rewrite condition (6) of Theorem 1 in a more suitable form. From (i) of Lemma 8, an equivalent condition is $s_{a+b+2^m} = s_{a+b} + s_{2^m} - s_0$, $m \in \{0, 1, \cdots, \nu\}$, $a \in \{0, 1, \cdots, 2^m - 1\}$, $b \in \{2^{m+1}l : l = 0, 1, \cdots, 2^{\nu-m} - 1\}$. Notice that $a + b$ represents any number $c \in \{0, 1, \ldots, 2^{\nu+1} - 1\}$ with the coefficient of 2^m in the binary representation of c equal to 0. Thus condition (6) of Theorem 1 holds if and only if for every such c,

$$s_{c+2^m} = s_c + s_{2^m} - s_0. \tag{18}$$

Assume that this conditions is satisfied. Then for any such number c,

$$s_{c+U2^m} = s_c + U(s_{2^m} - s_0),$$

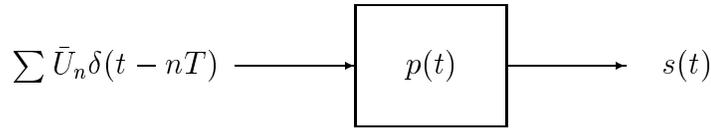$$\sum \bar{U}_n \delta(t - nT) \longrightarrow \boxed{p(t)} \longrightarrow s(t)$$

Figure 7: Pulse Amplitude Modulation

is a valid identity when $U = 1$ and it is trivially valid when $U = 0$. Using this repeatedly with decreasing values of c (starting with $c = \sum_{i=0}^{\nu-1} U_{-i} 2^{\nu-i}$), yields

$$
\begin{aligned}
s_I &= s_{0 + \sum_{i=0}^{\nu-1} U_{-i} 2^{\nu-i} + U_{-\nu} 2^0} \\
&= s_{0 + \sum_{i=0}^{\nu-1} U_{-i} 2^{\nu-i}} + U_{-\nu}(s_{2^0} - s_0) \\
&\;\;\vdots \\
&= s_0 + \sum_{i=0}^{\nu} U_{-i}(s_{2^{\nu-i}} - s_0).
\end{aligned}
$$

This proves the forward part of the theorem.

To prove the converse, let c be as specified above and let $C = c + 2^m$. (Hence the binary expansions of c and C are identical except for the coefficient of 2^m). Then

$$s_C - s_c = s_{2^m} - s_0,$$

where we used (17) on both, s_C and s_c. This shows that (18) holds.   □

IV   PAM Implementation

We now show that PAM always satisfies the condition of Theorem 1, an observation that has been reported in [7]. Consider PAM as shown in Fig. 7 where the complex pulse $p(t)$ extends over $\nu + 1$ symbol intervals. Without loss of generality we assume that the support set of $p(t)$ is $[0, (\nu + 1)T)$. The inputs are binary symbols $\bar{U}_i \in \{\pm 1\}$ and the output during

the nth symbol interval $t \in [nT, (n+1)T)$ is

$$s(t) = \sum_{i=-\infty}^{\infty} \bar{U}_i p(t - iT) \tag{19}$$

$$= \sum_{i=0}^{\nu} \bar{U}_{n-i} p(t - (n-i)T) \tag{20}$$

$$= \sum_{i=0}^{\nu} \bar{U}_{n-i} p_i(t - nT) \tag{21}$$

$$\tag{22}$$

where for $i = 0, 1, \ldots, \nu$,

$$p_i(\tau) = \begin{cases} p(\tau + iT), & \tau \in [0, T) \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

Without loss of generality, we denote by s_I, $I = \sum_{i=0}^{\nu} U_{-i} 2^{\nu-i}$, $U_i = \frac{1 - \bar{U}_i}{2}$, the signal set for the interval $t \in [0, T)$.

Lemma 2 For the PAM modulator shown in Fig. 7, the set of output signals satisfies both of the following conditions:

   (i) Condition (6) of Theorem 1

  (ii) $s_{2^{\nu+1}-1} = -s_0$.

Proof To prove that the first condition is fulfilled it is enough to show that $s_I, I = 0, 1, \ldots 2^{\nu+1} - 1$, satisfies (17). We check this for $t \in [0, T)$, but for notational convenience we do not explicitly show the time variable t.

$$s_I - s_0 \overset{(a)}{=} \sum_{i=0}^{\nu} \bar{U}_{-i} p_i - \sum_{i=0}^{\nu} p_i \tag{24}$$

$$= \sum_{i=0}^{\nu} (\bar{U}_{-i} - 1) p_i \tag{25}$$

$$= \sum_{i=0}^{\nu} -2 U_{-i} p_i \tag{26}$$

$$\overset{(b)}{=} \sum_{i=0}^{\nu} U_{-i} (s_{2^{\nu-i}} - s_0), \tag{27}$$

where in (a) we used the fact that $s_0 = \sum_{i=0}^{\nu} p_i$ (all relevant inputs to the PAM modulator are 1s) and in (b) that $\frac{s_0 - s_{2^{\nu-i}}}{2} = p_i$, as can be easily verified using superposition. The second condition is verified since the relevant $\bar{U}_{-i}$ are 1s for s_0 and -1s for $s_{2^{\nu+1}-1}$. □

Notice that in order to be decodable with delay νT, in addition to condition (6) of Theorem 1, the PAM signal must satisfy the equal-energy constraint. The next Lemma says when this is the case:

Lemma 3 The following conditions are equivalent:

(i) $|| \sum_{i=0}^{\nu} \bar{U}_{-i} p_i ||^2 = 1$ for all $\bar{U}_0, \ldots \bar{U}_{-\nu} \in \{\pm 1\}^{\nu+1}$

(ii) $\sum_{i=0}^{\nu} ||p_i||^2 = 1$ and $< p_i, p_j > = 0$, $i \neq j$.

