Harmonic Broadcasting Is Bandwidth-Optimal Assuming Constant Bit Rate

Lars Engebretsen[1,⋆] and Madhu Sudan[2,⋆⋆]

[1] Department of Numerical Analysis and Computer Science
Royal Institute of Technology
SE-100 44 Stockholm
SWEDEN
E-mail: enge@kth.se

[2] MIT Laboratory for Computer Science
Stata Center, Room G640
32 Vassar Street
Cambridge, Massachusetts 02139
USA
E-mail: madhu@mit.edu

November 2005

Abstract. Harmonic broadcasting was introduced by Juhn and Tseng in 1997 as a way to reduce the bandwidth requirements required for video-on-demand broadcasting. In this paper, we note that harmonic broadcasting is actually a special case of the priority encoded transmission scheme introduced by Albanese et al. in 1996 and prove–using an information theoretic argument–that it is impossible to achieve the design goals of harmonic broadcasting using a shorter encoding.

1 Introduction

One way to broadcast an m-minute movie in such a way that a viewer can start viewing the movie every m/k minutes is to simply allocate k channels and broadcast identical copies of the movie on each channel in such a way that there is one copy of the movie starting every m/k minutes. This requires a total of k channels. To reduce the bandwidth requirement, Juhn and Tseng [7] introduced the notion of *harmonic broadcasting*, a scheme where early parts of the movie are broadcasted more frequently than later parts. For a movie with length m minutes, they were able to reduce the waiting time to $2m/k - m/k^2$ minutes by using H_k channels, where H_k is the kth harmonic number. This

⋆This research was performed while the author was visiting MIT with support from the Marcus Wallenberg Foundation and the Royal Swedish Academy of Sciences.

⋆⋆Supported in part by NSF Career Award CCR 9875511, NSF Award CCR 9912342, and MIT-NTT Award MIT2001-04.

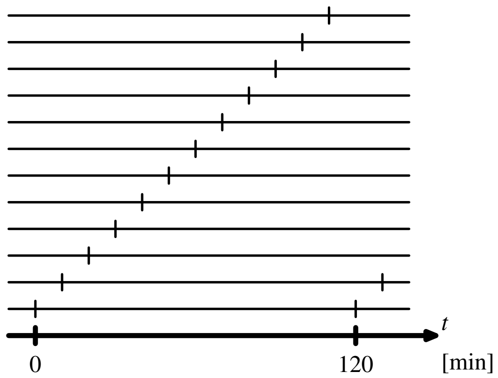

Figure 1. If a 2-hour movie is broadcasted in parallel on 12 channels, a user trying to watch the movie gets a maximum waiting time of 10 minutes. With Harmonic broadcasting, the same waiting time can be achieved using approximately 3 channels.

construction was improved by Pâris, Carter and Long [10, 11] to give a waiting time of m/k needing only slightly more than $\ln(k+1)$ channels (Figure 1). Pâris, Carter and Long [11] also identified the question regarding optimality of harmonic broadcasting, i.e., the question whether it is possible or not to achieve the design goals of harmonic broadcasting using even less bandwidth.

Question 1. Suppose that we want to broadcast an m-minute movie in such a way that the maximum waiting time is m/k minutes. What is the minimum number of channels we need?

In this paper, we study the above question under the assumption that the message is broadcasted with constant bit rate and that the clients have no particular bandwidth restrictions on their own. We first note that harmonic broadcasting is actually a special case of *priority encoded transmission*, a scheme proposed by Albanese et al. [1], and we show that a direct application of priority encoded transmission gives results comparable to those obtained by harmonic broadcasting. In their paper, Albanese et al. also provide a lower bound, showing that their encoding is optimal. Since harmonic broadcasting is a special case of priority encoded transmission, this lower bound is not directly applicable to the less general harmonic broadcasting.

The second contribution of this paper consists of two lower bounds. We first show that if a message of size m is divided into k equally sized blocks and those blocks are encoded in such a way that the ith message block can be recovered from any i consecutive blocks of the encoding, the total size of the encoding must be at least mH_k. We then study the setting where an m bit message is encoded as a bit stream in such a way that a client observing the bit stream has to wait at most w bits before it can start displaying the message, and prove that at least $\ln(1+m/w)+(2w(1+w/m))^{-1}+O(w^{-2})$ channels are needed in this case. With $w = m/k$, this resolves Question 1 and in fact also proves that the improved version of harmonic broadcasting due to Pâris, Carter and Long [11] is optimal.

