A Study on Universal Codes with Finite Blocklengths

Jun Shi, *Member, IEEE,* Richard D. Wesel, *Senior Member, IEEE*

***Abstract*—Based on random codes and typical set decoding, an alternative proof of Root and Varaiya's compound channel coding theorem for linear Gaussian channels is presented. The performance limit of codes with finite blocklength under a compound channel is studied through error bounds and simulation. Although the theorem promises uniform convergence of the probability of error as the blocklength approaches infinity, with short blocklengths the performance can differ considerably for individual channels. Simulation results show that universal performance can be a practical goal as the blocklengths become large.**

***Index Terms*—compound channel, random coding bound, sphere packing bound, universal code.**

I. Introduction

Traditional code design is often targeted at a specific channel. The performance of such channel-specific codes can deteriorate significantly when these codes are faced with unexpected channels. For example, an optimal additive white Gaussian noise (AWGN) code might not perform well under periodic erasure channels [1] or partial band jamming channels [2]. With space-time codes, the comparison in Fig. 6 of [3] shows that a code in [4] optimized for Rayleigh fading performs poorly in the special case of singular channel. In the other extreme, also in Fig. 6 of [3], an Alamouti space-time block code performs very well on singular channels but has poor performance on unitary channels.

One approach to solve this problem is to design individual optimal codes for each channel condition. However this scheme requires storage for all the possible codes at the transmitter and the receiver and the ability of both sides to intelligently identify and adapt to the environment. An alternative approach is to design a universal code, one that works reasonably well under most, if not all, possible scenarios. In this paper, we study both the theoretic and practical aspects of the latter approach, focusing on how blocklength affects the behavior of universal codes.

The rest of the paper is organized as follows. Section II reviews the compound channel coding theorem. Section III introduces the periodic erasure channel, a simple compound channel. Section IV discusses the figures of merit that are useful for performance evaluation on compound channels. Section V examines the finite-blocklength behavior of universal codes through error bounds and computer simulation. Section VI concludes the paper. The appendix gives the details of a typical-set-decoding proof of the linear Gaussian compound channel coding theorem.

Jun Shi is with Qualcomm Inc, San Diego, CA, 92122. Email: juns@qualcomm.com.

Richard D. Wesel is with Department of Electrical Engineering, the University of California, Los Angeles. Email: wesel@ee.ucla.edu.

This work was supported by the National Science Foundation through Grant CCR-0209110 and ST Microelectronics and the state of California through MICRO grant 03-031. The material in this paper was presented in part at MILCOM 2004, Monterey, CA.

II. Compound Channel Coding Theorem

A compound channel arises when users communicate under some channel uncertainty [5], i.e. users know the channel belongs to a family of channels but they do not know exactly what the channel is. Throughout this paper, we will restrict our discussion to discrete memoryless channels (DMC's).

A set of DMC's comprises compound channel as follows.

Definition 1: A compound channel is a family of channels indexed by $i \in \mathcal{I}$ denoted by

$$\{P(y|x,i), x \in \mathcal{X}, y \in \mathcal{Y}, i \in \mathcal{I}\}, \tag{1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabet, respectively. $\mathcal{I}$ is the channel index set which can be finite, countably infinite, or uncountably infinite. $P(y|x,i)$ is the conditional probability governing the channel with index i. ■

We assume that the channel index remains unchanged during the course of the transmission, or at least the time that the channel index stays the same is longer than the codeword blocklength. If the index varies arbitrarily from symbol to symbol, then such a channel is referred as an arbitrarily varying channel [5], which is not the focus of this paper.

The capacity of a compound channel is defined as

$$C(\mathcal{I}) = \sup_{p(x)\in\mathcal{S}} \inf_{i\in\mathcal{I}} I_i(x;y), \tag{2}$$

where $I(x;y)$ is the mutual information between the input and output random variables. The supremum in (2) is over all possible input distributions $p(x)$ in the set $\mathcal{S}$, which is usually specified by the input power constraint. Define the infimum of the individual channel capacities as

$$C_{\text{inf}}(\mathcal{I}) = \inf_{i\in\mathcal{I}} \sup_{p(x)\in\mathcal{S}} I_i(x;y). \tag{3}$$

Note that $C(\mathcal{I}) \leq C_{\text{inf}}(\mathcal{I})$. However it can be shown that $C(\mathcal{I}) = 0$ if and only if $C_{\text{inf}}(\mathcal{I}) = 0$ [6]. So any set of positive-capacity channels will have a positive compound channel capacity.

Blackwell, Breiman and Thomasian [7] proved that the capacity of a compound channel with a discrete alphabet can be achieved by a single sequence of codes. A similar result also appeared in [6]. This result was extended by Root and Varaiya [8] to $m \times m$ (square) linear Gaussian compound channels where the alphabet is continuous. A slight generalization of their theorem to the $m_r \times m_t$ (rectangular) MIMO channels is stated below.

Theorem 1: A family of real Gaussian multiple-input multiple-output (MIMO) channels is denoted as $\{\boldsymbol{H}_i, \boldsymbol{K}_i, i \in \mathcal{I}\}$, where the index set $\mathcal{I}$ is an arbitrary set. The channel input/output behavior is governed by $\boldsymbol{y} = \boldsymbol{H}_i \boldsymbol{x} + \boldsymbol{z}_i$, where $\boldsymbol{H}_i$ is an $m_r \times m_t$ real matrix representing the path gains between m_t transmitter antennas and m_r receiver antennas. The input $\boldsymbol{x}$ is an $m_t \times 1$ real random vector with zero mean and covariance matrix $\boldsymbol{S} \in \mathcal{S}$, where $\mathcal{S}$ is the set of positive symmetric matrices with trace less or equal to P_x. The output $\boldsymbol{y}$ is an $m_r \times 1$ real random vector. The noise $\boldsymbol{z}_i$ is a real Gaussian random vector of dimension $m_r \times 1$ with zero mean and covariance matrix $\boldsymbol{K}_i$. Assume there exist real numbers α_0, α_1 and α_2 satisfying $0 < \alpha_0$ and $0 < \alpha_1 < \alpha_2$, such that for each $i \in \mathcal{I}$,

1) $\|\boldsymbol{H}_i\| \le \alpha_0$, where $\|\cdot\|$ is the spectral norm of a matrix, i.e. the square root of the largest eigenvalue of $\boldsymbol{H}_i^\intercal \boldsymbol{H}_i$.
2) $\alpha_1 \le \frac{\boldsymbol{v}^\intercal \boldsymbol{K}_i \boldsymbol{v}}{\|\boldsymbol{v}\|^2} \le \alpha_2$, for any non-zero real $m_r \times 1$ vector $\boldsymbol{v}$.

where $\intercal$ stands for transpose. Then any rate $R < C(\mathcal{I})$ defined in (2) is achievable. i.e. There exists a sequence of $(2^{nR}, n)$ codes such that the probability of error under any channel in the family approaches zero as the blocklength approaches infinity.

Proof: The theorem for real square channel matrices first appeared in [8]. The decoding techniques used in the original proof complicate the error probability bounding. In the appendix, we give an alternative proof based on random codes and typical-set decoding. ■

Root and Varaiya's compound channel coding theorem for MIMO channels appeared in 1968, although it is only recently that MIMO channels have drawn considerable attention. A natural application of the compound channel coding theorem is universal space-time code design as in [3], [9], [10].

