Interactive Data Compression

Abbas El Gamal † and Alon Orlitsky †

Information Systems Laboratory
Electrical Engineering Department
Stanford University
Stanford, CA 94305

Abstract

Let X and Y be two random variables with probability distribution $p(x,y)$, joint entropy $H(X,Y)$ and conditional entropies $H(X|Y)$ and $H(Y|X)$. Person P_X knows X and person P_Y knows Y. They communicate over a noiseless two-way channel so that both know X and Y.

It is proved that, on the average, at least $H(X|Y)+H(Y|X)$ bits must be exchanged and that $H(X,Y)+2$ bits are sufficient. If $p(x,y)>0$ for all (x,y), then at least $H(X,Y)$ bits must be communicated on the average. However, if $p(x,y)$ is uniform over its support set, the average number of bits needed is close to $H(X|Y)+H(Y|X)$. Randomized protocols can reduce the amount of communication considerably but only when some probability of error is acceptable.

1. Introduction

Shannon's data compression theorem [1] states that if X is a random variable with entropy†† $H(X)$, then any variable length code that can communicate X over a noiseless channel must have expected length $\geq H(X)$. Moreover, codes with expected length $< H(X)+1$ exist. Later, Huffman [2] devised an elegant construction for optimal codes that achieve minimum expected code length.

In this paper we investigate the following two-way generalization of Shannon's data compression problem. Let X and Y be two random variables distributed over a finite set $\mathcal{X}\times\mathcal{Y}$ with joint entropy $H(X,Y)$, marginal entropies $H(X)$ and $H(Y)$, and conditional entropies $H(X|Y)=H(X,Y)-H(Y)$ and $H(Y|X)=H(X,Y)-H(X)$. Suppose that person P_X knows X and person P_Y knows Y. They communicate over noiseless two-way channel so that both know X and Y. How many bits on the average must they exchange? and what are the optimal codes?

† Work partially supported under NSF Grant ECS83-00988.

†† $H(X)=-\sum p(x)\cdot\log_2 p(x)$ where $p(x)$ is the probability that $X=x$.

We prove that, on the average, at least $H(X|Y)+H(Y|X)$ bits must be exchanged and that $H(X,Y)+2$ bits are sufficient. We also show that if $p(x,y)>0$ for all (x,y), then at least $H(X,Y)$ bits must be communicated on the average. However, if $p(x,y)$ is uniform over its support set, the average number of bits needed is close to $H(X|Y)+H(Y|X)$. The first of the last two results is somewhat disappointing as it precludes the search for efficient interactive data compression schemes in such cases. Yet the last result provides data compression schemes that may require considerably less than $H(X,Y)$ bits when the support set of $p(x,y)$ is a small subset of $\mathcal{X}\times\mathcal{Y}$. The following example illustrates one such case:

Suppose that each person has an n bit file and that the two files are known to differ in no more than K bits. Can they exchange the files using less than the obvious $n+\log(\sum_{k=0}^{K}\binom{n}{k})$ bits?

The lower bound can be used to show that at least $2\cdot\log(\sum_{k=0}^{K}\binom{n}{k})$ bits must be exchanged *in the worst case*. An upper bound on the number of bits exchanged (Theorem 4) ensures that $2\cdot\log(\sum_{k=0}^{K}\binom{n}{k})+\log n$ bits are always enough. The two bounds are asymptotically tight for every K (for more details see Example 5).

In the following section we formally define the two way data compression problem. In section 3, we prove general lower and upper bounds. In section 4 we prove the upper bound results for distributions that are uniform over their support set. In the last section we compare the performance of deterministic and randomized protocols. We show that if no errors are allowed then randomization doesn't help. If some probability of error is allowed then the number of bits required by a randomized scheme can be logarithmically smaller than that achieved by any deterministic one.

2. Definitions

In this section we introduce the communication model and define the complexity measures corresponding to the number of bits communicated. We begin by describing operations on sequences.

Let $a_1, \ldots, a_n$ be arbitrary elements. $<a_1 \cdots a_n>$ or, equivalently, $<a_i>_{i=1}^n$ denote the sequence consisting of these elements ($<a_i>_{i=1}^0$ is the empty sequence - consisting of 0 elements). If $A_1, \ldots, A_n$ are sequences, then $[A_1]$ is defined to be A_1 and, for $2 \leq m \leq n$, $[A_1 A_2 \cdots A_m]$ recursively denotes the sequence whose elements are the elements of $[A_1 A_2 \cdots A_{m-1}]$ followed by the elements of A_m. $[A_i]_{i=1}^m$ is an abbreviation for $[A_1 A_2 \cdots A_m]$ and $[A_i]_{i=1}^0$ denotes the empty sequence. Note that $[A_1]$ is the sequence A_1 whereas $<A_1>$ is the sequence whose only element is A_1 and, if A_2 is the empty sequence then $[A_1 A_2]=A_1$. If A is a sequence, let $|A|$ denote its length. Thus, $|<A_i>_{i=1}^n|=n$ while $|[A_i]_{i=1}^n|=\sum_{i=1}^n |A_i|$.

A sequence $<a_i>_{i=1}^m$ is said to be *a prefix* of a sequence $<b_i>_{i=1}^n$ if $m \leq n$ and for $i=1,\ldots,m$, $a_i=b_i$. It is said to be a *proper prefix* if, in addition, $m \neq n$. A set of sequences is said to be prefix free if no sequence in the set is a (proper) prefix of another.

The communication model we consider is the *Generalized Discrete Time Binary Channel.* A *message* for this channel is a finite sequence of bits (possibly, the empty sequence). At any time unit, both communicators can simultaneously transmit messages of arbitrary lengths. A *transmission descriptor* is an ordered pair of messages. A *codeword* is a finite sequence of transmission descriptors.

Let C be a function from a subset S of $\mathcal{X} \times \mathcal{Y}$ to the set of codewords. Then, $n(x,y)$ denotes the length of $C(x,y)$ (the number of transmission descriptors in the sequence). For $i=1,..,n(x,y)$, $C_i(x,y)$ denotes the i th transmission descriptor in $C(x,y)$; $b_i^X(x,y)$ denotes the first message in $C_i(x,y)$ and $b_i^Y(x,y)$ - the second. $C_1^i(x,y)$ is an abbreviation for $<C_j(x,y)>_{j=1}^i$. $C_1^0(x,y)$ is the empty sequence (and $C_1^{n(x,y)}(x,y)$ is just another name for $C(x,y)$).

