

Coherent and Multi-symbol Noncoherent CPFSK: Capacity and Code Design

Shi Cheng and Matthew C. Valenti
West Virginia University
Morgantown, WV
{shic,mvalenti}@csee.wvu.edu

Don Torrieri
U.S. Army Research Laboratory
Adelphi, MD
dtorrieri@arl.army.mil

ABSTRACT

The capacity of coded continuous-phase frequency-shift keying (CPFSK) is found for additive white Gaussian noise (AWGN) channels under the assumption that the symbols at the modulator input are independent and uniformly distributed. Two forms of reception are considered, coherent detection and multi-symbol noncoherent block detection. Calculating the coherent capacity of CPFSK is facilitated by considering the system as a finite-state Markov channel. A methodology is proposed for designing systems that approach the capacity by using an irregular repeat-accumulate (IRA) code. The code is optimized directly from the system's EXIT chart by using linear programming to determine the optimal variable-node degree distribution. Results are presented for a rate 1/2 MSK system that is within 0.43 dB and 0.33 dB of the coherent and 4-symbol noncoherent capacities, respectively.

INTRODUCTION

Continuous-phase modulation (CPM) is a general class of modulation that achieves high bandwidth efficiency by requiring a smooth phase transition between adjacent symbols. CPM is said to be *full-response* if the symbols at the input to the frequency modulator are represented by pulses that are entirely contained within one symbol interval. The simplest form of full-response CPM is *continuous-phase frequency-shift keying* (CPFSK), which has a rectangular pulse shaping function that spans the entire symbol.

Coherent detection of CPFSK is discussed in [1]. Unlike memoryless FSK, the phase of CPFSK is accumulated from symbol to symbol to maintain a smooth phase transition. When the modulation index h is a rational number, the accumulated phases take values from a finite set $\Phi \subset [0, 2\pi)$. In such a case, the phase trajectory can be viewed as a finite-state Markov random process, so that the modulator and additive white Gaussian noise (AWGN) channel can together be considered as a finite-state Markov channel (FSMC). This allows coherent detection to be performed on a trellis.

Finding the capacity of a FSMC is difficult, since it requires a maximization over the probability density function (pdf) of a long input sequence. Fortunately in practice, the input to the FSMC is usually preceded by an outer channel encoder, which typically produces uniformly distributed outputs. Arnold et

al. [2] use the forward recursion of the BCJR algorithm [3] to compute the *symmetric information rate* of the FSMC, which is the mutual information when the inputs are independent and uniformly distributed (i.u.d.). In this paper, we apply a similar approach to compute the symmetric information rate of coherently detected CPFSK, which is to our knowledge a new application of the techniques in [2]. For the remainder of the paper, we assume i.u.d. inputs and therefore use the term *capacity* to specifically mean the symmetric information rate.

A key advantage of using CPFSK is that it can be noncoherently detected. The capacity of symbol-by-symbol noncoherent detection of CPFSK is found in [4]. CPFSK can also be detected using a multi-symbol block noncoherent detector, as proposed by Simon and Divsalar in [5]. While prior work on multi-symbol block noncoherent detection of CPFSK has focused on its bit error rate (BER) analysis, we evaluate its *capacity* and show that the capacity approaches that of coherent detection as the block size increases. This result is analogous to the BER analysis of [5], which shows that the BER of multi-symbol block noncoherent detection approaches that of coherent detection for large block sizes. This result is also consistent with the asymptotic capacity analysis of generic noncoherent channels in [6] and [7].

Despite having a capacity that is lower than coherent detection, multi-symbol block noncoherent detection has significant complexity benefits. For coherent detection to be feasible, h must be a rational number, i.e. $h = P/Q$, where P and Q are relatively prime positive integers. When Q is large, the complexity of the coherent detector can be very high. However, for the noncoherent detector, complexity is independent of h , and thus h can be any real number. This allows a more flexible design, since values of h that might be convenient for coherent detection do not necessarily achieve capacity. Another benefit of the noncoherent detector is that it does not need to know the initial phase or even the set Φ to which it belongs. Furthermore, the coherent receiver requires Φ to be time invariant, while it may in fact drift due to offsets in the oscillators.

Having established the capacity of both coherent and multi-symbol block noncoherent detection of CPFSK, we turn our attention to the design of systems that are capable of approaching the corresponding capacity. A binary irregular repeat-accumulate (IRA) code is used along with iterative

demodulation and decoding. The IRA code is designed directly from the system's EXIT chart using a curve-fitting technique proposed by ten Brink et al. [8] and Roumy et al. [9] and previously applied to the noncoherent detection of orthogonal FSK by Guillén i Fàbregas in [10]. The combination of IRA codes and CPM has previously been considered in [11] for coherent detection, whereas in this paper we consider both coherent and noncoherent detection.

SYSTEM MODEL

In the following discussion, bold lowercase letters will be used to denote (column) vectors, e.g. \mathbf{x} , and bold uppercase letters will be used for matrices, e.g. \mathbf{X} . The scalar value $x_{i,j}$ is used to denote the $(i,j)^{th}$ entry of the matrix \mathbf{X} , while the scalar value x_i is used to denote the i^{th} element of the vector \mathbf{x} . All matrices and vectors are indexed starting at zero, $\mathbf{x} = [x_0, x_1, \dots, x_{M-1}]^T$. Matrices may be represented as a row of column vectors, e.g. $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{M-1}]$. The notation \mathbf{x}_i^j is used to represent the set $\{\mathbf{x}_i, \mathbf{x}_{i+1}, \dots, \mathbf{x}_j\}$.