The lower order terms in our lower bound come from the fact that we treat the message as a stream of bits. Indeed, there already exists a construction with a provably optimal bandwidth demand of exactly $\ln(k+1)$ under the assumtion that the broadcasted stream is a continuum of information rather than a stream of bits [6].

2 Harmonic Broadcasting

Suppose that we want to broadcast an m-minute movie in such a way that a client has a maximum waiting time of m/k until the movie can be viewed. The naive way to accomplish this goal is to simply allocate k channels and broadcast one copy of the movie on each channel in such a way that there is one copy of the movie starting every m/k minutes. This requires a bandwidth that is k times the bandwidth required to broadcast one move.

Harmonic broadcasting was invented by Juhn and Tseng [7] to reduce the bandwidth requirements for video-on-demand broadcasting. The harmonic broadcasting scheme can be viewed as follows: The movie is first divided into k equally sized segments $\langle M_1, M_2, \ldots, M_k \rangle$. Each segment but the first is then divided into equally sized subsegments; the ith segment is divided into the i subsegments $\langle M_{i,0}, M_{i,1}, \ldots, M_{i,i-1} \rangle$. An encoding $\langle E_1, E_2, \ldots, E_k \rangle$ consisting of k equally sized blocks is then transmitted. The ith block in the encoding is constructed by concatenating

$$E_i = M_1 M_{2,i \bmod 2} M_{3,i \bmod 3} \cdots M_{k,i \bmod k}.$$

By the encoding procedure, the total size of the encoded move is

$$k \cdot \sum_{i=1}^{k} \frac{m}{ki} = m \sum_{i=1}^{k} \frac{1}{i} = mH_k.$$

The crucial property of the above construction is that the client can reconstruct M_i from any i consecutive blocks from the encoding.

Proposition 1 [7]. *Harmonic broadcasting has the property that the client can reconstruct the ith segment of the movie given any i consecutive blocks from the encoding. Moreover, the total size of the encoding is at most mH_k, where H_k is the kth harmonic number.*

Juhn and Tseng [7] claimed that harmonic broadcasting gives a maximum waiting time of m/k at the client end, but this is not the case. It was observed by Pâris, Carter and Long [10] that the client actually may not get all data it needs to reconstruct M_i on time and that the maximum waiting time is in fact $2m/k - m/k^2$ at the client end before the movie can start playing.

In their papers [10, 11], Pâris, Carter and Long propose three protocols–cautious harmonic broadcasting, quasi-harmonic broadcasting and polyharmonic broadcasting–that can guarantee a maximum waiting time of m/k. Specifically, their polyharmonic broadcasting protocol [11] divides the message into s segments. To get a maximum waiting time of wm/s, the protocol uses $H_{s+w-1} - H_{w-1}$ channels. We establish in Corollary 2–using an information theoretic argument–that this protocol is optimal in the constant bit-rate model if the message is viewed as a bit stream.

3 Priority Encoded Transmission

Priority encoded transmission was introduced by Albanese et al. [1] as a way to transmit a message over a noisy channel in such a way that certain parts of the message are delivered more quickly than others. The message is assumed to consist of m words, a word consists of w bits.

Definition 1. *For a message of length m, a* priority encoded transmission system *with packet size ℓ, n packets and encoding length $e = n\ell$ consists of:*

1. *An encoding function that maps a message of length m onto an encoding of total length e consisting of n packets of ℓ words each.*
2. *A decoding function that maps sets of at most n packets onto m words.*
3. *A priority function ρ that maps $\{1, 2, \ldots, m\}$ to the interval $(0, 1]$.*

The guarantee of the system is that, for all messages of length m and for all $i \in \{1, 2, \ldots, m\}$, the decoding function is able to decode the ith message word from any ρ_i fraction of the n encoding packets.

In their paper, Albanese et al. show that given a priority function, it is possible to construct a priority encoded transmission system with a priority function that closely approximates the given one.