$B^X(x,y)$ is the sequence $[\, b_j^X(x,y)\,]_{j=1}^{n(x,y)}$, $B^Y(x,y)$ is $[\, b_j^Y(x,y)\,]_{j=1}^{n(x,y)}$; $b_i^{XY}(x,y)$ is the sequence $[\, b_i^X(x,y), b_i^Y(x,y)\,]$ and $B^{XY}(x,y)$ is the sequence $[\, b_j^{XY}(x,y)\,]_{j=1}^{n(x,y)}$.

The mapping C is said to be a *Generalized Binary Channel Code* (*G - code*) for S if it satisfies the following properties:

Prefix free messages.
For all $(x,y), (x,y') \in S$, $1 \leq i \leq \min(n(x,y), n(x,y'))$, $C_1^{i-1}(x,y) = C_1^{i-1}(x,y')$ implies that $b_i^Y(x,y)$ is not a proper prefix of $b_i^Y(x,y')$ and
for all $(x,y), (x',y) \in S$, $1 \leq i \leq \min(n(x,y), n(x',y))$, $C_1^{i-1}(x,y) = C_1^{i-1}(x',y)$ implies that $b_i^X(x,y)$ is not a proper prefix of $b_i^X(x',y)$.

Coordinated Termination.
For all $x \in X$, the set $\{C(x,y) : (x,y) \in S\}$ is prefix free and for all $y \in Y$, the set $\{C(x,y) : (x,y) \in S\}$ is prefix free.
(Note that (x,y), (x,y') can have the *same* codeword.)

Unique message.
For all $(x,y), (x,y') \in S$, $1 \leq i \leq \min(n(x,y), n(x,y'))$, $C_1^{i-1}(x,y) = C_1^{i-1}(x,y')$ implies $b_i^X(x,y)=b_i^X(x,y')$ and
for all $(x,y), (x',y) \in S$, $1 \leq i \leq \min(n(x,y), n(x',y))$, $C_1^{i-1}(x,y) = C_1^{i-1}(x',y)$ implies $b_i^Y(x,y)=b_i^Y(x',y)$.

The prefix free messages property ensures that the receiver knows how to interpret a received message (and that the length of the message is not used to transfer information). The coordinated termination property ensures that the communicators know when the communication ends. The unique message property ensures that the communication is "deterministic" i.e. the same inputs will always result in the same bits communicated. (See section 5 for randomized protocols).

Remark: One could avoid the coordinated termination property and shorten the description of the others by defining codewords to be infinitely long with only finitely many non empty messages. We, however, prefered the more intuitive, finite length messages.

Let $p(x,y)$ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$. Denote the marginal probability of $x \in \mathcal{X}$ by $p(x)$ and the marginal probability of $y \in \mathcal{Y}$ by $p(y)$. The support set S_p of p is defined by $S_p \triangleq \{(x,y) : p(x,y) > 0\}$. A code C is said to be a code for p if it is a code for S_p.

If C is a code for p, define the average length of C under p to be:

$$L_a(C,p) \triangleq \sum_{S_p} p(x,y) \cdot |B^{XY}(x,y)|$$

and the maximal length of C under p to be:

$$L_m(C,p) \triangleq \max_{S_p} \{|B^{XY}(x,y)|\} .$$

Let f be a function defined on S. A code C is said to *resolve* f for S if for all $(x,y), (x,y')\in S$, $C(x,y)=C(x,y')$ implies $f(x,y)=f(x,y')$ and for all $(x,y), (x',y)\in S$, $C(x,y)=C(x',y)$ implies $f(x,y)=f(x',y)$.

The codes discussed in this paper resolve the identity function $f(x,y) = (x,y)$. We call such codes *exchange codes.* The average complexity of a distribution p is defined as

$$L_a(p) \triangleq \min_{\{C:\ C \text{ is an exchange code for } p\}} L_a(C,p)$$

The maximal complexity of a distribution p is defined as

$$L_m(p) \triangleq \min_{\{C:\ C \text{ is an exchange code for } p\}} L_m(C,p)$$

$L_a(p)$ is the minimal number of bits that have to be exchanged on the average in order to exchange X and Y using any code that obeys the above properties. $L_m(p)$ has a similar interpretation.

3. Lower And Upper Bounds

We begin by proving some basic properties of codes. These properties follow directly from the definition of G-codes.

Lemma 1. Let C be a code for S. For every $(x,y), (x',y')\in S$, if $(x,y')\in S$ and one of $B^{XY}(x,y)$, $B^{XY}(x',y')$ is a prefix of the other then,

$$B^{XY}(x,y) = B^{XY}(x,y') = B^{XY}(x',y').$$

Proof: Without loss of generality, assume that $B^{XY}(x,y)$ is a prefix of $B^{XY}(x',y')$.
By definition, $C_1^0(x,y) = C_1^0(x,y') = C_1^0(x',y')$.
Also, if for some $1 \leq i \leq n(x,y)$,
$C_1^{i-1}(x,y)=C_1^{i-1}(x,y')=C_1^{i-1}(x',y')$ then,

(i) $[\, b_j^{XY}(x,y)\,]_{j=i}^{n(x,y)}$ is a prefix of $[\, b_j^{XY}(x',y')\,]_{j=i}^{n(x',y')}$.

(ii) By the coordinated termination property, $i\leq n(x,y')$ and, therefore, $i\leq n(x',y')$.

From (ii) and the unique message property, $b_i^X(x,y) = b_i^X(x,y')$. Hence, both $b_i^X(x,y')$ and $b_i^X(x',y')$ are prefixes of $[\, b_j^{XY}(x',y')\,]_{j=i}^{n(x',y')}$ and, thus, one is a prefix of the other. By the prefix free messages property, $b_i^X(x,y') = b_i^X(x',y')$.
Similarly, $b_i^Y(x,y) = b_i^Y(x,y') = b_i^Y(x',y')$.
Thus, $C_1^i(x,y)=C_1^i(x,y')=C_1^i(x',y')$ and, by induction, $C_1^{n(x,y)}(x,y)=C_1^{n(x,y)}(x,y')=C_1^{n(x,y)}(x',y')$.
From the coordinated termination property, $C(x,y) = C(x,y') = C(x',y')$ □

Corollary 1. Let C be an exchange code for S. If $(x,y), (x',y')$ are distinct members of S and $(x,y')\in S$ then neither one of $B^{XY}(x,y)$, $B^{XY}(x',y')$ is a prefix of the other (nor can they be equal).

