Distributed Function Computation over Fields and Rings via Linear Compression of Sources

K. Vinodh[1], V. Lalitha[1], N. Prakash[1], P. Vijay Kumar[1] AND S. Sandeep Pradhan[2]

Abstract | The setting considered in this paper is one of distributed function computation. More specifically, there is a collection of N sources possessing correlated information and a destination that would like to acquire a specific linear combination of the N sources. We address both the case when the common alphabet of the sources is a finite field and the case when it is a finite, commutative principal ideal ring with identity. The goal is to minimize the total amount of information needed to be transmitted by the N sources while enabling reliable recovery at the destination of the linear combination sought. One means of achieving this goal is for each of the sources to compress all the information it possesses and transmit this to the receiver. The Slepian-Wolf theorem of information theory governs the minimum rate at which each source must transmit while enabling all data to be reliably recovered at the receiver. However, recovering all the data at the destination is often wasteful of resources since the destination is only interested in computing a specific linear combination. An alternative explored here is one in which each source is compressed using a common linear mapping and then transmitted to the destination which then proceeds to use linearity to directly recover the needed linear combination. The article is part review and presents in part, new results. The portion of the paper that deals with finite fields is previously known material, while that dealing with rings is mostly new.

Attempting to find the best linear map that will enable function computation forces us to consider the linear compression of source. While in the finite field case, it is known that a source can be linearly compressed down to its entropy, it turns out that the same does not hold in the case of rings. An explanation for this curious interplay between algebra and information theory is also provided in this paper.

I. INTRODUCTION

We consider a distributed function computation problem in which there are multiple spatially separated sources of data, and a destination, which is interested in computing a deterministic function of these distributed sources. The goal is to determine how to efficiently compress (encode) these sources such that the receiver can reliably compute the function, given the compressed data from all the sources. Such distributed function computation problems occur in many engineering systems such as sensor networks [1], distributed video coding applications [2] and wireless cellular communication. A typical example could be that many different sensors in an area observe correlated readings of a parameter of interest like temperature and a central node is interested in finding out just the average of all these observations.

[1]Department of ECE, Indian Institute of Science, Bangalore - 560012, India

[2]Department of EECS, University of Michigan, Ann Arbor, MI 48109, USA

E-mail: [1]{kvinodh, lalitha, prakashn, vijay}@ece.iisc.ernet.in ,

[2]pradhanv@eecs.umich.edu

One simple strategy for compression is one in which every source encodes its own data into binary digits (bits) and transmits these bits to the receiver. The receiver then as a first step decompresses these bits to recover the data of all the sources and then computes the function of interest from this data. This strategy thus incurs a data rate needed to communicate all the source data, even though the receiver is only interested in a function of these sources. Such a strategy could sometimes be wasteful of resources. In many cases, it is possible to design compression schemes which allow the receiver to directly recover the function of interest, instead of having to recover the individual sources [3]. For example, if the function of interest is a linear combination of the various sources, it turns out that linear maps when used for compression, will allow the receiver to directly recover the linear combination of interest [4] [5]. We will explain this using an example.

Let (X,Y) be a pair of binary sources located in different geographical locations. When we say a source X, we mean the following. There is a sequence of discrete random variables $X_1,X_2,\dots$ such that $\forall i>1,\ X_i$ is independent of $X_1,\dots X_{i-1}$. For the case when there is pair of sources (X,Y), we will mean that there is a sequence of discrete random variables$(X_1,Y_1),(X_2,Y_2)\dots$ such that $\forall i>1,\ (X_i,Y_i)$ is independent of $(X_1,Y_1),\dots(X_{i-1},Y_{i-1})$. The random variables (X_i,Y_i) are identically distributed according to a distribution P_{XY}. Such a source is called a discrete memoryless source (DMS). The output of the source X and Y is a realization of these random variables. A binary source is one which takes values from the finite field $\mathbb{F}_2$. Let the destination be interested in computing the modulo two sum of the sources, $Z_i = X_i + Y_i$ mod 2. The encoder corresponding to each source operates on blocks of $n-$length output of the source and uses a matrix to carry out the compression as follows. Every $n-$length output of the source $\mathbf{x} = (x_1,\dots x_n)$ is multiplied by a $k \times n$ matrix A over the field $\mathbb{F}_2$ to obtain $A\mathbf{x}$. The encoder corresponding to Y also does a similar operation. The resulting $k-$length vectors $A\mathbf{x}$ and $A\mathbf{y}$ are presented to the receiver. If $k = \alpha n, 0 \leq \alpha \leq 1$ then α may be viewed as a crude measure of the amount of compression taking place at each encoder. The receiver will compute $A\mathbf{x}+A\mathbf{y}$ mod 2 to obtain $A\mathbf{z}$ and then finds an estimate of $Z_1,\dots,Z_n$ from $A\mathbf{z}$. If the matrix A is chosen properly then the estimate of $\{Z_i\}$ will be reliable. Thus we would have computed the sum of the sources without actually recovering any of the individual sources. In this paper we will consider such schemes in more detail and analyze the maximum amount of compression that can take place in the encoders.

In Section II, we elaborate on the system model that will be used for function computation, where the function is assumed to be a linear combination of the various sources. A brief introduction to various information theoretic concepts relevant to this paper, such as notions of entropy, typical sets and also the problem of source compression, distributed source compression, is presented in Appendix A. For more details, the reader is referred to [6]. In Section III we discuss a system model for point-point source compression under linear encoders. This system model will be derived from the system model for the function computation problem. The point-point source compression problem for the case when the alphabet of the source is a finite field $\mathbb{F}_q$, where q is a power of prime is discussed in Section V. The cases when the

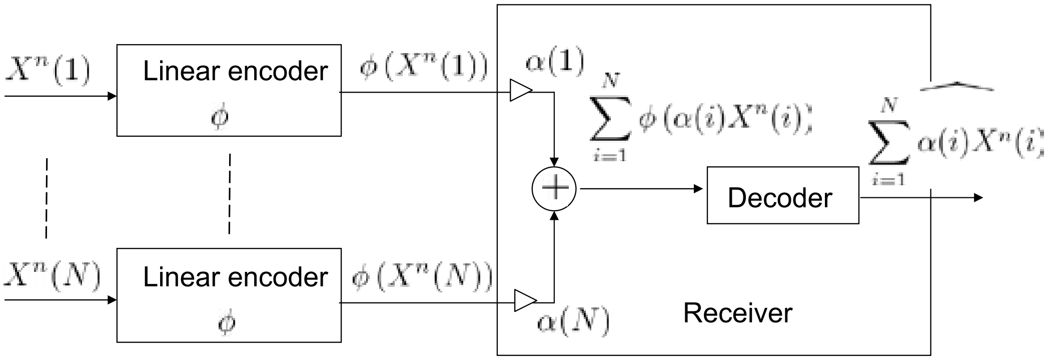

Fig. 1. System model for function computation

alphabets are the ring $\mathbb{Z}_{p^r}$ (the ring of integers modulo p^r, p a prime and $r \geq 1$), chain rings and principal ideal rings are respectively discussed in Sections VI, VII and VIII.

II. System Model for Function Computation

Consider the distributed source compression problem (See Fig 1) involving N correlated but memoryless sources and a receiver that is only interested in reliably computing a function of the sources. Let $X(1), X(2), \ldots, X(N)$ denote the random variables (r.v.) corresponding to the N sources. We assume that all the r.v. have the same finite alphabet $\mathcal{A}$. The alphabet $\mathcal{A}$ will be assumed to be a finite field or a finite commutative principal ideal ring with identity, which we will simply denote as PIR. Let $X^n(i)$ denote the random variables corresponding to an $n-$length output sequence of the i^{th} source and let $\mathbf{x}(i)$ denote a realization of $X^n(i)$, $i = 1, \ldots, N$. The sequence of r.v. $(X^n(1), \ldots, X^n(N))$ is assumed to be independent and identically distributed (i.i.d.) $\sim P_{X(1)X(2)\ldots X(N)}$. The receiver is interested in computing the linear combination X of the source outputs, where $X = \sum_{i=1}^{N} \alpha(i)X(i),\ \alpha(i) \in \mathcal{A}$. In this article, we will denote the realization of random variable X^n by bold face $\mathbf{x}$.

Encoder : The encoding is carried out using an $\mathcal{A}-$ module homomorphism, $\phi^{(n)}$,

$$\phi^{(n)} : \mathcal{A}^n \longrightarrow \mathcal{M}\ , \tag{1}$$

where the co-domain $\mathcal{M}$ is an $\mathcal{A}-$module. If $\mathcal{M} = \mathcal{A}^k$, then $\phi^{(n)}$ will be replaced by the matrix $A^{(n)}$ corresponding to the homomorphism, where $A^{(n)} \in M_{k\times n}(\mathcal{A})$, the set of $k \times n$ matrices over $\mathcal{A}$. In this case, the output of the i^{th} encoder is $A^{(n)}X^n(i)$, left multiplication by the matrix $A^{(n)}$. Note that we use the same $\mathcal{A}-$ linear map $\phi^{(n)}$ for all encoders. For notational convenience, we shall use ϕ, A in place of $\phi^{(n)}$, $A^{(n)}$ when there is no ambiguity.

Receiver : Since the function of interest corresponds to a linear combination of the sources, the first step taken by the receiver is to take linear combination of the outputs of the encoders to obtain

$$\begin{aligned} \sum_{i=1}^{N} \alpha(i)\phi(X^n(i)) &\overset{(a)}{=} \phi\left(\sum_{i=1}^{N} \alpha(i)X^n(i)\right) \\ &= \phi(X^n) \ , \end{aligned} \tag{2}$$

where (a) follows since ϕ is $\mathcal{A}-$linear. Thus the input to the decoder is $\phi(X^n)$. Given $\phi(X^n)$, the decoder is expected to output a reliable estimate, $\hat{X}^n$, of X^n. An error occurs if $\hat{X}^n \neq X^n$. We will use $P_e^{(n)}$ to denote the probability of error, averaged over all source symbols i.e., $P_e^{(n)} = P(\hat{X}^n \neq X^n)$, the probability of the event $\hat{X}^n \neq X^n$.

Rate: The rate of any encoder, in bits per symbol, is given by

$$R^{(n)} = \frac{\log_2 |\mathrm{Im}(\phi^{(n)})|}{n} \ , \tag{3}$$

where $|\mathrm{Im}(\phi^{(n)})|$ denotes the cardinality of the image of $\phi^{(n)}$. Note that the rate $R^{(n)}$ per encoder translates to a sum rate of $NR^{(n)}$ for the whole system. The objective of the system is to allow for reliable recovery of X^n, with as small sum rate as possible. This notion is quantified below.

