On Distribution Shaping Codes for Partial-Response Channels*

Joseph B. Soriaga and Paul H. Siegel
Dept. of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92037
{jsoriaga,psiegel}@ucsd.edu

Abstract
We consider a concatenated coding architecture for partial-response channels, comprising an inner modulation code and an outer parity-check code. In particular, we focus on designing inner finite-state encoders whose output statistics are similar to a near-capacity-achieving distribution for the channel. We show that both invertible and non-invertible finite-state inner encoders may be used to maximize the rate achievable with an outer parity-check code, which can be separately designed to ensure reliable communication. We then present two methods for the inner encoder design. The first method relates typicality with constraint graphs, and uses the state-splitting algorithm, for example, to produce encoders. For the second method, we group realizations of equiprobable binary $k$-tuples to produce a distribution with probabilities in multiples of $2^{-k}$, leading to an encoder that may be non-invertible. We use both methods to construct inner codes for the decoder channel, resulting in achievable rates near the best known lower bounds on capacity.

1 Introduction

Recently, significant advances have been made toward determining the capacity of noisy partial-response channels. We characterize these channels by the discrete-time relation

$$r_t = \sum_{i=0}^{\nu} h_i x_{t-i} + n_t$$

where $x_t \in \{\pm 1\}$ is the input, $r_t$ is the output, $\{h_0, \ldots, h_\nu\}$ is the channel impulse response, and $n_t$ is additive white Gaussian noise (AWGN) with mean zero and variance $\sigma^2$. In particular, several authors have independently discovered an efficient method for estimating the mutual information rate when the input is a finite-state Markov process (FSMP), and using this technique they have been able to improve existing lower bounds by optimizing the input FSMPs [1, 2, 3]. Additionally, methods to improve previous

*This research is supported in part by the Center for Magnetic Recording Research at the University of California, San Diego, and by the National Science Foundation under Grant CCR-0219582, part of the Information Technology Research program.
upper bounds have been presented in [4, 5], and these results suggest that the improved lower bounds are practically at capacity.

Currently the only system to closely approach these improved lower bounds on capacity is that of Kavčič, et al. [6, 7]. It consists of a serial concatenation in which the inner encoder produces an output process resembling the optimal input FSMP, and the outer low-density parity-check code reduces the overall error rate. However, the inner code design methodology uses randomized algorithms whose convergence is not well-characterized, and, furthermore, it may not be well-suited to channels with a large number of states.

In this paper, we present two methods for designing finite-state encoders that, when driven with equiprobable binary inputs, generate outputs that resemble the optimized input FSMPs. As demonstrated in [6, 7], under certain conditions one may combine an inner finite-state encoder of rate \( k : n \) with an outer parity-check code so that reliable communication can be achieved at any rate less than

\[
I(X; R) = \lim_{N \to \infty} \frac{1}{nN} I(X_1, \ldots, X_{nN}; R_1, \ldots, R_{nN} | s_0),
\]

where \( \{X_t\} \) is the process induced by the inner code, \( \{R_t\} \) is the channel output process, and \( s_0 \) is the initial state for the combined trellis of the channel and inner code. Moreover, (1) can readily be calculated using the Monte Carlo technique of [1, 2], with a slight modification to account for multiple symbols per edge. Consequently, we use (1) as a figure of merit when evaluating inner finite-state encoders, with the ultimate goal of designing a code whose achievable rate approaches that of an optimized input FSMP [1, 2, 3]. (However, due to space limitations, we do not address in this paper the design of the outer parity-check codes).