Proof: If one is a prefix of the other then, from Lemma 1, $C(x,y) = C(x,y') = C(x',y')$. Since C is an exchange code, this implies that $(x,y) = (x,y') = (x',y')$ which contradicts the assumption □

Corollary 2. If C is an exchange code for S then for every $x\in\mathcal{X}$, $\{B^Y(x,y) : (x,y)\in S\}$ is prefix free with cardinality $|\{y : (x,y)\in S\}|$ and for every $y\in\mathcal{Y}$, $\{B^X(x,y) : (x,y)\in S\}$ is prefix free with cardinality $|\{x : (x,y)\in S\}|$.

Proof: Assume that for some $(x,y), (x,y')\in S$, $y\neq y'$, $B^Y(x,y)$ is a prefix of $B^Y(x,y')$. Then, by induction as in Lemma 1, $C(x,y)$ is a prefix of $C(x,y')$ contradicting Corrolary 1 □

Using these properties, we can apply one-way-communication results to prove the following lower and upper bounds on $L_a(p)$.

Theorem 1: $H(X|Y) + H(Y|X) \leq L_a(p) \leq H(X,Y)+2$.

Proof: Upper Bound: Using a Huffman code [2], P_X encodes X with average length $\leq H(X)+1$. Then, P_Y, knowing X, encodes Y with average length $\leq H(Y|X)+1$.
Lower Bound: By Corollary 2, if C is an exchange code for p then for every $x\in\mathcal{X}$, $\{B^Y(x,y) : (x,y)\in S_p\}$ is a prefix free code [3] for $\{y : (x,y)\in S_p\}$.
Therefore, for every $x\in\mathcal{X}$,

$$\sum_{\{y\,:\,(x,y)\in S_p\}} p(y|x)\cdot|B^Y(x,y)| \geq H(Y|X=x).$$

Similarly, for every $y\in\mathcal{Y}$,

$$\sum_{\{x\,:\,(x,y)\in S_p\}} p(x|y)\cdot|B^X(x,y)| \geq H(X|Y=y). \quad \text{Thus,}$$

$$\begin{aligned}
L_a(C,p) &= \sum_{(x,y)\in S_p} p(x,y)\cdot|B^{XY}(x,y)| \\
&= \sum_{(x,y)\in S_p} p(x,y)\cdot|B^X(x,y)| + \sum_{(x,y)\in S_p} p(x,y)\cdot|B^Y(x,y)| \\
&= \sum_{y\in\mathcal{Y}} p(y) \sum_{\{x\,:\,(x,y)\in S_p\}} p(x|y)\cdot|B^X(x,y)| \\
&\quad + \sum_{x\in\mathcal{X}} p(x) \sum_{\{y\,:\,(x,y)\in S_p\}} p(y|x)\cdot|B^Y(x,y)| \\
&\geq \sum_{y\in\mathcal{Y}} p(y)\cdot H(X|Y=y) + \sum_{x\in\mathcal{X}} p(x)\cdot H(Y|X=x) \\
&= H(X|Y)+H(Y|X) \quad \square
\end{aligned}$$

The following theorem shows that the two way Huffman scheme described in Theorem 1 is nearly optimal when $p(x,y)>0$ for all $(x,y)\in \mathcal{X}\times\mathcal{Y}$.

Theorem 2. If for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$, $p(x,y)>0$ then,

$$H(X,Y) \le L_a(p) \le H(X,Y)+2.$$

Proof: Upper Bound: As in Theorem 1.
Lower Bound: Let C be an exchange code for p. Since $S_p=\mathcal{X}\times\mathcal{Y}$, every (x,y), $(x',y')\in\mathcal{X}\times\mathcal{Y}$ satisfy the requirements of Corollary 1. Hence, $\{B^{XY}(x,y)\}$ is a prefix free code for $\mathcal{X}\times\mathcal{Y}$ and,

$$L_a(p) \triangleq \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\cdot|B^{XY}(x,y)| \ge H(X,Y) \quad \square$$

According to Theorem 1, $L_a(p)$ lies between $H(X|Y)+H(Y|X)$ and $H(X,Y)+2$. The following examples describe distributions of complexities achieving the lower bound (Example 1), at most two bits less than the upper bound (Example 2) and strictly in between (Example 3).

Example 1. ($L_a(p)=0=H(X|Y)+H(Y|X)$)
Let $\mathcal{X}=\mathcal{Y}=\{1,\ldots,n\}$ and

$$p(x,y)=\begin{cases}1 & x=y\\ 0 & x\neq y\end{cases}.$$

Then, $X=Y$ so no bit needs to be exchanged.

Example 2.
($L_a(p)\ge \log n + h(\epsilon)+\epsilon\cdot\log(n-1) = H(X,Y)$)
Let $\mathcal{X}=\mathcal{Y}=\{1,\ldots,n\}$, $\epsilon>0$ and

$$p(x,y)=\begin{cases}\frac{1-\epsilon}{n} & x=y\\ \frac{\epsilon}{n(n-1)} & x\neq y\,.\end{cases}$$

Then, by Theorem 2 $L_a(p)=H(X,Y)$.

Example 3.
($L_a(p)$ strictly between $H(X|Y)+H(Y|X)$ and $H(X,Y)$)
Let $\mathcal{X}=\mathcal{Y}=\{1,\ldots,2n\}$, $\epsilon>0$ and

$$p(x,y)=\begin{cases}P\cdot\frac{1}{n} & x<n\ y<n\ x=y\\ \overline{P}\cdot\frac{1-\epsilon}{n} & x>n\ y>n\ x=y\\ \overline{P}\cdot\frac{\epsilon}{2n(n-1)} & x>n\ y>n\ x\neq y\\ 0 & \text{otherwise}\,.\end{cases}$$

Then,

$$H(X|Y)+H(Y|X)=2\cdot\overline{P}\cdot(\,h(\epsilon)+\epsilon\cdot\log(n-1)\,),\ \text{and}$$
$$H(X,Y)=\log n+\overline{P}\cdot h(\epsilon)+\overline{P}\cdot\epsilon\cdot\log(n-1)+h(p)\,.$$

Combining the previous examples we obtain

$$\overline{P}\cdot\big(\log n+h(\epsilon)+\epsilon\cdot\log(n-1)\big) \le L_a(p)$$
$$\le \overline{P}\cdot\big(\log n+h(\epsilon)+\epsilon\cdot\log(n-1)\big)+2\quad.$$

By letting $\epsilon\to 0$,

$$H(X|Y)+H(Y|X)\approx 0$$
$$H(X,Y)\approx\log n$$
$$L_a(p)\approx\overline{P}\cdot\log n$$

Where $A\approx B$ means that $|A-B|\le 2$.