Achievability : A rate R, per encoder, is said to be achievable, if there exists a sequence of $\mathcal{A}-$linear maps $\{\phi^{(n)}\}$ such that

$$\lim_{n\to\infty} R^{(n)} = R \quad \text{and} \quad \lim_{n\to\infty} P_e^{(n)} = 0. \tag{4}$$

We define the term *achievable rate region* to be the closure of the set of all achievable rates.

III. An Equivalent System Model

Since we assume that the first step in the receiver is to form the linear combination $\phi(X^n)$ from the various encoder outputs, as far as the decoder is concerned, one can consider an equivalent system model (see Fig. 2) in which a source directly outputs the linear combinations X^n i.i.d. $\sim P_X$ where

$$P_X(x) = \sum_{(x(1),\ldots,x(N)) \ : \ \sum_{i=1}^{N} \alpha(i)x(i)=x} P_{X(1)\ldots X(N)}(x(1),\ldots,x(N)) \ , \tag{5}$$

and is encoded by the map ϕ.

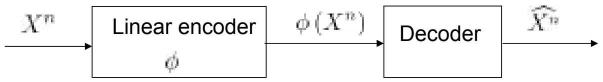

Fig. 2. Equivalent single source system model

It is clear that for a fixed encoder ϕ, the probability of error in recovering X^n in the original system is same as the probability of error in recovering X^n in the equivalent system. Hence from now on, we shall only consider this equivalent system for probability of error analysis. Note that an achievable rate R in the equivalent system translates to a sum rate NR in the original system.

IV. Summary of the results

The goal in the rest of the document is to characterize achievable rate regions for this equivalent system for various choices of the alphabet $\mathcal{A}$.

1) We will start with the simplest case when the alphabet $\mathcal{A}$ is the finite field $\mathbb{F}_q$, where $q = p^r$, for some prime p and $r \in \mathbb{N}^+$. We will review the well known result [7], [8] which states that a source X whose alphabet is a finite field can be compressed down to its entropy, $H(X)$, using a linear encoder. It should be noted that this is also the optimal compression rate for the source X using any encoder, not necessarily linear.
2) The second alphabet that we consider is the ring $\mathbb{Z}_{p^r}$. Surprisingly, unlike in the case of fields, we will see here that compression down to entropy is not always possible using linear encoders. For example, consider the case when $p = 2$ and $r = 2$ i.e., the ring $\mathbb{Z}_4$. Then, it turns out that the achievable rate region $R = \max\{H(X), 2H(X|[X]_1\}$ where $[X]_1 = X \bmod 2$. We will first review the achievability part of the result for the ring $\mathbb{Z}_{p^r}$ [5]. We then prove a converse [9] where we show that the presence of non-trivial ideals in $\mathbb{Z}_{p^r}$ is the reason for the suboptimality of compression under linear encoders.
3) Finally, we consider the case when the alphabet is any PIR. Using the decomposition theorem for PIRs, which states that any PIR is isomorphic to a direct product of chain rings, we will see that characterizing rate regions for chain rings and their direct products amount to characterizing rate regions for PIRs. Chain rings are rings in which all the ideals form a chain by set inclusion. It turns out that the characterization of the rate region for chain rings can be carried out along similar lines as for the ring $\mathbb{Z}_{p^r}$. The similarity comes in due to the fact that both rings allow for component-wise expansion of elements in them. Whereas, in the case of the ring $\mathbb{Z}_{p^r}$, every element has a $p-$ary expansion, in chain rings, every element has a $\theta-$ary expansion, where θ is a generator of the maximal ideal of the chain ring. The characterization of the rate region for a general finite ring remains unsolved.

The characterization of the rate region in all the cases involves two steps. In the first step, we show the achievability of a region $\mathcal{R}$, i.e., for every $R \in \mathcal{R}$, we show the existence of a sequence of $\mathcal{A}-$linear maps $\{A^n\}$ satisfying (4) under a *typical set decoder*, which is explained below. A typical set corresponding to a source X, denoted by $A_\varepsilon^{(n)}(X)$, roughly speaking, is the set of all $n-$length realizations of the source whose empirical frequencies are close to its symbol probabilities. For example, if X is a binary source with $P(0) = \frac{1}{4}$ and $P(1) = \frac{3}{4}$, then the typical set is the set of all $n-$length binary sequences whose ratio of the number of zeros to n is close to $\frac{1}{4}$. This set is called the typical set because we expect the output of the source to be only such sequences. A typical set decoder (in this paper) is a decoder which upon receiving a $k-$length vector searches for a unique typical sequence among the set of all source sequences which when multiplied by the encoder matrix results in this $k-$length vector.

If a such a sequence exists it is declared to be the source sequence, else an error is declared.

In the second step, we will prove a converse to this achievability, meaning that no rate point outside this region is achievable. The converse will be independent of any particular decoding method.

V. Linear compression over finite fields

In this section, we consider the case when the alphabet $\mathcal{A}$ is the finite field $\mathbb{F}_q$. The achievable rate region is characterized by the following theorem.

Theorem 1. *For the source X drawn i.i.d. $\sim P_X$ and whose alphabet is $\mathbb{F}_q$, the achievable rate region under linear encoding is given by*

$$R \geq H(X) \ . \tag{6}$$

Proof: The achievability part is shown by a random coding argument by averaging over the set of all linear encoders of the form

$$A^{(n)} : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^k \ , \tag{7}$$

where $A^{(n)}$ is assumed to be a realization of the random matrix $\mathbf{A}^{(n)}$, distributed uniformly on the ensemble $M_{k\times n}(\mathbb{F}_q)$. In this process, we calculate the probability of error $P_e^{(n)}$ averaged over the source symbols and also over all realizations of the random matrix $\mathbf{A}^{(n)}$. We show that if we assume k as a function of n, say $k(n)$, such that

$$\frac{k(n)}{n}\log(q) > H(X) \ , \tag{8}$$

then $P_e^{(n)} \to 0$. This will prove the existence of a particular sequence of matrix encoders $\{A^{(n)}\}$ in the ensemble of all sequences of encoders which achieve the rate $H(X)$.

The decoder will be assumed to be a typical set decoder [1]. Assuming that the transmitted sequence is $\mathbf{x}$, the receiver will make an error when any one of the following events occur:

$$E_1 \ : \ \mathbf{x} \notin A_\varepsilon^n(X) \tag{9}$$

$$E_2 \ : \ \exists \mathbf{y} \in A_\varepsilon^n(X) \text{ such that } \mathbf{y} \neq \mathbf{x} \text{ and } \mathbf{Ay} = \mathbf{Ax} \tag{10}$$

The average probability of error $P_e^{(n)}$ is then given by

$$P_e^{(n)} = P(E_1 \cup E_2) \tag{11}$$

$$\leq P(E_1) + P(E_2) \ . \tag{12}$$

[1] For definitions and other facts regarding typical sets used in this proof, please refer Appendix A

For a fixed ε, the probability that an element is not typical can be made arbitrarily small by choosing a large n, i.e, $P(E_1) \leq \delta_n$ with $\delta_n \overset{n\to\infty}{\longrightarrow} 0$. Now,

$$\begin{aligned} P(E_2) &= \sum_{\mathbf{x}\in A_\varepsilon^n(X)} P(\mathbf{x}) \sum_{\substack{\mathbf{y}\neq\mathbf{x} \\ \mathbf{y}\in A_\varepsilon^n(X)}} P(\mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{x}) \qquad (13) \\ &:= \sum_{\mathbf{x}\in A_\varepsilon^n(X)} P(\mathbf{x})\Delta(\mathbf{x}) \ . \qquad (14) \end{aligned}$$

Noting that $\mathbf{A}$ is uniform over $M_{k\times n}(\mathbb{F}_q)$, we can write $\Delta(\mathbf{x})$ as

$$\Delta(\mathbf{x}) = q^{-nk} \sum_{\substack{\mathbf{y}:\mathbf{y}-\mathbf{x}\neq 0 \\ \mathbf{y}\in A_\varepsilon^n(X)}} \sum_{\substack{A\in M_{k\times n}(\mathbb{F}_q): \\ A(\mathbf{y}-\mathbf{x})=0}} 1 \ . \tag{15}$$

We shall now compute $\sum_{\substack{A\in M_{k\times n}(\mathbb{F}_q): \\ A(\mathbf{y}-\mathbf{x})=0}} 1$. Let $A = [\mathbf{t_1} \dots \mathbf{t_n}]$ where $\mathbf{t_i}$ are the columns of the matrix A. Let $\mathbf{z} = \mathbf{y} - \mathbf{x}$ and $\mathbf{z} = [z_1 \dots z_n]^t$. Since $\mathbf{z} = \mathbf{y} - \mathbf{x} \neq 0$, there exists a $z_j \in \mathbb{F}_q^*$. Thus the condition $A\mathbf{z} = 0$ would demand that

$$\mathbf{t_j} = z_j^{-1} \sum_{i\neq j} \mathbf{t_i} z_i \ . \tag{16}$$

Hence the number of ways to choose A such that $A\mathbf{z} = 0$ is same as the number of choices of the columns of A, excluding the column $\mathbf{t}_i$. Thus we get

$$\sum_{\substack{A\in M_{k\times n}(\mathbb{F}_q): \\ A(\mathbf{y}-\mathbf{x})=0}} 1 = q^{k(n-1)} \ . \tag{17}$$

Using this in (15), we get

$$\begin{aligned} \Delta(\mathbf{x}) &= q^{-nk} \sum_{\substack{\mathbf{y}:\mathbf{y}-\mathbf{x}\neq 0 \\ \mathbf{y}\in A_\varepsilon^n(X)}} q^{k(n-1)} \qquad (18) \\ &\leq q^{-k} |A_\varepsilon^n(X)| \ . \qquad (19) \end{aligned}$$

Substituting the above upper bound on $\Delta(\mathbf{x})$ in (14) we get,

$$\begin{aligned} P(E_2) &\leq q^{-k} |A_\varepsilon^n(X)| \sum_{\mathbf{x}\in A_\varepsilon^n(X)} P(\mathbf{x}) \qquad (20) \\ &\leq q^{-k} |A_\varepsilon^n(X)| \ . \qquad (21) \end{aligned}$$

Now, since we know that $|A_\varepsilon^n(X)| \leq 2^{n(1+\varepsilon)H(X)}$, we get

$$P(E_2) \leq q^{-k} 2^{(1+\varepsilon)nH(X)} \ . \tag{22}$$

Hence, substituting the upper bound on $P(E_1)$ and $P(E_2)$ in (12) we get,

$$P_e^{(n)} \leq \delta_n + 2^{-n\left\{\frac{k}{n}\log q - (1+\varepsilon)H(X)\right\}} \ . \tag{23}$$

Hence, if $\frac{k}{n}\log q > H(X) + \varepsilon H(X)$ then $P_e^{(n)} \to 0$ as $n \to \infty$. Since, $R^{(n)} = \frac{\log|\mathrm{Im}(A^{(n)})|}{n} \leq \frac{k}{n}\log q$, the achievability part of the theorem follows.