In Section 2, we show that (1) may be achieved with an outer parity-check code and an inner finite-state encoder which is possibly non-invertible, as long as the combination of the channel and inner encoder leads to an indecomposable "superchannel." Allowing the use of non-invertible inner encoders provides a new degree of flexibility in system design. In Section 3, we then provide two methods for finite-state encoder construction. The first method relates typicality and asymmetric running-digital-sum constraint graphs. From these graphs, we can design encoders for typical sequences using, for example, the state-splitting algorithm [8]. However, this method is only practical for single-parameter FSMPs with a small number of states. In the second method, we group realizations of an i.i.d. equiprobable binary \( k \)-tuple to create a distribution with probabilities in multiples of \( 2^{-k} \). By appropriately assigning these groups and probabilities to the edges of a graph representing the target FSMP, we produce an encoder whose output process approximates the desired process. Although this design approach may yield non-invertible encoders, we demonstrate its potential usefulness. Finally, we apply both methods to construct finite-state encoders for the dicode channel, the partial-response channel with impulse response \( \{h_0, h_1\} = \{1/\sqrt{2}, -1/\sqrt{2}\} \).

## 2 Achievable Rates with Concatenated Codes

Henceforth, we denote vectors with bold letters, e.g., \( X_N^N = (X_1 \ldots X_N) \), and use uppercase letters for random variables and lowercase for their realizations. We define a rate \( k : n \) finite-state encoder \( C \) as a finite-state machine whose state transitions depend on the current state \( s \) and input block \( u_t^k \in \{\pm 1\}^k \), i.e., \( s' = \tau_C(s, u_t^k) \), and whose output
blocks $x^n_i \in \{\pm 1\}^n$ depend on the input block and current state, i.e., $x^n_i \in E_C(s, u^t_i)$. This form of encoder is often represented by means of a state diagram or trellis.

Of course, the channel input-output relation can also be described by a finite-state-machine, with states $q_t = x^{t-1}_{t-1}$, and outputs $y_t = \sum_{i=0}^{\nu} h_i x_{t-i}$. The noisy channel output is given by $r_t = y_t + n_t$. One can combine the finite-state encoder and the channel; following [6, 7], and the references therein, we refer to this as a “superchannel.” This concept is illustrated in Figure 1(a), and an example of a superchannel trellis - corresponding to the biphase-coded dicode channel - is shown in Figure 1(b).

![Figure 1](image)

**Figure 1:** (a) Illustration of a superchannel, i.e., a combination of an inner finite-state encoder and channel. (b) Superchannel trellis for the biphase inner code (i.e., $u_j = (u_j, -u_j)$) on the dicode channel, labeled with inputs, biphase code outputs, and channel outputs, i.e., $u_j/(x_{2j-1}, x_{2j})/(y_{2j-1}, y_{2j})$.

**Lemma 1.** Consider a rate $k : n$ finite-state encoder and partial-response channel for which the corresponding superchannel is indecomposable [9, p. 106]. Let $\{X_i\}$ be the encoder output process generated by i.i.d. equiprobable inputs $\{U_i\}$, and let $I(X; R)$ be the mutual information rate of the coded channel.

If we define

$$I(U; R) = \lim_{N \to \infty} \frac{1}{nN} I(U_1^{kN}; R_1^{nN} | s_0),$$

where $s_0$ is any initial superchannel state, then $I(U; R) = I(X; R)$.

**Corollary 1.** By further serially concatenating an outer parity-check code (or coset) of rate $R_{\text{outer}}$, reliable communication can be achieved for any $kR_{\text{outer}}/n < I(U; R)$.

These concatenated coding results are similar to, but more general than, those in [6, 7], in that they can be applied to non-invertible encoders with $k > n$. It is somewhat surprising that, in this case, reliable communication may still be achieved with an appropriately chosen outer code. Allowing the encoder to be non-invertible provides added flexibility in the inner code design, as we demonstrate in Section 3.2.

**Proof of Lemma 1.** The lemma considers only finite-state encoders leading to indecomposable superchannels (i.e., for which the combined transition mapping can be represented by an irreducible aperiodic graph [8]) to ensure that the limits in (1) and (2) exist and are independent of the initial state $s_0$ [9, p. 109].