The essential fact about the compound channel coding theorem is that the probability of error goes to zero uniformly as long as the code rate is less than the compound channel capacity no matter what channel the sequence of codes is actually encountering. This uniform convergence does not mean that the error probability of each channel in the compound channel goes to zero at exactly the same speed, but the speed is at least lower bounded. The difference is negligible at large blocklengths, but significant for codes with relatively short blocklengths.

III. Periodic Erasure Channels

In order to evaluate the finite-blocklength performance of a single code under various channels, we need a relatively simple compound channel. In Theorem 1, the compound channels are matrix channels. Although these matrix channels provide plenty of flexibility, they complicate the mathematical analysis. One simple compound channel involves erasing transmitted symbols periodically with different patterns. Each distinct erasure pattern generates a distinct channel.

As far as the effect of blocklength is concerned, sending a codeword through a periodic erasure channel is equivalent to puncturing the codewords first then sending through a standard (unerased) channel. This equivalence provides great convenience in computing the error bounds as explained in the rest of the paper. As a result, We are able to gain insight into the short-blocklength behavior.

Consider a binary symmetric channel (BSC) with input symbols $x_i \in \{-1, 1\}$. The channel is expressed by

$$y_i = f(x_i, p) = \begin{cases} x_i & \text{with probability } 1-p \\ -x_i & \text{with probability } p \end{cases}.$$

For a BSC with periodic erasures, $y_i = a_i f(x_i, p)$, where $a_i \in \{0, 1\}$ are the erasing coefficients with period T, i.e. $a_i = a_{i+T}$. Similarly an AWGN periodic erasure channel can be formulated as $y_i = a_i x_i + n_i$, where n_i is the Gaussian noise.

The periodic erasure channel can be regarded as a very simple matrix channel that is a diagonal matrix containing only ones and zeros. It is a simplified model for frequency-hopped or OFDM systems where partial band interference arises due to frequency dependent disturbance or jamming [2].

IV. Figure of Merit

Before we analyze the performance of an error-correcting code under various channels, a fair and convenient figure of merit is needed. This figure should automatically take channel conditions into consideration and act consistently across all channels.

Some figures of merit depend on the regime of concern. For example, in the low-SNR regime and the wide-band regime, Verdú proposes the normalized energy per information bit E_b/N_0 to be the figure of merit [11]. In this paper, we do not limit analysis in the low-SNR regime, thus precluding the use of E_b/N_0. Conventionally, the code performance in an AWGN channel is gauged by the signal-to-noise-ratio (SNR) required to achieve a certain target bit error rate (BER) or frame error rate (FER). SNR can be used to evaluate different codes under the same channel condition and the same rate. However as will be shown later, it is not suitable to evaluate codes under multiple channel conditions or compound channels. One possible metric is the normalized SNR proposed by Forney [12], defined as

$$\text{SNR}_{\text{norm}} = \frac{\text{SNR}}{2^{2R} - 1}, \tag{4}$$

where R is the code rate, and $10 \log_{10}(\text{SNR}_{\text{norm}})$ is often called the SNR gap of a code. The value of the SNR gap indicates how far a system is operating from the Shannon limit.

In [1] excess mutual information (EMI) was proposed as the figure of merit for the purpose of universal code design. It is defined as:

$$\text{EMI}(\text{SNR}, R) = I(\text{SNR}) - R, \tag{5}$$

where R is the rate of the code and SNR is the signal-to-noise-ratio at which the code achieves a certain target probability of error, $I(\text{SNR})$ is the mutual information of the channel at that SNR given the input distribution. For the AWGN channel with Gaussian input, $I(\text{SNR})$ coincides with the capacity of the channel. EMI indicates the penalty in terms of untransmitted information that is paid due to the imperfectness of the code.

For an AWGN channel, EMI and SNR gap can be approximately related by a constant factor in the high-SNR and high-rate regime. Assuming a real AWGN channel with Gaussian input, SNR gap can be written as

$$\begin{aligned} \text{SNR gap} &= 10\log_{10}(\text{SNR}) - 10\log_{10}(2^{2R}-1) \\ &\approx 10\log_{10}(\text{SNR}) - 20R\log_{10}(2). \end{aligned} \tag{6}$$

EMI can be written as

$$\begin{aligned} \text{EMI} &= \frac{1}{2}\log_2(1+\text{SNR}) - R \\ &\approx \frac{1}{2}\log_2(\text{SNR}) - R. \end{aligned} \tag{7}$$

Thus SNR gap $\approx 20\log_{10}(2)$ EMI. The approximations in (6) and (7) are valid when R and SNR are both large. However, if we perform the same analysis for a channel where every other symbol is erased, we see that SNR gap $\approx 40\log_{10}(2)$ EMI. Thus SNR gap and EMI are fundamentally different figures of merit, and one must choose between them.

We prefer EMI to SNR gap because EMI indicates how the error probability changes as blocklength increases. For typical-set decoding in a Gaussian channel, the error probability of a Gaussian code book of blocklength n is upper bounded by [13, p. 245]

$$\begin{aligned} P_e^{(n)} &\le 2\epsilon + 2^{3n\epsilon}2^{-n(I(X;Y)-R)} \\ &= 2\epsilon + 2^{3n\epsilon}2^{-n\text{EMI}}. \end{aligned} \tag{8}$$

Although EMI is not the true error exponent for typical set decoding due to the fact that the ϵ in (8) hides too much information, the appearance of EMI in the exponent in (8) manifests its importance in determining the error probability.

V. Performance Evaluation

A finite blocklength code is often compared to the Shannon capacity to measure its imperfectness. However this comparison is not completely fair; in most cases, the Shannon capacity can only be achieved as the blocklength goes to infinity. Among available finite-blocklength analysis tools are the sphere-packing bound (SPB) [14] and the random-coding bound (RCB) [15]. The probability of error for codes with finite blocklength is lower bounded by the SPB.

The RCB, characterizing the average performance of randomly selected codes, serves as an upper bound on the probability of error for an optimal code. However in reality it might be the case that even the RCB cannot be achieved by a carefully designed code due to the increasing decoding complexity. A fair assessment of a finite-blocklength code can be made by measuring its EMI against the EMI of the SPB or the RCB.

In what follows, we will state the SPB and the RCB for the BSC and the AWGN channel, then use examples to illustrate the finite-blocklength behavior of universal codes.

A. Binary Symmetric Channels

1) Sphere-Packing Bound: The derivation of the SPB for the probability of codeword error $P_w(n,k,p)$ on the BSC is combinatorial. For an $(n,\ k)$ binary code, the bound can be written as [15], [16]

$$\begin{aligned} P_w(n,k,p) \ge & p^{r+1}(1-p)^{n-r-1}\left(\sum_{i=0}^{r+1}\binom{n}{i} - 2^{n-k}\right) \\ & + \sum_{i=r+2}^{n}\binom{n}{i}p^i(1-p)^{n-i} \\ = & \sum_{i=r+1}^{n}\binom{n}{i}p^i(1-p)^{n-i} \\ & - p^{r+1}(1-p)^{n-r-1}\left(2^{n-k} - \sum_{i=0}^{r}\binom{n}{i}\right), \end{aligned} \tag{9}$$

where p is the crossover probability of the BSC and r is the maximum integer such that $\sum_{i=0}^{r}\binom{n}{i} \le 2^{n-k}$.

2) Random Coding Bound and Error Exponent: The RCB for the BSC is computed as [15, p. 146]

$$P_w^n \le e^{-nE_r(R)}, \tag{10}$$

where $E_r(R)$ is the random coding exponent. However, even for such a simple channel, $E_r(R)$ does not have a simple explicit form. We need first to compute an intermediate parameter δ whose relation with the code rate R (in nats) is given as:

$$R = \ln 2 - H(\delta), \tag{11}$$

where $H(\delta)$ is the binary entropy function (in nats). The possible range of δ is $p \le \delta \le 1/2$.

