Nested linear codes achieve Marton's inner bound for general broadcast channels

Arun Padakandla and S. Sandeep Pradhan
Dept. of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, USA
{arunpr,pradhanv}@umich.edu

***Abstract*— In several multi-terminal communication systems, it has been noted that the average performance of linear code ensemble is better than that of the standard unstructured code ensemble. However, it is well-known that linear code ensembles cannot achieve the point-to-point capacity of an arbitrary discrete memoryless channel. In this paper, we study nested linear codes and prove they achieve capacity of arbitrary discrete memoryless point to point channel with and without channel state information at the transmitter. Furthermore, we prove nested linear codes achieve Marton's inner bound, the largest known inner bound for the general discrete broadcast channel.**

I. Introduction

The broadcast channel was defined and the problem of characterizing it's capacity region was proposed by Cover [1]. Using a binning technique similar to that proposed by Gelfand and Pinsker [2] and superposition coding [1], an inner bound to the capacity region was derived by Marton [3]. A generalization [4, p. 391, Problem 10(c)], [5] of Marton's inner bound to accommodate common information is the largest known inner bound to the capacity region. Having established computability of Marton's inner bound, Gohari and Anantharam [6] have identified channels for which Marton's inner bound and the only known computable outer bound, derived by Nair and El Gamal [7], do not match. The problem of characterizing the capacity region of the general broadcast channel thus remains open.

The above inner bounds have been obtained by averaging error probability over the entire collection of code books, not restricting to any particular sub collection. While this is the prevalent technique in information theory, strictly larger rate regions have been achieved for certain problems by restricting averaging to structured code ensembles. Körner and Marton [8] have derived larger achievable rate regions for the problem of reconstructing modulo-2 sum of distributed binary sources, by restricting to linear codes. Philosof and Zamir [9] have proved nested linear codes achieve capacity of a particular binary additive multiple access channel with distributed side information. The above works study particular problem settings - reconstruction of binary sum, additive channels. Based on correlated binning of source/reconstruction vectors using algebraic codes, Krithivasan and Pradhan [10] have proposed a new framework for communicating information from distributed encoders observing correlated sources to a centralized decoder. Firstly, this framework is applicable to a large class of problems including distributed function computation, joint quantization of distributed sources etc. Secondly, rate regions derived using this technique subsumes Berger-Tung [11] rate region and strictly improves it for certain problems. These results indicate potential gains achievable through structured code ensembles for multi-terminal communication settings.

Motivated by these and other [12], [13] such results, we derive a rate region achievable by nested linear codes for the general broadcast channel. In particular, we prove nested linear codes achieve Marton's inner bound. As a simple corollary, this proves nested linear codes achieve capacity of point to point channels (PTP) and PTP with state known at transmitter (PTP-STx). The significance of our contribution is three fold. Firstly, this maybe viewed as a first step in deriving potentially improved inner bounds to the broadcast channel capacity region using structured code ensembles. Secondly, codes that achieve capacity of PTP, and binning technique proposed in [2] for PTP-STx have proved to be building blocks for designing codes for multi-terminal channels. Our study establishes nested linear code analogues for these building blocks for potential use in other multi-terminal settings. Thirdly, nested linear codes are of independent interest.

Linear and nested linear codes are subjects of considerable interest. Elias [14] proved linear codes achieve capacity of binary symmetric channels and a reformulation [15], [8] of this result proved linear codes can compress a source down to it's entropy. Wyner [15] proved linear codes achieve Slepian-Wolf [16] rate region for a pair of binary symmetric sources. The above studies were restricted to particular settings - symmetric sources and channels. Indeed, Ahlswede [17] proved linear codes are suboptimal for arbitrary PTP. Gallager [18] proved that linear codes followed by a nonlinear mapping achieve capacity of arbitrary PTP. A benign relaxation of linearity enables Krithivasan and Pradhan [10] prove nested linear codes achieve the Berger-Tung [11] rate distortion region for arbitrary sources and in particular the Shannon rate distortion function of a single source.

Our encoding and decoding techniques are similar to that proposed by Marton. The encoding proposed by Marton may be viewed as a generalization of that proposed in [2], where the

This work was supported by NSF grant CCF-0915619.

codeword of one user can be thought of channel state. Each receiver performs single user decoding. Essential aspects of the proof are thus captured in proving achievability of PTP-STx capacity. We therefore prove nested linear codes achieve capacity of PTP-STx in section III and in the interest of brevity restrict proof of achievability of Marton's inner bound to an outline in section V.

II. Definitions : PTP-STx, nested linear codes

We begin with some remarks on notation. Unless otherwise stated, $\mathbb{F}_q$ will denote the finite field with q elements. Logarithms and exponentials are to base q. Entropy is measured in units of q-bits, where 1 q-bit is $\log_2 q$ bits. For $M \in \mathbb{N}$, $[M] := \{1, 2, \cdots, M\}$. If $\mathcal{A}$ and $\mathcal{B}$ are finite sets and $f : \mathcal{A} \to \mathcal{B}$ is a map, the n-letter extension of f denoted $f^n : \mathcal{A}^n \to \mathcal{B}^n$ is defined $f(a^n) := (f(a_i) : i = 1, 2, \cdots, n)$.

A PTP-STx is a single input single output channel whose channel transition probabilities depend on a state variable S. A precise definition is provided below.