Notice that $I(U_1^{kN}; R_1^{nN} | s_0) = I(X_1^{nN}; R_1^{nN} | s_0)$, for any $N > 0$. Specifically, since $R_1^{nN}$ is independent of $U_1^{kN}$ when conditioned on both $X_1^{nN}$ and $s_0$, we find from [10, p. 27] that $I(U_1^{kN}; R_1^{nN} | X_1^{nN}, s_0) = 0$. Thus,

$$I(U_1^{kN}, X_1^{nN}; R_1^{nN} | s_0) = I(X_1^{nN}; R_1^{nN} | s_0) + I(U_1^{kN}; R_1^{nN} | X_1^{nN}, s_0) = I(X_1^{nN}; R_1^{nN} | s_0).$$

470
Similarly, $I(X_1^{nN}; R_1^{nN} | U_1^{kN}, s_0) = 0$, because $X_1^{kN}$ is a function of $U_1^{kN}$ and $s_0$. Thus,

$$I(U_1^{kN}; X_1^{nN}; R_1^{nN} | s_0) = I(U_1^{kN}; R_1^{nN} | s_0) + I(X_1^{nN}; R_1^{nN} | U_1^{kN}, s_0) = I(U_1^{kN}; R_1^{nN} | s_0),$$

and since this implies that the limits in (1) and (2) are equal, the lemma is proved.  \(\square\)

**Proof of Corollary 1.** Theorem 6.2.1 in [9, p. 206] states that, for discrete memoryless channels, both the random i.i.d. equiprobable codeword ensemble and the random coset code ensemble exhibit a vanishing average probability of error for any rate less than the mutual information rate corresponding to an i.i.d. equiprobable input process. The corollary follows from a straightforward extension of this theorem.  \(\square\)

## 3 Modulation Code Design

In this section, we present two methods to construct a finite-state encoder whose outputs resemble a specific FSM. As a figure of merit, we use $I(U; R)$, the achievable rate of the channel with a finite-state encoder and i.i.d. equiprobable binary inputs. Both construction methods are demonstrated for the dicode channel.

### 3.1 Encoders from Constraint Graphs for Typical Sequences

We first describe constraint graphs for typical sequences of two classes of Markov processes. In Section 3.1.1, the graph corresponds to strongly typical sequences generated by an i.i.d. Bernoulli process. In Section 3.1.2, we provide an extension from the Bernoulli case to that of a binary symmetric 2-state Markov process, i.e., a binary Markov process $\{X_t\}$ for which $P(X_t = 1|X_{t-1} = 1) = P(X_t = -1|X_{t-1} = -1) = p$.

By applying the state-splitting algorithm [8] to such a constraint graph, one can design invertible finite-state encoders at any rational rate less than or equal to the Shannon capacity of the constraint. We illustrate this inner encoder design methodology for a specific binary symmetric Markov process on the dicode channel.

#### 3.1.1 I.I.D. Bernoulli Processes

Suppose $\{X_t\}$ is a sequence of i.i.d. random variables, with $P(X_t = 1) = p$. A sequence $x_1^N$ is defined to be $\epsilon$-strongly typical with respect to this distribution if

$$\left| \frac{N_1(x_1^N)}{N} - p \right| < \frac{\epsilon}{2},$$

where $N_1(x_1^N)$ equals the number of 1’s in the sequence [10, p. 288]. We let $A_\epsilon^{(N)}$ denote the set of all typical sequences of length $N$.

Now consider the labeled graph $G_{K,r,s}$ shown in Figure 2, and let $S_N(G_{K,r,s})$ denote the set of all binary length-$N$ sequences generated from paths in $G_{K,r,s}$. Any $x_1^N \in S_N(G_{K,r,s})$ clearly must satisfy the bound $|r N_1(x_1^N) - s N_{-1}(x_1^N)| < 2K$. After substituting $N_{-1}(x_1^N) = N - N_1(x_1^N)$, this bound can be rewritten as

$$\left| \frac{N_1(x_1^N)}{N} - \frac{s}{r+s} \right| < \frac{2K}{N(r+s)};$$
and we find that for any $K$ and any $\epsilon > 0$, all generated sequences of length $N > 4K/\epsilon(r + s)$ are $\epsilon$-strongly typical of an i.i.d. Bernoulli sequence $\{X_i\}$ with $p = s/(s + r)$. Notice that because $S_N(G_{K,r,s}) \subseteq A^\epsilon(N)$ for all sufficiently large $N$, it follows from the asymptotic equipartition property \cite[p. 51]{10} that $\text{Cap}(G_{K,r,s}) \leq h(s/(s + r))$, where $h(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$.