The converse follows from standard information theoretic arguments [6]. As noted previously, linear encoders indeed achieve optimal compression when the source alphabet is a finite field. □

A. Computation of modulo two sum of two binary sources [4]

We will now apply the results of the compression of finite fields to the function computation problem described in Section II (also, see Fig 1) and show the rate benefits compared to the Slepian-Wolf encoding.

Consider an example where $N = 2$ and $X(1)$ and $X(2)$ are binary random variables i.e., $\mathcal{A} = \mathbb{F}_2$. Let the receiver be interested in computing the linear combination $X = X(1) + X(2)$. Assume the sources to have joint distribution given by $P(0,0) = P(1,1) = p/2$, $P(0,1) = P(1,0) = (1-p)/2,\ 0 < p < 1/2$. The distribution of X is then given by $P(0) = p$ and $P(1) = 1-p$. Hence, $H(X) = -p\log p - (1-p)\log(1-p) := h(p)$. Thus, if we use the matrix A in the encoder according to the distribution of X then the rate of each encoder will be $h(p)$ bits, giving a total sum rate of $2h(p)$ bits. However, if we resort to the method of recovering both $X(1)$ and $X(2)$ at the receiver and then compute $X(1) + X(2)$ (which is the Slepian-Wolf encoder) then the sum rate incurred will be $H(X(1), X(2)) = 1 + h(p)$ bits. It can be shown that $2h(p)$ is less than $1 + h(p)$ if $0 < p < \frac{1}{2}$.

VI. Linear compression over the ring $\mathbb{Z}_{p^r}$

In this section we consider the case when linear compression has to be done over the source alphabet $\mathcal{A} = \mathbb{Z}_{p^r}$. As it turns out, the ideals of $\mathbb{Z}_{p^r}$ play a major role in determining its compressibility using linear encoders. We therefore highlight a few quick facts regarding the ideal structure of $\mathbb{Z}_{p^r}$.

The ideals of $\mathbb{Z}_{p^r}$ are $p^i\mathbb{Z}_{p^r}, 0 \le i \le r$. The ideal $p^i\mathbb{Z}_{p^r}$ is isomorphic to $\mathbb{Z}_{p^{r-i}}$. The quotient ring $\mathbb{Z}_{p^r}/p^i\mathbb{Z}_{p^r}$, comprised of the cosets of $p^i\mathbb{Z}_{p^r}$ in $\mathbb{Z}_{p^r}$, is isomorphic to $\mathbb{Z}_{p^i}$ and hence, we will identify $\mathbb{Z}_{p^i}$ with the coset representatives of $\mathbb{Z}_{p^r}/p^i\mathbb{Z}_{p^r}$.

As in the system model, $X \sim P_X$ denotes the source random variable, defined over $\mathbb{Z}_{p^r}$. Define a new random variable, $[X]_i = X \bmod p^i$ and let $P_{[X]_i}$ denote the induced distribution on $[X]_i$. For example, if the ring is $\mathbb{Z}_4$, then $[X]_1 \sim (P_X(0) + P_X(2), P_X(1) + P_X(3))$. Note that X and $[X]_i$ are jointly distributed according to

$$P_{X,[X]_i}(x,y) = \begin{cases} P_X(x) & \text{if } y = x \bmod p^i \ , \\ 0 & \text{else} \ . \end{cases} \tag{24}$$

Also, if a sequence $\mathbf{x} \in p^i\mathbb{Z}_{p^r}^n \backslash p^{i+1}\mathbb{Z}_{p^r}^n$ we denote it as $p^i || \mathbf{x}$.

Theorem 2. *[5] [9] For the source X drawn i.i.d. $\sim P_X$ and whose alphabet is $\mathbb{Z}_{p^r}$, the achievable rate region under linear encoding is given by*

$$R \ \ge \ \max_{0 \le i < r} \left(\frac{r}{r-i} \right) H(X|[X]_i) \ . \tag{25}$$

We will need the following two lemmas to prove the achievability part of this theorem.

Lemma 3. *Let $\mathbf{z} \in \mathbb{Z}_{p^r}^n$ and $p^i || \mathbf{z}$. Then,*

$$|\{A \in M_{k \times n}(\mathbb{Z}_{p^r}) : A\mathbf{z} = 0\}| = p^{r(n-1)k} p^{ik}, \ 0 \le i \le r \ . \tag{26}$$

Proof: Please see Appendix B □

Lemma 4. *Consider a sequence* $\mathbf{y} \in A_\varepsilon^n([X]_i)$ *and let* $C_\mathbf{y} = \mathbf{y} + p^i\mathbb{Z}_{p^r}^n$ *be a coset of* $p^i\mathbb{Z}_{p^r}^n$*. Then,* $A_\varepsilon^n(X) \cap C_\mathbf{y} = A_\varepsilon^n(X|\mathbf{y})$.

Proof: We only give a proof sketch here. A detailed proof is presented in Appendix C. Consider a sequence $\mathbf{x}$ that is typical and belongs to the coset $C_\mathbf{y}$. Since $\mathbf{y} = \mathbf{x} \bmod p^i$ is a deterministic function of $\mathbf{x}$, if $\mathbf{x}$ is likely to occur then $\mathbf{y}$ is also likely to occur i.e., $\mathbf{x}$ and $\mathbf{y}$ are jointly typical. Now, consider a sequence $\mathbf{x}$ that is jointly typical with $\mathbf{y}$, which means $\mathbf{x}$ is also typical. Also, since the cosets of $p^i\mathbb{Z}_{p^r}^n$ are disjoint, $\mathbf{x}$ cannot be jointly typical with any $\mathbf{y}^{'} \neq \mathbf{y}$ and hence $\mathbf{x} \in C_\mathbf{y}$.

□

Corollary 5. *Consider a sequence* $\mathbf{x} \in A_\varepsilon^n(X)$*. Then*

$$|\mathbf{x} + p^i\mathbb{Z}_{p^r}^n \backslash p^{i+1}\mathbb{Z}_{p^r}^n \cap A_\varepsilon^n(X)| \leq 2^{nH(X|[X]_i)(1+\varepsilon)} .$$

Proof: The left hand side can be upper bounded as

$$|\mathbf{x} + p^i\mathbb{Z}_{p^r}^n \backslash p^{i+1}\mathbb{Z}_{p^r}^n \cap A_\varepsilon^n(X)| \leq |\mathbf{x} + p^i\mathbb{Z}_{p^r}^n \cap A_\varepsilon^n(X)| .$$

Now, let $\mathbf{y} \in A_\varepsilon^n([X]_i)$ be the representative for the coset $\mathbf{x} + p^i\mathbb{Z}_{p^r}^n$. Thus

$$\begin{aligned} |\mathbf{x} + p^i\mathbb{Z}_{p^r}^n \backslash p^{i+1}\mathbb{Z}_{p^r}^n \cap A_\varepsilon^n(X)| &\leq |\mathbf{y} + p^i\mathbb{Z}_{p^r}^n \cap A_\varepsilon^n(X)| \\ &\overset{(a)}{=} |A_\varepsilon^n(X|\mathbf{y})| \\ &\leq 2^{nH(X|[X]_i)(1+\varepsilon)} . \qquad (27) \end{aligned}$$

where (a) follows from Lemma 4 and the last inequality follows from the upper bound on the size of the conditional typical set.

□

Achievability of Theorem 2

As in Section VI, the achievability part is once again shown by a random coding argument. The averaging is done over the set of all linear encoders of the form

$$A^{(n)} : \mathbb{Z}_{p^r}^n \longrightarrow \mathbb{Z}_{p^r}^k , \qquad (28)$$

where $A^{(n)}$ is assumed to be a realization of the random matrix $\mathbf{A}^{(n)}$, distributed uniformly on the ensemble $M_{k\times n}(\mathbb{Z}_{p^r})$. The decoder will also be a typical set decoder. Error events can be defined in the same way as in the field case and the average probability of error in decoding can be upper bounded as

$$\begin{aligned} P_e^{(n)} &\leq \delta_n + \sum_{\mathbf{x}\in A_\varepsilon^n(X)} P(x) \sum_{\substack{\mathbf{y}\neq\mathbf{x} \\ \mathbf{y}\in A_\varepsilon^n(X)}} P(\mathbf{Ay} = \mathbf{Ax}) \\ &= \delta_n + \sum_{\mathbf{x}\in A_\varepsilon^n(X)} P(x)\Delta(\mathbf{x}) , \qquad (29) \end{aligned}$$

where,

$$\Delta(\mathbf{x}) \;=\; \sum_{i=0}^{r-1} \sum_{\substack{\mathbf{y}\neq\mathbf{x}\\ \mathbf{y}\in A_\varepsilon^n(X)\\ p^i\|(\mathbf{y}-\mathbf{x})}} P(\mathbf{A}(\mathbf{y}-\mathbf{x})=0)\ . \tag{30}$$

Using the fact that $\mathbf{A}^{(n)}$ is distributed uniformly on the ensemble $M_{k\times n}(\mathbb{Z}_{p^r})$ and applying Lemma 3, we get

$$\begin{aligned}\Delta(\mathbf{x}) &= \sum_{i=0}^{r-1} p^{-(r-i)k} \sum_{\substack{\mathbf{y}\neq\mathbf{x}\\ \mathbf{y}\in A_\varepsilon^n(X)\\ p^i\|(\mathbf{y}-\mathbf{x})}} 1\\ &\le \sum_{i=0}^{r-1} p^{-(r-i)k} |\mathbf{x}+p^i\mathbb{Z}_{p^r}^n \backslash p^{i+1}\mathbb{Z}_{p^r}^n \cap A_\varepsilon^n(X)|\ .\end{aligned}$$

Applying Corollary 5 to the above equation and substituting the resulting expression in (29) we get,

$$P_e^{(n)} \;\le\; \delta_n + \sum_{i=0}^{r-1} p^{-(r-i)k} 2^{nH(X|[X]_i)(1+\varepsilon)}\ . \tag{31}$$

Thus if, $\frac{k}{n}\log p^r > \left(\frac{r}{r-i}\right) H(X|[X]_i)(1+\varepsilon),\ 0\le i<r$, then $P_e^{(n)}\to 0$ as $n\to\infty$. Since, $R^{(n)} = \frac{\log|\mathrm{Im}(A^{(n)})|}{n} \le \frac{k}{n}\log p^r$, the achievable part of the theorem follows.