Proposition 2 [1, Theorem 4.3]. *On input message length m, packet length ℓ, a k-partition of the message, i.e., positive integers $m_1, m_2, \ldots, m_k$ such that $m = m_1 + m_2 + \cdots + m_k$, and corresponding priority values $\rho_1, \rho_2, \ldots, \rho_k$, there is an efficient procedure that produces a priority encoded transmission system with priority function ρ' and n encoding packets such that the total encoding length is*

$$n\ell \leq \frac{1}{1 - k/\ell} \sum_{i=1}^{k} \frac{m_i}{\rho_i} + \ell$$

and that all words of the message in the ith block of the k-partition have priority value $\rho'_i \leq \rho_i + \ell/m$.

In the video-on-demand setting, we have a message of size m which we divide into k equally sized blocks. We want to encode this message in such a way that we can always decode the ith block from a fraction i/k of consecutive encoding packets.

Theorem 1. *For any packet length ℓ such that $k < \ell < m/k$, there exists a priority encoded transmission system fulfilling the design goals of harmonic broadcasting. Moreover, the total length of the encoding is at most*

$$\frac{m}{1-k/\ell}\left(H_k + \frac{\pi^2/6}{m/k\ell - 1}\right) + \ell$$

where H_k is the kth harmonic number.

Before proving the theorem, we note that typical values of k, ℓ and m satisfy $k \ll \ell \ll m/k$; in this case the bound given above is approximately mH_k. Also, since the client has to receive a fraction $1/k$ of the packets in the encoding before starting to view the movie, the maximum waiting time is less than $\ell + m/k$, which is approximately m/k when $\ell \ll m/k$.

Proof of Theorem 1. By Proposition 2 with $m_i = m/k$ and priorities $\rho_i = i/k - \ell/m$, there is an efficient procedure that produces a priority encoded transmission system with a priority function ρ' satisfying $\rho'_i \leq i/k$ and having a total encoding length at most

$$\frac{1}{1-k/\ell}\sum_{i=1}^{k}\frac{m_i}{\rho_i} + \ell = \frac{m}{1-k/\ell}\sum_{i=1}^{k}\frac{1}{i - k\ell/m} + \ell. \tag{1}$$

To give the above expression a more convenient form, we let $\varepsilon = k\ell/m < 1$ and rewrite the sum in (1) as

$$\sum_{i=1}^{k}\frac{1}{i-\varepsilon} = \sum_{i=1}^{k}\frac{1}{i(1-\varepsilon/i)} = \sum_{i=1}^{k}\frac{1}{i}\sum_{j=0}^{\infty}(\varepsilon/i)^j = \sum_{j=0}^{\infty}\varepsilon^j\sum_{i=1}^{k}\frac{1}{i^{j+1}}\,.$$

The term corresponding to $j = 0$ above evaluates to H_k; the rest of the sum can be upper bounded by

$$\sum_{j=1}^{\infty}\varepsilon^j\sum_{i=1}^{k}\frac{1}{i^{j+1}} \leq \sum_{j=1}^{\infty}\varepsilon^j\sum_{i=1}^{k}\frac{1}{i^2} \leq \sum_{j=1}^{\infty}\varepsilon^j\sum_{i=1}^{\infty}\frac{1}{i^2} = \frac{\pi^2}{6}\sum_{j=1}^{\infty}\varepsilon^j = \frac{\pi^2}{6}\frac{\varepsilon}{1-\varepsilon}\,.$$

The first of the two equalities above follows, e.g., from [4, § 1.443, equation (3)] with $x = 0$. To conclude, the entire sum in (1) is at most

$$H_k + \frac{\pi^2}{6}\frac{k\ell/m}{1-k\ell/m} = H_k + \frac{\pi^2/6}{m/k\ell - 1}$$

and therefore the total encoding length is at most

$$\frac{m}{1-k/\ell}\left(H_k + \frac{\pi^2/6}{m/k\ell - 1}\right) + \ell. \qquad \blacksquare$$

4 Optimality Of Harmonic Broadcasting

Albanese et al. [1] also provide a lower bound on the total length of the transmitted message in their paper.

Proposition 3 [1, Theorem 3.3]. *For any given priorities* $0 < \rho_1 \le \rho_2 \le \ldots \le \rho_m \le 1$*, if there is a priority encoded transmission system with these priorities, the total encoding length is at least* $1/\rho_1 + 1/\rho_2 + \cdots + 1/\rho_m$.

Applied to harmonic broadcasting, i.e., a partition of m into k equally sized blocks and priorities such that all words in the ith block have priority i/k, the above theorem lower bounds the size of the encoding by mH_k.