Proof

$$1 = || \sum_{i=0}^{\nu} \bar{U}_{-i} p_i ||^2 \tag{28}$$

$$= < \sum_{i=0}^{\nu} \bar{U}_{-i} p_i, \sum_{i=0}^{\nu} \bar{U}_{-i} p_i > \tag{29}$$

$$= \sum_{i=0}^{\nu} ||p_i||^2 + \sum_{i \neq j} \bar{U}_{-i} \bar{U}_{-j} < p_i, p_j > . \tag{30}$$

$$\tag{31}$$

But the last line equals 1 regardless of $\bar{U}_{-i}$ and $\bar{U}_{-j}$ iff the conditions of the Lemma are satisfied. □

The following lemma considers the reverse problem.

Lemma 4 The modulator of Fig. 1 may be implemented as a pulse amplitude modulator as shown in Fig. 7 iff

(i) Condition (6) of Theorem 1 is satisfied

(ii) $s_0 = -s_{2^{\nu+1}-1}$.

Proof We already know from Lemma 2 that a PAM modulator satisfies both of the above conditions. Hence the conditions are necessary. Now we show that they are sufficient.

Suppose we are given a modulator as shown in Fig. 1 and that for this modulator the conditions above are satisfied. Let $I = \sum_{i=0}^{\nu} U_{-i} 2^{\nu-i}$. Then

$$s_I \overset{(a)}{=} \sum_{i=0}^{\nu} U_{-i}\big(s_{2^{\nu-i}} - s_0\big) + s_0 \tag{32}$$

$$= \sum_{i=0}^{\nu} (1 - 2U_{-i})\frac{s_0 - s_{2^{\nu-i}}}{2} + \sum_{i=0}^{\nu} \frac{s_{2^{\nu-i}} - s_0}{2} + s_0 \tag{33}$$

$$\overset{(b)}{=} \sum_{i=0}^{\nu} (1 - 2U_{-i})\frac{s_0 - s_{2^{\nu-i}}}{2} + \frac{s_{2^{\nu+1}-1} - s_0}{2} + s_0 \tag{34}$$

$$= \sum_{i=0}^{\nu} (1 - 2U_{-i})\frac{s_0 - s_{2^{\nu-i}}}{2} + \frac{s_{2^{\nu+1}-1} + s_0}{2} \tag{35}$$

$$\overset{(c)}{=} \sum_{i=0}^{\nu} (1 - 2U_{-i})\frac{s_0 - s_{2^{\nu-i}}}{2} \tag{36}$$

$$\overset{(d)}{=} \sum_{i=0}^{\nu} \bar{U}_{-i} p_i \tag{37}$$

$$\tag{38}$$

Where in (a) and (b) we used Lemma 1 (which is equivalent to condition (i) above), (c) follows from the second condition, and (d) from letting

$$p_i = \frac{s_0 - s_{2^{\nu-i}}}{2} \tag{39}$$

and $\bar{U}_{-i} = 1 - 2U_{-i}$.   □

Comments: (1) We have already seen that the first condition of the above Lemma guarantees bounded decoding delay. We will see in the next section that the second condition maximizes the minimum squared Euclidean distance; (2) When we change the sign of all $\nu+1$ consecutive inputs that determine an output waveform, say s_I, of a PAM implementation, the output waveform changes sign. But changing $\bar{U}_{-i}$ from 1 to -1 corresponds to changing U_{-i} from 0 to 1. This means that the signal with index I is the negative of the signal with index $2^{\nu+1} - 1 - I$. Indeed, the sum of the two indices has to be $2^{\nu+1} - 1$ since it has the all-one binary expansion.

Example 3 We know already that MSK satisfies (i) of the above Lemma. Using Table 1 we see that it also satisfies (ii). Hence, MSK may be implemented as PAM. This is also a

known fact (see e.g. [4]). Using $p(t) = p_0(t) + p_1(t - T)$, (39), and Table 1, we obtain

$$p_0(t) = \begin{cases} \sqrt{\frac{1}{2T}} \left(\cos 2\pi f_0 t + \cos 2\pi f_1 t \right), & t \in [0, T) \\ 0, & \text{otherwise,} \end{cases}$$

$$p_1(t) = \begin{cases} \sqrt{\frac{1}{2T}} \left(\cos 2\pi f_0 t - \cos 2\pi f_1 t \right), & t \in [0, T) \\ 0, & \text{otherwise,} \end{cases}$$

$$p_1(t - T) = \begin{cases} \sqrt{\frac{1}{2T}} \left(\cos 2\pi f_0 t + \cos 2\pi f_1 t \right), & t \in [T, 2T) \\ 0, & \text{otherwise,} \end{cases}$$

$$p(t) = \begin{cases} \sqrt{\frac{1}{2T}} \left(\cos 2\pi f_0 t + \cos 2\pi f_1 t \right), & t \in [0, 2T) \\ 0, & \text{otherwise.} \end{cases}$$

V   Bounds on the Minimum Euclidean Distance

For the AWGN channel, the probability of error of a TCM scheme is essentially determined by the minimum squared Euclidean distance D^2_{min} between any pair of output signals. In this section we upper-bound D^2_{min} and give necessary and sufficient conditions to achieve the upper-bound.