Converse of Theorem 2

We will show that if for any sequence of linear encoders $\{A^{(n)}\}$ of the form given by

$$A^{(n)} : \mathbb{Z}_{p^r}^n \;\longrightarrow\; \mathbb{Z}_{p^r}^k\ , \tag{32}$$

and decoders $\{D^{(n)}\}$, the average probability of error $P_e^{(n)}\to 0$, then

$$\lim_{n\to\infty} \frac{k}{n}\log p^r \;\ge\; \max_{0\le i<r} \left(\frac{r}{r-i}\right) H(X|[X]_i)\ . \tag{33}$$

The converse thus assumes that the co-domain $\mathcal{M} = \mathbb{Z}_{p^r}^k$ and that the rate of the encoder is given by $\frac{k}{n}\log(p^r)$. The proof with these assumptions will help us in highlighting the fact that the presence of non-trivial ideals is the reason for suboptimality of linear compression over rings. A rigorous proof without these assumptions will be presented in the next section when we discuss linear compressibility of chain rings, of which the ring $\mathbb{Z}_{p^r}$ is a special case. Note that we do not make any assumption on the nature of the decoder.

Consider the sub-module $p^i\mathbb{Z}_{p^r}^n$ of $\mathbb{Z}_{p^r}^n$. Any vector $\mathbf{x}\in\mathbb{Z}_{p^r}^n$ can be written as

$$\begin{aligned}\mathbf{x} &= \mathbf{x} \bmod p^i + p^i\mathbf{x}_0\\ &= [\mathbf{x}]_i + p^i\mathbf{x}_0\ ,\end{aligned} \tag{34}$$

where $\mathbf{x}_0 \in \mathbb{Z}_{p^r}^n$. Thus

$$(35) \qquad A^{(n)}(\mathbf{x}) \;=\; A^{(n)}([\mathbf{x}]_i) + A^{(n)}(p^i\mathbf{x}_0) \ .$$

Now, consider a second system, as shown in Fig. 3 which is derived from the original system. The new system also has X^n as the source output, but it only encodes $p^iX_0^n$,

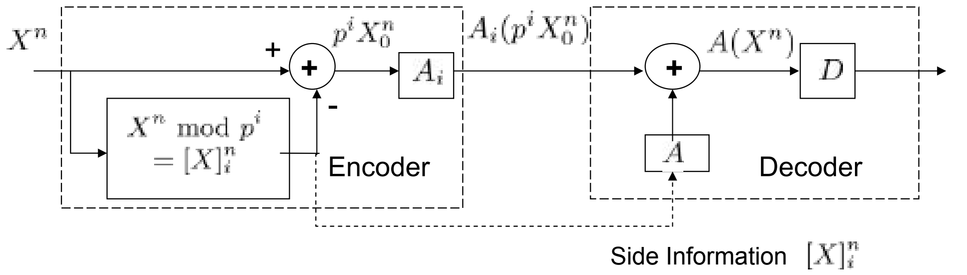

Fig. 3. An alternate system that aids in the proof of Theorem 2

where $X^n = [X]_i^n + p^iX_0^n$. The encoding is carried out using the restricted map $A_i^{(n)}$, where

$$(36) \qquad A_i^{(n)} \;=\; A^{(n)}\Big|_{p^i\mathbb{Z}_{p^r}^n} : p^i\mathbb{Z}_{p^r}^n \;\longrightarrow\; \mathbb{Z}_{p^r}^k \ .$$

At the receiver the missing information $[X]_i^n$ is given as a side information, so that together with the encoded output the receiver could first form the sum

$$(37) \qquad A^{(n)}(X^n) \;=\; A^{(n)}([X]_i^n) + A^{(n)}(p^iX_0^n) \ .$$

Now supposing that we use the decoder $D^{(n)}$ in this system, then X^n can be reliably decoded by this new system as well. But for this to be true, the theorem of source coding with side information (see Appendix A) says that rate of the system must be higher than the entropy of the source output conditioned on the side information, i.e.,

$$(38) \qquad \lim_{n\to\infty} \frac{1}{n} \log \left|\mathrm{Im}(A_i^{(n)})\right| \geq H(X|[X]_i) \ .$$

Now, let I be any ideal of $\mathbb{Z}_{p^r}$. Since A is $\mathbb{Z}_{p^r}-$linear, $A(I^n) \subseteq I^k$. Applying this to the ideal $p^i\mathbb{Z}_{p^r}$, we get

$$(39) \qquad A(p^i\mathbb{Z}_{p^r}^n) \;\subseteq\; p^i\mathbb{Z}_{p^r}^k \ .$$

Since $A(p^i\mathbb{Z}_{p^r}^n) = A_i(p^i\mathbb{Z}_{p^r}^n)$, $\left|A_i(p^i\mathbb{Z}_{p^r}^n)\right| \leq |p^i\mathbb{Z}_{p^r}^k| = p^{(r-i)k}$. Using this inequality in (38) we get,

$$(40) \qquad \lim_{n\to\infty} \frac{k}{n} \log p^r \;\geq\; \left(\frac{r}{r-i}\right) H(X|[X]_i) \ .$$

Since the above sequence of arguments in the converse can be carried out for every $i \in \{0,\ldots,r-1\}$, the converse follows.

VII. Linear compression over chain rings

In this section we consider the case when linear compression has to be done when the source alphabet $\mathcal{A}$ is a chain ring. We start with a brief introduction to chain rings.

Ring $\mathcal{A}$ will be called a chain ring if all ideals of the ring form a chain. Thus, if M denotes the maximal ideal of the ring, then all ideals are included in the chain $M \supseteq M^2 \supseteq \ldots$. Following are some properties of chain rings, which we will call upon during the sequel. For details, the reader is referred to [10] [11] [12].

(P.1) A ring is a chain ring if and only if it is a local principal ideal ring. A ring is called local if it has a unique maximal ideal.
(P.2) The characteristic of the ring is a power of prime; denote it by p^m.
(P.3) The maximal ideal, M, is exactly the set of nil-potent elements of the ring. Let θ denote its generator (such a generator exists as the ring is a PIR) and β its nil-potency index. Thus the chain of ideals in the ring is $(\theta) \supseteq (\theta)^2 \supseteq \ldots (\theta)^{\beta-1} \supseteq (\theta)^{\beta} = (0)$.
(P.4) For any $a \in \mathcal{A}, \exists! \;\; i$ such that $a = u\theta^i, \;\; 0 \leq i \leq \beta$, where u is a unit.
(P.5) There is a unique subring S of $\mathcal{A}$ such that S is isomorphic to the Galois ring $GR(p^m, r)$ for some r. Also $\mathcal{A} = S \oplus S\theta \oplus \ldots \oplus S\theta^{l-1}$ is an $S-$ module direct sum, where l is such that $p = u\theta^l$ for some unit u. Recall that the Galois ring $GR(p^m, r)$ is defined as the ring $\mathbb{Z}_{p^m}[x]/(f(x))$ where $f(x) \in \mathbb{Z}_{p^m}[x]$ is a monic polynomial of degree r and is irreducible modulo p.
(P.6) The quotient $\mathcal{A}/M$ is isomorphic to the Galois field $\mathbb{F}_q$, where $q = p^r$. Let V denote a set of coset representatives of M in $\mathcal{A}$. Then $\forall a \in \mathcal{A}, \exists! \;\; a_0, \ldots a_{\beta-1} \in V$ such that $a = \sum_{i=0}^{\beta-1} a_i \theta^i$. Thus $|(\theta^j)| = |V|^{\beta-j} = q^{\beta-j}$.

Examples of chain rings

1) Galois rings are well known examples of chain rings. It is known [11] that a ring S is isomorphic to $GR(p^m, r)$ if and only if S is a chain ring of characteristic p^m whose maximal ideal is pS. As special cases, Galois rings include $GR(p^m, 1) = \mathbb{Z}_{p^r}$ and $GR(p, r) = \mathbb{F}_{p^r}$.
2) Our second example is the ring $\mathcal{A} = \mathbb{Z}[i]/4\mathbb{Z}[i] \cong \mathbb{Z}_4[i]$, where $1 + i^2 = 0$. The only maximal ideal is generated by $\theta = 1 + i$, with $\beta = 4$. If expressed in the form as given in (P.5), $\mathcal{A} = \mathbb{Z}_4 \oplus \mathbb{Z}_4(1+i)$ as a $\mathbb{Z}_4$ module.
3) Our last example is the ring $\mathcal{A} = \mathbb{Z}[\omega]/3\mathbb{Z}[\omega] \cong \mathbb{Z}_3[\omega]$, where $1 + \omega + \omega^2 = 0$. $M = (1 + 2\omega), \;\; \beta = 2, \mathcal{A} = \mathbb{Z}_3 \oplus \mathbb{Z}_3(1 + 2\omega)$.

Theorem 6. *Consider a source X drawn i.i.d. $\sim P_X$ whose alphabet is the chain ring $\mathcal{A}$. Let $M = (\theta)$ denote the maximal ideal of $\mathcal{A}$ and let β be its nilpotency index. The achievable rate region under $\mathcal{A}-$module homomorphic encoding is given by*

$$R \geq \max_{0 \leq i < \beta} \left(\frac{\beta}{\beta - i} \right) H(X|[X]_i) \;, \tag{41}$$

where $[X]_i = X \bmod \theta^i, 0 \leq i < \beta$.

Remark 1. *The rate region in the Theorem 6, when specialized to the case when $\mathcal{A}$ is the ring $\mathbb{Z}_{p^r}$ leads to Theorem 2.*

Before we proceed to the proof of Theorem 6 we first state a lemma that will be used in calculating the average probability of error.