Since a priority encoded transmission system with the above priorities satisfies the requirement that it should be possible to reconstruct the first k parts from *any* collection of k different received blocks, it is conceivable that there exists a shorter encoding that still satisfies the weaker requirement that it should be possible to reconstruct the first k parts from any sequence of k *consecutive* received blocks. In this section, we rule out that possibility by proving that also under this latter, weaker, requirement, the above lower bound holds.

Theorem 2. *Suppose that a message* $\langle M_1, M_2, \ldots, M_m \rangle \in \tau_1 \times \tau_2 \times \cdots \times \tau_m$ *is encoded as* $\langle E_1, E_2, \ldots, E_n \rangle \in \sigma_1 \times \sigma_2 \times \cdots \times \sigma_n$ *in such a way that the value of* M_i *can be recovered from any* $\rho_i n$ *consecutive packets from the encoding, where* $0 < \rho_1 \le \rho_2 \le \cdots \le \rho_m = 1$*. Then*

$$\sum_{i=1}^{m} \frac{\log_2 |\tau_i|}{\rho_i} \le \sum_{i=1}^{n} \log_2 |\sigma_i|.$$

Before proving Theorem 2, we state some applications of it. The first application shows that given the assumption that we encode equally sized blocks of the message in such a way that the ith block can be recovered from i consecutive packets of the encoding, it is impossible to achieve a shorter encoding than that achieved by harmonic broadcasting.

Corollary 1. *Suppose that a message containing* m *bits is divided into* k *equally sized blocks* $\langle M_1, M_2, \ldots, M_k \rangle$ *and that these blocks are then encoded into* k *packets* $\langle E_1, E_2, \ldots, E_k \rangle$ *in such a way that it is possible to recover* M_i *from any* i *consecutive packets of the encoding. Then the encoding contains at least* mH_k *bits where* H_k *is the* k*th harmonic number.*

Proof. Use Theorem 2 with $m = n = k$, $\tau_i = \{0,1\}^{m/k}$ and $\rho_i = i/k$. Then the total number of bits in the encoding is at least

$$\sum_{i=1}^{k} \frac{\log_2 |\tau_i|}{\rho_i} = \sum_{i=1}^{k} \frac{m/k}{i/k} = m \sum_{i=1}^{k} \frac{1}{i} = mH_k. \qquad \blacksquare$$

In the second application of Theorem 2 we drop the assumption on the block size from Corollary 1. Instead, we assume that we want a certain maximum waiting time at the client end and prove a lower bound on the number of channels needed to achieve this waiting time. This answers Question 1 and establishes, with the substitution $w = m/k$ in Corollary 2 below, that polyharmonic broadcasting is optimal.

Corollary 2. *Suppose that we want to transmit a movie consisting of m bits at constant bit rate α in such a way that the maximum waiting time until the client can start viewing the movie is w. Then α must be at least*

$$H_{m+w-1} - H_{w-1} = \ln(1 + m/w) + \frac{1}{2w(1 + w/m)} + O(w^{-2})$$

times the bit rate needed to transmit one copy of the movie.

Proof. If we let $\tau_i = \{0, 1\}$ and $\sigma_i = \{0, 1\}$ in Theorem 2, we get the bound

$$\sum_{i=1}^{m} \frac{1}{\rho_i n} \leq 1. \tag{2}$$

Now suppose that we use α channels, i.e., a total bandwidth of α times the bandwidth required to transmit one copy of the movie. Since we assumed a maximum waiting time of w, $\rho_1 n \leq \alpha w$ since we must be able to decode the first bit in the message after time w. Similarly, $\rho_i n \leq \alpha(w + i - 1)$, since we must be able to decode the ith bit in the message after time $w + i - 1$. Therefore,

$$\sum_{i=1}^{m} \frac{1}{\rho_i n} \geq \sum_{i=1}^{m} \frac{1}{\alpha(w + i - 1)}. \tag{3}$$