4. Distributions For Which $L_a(p)$ Is Close To $H(X|Y)+H(Y|X)$

In this section, we show that for almost all distributions p which are uniform over a subset of $\mathcal{X}\times\mathcal{Y}$, $L_a(p)$ is very close to $H(X|Y)+H(Y|X)$. As a first result, we have the following.

Lemma 2. [4] If for all $x\in\mathcal{X}$, $|\{y: p(x,y)>0\}|\le n$ and for all $y\in\mathcal{Y}$, $|\{x: p(x,y)>0\}|\le m$; then,

$$L_m(p)\le\lceil\log(m\cdot n)\rceil+\lceil\log(\min(m,n))\rceil\ .$$

Proof: Without loss of generality, assume $m\le n$. Create a graph $G=<V,E>$ with $V=\{y: p(y)>0\}$ and $E=\{(y_1,y_2): y_1\neq y_2$ and for some x, $p(x,y_1)>0$ and $p(x,y_2)>0\}$. Clearly, the degree of each vertex is at most $m\cdot(n-1)\le m\cdot n-1$. Therefore there exists a vertex coloring of G using $\le m\cdot n$ colors. P_X and P_Y agree in advance on such a coloring. P_Y transmits the color of Y (using $\le\lceil\log mn\rceil$ bits). With this information, P_X knows Y. He, then, sends P_Y the index of X in the set $\{x: p(x,Y)>0\}$ (using $\le\lceil\log m\rceil$ bits) $\square$

To improve the result of this Lemma, we need some results concerning hypergraph partitioning.

Lemma 3: Let V be a set of even size $v>0$ and $\{E_i\}$, $i=1,\ldots,e$ a collection of subsets of V such that $|E_i|\le m$. Then, there exists a partition X, $\overline{X}$ of V such that $|X|=|\overline{X}|=v/2$ and, for $i=1,\ldots,e$,

$$|X\cap E_i|,\ |\overline{X}\cap E_i| < \frac{m}{2}+\sqrt{m\cdot\ln(e\cdot\sqrt{mv})}.$$

Proof: Without loss of generality, assume $|E_i|=m$ for all i. Denote $\sqrt{\frac{m}{2}\ln(e\cdot\sqrt{mv})}$ by α. Call a subset of V a *half subset* if its cardinality is $v/2$. We prove that if $\lfloor\frac{m}{2}-\alpha\rfloor \geq 2$ then there exists a half subset X of V such that for all $0\leq i\leq e$, $\frac{m}{2}-\alpha<|X\cap E_i|<\frac{m}{2}+\alpha$ †.

It is easy to show that the number of half subsets of V that *don't* have the above property is at most $e\cdot 2^{v-m}\lfloor\frac{m}{2}-\alpha\rfloor\binom{m}{\lfloor m/2-\alpha\rfloor}$ while the number of half subsets is $\binom{v}{v/2}$.

In the rest of the proof, we show that

$$\binom{v}{v/2} > e\cdot 2^{v-m}\lfloor\frac{m}{2}-\alpha\rfloor\binom{m}{\lfloor m/2-\alpha\rfloor}$$

so there must be at least one half subset with the required properties.

By expanding $1-h\left(\frac{1}{2}-x\right)$ around $x=0$, we get

$$1-h\left(\frac{1}{2}-x\right) = \sum_{k=2,4,..}\frac{(2x)^k}{(ln2)}\cdot k(k-1) > \frac{2x^2}{\ln 2}.$$

Thus,

$$m\cdot(1-h(\frac{1}{2}-\frac{\alpha}{m})) > \log\,(e\cdot\sqrt{mv}).$$

Raising both sides to to the power of 2,

$$2^{m\cdot(1-h(\frac{1}{2}-\frac{\alpha}{m}))} > e\cdot\sqrt{mv}$$

$$> 2e\sqrt{v}\lfloor m/2-\alpha\rfloor\cdot\sqrt{\frac{m}{2\pi\lfloor m/2-\alpha\rfloor\cdot\lceil m/2+\alpha\rceil}}$$

Using the right hand side of the inequality [3]:

$$\sqrt{\frac{n}{8\cdot j\cdot(n-j)}} \leq \binom{n}{j}2^{-n\cdot h(j/n)} < \sqrt{\frac{n}{2\cdot\pi\cdot j\cdot(n-j)}} \quad (1)$$

we obtain

$$2^m > 2e\sqrt{v}\lfloor\frac{m}{2}-\alpha\rfloor\binom{m}{\lfloor m/2-\alpha\rfloor}$$

And the other side of (1) yields:

$$\binom{v}{v/2} \geq 2^v\cdot\sqrt{\frac{v}{8\cdot(v/2)\cdot(v/2)}} > 2^{v-m}\cdot e\lfloor\frac{m}{2}-\alpha\rfloor\cdot\binom{m}{\lfloor m/2-\alpha\rfloor} \quad \square$$

Lemma 4: Let g be a real valued continuous function and $a>0$ such that for some $\delta,c,\epsilon>0$, $g(x)\geq\max\{2+\delta\,,\,c\cdot ln^{1+\epsilon}x\}$ for all $x\geq a$.

Then the sequence $\{a_i\}_{i=0}^{\infty}$ given by, $a_o \triangleq a$, $a_{i+1}\triangleq\max\{x\!: a_i = \frac{x}{2}+\frac{x}{g(x)}\}$ is well defined and satisfies

(i) $a<a_i<a_{i+1}<2a_i$

(ii) there exists a constant $b>0$ such that $\frac{a_i}{a}>\frac{2^i}{b}$.

Proof: See [5] $\square$

We now combine the last 2 lemmas to prove the main result of this section.

Theorem 3: Let V be a set of size v and for $i=1,...,e$ $E_i\subseteq V$ and $|E_i|\leq m$.
Then, given $\epsilon>0$, there exists $C(\epsilon)$ such that for all $p\geq(\ln\sqrt{ev})^{1+\epsilon}$, $p>1$ it is possible to find a partition $V_1,\ldots,V_{\lceil C(\epsilon)\cdot m/p\rceil}$ of V such that $|V_j\cap E_i|<p$ for $i=1,...,e$ $j=1,...,\lceil C(\epsilon)\cdot m/p\rceil$.