For δ in the range

$$p \le \delta \le \frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}}, \tag{12}$$

where $p \le 1/2$ is the crossover probability, the random coding exponent of the BSC is

$$E_r(R) = -\delta\ln p - (1-\delta)\ln(1-p) - H(\delta). \tag{13}$$

For

$$\frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}} < \delta \le 1/2, \tag{14}$$

the exponent becomes

$$E_r(R) = H(\delta) - 2\ln(\sqrt{p}+\sqrt{1-p}). \tag{15}$$

3) Extension to Periodic Erasure Channels: As mentioned earlier, erasures by the channel are equivalent to punctures by the transmitter as long as the receiver has full knowledge which symbols are erased or punctured. We also assume the erasure period is much shorter than the blocklength so that the effect of the last (possibly fractional) period can be negligible. The previously-listed bounds and error exponents can be extended to erasure channels by this equivalence. For example, the RCB of a rate-$1/4$ length-N code with erasure pattern “10” is the same as that of a rate-$1/2$ length-$N/2$ code without erasures. The same is true for the SPB.

4) Numerical Analysis: The random-coding error exponent indicates the rate at which the error probability of the ensemble codes approaches zero as blocklength grows. We consider a BSC compound channel with erasure patterns "11111" and "00111". For a fair comparison, the BSC's have a crossover probability 0.11 for the first erasure pattern and 0.0246 for the second pattern such that both channels have capacity 0.5 bit. The random-coding error exponents are plotted in Fig. 1 for various code rates. It is clear that the erased channel is more favorable than the unerased channel from the perspective of error exponent vs. transmitted rate on a channel with a fixed capacity.

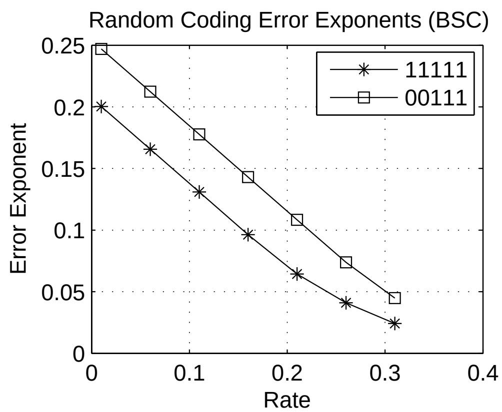


Fig. 1. Random-coding error exponents for a BSC with capacity 0.5 bit. 11111–unerased channel, 00111–erasure channel.

We now show the error bounds vs. EMI. Suppose CH1 is the standard BSC with crossover probability p and CH2 is a BSC with the same crossover probability p but with periodic erasure pattern "00111". The capacities of two channels are:

$$C_1(p) = 1 - H(p), \qquad C_2(p) = \frac{3}{5}(1 - H(p)), \tag{16}$$

where $H(p) = -(1-p)\log_2(1-p) - p\log_2(p)$ is the binary entropy function in bits. In both cases, the code rate is $1/4$. According to (5), the EMI for CH1 is calculated by $\text{EMI}_1 = C_1(p) - 0.25$ and the EMI for CH2 is $\text{EMI}_2 = C_2(p) - 0.25$. Bounds on the frame error rate for blocklength 100 are plotted against EMI in Fig. 2. Again the erasure channel is a more EMI-efficient channel according to the SPB and the RCB. So one would expect a short-blocklength universal code to have better performance in an erasure channel than in a standard channel in terms of EMI.

B. AWGN Channels

1) Sphere-Packing Bound: The derivation of Shannon's SPB for the AWGN channel is essentially geometric. The codewords of blocklength n are regarded as points on the surface of an n-dimensional sphere with radius $\sqrt{nE_s}$. The error probability is lower bounded by the probability that an n-dimensional Gaussian random variable falls outside a cone whose cap area corresponds to that of the Vonoroi region of the transmitted codeword. The error probability is given as [14], [17]:

$$P_w \geq Q_n(\theta_s, A), \tag{17}$$

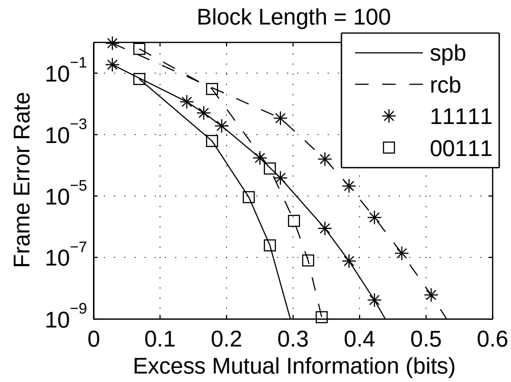


Fig. 2. Sphere-packing bound (SPB) and random-coding bound (RCB) for BSC with blocklength 100. 11111–unerased channel, 00111–erasure channel.

where $A = \sqrt{E_s/N_o}$, θ_s is the half-cone angle of a cone whose normalized solid angle is $1/2^k$. i.e. Its solid angel is $1/2^k$ of the total solid angle of the n-dimensional sphere, where k is the information bit length. The half-cone angle θ_s satisfies

$$\Omega_n(\theta_s) = \int_0^{\theta_s} \frac{(n-1)\Gamma(\frac{n}{2}+1)(\sin\phi)^{(n-2)}}{n\Gamma(\frac{n+1}{2})\sqrt{\pi}} d\phi = \frac{1}{2^k}. \tag{18}$$

$Q_n(\theta_s, A)$ is given as

$$Q_n(\theta_s, A) = \int_{\theta_s}^{\pi} \frac{(n-1)(\sin\phi)^{(n-2)}}{2^{n/2}\sqrt{\pi}\Gamma(\frac{n+1}{2})} \int_0^{\infty} s^{(n-1)} e^{-(s^2 + nA^2 - 2s\sqrt{n}A\cos\phi)/2} ds d\phi. \tag{19}$$

The computation in (18) and (19) becomes numerically unstable when n becomes large. The following asymptotic approximations should be used for large n:

$$\Omega_n(\theta_s) \approx \frac{\Gamma(\frac{n}{2}+1)(\sin\theta_s)^{(n-1)}}{n\Gamma(\frac{n+1}{2})\sqrt{(\pi)}\cos\theta_s} \approx \frac{(\sin\theta_s)^{(n-1)}}{\sqrt{2\pi n}\cos\theta_s}, \tag{20}$$

$$Q_n(\theta_s, A) \approx \frac{1}{\sqrt{n\pi}\sqrt{1+G^2(\theta_s,A)}\sin\theta_s} \times \frac{[G(\theta_s,A)\sin\theta_s e^{-(A^2 - AG(\theta_s,A)\cos\theta_s)/2}]^n}{AG(\theta_s,A)\sin^2\theta_s - \cos\theta_s}, \tag{21}$$

where $G(\theta_s, A) = (1/2)[A\cos\theta_s + \sqrt{A^2\cos^2\theta_s + 4}]$.