Definition 1: A point to point channel with state known at transmitter $\left(\mathbb{S}, p_S, \mathbb{X}, \mathbb{Y}, p_{Y|X,S}\right)$, abbreviated PTP-STx, consists of (i) a finite set $\mathbb{S}$ of states, (ii) a probability distribution p_S defined on $\mathbb{S}$, (iii) a finite input alphabet set $\mathbb{X}$, (iv) a finite output alphabet set $\mathbb{Y}$, and (v) a collection of probability mass functions $p_{Y|X,S}(\cdot|x, s)$ defined on $\mathbb{Y}$, one for each pair $(x, s) \in \mathbb{X} \times \mathbb{S}$. □

Throughout this paper, we assume all channels are (i)time invariant, (ii) memoryless, and (iii) used without feedback.

We assume knowledge of state sequence at transmitter non causally, and thus transmitted vector is a function of message and channel state. A precise definition follows.

Definition 2: An PTP-STx code $(n, \mathcal{M}, e, d)$ consists of (i) an index set $\mathbb{M} = [\mathcal{M}]$ of messages, (ii) an encoder map $e : \mathbb{S}^n \times \mathbb{M} \to \mathbb{X}^n$, and (iii) a decoder map $d : \mathbb{Y}^n \to \mathbb{M}$. □

Assuming a uniform distribution on the set of messages, we define the average error probability of a code as follows.

Definition 3: The error probability of the code $(n, \mathcal{M}, e, d)$ conditioned on the message $m \in \mathbb{M}$ is defined as

$$P_{\xi,m}(e, d) = \sum_{s^n \in \mathbb{S}^n} \sum_{\substack{y^n : d(y^n) \\ \neq m}} p_{s^n}(s^n) p_{Y^n|X^n,S^n}(y^n|e(m), s^n).$$

The average error probability of the code $(n, \mathcal{M}, f, g)$ is defined as $P_\xi(e, d) = \sum_{m=1}^{\mathcal{M}} \frac{1}{\mathcal{M}} P_{\xi,m}(e, d)$. □

A nested linear code over $\mathbb{F}_q$ consists of two linear codes, an inner and outer code. The inner code is contained within the outer code. Coset shifts of inner code within the outer code form a collection of bins. For technical reasons, we permit the outer code to be shifted by a constant bias.

Definition 4: A *nested linear code* $(n, k, l, g, \Delta g, b^n)$ over $\mathbb{F}_q$ consists of (i) a collection of vectors $\left\{a^k g + m^l \Delta g + b^n : a^k \in \mathbb{F}_q^k, m^l \in \mathbb{F}_q^l\right\}$ that defines an outer linear code, and (ii) the collection $\left\{a^k g + b^n : a^k \in \mathbb{F}_q^k\right\}$ that defines the inner linear code. □

The encoding technique and structure of our code is similar to that proposed in [2]. A code therein is based on an auxiliary code defined over an auxiliary alphabet set $\mathcal{U}$. This auxiliary code is a collection of bins, one for each message in the set of messages. Each bin is a collection of vectors in $\mathcal{U}^n$. The encoder uses the message to index a bin and chooses a codeword within this bin as a function of the channel state S^n. This codeword is mapped to a input vector on $\mathbb{X}^n$ using a map evaluated component wise. The auxiliary code and the map define a code for the PTP-STx. We restrict attention to those codes whose auxiliary code is a nested linear code over some finite field. With an abuse of notation, we define a code for PTP-STx as nested linear if it is based on an auxiliary nested linear code.

Definition 5: A PTP-STx code $\left(n, q^l, e, d\right)$ is nested linear if there exists (i) a nested linear code $(n, k, l, g, \Delta g, b^n)$ over $\mathbb{F}_q$ and (ii) a map $f : \mathbb{F}_q \times S \to \mathbb{X}$ such that $e(m^l, s^n) \in \left\{f^n\left(a^k g + m^l \Delta g + b^n\right) : a^k \in \mathbb{F}_q^k\right\}$. □

III. Nested linear codes achieve capacity of PTP-STx

We now proceed towards defining a rate region achievable using codes that are nested linear.

Definition 6: For a PTP-STx $\left(\mathbb{S}, p_S, \mathbb{X}, \mathbb{Y}, p_{Y|X,S}\right)$, let $\mathcal{P}\left(p_S, p_{Y|X,S}\right)$ be the collection of random vectors $(U, X, S, Y) : \Omega \to \mathcal{U} \times \mathbb{X} \times \mathbb{S} \times \mathbb{Y}$ such that (i) $\mathcal{U} = \mathbb{F}_q$ is a finite field, (ii) $P(S = s) = p_S(s)$, (iii) $P(Y = y|X = x, S = s) = p_{Y|X,S}(y|x, s)$, (iv) $U - (X, S) - Y$ is a Markov chain, and (v) there exists a map $f : \mathcal{U} \times \mathbb{S} \to \mathbb{X}$ such that $X = f(U, S)$.

For $Z = (U, X, S, Y) \in \mathcal{P}\left(p_S, p_{Y|X,S}\right)$, let $\alpha(Z) := [0, I(U; Y) - I(U; S)]$ and $\alpha\left(p_S, p_{Y|X,S}\right)$ be the closure of $\cup_{Z \in \mathcal{P}\left(p_S, p_{Y|X,S}\right)} \alpha(Z)$ □

We now prove achievability of $\alpha\left(p_S, p_{Y|X,S}\right)$ using codes that are nested linear.