Proof: Assume first that v is a power of 2. Let $\epsilon'\triangleq\frac{\epsilon}{2+2\epsilon}$ and define $<n_i>_{i=0}^{\infty}$ recursively:

$$n_0 = \lfloor 2\cdot 2^{1/\epsilon'}\rfloor = \lfloor 8\cdot 4^{1/\epsilon}\rfloor$$

$$n_{k+1}\triangleq\max\{x: n_k=\frac{x}{2}+\frac{x}{x^{\epsilon'}}\}.$$

By Lemma 4, there exists $C'(\epsilon)$ such that for all k, $\frac{n_k}{n_0}>\frac{2^k}{C'(\epsilon)}$.

Let $C''(\epsilon)\triangleq\max\{\,2\cdot C'(\epsilon), \exp\,(16\cdot 4^{1/\epsilon})\,\}$. We show that $C''(\epsilon)$ satisfies the requirements of the theorem (when v is a power of 2).

We distinguish between two cases:

I. $p\leq 8\cdot 4^{1/\epsilon}$.
In this case, $(ln\sqrt{ev})^{1+\epsilon}\leq p\leq 8\cdot 4^{1/\epsilon}$. If $m<p$, take $V_1=V$, $V_2=...=V_{\lceil C''(\epsilon)\frac{m}{p}\rceil}\triangleq\Phi$. If $m\geq p$ then $\ln(ev)/2\leq(ln\sqrt{ev})^{1+\epsilon}\leq 8\cdot 4^{1/\epsilon}$ implies $v\leq ev\leq\exp(16\cdot 4^{1/\epsilon})\leq C''(\epsilon)\leq\lceil C''(\epsilon)\cdot\frac{m}{p}\rceil$ so let each of $V_1,\ldots,V_v$ consist of a single element of V and the rest of the V_i's be empty.

II. $p>8\cdot 4^{1/\epsilon}$.
Let $m_0=p$, $m_{k+1}\triangleq\max\{x: m_k=\frac{x}{2}+\frac{x}{x^{\epsilon'}}\}$.
For all $x\geq m_0$, $x^{\epsilon'}\geq m_0^{\epsilon'}>(2^{1/\epsilon'})^{\epsilon'}=2$ so, by Lemma 4, m_k is well defined. First, we show by induction on k that:

† This is actually stronger than the claim of the lemma ($\frac{m}{2}-\sqrt{2}\cdot\alpha<|X\cap E_i|<\frac{m}{2}+\sqrt{2}\cdot\alpha$). The extra $\sqrt{2}$ factor takes care of the case $v=2$, $m=e=1$. Note that we omitted the proof for $\lfloor\frac{m}{2}-\alpha\rfloor<2$.

If $p > 8\cdot 4^{1/\epsilon}$ and $m<m_k$ then there exists a partition $V_1, V_2,..., V_{2^k}$ of V such that $|V_j \bigcap E_i| < p$.
Induction basis: If $m<m_0$ then $m<p$, so $V_1=V$ will do.
Induction step: If $m_k \leq m < m_{k+1}$ then $m \geq m_0 = p \geq (ln\sqrt{ev})^{1+\epsilon}$. Therefore,

$$\frac{m}{m^{\epsilon'}} = \sqrt{mm^{\frac{1}{2+2\epsilon}}} \geq \sqrt{m(ln\sqrt{ev})^{\frac{1+\epsilon}{2+2\epsilon}}} = \sqrt{m\cdot \ln\sqrt{ev}}.$$

By Lemma 3, V can be partitioned into two sets X_1, X_2 such that for $j=1, 2$

$$|X_j| = \frac{v}{2} \text{ and },$$

$$|X_j \bigcap E_i| < \frac{m}{2} + \sqrt{m\cdot \ln\sqrt{ev}} \leq \frac{m}{2} + \frac{m}{m^{\epsilon'}}$$

$$< \frac{m_{k+1}}{2} + \frac{m_{k+1}}{m_{k+1}^{\epsilon'}} = m_k .$$

Now, $|X_1| = \frac{v}{2}$; $|X_1 \bigcap E_i| < m_k$, and, still, $p > 8\cdot 4^{1/\epsilon}$ so, by induction hypothesis, there exists a partition $V_1^1, ..., V_{2^k}^1$ of X_1 such that $|V_j^1 \bigcap (X_1 \bigcap E_i)| < p$ for $1 \leq j \leq 2^k$, $1 \leq i \leq e$. Similarly, there exists a partition $V_1^2, ..., V_{2^k}^2$ of X_2 such that $|V_j^2 \bigcap (X_2 \bigcap E_i)| < p$ for $1 \leq j \leq 2^k$, $1 \leq i \leq e$. Therefore, define

$$V_j = \begin{cases} V_j^1 & \text{for } 1 \leq j \leq 2^k \\ V_{j-2^k}^2 & \text{for } 2^k+1 \leq j \leq 2^{k+1} \end{cases}$$

to get a partition $V_1, \ldots, V_{2^{k+1}}$ of V such that $|V_j \bigcap E_i| < p$. This proves the induction.

Next, we show that for all k, $\frac{m_k}{m_0} \geq \frac{n_k}{n_0}$.
Since $m_0 = p \geq 8\cdot 4^{1/\epsilon} \geq n_0$ and $\frac{x}{2} + x^{1-\epsilon'}$ is an increasing function of x, then by induction, $m_k \geq n_k$. Also,

$$\frac{m_k}{m_{k-1}} = \frac{1}{\frac{1}{2} + \frac{1}{m_k^{\epsilon'}}} \geq \frac{1}{\frac{1}{2} + \frac{1}{n_k^{\epsilon'}}} = \frac{n_k}{n_{k-1}}$$

thus, again by induction, $\frac{m_k}{m_0} \geq \frac{n_k}{n_0}$.
Finally, let k_0 be the first integer such that $m_{k_0} > m$ (k_0 exists since $m_k \to \infty$). It follows that

$$2^{k_0} < \frac{n_{k_0}}{n_0}\cdot C'(\epsilon) \leq \frac{m_{k_0}}{m_0}\cdot C'(\epsilon) = \frac{m_{k_0}}{p}\cdot C'(\epsilon)$$

and, from part (i) of Lemma 4, $m_{k_0} < 2\cdot m$.

Thus, $2^{k_0} < \frac{m}{p}\cdot 2\cdot C'(\epsilon) \leq \lceil \frac{m}{p}\cdot C''(\epsilon) \rceil$. Since $m < m_{k_0}$, the induction implies the existance of a partition $V_1, \ldots, V_{2^{k_0}}$ of V satisfying the requirements of the theorem. To get the right number of subsets, this partition need only be augmented by $\lceil \frac{m}{p}\cdot C''(\epsilon) \rceil - 2^{k_0}$ empty sets.