Lemma 7. *Let $\mathbf{z} \in \mathcal{A}^n$. Let $\theta^i || \mathbf{z}$ i.e., $\mathbf{z} \in \theta^i \mathcal{A}^n \backslash \theta^{i+1} \mathcal{A}^n$. Then,*

$$|\{A \in M_{k\times n}(\mathcal{A}) : A\mathbf{z} = 0\}| = q^{ik} q^{\beta k(n-1)} \ , \ 0 \le i \le \beta \ . \tag{42}$$

Proof: The proof is similar to the proof of Lemma 3. While in Lemma 3 we used the fact that every element in $\mathbb{Z}_{p^r}$ has a unique $p-$ary expansion, herein for a chain ring, we should use the fact every element has a unique $\theta-$ary expansion (due to Property (P.6)). □

Achievability of Theorem 6

Proceeding along the same lines as those in the $\mathbb{Z}_{p^r}$ case, we get

$$\begin{aligned} P_e^{(n)} &\le \delta_n + \sum_{\mathbf{x} \in A_\varepsilon^n(X)} P(\mathbf{x}) \sum_{\substack{\mathbf{y} \neq \mathbf{x} \\ \mathbf{y} \in A_\varepsilon^n(X)}} P(\mathbf{Ay} = \mathbf{Ax}) \\ &:= \sum_{\mathbf{x} \in A_\varepsilon^n(X)} P(\mathbf{x}) \Delta(\mathbf{x}) \ . \end{aligned} \tag{43}$$

Now,

$$\Delta(\mathbf{x}) = \sum_{i=0}^{\beta-1} \sum_{\substack{\mathbf{y} \neq \mathbf{x} \\ \mathbf{y} \in A_\varepsilon^n(X) \\ \theta^i || (\mathbf{y}-\mathbf{x})}} P(\mathbf{A}(\mathbf{y}-\mathbf{x}) = 0) \tag{44}$$

$$\overset{(a)}{=} \sum_{i=0}^{\beta-1} q^{-(\beta-i)k} \sum_{\substack{\mathbf{y} \neq \mathbf{x} \\ \mathbf{y} \in A_\varepsilon^n(X) \\ \theta^i || (\mathbf{y}-\mathbf{x})}} 1 \ , \tag{45}$$

where (a) follows from Lemma 7.

Similar to Corollary 5, it can be shown that

$$|\{\mathbf{y} : \theta^i || (\mathbf{y}-\mathbf{x}), \mathbf{y} \in A_\varepsilon^n(X)\}| \le 2^{nH(X|[X]_i)(1+\varepsilon)}.$$

Substituting the above expression in (45), we get an upper bound on the probability of error as

$$P_e^{(n)} \le \delta_n + \sum_{i=0}^{\beta-1} q^{-(\beta-i)k} 2^{nH(X|[X]_i)(1+\varepsilon)} \ . \tag{46}$$

Thus if, $\frac{k}{n}\log q^{\beta} \;>\; \left(\frac{\beta}{\beta-i}\right)H(X|[X]_i),\; 0 \le i < \beta$, then $P_e^{(n)} \to 0$ as $n \to \infty$. Since, $R^{(n)} = \frac{\log|\mathrm{Im}(A^{(n)})|}{n} \le \frac{k}{n}\log q^{\beta}$, the achievable part of the theorem follows.

Converse of Theorem 6

We will show that any sequence of $\mathcal{A}-$module homomorphisms $\{\phi^{(n)}\}$ of the form given in (1), for which $P_e^{(n)} \to 0$, must satisfy

$$\lim_{n\to\infty} R^{(n)} \ge \left(\frac{\beta}{\beta-i}\right)H(X|[X]_i),\; 0 \le i < \beta \;. \tag{47}$$

Let $\phi_i^{(n)}$ denote the restriction of $\phi^{(n)}$ to the ideal $\theta^i\mathcal{A}^n$ of $\mathcal{A}^n$ and let $R_i^{(n)}$ be the rate of the restriction, i.e.,

$$R_i^{(n)} = \frac{\log(|\mathrm{Im}(\phi_i^{(n)})|)}{n} \;. \tag{48}$$

A necessary condition on $R_i^{(n)}$, which follows from arguments similar to those that led to (38), is given by

$$\lim_{n\to\infty} R_i^{(n)} \;\ge\; H(X|[X]_i) \;. \tag{49}$$

The following lemma, derived from algebraic arguments, will relate the rates $R^{(n)}$ and $R_i^{(n)}$.

Lemma 8. *For any n and $0 \le i < \beta$ we have,*

$$\log(|Ker(\phi^{(n)})|) \;\le\; \left(\frac{\beta}{\beta-i}\right)\log(|Ker(\phi_i^{(n)})|) \;.$$

Proof: We drop the superscript on $\phi^{(n)}$ and $\phi_i^{(n)}$ for convenience. Clearly $\mathrm{Ker}(\phi)$ is finitely generated and let $(\mathbf{t_1},\mathbf{t_2},\ldots,\mathbf{t}_\ell)$ be a set of generators, where $\mathbf{t_j} \in \mathcal{A}^n,\; 1 \le j \le \ell$. Equivalently, $\mathrm{Ker}(\phi)$ is the column space of the matrix $S = [\mathbf{t_1}\mathbf{t_2}\ldots\mathbf{t}_\ell]$. Let us denote the column space of S by $\mathrm{Col}(S)$. Now $\mathrm{Ker}(\phi_i) = \mathrm{Col}(S) \cap \theta^i\mathcal{A}^n$.

Let $\bar{S}$ be the image of S under any elementary row or column transformation. It is easy to show the following equalities.

$$|\mathrm{Col}(S)| \;=\; |\mathrm{Col}(\bar{S})| \tag{50}$$

$$|\mathrm{Col}(S)\cap\theta^i\mathcal{A}^n| \;=\; |\mathrm{Col}(\bar{S})\cap\theta^i\mathcal{A}^n| \;. \tag{51}$$

i.e., $|\mathrm{Ker}(\phi)|$ and $|\mathrm{Ker}(\phi_i)|$ remain invariant to elementary row and column transformations.

Also due to Property (P.4), we know that every element in the matrix S is of the form $u\theta^j$ for some $j,\; 0 \le j \le \beta$, and unit u. With this, it can be shown that after a series of elementary row and column operations, the matrix S can be transformed to the form

$$
\begin{bmatrix}
I_{\ell_0} & & & & \\
& \theta I_{\ell_1} & & & \\
& & \ddots & & \\
& & & \theta^{\beta-1} I_{\theta_{\beta-1}} & \\
& & & & 0_{\ell_\beta \times \ell_\beta} \\
& 0_{n-\ell \times \ell} & & &
\end{bmatrix}, \tag{52}
$$

where $\sum_{k=0}^{\beta} \ell_k = \ell$. Then we have,

$$
\begin{aligned}
|\mathrm{Ker}(\phi)| &= q^{\ell_0\beta+\ell_1(\beta-1)+\ldots+\ell_i(\beta-i)+\ell_{i+1}(\beta-i-1)+\ldots\ell_{\beta-1}} \\
|\mathrm{Ker}(\phi_i)| &= q^{\ell_0(\beta-i)+\ell_1(\beta-i)+\ldots+\ell_i(\beta-i)+\ell_{i+1}(\beta-i-1)+\ldots\ell_{\beta-1}} \; .
\end{aligned}
$$

The Lemma now follows by taking logarithm on both sides and comparing the terms. □

Corollary 9.

$$
R^{(n)} \geq \left(\frac{\beta}{\beta-i}\right) R_i^{(n)} \; . \tag{53}
$$

Proof: For any homomorphism $\psi : M \to N$, we have $|\mathrm{Im}(\psi)||\mathrm{Ker}(\psi)| = |M|$. When applied to the maps ϕ and ϕ_i, this gives

$$
|\mathrm{Im}(\phi)||\mathrm{Ker}(\phi)| = q^{\beta n} \tag{54}
$$

$$
|\mathrm{Im}(\phi_i)||\mathrm{Ker}(\phi_i)| = q^{(\beta-i)n} \; . \tag{55}
$$

The corollary follows by substituting the above two equations in the definition of rates R^n and $R_i^{(n)}$. □

Now, using the relation of rates given by Corollary 9 in (49), we get

$$
\lim_{n\to\infty} R^{(n)} \geq \left(\frac{\beta}{\beta-i}\right) H(X|[X]_i) \; . \tag{56}
$$

Since the above sequence of arguments in the converse can be carried out for every $i \in \{0, \ldots, \beta-1\}$, the converse follows.

VIII. Linear compression over Principal Ideal Rings

Theorem 10. *[10] Let $\mathcal{A}$ be a PIR. Then $\mathcal{A}$ decomposes as direct product of finite chain rings. Further, the decomposition is unique upto ordering.*

Let $\mathcal{A} \cong \mathcal{A}_1 \times \ldots \times \mathcal{A}_t$ be the direct product decomposition of the PIR $\mathcal{A}$ as per Theorem 10, where $\mathcal{A}_i, \forall\ i$ are chain rings. Thus finding the achievable rate region of $\mathcal{A}-$module homomorphisms of $\mathcal{A}^n$ is same as characterizing achievable rate region for $\mathcal{A}_1 \times \ldots \times \mathcal{A}_t-$module homomorphisms of $\mathcal{A}_1^n \times \ldots \times \mathcal{A}_t^n$. Hence without loss of generality, we shall assume that $\mathcal{A} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_t$. Note that scalar multiplication in the $\mathcal{A}_1 \times \ldots \times \mathcal{A}_t-$module $\mathcal{A}_1^n \times \ldots \times \mathcal{A}_t^n$ is defined as

$$(\alpha_1, \ldots, \alpha_t) \circ (\mathbf{a}_1, \ldots, \mathbf{a}_t) \;=\; (\alpha_1 \mathbf{a}_1, \ldots, \alpha_t \mathbf{a}_t),\ \mathbf{a}_i \in \mathcal{A}_i^n,\ \alpha_i \in \mathcal{A}_i,\ \forall i\ . \tag{57}$$

In what follows, we will assume $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$. Extension to more number of components will follow in a straight forward manner.

The random variable on the source will be denoted by $X = (X_1, X_2)$. Let $M_k = (\theta_k)$ denote the maximal ideal of $\mathcal{A}_k$ and β_k denote the nil-potency index of M_k, $k = 1,2$. Let $[X]_{i,j} = ([X_1]_i, [X_2]_j)$ denote the derived random variable on the quotient $\mathcal{A}_1 \times \mathcal{A}_2 / M_1^i \times M_2^j$, $0 \le i \le \beta_1$, $0 \le j \le \beta_2$. The rate region is characterized by the following theorem.