Theorem 1: Let $\left(\mathbb{S}, p_S, \mathbb{X}, \mathbb{Y}, p_{Y|X,S}\right)$ be a PTP-STx. If $R \in \alpha\left(p_S, p_{Y|X,S}\right)$, then for every $\eta > 0$, $\epsilon > 0$ and sufficiently large n, there exists a PTP-STx code $(n, \mathcal{M}, e, d)$ that satisfies (i) the PTP-STx code $(n, \mathcal{M}, e, d)$ is nested linear, (ii) $\frac{\log \mathcal{M}}{n} \geq R - \eta$, and (iii) $P_\xi(e, d) \leq \epsilon$. □

Proof: We begin with an outline. As is typical in information theory, we prove existence by averaging the error probability over the ensemble of nested linear codes. Let $R \in \alpha(Z)$, where $Z = (U, X, S, Y) : \Omega \to \mathcal{U} \times \mathbb{X} \times \mathbb{S} \times \mathbb{Y}$ is a random vector, $X = f(U, S)$ and $\mathcal{U} = \mathbb{F}_q$. Pick generator matrices $G \in \mathbb{F}_q^{k \times n}$, $\Delta G \in \mathbb{F}_q^{l \times n}$ and bias vector $B^n \in \mathbb{F}_q^n$ mutually independently and uniformly from their respective range spaces. G and ΔG are generator matrices of inner code and it's shifts within the outer code respectively. The message set is $\mathbb{F}_q^l$, outer code is $\left\{U^n(a^k, m^l) := a^k G + m^l \Delta G + B^n : a^k \in \mathbb{F}_q^k, m^l \in \mathbb{F}_q^l\right\}$ and the m^l-th bin (coset) is $\left\{U^n(a^k, m^l) : a^k \in \mathbb{F}_q^k\right\}$. Having observed message M^l and channel state sequence S^n, the encoder looks for a vector in M^l-th bin that is jointly typical with S^n. If it finds one such vector, say $U^n(a^k, M^l)$, then $f^n(U^n(a^k, M^l), S^n)$ is transmitted. Else the encoder declares an error. The decoder observes received vector Y^n and identifies bins that contains a vector jointly typical with Y^n.

If there is exactly one such bin, the corresponding bin index is the decoded message. Else an error is declared.

We now characterize error events. The encoder declares error if no vector in M^l-th bin is jointly typical with S^n. Let $\theta_\delta(S^n) := \sum_{a^k \in \mathbb{F}_q^k} 1_{\{(U^n(a^k,M^l),S^n)\in T_\delta^n\}}$. An error occurs if $\theta_{\frac{\delta}{4}}(S^n) = 0$. We prove that if $\frac{k}{n} > 1 - H(U|S)$, then $P(\theta_{\frac{\delta}{4}}(S^n) = 0)$ is arbitrarily small for sufficiently large n.

We impose the above constraint on rate of the inner code and assume the encoder transmits a typical sequence with high probability. The decoder declares an error if either it finds (i) no vector in the outer code jointly typical with Y^n or (ii) vectors in multiple cosets jointly typical with Y^n. The probability of the former event falls exponentially with block length. This follows from Markov chain condition and conditional frequency typicality. We now characterize the latter event. Let $\xi\left(m^l, Y^n\right) = \sum_{a^k \in \mathbb{F}_q^k} 1_{\{(U^n(a^k,m^l),Y^n)\in T_\delta^n\}}$. An error occurs if $\xi\left(\tilde{m}^l, Y^n\right) \geq 1$ for any $\tilde{m}^l \neq M^l$. We prove that if $\frac{k+l}{n} < 1 - H(U|Y)$, then $P\left(\cup_{\tilde{m}^l \neq M^l}\left\{\xi\left(\tilde{m}^l, Y^n\right) \geq 1\right\}\right)$ falls exponentially with n. By choosing $\frac{k}{n}$ and $\frac{k+l}{n}$ arbitrarily close to $1 - H(U|S)$ and $1 - H(U|Y)$ respectively, $\frac{l}{n} \approx H(U|S) - H(U|Y) = I(U;Y) - I(U;S)$ can be achieved.

We now get to the details. Recall $G \in \mathbb{F}_q^{k\times n}$, $\Delta G \in \mathbb{F}_q^{l\times n}$, $B^n \in \mathbb{F}_q^n$ and $M^l \in \mathbb{F}_q^l$ are mutually independent random objects uniformly distributed on their respective range spaces. We begin with some preliminaries.

Remark 1: For $a^k \in \mathbb{F}_q^k$, $m^l \in \mathbb{F}_q^l$, $U^n(a^k, m^l)$ is uniformly distributed. Mutual independence and uniform distribution of random objects involved enable us verify this by a counting argument. For any $g \in \mathbb{F}_q^{k\times n}$, $\Delta g \in \mathbb{F}_q^{l\times n}$, there exists a unique $b^n \in \mathbb{F}_q^n$ such that $a^k g + m^l \Delta g + b^n = u^n$. The probability in question is therefore $\frac{q^{kn}q^{ln}}{q^{kn}q^{ln}q^n} = \frac{1}{q^n}$. It is easy to see $P\left(U^n(0^k, m^l) = u^n\right) = P\left(U^n(\tilde{a}^k, M^l) = \tilde{u}^n\right) = P\left(U^n(\hat{a}^k, m^l) = \hat{u}^n\right) = \frac{1}{q^n}$ for any $a^k, \tilde{a}^k, \hat{a}^k \in \mathbb{F}_q^k$ and $u^n, \tilde{u}^n, \hat{u}^n \in \mathbb{F}_q^n$. □