2) Random-Coding Bound and Error Exponent: Assuming a Gaussian input, the RCB for the AWGN channel has an explicit form [15, p. 340]. As in the BSC case, we need to compute the error exponent. For the rate R (in nats) in the range:

$$\frac{1}{2}\ln\left(\frac{1}{2} + \frac{A}{4} + \frac{1}{2}\sqrt{1 + \frac{A^2}{4}}\right) \leq R \leq \frac{1}{2}\ln(1+A), \tag{22}$$

where $A = E_s/N_o$. The error exponent is

$$E_r(R) = \frac{A}{4\beta}\left[\beta + 1 - (\beta-1)\sqrt{1+\frac{4\beta}{A(\beta-1)}}\right] + \frac{1}{2}\ln\left[\beta - \frac{A(\beta-1)}{2}\left(\sqrt{1+\frac{4\beta}{A(\beta-1)}}-1\right)\right], \tag{23}$$

where $\beta = e^{2R}$. When R is less than the left hand side of (22), the error exponent becomes

$$E_r(R) = 1 - \beta + \frac{A}{2} + \frac{1}{2}\ln\left(\beta - \frac{A}{2}\right) + \frac{1}{2}\ln\beta - R, \tag{24}$$

where

$$\beta = \frac{1}{2}\left(1 + \frac{A}{2} + \sqrt{1+\frac{A^2}{4}}\right). \tag{25}$$

These bounds can again be extended to erasure channels through the erasure-puncture equivalence. It is worth noting that the equivalence only exists for erasure channels. If the scaling coefficients are arbitrary real numbers, then the channels become periodic fading channels. In this case, the optimal codewords lie on the surface of an ellipsoid. Thus, the SPB literally becomes an ellipsoid-packing bound, which is extremely difficult to compute.

3) Numerical Analysis and Simulation Results: In parallel with the BSC case, we compute the random-coding exponents for two channels with erasure pattern “11111” and “00111”. We set the SNR to be 1.0 for the first pattern and 2.1748 for the second such that both channels have capacity 0.5 bit. The random-coding error exponents are plotted in Fig. 3 for various code rates. Again, the erasure channel is more favorable in terms of error exponent.

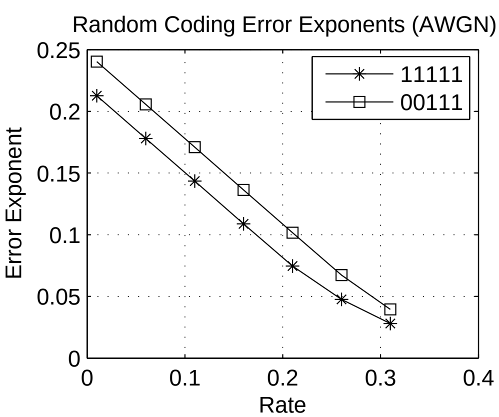


Fig. 3. Random-coding error exponents for an AWGN channel with capacity 0.5 bit. 11111–unerased channel, 00111–erasure channel.

In what follows, we compare the SPB and the RCB to the simulation results from three different codes: a trellis code, an LDPC code and a turbo code.

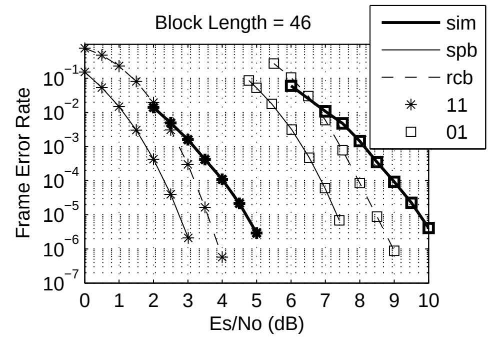
Example 1: A trellis code

The rate-1/3 trellis code [171 46 133] (in octal) proposed in [1] as a universal code for periodic erasure channels was simulated. This code was designed by minimizing the sum of the residual Euclidean distances and the sum of the SNRs over all the possible erasure patterns. A Gray-labeled 8PSK constellation was employed. The blocklength was 46 symbols. We denote the standard AWGN channel by CH1 and the one with erasure pattern “01” by CH2. The EMI for CH1 is (per complex symbol)

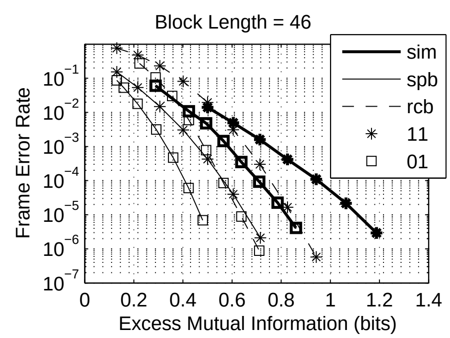
$$\mathrm{EMI}_1 = \log(1+\mathrm{SNR}) - R. \tag{26}$$

And the EMI for CH2 is

$$\mathrm{EMI}_2 = 1/2\log(1+\mathrm{SNR}) - R. \tag{27}$$



(a) FER vs SNR



(b) FER vs EMI

Fig. 4. Sphere-packing bound (spb), random-coding bound (rcb) and trellis code simulation results (sim) for AWGN channel with blocklength 46. 11–unerased channel, 01–erasure channel.

The bounds on frame error rate together with the simulation results are plotted against SNR in Fig. 4(a) and against EMI in Fig. 4(b). If only looking at the SNR plot and ignoring the SPB and RCB, one might assume that the code performance is much worse in the erasure channel because it needs a larger SNR. However this is not correct. The erasure channel inherently requires a larger SNR than the AWGN channel to provide the same capacity. Like the BSC case, the erasure channel requires less EMI according to both the bounds and the simulation results. Fig. 4 shows that the bounds for the two channels differ considerably at short blocklength, meaning that constant EMI is not possible across both channels. This gap becomes much smaller at longer blocklengths.

Example 2: An LDPC code

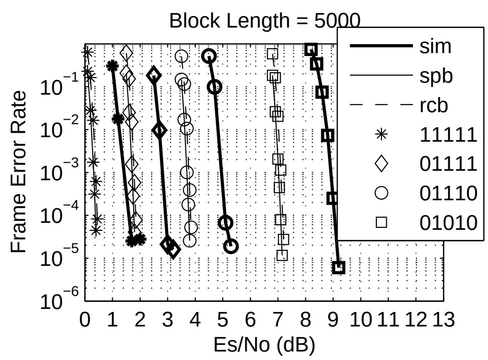
As shown in Fig. 5, the difference between the SPB and RCB becomes negligible as the blocklength becomes large. Also shown in the Fig. 5 are simulation results of a rate-$1/4$ blocklength-20000 binary LDPC code mapped to 5000

16QAM symbols. This code was optimized simply for the AWGN channel. Based on [18], this is not unreasonable as a universal LDPC code design choice. Its parity-check matrix was generated according to the following degree distribution,

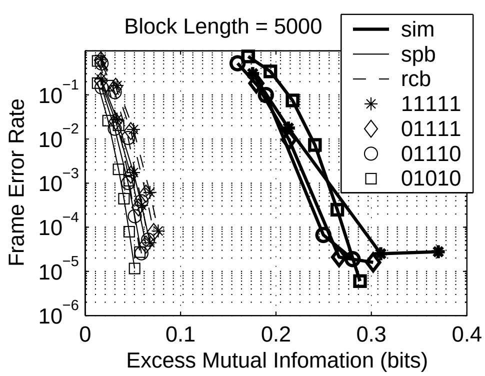
$$\begin{aligned}\lambda(x) &= 0.356x + 0.219x^2 + 0.175x^5 \\ &\quad + 0.057x^6 + 0.1x^{15} + 0.0926x^{16} \\ \rho(x) &= 0.5x^3 + 0.5x^4.\end{aligned} \tag{28}$$

The graph-conditioning methods in [19], [20] and [21] were used to lower the error floor. Systematic design of LDPC codes for periodic erasure channel through density evolution can be found in [22].