Theorem 11. *Consider a source $X = (X_1, X_2)$ drawn i.i.d. $\sim P_X = P_{X_1,X_2}$ whose alphabet is the principal ideal ring $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$. The achievable rate region under $\mathcal{A}-$module homomorphic encoding is given by*

$$\mathcal{R} \;=\; \left\{ R_1 + R_2 \mid \left(\frac{\beta_1 - i}{\beta_1}\right) R_1 + \left(\frac{\beta_2 - j}{\beta_2}\right) R_2 \;\ge H(X|[X]_{i,j}), 0 \le i \le \beta_1,\ 0 \le j \le \beta_2 \right\}$$

Achievability of Theorem 11

Consider the $\mathcal{A}_1$ and $\mathcal{A}_2-$module homomorphisms, A_1 and A_2, respectively, as follows.

$$A_1 : \mathcal{A}_1^n \to \mathcal{A}_1^{k_1} \quad , \quad A_2 : \mathcal{A}_2^n \to \mathcal{A}_2^{k_2}\ . \tag{58}$$

Thus $A_1 \in M_{k \times n}(\mathcal{A}_1)$ and $A_2 \in M_{k \times n}(\mathcal{A}_2)$. Use these maps to construct the $\mathcal{A}_1 \times \mathcal{A}_2$ module homomorphism A as follows.

$$\begin{aligned} A : \mathcal{A}_1^n \times \mathcal{A}_2^n \;&\longrightarrow\; \mathcal{A}_1^{k_1} \times \mathcal{A}_2^{k_2} \\ (\mathbf{x}_1, \mathbf{x}_2) \;&\rightsquigarrow\; (A_1 \mathbf{x}_1, A_2 \mathbf{x}_2)\ , \end{aligned} \tag{59}$$

where the scalar multiplication in the $\mathcal{A}_1 \times \mathcal{A}_2-$module $\mathcal{A}_1^{k_1} \times \mathcal{A}_2^{k_2}$ is defined as

$$(\alpha_1, \alpha_2) \circ (\mathbf{a}_1, \mathbf{a}_2) \;=\; (\alpha_1 \mathbf{a}_1, \alpha_2 \mathbf{a}_2),\ \mathbf{a}_1 \in \mathcal{A}_1^{k_1},\ \mathbf{a}_2 \in \mathcal{A}_2^{k_2}, \alpha_1 \in \mathcal{A}_1,\ \alpha_2 \in \mathcal{A}_2\ . \tag{60}$$

The achievability uses a random coding argument by averaging over set of all $\mathcal{A}-$module homomorphisms of the form given in (59). Assuming that the source transmits $(\mathbf{x}_1, \mathbf{x}_2)$, the decoder searches for a unique $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) \in A_\varepsilon^{(n)}(X_1, X_2)$ such that $A_1 \hat{\mathbf{x}}_1 = A_1 \mathbf{x}_1$ and $A_2 \hat{\mathbf{x}}_2 = A_2 \mathbf{x}_2$. The following lemma is the analogue of Corollary 5, for the direct product of chain rings.

Lemma 12. *For any $(\mathbf{x_1}, \mathbf{x_2}) \in A_\varepsilon^{(n)}(X_1, X_2)$, we have*

$$\begin{aligned} \left| A_\varepsilon^{(n)}(X_1, X_2) \;\cap\; (\mathbf{x_1}, \mathbf{x_2}) + \theta_1^i \mathcal{A}_1^n \times \theta_2^j \mathcal{A}_2^n \right| \;\le 2^{nH(X|[X]_{i,j})(1+\varepsilon)}, \\ 0 \le i \le \beta_1,\ 0 \le j \le \beta_2\ . \end{aligned} \tag{61}$$

Using the above lemma, calculation of the average probability of error now could be done in the same way as was done in section VII for the single component case. □

Converse of Theorem 11

Let $\phi^{(n)} : \mathcal{A}_1^{(n)} \times \mathcal{A}_2^{(n)} \to \mathcal{M}$ be a sequence of $\mathcal{A}_1 \times \mathcal{A}_2-$module homomorphisms that achieves a rate R. We will show that $R \in \mathcal{R}$, where $\mathcal{R}$ is as defined in Theorem 11. Define the component maps $\phi_1^{(n)}$ and $\phi_2^{(n)}$ as follows.

$$\phi_1^{(n)} = \phi^{(n)}|_{\mathcal{A}_1^{(n)} \times 0} \qquad \phi_2^{(n)} = \phi^{(n)}|_{0 \times \mathcal{A}_2^{(n)}} \ . \tag{62}$$

Now, consider the ideal $D_{i,j} = \theta_1^i \mathcal{A}_1^{(n)} \times \theta_2^j \mathcal{A}_2^{(n)}$ of $\mathcal{A}_1^{(n)} \times \mathcal{A}_2^{(n)}$ and let $\psi^{(n)}$ denote the restriction of $\phi^{(n)}$ to $D_{i,j}$, i.e;

$$\psi^{(n)} = \phi^{(n)}|_{D_{i,j}} : D_{i,j} \longrightarrow \mathcal{M} \tag{63}$$

Also define the component maps $\psi_1^{(n)}$ and $\psi_2^{(n)}$ as follows.

$$\psi_1^{(n)} = \psi^{(n)}|_{\theta_1^i \mathcal{A}_1^{(n)} \times 0} \qquad \psi_2^{(n)} = \psi^{(n)}|_{0 \times \theta_2^j \mathcal{A}_2^{(n)}} \ . \tag{64}$$

By applying Lemma 8 and Corollary 9 to the component maps $\psi_1^{(n)}$ and $\psi_2^{(n)}$, we get

$$R_{\psi_1}^{(n)} \leq \left(\frac{\beta_1 - i}{\beta_1} \right) R_{\phi_1}^{(n)} \ , \ R_{\psi_2}^{(n)} \leq \left(\frac{\beta_2 - j}{\beta_2} \right) R_{\phi_2}^{(n)} \ . \tag{65}$$

These equations, along with the fact that $R_{\psi_1}^{(n)} + R_{\psi_2}^{(n)} \geq R_{\psi}^{(n)}$ gives

$$\left(\frac{\beta_1 - i}{\beta_1} \right) R_{\phi_1}^{(n)} \ + \ \left(\frac{\beta_2 - j}{\beta_2} \right) R_{\phi_2}^{(n)} \ \geq \ R_{\psi}^{(n)} \ . \tag{66}$$

Similar to what was shown in (49), the rate $R_{\psi}^{(n)}$ can be shown to be constrained as

$$\lim_{n \to \infty} R_{\psi}^{(n)} \ \geq \ H(X|[X]_{i,j}) \ . \tag{67}$$

Combining the above two equations, we get

$$\left(\frac{\beta_1 - i}{\beta_1} \right) R_1 \ + \ \left(\frac{\beta_2 - j}{\beta_2} \right) R_2 \ \geq \ H(X|[X]_{i,j}), \tag{68}$$

where $R_1 = \lim_{n \to \infty} R_{\phi_1}^{(n)}$ and $R_2 = \lim_{n \to \infty} R_{\phi_2}^{(n)}$

The converse could now be completed by observing the following algebraic fact.

Lemma 13.

$$R^{(n)} = R^{(n)}_{\phi_1} + R^{(n)}_{\phi_2}, \ \forall n \ . \tag{69}$$

Proof: By assumption, $A^{(n)} : \mathcal{A}_1^{(n)} \times \mathcal{A}_2^{(n)} \rightarrow \mathcal{M}$ is an $\mathcal{A}_1 \times \mathcal{A}_2-$ module homomorphism and thus by definition of a module homomorphism, the co-domain $\mathcal{M}$ is an $\mathcal{A}_1 \times \mathcal{A}_2-$module. Define the sets $\mathcal{M}_1$ and $\mathcal{M}_2$ as follows.

$$\begin{aligned} \mathcal{M}_1 &= \{(a_1,0).m, \ \forall \ a_1 \in \mathcal{A}_1, \ m \in \mathcal{M}\} \\ \mathcal{M}_2 &= \{(0,a_2).m, \ \forall \ a_2 \in \mathcal{A}_2, \ m \in \mathcal{M}\} \ . \end{aligned} \tag{70}$$

Clearly $\mathcal{M}_1$ and $\mathcal{M}_2$ are submodules of $\mathcal{M}$ and further, it can be checked that $\mathcal{M}$ is the internal direct sum of $\mathcal{M}_1$ and $\mathcal{M}_2$, i.e; $\mathcal{M} = \mathcal{M}_1 \oplus \mathcal{M}_2$. The Lemma now follows by checking that $\mathrm{Im}(\phi_1^{(n)}) \subseteq \mathcal{M}_1$ and $\mathrm{Im}(\phi_2^{(n)}) \subseteq \mathcal{M}_2$. □

The converse now follows by combining (68) and Lemma 13.

Acknowledgment

This research is supported by the DRDO-IISc Program on Advanced Research in Mathematical Engineering and Microsoft Corporation and Microsoft Research India under the Microsoft Research India PhD Fellowship Award.

References

[1] The SmartDetect Project Team, "Wireless sensor networks for human intruder detection," *Journal of the Indian Institute of Science, Special issue on Advances in Electrical Science*, vol. 90, no. 3, pp. 471–480, July-September 2010 (invited).

[2] R. Puri, A. Majumdar, and K. Ramchandran, "Prism: A video coding paradigm with motion estimation at the decoder," *Image Processing, IEEE Transactions on*, vol. 16, no. 10, pp. 2436–2448, Oct. 2007.

[3] A. Orlitsky and J.R. Roche, "Coding for computing," *Information Theory, IEEE Transactions on*, vol. 47, no. 3, pp. 903–917, 2002.

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

[5] D. Krithivasan and S. Pradhan, "Distributed Source Coding using Abelian Group Codes," *to appear in IEEE Transactions on Information Theory, 2010/2011, available: arxiv: 0808.2659v1[cs.IT].*

[6] T.M. Cover, J.A. Thomas, and J. Wiley, *Elements of information theory*. Wiley Online Library, 1991.

[7] I. Csiszar, "Linear codes for sources and source networks: Error exponents, universal coding," *Information Theory, IEEE Transactions on*, vol. 28, no. 4, pp. 585 – 592, Jul. 1982.

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

[9] K. Vinodh, V. Lalitha, N. Prakash, P. Vijay Kumar, and S. Sandeep Pradhan, "On the achievable rates of sources having a group alphabet in a distributed source coding setting," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, 29 2010.

[10] B.R. McDonald, *Finite rings with identity*. M. Dekker, 1974.