Remark 2: If $a^k, \tilde{a}^k \in \mathbb{F}_q^k$, $a^k \neq \tilde{a}^k$, $m^l \in \mathbb{F}_q^l$ and $u^n, \tilde{u}^n \in \mathbb{F}_q^n$, we claim $P\left(U^n(a^k, m^l) = u^n, U^n(\tilde{a}^k, m^l) = \tilde{u}^n\right) = \frac{1}{q^{2n}} = P\left(U^n(a^k, m^l) = u^n\right) P\left(U^n(\tilde{a}^k, m^l) = \tilde{u}^n\right)$. We first note $P\left(U^n(a^k, m^l) = u^n, U^n(\tilde{a}^k, m^l) = \tilde{u}^n\right) = P\left(U^n(a^k, m^l) = u^n, (\tilde{a}^k - a^k)G = (\tilde{u}^n - u^n)\right)$. We now employ a similar counting argument. Since $\tilde{a}^k \neq a^k$, $\exists j \ni \tilde{a}_j \neq a_j$. Given arbitrary $\Delta g \in \mathbb{F}_q^{k\times n}$ and rows $\underline{g}_1, \cdots, \underline{g}_{j-1}, \underline{g}_{j+1}, \cdots, \underline{g}_k$ of $g \in \mathbb{F}_q^{k\times n}$, choose $\underline{g}_j := (\tilde{a}_j - a_j)^{-1}(\tilde{u}^n - u^n - \sum_{i\neq j}(\tilde{a}_i - a_i)\underline{g}_i)$. This choice uniquely determines b^n (as in remark 1) and therefore the probability in question is $\frac{q^{(k-1)n}q^{ln}}{q^{kn}q^{ln}q^n} = \frac{1}{q^{2n}}$. Following a similar argument, one can prove $P\left(U^n(0^k, m^l) = u^n, U^n(0^k, \hat{m}^l) = \tilde{u}^n\right) = \frac{1}{q^{2n}} = P\left(U^n(0^k, m^l) = u^n\right) P\left(U^n(0^k, \hat{m}^l) = \tilde{u}^n\right)$. □

We now upper bound $P(\theta_{\frac{\delta}{4}}(S^n) = 0)$. We employ a second moment method similar to that used in [19]. By frequency typicality, $P(S^n \notin T_{\frac{\delta}{8}})$ falls exponentially with n, and therefore we restrict attention to $P(\theta_{\frac{\delta}{4}}(S^n) = 0, S^n \in T_{\frac{\delta}{8}})$.

$$\sum_{s^n \in T_{\frac{\delta}{8}}} P\begin{pmatrix} S^n = s^n, \\ \theta_{\frac{\delta}{4}}(s^n) = 0 \end{pmatrix} = \sum_{s^n \in T_{\frac{\delta}{8}}} P(S^n = s^n) P(\theta_{\frac{\delta}{4}}(s^n) = 0) \quad (1)$$

$$\leq \sum_{s^n \in T_{\frac{\delta}{8}}} P(S^n = s^n) P(|\theta_{\frac{\delta}{4}}(s^n) - \mathbb{E}\theta_{\frac{\delta}{4}}(s^n)| \geq \mathbb{E}\theta_{\frac{\delta}{4}}(s^n))$$

$$\leq \sum_{s^n \in T_{\frac{\delta}{8}}(S)} P(S^n = s^n) \frac{\mathrm{Var}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\}}{\left\{\mathbb{E}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\}\right\}^2}, \quad (2)$$

where (1) follows because $\theta_{\frac{\delta}{4}}(s^n)$ is a function of G, ΔG and B^n and these random objects are independent of S^n, and (2) from Cheybyshev inequality.

We now evaluate $\mathrm{Var}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\}$ and $\left\{\mathbb{E}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\}\right\}^2$.

$$\mathbb{E}\theta_{\frac{\delta}{4}}(s^n) = \sum_{\substack{u^n \in \\ T^n_{\frac{\delta}{4}}(U|s^n)}} \sum_{a^k \in \mathbb{F}_q^k} P\left(U^n(a^k, M^l) = u^n\right) = \frac{|T^n_{\frac{\delta}{4}}(U|s^n)|}{q^{n-k}},$$

where the last equality follows from Remark 1. We have

$$\mathbb{E}\theta^2_{\frac{\delta}{4}}(s^n) = \sum_{\substack{u^n, \tilde{u}^n \in \\ T^n_{\frac{\delta}{4}}(U|s^n)}} \sum_{\substack{a^k, \tilde{a}^k \\ \in \mathbb{F}_q^k}} P\begin{pmatrix} U^n(a^k, M^l) = u^n, \\ U^n(\tilde{a}^k, M^l) = \tilde{u}^n \end{pmatrix}$$

$$= \sum_{\substack{u^n \in \\ T^n_{\frac{\delta}{4}}(U|s^n)}} \sum_{a^k \in \mathbb{F}_q^k} P\left(U^n(a^k, M^l) = u^n\right)$$

$$+ \sum_{\substack{u^n, \tilde{u}^n \in \\ T^n_{\frac{\delta}{4}}(U|s^n)}} \sum_{\substack{a^k, \tilde{a}^k \in \\ \mathbb{F}_q^k, a^k \neq \tilde{a}^k}} P\left(U^n(a^k, M^l) = u^n, U^n(\tilde{a}^k, M^l) = \tilde{u}^n\right)$$

$$= \frac{q^k \left|T^n_{\frac{\delta}{4}}(U|s^n)\right|}{q^n} + \frac{\left|T^n_{\frac{\delta}{4}}(U|s^n)\right|^2 q^k\left(q^k - 1\right)}{q^{2n}},$$