This completes the proof for v's that are a power of 2.

If v is not a power of 2, partition V into two sets one of which having size $2^{\lfloor \log v \rfloor}$ and extend the other set to have the same size. Use the result separately on each of the sets and combine the sets to get a partition with at most $2\lceil C''(\epsilon)\frac{m}{p} \rceil$ sets. Thus, $C(\epsilon) \triangleq 2.1\cdot C''(\epsilon)$ satisfies the requirements for all cases □

This theorem has the following hypergraph coloring interpretation: Let $(V, \{E_i\}_{i=1}^{e})$ be a hypergraph such that each edge contains at most m vertices. Given a number $p > (ln\sqrt{ev})^{1.1}$, there exists a $\lceil C\cdot\frac{m}{p} \rceil$ coloring of the vertices such that no more than p vertices in each edge have the same color. Note that $\lceil \frac{m}{p} \rceil$ is the minimum number required by any edge containing m vertices. Hence the number of colors required is never more than a constant times larger than that required by the largest edge.

If for every x, the number of possible y's is roughly the same, the theorem combined with Lemma 4 can be used to derive an exchange code with good maximal length:

Theorem 4. Given $\epsilon>0$ there exists $C(\epsilon)$ such that for all probability distributions p satisfying $|\{y : p(x,y)>0\}| \leq n$ for all $x \in \mathcal{X}$ and $|\{x : p(x,y)>0\}| \leq m$ for all $y \in \mathcal{Y}$,

$$L_m(p) \leq \log(m\cdot n) + (1+\epsilon)\cdot \log\log(max(|\mathcal{X}|,|\mathcal{Y}|)) + C(\epsilon).$$

Proof: Let $p = (\ln\sqrt{|\mathcal{X}|\cdot|\mathcal{Y}|})^{1+\epsilon}$ and denote $C(\epsilon)$ of Theorem 3 by $C'(\epsilon)$. P_X and P_Y agree on a partition $\mathcal{X}_1, \ldots, \mathcal{X}_{\lceil \frac{m}{p}\cdot C'(\epsilon) \rceil}$ of $\mathcal{X}$ such that for all y $|X_i \bigcap \{x : p(x,y) > 0\}| < p$ and on a similar partition $\mathcal{Y}_1, \ldots, \mathcal{Y}_{\lceil \frac{n}{p}\cdot C'(\epsilon) \rceil}$.
First, using $\leq \lceil \log\frac{m}{p}\cdot C'(\epsilon) \rceil$ bits, P_X transmits the index of the $\mathcal{X}$ subset that X is in. Then, P_Y transmits the index of his subset using $\leq \lceil \log\frac{n}{p}\cdot C'(\epsilon) \rceil$ bits. Now, they

can restrict themselves to a subset of $\mathcal{X}\times\mathcal{Y}$ with at most p non zero-probability elements for each $x \in \mathcal{X}$ and for each $y \in \mathcal{Y}$. By Lemma 2, $3\cdot\lceil \log p \rceil$ bits are enough to inform each of the other's value.
The total number of bits transmitted is at most

$$\lceil \log (\frac{m}{p}\cdot C'(\epsilon)) \rceil + \lceil \log(\frac{n}{p}\cdot C'(\epsilon)) \rceil + 3\cdot\lceil \log p \rceil$$

$$\leq \log(mn) + 2\cdot\log C'(\epsilon) + 5 + \log p$$

$$\triangleq \log(mn) + C(\epsilon) + \log \left(\ln \sqrt{|\mathcal{X}|\cdot|\mathcal{Y}|}\right)^{1+\epsilon}$$

$$\leq \log(mn) + C(\epsilon) + (1+\epsilon)\cdot\log\log(\max(|\mathcal{X}|, |\mathcal{Y}|)) \quad \square$$

The following two examples demonstrate the use of results obtained so far.

Example 4: Shifts. (Suggested by T. Cover)
Two persons have sequences that are cyclic shifts of each other. They wish to exchange these sequences. Formally, let $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$, $S \triangleq \{(x,y): x \text{ is a cyclic shift of } y\}$, and p_0 be the uniform probability distribution over S.
Since $S_p = S_{p'}$ implies $L_m(p) = L_m(p')$ we obtain, from Theorem 1 that for all probability distributions p with $S_p = S$, the *worst case* complexity $L_m(p) = L_m(p_0) \geq$ $\geq L_a(p_0) \geq H(X|Y) + H(Y|X) \geq 2\cdot(\log n - \frac{2\cdot\log n}{2^{n/2}})$ which is larger than $2\cdot\log n - 1$ for all $n \geq 8$. Since the number of possible y's for every x is $\leq n = \log|\mathcal{X}|$, Theorem 4, ensures that for all distributions p with $S_p = S$, $L_m(p) \leq (3+\epsilon)\cdot\log n + C(\epsilon)$. The two bounds are asymptotically tight. However, the upper bound is 1.5 times larger than the lower bound. The following scheme reduces the upper bound to within 3 bits above the lower bound: Let Z be the largest sequence among all cyclic shifts of X. Then Z is also the largest sequence among all cyclic shifts of Y. Both P_X and P_Y find Z. Then, P_X transmits to P_Y the number of times Z should be right shifted to obtain X (log n bits) and P_Y does the same.

Example 5: K errors.
In this example, described in the introduction, $\mathcal{X}=\mathcal{Y}=\{0,1\}^n$, $S=\{(x,y) : d_H(x,y)\leq K\}$. By theorem 4 and the discussion in Example 4, the *worst case* complexity satisfies $L_m(p) \geq 2\cdot\log(\sum_{k=0}^{K}\binom{n}{k})$ for all distributions p with support S. While, using Theorem 4, $L_m(p) \leq 2\cdot\log(\sum_{k=0}^{K}\binom{n}{k}) + \log n$. Again, the two bounds are asymptotically tight for every K. Moreover, for K's growing with n, the ratio between the upper and lower bounds approaches 1. However, for fixed values of K, there is a (very small) ratio between the two. For K=1, 2, 3 the ratios are 1.5, 1.25, 1.166 respectively. The following schemes achieve worst case complexities of $2\cdot(1+\log n)$ for $K=1$ and $2\cdot K\cdot\log n$ for K=2,3 Thus reducing the ratio between the upper and lower bounds to 1 for these cases. (For simplicity, we assume that n is a power of 2).