Suppose the input sequence to a CPFSK modulator is \mathbf{q} , whose elements are i.i.d. over the integers from 0 to $M-1$. For every entry of \mathbf{q} , the modulated signal $x_i(t)$ is chosen as the q_i^{th} signal of the set $\mathcal{S} = \{s_k(t), k = 0, 1, \dots, M-1\}$, where

$$s_k(t) = \frac{1}{\sqrt{T_s}} \exp \left\{ \frac{j2\pi k h t}{T_s} \right\}, \quad t \in [0, T_s), \quad (1)$$

and h is the *modulation index*. In order to satisfy the continuous-phase constraint, the phase of each modulated symbol is accumulated as

$$\phi_{i+1} \triangleq \phi_i + 2q_i h \pi, \quad (2)$$

where ϕ_i is the accumulated phase at the start of the i^{th} symbol [12]. The continuous-time waveform $\sqrt{\mathcal{E}_s} e^{j\phi_i} x_i(t)$ is transmitted, and the corresponding received signal is

$$y_i(t) = \sqrt{\mathcal{E}_s} e^{j\phi_i} x_i(t) + n_i(t), \quad (3)$$

where $n_i(t)$ is a circularly symmetric complex AWGN process with noise spectral density N_0 , and \mathcal{E}_s is the energy per symbol.

Given the initial phase ϕ_i at the start of the i^{th} interval, the front-end of the coherent receiver determines the log-likelihood of receiving waveform $y_i(t)$, for each of the M possible transmitted waveforms. Since this process is the same for every received symbol, we drop the index i for the remainder of this section. The received signal $y(t), 0 \leq t \leq T_s$, is first passed through a bank of M pairs of matched filters, with one pair matched to the in-phase and quadrature components of each tone, and then sampled at the symbol epoch. The sampled signal can be written in vector form as

$$\mathbf{y} = e^{j\phi} \sqrt{\mathcal{E}_s} \mathbf{x} + \mathbf{n}, \quad (4)$$

where the elements of \mathbf{y} , \mathbf{x} and \mathbf{n} are

$$y_k = e^{j\phi} \sqrt{\mathcal{E}_s} x_k + n_k \quad (5)$$

$$x_k = \int_0^{T_s} x(t) s_k^*(t) dt \quad (6)$$

$$n_k = \int_0^{T_s} n(t) s_k^*(t) dt, \quad (7)$$

and $k = \{0, 1, \dots, M-1\}$. The noise vector \mathbf{n} is Gaussian with a covariance matrix $\mathbf{R} = \mathbf{E}(\mathbf{n}\mathbf{n}^H)$ with $(k, i)^{th}$ element

$$\begin{aligned} r_{k,i} &= N_0 \int_0^{T_s} s_k^*(t) s_i(t) dt \\ &= N_0 \rho_{k,i}, \end{aligned} \quad (8)$$

where

$$\rho_{k,i} = \frac{\sin(\pi(i-k)h)}{\pi(i-k)h} e^{j\pi(i-k)h}. \quad (9)$$

When conditioned on both \mathbf{x} and ϕ , the vector \mathbf{y} is Gaussian with mean \mathbf{x} and covariance \mathbf{R} , and has conditional pdf

$$p(\mathbf{y}|\mathbf{x}, \phi) = \frac{1}{\pi^M \det(\mathbf{R})} e^{-(\mathbf{y} - e^{j\phi} \sqrt{\mathcal{E}_s} \mathbf{x})^H \mathbf{R}^{-1} (\mathbf{y} - e^{j\phi} \sqrt{\mathcal{E}_s} \mathbf{x})}. \quad (10)$$

The exponent can be simplified as

$$\begin{aligned} &-(\mathbf{y} - e^{j\phi} \sqrt{\mathcal{E}_s} \mathbf{x})^H \mathbf{R}^{-1} (\mathbf{y} - e^{j\phi} \sqrt{\mathcal{E}_s} \mathbf{x}) \\ &= -\mathbf{y}^H \mathbf{R}^{-1} \mathbf{y} - \mathcal{E}_s \mathbf{x}^H \mathbf{R}^{-1} \mathbf{x} + 2\text{Re}(e^{-j\phi} \sqrt{\mathcal{E}_s} \mathbf{x}^H \mathbf{R}^{-1} \mathbf{y}). \end{aligned} \quad (11)$$

Define $\mathbf{K} \triangleq \frac{1}{N_0} \mathbf{R}$, i.e. a normalized version of \mathbf{R} . Note that when $x(t) = s_\nu(t)$, \mathbf{x} is the ν^{th} column of \mathbf{K} , i.e. $\mathbf{x} = \mathbf{k}_\nu$. Therefore, when $x(t) = s_\nu(t)$, the exponent becomes

$$-\frac{\mathbf{y}^H \mathbf{K}^{-1} \mathbf{y} + \mathcal{E}_s}{N_0} + 2 \frac{\sqrt{\mathcal{E}_s}}{N_0} \text{Re}(e^{-j\phi} y_\nu). \quad (12)$$

Taking the log and discarding factors that are common to all hypotheses, the log-likelihood for coherent reception can be expressed as

$$\log p(\mathbf{y}|\mathbf{x} = \mathbf{k}_\nu, \phi) \propto 2 \frac{\sqrt{\mathcal{E}_s}}{N_0} \text{Re}(e^{-j\phi} y_\nu). \quad (13)$$

For noncoherent reception, the value of ϕ is unknown and therefore must be marginalized out of the above expression [4].