Four different channels were considered, including the standard AWGN channel and three erasure channels. The simulation performance of this code in terms of EMI improves as fraction of erasures increases until the most severe erasure channel, where the actual rate per unerased symbol grows so large for the 16QAM constellation that the effect of non-Gaussian input distribution becomes appreciable.



(a) FER vs SNR

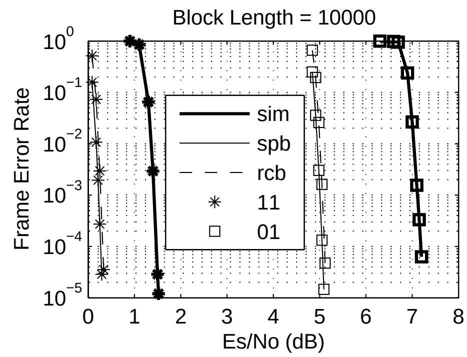


(b) FER vs EMI

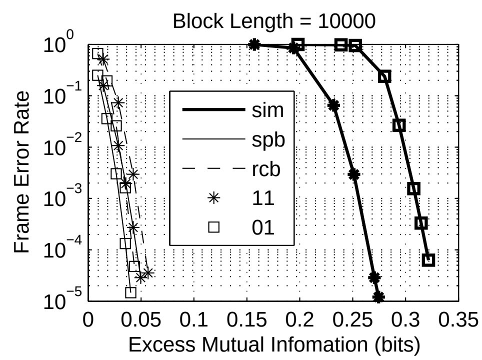
Fig. 5. Sphere-packing bound (spb), random-coding bound (rcb) and LDPC code simulation results (sim) for blocklength 5000 on an AWGN channel. 11111–unerased channel, 01110, 01110, 01010–three erasure channels.

Example 3: A turbo code

Performance of the rate-1/3 turbo code SC-5 proposed in [10], [23] is shown in Fig. 6. This code was found through computer search by optimizing the threshold of the constituent codes over both erasure patterns. The blocklength was 10000 8PSK symbols. The bounds still suggest that the erasure channel requires less EMI, but the simulation result shows the opposite. This is because the 8PSK constellation size is simply too small when half of the symbols are erased. Each 8PSK symbol in the erasure channel carries 2 bits of information. At this rate, the uniform distributed 8PSK achieves significantly lower capacity than the Gaussian distributed input. The same phenomenon was observed in the previous LDPC example, where the code performs worse in terms of EMI in the most erased channel. However, it was less severe in that case because of a larger constellation.



(a) FER vs SNR



(b) FER vs EMI

Fig. 6. Sphere-packing bound (spb), random-coding bound (rcb) and turbo code simulation results (sim) for blocklength 10000 on an AWGN channel. 11–unerased channel, 10–erasure channel.

VI. Conclusion

This paper begins by reviewing compound-channel coding theorem. A new proof of the theorem for linear Gaussian channels is presented. Like most coding theorems, only the asymptotic behavior of codes for a compound channel is stated in the theorem. We investigate the performance limits of universal codes with finite blocklengths by using the random-coding bound and the sphere-packing bound. It is shown that although the probability of error approaches zero uniformly under all the channels in the family, the difference in performance for different channels can be significant at finite blocklengths.

In particular, we observe that short-blocklength channels with erasures are more EMI-efficient than channels without erasures. This was also illustrated in Fig. 5 of [1], Fig. 6 of [3] and Fig. 3 of [9] where the erasure channels (singular channel for [3] and [9]) were more favorable channels. Once the input blocklength is on the order of 5000 bits, the bounds indicate that uniform behavior over all channels is a practical goal. However, we note that a sufficiently large constellation is required to avoid degradation on erasure channels. Furthermore, LDPC codes seem well-suited to provide this behavior. It is interesting to notice that in all three examples, regardless of blocklength the codes perform approximately 0.2 bit EMI away from the SPB. So with careful design, the short-blocklength trellis codes can be as universal as the long-blocklength turbo and LDPC codes.

VII. Acknowledgement

The authors would like to thank Chris Jones, Andres Vila Casado and Aditya Ramamoorthy for their help with the LDPC code simulations and Wen-Yen Weng for providing the turbo code simulation results.

VIII. Appendix: Proof of Theorem 1

We start with the case $m_r = m_t = m$. The result will be generalized to $m_r \neq m_t$ afterwards. Following [8], we divide the proof into two steps. We will show the theorem is true when the set of the channels is finite, then extend the result to uncountably many channels by bounding the error probability of an arbitrary channel.

Definition 1: [13] The set $A_\epsilon^{(n)}$ of jointly-typical sequences $(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)})$ with respect to the distribution $p(x,y)$ is the set of n-sequences with empirical differential entropies ϵ-close to the true entropies, i.e.,

$$\begin{aligned} A_\epsilon^{(n)} = \{ \; & (\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}) \in \mathcal{X}^{(n)} \times \mathcal{Y}^{(n)} : \\ & \left| -\frac{1}{n} \ln p(\boldsymbol{x}^{(n)}) - h(\boldsymbol{X}) \right| < \epsilon, \\ & \left| -\frac{1}{n} \ln p(\boldsymbol{y}^{(n)}) - h(\boldsymbol{Y}) \right| < \epsilon, \\ & \left| -\frac{1}{n} \ln p(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}) - h(\boldsymbol{X}, \boldsymbol{Y}) \right| < \epsilon, \} \end{aligned} \tag{29}$$

where

$$p(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i, \boldsymbol{y}_j). \tag{30}$$

For a given covariance matrix $\boldsymbol{S}$, the Gaussian distribution maximizes the entropy. Thus to prove Theorem 1, we only need to consider Gaussian input due to the Gaussianity of the noise.

Lemma 1: Let $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{z}$, where $\boldsymbol{H}$ is a deterministic matrix of dimension $m \times m$ and $\boldsymbol{x}$ and $\boldsymbol{z}$ are independent Gaussian random vectors of dimension $m \times 1$ with zero mean and covariance matrices $\boldsymbol{S}$ and $\boldsymbol{K}$, respectively. Consider i.i.d. length-n sequence $(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)})$ of dimension $2mn \times 1$ drawn according to $p(\boldsymbol{x}, \boldsymbol{y})$,

$$\Pr\left\{ (\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}) \notin A_\epsilon^{(n)} \right\} < 6 e^{-\frac{n\epsilon^2}{4m}}. \tag{31}$$

Proof: We bound the probability that the first inequality in (29) is violated.

$$\begin{aligned} & \Pr\left\{ -\ln p(\boldsymbol{x}^{(n)}) \leq nh(\boldsymbol{X}) - n\epsilon \right\} \\ & \overset{(a)}{=} \Pr\left\{ \frac{nm}{2} - n\epsilon - \frac{1}{2} \boldsymbol{x}^{(n)\mathsf{T}} \boldsymbol{K}^{(n)^{-1}} \boldsymbol{x}^{(n)} \geq 0 \right\} \\ & \overset{(b)}{\leq} \left(e^{t(\frac{m}{2} - \epsilon)} \mathsf{E} e^{-\frac{t}{2} \boldsymbol{x}^\mathsf{T} \boldsymbol{K}^{-1} \boldsymbol{x}} \right)^n \\ & = \left(e^{\frac{mt}{2} - \epsilon t - \frac{m}{2} \ln(1+t)} \right)^n \\ & \overset{(c)}{\leq} \left(e^{\frac{mt^2}{4} - \epsilon t} \right)^n \\ & \overset{(d)}{=} \left(e^{-\frac{\epsilon^2}{m}} \right)^n, \end{aligned} \tag{32}$$

where $\boldsymbol{K}^{(n)}$ in (a) is a block diagonal matrix of dimension $nm \times nm$ with $\boldsymbol{K}$ along the diagonal, and (b) is due to the Chernoff bound and the fact $\boldsymbol{x}^{(n)}$ is i.i.d. The variable t assumes positive real values and E stands for expectation. The inequality (c) is because $t - \ln(1+t) \leq t^2/2$ for $0 \leq t < 1$. We substitute $t = 2\epsilon/m < 1$ in (c) and arrive at (d).