Whenever $K$ is a multiple of both $s$ and $r$, the graph $G_{K,r,s}$ supports any cycle (i.e., a sequence of edges beginning and ending at the same state) which originates from state 0 and has $K/r$ forward steps, $K/s$ backward steps, and thus an overall length of $N = K/r + K/s$. The total number of such cycles is

$$\binom{K/r + K/s}{K/r} = \binom{N}{sN/(s + r)} \geq \frac{1}{N + 1} 2^{Nh(s/(s + r))}.$$

(The lower bound can be found in \cite[Example 1.2.1.3]{10}.) Since cycles can be concatenated to form longer allowable sequences, we conclude that

$$\text{Cap}(G_{K,r,s}) \geq h(s/(s + r)) - \frac{1}{N} \log_2 (N + 1).$$

Hence, the capacity of the graph converges to $h(s/(s + r))$ as $K$ increases. By choosing $G_{K,r,s}$ corresponding to a large value of $K$, we can, in principle, use the state-splitting algorithm to design a finite-state encoder with rate arbitrarily close to $\text{Cap}(G_{K,r,s})$, thus approximating an ideal probability transformer from an i.i.d. Bernoulli$(1/2)$ process to an i.i.d. Bernoulli$(s/(s + r))$ process.

The properties of these constraint graphs are summarized in the theorem below.

**Theorem 1.** Let $A^\epsilon(N)$ be the set of binary length-$N$ sequences that are $\epsilon$-strongly typical with respect to an i.i.d. Bernoulli process $\{X_i\}$ with $P(X_i = 1) = s/(s + r)$, where $s$ and $r$ are positive integers.

For the sequence of graphs $\{G_{K,r,s}\}_{K > 0}$, and for any $\epsilon > 0$, there exists a $K'$ such that

$$\left| \text{Cap}(G_{K,r,s}) - h\left(\frac{r}{r + s}\right) \right| < \epsilon,$$

for all $K > K'$,

and for each such $K$ there exists an $N'$ such that $S_N(G_{K,r,s}) \subseteq A^\epsilon(N)$ for all $N > N'$.

The limiting capacity for the graphs $\{G_{K,r,s}\}$, as $K \to \infty$, was determined earlier by Janssen and Immink \cite{11}, who investigated the entropy and power spectrum of asymmetric runlength constraints. The proof given here is arguably simpler, and it can be easily applied to more general cases (not shown in this paper). However, the capacity bound in \cite{11} converges more rapidly to the limiting value than does ours.
3.1.2 Binary Symmetric 2-State Markov Process

We now extend this idea to a binary symmetric 2-state Markov process, as shown in Figure 3(a), by introducing the constraint graph $H_{K,r,s}$ shown in Figure 3(b). Notice that the bottom states in $H_{K,r,s}$ all have incoming edges labeled with $-1$, and the top states have incoming edges labeled with $1$. For $x_1^N \in S_N(H_{K,r,s})$, let $N_{i,j} = N_{i,j}(x_1^N)$ be the number of edges in the sequence $x_1^N$ that are labeled with a $j$ but originate from a state with incoming edges labeled $i$. Clearly then, $x_1^N$ must satisfy

$$\left| \frac{N_{\text{trans}}}{N} - \frac{s}{r+s} \right| < \frac{2K}{N(r+s)}, \quad \left| \frac{N_{\text{non-trans}}}{N} - \frac{r}{r+s} \right| < \frac{2K}{N(r+s)}.$$ 