By combining the bounds (2) and (3), we obtain

$$\alpha \geq \sum_{i=1}^{m} \frac{1}{i + w - 1} = \sum_{i=0}^{m-1} \frac{1}{i + w} = \psi(m + w) - \psi(w),$$

where ψ is the digamma function [4, § 8.36]. Since

$$\psi(x) = \ln x - \frac{1}{2x} - 2\int_0^{\infty} \frac{t\,dt}{(t^2 + x^2)(e^{2\pi t} - 1)} = \ln x - \frac{1}{2x} + O(x^{-2})$$

for positive x we obtain the bound

$$\alpha \geq \ln(1 + m/w) + \frac{1}{2w(1 + w/m)} + O(w^{-2}). \qquad \blacksquare$$

We now turn to the proof of Theorem 2. The proof uses an information theoretic argument along the lines of Albanese et al. [1], the main difference being that we consider only consecutive blocks of the encoding instead of arbitrary sets of blocks. This requires us to derive an information theoretic inequality that, to our knowledge, has not been given explicitly in the literature before and might be of independent interest.

Definition 2. *For a random variable X with probability density P, the* binary entropy of X *is* $H(X) = \mathrm{E}_P[-\log_2 P(X)]$.

Definition 3. *Let $X = \langle X_1, X_2, \ldots, X_k \rangle$. For any set $Q \subseteq \{1, 2, \ldots, k\}$, let $X_Q = \langle X_i \rangle_{i \in Q}$.*

Definition 4. *Let $B_{q,k}$ be the set of all blocks of size q from $\{1, 2, \ldots, k\}$, wrapping is allowed. I.e.,*

$$B_{q,k} = \{\{i, i+1 \bmod k, i+2 \bmod k, \ldots, i+q-1 \bmod k\} : i \in [k]\}$$

where $[k]$ denotes the set of integers from 1 to k.

Definition 5. *Let $X = \langle X_1, X_2, \ldots, X_k \rangle$ be a random variable. For any $q \in \{1, 2, \ldots, k\}$, define*

$$H_q(X) = \frac{k/q}{|B_{q,k}|} \sum_{Q \in B_{q,k}} H(X_Q).$$

Intuitively, the quantity $qH_q(X)/k$ expresses the entropy of a randomly selected block of q consecutive random variables and $H_q(X)$ expresses the average entropy per block in a random interval of q successive blocks. The following lemma is a modification of an inequality due to Han [5]; we provide a proof along the lines of Chung et al. [2] in Sec. 5.

Lemma 1. *Let $X = \langle X_1, X_2, \ldots, X_k \rangle$ and let P be a probability distribution on X. Then $H_1(X) \geq H_2(X) \geq \cdots \geq H_k(X) = H(X)$ where the entropy is computed according to the distribution P.*

Using the above lemma together with the identity $H(X, Y \mid Z) = H(Y \mid Z) + H(X \mid Y, Z)$, the proof of Theorem 2 proceeds by relating the entropy of the encoding to the entropy on the message.

Lemma 2. *Given a message space $\tau_1 \times \tau_2 \times \cdots \times \tau_m$, a probability distribution on this space, and an encoding space $\sigma_1 \times \sigma_2 \times \cdots \times \sigma_n$, suppose that a message $\langle M_1, M_2, \ldots, M_m \rangle$ from the message space is encoded as $\langle E_1, E_2, \ldots, E_n \rangle$ in such a way that the value of M_i can be recovered from any $\rho_i n$ consecutive encoded packets, where $0 < \rho_1 \leq \rho_2 \leq \cdots \leq \rho_m = 1$. Then*

$$\sum_{i=1}^{n} H(E_i) \geq \sum_{i=1}^{m} \frac{H(M_i \mid M_{[i-1]})}{\rho_i}.$$

where $[j]$ denotes the set of integers from 1 to j and the entropies are computed according to the probability distribution on the message space.

Proof. Define $\rho_0 = 1/n$. We first prove by induction on j that

$$H_{\rho_j n}(E \mid M_{[j]}) \leq H_1(E) - \sum_{i=1}^{j} \frac{H(M_i \mid M_{[i-1]})}{\rho_i} \tag{4}$$

for all $j \in \{0, 1, \ldots, m\}$. The base case $j = 0$ is clear since $H_{\rho_0 n}(E \mid M_{[0]}) = H_1(E)$. Now we assume that (4) holds for j and prove that this implies the inequality for $j + 1$. Define $q = \rho_{j+1} n$ and fix a $Q \in B_{q,n}$. Then