The other direction goes similarly.

$$\begin{aligned} & \Pr\left\{ -\ln p(\boldsymbol{x}^{(n)}) \geq nh(\boldsymbol{X}) + n\epsilon \right\} \\ & = \Pr\left\{ -\frac{nm}{2} - n\epsilon + \frac{1}{2} \boldsymbol{x}^{(n)\mathsf{T}} \boldsymbol{K}^{(n)^{-1}} \boldsymbol{x}^{(n)} \geq 0 \right\} \\ & \leq \left(e^{-t(\frac{m}{2} + \epsilon)} \mathsf{E} e^{\frac{t}{2} \boldsymbol{x}^\mathsf{T} \boldsymbol{K}^{-1} \boldsymbol{x}} \right)^n \\ & = \left(e^{-\frac{mt}{2} - \epsilon t - \frac{m}{2} \ln(1-t)} \right)^n \\ & \overset{(a)}{\leq} \left(e^{\frac{mt^2}{2} - \epsilon t} \right)^n \\ & \overset{(b)}{=} \left(e^{-\frac{\epsilon^2}{2m}} \right)^n, \end{aligned} \tag{33}$$

where (a) is because $t + \ln(1-t) \geq -t^2$ for $0 \leq t \leq 0.5$. We obtain (d) by substituting $t = \epsilon/m$ in (a). Combine (32) and (33) we get

$$\Pr\left\{ \left| -\frac{1}{n} \ln p(\boldsymbol{x}^{(n)}) - h(\boldsymbol{X}) \right| \geq \epsilon \right\} < 2e^{-\frac{n\epsilon^2}{2m}} \leq 2e^{-\frac{n\epsilon^2}{4m}}. \tag{34}$$

Similarly, we can prove

$$\Pr\left\{ \left| -\frac{1}{n} \ln p(\boldsymbol{y}^{(n)}) - h(\boldsymbol{Y}) \right| \geq \epsilon \right\} < 2e^{-\frac{n\epsilon^2}{2m}} \leq 2e^{-\frac{n\epsilon^2}{4m}}, \tag{35}$$

$$\Pr\left\{ \left| -\frac{1}{n} \ln p(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}) - h(\boldsymbol{X}, \boldsymbol{Y}) \right| \geq \epsilon \right\} < 2e^{-\frac{n\epsilon^2}{4m}}. \tag{36}$$

Finally, we arrive at (31) by the union bound. ■

The following lemma gives bounds on the power of input and output of the channel.

Lemma 2: With the same setup as in Lemma 1, denote the total power of vector $\boldsymbol{x}^{(n)}$ by $P_x = \mathrm{Tr}(\boldsymbol{S})$. Then for any $\epsilon > 0$,

$$\Pr\left\{ \| \boldsymbol{x}^{(n)} \|^2 \geq n(P_x + \epsilon) \right\} \leq e^{-c_1 n}, \tag{37}$$

where $c_1 = \frac{1}{2}[\frac{\epsilon}{P_x} - \ln(1 + \frac{\epsilon}{P_x})]$. Furthermore, if the channel satisfies the conditions in Theorem 1, for any input complying with the power constraint $\| \boldsymbol{x}^{(n)} \|^2 \leq nP_x$, the output satisfies,

$$\Pr\left\{\| \boldsymbol{y}^{(n)} \|^2 \geq nP_y \mid \boldsymbol{x}^{(n)}\right\} \leq e^{-c_2 n}, \tag{38}$$

where $P_y = 2\alpha_0^2 P_x + 2m\alpha_2 + 2$ and $c_2 = \frac{1}{2}[\frac{1}{m\alpha_2} - \ln(1 + \frac{1}{m\alpha_1})]$. The constants α_0, α_1, α_2 are defined in Theorem 1, and m is defined in Lemma 1.

Proof: This is a direct result of the Chernoff bound. For details, see Lemma 5,8 of [8] and Lemma 5 of [24]. ■

In the following lemma, we prove the existence of universal codes for a finite set of channels.

Lemma 3: With the same setup up as Theorem 1, denote $L = |\mathcal{I}|$, the cardinality of the index set $\mathcal{I}$. Assume L to be finite, any rate $R < C(\mathcal{I})$ is achievable.

Proof: We will use the same idea as in [13], namely random codes and joint typical-set decoding. We generate i.i.d. codewords according to the distribution of $\boldsymbol{x}$. The codewords are denoted by $\boldsymbol{x}^{(n)}(w), w = 1, 2, ..., 2^{nR}$. The receiver looks for codewords that are jointly typical with the received vector given knowledge of the channel index at the receiver. If a single codeword is found, it is declared to be the transmitted codeword. Otherwise an error is declared. The receiver also declares an error if the chosen codeword does not satisfy the power constraint. Without loss of generality, assume that codeword 1 is sent.

Define the following events:

$$\mathrm{E}_0 = \left\{\| \boldsymbol{x}^{(n)}(1) \|^2 > P_x\right\}, \tag{39}$$

$$\mathrm{E}_i = \left\{(\boldsymbol{x}^{(n)}(i), \boldsymbol{y}^{(n)}) \in A_\epsilon^{(n)})\right\}. \tag{40}$$

Let P_s^n be the sum of error probabilities of individual channels when codeword 1 is sent.

$$\begin{aligned} P_s^n &= \sum_{i=1}^{L} P_i\left\{\mathcal{E}|W=1\right\} \\ &\stackrel{(a)}{\leq} \sum_{i=1}^{L}\left(P_i(\mathrm{E}_0) + P_i(\mathrm{E}_1^c) + \sum_{j=2}^{2^{nR}} P_i(\mathrm{E}_j)\right) \\ &\stackrel{(b)}{\leq} \sum_{i=1}^{L}\left(e^{-c_1 n} + 6e^{-\frac{n\epsilon^2}{4m}} + e^{-n(\ln 2)(I_i(x;y)-R-3\epsilon)}\right) \\ &\stackrel{(c)}{\leq} L\left(e^{-c_1 n} + 6e^{-\frac{n\epsilon^2}{4m}} + e^{-n\epsilon\ln 2}\right) \longrightarrow 0, \text{as } n \longrightarrow \infty. \end{aligned} \tag{41}$$

where (a) is due to the union bound, $P_i(\cdot)$ is the probability of an error event under the ith channel. The first two terms of (b) are due to Lemma 1 and 2; the third term (which also appears in (8)) upper-bounds the probability that a wrong codeword is jointly typical with the transmitted codeword. Its proof is discussed in [13]. Since $R < C(\mathcal{I}) \leq I_i(x,y), \forall i \in \mathcal{I}$, we can select small positive ϵ to satisfy $R < C(\mathcal{I}) - 4\epsilon$, which results in (c). Thus, for any individual channel in the compound channel, the error probability also approaches zero. By deleting the worst half of the codewords we obtain a code with low maximal probability of error. ■

To extend the result to arbitrary set $\mathcal{I}$, we need to form a dense finite subset and establish the relationship of the error probabilities between an arbitrary channel and its neighbor in the subset. The following lemma reveals the relationship.