We begin with the following:

Lemma 5 For any modulator as in Fig. 1 that satisfies condition (6) of Theorem 1, the squared Euclidean distance satisfies $D^2_{min} \leq 4$, where the inequality is strict unless $s_0 = -s_{2^{\nu+1}-1}$, i.e., unless the modulator admits a PAM implementation.

Proof It is sufficient to consider any two input sequences that differ in exactly one position.

The corresponding squared Euclidean distance is

$$\sum_{i=1}^{\nu} ||s_0 - s_{2^i}||^2 = \sum_{i=1}^{\nu} < s_0 - s_{2^i}, s_0 - s_{2^i} > \tag{40}$$

$$= \sum_{i=1}^{\nu} \left\{ ||s_0||^2 + ||s_{2^i}||^2 - 2\Re\{< s_0, s_{2^i} >\} \right\} \tag{41}$$

$$= \sum_{i=1}^{\nu} \left\{ 2||s_0||^2 - 2\Re\{< s_0, s_{2^i} >\} \right\} \tag{42}$$

$$= 2\Re\{< s_0, s_0 - s_{2^{\nu+1}-1} >\} \tag{43}$$

$$= 2||s_0||^2 - 2\Re\{< s_0, s_{2^{\nu+1}-1} >\} \tag{44}$$

$$\overset{(a)}{\leq} 4||s_0||^2 \tag{45}$$

$$= 4, \tag{46}$$

where in (a) we used Cauchy inequality. Equality holds if and only if $s_0(\tau) = k s_{2^{\nu+1}-1}(\tau)$ for some negative number k. Since the signals have unit energy, k must be -1. □

Now we proceed to prove that when the modulator of Fig. 1 admits a PAM implementation, then $D_{min}^2 = 4$. The following Lemma holds also when no PAM implementation exists.

Lemma 6 Whenever condition (6) of Theorem 1 holds,

$$s_{\sum_{i=0}^{\nu} a_{-i}2^{\nu-i}} - s_{\sum_{i=0}^{\nu} b_{-i}2^{\nu-i}} = \sum_{i=0}^{\nu} (a_{-i} - b_{-i})(s_{2^{\nu-i}} - s_0).$$

Proof From Lemma 1

$$s_{\sum_{i=0}^{\nu} a_{-i}2^{\nu-i}} - s_{\sum_{i=0}^{\nu} b_{-i}2^{\nu-i}} = \sum_{i=0}^{\nu} a_{-i}(s_{2^{\nu-i}} - s_0) - \sum_{i=0}^{\nu} b_{-i}(s_{2^{\nu-i}} - s_0)$$

$$= \sum_{i=0}^{\nu} (a_{-i} - b_{-i})(s_{2^{\nu-i}} - s_0)$$

□

Lemma 7 Whenever the modulator of Fig. 1 admits a PAM implementation as in Fig. 7 (i.e., the conditions of Lemma 4 are satisfied),

$$\left|\left| s_{\sum_{i=0}^{\nu} a_{-i}2^{\nu-i}} - s_{\sum_{i=0}^{\nu} b_{-i}2^{\nu-i}} \right|\right|^2 = 4 \sum_{i=0}^{\nu} (a_{-i} - b_{-i})^2 ||p_i||^2.$$

Proof It follows immediately from the previous lemma and the fact that p_i and p_j are orthogonal when $i \neq j$ (Lemma 3).   □

The above Lemma says that if I and J differ in the ith position, then this difference contributes to $||s_I - s_J||^2$ by the amount $4||p_i||^2$. Hence, we have the following immediate consequence:

Corollary 1 Consider two infinite input sequences to the pulse amplitude modulator of Fig. 7. Each position in which the two sequences differ contributes to the squared Euclidean distances by the amount $4 \sum_{i=1}^{\nu} ||p_i||^2 = 4$. Hence the minimum squared Euclidean distance is the distance between any two outputs whose input sequences differ in exactly one position.

□

Many authors use the *normalized minimum squared Euclidean distance* d_{min}^2 given by

$$d_{min}^2 = \frac{D_{min}^2}{2E_b} = \frac{D_{min}^2}{2},\tag{47}$$

where the second equality comes from the fact that $E_b = 1$ for the schemes under consideration. We have shown that for the modulator of Fig. 1, $d_{min}^2 \leq 2$ with equality if and only if it admits a PAM implementation. Notice that $d_{min}^2 = 2$ for antipodal modulation. Also for MSK $d_{min}^2 = 2$, as it must be since MSK can be implemented as in Fig. 1 as PAM.

VI   Optimization of Bandwidth Efficiency

In this section we focus on bandwidth efficiency optimization, applied to the subset of the modulation schemes of Fig. 1 for which $d_{min}^2 = 2$. As we have seen in the previous section, these are the modulation schemes that admit the PAM implementation of Fig. 7.