$$\begin{aligned} H(E_Q \mid M_{j+1}, M_{[j]}) &= H(E_Q, M_{j+1} \mid M_{[j]}) - H(M_{j+1} \mid M_{[j]}) \\ &= H(E_Q \mid M_{[j]}) - H(M_{j+1} \mid M_{[j]}), \end{aligned}$$

where the last equality follows since

$$\begin{aligned} H(E_Q, M_{j+1} \mid M_{[j]}) &= H(E_Q, \mid M_{[j]}) + H(M_{j+1} \mid E_Q, M_{[j]}) \\ &= H(E_Q \mid M_{[j]}), \end{aligned}$$

which, in turn, follows since the value of E_Q determines that of M_{j+1}. Thus,

$$\sum_{Q \in B_{q,n}} H(E_Q \mid M_{[j+1]}) = \sum_{Q \in B_{q,n}} H(E_Q \mid M_{[j]}) - |B_{q,n}| H(M_{j+1} \mid M_{[j]}).$$

Dividing this relation with $q|B_{q,n}|/n = \rho_{j+1}|B_{q,n}|$, we obtain

$$H_q(E \mid M_{[j+1]}) = H_q(E \mid M_{[j]}) - \frac{H(M_{j+1} \mid M_{[j]})}{\rho_{j+1}}$$

by the definition of H_q and the fact that $n/q = 1/\rho_{j+1}$. Since, by Lemma 1, $H_q(E \mid M_{[j]}) \leq H_r(E \mid M_{[j]})$ for all $q \geq r$,

$$H_q(E \mid M_{[j+1]}) = H_{\rho_{j+1} n}(E \mid M_{[j+1]}) \leq H_{\rho_j n}(E \mid M_{[j]}) - \frac{H(M_{j+1} \mid M_{[j]})}{\rho_{j+1}}$$

and when this is combined with the induction hypothesis, we obtain

$$H_{\rho_{j+1} n}(E \mid M_{[j+1]}) \leq H_1(E) - \sum_{i=1}^{j+1} \frac{H(M_i \mid M_{[i-1]})}{\rho_i}$$

and conclude that (4) holds for all $j \in \{0, 1, \ldots, m\}$. The upper bound on j follows from the fact that $\rho_m = 1$.

Setting $j = m$ in (4) gives $H_{\rho_m n}(E \mid M_{[m]}) = 0$ since the encoding is completely determined by the message, and thus

$$H_1(E) - \sum_{i=1}^{m} \frac{H(M_i \mid M_{[i-1]})}{\rho_i} \geq 0.$$

Since

$$H_1(E) = \sum_{i=1}^{n} H(E_i)$$

we have obtain the bound we set out to prove. ∎

The proof of Theorem 2 is a direct application of the above result.

Proof of Theorem 2. Apply Lemma 2 to the case when $\langle M_1, M_2, \ldots, M_m \rangle$ attains each value in $\tau_1 \times \tau_2 \times \cdots \times \tau_m$ with equal probability. Then $H(M_i \mid M_{[i-1]}) = \log_2 |\tau_i|$. Furthermore, $H(E_i)$ is always at most $\log_2 |\sigma_i|$, no matter the distribution of the encodings. ∎

5 Proof of Lemma 1

The proof uses the following basic entropy identities:

$$H(Y \mid X, Z) \leq H(Y \mid X), \tag{5}$$
$$H(X, Y) = H(X) + H(Y \mid X), \tag{6}$$

where X, Y and Z can be arbitrary collections of random variables [8, 3]. There are two cases–notice that $|B_{q,k}| = k$ for all q such that $1 \leq q \leq k-2$ and that $|B_{k,k}| = 1$. Let us first show that $H_{k-1}(X) \geq H_k$.

Lemma 3. *Let $X = \langle X_1, X_2, \ldots, X_k \rangle$ and let P be a probability distribution on X. Then $H_{k-1}(X) \geq H_k(X)$ where the entropy is computed according to the distribution P.*