Lemma 4: Let ($\boldsymbol{H}_1$,$\boldsymbol{K}_1$) and ($\boldsymbol{H}_2$,$\boldsymbol{K}_2$) be two channels satisfying the constraints in Theorem 1. Denote $\boldsymbol{x}^{(n)}$ and $\boldsymbol{y}^{(n)}$ to be the input and output n-sequence of m-dimensional vectors, respectively. Let $P_{\boldsymbol{H}_1,\boldsymbol{K}_1}\{\boldsymbol{y}^{(n)}|\boldsymbol{x}^{(n)}\}$ and $P_{\boldsymbol{H}_2,\boldsymbol{K}_2}\{\boldsymbol{y}^{(n)}|\boldsymbol{x}^{(n)}\}$ be the nm-variate probability densities for the output signal sequence $\boldsymbol{y}^{(n)}$ given $\boldsymbol{x}^{(n)}$, for the n-extension of the two channel ($\boldsymbol{H}_1$,$\boldsymbol{K}_1$) and ($\boldsymbol{H}_2$,$\boldsymbol{K}_2$), respectively. Then for those $\boldsymbol{x}^{(n)}$ satisfying $\|\boldsymbol{x}^{(n)}\|^2 \leq nP_x$ and $\boldsymbol{y}^{(n)}$ satisfying $\|\boldsymbol{y}^{(n)}\|^2 \leq nP_y$,

$$\frac{P_{\boldsymbol{H}_1,\boldsymbol{K}_1}\{\boldsymbol{y}^{(n)}, \boldsymbol{x}^{(n)}\}}{P_{\boldsymbol{H}_2,\boldsymbol{K}_2}\{\boldsymbol{y}^{(n)}, \boldsymbol{x}^{(n)}\}} \leq e^{n(c_3(\delta,\eta)+c_4)}, \tag{42}$$

where

$$\begin{aligned} c_3(\delta,\eta) = &\frac{1}{2\alpha_1^2}\left(P_y + \alpha_0^2 P_x + \alpha_0\sqrt{P_x P_y}\right)\delta \\ &+ \frac{1}{\alpha_1^2}\left(\alpha_0 P_x + \sqrt{P_x P_y}\right)\eta, \end{aligned} \tag{43}$$

$$c_4 = \frac{1}{2}\left(\ln\det(\boldsymbol{K}_2) - \ln\det(\boldsymbol{K}_1)\right), \tag{44}$$

The variables $\delta = \|\boldsymbol{K}_1 - \boldsymbol{K}_2\|$, $\eta = \|\boldsymbol{H}_1 - \boldsymbol{H}_2\|$, the numbers α_0 and α_1 are defined in Theorem 1.

Proof: See Lemma 7 of [8]. ■

Now we are ready to prove the theorem. Define the δ-neighborhood of the channel $(\boldsymbol{H}, \boldsymbol{K}) \in \mathcal{I}$ to be the set of channels $(\boldsymbol{H}', \boldsymbol{K}') \in \mathcal{I}$ satisfying $\|\boldsymbol{K} - \boldsymbol{K}'\| \leq \delta$ and $\|\boldsymbol{H} - \boldsymbol{H}'\| \leq \delta$. The conditions in Theorem 1 guarantee that the channel space is compact. We can select a finite subset $\mathcal{I}' \subset \mathcal{I}$ such that for an arbitrary channel in $\mathcal{I}$, in its neighborhood there exists at least one channel belonging to $\mathcal{I}'$. We denote $|\mathcal{I}'|$ to be L_δ to emphasize its relationship with δ.

For any $R < C(\mathcal{I}) \leq C(\mathcal{I}'))$, by Lemma 3, we can find a sequence of codes whose probability of error over $\mathcal{I}'$ vanishes as the blocklength grows. The code can be applied to the whole channel space in the following manner. If the channel is in $\mathcal{I}'$, then the receiver uses its own typical set decoder described in Lemma 3, otherwise the receiver uses the typical set decoder from its neighbor that is in $\mathcal{I}'$. The probability of error when the receiver uses its neighbor's decoder can be bounded. To be specific, let $(\boldsymbol{H}, \boldsymbol{K}) \in \mathcal{I}$ and $(\boldsymbol{H}', \boldsymbol{K}') \in \mathcal{I}'$ satisfy $\|\boldsymbol{K} - \boldsymbol{K}'\| \leq \delta$ and $\|\boldsymbol{H} - \boldsymbol{H}'\| \leq \delta$, by Lemma 4 we get,

$$P_{\boldsymbol{H},\boldsymbol{K}}\left\{\mathcal{E} \cap \mathrm{E}_0^c \cap \mathrm{E}_y^c\right\} \leq e^{n(c_3(\delta,\delta)+c_4)} P_{\boldsymbol{H}',\boldsymbol{K}'}\left\{\mathcal{E} \cap \mathrm{E}_0^c \cap \mathrm{E}_y^c\right\}, \tag{45}$$

where $\mathcal{E}$ is the event that the receiver makes an error, E_0 is defined in (39) and E_y is the event that $\boldsymbol{y}^{(n)}$ violates the power constraints.

Using Taylor expansion of $\det(\boldsymbol{K}')$ at $\boldsymbol{K}$ we can show that $\det(\boldsymbol{K}')/\det(\boldsymbol{K}) \leq 1 + P(\delta)/\alpha_1^m$, where $P(\delta)$ is a polynomial with $P(0) = 0$. Then, (45) becomes

$$\begin{aligned}
&P_{\boldsymbol{H},\boldsymbol{K}}\left\{\mathcal{E}\cap \mathrm{E}_0^c \cap \mathrm{E}_y^c\right\} \\
&\le \exp\left\{n\left[c_3(\delta,\delta)+\frac{1}{2}\ln\left(1+\frac{P(\delta)}{\alpha_1^p}\right)\right]\right\} P_{\boldsymbol{H}',\boldsymbol{K}'}\left\{\mathcal{E}\cap \mathrm{E}_0^c \cap \mathrm{E}_y^c\right\} \\
&\le \exp\left\{n\left[c_3(\delta,\delta)+\frac{1}{2}\ln\left(1+\frac{P(\delta)}{\alpha_1^p}\right)\right]\right\} P_{\boldsymbol{H}',\boldsymbol{K}'}\left\{\mathcal{E}\right\} \\
&\le \exp\left\{n\left[c_3(\delta,\delta)+\frac{1}{2}\ln\left(1+\frac{P(\delta)}{\alpha_1^p}\right)\right]\right\}. \\
&\quad L_\delta\left(\exp\{-c_1 n\}+6\exp\left\{-\frac{n\epsilon^2}{4p}\right\}+\exp\{-n\epsilon\ln 2\}\right)
\end{aligned} \tag{46}$$

The last inequality is due to Lemma 3. Since $c_3(\delta,\delta)$ and $P(\delta)$ approach zero when δ goes zero and L_δ is independent of n, we can select sufficiently small δ to ensure the overall exponent in the last expression to be negative. Then as n goes zero, $P_{\boldsymbol{H},\boldsymbol{K}}\left\{\mathcal{E}\cap \mathrm{E}_0^c \cap \mathrm{E}_y^c\right\}$ vanishes. Now we use the union bound,

$$P_{\boldsymbol{H},\boldsymbol{K}}\{\mathcal{E}\} \le P_{\boldsymbol{H},\boldsymbol{K}}\{\mathrm{E}_0\}+P_{\boldsymbol{H},\boldsymbol{K}}\{\mathrm{E}_y\}+P_{\boldsymbol{H},\boldsymbol{K}}\left\{\mathcal{E}\cap \mathrm{E}_0^c \cap \mathrm{E}_y^c\right\}. \tag{47}$$

According to Lemma 2, the first two terms vanish as the blocklength approaches infinity, thus the code works for any channel in $\mathcal{I}$. The converse is due to the fact that there exists a channel in $\mathcal{I}$ whose mutual information is less than $C+\epsilon$. So any rate greater than $C+\epsilon$ will not be achievable for that channel.