COHERENT CAPACITY

Trellis-based detection of CPFSK requires that the modulation index h be a rational number so that the accumulated phase ϕ takes on values from a finite set. Suppose $h = P/Q$, where P and Q are relatively prime positive integers. The total number of unambiguous values that ϕ can assume is Q [12]. Thus, demodulation can be performed over a trellis with Q states and QM branches per trellis section.

The capacity of coherently detected CPFSK is found by first evaluating the average mutual information $I(\mathbf{x}_0^{N-1}, \mathbf{y}_0^{N-1})$ between \mathbf{x}_0^{N-1} and \mathbf{y}_0^{N-1} , and then taking the average as the sequence length N goes to infinity,

$$C^{(c)} = \lim_{N \rightarrow \infty} \frac{1}{N} I(\mathbf{x}_0^{N-1}, \mathbf{y}_0^{N-1}). \quad (14)$$

From the chain rule of entropy,

$$\begin{aligned} I(\mathbf{x}_0^{N-1}, \mathbf{y}_0^{N-1}) &= H(\mathbf{x}_0^{N-1}) - H(\mathbf{x}_0^{N-1} | \mathbf{y}_0^{N-1}) \\ &= \sum_{i=0}^{N-1} H(\mathbf{x}_i) - \sum_{i=0}^{N-1} H(\mathbf{x}_i | \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1}), \end{aligned} \quad (15)$$

where $H(\mathbf{x}_i | \mathbf{x}_0^{i-1}) = H(\mathbf{x}_i)$ is used. Because the input \mathbf{x}_i is i.u.d. over M constellation points, $H(\mathbf{x}_i) = \log_2 M$ bits, and all that remains to be calculated is $H(\mathbf{x}_i | \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1})$. From the definition of conditional entropy,

$$H(\mathbf{x}_i | \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1}) = -E [\log_2 p(\mathbf{x}_i | \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1})]. \quad (16)$$

The above expectation can be found using Monte Carlo integration.

To compute the probability $p(\mathbf{x}_i | \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1})$, first apply Bayes' rule to obtain

$$p(\mathbf{x}_i | \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1}) = \frac{p(\mathbf{x}_i, \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1})}{\sum_{\mathbf{x}_i} p(\mathbf{x}_i, \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1})}. \quad (17)$$

Similar to [2], a BCJR-like method can be used to compute $p(\mathbf{x}_i, \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1})$. Rather than explicitly calculating the denominator in (17), its value is found such that $\sum_{\mathbf{x}_i} p(\mathbf{x}_i | \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1}) = 1$.

Assume ϕ takes on values from the set Φ , whose cardinality is Q . Similar to the BCJR algorithm, we define α , β , γ as

$$\alpha_i(\phi_i) \triangleq p(\phi_i, \mathbf{y}_0^{i-1}, \mathbf{x}_0^{i-1}) \quad (18)$$

$$\beta_{i+1}(\phi_{i+1}) \triangleq p(\mathbf{y}_{i+1}^{N-1} | \phi_{i+1}) \quad (19)$$

$$\gamma(\phi_i \rightarrow \phi_{i+1}, \mathbf{y}_i, \mathbf{x}_i) \triangleq p(\mathbf{y}_i, \phi_{i+1} | \phi_i, \mathbf{x}_i). \quad (20)$$

Note that $\gamma(\phi_i \rightarrow \phi_{i+1}, \mathbf{y}_i, \mathbf{x}_i)$ is nonzero only when \mathbf{x}_i causes the state transition from ϕ_i to ϕ_{i+1} . Therefore, it may be written as

$$\begin{aligned} \gamma(\phi_i \rightarrow \phi_{i+1}, \mathbf{y}_i, \mathbf{x}_i = \mathbf{k}_\nu) &= p(\phi_{i+1} | \phi_i, \mathbf{x}_i = \mathbf{k}_\nu) p(\mathbf{y}_i | \phi_{i+1}, \phi_i, \mathbf{x}_i = \mathbf{k}_\nu) \\ &= \begin{cases} p(\mathbf{y}_i | \phi_i, \mathbf{x}_i = \mathbf{k}_\nu) & \phi_{i+1} = \phi_i + 2\nu h\pi \\ 0 & \phi_{i+1} \neq \phi_i + 2\nu h\pi. \end{cases} \end{aligned} \quad (21)$$

As with the BCJR algorithm, α can be calculated in a forward recursion as

$$\alpha_{i+1}(\phi_{i+1}) \propto \sum_{\phi_i \in \Phi} \alpha_i(\phi_i) \gamma(\phi_i \rightarrow \phi_{i+1}, \mathbf{y}_i, \mathbf{x}_i = \mathbf{k}_{q_i}), \quad (22)$$

where the constant of proportionality is chosen to make α a valid probability. Similarly, β can be calculated in a backward recursion as

$$\beta_i(\phi_i) \propto \sum_{\substack{\mathbf{x}_i, \\ \phi_{i+1} \in \Phi}} \beta_{i+1}(\phi_{i+1}) \gamma(\phi_i \rightarrow \phi_{i+1}, \mathbf{y}_i, \mathbf{x}_i). \quad (23)$$

Note that \mathbf{x}_i is marginalized out of the summand since $\beta_i(\phi_i)$ does not depend on it.