where second term in (3) follows from Remark 2. Since $\mathrm{Var}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\} = \mathbb{E}\left\{\theta^2_{\frac{\delta}{4}}(s^n)\right\} - \mathbb{E}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\}^2$, we have

$$\mathrm{Var}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\} = \frac{q^k \left|T^n_{\frac{\delta}{4}}(U|s^n)\right|}{q^n}\left(1 - \frac{\left|T^n_{\frac{\delta}{4}}(U|s^n)\right|}{q^n}\right),$$

and for $s^n \in T_{\frac{\delta}{8}}$, by conditional frequency typicality

$$\frac{\mathrm{Var}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\}}{\mathbb{E}\left\{\theta_{\frac{\delta}{4}}(s^n)\right\}^2} \leq \frac{q^{n-k}}{|T^n_{\frac{\delta}{4}}(U|s^n)|} \leq q^{-n\left(\frac{k}{n} - (1 - H(U|S)) - \frac{3\delta}{8}\right)}. \quad (3)$$

Substituting (3) in (2), we obtain $P(\theta_{\frac{\delta}{4}}(S^n) = 0) \leq q^{-n\left(\frac{k}{n} - (1-H(U|S)) - \frac{3\delta}{8}\right)} + \frac{\epsilon}{4}$. By choosing $\delta > 0$ sufficiently small, $\frac{k}{n}$ can be made arbitrarily close to $1 - H(U|S)$ and probability of encoding error can be made arbitrarily small by choosing a sufficiently large block length.

It remains to upper bound probability of decoding error. We claim statistical independence of a bin, say $\left(U^n(a^k, m^l) : a^k \in \mathbb{F}_q^k\right)$ and a vector in a different bin, say

$U^n(\hat{a}^k, \hat{m}^l)$, $\hat{m}^l \neq m^l$. Let $u^n_{a^k} \in \mathbb{F}_q^n$ for each $a^k \in \mathbb{F}_q^k$, and $\hat{u}^n \in \mathbb{F}_q^n$. We claim

$$P\left(U^n(a^k, m^l) = u^n_{a^k} : a^k \in \mathbb{F}_q^k, U^n(\hat{a}^k, m^l) = \hat{u}^n\right)$$
$$= P\left(U^n(a^k, m^l) = u^n_{a^k} : a^k \in \mathbb{F}_q^k\right) P\left(U^n(\hat{a}^k, m^l) = \hat{u}^n\right).$$

If $(u^n_{a^k+\hat{a}^k} - u^n_{0^k}) \neq (u^n_{a^k} - u^n_{0^k}) + (u^n_{\hat{a}^k} - u^n_{0^k})$ for some pair a^k, $\hat{a}^k \in \mathbb{F}_q^k$, the LHS and first term of RHS are zero and equality holds. Else, we have

$$\begin{aligned}
&P(U^n(a^k, m^l) = u^n_{a^k} : a^k \in \mathbb{F}_q^k, U^n(\hat{a}^k, m^l) = \hat{u}^n) \\
&= P(a^kG = u^n_{a^k} - u^n_{0^k} : a^k \in \mathbb{F}_q^k, m^l\Delta G + B^n = u^n_{0^k}, \\
&\qquad \hat{m}^l\Delta G + B^n = \hat{u}^n - u^n_{\hat{a}^k}) \\
&= P(a^kG = u^n_{a^k} - u^n_{0^k} : a^k \in \mathbb{F}_q^k) P(m^l\Delta G + B^n = u^n_{0^k}, \\
&\qquad \hat{m}^l\Delta G + B^n = \hat{u}^n - u^n_{\hat{a}^k}) \quad (4) \\
&= P(a^kG = u^n_{a^k} - u^n_{0^k} : a^k \in \mathbb{F}_q^k) P(m^l\Delta G + B^n = u^n_{0^k}) \\
&\qquad P(\hat{m}^l\Delta G + B^n = \hat{u}^n - u^n_{\hat{a}^k}) \quad (5) \\
&= P(a^kG = u^n_{a^k} - u^n_{0^k} : a^k \in \mathbb{F}_q^k, m^l\Delta G + B^n = u^n_{0^k}) \\
&\qquad P(\hat{m}^l\Delta G + B^n = \hat{u}^n - u^n_{\hat{a}^k}) \quad (6) \\
&= P(U^n(a^k, m^l) = u^n_{a^k} : a^k \in \mathbb{F}_q^k) P(U^n(\hat{a}^k, m^l) = \hat{u}^n,)
\end{aligned}$$

where (4) and (6) follow from independence of ΔG, B^n and G (5) follows from Remark 2, and the last equality follows from Remark 1. We summarize the key aspect of the above observation in the following remark.