There are various definitions of bandwidth that one can use. We are interested in the bandwidth that contains a certain percent of the total energy. We say that a modulation scheme has $100\eta\%$ bandwidth B, where $0 < \eta < 1$, if the width of the smallest frequency interval that contains $100\eta\%$ of the signal's energy is B. Mathematically, a modulation scheme that produces a process of power spectral density $G(f)$ has $100\eta\%$ bandwidth B if

$$\eta = \frac{\max_{a \in \mathbb{R}} \int_{a-\frac{B}{2}}^{a+\frac{B}{2}} G(f)df}{\int_{-\infty}^{\infty} G(f)df}.$$

The maximization over the *center frequency* a is needed since it is not guaranteed that the spectrum is symmetric with respect to $f = 0$. We cannot even assume that the spectrum be symmetric with respect to some frequency $f = a$.

Instead of fixing η and finding the smallest B that solves the above equation for one modulation scheme in the class of interest, it is more convenient to fix B and find the largest η over the same class of modulation schemes. Either way, for each ν we obtain a function that relates η and B. Hence we define

$$\eta(B) = \max \frac{\max_{a \in \mathbb{R}} \int_{a-\frac{B}{2}}^{a+\frac{B}{2}} G(f) df}{\int_{-\infty}^{\infty} G(f) df} \tag{48}$$

where the maximum is over the modulation schemes with a given ν. We will carry out the maximization numerically for $\nu = 1$ and $\nu = 2$.

Now we proceed to specify the power spectral density (PSD) $G(f)$ of the process produced by the modulation shown in Figure 7. Letting $P(f)$ be the Fourier transform of the impulse response $p(t)$ and assuming that input symbols are i.i.d., we obtain

$$G(f) = \frac{1}{T}|P(f)|^2. \tag{49}$$

Furthermore, using (23) and (39) we obtain

$$P(f) = \sum_{i=0}^{\nu} P_i(f) e^{-j2\pi i f T} = \frac{1}{2} \sum_{i=0}^{\nu} [S_0(f) - S_{2\nu-i}(f)] e^{-j2\pi i f T},$$

where $S_i(f)$ is the Fourier transform of $s_i(t)$. In particular, the PSD for modulation schemes with $\nu = 1$ is given by

$$G(f) = \frac{1}{4T}|S_0(f) + S_1(f) + [S_0(f) - S_1(f)] e^{-j2\pi f T}|^2,$$

and for $\nu = 2$ by

$$G(f) = \frac{1}{4T}|S_1(f) + S_2(f) + [S_0(f) - S_2(f)] e^{-j2\pi f T} + [S_0(f) - S_1(f)] e^{-j4\pi f T}|^2.$$

Since scaling the time axis scales the frequency axis leaving BT constant, without loss of generality we may normalize the time axis so that $T = 1$.

The right side of (48) is carried out numerically using the the Matlab optimization toolbox function "fmincon"[2]. The optimal signals are represented by 25 uniformly spaced samples in the interval (0,1).

---

[2]It should be noted that the solution found is not guaranteed to be global.

Figure 8 plots the minimum fractional out-of-band power (i.e. $1 - \eta(B)$) for modulation schemes with $\nu = 1$ and 2. Also shown in the figure is the fractional out-of-band power for MSK. We see that even for $\nu = 1$ the scheme obtained form the optimization procedure is more bandwidth efficient than MSK.

For $\nu = 1$ and $B = 0.8$, the power spectrum $G(f)$ and the optimal signals $s_0(\tau)$ and $s_1(\tau)$ are shown in Figure 9. The power spectrum $G(f)$ and the optimal signals $s_0(\tau)$ and $s_1(\tau)$ for $\nu = 1$ and $B = 1.2$ are shown in Figure 10. For $\nu = 2$ and $B = 1.2$, the power spectrum $G(f)$ and the optimal signals $s_0(\tau)$, $s_1(\tau)$, and $s_2(\tau)$ are shown in Figure 11.
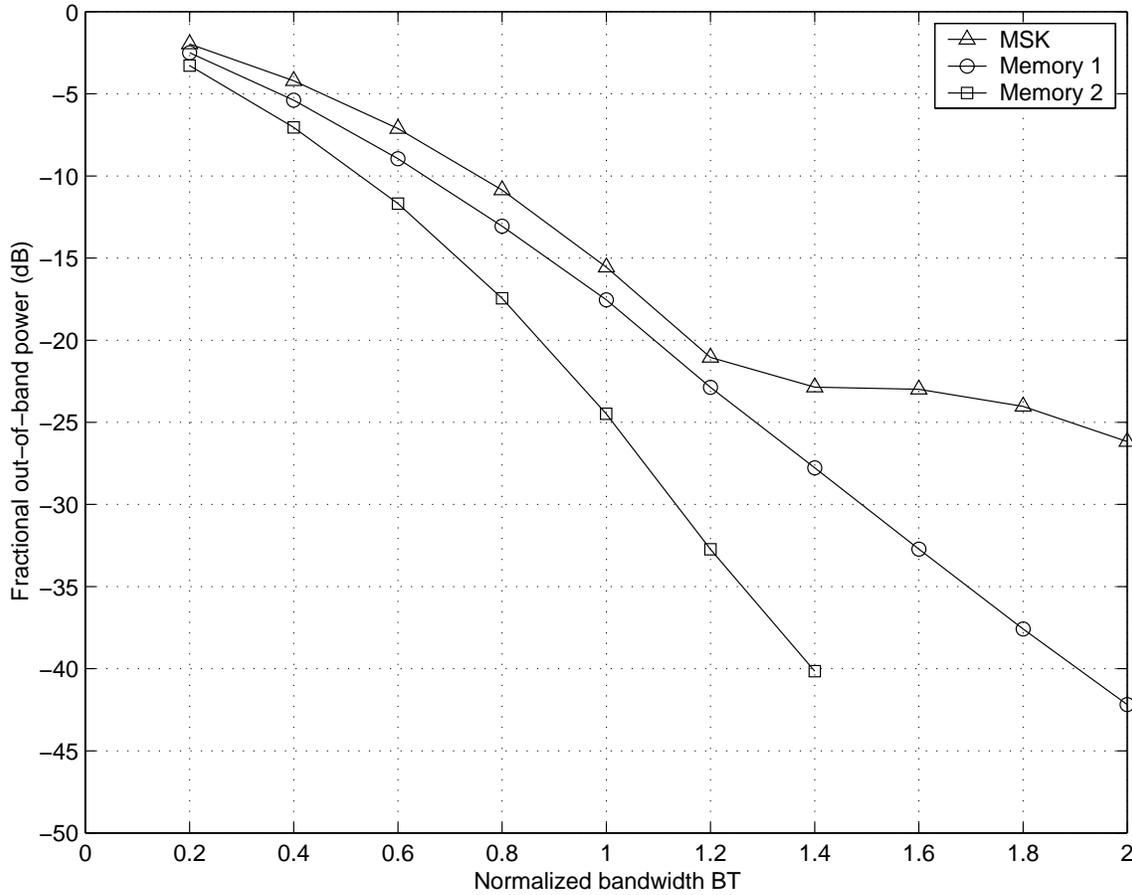