In the absence of knowing the starting and ending states, both α_0 and β_N can be initialized assuming equally likely states, i.e. $\alpha_0(\phi) = \beta_N(\phi) = 1/M$, $\forall \phi \in \Phi$. Alternatively, if the initial phase ϕ_0 is known to the detector, α_0 can be set to all zeros except a one at the corresponding entry. Obviously, the effect of the initial states of α_0 and β_N diminish as N approaches infinity.

Given the above definitions, $p(\mathbf{x}_i, \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1})$ is found from

$$\begin{aligned} &p(\mathbf{x}_i, \mathbf{x}_0^{i-1}, \mathbf{y}_0^{N-1}) \\ &\propto \sum_{\phi_i, \phi_{i+1} \in \Phi} \alpha_i(\phi_i) \beta_{i+1}(\phi_{i+1}) \gamma(\phi_i \rightarrow \phi_{i+1}, \mathbf{y}_i, \mathbf{x}_i). \end{aligned} \quad (24)$$

Fig. 1 compares the capacities of four different CPFSK designs: (1) $M = 2$, $h = 1/2$, i.e. minimum-shift keying (MSK), (2) $M = 4$, $h = 1/2$, (3) $M = 8$, $h = 3/8$, and (4) $M = 16$, $h = 1/4$. These capacities were calculated by Monte Carlo simulation with at least one million simulated symbols per SNR point. The figure displays the minimum \mathcal{E}_b/N_0 required as a function of binary code rate r , where $\mathcal{E}_b = \mathcal{E}_s/r \log_2 M$. All four systems have roughly the same uncoded bandwidth efficiency at 1 bps/Hz, found by integrating the power spectral densities given in Chapter 4 of [13]. For this bandwidth efficiency, $M = 8$ offers the best energy efficiency over the entire range of code rates. However, the gap between the energy efficiencies closes as the code rate decreases. At code rate $r = 0.5$, the required \mathcal{E}_b/N_0 for MSK is 0.2 dB while for $M = 8, h = 3/8$, it is only 0.85 dB lower. This marginal gain in energy efficiency must be weighed against the additional complexity of using $M = 8, h = 3/8$, which has 16/3 times more edges in the trellis per bit than MSK.

NONCOHERENT CAPACITY

Consider an L -symbol block noncoherent detector [5]. Let the block of modulated symbols be \mathbf{x}_0^{L-1} and the block of received symbols be \mathbf{y}_0^{L-1} . The block noncoherent detector computes the probability $p(\mathbf{y}_0^{L-1} | \mathbf{x}_0^{L-1})$ for each of the M^L possible \mathbf{x}_0^{L-1} . If the initial phase ϕ_0 at the start of the block is given, the conditional probability can be represented by the chain rule as

$$\begin{aligned} p(\mathbf{y}_0^{L-1} | \mathbf{x}_0^{L-1}, \phi_0) &= \prod_{i=0}^{L-1} p(\mathbf{y}_i | \mathbf{y}_0^{i-1}, \mathbf{x}_0^{L-1}, \phi_0) \\ &= \prod_{i=0}^{L-1} p(\mathbf{y}_i | \mathbf{x}_i, \phi_i), \end{aligned} \quad (25)$$

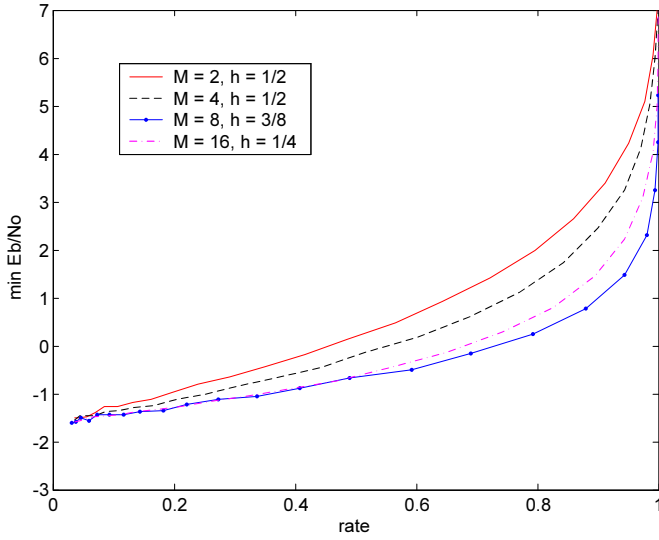


Fig. 1. Coherent capacity of CPFSK.

where the second equality comes directly from the properties of Markov chains, and ϕ_i is recursively updated using (2).