Remark 3: The transmitted vector, denoted $E\left(S^n, M^l\right)$, is a function of m^l-th bin. The above claim implies $E\left(S^n, m^l\right)$ and $U^n(\hat{a}^k, \hat{m}^l)$ are statistically independent if $m^l \neq \hat{m}^l$. □

We now upper bound $P\left(\cup_{\hat{m}^l \neq M^l} \left\{\xi\left(\hat{m}^l, Y^n\right) \geq 1\right\}\right)$. By the union bound, we have

$$\begin{aligned}
&P(\cup_{\hat{m}^l \neq M^l} \left\{\xi(\hat{m}^l, Y^n) \geq 1\right\}) \leq \sum_{m^l \neq M^l} P(\xi(\hat{m}^l, Y^n) \geq 1) \\
&\leq \sum_{m^l \neq M^l} \sum_{\hat{a}^k \in \mathbb{F}_q^k} P(U^n(\hat{a}^k, \hat{m}^l) \in T_\delta(U|Y^n)) \\
&\leq \sum_{\substack{m^l, \hat{m}^l \\ \in \mathbb{F}_q^n \\ m^l \neq \hat{m}^l}} \sum_{\hat{a}^k \in \mathbb{F}_q^k} \sum_{\substack{(x^n, s^n, y^n) \\ \in T_{\frac{\delta}{2}}}} \sum_{\substack{u^n \in \\ T_\delta(U|y^n)}} P(M^l = m^l, S^n = s^n, \\
&\qquad U^n(\hat{a}^k, \hat{m}^l) = u^n, E(s^n, m^l) = x^n, Y^n = y^n) + \frac{\epsilon}{4} \quad (7) \\
&= \sum_{\substack{m^l, \hat{m}^l \\ \in \mathbb{F}_q^n \\ m^l \neq \hat{m}^l}} \sum_{\hat{a}^k \in \mathbb{F}_q^k} \sum_{\substack{(x^n, s^n, y^n) \\ \in T_{\frac{\delta}{2}}}} \sum_{\substack{u^n \in \\ T_\delta(U|y^n)}} \frac{1}{q^n} P(M^l = m^l, \\
&\qquad S^n = s^n, E(s^n, m^l) = x^n, Y^n = y^n) + \frac{\epsilon}{4} \quad (8) \\
&\leq \sum_{\substack{m^l, \hat{m}^l \\ \in \mathbb{F}_q^n \\ m^l \neq \hat{m}^l}} \sum_{\hat{a}^k \in \mathbb{F}_q^k} \sum_{\substack{(x^n, s^n, y^n) \\ \in T_{\frac{\delta}{2}}}} \frac{q^{n(H(U|Y)+3\frac{\delta}{2})}}{q^n} P(M^l = m^l, \\
&\qquad S^n = s^n, E(s^n, m^l) = x^n, Y^n = y^n) + \frac{\epsilon}{4} \quad (9) \\
&\leq \frac{q^{k+l+n(H(U|Y)+\frac{3\delta}{2})}}{q^n} = q^{-n\left(1-H(U|Y)-\frac{3\delta}{2}-\frac{k+l}{n}\right)}. \quad (10)
\end{aligned}$$

Since $\frac{k}{n} > 1 - H(U|S)$, $(E^n(S^n, M^l)$ and S^n are jointly typical. By conditional frequency typicality $Y^n \in T_\delta(Y|E^n(S^n, M^l), S^n)$ with high probability and therefore (7) is true. To verify (8), we note

$$\begin{aligned}
&P(M^l = m^l, S^n = s^n, E(s^n, m^l) = x^n, U^n(\hat{a}^k, \hat{m}^l) = u^n, \\
&\qquad Y^n = y^n) = P(M^l = m^l, S^n = s^n, E(s^n, m^l) = x^n) \\
&\qquad P(U^n(\hat{a}^k, \hat{m}^l) = u^n) P(Y^n = y^n | M^l = m^l, S^n = s^n, \\
&\qquad E(s^n, m^l) = x^n) \quad (11) \\
&= \frac{1}{q^n} P(M^l = m^l, S^n = s^n, E(s^n, m^l) = x^n, Y^n = y^n)
\end{aligned}$$

where (11) follows from Remark 3. (9) follows from the bound on conditional typical set $T_\delta(U|y^n)$ when $y^n \in T_{\frac{\delta}{2}}$. From (10), we note that $\frac{k+l}{n}$ can be made arbitrarily close to $1 - H(U|Y)$ by choosing δ sufficiently small. Decoding error probability can be made arbitrarily small by choosing n sufficiently large. This completes the proof. ■

IV. Definitions: Broadcast channel

Definition 7: A two user discrete broadcast channel $\left(\mathbb{X}, \mathbb{Y}_1, \mathbb{Y}_2, p_{Y_1,Y_2|X}\right)$, abbreviated DBC consists of (i) a finite input alphabet set $\mathbb{X}$, (ii) two finite output alphabet sets $\mathbb{Y}_1$ and $\mathbb{Y}_2$, and (iii) a collection of probability mass functions $p_{Y_1,Y_2|X}(\cdot, \cdot|x)$ defined on $\mathbb{Y}_1 \times \mathbb{Y}_2$, one for each $x \in \mathbb{X}$. □

Definition 8: A DBC code $(n, \mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, e, d_1, d_2)$ consists of (i) three index sets $\mathbb{M}_i = [\mathcal{M}_i] : i = 0, 1, 2$, of messages, (ii) an encoder map $e : \mathbb{M}_0 \times \mathbb{M}_1 \times \mathbb{M}_2 \to \mathbb{X}^n$, and (iii) two decoder maps $d_i : \mathbb{Y}_i^n \to \mathbb{M}_0 \times \mathbb{M}_i$ for $i = 1, 2$. □

Definition 9: The error probability of DBC code $(n, \mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, e, d_1, d_2)$ conditioned on message triple $(m_0, m_1, m_2) \in \mathbb{M}_0 \times \mathbb{M}_1 \times \mathbb{M}_2$ is defined as

$$\zeta_{m_0,m_1,m_2} = \sum_{\substack{d_i(y_i^n) \\ \neq (m_0, m_i)}} p_{Y_1^n, Y_2^n|X^n}(y_1^n, y_2^n | e(m_0, m_1, m_2)).$$