K=1. If n=1 then exchanging X,Y achieves $2\cdot(1+\log n)$. Assume that for sequences of length $n/2$ the algorithm has maximal length $2\cdot(1+\log n/2)$. Given a sequence of length n, P_X transmits to P_Y the parity of the first $n/2$ bits of X and P_Y transmits to P_X the parity of the first $n/2$ bits of Y. If the parities differ, P_X and P_Y know that there exists $1\leq i\leq n/2$ such that $X_i\neq Y_i$ and they use the $n/2$ algorithm on the subsequences $\langle X_i\rangle_{i=1}^{n/2}$ and $\langle Y_i\rangle_{i=1}^{n/2}$. If the parities are the same, there is at most one i, $n/2<i\leq n$ such that $X_i\neq Y_i$ so they use the algorithm on the subsequences $\langle X_i\rangle_{i=n/2+1}^{n}$ and $\langle Y_i\rangle_{i=n/2+1}^{n}$. In either case, the total number of bits is at most $2+2\cdot(1+\log n/2) = 2\cdot(1+\log n)$.

K=2. For m=1,...,$\log n$, let $A_m \triangleq \{i : 0\leq i<n$ and the m th least significant bit in the binary representation of i is 1 $\}$. For m=1,...,$\log n$, P_X transmits $\bigoplus_{i\in A_m} X_i$ to P_Y and P_Y transmits $\bigoplus_{i\in A_m} Y_i$ to P_X ($\oplus$ denotes exclusive or). Let B be the log n bit long binary number whose m th least significant bit is one iff the parities corresponding to A_m are different. If all parities are the same ($B=0$) then either $X=Y$ or $X_0\neq Y_0$ so P_X and P_Y exchange the 0 th bit to know which is the case. If the parities are not all equal, let M be any integer such that the parities for A_M differ (the M th least significant bit of B is one). There is at most one $i\notin A_M$ such that $X_i\neq Y_i$ so P_X and P_Y use the scheme for K=1 described above on the subsequences $\langle X_i\rangle_{i\notin A_M}$ and $\langle Y_i\rangle_{i\notin A_M}$. If they find that the two subsequences are equal then $X_B\neq Y_B$ and all other bits are the same. If they find that $X_C\neq Y_C$ say, then also, $X_{C\oplus B}\neq Y_{C\oplus B}$ and all the other bits are equal. The total number of bits exchanged is $4\cdot\log n$.

K=3. An easy combination of the above algorithms results in maximal length $< 6\cdot\log n$.

A probability distribution p is called *equiprobable* if $p\neq 0$ implies that $p = \frac{1}{|S_p|}$. If p is equiprobable and S_p is regular (i.e. about the same number of possible y's for every x and about the same number of possible x's for every y), Theorem 4 can be rephrased to show that $L_a(p) \approx H(X|Y) + H(Y|X)$. If, however, S_p is not regular, the corresponding code will be good only in the *maximal* length sense. The next theorem takes care of this case.

Theorem 5: For every $\epsilon>0$ there exists $C(\epsilon)$ such that if p is equiprobable then

$$L_a(p) \leq H(X|Y) + H(Y|X) + (3+\epsilon)\cdot\log\log(\max(|\mathcal{X}|, |\mathcal{Y}|)) + C(\epsilon) .$$

Proof: Let $A_1 \triangleq \{x : |\{y : p(x,y)>0\}| \leq 2\}$ and for $i=2,..,\lceil \log|\mathcal{Y}| \rceil$ let $A_i \triangleq \{x : 2^{i-1} < |\{y : p(x,y)>0\}| \leq 2^i\}$. Define B_i, $i=1,...,\lceil \log|\mathcal{X}| \rceil$ symmetrically.

The protocol proceeds as follows:

i) P_X transmits the index of the set A_i containing X ($\lceil \log\log|\mathcal{Y}| \rceil$ bits.)

ii) P_Y transmits the index of the set B_j containing Y ($\lceil \log\log|\mathcal{X}| \rceil$ bits.)

iii) P_X and P_Y can now restrict themselves to a submatrix of $\mathcal{X}\times\mathcal{Y}$ having at most 2^i possible Y values for every x and at most 2^j possible X values for every y. They use the protocol of Theorem 4 to find each other's value using at most $\log 2^i + \log 2^j + (1+\epsilon)\cdot\log\log(\max(|\mathcal{X}|, |\mathcal{Y}|)) + C'(\epsilon)$ bits.

The average number of bits transmitted is therefore

$$L_a(p) = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\cdot L(x,y)$$

$$= \sum_{i=1}^{\lceil \log|\mathcal{Y}| \rceil} \sum_{x\in A_i} \sum_{j=1}^{\lceil \log|\mathcal{X}| \rceil} \sum_{y\in B_j} p(x,y)\cdot L(x,y)$$

$$= \sum_{i=1}^{\lceil \log|\mathcal{Y}| \rceil} \sum_{x\in A_i} \sum_{j=1}^{\lceil \log|\mathcal{X}| \rceil} \sum_{y\in B_j} p(x,y)\cdot\Big(\lceil \log\log|\mathcal{X}| \rceil + \lceil \log\log|\mathcal{Y}| \rceil + \log 2^i + \log 2^j + (1+\epsilon)\lceil \log\log(\max(|\mathcal{X}|, |\mathcal{Y}|)) \rceil + C'(\epsilon)\Big)$$

$$= \sum_{i=1}^{\lceil \log|\mathcal{Y}| \rceil} \sum_{x\in A_i} p(x)\cdot i + \sum_{j=1}^{\lceil \log|\mathcal{X}| \rceil} \sum_{y\in B_j} p(y)\cdot j + \lceil \log\log|\mathcal{X}| \rceil + \lceil \log\log|\mathcal{Y}| \rceil + (1+\epsilon)\cdot\lceil \log\log(\max(|\mathcal{X}|, |\mathcal{Y}|)) \rceil + C'(\epsilon)$$

$$\leq \sum_{i=1}^{\lceil \log|\mathcal{Y}| \rceil} \sum_{x\in A_i} p(x)\cdot(H(Y|X=x) + 1) + \sum_{j=1}^{\lceil \log|\mathcal{X}| \rceil} \sum_{y\in B_j} p(y)\cdot(H(X|Y=y) + 1) + \lceil \log\log|\mathcal{X}| \rceil + \lceil \log\log|\mathcal{Y}| \rceil + (1+\epsilon)\cdot\lceil \log\log(\max(|\mathcal{X}|, |\mathcal{Y}|)) \rceil + C'(\epsilon)$$

$$\leq H(X|Y)+H(Y|X)+(3+\epsilon)\cdot\log\log(\max(|x|,|y|))+C(\epsilon) \ \square$$

Note that for almost all equiprobable distributions, $H(X|Y) + H(Y|X) >> \log\log(\max(|\mathcal{X}|,|\mathcal{Y}|))$ making the lower bound of Theorem 1 and the upper bound of Theorem 5 very tight.