From (25) and (13), if the input to the modulator is $\mathbf{q} = [q_0, \dots, q_{L-1}]$, then the conditional pdf is

$$p(\mathbf{y}_0^{L-1} | \mathbf{q}, \phi_0) \propto \exp \left(2 \frac{\sqrt{\mathcal{E}_s}}{N_0} \text{Re} \{ e^{-j\phi_0} \mu(\mathbf{q}) \} \right) \quad (26)$$

where

$$\mu(\mathbf{q}) = \sum_{i=0}^{L-1} y_{q_i} e^{-2h\pi \sum_{k=0}^{i-1} q_k}. \quad (27)$$

The noncoherent detector assumes ϕ_0 has a uniform distribution over $[0, 2\pi)$. Marginalizing $p(\mathbf{y}_0^{L-1} | \mathbf{q}, \phi_0)$ with respect to ϕ_0 yields

$$p(\mathbf{y}_0^{L-1} | \mathbf{q}) \propto I_0 \left(2 \frac{\sqrt{\mathcal{E}_s}}{N_0} |\mu(\mathbf{q})| \right), \quad (28)$$

where $I_0(\cdot)$ is the 0th order modified Bessel function of the first kind.

The capacity can then be calculated from

$$C_L^{(n)} = \log_2 M + \frac{1}{L} E \left[\log_2 \frac{I_0 \left(\frac{2\sqrt{\mathcal{E}_s} |\mu(\mathbf{q})|}{N_0} \right)}{\sum_{\mathbf{q}' \in \mathcal{Q}} I_0 \left(\frac{2\sqrt{\mathcal{E}_s} |\mu(\mathbf{q}')|}{N_0} \right)} \right], \quad (29)$$

where \mathcal{Q} is the set of M^L possible values of \mathbf{q} and the expectation is taken over the ensemble of all possible transmitted \mathbf{q} and received \mathbf{y}_0^{L-1} . As in the coherent case, the above expectation can be found using Monte Carlo integration.

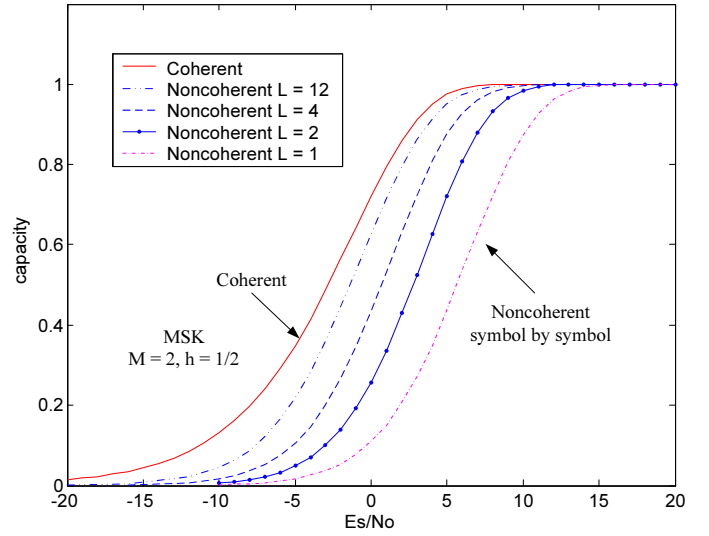


Fig. 2. Capacity of MSK using multi-symbol noncoherent and coherent detection.

As an example, Fig. 2 shows the capacity of multi-symbol noncoherent detection of MSK for several different block sizes. The rightmost curve ($L = 1$) is the capacity of symbol-by-symbol noncoherent detection previously reported in [4], while the leftmost curve is the coherent capacity. By increasing L from 1 to 4, the gain at code rate 0.5 is about 5 dB, and it is only 3.5 dB worse than the coherent detection. When the block size L is larger, this capacity of noncoherent detection gets closer to that of coherent detection. When L increases to infinity, we conjecture that the noncoherent capacity converges to the coherent capacity, which would be consistent with the asymptotic capacity analysis [6], [7] and the BER performance in [5]. More generally, the capacity of multi-symbol noncoherent detection can be found for any arbitrary value of h , M , and L using the same methodology used to generate the MSK curves shown in Fig. 2.

CODE DESIGN

EXIT charts are often used to analyze the convergence behavior of iterative decoding systems. In [8] a curve-fitting technique was applied that allows EXIT charts to be directly used as a code design methodology. This technique was later applied to the design of IRA codes in [9] and systems using orthogonal FSK with symbol-by-symbol noncoherent detection in [10]. Here, we apply the EXIT curve-fitting technique to design nonsystematic IRA codes for CPFSK with coherent and multi-symbol noncoherent detection.

The structure of the coded system is shown in Fig. 3. The system is a serial concatenation of two codes separated by an interleaver Π . The outer code is a mixture of repetition codes represented by variable nodes “ $=$ ”. The *degree* of a variable node is the number of times that the corresponding message bit is repeated. Since the code is irregular, the variable nodes

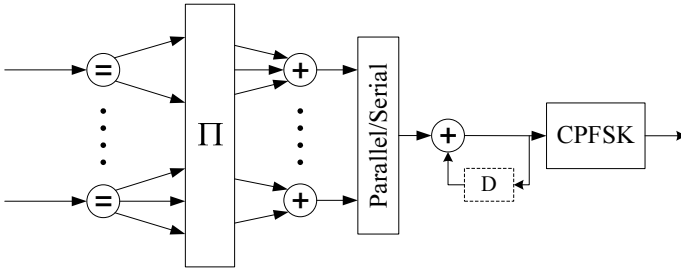


Fig. 3. Nonsystematic IRA coding structure. “=” corresponds to variable nodes and “+” corresponds to single parity-check nodes. The optional accumulator is not used when reception is coherent.

do not all require the same degree. The entire set of repeated bits is interleaved and sent to the check nodes, represented by “+”. Each check node forms a single parity-check (SPC) on a distinct subset of interleaved bits. The degree of a check node is the number of bits used to form the check. The outputs of the SPC nodes are grouped together and passed to the modulator. As in [10], the code is nonsystematic, and therefore unlike [9], the message bits are not modulated.