The average error probability of code $(n, \mathcal{M}_1, \mathcal{M}_2, e, d_1, d_2)$ is defined as $\zeta(e, d_1, d_2) = \sum_{m_0=1}^{\mathcal{M}_0} \sum_{m_1=1}^{\mathcal{M}_1} \sum_{m_2=1}^{\mathcal{M}_2} \frac{1}{\mathcal{M}_0\mathcal{M}_1\mathcal{M}_2} \zeta_{m_0,m_1,m_2}$. □

Definition 10: A DBC code $(n, \mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, e, d_1, d_2)$ is nested linear if there exists (i) a triple $(n, k_i, l_i, g_i, \Delta g_i, d_i^n) : i = 0, 1, 2$ of nested linear codes over $\mathbb{F}_q$ and (ii) a map $f : \mathbb{F}_q \times \mathbb{F}_q \times \mathbb{F}_q \to \mathbb{X}$ such that $e(m_0, m_1, m_2) \in \left\{f^n(a_i^k g_i + m_i^{l_i} \Delta g_i + d_i^n : 1 \leq i \leq 3) : a_i^k \in \mathbb{F}_q^k : 0 \leq i \leq 2\right\}$. □

V. Nested linear codes achieve Marton's inner bound

We now proceed towards defining an achievable rate region using nested linear codes.

Definition 11: For DBC $(\mathbb{X}, \mathbb{Y}_1, \mathbb{Y}_2, p_{Y_1,Y_2|X})$, let $\mathcal{Q}(p_{Y_1,Y_2}|X)$ be the collection of random vectors $(W, U, V, X, Y_1, Y_2) : \Omega \to \mathbb{W} \times \mathcal{U} \times \mathbb{V} \times \mathbb{X} \times \mathbb{Y}_1 \times \mathbb{Y}_2$ that satisfy (i) $\mathbb{W} = \mathcal{U} = \mathbb{V} = \mathbb{F}_q$, (ii) $P(Y_1 = y_1, Y_2 = y_2|X = x) = p_{Y_1Y_2|X}(y_1, y_2|x)$, (iii) $(W, U, V) - X - (Y_1, Y_2)$ is a Markov chain, and (iv) there exists a map $f : \mathbb{W} \times \mathcal{U} \times \mathbb{V} \to \mathbb{X}$ such that $X = f(W, U, V)$.

For $T = (W, U, V, X, Y_1, Y_2) \in \mathcal{Q}(p_{Y_1,Y_2}|X)$, let

$$r(T) = \left\{ \begin{array}{l} (R_0, R_1, R_2) : R_0 \leq \min_{i=1,2} \{I(W;Y_i)\} \\ R_0 + R_1 \leq I(UW;Y_1) R_0 + R_2 \leq I(VW;Y_2) \\ R_0 + R_1 + R_2 \leq \min\{I(W;Y_1), I(W;Y_2)\} \\ +I(U;Y_1|W) + I(V;Y_2|W) - I(U;V|W) \end{array} \right\}$$

and $r(p_{Y_1,Y_2|X}) := \cup_{T \in \mathcal{Q}(p_{Y_1,Y_2|X})} r(T)$. □

the following theorem contains our main result.

Theorem 2: Let $(\mathbb{X}, \mathbb{Y}_1, \mathbb{Y}_2, p_{Y_1,Y_2|X})$ be a DBC and $(R_0, R_1, R_2) \in r(p_{Y_1,Y_2|X})$. For every $\eta > 0$, $\epsilon > 0$ and sufficiently large n, there exists a DBC code $(n, \mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, e, d_1, d_2)$ that satisfies (i) the DBC code is nested linear, (ii) $\frac{\log \mathcal{M}_i}{n} \geq R_i - \eta$ for $i = 0, 1, 2$, and (iii) $\zeta(e, d_1, d_2) \leq \epsilon$. □

We only outline a proof. Build three nested linear codes $(n, k_i, l_i, G_i, \Delta G_i, B_i^n) : i = 0, 1, 2$ over $\mathbb{F}_q (= \mathbb{W} = \mathcal{U} = \mathcal{V})$, denoted $\mathcal{C}_i$ by picking each of $G_i \in \mathbb{F}_q^{k_i \times n}$, $\Delta G_i \in \mathbb{F}_q^{l_i \times n}$ and $B_i \in \mathbb{F}_q^n$ independently and uniformly over their respective range spaces. We note $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2$ are statistically independent. We recall, inner code of $\mathcal{C}_i$ contains q^{k_i} vectors (in each bin) and $\mathcal{C}_i$ contains q^{l_i} such bins. $\mathbb{F}_q^{l_0}$ is the common message set and $\mathbb{F}_q^{l_i}$ is private message set of user i. Encoder observes message triple $(m_0^{l_0}, m_1^{l_1}, m_2^{l_2})$ and looks for a jointly typical triple $(w^n, u^n, v^n) \in \mathbb{W}^n \times \mathcal{U}^n \times \mathcal{V}^n$ such that w^n is a vector in $m_0^{l_0}$-th bin of $\mathcal{C}_0$, u^n in $m_1^{l_1}$-th bin of $\mathcal{C}_1$ and v^n in $m_2^{l_2}$-th bin of $\mathcal{C}_2$. If it finds such a triple, say (w^n, u^n, v^n), then $f^n(w^n, u^n, v^n)$ is transmitted. Else it declares an error. Having received Y_i^n, decoder of user i, identifies all bin pairs in $\mathcal{C}_0 \times \mathcal{C}_i$ that contains a pair of vectors jointly typical with Y_i^n. If there is exactly one such bin pair, it declares the indices of this bin pair as the decoded message. Else it declares an error.