Figure 8: Fractional Out-of-band Power for Optimal Signal Sets for $\nu = 1$ and $\nu = 2$.

The bandwidth efficiency is commonly defined to be R/B in bits/s/Hz, where R is the data rate in bits/s and B is the bandwidth in Hz [8]. Since a signal of duration T is transmitted for every information bit, the bandwidth efficiency of the TCH scheme under consideration is given by

$$R/B = \frac{1}{BT} \text{ [bits/s/Hz]}. \tag{50}$$

The minimum values of $B_{99}T$ and $B_{99.9}T$ attainable by systems with $\nu = 1$ and 2 may be found in Figure 8.[3] The maximum bandwidth efficiency in terms of the 99% and 99.9% bandwidths are given in Table 2. We observe that a significant improvement in bandwidth efficiency over MSK is obtainable with the optimized signal set without loss in d_{min}^2 and for a prescribed decoding delay. It is expected that further improvement in bandwidth efficiency is achievable for systems with larger memory.

Table 2: Bandwidth Efficiency in terms of Normalized 99% and 99.9% Power Bandwidth.

	Optimum Signals		MSK
	$\nu = 2$	$\nu = 1$	
$1/B_{99}T$ [bits/s/Hz]	1.15	0.91	0.84
% gain over MSK	36.9	8.3	
$1/B_{99.9}T$ [bits/s/Hz]	0.88	0.67	0.36
% gain over MSK	144.4	86.1	

VII   Conclusion

We have considered a class of trellis-coded modulation schemes that can be described as shown in Figure 1. For any such scheme, a set of necessary and sufficient conditions have been found so that the Viterbi decoder can decode an information symbol upon observation of the received signal in $\nu + 1$ consecutive symbol intervals, where ν is the memory of the modulation. It has also been found that when these conditions hold, the minimum squared Euclidean distance is upper bounded by that of antipodal modulation or, equivalently, of MSK. Necessary and sufficient conditions for achieving the upper bound have been found and it has been shown that these conditions are satisfied if and only if the modulation scheme is implementable as pulse amplitude modulation. Hence, if we require that the maximum likelihood sequence estimator of a ν-memory constant-energy modulation scheme be able to decode with delay νT, then the memory may be used to shape the spectrum of the transmitted signal but not to increase the minimum squared Euclidean distance beyond that of antipodal modulation. Signal sets that maximize the minimum squared Euclidean distance and minimize the transmitted signal's bandwidth have been found numerically for $\nu = 1$ and 2. Their bandwidth efficiency exceeds that of MSK. It is expected that further improvements are attainable with larger ν.