To extend it to $m_r \neq m_t$, let $m = \max(m_t, m_r)$ we can expand $\boldsymbol{H}$ to be of dimension $m \times m$ by padding zero columns or rows. Simultaneously we expand $\boldsymbol{y}$ and $\boldsymbol{z}$ by appending zeros when $m_r < m_t$ or expand $\boldsymbol{x}$ when $m_r > m_t$. If a code works for the expanded compound channel, it also works for the original compound channel.

Remark 1: With proper modification, the proof can be extended to the complex case where the input and noise are circularly symmetric Gaussian random vectors.

Remark 2: The above proof simplifies the error probability computation by using typical set decoding. This requires channel side information at the receiver. The original proof in [8] is stronger since it does not have this assumption. One could, however, argue that the rate loss due to channel estimation is negligible when the blocklength goes to infinity.

References

[1] R. D. Wesel, X. Liu, and W. Shi. Trellis Codes for Periodic Erasures. *IEEE Trans. on Comm.*, 28(1):938–947, Jun. 2000.

[2] M. Pursely and J. Skinner. Decoding strategies for turbo product codes in frequency-hop wireless communications. In *Proc. IEEE ICC–03*, May 2003.

[3] C. Köse and R. D. Wesel. Universal Space-Time Trellis Codes. *IEEE Trans. on Info. Th.*, 49(10):2717–2727, Oct. 2003.

[4] Q. Yan and R. S. Blum. Improved Space-time Convolutional Codes for Quasistatic Slow Fading Channels. *IEEE Trans. on Wireless Comm*, 1(4):563–571, 2002.

[5] A. Lapidoth and P. Narayan. Reliable Communication Under Channel Uncertainty. *IEEE Trans. on Info. Th.*, 44(6):2148–2177, Oct. 1998.

[6] J. Wolfowitz. *Coding Theorems of Infomation Theory*, volume 3rd ed. Springer-Verlag, Berlin, Germany, 1978.

[7] D. Blackwell, L. Breiman, and A. J. Thomasian. The Capacity of a Class of Channels. *Ann. Math. Statist.*, 30:1229–1241, 1959.

[8] W. L. Root and P. P. Varaiya. Capacity of Classes of Gaussian Channels. *SIAM J. Appl. Math.*, 16(6):1350–1393, Nov. 1968.

[9] A. Matache and R. D. Wesel. Universal Trellis Codes for Diagonally Layered Space-Time Systems. *IEEE Trans. on Signal Proc.*, 51(11):2773–2783, Nov. 2003.

[10] W. Weng, C. Köse, B. Xie, and R. D. Wesel. Universal Serially Concatenated Trellis Coded Modulation for Space-Time Channels. *IEEE Trans. on Comm.*, accepted, April 2007.

[11] S. Verdu. Spectral Efficiency in the Wideband Regime. *IEEE Trans. on Info. Th.*, 48(6):1319–1343, June. 2002.

[12] G. D. Forney. Modulation and Coding for Linear Gaussian Channels. *IEEE Trans. on Info. Th.*, 44(6):2384–2415, Oct. 1998.

[13] T. M. Cover and J. A. Thomas. *Elements of Infomation Theory*. Wiley, New York, 1991.

[14] C. E. Shannon. Probability of Error for Optimal Codes in a Gaussian Channel. *Bell Syst. Tech. J.*, 38:611–656, 1959.

[15] R. Gallager. *Information Theory and Reliabile Communication*. John Wiley & Sons, New York, 1968.

[16] S. J. MacMullan and O. M. Collins. A Comparison of Known Codes, Random Codes and the Best Codes. *IEEE Trans. on Info. Th.*, 44(7):3009–3022, Nov. 1998.

[17] S. Dolinar, D. Divsalar, and F. Pollara. Code Performance as a Function of Block Size. *TMO Progress Report 42-133*, May 1998.

[18] F. Peng, W. E. Ryan, and R. D. Wesel. Surrogate-Channel Design of Universal LDPC Codes. *IEEE Comm. Let.*, 10(6):480–482, June 2006.

[19] T. Tian, C. Jones, J. Villasenor, and R. Wesel. Construction of Irregular LDPC Codes with Low Error Floors. In *Proc. IEEE ICC-03*, Anchorage, May 2003.

[20] A. Ramamoorthy and R. Wesel. Construction of Short Block Length Irregular LDPC Codes. In *Proceedings of ICC*, Jun. 2004.

[21] T. Tian, C. Jones, J. D. Villasenor, and R. D. Wesel. Selective Avoidance of Cycles in Irregular LDPC Code Construction. *IEEE Trans. on Comm.*, 52(8):1242–1247, Aug. 2004.

[22] C. Jones, A. Matache, T. Tian, J. Villasenor, and R. Wesel. The Universality of LDPC Codes on Wireless Channels. In *Proceedings of Military Communications Conference*, Oct. 2003.

[23] C. Köse, W. Weng, and R. D. Wesel. Serially Concatenated Trellis Coded Modulation for the Compound Periodic Erasures Channel. In *Proceedings of ICC*, May 2003.

[24] W. L. Root and P. P. Varaiya. Capacity of Classes of Gaussian Channels Part 1: Discrete-time. *Tech. Memo M-208*, Electronics Research Lab., University of California, Berkeley, 1967.

PLACE PHOTO HERE

Jun Shi received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1997 and 2000, respectively, and the Ph.D. degree from the University of California, Los Angeles in 2005, all in electrical engineering.

From 2005 to 2007, he was with Intel Corp. Since 2007, he has been with Qualcomm Inc. His research interests are in the area of channel coding, MIMO systems and network information theory.

PLACE
PHOTO
HERE

Richard D. Wesel was born in Marietta, OH, in 1966. He received the Bachelors and Masters degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, both in 1989, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1996.

From 1989 to 1991, he was a Member of Technical Staff with AT&T Bell Laboratories, where his work resulted in two patents. Since July of 1996, he has been on the faculty of the University of California at Los Angeles (UCLA). In July 2002, he was promoted to Associate Professor, and in July 2006, he was promoted to Professor. Since July 2005, he has been serving as Vice Chair of Undergraduate Affairs for the UCLA Electrical Engineering Department. His research interests include parallel concatenated codes, low-density parity-check codes, spaceC-time codes, universal codes, and codes for asymmetric channels. Specific applications include indoor wireless communication systems, wireless local loops, asynchronous digital subscriber lines, and digital video broadcast.

Dr. Wesel has received the National Science Foundation CAREER Award, an Okawa Foundation Award, and the UCLA Henry Samueli School of Engineering and Applied Science Excellence in Teaching Award. He has been an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS since 1999 in the area of coding and communication theory.