By observing that

$$\frac{1}{N} \log_2 P(X_1^N = x_1^N) = \frac{N_{\text{trans}}}{N} \log_2 \left( \frac{s}{r+s} \right) + \frac{N_{\text{non-trans}}}{N} \log_2 \left( \frac{r}{r+s} \right),$$

we can write

$$\left| -\log_2 P(X_1^N = x_1^N) - h \left( \frac{r}{r+s} \right) \right| < \frac{2K}{N(r+s)} \log_2 \left( \frac{(r+s)^2}{rs} \right),$$

and one can determine an $N'$ such that $S_N(H_{K,r,s}) \subseteq A_i^{(N)}$ for all $N > N'$.

Cycle counting arguments similar to those used in Section 3.1.1 yield the following extension of Theorem 1.
Theorem 2. Let \( \{X_t\} \) be a binary symmetric 2-state Markov process with \( P(X_t = x \mid X_{t-1} = x) = r/(r + s) \) for \( x = \pm 1 \), where \( s \) and \( r \) are positive integers, and let \( A_{\epsilon}^{(N)} \) be the set of \( \epsilon \)-weakly typical sequences with respect to this distribution, i.e.,

\[
A_{\epsilon}^{(N)} = \left\{ x_1^N : \left| \frac{1}{N} \log_2 P(X_1^N = x_1^N) - h \left( \frac{r}{r+s} \right) \right| < \epsilon \right\}.
\]

For the sequence of graphs \( \{H_{K,r,s}\}_{K>0} \) and for any \( \epsilon > 0 \), there exists a \( K' \) such that

\[
\left| \text{Cap}(H_{K,r,s}) - h \left( \frac{r}{r+s} \right) \right| < \epsilon, \quad \text{for all } K > K',
\]

and for each such \( K \) there exists an \( N' \) such that \( S_N(H_{K,r,s}) \subseteq A_{\epsilon}^{(N)} \) for all \( N > N' \).

As an example, we considered constraint graphs \( H_{K,1,2} \), for various \( K \), and used the state-splitting algorithm to design encoders\(^1\) for a target binary symmetric 2-state Markov process with \( P(X_t = x \mid X_{t-1} = x) = 1/3 \). The achievable rates of these encoders on the decode channel are plotted in Figure 4(a). We observe that, as the encoder rate increases, \( I(U; R) \) approaches the \( I(X; R) \) curve for the target process. In Figure 4(b), we show the achievable rates on the decode channel for encoders derived from \( H_{K,r,s} \) for selected other values of \( K, r, \) and \( s \). We see that this method realizes a large part of the gain possible with shaped inputs on the decode channel.

3.2 Encoders for FSMPs using Simple Bit Groupings

The second method we present is based on the observation that grouping realizations of i.i.d. equiprobable binary \( k \)-tuples can produce distributions with probabilities that are multiples of \( 2^{-k} \). For instance, suppose we want to generate the FSMP \( \{X_t\} \) shown in Figure 5(a), but are given only an i.i.d. equiprobable sequence \( \{U_t\} \). Observing that the edge probabilities in the FSMP are integer multiples of \( 1/4 \), we construct the finite-state encoder in Figure 5(b) to accomplish the task. However, this encoder is non-invertible, because there exist parallel edges with the same output value. Nevertheless, Lemma 1 proves that this encoder can still be combined with an outer parity-check code to achieve reliable performance at any rate less than \( I(X; R) \).

The technique of using such lossy shaping encoders in serially concatenated systems with an outer parity-check code was actually presented much earlier by Gallager [9, p. 208] in the context of coding for discrete memoryless channels. Gallager also suggests that any distribution can be approximated with \( 2^{-k} \) precision, and by increasing \( k \) we expect pointwise convergence to the target distribution. Of course, for a binary FSMP with one symbol per edge, as the inner code rate \( k : 1 \) increases, the corresponding outer code rate must decrease accordingly (Corollary 1). To avoid such low outer code rates, we instead approximate the \( n \)-symbol-per-edge distribution to the nearest \( 2^{-k} \), leading to a rate \( k : n \) encoder. Interestingly, as shown in Figure 6, the achievable rate of the target process can still be very closely approached as \( k/n \) approaches 1 from above, suggesting that pointwise convergence is not a necessary criterion for good shaping. (More details on this will be presented in forthcoming work.)