---

[3]It should be pointed out that the lines connecting points in Figure 8 have been drawn for convenience only and do not guarantee the existence of signal sets with the corresponding characteristics.
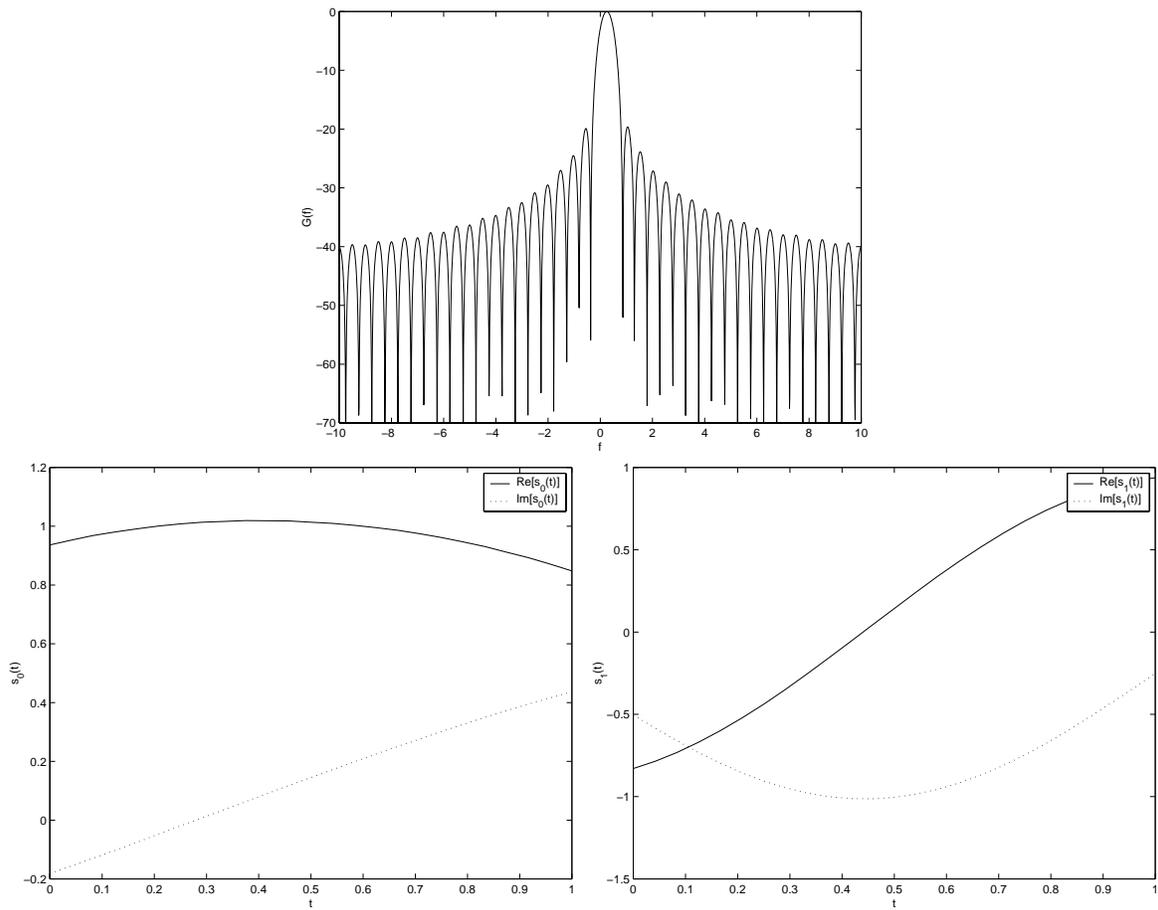
Figure 9: Optimization results for $\nu = 1$ and $BT = 0.8$: PSD (top); Optimal signal $s_0(\tau)$ (bottom left); Optimal signal $s_1(\tau)$ (bottom right).