The inner code of an IRA code must be recursive [14]. When the modulation is memoryless, recursive coding is achieved by passing the parity bits through an accumulator prior to modulation [9]. However, as evident by (2), CPFSK with noninteger h is already recursive. A coherent system does not need an accumulator because the receiver tracks the memory in the modulation [11]. With a block noncoherent detector, the memory within each L block is exploited by the receiver, though the memory from block-to-block is lost. Preceding the CPFSK modulator with an accumulator helps to restore the memory from block-to-block, which may be beneficial for small values of L . Together, the combination of SPC nodes, accumulator (if present), and CPFSK modulator comprise the inner code.

At the receiver, decoders for each of the inner and outer codes exchange extrinsic information using a turbo-like schedule [9]. An EXIT chart is created for a particular SNR by drawing the *information transfer* functions for the inner and outer codes on the same plot. The information transfer function for an outer repetition code of degree d is [15]

$$I_{E,d}^{(o)} = J\left(\sqrt{d-1}J^{-1}(I_A^{(o)})\right), \quad (30)$$

where the superscript (o) denotes the outer code, and the subscripts A and E represent the *a priori* input and extrinsic information output. The function $J(\cdot)$ is defined in [15] as

$$J(\sigma) = \int \frac{1}{2\pi\sigma} e^{-\frac{(x-\sigma^2)^2}{2\sigma^2}} \log_2(1 + e^{-x}) dx, \quad (31)$$

and can be predetermined by numerical or Monte Carlo integration.

When an IRA code is used, the variable nodes do not all have the same degree. Let $\{\lambda_d\}, 3 \leq d \leq d_v$ (degree 2

variable nodes are not stable, and lead to high BER floor [15]), represent the edge-perspective degree distribution of the variable nodes, which is the fraction of edges connected with a degree- d variable node, and d_v represent the maximum variable-node degree. The overall information transfer function for the outer code can be approximated by using $\{\lambda_d\}$ to linearly combine the component information transfer functions according to [15]:

$$I_E^{(o)} = \sum_{d=3}^{d_v} \lambda_d I_{E,d}^{(o)}. \quad (32)$$

As with the outer code, the overall information transfer function $I_E^{(i)}$ of the inner code can be found by linearly combining the component information transfer functions $I_{E,d}^{(i)}$ for each degree d . Let $\{\rho_d\}, 1 \leq d \leq d_c$ represent the edge-perspective degree distribution of the check nodes and d_c represent the maximum check-node degree. The overall information transfer characteristic for the inner code can then be approximated by

$$I_E^{(i)} = \sum_{d=1}^{d_c} \rho_d I_{E,d}^{(i)}. \quad (33)$$

What remains is the calculation of the $I_{E,d}^{(i)}$ for each d . Unlike the outer code, the component information transfer functions $I_{E,d}^{(i)}$ cannot be easily expressed in integral form like (30)-(31), and therefore must be found via Monte Carlo simulation.

Once the information transfer functions for the inner and outer codes have been found, they are drawn on the same plot. The inner code's information transfer function is drawn with $I_A^{(i)}$ as its horizontal axis and $I_E^{(i)}$ as its vertical axis, while the outer code's information transfer function is drawn with $I_E^{(o)}$ as its horizontal axis and $I_A^{(o)}$ as its vertical axis. The plot showing both of these curves constitutes the system's EXIT chart. The code is said to *converge* if there is a gap between the two curves, and the *convergence threshold* is the minimum SNR for which the two curves just barely touch. The design objective is to minimize this threshold through the proper selection of the degree distributions.

Suppose that arbitrary degree distributions are selected for both the inner and outer code. By drawing the EXIT chart, the convergence threshold is found. The convergence threshold could be lowered by changing the degree distributions of either the inner or outer code in such a way that the two information transfer functions no longer touch. This process could be repeated until the curves can no longer be reshaped, at which point the two curves will be nearly identical and thus the area between the two curves will be arbitrarily small.

Following the above argument, a reasonable, albeit heuristic, way to design a system is to pick degree distributions that minimize the area between the two information transfer functions. To do this, we fix the degree distribution of the inner code, and find the degree distribution of the outer code that

TABLE I

OPTIMIZED CODES FOR MSK AT RATE $r = 0.5$. FOR EACH OF THE COHERENT AND MULTI-SYMBOL NONCOHERENT ($L = 4$) DETECTORS, THE DEGREE DISTRIBUTIONS, CAPACITY, AND THRESHOLDS ARE LISTED.