We now outline an analysis of error probability. The encoder declares an error if no triple of vectors in the corresponding bins are jointly typical. If $\frac{k_0}{n} > 1 - H(W)$, $\frac{k_1}{n} > 1 - H(U|W)$ and $\frac{k_2}{n} > 1 - H(V|U, W)$, it can be shown by second moment method (proof of theorem 2), the probability of this event falls exponentially with n. Decoder of user i declares error if it finds either (ii) no pair in $\mathcal{C}_0 \times \mathcal{C}_i$, or (ii) pairs of vectors in multiple bin pairs, jointly typical with Y_i^n. By frequency typicality and Markov chain condition, probability of former event falls exponentially with n. If $\frac{k_0+l_0}{n} < \min\{1 - H(W|Y_1), 1 - H(W|Y_2)\}$, $\frac{k_1+l_1}{n} < 1 - H(U|W, Y_1)$ and $\frac{k_1+l_1}{n} < 1 - H(V|W, Y_2)$, the probability of the latter event falls exponentially with n. Therefore probability of decoding error can be made arbitrarily small by choosing n sufficiently large. We now compute the rates achieved. The rate of common message can be made arbitrarily close to $\frac{l_0}{n} \approx \min\{1 - H(W|Y_1), 1 - H(W|Y_2)\} - (1 - H(W)) = \min\{I(W;Y_1), I(W;Y_2)\}$. Private information can be sent to user 1 at rate arbitrarily close to $\frac{l_1}{n} \approx 1 - H(U|W, Y_1) - (1 - H(U|W)) = I(U;Y_1|W)$ and to user 2 at rate $\frac{l_2}{n} \approx 1 - H(V|W, Y_2) - (1 - H(V|U, W)) = I(V;Y_2|W) - I(U;V|W)$.

VI. Concluding remarks

(1) If $\mathcal{U} = \mathbb{X} = \mathbb{S} = \mathbb{F}_q$ and $f(u, s) = u + s$ modulo-q, then $\{f^n(U^n(a^k, m^l)) : a^k \in \mathbb{F}_q^k, m^l \in \mathbb{F}_q^l\}$ is a nested linear code. Else, an *embedding* into a sufficiently large finite field, analogous to that proposed by Krithivasan and Pradhan [10], results in a nested linear code over the resulting finite field.

(2) The outer code over $\mathbb{F}_q$ contains $q^{n(1-H(U|Y))}$ vectors which in general is larger than $q^{nI(U;Y)}$, yet decoding is successful. A linear code of rate R contains an exponentially smaller fraction of typical vectors. Indeed, a code of rate $R (> 1 - H(U))$ contains $q^{n(R-(1-H(U)))}$ typical sequences with high probability. Hence the outer code contains $q^{n((1-H(U|Y))-(1-H(U)))} = q^{nI(U;Y)}$ individually typical vectors. Therefore, one is able to stack more vectors in the outer code and yet decode successfully. Alternatively, one is forced to enlarge the bins to find a single typical sequence.

References

[1] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, pp. 2–14, Jan. 1972.

[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 19, no. 1, pp. 19–31, 1980.

[3] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 3, pp. 306 – 311, May 1979.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Budapest: Academic Press, 1986.

[5] S. I. Gel'fand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Peredachi Inf.*, vol. 16, pp. 24–34, 1980.

[6] A. A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," *submitted to IEEE Trans. on Inform. Th.*, April 2009, available at http://arxiv.org/abs/0904.4541.

[7] A. Nair, C.; El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 350–355, Jan. 2007.

[8] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (corresp.)," *IEEE Trans. Inform. Theory*, vol. 25, no. 2, pp. 219 – 221, Mar 1979.

[9] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," *IEEE Trans. Inform. Theory*, vol. 55, pp. 2442–2454, June 2009.

[10] D. Krithivasan and S. S. Pradhan, "Distributed source coding using abelian group codes," *submitted to IEEE Trans. on Inform. Th.*, August 2008, available at http://arxiv.org/abs/0808.2659.

[11] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of electrical engineering, Cornell University, Ithaca, NY, May 1978.

[12] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *submitted to IEEE Trans. on Inform. Th.*, August 2009, available at http://arxiv.org/abs/0908.2119.

[13] B. Nazer, M. Gastpar, S. A. Jafar, and S. Vishwanath, "Ergodic interference alignment," *IEEE Intl. Symp. on Inform. Th.*, June 2009.

[14] P. Elias, "Coding for noisy channels," *IRE Conv. Rec.*, pp. 37–46, 1955.

[15] A. Wyner, "Recent results in the shannon theory," *Information Theory, IEEE Transactions on*, vol. 20, no. 1, pp. 2 – 10, Jan 1974.

[16] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *Information Theory, IEEE Transactions on*, vol. 19, no. 4, pp. 471 – 480, July 1973.

[17] R. Ahlswede, "Group codes do not achieve shannon's channel capacity for general discrete channels," *The Annals of Mathematical Statistics*, vol. 42, no. 1, pp. 224–240, February 1971.

[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[19] A. El Gamal and E. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel (corresp.)," *Information Theory, IEEE Transactions on*, vol. 27, no. 1, pp. 120 – 122, Jan 1981.