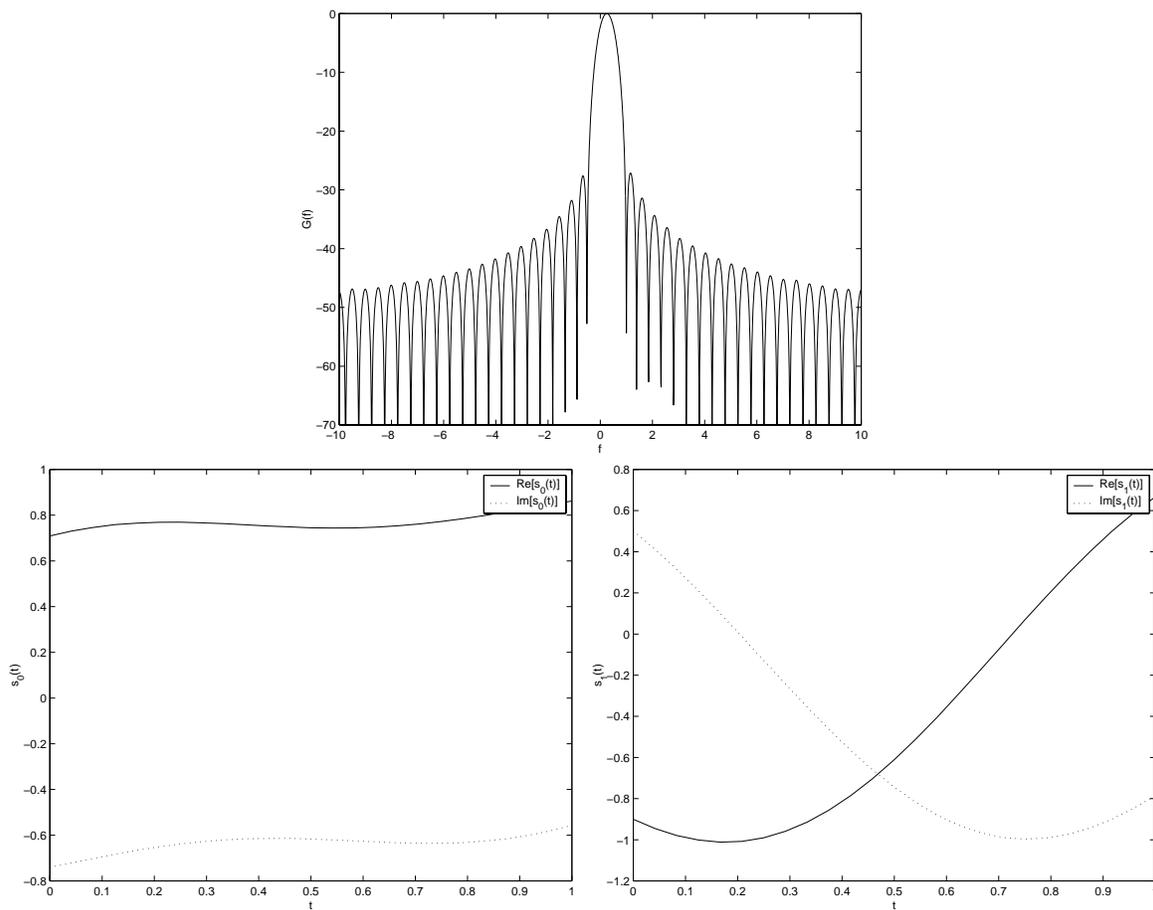
Figure 10: Optimization results for $\nu = 1$ and $BT = 1.2$: PSD (top); Optimal signal $s_0(\tau)$ (bottom left); Optimal signal $s_1(\tau)$ (bottom right).
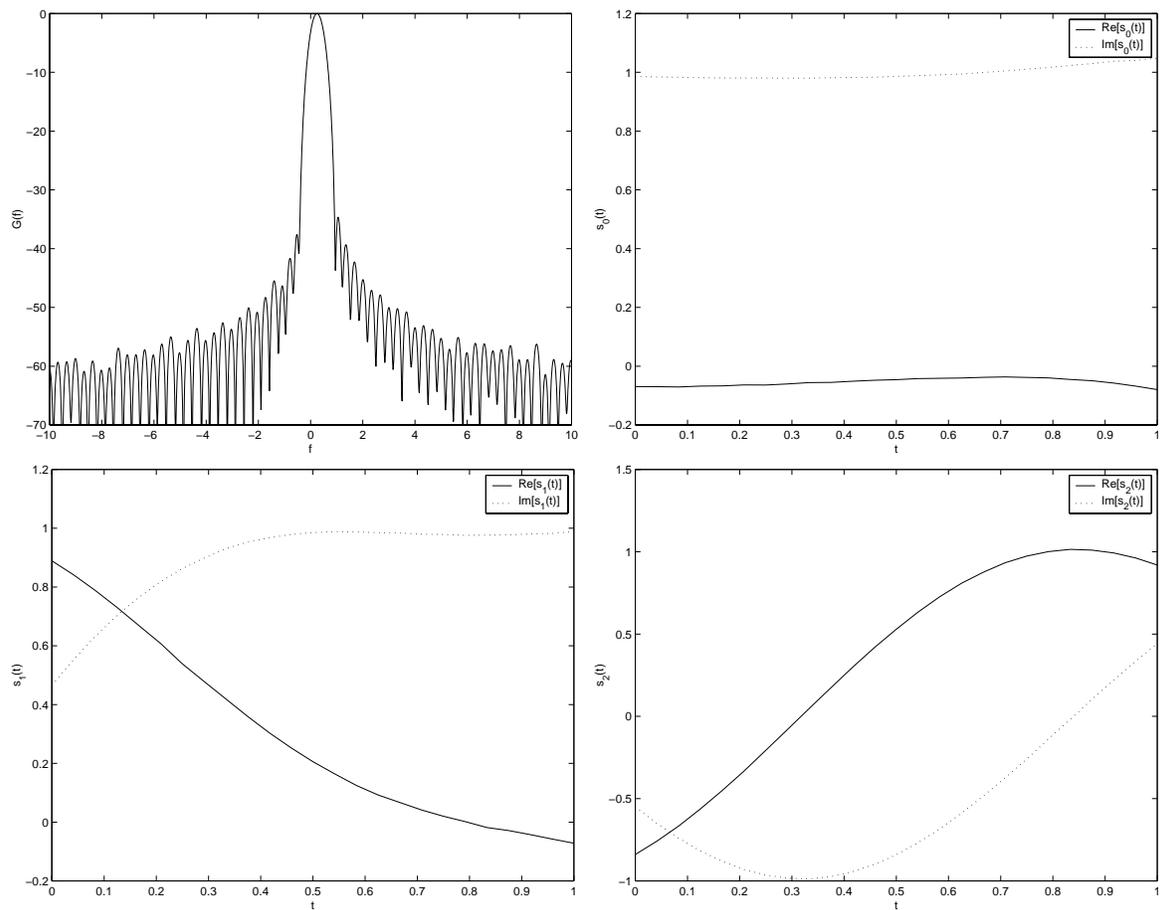
Figure 11: Optimization results for $\nu = 2$ and $BT = 1.2$: PSD (top left); Optimal signal $s_0(\tau)$ (top right); Optimal signal $s_1(\tau)$ (bottom left); Optimal signal $s_2(\tau)$ (bottom right).