	Coherent	Noncoherent ($L = 4$)
Variable Node	$\lambda_3 = 0.570$ $\lambda_7 = 0.365$ $\lambda_8 = 0.065$	$\lambda_3 = 0.414$ $\lambda_5 = 0.439$ $\lambda_6 = 0.147$
Check Node	$\rho_1 = 0.001$ $\rho_2 = 0.999$	$\rho_1 = 0.001$ $\rho_2 = 0.999$
Capacity(\mathcal{E}_b/N_0)	0.2 dB	3.7 dB
Threshold(\mathcal{E}_b/N_0)	0.48 dB	3.94 dB

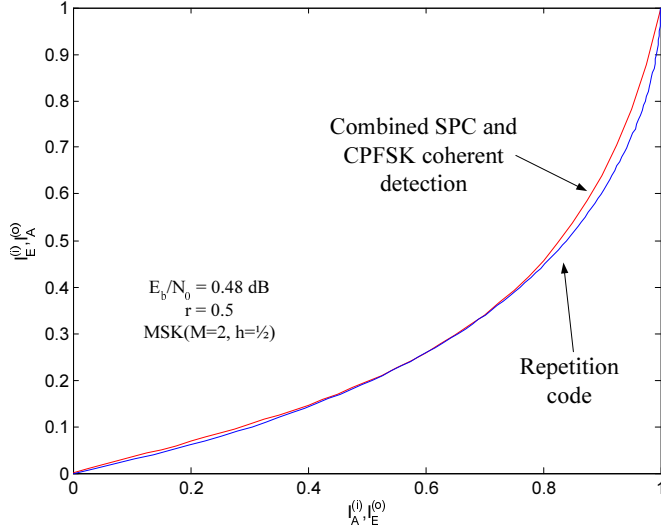


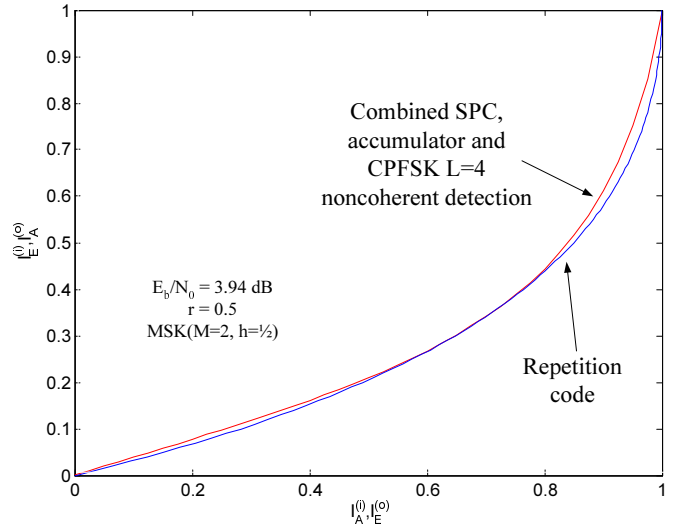
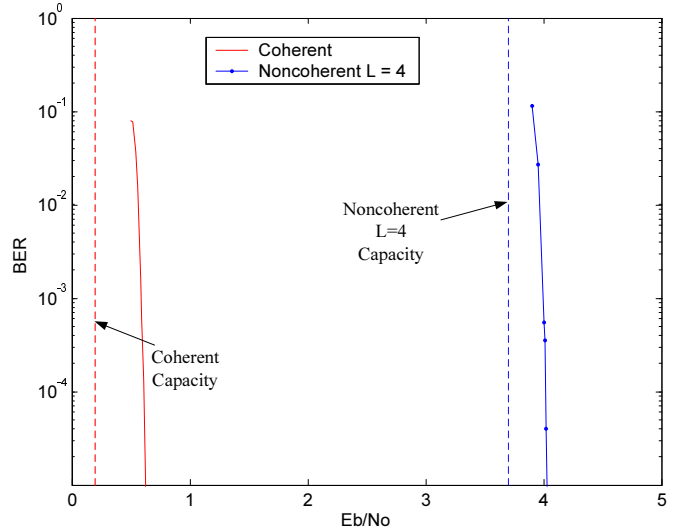
Fig. 4. Optimized EXIT curve for coherent detection of MSK

minimizes the area between the curves subject to the following constraints: (1) The outer code's information transfer function does not cross the inner code's; (2) the maximum variable-node degree is d_v ; and (3) The degree distributions are chosen such that the desired rate r is attained, where

$$r = \frac{\sum_{d=1}^{d_v} \frac{\lambda_d}{d}}{\sum_{d=1}^{d_c} \frac{\rho_d}{d}}. \quad (34)$$

Linear programming [9] may be used for the optimization process.

A code optimization was performed for MSK modulation with rate $r = 1/2$ coding. Both coherent and $L = 4$ blockwise noncoherent detection were considered. For both cases, the inner code's degree distribution was first set to $\rho_1 = 0.001$ and $\rho_2 = 0.999$. The maximum check node degree was set to 2 because it has a low computational complexity, yet still achieves good performance. A very small number of degree 1 check nodes are needed to allow the iterative decoding process to start properly. Otherwise, the decoding process always stays at origin of the EXIT chart [11]. Having fixed the inner code's degree distribution, the outer code's degree distribution was found under the constraint that the maximum outer-code degree is $d_v = 20$. The resulting optimized

Fig. 5. Optimized EXIT curve for $L = 4$ noncoherent detection of MSKFig. 6. BER of optimized MSK with rate $r = 0.5$ IRA coding.

degree distributions are shown in Table I. EXIT curves for the optimized coherent and $L = 4$ blockwise noncoherent systems are shown in Fig. 4 and 5, respectively. From the EXIT curves, the convergence thresholds are found to be $\mathcal{E}_b/N_0 = 0.48$ dB and $\mathcal{E}_b/N_0 = 3.94$ dB for coherent and $L = 4$ blockwise noncoherent reception, respectively.