\(^1\)No attempt was made to optimize the sequence of state-splittings.
Figure 4: Achievable rates for various FSMPs and encoders on the dicode channel. In both plots, the upper heavy solid line (•) is the capacity lower bound, and the bottom heavy solid line (•) is \( I(\mathcal{X}; \mathcal{R}) \) when \( \{X_t\} \) are i.i.d. equiprobable. (See [2]).

(a) The dashed line (—) corresponds to a binary symmetric 2-state Markov process, with \( p = 1/3 \), and the thin solid lines (—) correspond to the maxentropic measures on the constraint graphs \( H_{5,1,2} \) and \( H_{9,1,2} \) (see Figure 3(b)), with \( \text{Cap}(H_{5,1,2}) \approx 0.528 \), and \( \text{Cap}(H_{9,1,2}) \approx 0.780 \). Finally, the dash-dot lines (---) show \( I(U; R) \) for a rate 1:2, 11-state encoder derived from \( H_{5,1,2} \), and a rate 3:4, 38-state encoder derived from \( H_{9,1,2} \).

(b) Achievable rates for encoders derived from \( H_{K, r, s} \) for selected values of \( K, r, \) and \( s \). The legend also shows the encoder rate, \( k : n \), and the number of encoder states.

4 Conclusion

We considered the design of finite-state encoders to maximize the achievable rate on a noisy partial-response channel using an outer parity-check code. Letting \( \{X_t\} \) denote the output process produced when the inner encoder is driven by i.i.d. equiprobable inputs \( \{U_t\} \), and letting \( \{R_t\} \) denote the corresponding noisy channel output process, we showed that the mutual information rate \( I(\mathcal{X}; \mathcal{R}) \) is achievable with an appropriately designed outer parity-check code when the superchannel consisting of the encoder and channel is indecomposable. Notably, this result does not preclude inner encoders that are non-invertible, thus providing greater flexibility in the overall system design.

We then presented two methods for constructing the inner encoders. The first relies on constraint graphs which describe typical sequences for single-parameter, 1-state and 2-state Markov processes, from which one can, in principle, design efficient finite-state encoders using, for example, the state-splitting algorithm. This method becomes impractical for more complex processes.

The second method involves grouping realizations of an i.i.d. equiprobable binary \( k \)-tuple and suitably assigning them to edges of a graph representation of the desired FSMP, thereby defining an encoder that approximates the desired transition probabilities with integer multiples of \( 2^{-k} \). Although this method may result in a non-invertible encoder, it can nevertheless be used to closely approximate any FSMP.

The remaining task to be considered in future work is the design of an outer parity-check code and the development of a decoder such that reliable communication can
be achieved near $I(U; R)$ with these inner codes. One approach would be to use the technique of [6, 7] to design outer low-density parity-check codes for use with turboequalization. We are also considering an alternative approach based upon multilevel encoding with multistage decoding [2, 12].

5 Acknowledgements

The authors are grateful to Dr. Henry Pfister for many helpful discussions.

References


Figure 6: Achievable rates on the dicode channel for simple rate $k : n$ encoders obtained from Section 3.2, where the target process is an 8-state binary Markov process optimized for $\sigma = 0.95$ (i.e., where $C_{LB} \approx 0.5$). The upper heavy solid line ($\cdot$) is the capacity lower bound, and the bottom heavy solid line ($\cdot$) is $I(X; R)$ when $\{X_t\}$ is i.i.d. and equiprobable (from [2]). The thin solid line (---), dashed line (- -), dash-dot line (---) represent achievable rates with inner 8-state encoders of rate 2:1, 3:2, and 4:3, respectively.