[11] W.E. Clark and D.A. Drake, "Finite chain rings," in *Abhandlungen aus dem mathematischen Seminar der Universitat Hamburg*, vol. 39, no. 1. Springer, 1973, pp. 147–153.

[12] G.H. Norton and A. Sălăgean, "On the structure of linear and cyclic codes over a finite chain ring," *Applicable algebra in engineering, communication and computing*, vol. 10, no. 6, pp. 489–506, 2000.

[13] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.

Appendix A
Primer on Information Theory

The goal of this section is to introduce basic concepts in information theory like notions of entropy and typical sets. We will also see how these concepts naturally arise in various source compression problems.

A. Entropy

Consider a discrete random variable(r.v.) X, having distribution P_X on a set $\mathcal{A}$. The set $\mathcal{A}$ will be referred to as the alphabet of the r.v. X. Entropy of X, denoted by $H(X)$, is defined as

$$H(X) = \sum_{x \in \mathcal{A}} P_X(x) \log_2 \frac{1}{P_X(x)}. \tag{71}$$

$H(X)$ could be considered as a measure of the amount of uncertainty in the r.v. X. The log is to the base 2 and the unit of entropy will be bits. It can be checked that $0 \leq H(X) \leq |\mathcal{A}|$. $H(X) = 0$, iff the distribution P_X has the form

$$P_X(x) \begin{cases} = 1 & \text{if } x = x_0 \\ = 0 & \text{if } x \neq x_0 \ , \end{cases} \tag{72}$$

i.e; there is no randomness in X. In a similar manner, $H(X) = |\mathcal{A}|$ iff P_X is uniform over the alphabet $\mathcal{A}$.

The notion of entropy extends to a collection of random variables as well, wherein we measure the overall uncertainty of the collection of random variables. For example, for a pair of random variables (X,Y) having the alphabet $\mathcal{A} \times \mathcal{B}$, the joint entropy $H(X,Y)$ of the pair (X,Y) is defined as

$$H(X,Y) = \sum_{(x,y) \in \mathcal{A} \times \mathcal{B}} P_{XY}(x,y) \log \frac{1}{P_{XY}(x,y)} \ . \tag{73}$$

For the pair $H(X,Y)$, one could also ask for the entropy of X given the event $Y = y$. This is denoted by $H(X|Y = y)$ and defined as

$$H(X|Y = y) = \sum_{x \in \mathcal{A}} P_{X|Y}(x|y) \log \frac{1}{P_{X|Y}(x|y)} \ . \tag{74}$$

Further, one averages $H(X|Y = y)$ over the conditioning random variable Y to obtain the conditional entropy $H(X|Y)$, i.e;

$$\begin{aligned} H(X|Y) &= \sum_{y \in \mathcal{B}} P_Y(y) H(X|Y = y) \\ &= \sum_{(x,y) \in \mathcal{A} \times \mathcal{B}} P_{XY}(x,y) \log \frac{1}{P_{X|Y}(x|y)} \ . \qquad (75) \end{aligned}$$

It is easy to verify that joint and conditional entropy of the pair (X,Y) are related as

$$H(X,Y) \ = \ H(X) + H(Y|X) \ = \ H(Y) + H(X|Y) \ . \tag{76}$$

B. Typical Sequences

Let $X^n = X_1, X_2, \ldots X_n$ be n independent random variables, identically distributed according to P_X. Let $\mathbf{x} = (x_1, x_2, \ldots x_n)$ denote a realization of X^n. We are interested in the predicting the realizations of X^n which are most likely to occur. It turns out that for large n, these are exactly those sequences whose empirical symbol frequencies are close to the actual distribution P_X. Such sequences will be called typical sequences and their collection, the typical set. More formally, let $N(a|\mathbf{x})$ denote the number of occurrences of the symbol a in the realization $\mathbf{x}$. The typical set corresponding to the random variable X is given by

$$A_\varepsilon^{(n)}(X) \;=\; \left\{\mathbf{x} : \left|\frac{N(a|\mathbf{x})}{n} - P_X(a)\right| \leq \varepsilon P_X(a), \forall a \in \mathcal{A}\right\} . \tag{77}$$

Some important properties of the typical set are stated next[13].

(P. 1)

$$P(A_\varepsilon^{(n)}(X)) \;\geq\; 1 - \delta_n \;, \tag{78}$$

where $\delta_n \to 0$ as $n \to \infty$

(P. 2) For any $\mathbf{x} \in A_\varepsilon^{(n)}(X)$,

$$2^{-(1+\varepsilon)nH(X)} \leq P(\mathbf{x}) \leq 2^{-(1-\varepsilon)nH(X)} \;. \tag{79}$$

(P. 3)

$$(1-\delta_n)2^{(1-\varepsilon)nH(X)} \leq |A_\varepsilon^{(n)}(X)| \leq 2^{(1+\varepsilon)nH(X)} \;. \tag{80}$$

The concept of typical sets and typical sequences extends to collection of random variables as well. In the case of a pair of r.v.s (X,Y), the joint typical set corresponding to the pair (X,Y) is defined as

$$A_\varepsilon^{(n)}(X,Y) \;=\; \left\{\mathbf{x} : \left|\frac{N(a,b|\mathbf{x},\mathbf{y})}{n} - P_{XY}(a,b)\right| \leq \varepsilon P_{XY}(a,b), \forall (a,b) \in \mathcal{A} \times \mathcal{B}\right\}. \tag{81}$$

We also need the notion of a conditional typical set. The conditional typical set $A_\varepsilon^n(X|\mathbf{y})$ is defined as the set of all sequences $\mathbf{x}$ which are jointly typical given the given realization of $Y^n = \mathbf{y}$; i.e;

$$A_\varepsilon^{(n)}(X|\mathbf{y}) \;=\; \left\{\mathbf{x} : (\mathbf{x},\mathbf{y}) \in A_\varepsilon^{(n)}(X,Y)\right\} \;. \tag{82}$$

Properties similar to (78)-(80) could be written for both joint and conditional typical sets. The analogue of (80) give us the following upper bound on the size of the conditional typical set

$$|A_\varepsilon^{(n)}(X|\mathbf{y})| \;\leq\; 2^{(1+\varepsilon)nH(X|Y)} \;. \tag{83}$$

This will be used in proving various theorems in the main text.

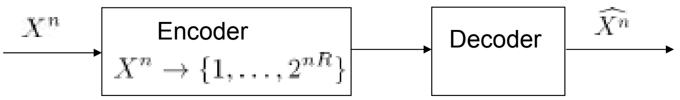

Fig. 4. System model for point-to-point source compression

C. Point-to-Point Source Compression

Consider a system in which a source outputs a sequence of i.i.d. random variables $X_1, X_2, \dots$ distributed according to P_X. The goal is to compress the output of the source by an encoder in such a manner that given the output of the encoder, one can decode the actual source reliably. This system is shown in Fig 4.

We will restrict our attention to block encoders. A block encoder assigns an index $i \in \{1,2,\dots,2^{nR}\}$ to every block of $n-$length source output. R is the rate of the encoder in bits per symbol. Clearly $R = \log|\mathcal{A}|$ bits suffice to encode the source such that reliable decoding is possible. But if we relax our requirement that we only need the probability of decoding error be made arbitrarily small, better compression rates can be achieved. To see how this is possible, note that by Property (P.1) of typical sets, for large n, the source output is typical with very high probability. Thus the encoder can restrict its attention to the typical set and ensure that every sequence in the typical set gets a distinct index. Then by Property (P.3), the number of encoder indices is upper bounded by $2^{(1+\varepsilon)nH(X)}$. Given the encoder output, the decoder now searches for a typical sequence having the corresponding index. An error occurs at the decoder only when the original source sequence itself is not typical. Since the probability of this can be made arbitrarily small by choosing a large enough block length n, we have thus achieved almost lossless compression at a rate $R = (1+\varepsilon)H(X)$. Put in another way, almost lossless compression at any rate $R > H(X)$ is possible. Conversely, it can also be shown by information theoretic arguments that lossless compression at any rate less than $H(X)$ is not possible.

D. Source Compression with Side Information

Consider the same problem as above, wherein we need to compress and communicate a source X to a decoder, but now assume that the decoder also has access to another random variable Y, where (X,Y) are jointly distributed according to P_{XY} (see Fig 5). This problem is called as source coding with side information at the decoder. Intuitively, decoder knows something about X when Y is made available and hence we may expect

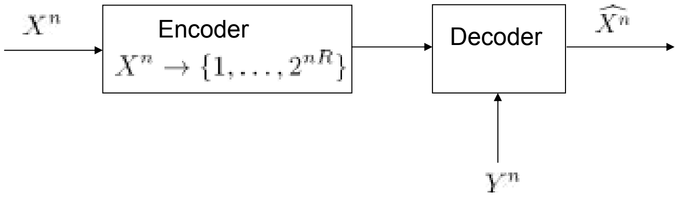

Fig. 5. System model for source compression with side information

that we will need a rate less than $H(X)$ for lossless compression. Indeed this turns out to be true and it can be shown that compression at any rate higher than $H(X|Y)$ is possible. To see how this is true, given the realization $Y^n = \mathbf{y}$, it can be shown that the probability of conditional typical set of X is almost 1 (analogous to (78)). Thus the decoder can restrict its attention to the conditional typical set of X, whose size is approximately $2^{nH(X|Y)}$ (analogous to (80)). If for a moment we assume that the source also knew the realization $Y^n = \mathbf{y}$, then the source only needs to encode the set $A_\varepsilon^{(n)}(X|\mathbf{y})$ and thus a rate of $H(X|Y)$ can be achieved. The surprising thing that can be shown is that even when the source does not access to Y^n, this rate could be achieved. This is shown by using a random coding argument, wherein at the encoder we randomly assign indices to the source outputs. One can then show that the probability that two source sequences which are both jointly typical with $Y^n = y^n$ becomes almost nil, as long as the rate of the encoder is higher than $H(X|Y)$. Note that this argument will only prove the existence of an encoder which achieves compression at a rate $H(X|Y)$. A converse could also be proved using information theoretic arguments which says that compression at any rate less than $H(X|Y)$ is not possible.

E. Distributed Source Compression

Consider two sources (X,Y) with joint distribution P_{XY}, which are spatially separated. Both sources are compressed separately by different encoders. The decoder is interested to recovering the pair (X,Y), in an almost lossless manner. The system is shown in Fig 6.