Fig. 6 shows the bit error rate (BER) of the two optimized systems in Table I. In both systems, 100,000 uncoded bits are used and 200 iterations of sum-product [16] decoding are applied. For the coherent system, the numbers of degree 1 and 2 check nodes are 399 and 199602, respectively, while the numbers of degree 3, 7, and 8 variable nodes are 75911, 20842, and 3247 respectively. For the noncoherent system, the numbers of degree 1 and 2 check nodes are 399 and 199600, respectively, while the numbers of degree 3, 5, and 6 variable nodes are 55109, 35074, and 9817, respectively.

The noncoherent system uses the optional accumulator shown in Fig. 3. For coherent detection, the \mathcal{E}_b/N_0 required to achieve BER 10^{-5} is 0.63 dB, about 0.15 dB away from the estimated threshold and 0.43 dB away from the capacity. For $L = 4$ noncoherent detection, a more perfectly matched coding can achieve BER 10^{-5} at $\mathcal{E}_b/N_0 = 4.03$ dB, which is only 0.09 dB and 0.33 dB away from the threshold and capacity, respectively.

CONCLUSION

Coherent demodulation of CPFSK on a trellis requires that the modulation index h be rational, in which case the combination of modulator and AWGN channel is a finite-state Markov channel. The coherent demodulator also needs to know the set Φ of postulated accumulated phases, and in the absence of an additional tracking mechanism, this set must remain fixed for the duration of the sequence. Furthermore, the demodulator must know the initial phase at the start of the first symbol, though this requirement may be relaxed for long sequences. Under these assumptions, the capacity of coherent CPFSK with independent and uniformly distributed inputs can be found using a BCJR-like algorithm. While the results in this paper have concentrated on full-response CPFSK with rectangular phaseform shaping, the concepts can be easily extended to partial-response CPM with other phaseforms. All that is needed is a trellis representation of the phase trajectory.

Noncoherent demodulation is an attractive alternative to coherent demodulation. Unlike the coherent case, the complexity of the noncoherent detector does not depend on h , which may be any real number. This provides more design flexibility, especially in narrowband systems that tend to require small values of h that would result in very complex coherent detectors. Furthermore, noncoherent detection does not require knowledge of the set of phases Φ or the initial phase, and the set of phases may evolve due to, for instance, oscillator offsets or Doppler.

The main drawback of symbol-by-symbol noncoherent CPFSK is a very large penalty in energy efficiency. For example, the symbol-by-symbol noncoherent capacity of MSK at code rate $r = 0.5$ is 8.5 dB worse than the coherent capacity. Much of this loss can be recovered by using multi-symbol noncoherent block detection. For instance by performing detection over a block as small as $L = 4, 5$ dB

of the loss relative to coherent reception can be recovered.

For both the coherent and noncoherent cases, the capacity can be approached by using an IRA code designed using a curve-fitting technique. For the coherent case, no accumulator is needed since the modulation has memory. However, since noncoherent demodulation destroys the inter-block memory in the modulation, a differential precoder is recommended when the block length L is small.

REFERENCES

- [1] J. B. Anderson, T. Aulin, and C. E. Sundberg, *Digital Phase Modulation (Applications of Communications Theory)*. Springer, 1986.
- [2] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, pp. 3498 – 3508, Aug. 2006.
- [3] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.
- [4] S. Cheng, R. I. Seashadri, M. Valenti, and D. Torrieri, "The capacity of noncoherent continuous-phase frequency shift keying," in *Proc. Conference on Information Sciences and Systems (CISS)*, (Baltimore, MD), March 2007.
- [5] M. K. Simon and D. Divsalar, "Maximum-likelihood block detection of noncoherent continuous phase modulation," *IEEE Trans. Commun.*, vol. 41, pp. 90–98, Jan. 1993.
- [6] Y. Liang and V. Veeravalli, "Capacity of noncoherent time-selective rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3095 – 3110, Dec. 2004.
- [7] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2257 – 2270, Oct. 2004.
- [8] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, pp. 670–678, Apr. 2004.
- [9] A. Roumy, S. Guemghar, G. Caire, and S. Verdu, "Design methods for irregular repeat-accumulate codes," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1711 – 1727, Aug. 2004.
- [10] A. Guillén i Fàbregas and A. Grant, "Capacity approaching codes for the non-coherent FSK channel," in *Proc. 2006 Australian Communications Theory Workshop*, (Perth), Feb. 2006.
- [11] M. Xiao and T. Aulin, "Irregular repeat continuous phase modulation," *IEEE Commun. Letters*, vol. 9, pp. 723–725, Aug. 2005.
- [12] B. Rimoldi, "A decomposition approach to CPM," *IEEE Trans. Inform. Theory*, vol. 34, pp. 260 – 270, Mar. 1988.
- [13] J. Proakis, *Digital Communications*. New York, NY: McGraw-Hill, Inc., fourth ed., 2001.
- [14] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in *Proc. Int. Symp. on Turbo Codes and Related Topics*, (Brest, France), pp. 1–8, Sept. 2000.
- [15] S. ten Brink and G. Kramer, "Design of repeat-accumulate codes for iterative detection and decoding," *IEEE Trans. Signal Proc.*, vol. 51, pp. 2764–2772, Nov. 2003.
- [16] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498 – 519, Feb. 2001.