A   Appendix

The following lemma is used in the proof of Theorem 1.

Lemma 8 *The following three conditions are equivalent:*
(i) $s_0(\tau) - s_{2^m}(\tau) = s_{a+b}(\tau) - s_{a+b+2^m}(\tau)$,

$\quad m \in \{0, 1, \cdots, \nu\}, \qquad a \in \{0, 1, \cdots, 2^m - 1\}, \qquad b \in \{2^{m+1}l : l = 0, 1, \cdots, 2^{\nu-m} - 1\}.$

(ii) $s_0(\tau) - s_{a+2^m}(\tau) = s_{a+b}(\tau) - s_{a+b+2^m}(\tau)$,

$m \in \{0, 1, \cdots, \nu - 1\}, \qquad a \in \{0, 1, \cdots, 2^m - 1\}, \qquad b \in \{2^{m+1}l : l = 0, 1, \cdots, 2^{\nu-m} - 1\}.$

(iii) $s_0(\tau) - s_{2^m}(\tau) = s_b(\tau) - s_{b+2^m}(\tau)$,

$\quad\quad\quad m \in \{0, 1, \cdots, \nu - 1\}, \qquad b \in \{2^{m+1}l : l = 0, 1, \cdots, 2^{\nu-m} - 1\}.$

Proof (i) $\Rightarrow$ (ii): Assume that (i) is true. The left side of (i) does not depend on b. Hence also the right side must be independent of b. In particular, $s_{0+a}(\tau) - s_{0+a+2^m}(\tau) = s_{a+b}(\tau) - s_{a+b+2^m}(\tau)$, proving that (ii) is also true.

(ii) $\Rightarrow$ (iii): This is true since (iii) is (ii) with $a = 0$.

(iii) $\Rightarrow$ (i): We prove this by induction on m. For $m = 0$, (iii) and (ii) are identical since $m = 0$ implies $a = 0$. Assume that (iii) holds and that (i) holds for $m = 1, 2, \cdots, M - 1 \le \nu - 1$, and we show that it also holds for $m = M$, i.e.,

$$s_0(\tau) - s_{2^M}(\tau) = s_{a+b}(\tau) - s_{a+b+2^m}(\tau), \tag{51}$$

$$a \in \{0, 1, \cdots, 2^M - 1\}, \qquad b \in \{2^{M+1}l : l = 0, 1, \cdots, 2^{\nu-M} - 1\}.$$

From induction hypothesis, we have

$$s_{a+b}(\tau) - s_{a+b+2^m}(\tau) = s_{a+b'}(\tau) - s_{a+b'+2^m}(\tau),$$

$m \in \{0, 1, \cdots, M - 1\}, \qquad a \in \{0, 1, \cdots, 2^m - 1\}, \qquad b, b' \in \{2^{m+1}l : l = 0, 1, \cdots, 2^{\nu-m} - 1\}.$

Let $b \in \{2^{M+1}l : l = 0, 1, \cdots, 2^{\nu-M} - 1\} \subset \{2^{m+1}l : l = 0, 1, \cdots, 2^{\nu-m} - 1\}$, and $b' = b + 2^M \in \{2^{m+1}l : l = 0, 1, \cdots, 2^{\nu-m} - 1\}$. The above equality becomes

$$s_{a+b}(\tau) - s_{a+b+2^m}(\tau) = s_{a+b+2^M}(\tau) - s_{a+b+2^m+2^M}(\tau).$$

Rearranging terms, we get

$$s_{a+b}(\tau) - s_{a+b+2^M}(\tau) = s_{a+b+2^m}(\tau) - s_{a+b+2^m+2^M}(\tau). \tag{52}$$

$$m \in \{0, 1, \cdots, M-1\}, \qquad a \in \{0, 1, \cdots, 2^m - 1\}, \qquad b \in \{2^{M+1}l : l = 0, 1, \cdots, 2^{\nu-M} - 1\}.$$

Now let $a \in \{0, 1, \cdots, 2^M - 1\}$, $b \in \{0, 1, \cdots, 2^M - 1\}$, and write $a = \sum_{n=0}^{M-1} a_n 2^n$, $a_n \in \{0, 1\}$, $n = 1, 2, \ldots, M-1$. We use (52) M times and obtain

$$
\begin{aligned}
s_b(\tau) - s_{b+2^M}(\tau) &= s_{a_0 2^0 + b}(\tau) - s_{a_0 2^0 + b + 2^M}(\tau) \\
&= s_{a_0 2^0 + a_1 2^1 + b}(\tau) - s_{a_0 2^0 + a_1 2^1 + b + 2^M}(\tau) \\
&\ \ \vdots \\
&= s_{a_0 2^0 + a_1 2^1 + \cdots + a_{M-1} 2^{M-1} + b}(\tau) - s_{a_0 2^0 + a_1 2^1 + \cdots + a_{M-1} 2^{M-1} + b + 2^M}(\tau) \\
&= s_{a+b}(\tau) - s_{a+b+2^M}(\tau).
\end{aligned}
$$

This and (iii) with $m = M$ imply (51).   □

References

[1] B. Rimoldi, "A decomposition approach to CPM," *IEEE Trans. Inform. Theory*, vol. IT–34, pp. 260–270, Mar. 1988.

[2] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[3] J. L. Massey, "A generalized formulation of minimum shift keying modulation," in *Proceedings IEEE Int. Commun. Conf.*, (Seattle, WA), pp. 26.5.1–26.5.5, June 1980.

[4] B. Rimoldi, "Five views of differential MSK: A unified approach," in *Communications and cryptography: two sides of one tapestry* (R. E. Blahut, D. J. Costello Jr., U. Maurer, and T. Mittelholzer, eds.), pp. 333–342, Kluwer Academic Publishers, 1994.

[5] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. New York: Macmillan, 1991.

[6] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill, 1979.

[7] M. K. Simon, S. Darden, and M. Fong, "Bandwidth efficient trellis coded modulation with prescribed decoding delay - new interpretations and results," in *ICC Conf. Rec.*, 2001. To be presented.

[8] B. Sklar, *Digital Communications Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice Hall, 1988.