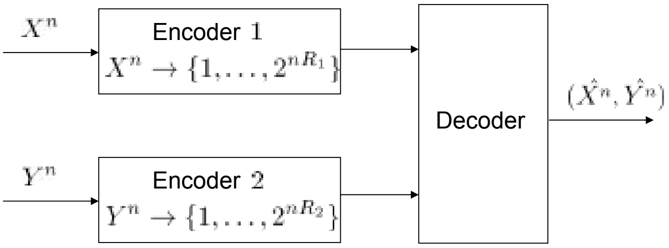
X^n
Encoder 1
$X^n \to \{1,\ldots,2^{nR_1}\}$
Y^n
Encoder 2
$Y^n \to \{1,\ldots,2^{nR_2}\}$
Decoder
$(\hat{X^n},\hat{Y^n})$

Fig. 6. System model for distributed source compression

The problem is referred to as the Slepian-Wolf source coding problem of two sources. The rate region in this case is given by

$$\begin{aligned} R_1 &\geq H(X|Y) \\ R_2 &\geq H(Y|X) \\ R_1 + R_2 &\geq H(X,Y)\ . \end{aligned} \tag{84}$$

Observe from the last of the inequalities that a sum rate of the joint entropy of the pair (X,Y) is achievable even though the sources are encoded separately. This is made possible by designing the encoders in such a manner that will allow the decoder to utilize the correlation properties between the two sources, while decoding. We will illustrate this by the following example. Assume that (X,Y) are temperature recordings in two nearby places both of which need to be communicated to a weather station.

Let us further assume that Y is either X or $X+1$. An optimal encoding strategy in this case is to allow the first encoder to compress X at a rate $H(X)$ and let the second encoder sends only one bit; 0 if Y is even and 1 if Y is odd. At the decoder, we first recover X, since $H(X)$ bits suffice to recover X almost losslessly. Clearly, now Y can also be recovered given its parity and X.

Appendix B
Proof of Lemma 3

Let $[a_1, a_2, \dots, a_n]$ be some row of A. Note that $a_i \in \mathbb{Z}_{p^r}$. Let $\mathbf{z} = [z_1, \dots, z_n]^t$. If $\mathbf{z} = 0$, then $p^r || \mathbf{z}$. Hence, for every matrix A, we have $A\mathbf{z} = 0$ and the statement of Lemma follows easily for this case. Now, consider the case when $\mathbf{z} \neq 0$. Since $p^i || \mathbf{z}$ we can write $\mathbf{z} = p^i [w_1, \dots, w_n]^t$, where $w_j \in \mathbb{Z}^*_{p^r}$ for some $1 \le j \le n$. Without loss of generality let us assume $w_1 \in \mathbb{Z}^*_{p^r}$. Hence,

$$\begin{aligned}
[a_1, \dots, a_n] p^i [w_1, \dots, w_n]^t &= 0 \\
p^i \sum_{j=1}^{n} a_j w_j &= 0 \\
p^i a_1 &= (w_1)^{-1} \sum_{j=2}^{n} a_j w_j \ . \qquad (85)
\end{aligned}$$

Since $a_1 \in \mathbb{Z}_{p^r}$ there is a unique p-ary expansion for a_1 as follows.

$$a_1 = \sum_{\ell=0}^{r-1} a_{1\ell} p^{\ell} \ , \tag{86}$$

for some $a_{1\ell} \in \mathbb{Z}_p$. Substituting the $p-$ary expansion of a_1 in (85) we get,

$$\sum_{\ell=0}^{r-i-1} a_{1\ell} p^{\ell+i} = (w_1)^{-1} \sum_{j=2}^{n} a_j w_j \ . \tag{87}$$

From the above equation we see that for a given $\mathbf{z}$ we can freely choose $\{w_j\}_{j=2}^n$ and $\{a_{1\ell}\}_{\ell=r-i}^{r-1}$ (the choice of which determines $\{a_{1\ell}\}_{\ell=0}^{r-i-1}$). Since, $t_{1\ell} \in \mathbb{Z}_p, w_j \in \mathbb{Z}_{p^r}$ and each row of the matrix A can be chosen independently we have $(p^{r(n-1)} p^i)^k$ choices for the matrix A. Hence, the lemma follows.

Appendix C
Proof of Lemma 4

We will first show that

$$A^n_{\varepsilon}(X) \ \cap \ C_{\mathbf{y}} \ \subseteq \ A^n_{\varepsilon}(X|\mathbf{y}) \ . \tag{88}$$

Let $\mathbf{x} \in A^n_{\varepsilon}(X) \ \cap \ C_{\mathbf{y}}$. Say $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$. Since $\mathbf{x} \in C_{\mathbf{y}}$ we have $x_k \bmod p^i = y_k, 1 \le k \le n$. Then, for any $a \in \mathbb{Z}_{p^r}, b \in \mathbb{Z}_{p^i}$,

$$N(a, b|\mathbf{x}, \mathbf{y}) = \begin{cases} N(a|\mathbf{x}), & a \bmod p^i = b \\ 0, & a \bmod p^i \neq b \end{cases} \ , \tag{89}$$

$$P(a,b) = \begin{cases} P(a), & a \bmod p^i = b \\ 0, & a \bmod p^i \neq b \end{cases} . \tag{90}$$

If $a \bmod p^i = b$,

$$\left|\frac{N(a,b|\mathbf{x},\mathbf{y})}{n} - P(a,b)\right| \overset{(a)}{=} \left|\frac{N(a|\mathbf{x})}{n} - P(a)\right|$$

$$\overset{(b)}{\leq} \varepsilon P(a) \tag{91}$$

$$\overset{(c)}{=} \varepsilon P(a,b)\ . \tag{92}$$

If $a \bmod p^i \neq b$,

$$\left|\frac{N(a,b|\mathbf{x},\mathbf{y})}{n} - P(a,b)\right| \overset{(a)}{=} |0-0| \tag{93}$$

$$\overset{(b)}{=} \varepsilon.0 \tag{94}$$

$$\overset{(c)}{=} \varepsilon P(a,b)\ , \tag{95}$$

where (a) follows from (89) and (90), (b) follows from (77) and the assumption that $\mathbf{x} \in A_\varepsilon^n(X)$ and (c) follows from (90). Hence from (92) and (95) and from the definition of typical sets we have $\mathbf{x} \in A_\varepsilon^n(X|\mathbf{y})$.

We will now show that $A_\varepsilon^n(X|\mathbf{y}) \subseteq A_\varepsilon^n(X)\ \cap\ C_\mathbf{y}$. Let $\mathbf{x} \in A_\varepsilon^n(X|\mathbf{y})$. Then, $(\mathbf{x},\mathbf{y}) \in A_\varepsilon^n(X,[X]_i)$. Thus, $\mathbf{x} \in A_\varepsilon^n(X)$. Say $\mathbf{x} = (x_1,\ldots,x_n)$ and $\mathbf{y} = (y_1,\ldots,y_n)$. We claim, $\mathbf{x} \in C_\mathbf{y}$. Since, if it is not true then for some $k, 1 \leq k \leq n$, $x_k \bmod p^i \neq y_k$. Hence,

$$N(x_k,y_k|\mathbf{x},\mathbf{y}) \ \geq\ 1\ . \tag{96}$$

However, since $(\mathbf{x},\mathbf{y}) \in A_\varepsilon^n(X,[X]_i)$, from we have,

$$\left|\frac{N(x_k,y_k|\mathbf{x},\mathbf{y})}{n} - P(x_k,y_k)\right| \ \leq\ \varepsilon P(x_k,y_k)\ . \tag{97}$$

Since $x_k \bmod p^i \neq y_k$, from (90) we get

$$P(x_k,y_k) \ =\ 0\ . \tag{98}$$

Substituting (98) in (97) we get $N(x_k,y_k|\mathbf{x},\mathbf{y}) = 0$ which contradicts (96). Hence, $\mathbf{x} \in C_\mathbf{y}$. We have shown that if $\mathbf{x} \in A_\varepsilon^n(X|\mathbf{y})$ then $\mathbf{x} \in C_\mathbf{y}$ and $\mathbf{x} \in A_\varepsilon^n(X)$ i.e.,

$$A_\varepsilon^n(X|\mathbf{y}) \ \subseteq\ A_\varepsilon^n(X)\ \cap\ C_\mathbf{y}\ . \tag{99}$$

Hence, from (88) and (99), the statement of the Lemma follows. □

Received 4 February 2011; Revised 28 February 2011

K. Vinodh received the B.E. degree in Electronics and Communication Engineering from the PSG College of Technology, Coimbatore, in 2003 and the M.Sc.(Engg.) degree from the Indian Institute of Science (IISc), Bangalore, in 2008. He is currently a Ph.D. student at the IISc, Bangalore. His research interests include distributed function computation, space-time codes and cooperative communications.

V. Lalitha received her Master's in Signal Processing from the Indian Institute of Science, Bangalore in 2005. She is currently pursuing her Ph.D. at the Indian Institute of Science, Bangalore. Her research interests include information theory, wireless sensor networks and coding theory.

N. Prakash received his Masters in Communication Engineering from the Indian Institute of Technology, Madras in 2006. He is currently working towards his doctoral degree at the Indian Institute of Science, Bangalore. His research interests include Information and coding theory, and their applications to wireless communication systems.

P. Vijay Kumar received the B.Tech. and M.Tech. degrees from the Indian Institutes of Technology (Kharagpur and Kanpur),and the Ph.D. Degree from the University of Southern California (USC) in 1983, all in Electrical Engineering. From 1983-2003 he was on the faculty of the EE-Systems Department at USC. Since 2003 he has been on the faculty of the Indian Institute of Science, Bangalore and also holds the position of adjunct research professor at USC. His current research interests include codes for distributed storage, distributed function computation, sensor networks and space-time codes for MIMO and cooperative communication networks. He is a fellow of the IEEE and an ISI highly-cited author. He is co-recipient of the 1995 IEEE Information Theory Society prize paper award as well as of a best paper award at the DCOSS 2008 conference on sensor networks.

S. Sandeep Pradhan obtained his M.E. degree from the Indian Institute of Science (IIS), India in 1996 and Ph.D. from the University of California at Berkeley in 2001. From 2002 to 2008 he was an assistant professor in the Department of Electrical Engineering and Computer Science at the University of Michigan at Ann Arbor. Currently he is an associate professor. He is the recipient of 2001 Eliahu Jury award given by the University of California at Berkeley for outstanding research in the areas of systems, signal processing, communications and control, the CAREER award given by the National Science Foundation (NSF), and the Outstanding achievement award for the year 2009 from the University of Michigan. His research interests include sensor networks, multi-terminal communication systems, coding theory, quantization, information theory.