Remark: A corollary of the Slepian Wolf Theorem [6] states that if $<(X_i, Y_i)>_{i=1}^{n}$ is a sequence of independent identically distributed random variables, P_X knows $<X_i>_{i=1}^{n}$ and P_Y knows $<Y_i>_{i=1}^{n}$, then, given $\epsilon>0$, for all sufficiently large n, P_X and P_Y can exchange $<X_i>_{i=1}^{n}$ and $<Y_i>_{i=1}^{n}$ with probability of error $<\epsilon$ using $n\cdot\{ H(X_1|Y_1) + H(Y_1|X_1) + c\epsilon \}$ bits (c being a fixed constant).

The standard proof [7] proceeds in two steps. In the first, a set of "typical" $<(x_i, y_i)>_{i=1}^{n}$'s is defined such that

1) The probability of the set is $1-\frac{\epsilon}{2}$.

2) All elements in this set have about the same probability.

3) For each $<x_i>_{i=1}^{n}$ in the "x projection" there are about the same number of $<y_i>_{i=1}^{n}$'s such that $<(x_i, y_i)>_{i=1}^{n}$ is typical. and vice versa.

In the second step it is proved that for all sufficiently large n, if $<(X_i, Y_i)>_{i=1}^{n}$ is in the typical set then P_X and P_Y can exchange $<X_i>_{i=1}^{n}$ and $<Y_i>_{i=1}^{n}$ with probability of error $< \frac{\epsilon}{2}$ using $n\cdot\{ H(X_1|Y_1) + H(Y_1|X_1) + c\epsilon \}$ bits. Theorem 5 can be used to strengthen this part of the proof in three ways:

1) The assumption that there are the same number in each row and the same number in each column can be dropped.

2) The number of bits exchanged is $n\cdot\{ H(X_1|Y_1) + H(Y_1|X_1) \} + c'\cdot\log n$.

3) The probability of error is 0.

5. Randomized Codes.

So far, we have only discussed deterministic protocols. Randomized protocols can be defined similarly. The only difference being that the transmitter's value and previous transmissions determine a real number $\in[0,1]$ rather than an integer $\in\{0,1\}$. (This number denotes the probability that the next transmitted bit is a "1"). We consider the advantages of using randomized codes in 3 cases:

1) P_X and P_Y are required to always know X and Y.

2) P_X and P_Y are required to know X and Y with average probability of error $<\epsilon$.

3) P_X and P_Y are required to know X and Y with probability of error $<\epsilon$ for all instances of X, Y.

We restrict the consideration to the average lengths of the codes. Let $L_i^{\det}(p)$ denote the shortest average length of a deterministic code satisfying the requirements of case i when $p(x,y)$ is the underlying distribution of X, Y and let $L_i^{ran}(p)$ denote the same for randomized codes. The difference between $L_i^{\det}(p)$ and $L_i^{ran}(p)$ (which indicates the advantage of using randomized codes over deterministic ones) increases as we progress through the cases:

1) Since it is always possible to toss all coins prior to the commencement of communication, we transform each randomized code to a deterministic one with shorter or equal average length by looking at all combinations of coin tosses and using the one that minimizes the average code length for the deterministic code. Thus, in this case, $L_1^{\det}(p) = L_1^{ran}(p)$.

2) In a manner similar to that described in [8], a randomized protocol of length L having probability of error $<\epsilon$ implies the existance of a deterministic protocol of length $< 2\cdot L$ with probability of error $< 2\cdot\epsilon$. Thus $L_2^{\det}(p,\epsilon) \leq 2\cdot L_2^{ran}(p,\epsilon/2)$.

3) The following example shows that $L_3^{\det}(p,\epsilon)$ can be as high as $2^{L_3^{ran}(p,\epsilon)/C(\epsilon)}$.

Fix $0<\epsilon<1$ and let $\mathcal{X} = \mathcal{Y} = \{1,\dots,n\}$,

$$p(x,y) = \begin{cases} \frac{1-\delta}{n} & x=y \\ \frac{\delta}{n(n-1)} & x\neq y\,. \end{cases}$$

Using a deterministic protocol, $P(\text{error} \mid X=x, Y=y)$ is always either 0 or 1 so, for it to be less than ϵ, we need $P(\text{error} \mid X=x, Y=y) = 0$ for all x, y. By Example 2, $L_3^{\det}(p) \geq H(X,Y) \approx \log n$.

On the other hand, the following randomized protocol, achieves $L \approx C(\epsilon)\cdot\log\log n$. P_X and P_Y first use the randomized protocol of [9] to find out if $X=Y$ (using $C(\epsilon)\cdot\log\log n$ bits). If this is the case, they stop communicating. Otherwise, they use $2\cdot\log n$ bits to communicate X and Y completely. The average length here is $L = C(\epsilon)\cdot\log\log n + \delta\cdot 2\log n \rightarrow C(\epsilon)\cdot\log\log n$ as $\delta\rightarrow 0$.

References

[1] C. E. Shannon, 'A Mathematical Theory of Communication', *Bell Syst. Tech. J.,* July 1948.

[2] D. A. Huffman, 'A Method for the Construction of Minimum Redundancy Codes.' *Proc. IRE,* Sept. 1952

[3] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, 1968.

[4] R. Ahlswede, 'Coloring Hypergraphs: A New Approach to Multi-user Coding', *J. of Combinatorics, Information & System Sciences',* Vol 4, No 1, pp. 76-115.

[5] A. Orlitsky and A. El Gamal, 'Communication with Secrecy Constraints', *Proc. of the Sixteenth Annual ACM Symposium on Theory of Computing'*, 1984.

[6] D. Slepian and J. K. Wolf, 'Noiseless Coding of Correlated Information Sources', *IEEE Trans. Info. Theory,* July 1973.

[7] T. Cover, 'A proof of the Data Compression Theorem of Slepian and Wolf for for Ergodic Sources',

[8] A. C. Yao, 'Probabilistic Computations: Towards a Unified Measure of Complexity', *Proc. 18th Symp on Foundations of Computer Science*, Oct. 1977, pp. 222-227.

[9] A. C. Yao, 'Some Complexity Questions Related to Distributive Computing', *Proc. of the Eleventh Annual ACM Symposium on Theory of Computing,* 1979.