Proof. We first rewrite $H(X_1, X_2, \ldots, X_k)$ by conditioning on X_i according to equation (6): $H(X_1, X_2, \ldots, X_k) = H(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k) + H(X_i \mid X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$ for any i such that $1 \leq i \leq k$. Applying the bound (5) to the last term above then gives $H(X_1, X_2, \ldots, X_k) \leq H(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k) + H(X_i \mid X_1, X_2, \ldots, X_{i-1})$ for any i such that $1 \leq i \leq k$. Summing the above inequality over all i gives the inequality

$$kH(X_1, X_2, \ldots, X_k) \leq \sum_{Q \in B_{k-1,k}} H(X_Q) + \sum_{i=1}^{k} H(X_i \mid X_1, X_2, \ldots, X_{i-1})$$

$$= \sum_{Q \in B_{k-1,k}} H(X_Q) + H(X_1, X_2, \ldots, X_k),$$

where the last equality follows from equation (6). To conclude,

$$(k-1)H(X_1, X_2, \ldots, X_k) \leq \sum_{Q \in B_{k-1,k}} H(X_Q),$$

which is equivalent to $H_{k-1}(X) \geq H_k(X)$. ∎

The remaining case follows by a slight extension of the above argument.

Lemma 4. *Let $X = \langle X_1, X_2, \ldots, X_k \rangle$, q be an integer such that $1 \leq q \leq k-2$, and P be a probability distribution on X. Then $H_q(X) \geq H_{q+1}(X)$ where the entropy is computed according to the distribution P.*

Proof. In this proof, the indices are cyclic: $k + j$ is interpreted as j. By equation (6)

$$\begin{aligned} &H(X_i, X_{i+1}, \ldots, X_{i+q}) \\ &\qquad = H(X_{i+1}, X_{i+2}, \ldots, X_{i+q}) + H(X_i \mid X_{i+1}, X_{i+2}, \ldots, X_{i+q}) \end{aligned}$$

for any integer i such that $1 \leq i \leq q$. Summing the above equation over all i and multiplying by $q + 1$ gives

$$\begin{aligned} &(q+1) \sum_{Q \in B_{q+1,k}} H(X_Q) \\ &\qquad = (q+1) \sum_{Q \in B_{q,k}} H(X_Q) + (q+1) \sum_{i=1}^{k} H(X_i \mid X_{i+1}, X_{i+2}, \ldots, X_{i+q}). \end{aligned}$$

We now claim that

$$(q+1) \sum_{i=1}^{k} H(X_i \mid X_{i+1}, X_{i+2}, \ldots, X_{i+q}) \leq \sum_{Q \in B_{q+1,k}} H(X_Q). \tag{7}$$

This is enough to complete the proof, since then

$$q \sum_{Q \in B_{q+1,k}} H(X_Q) \leq (q+1) \sum_{Q \in B_{q,k}} H(X_Q).$$

To see that (7) holds, we rewrite the left-hand side as

$$\sum_{j=0}^{q} \sum_{i=1}^{k} H(X_i \mid X_{i+1}, X_{i+2}, \ldots, X_{i+q})$$

and reorder the terms in the sum to get the equivalent expression

$$\sum_{i=1}^{k} \sum_{j=0}^{q} H(X_{i+j} \mid X_{i+j+1}, X_{i+j+2}, \ldots, X_{i+j+q}).$$

By the bound (5), this sum is at most

$$\sum_{i=1}^{k} \sum_{j=0}^{q} H(X_{i+j} \mid X_{i+j+1}, X_{i+j+2}, \ldots, X_{i+q}),$$

which, by equation (6), can be rewritten as

$$\sum_{i=1}^{k} H(X_i, X_{i+1}, \ldots, X_{i+q}) = \sum_{Q \in B_{q+1,k}} H(X_Q). \qquad \blacksquare$$

Acknowledgment

We thank the anonymous referees for many pointers and valuable comments. Gérard Battail found an error in an early version of the proof of Theorem 1 and provided an elegant fix.

References

1. Andres Albanese, Johannes Blömer, Jeff Edmonds, Michael Luby, and Madhu Sudan. Priority encoding transmission. *IEEE Transactions on Information Theory*, 42(6):1737–1744, November 1996.
2. Fan R. K. Chung, Ronald L. Graham, Peter Frankl, and James B. Shearer. Some intersection theorems for ordered sets and graphs. *Journal of Combinatorial Theory, Series A*, 43(1):23–37, September 1986.
3. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* John Wiley & Sons, New York, 1991.
4. Izrail' S. Gradstein and Iosif M. Ryzhik. Alan Jeffrey and Daniel Zwillinger, editors. *Tables of Integrals, Series, and Products.* Academic Press, San Diego, sixth edition, 2000.
5. Te Sun Han. Nonnegative entropy measures of multivariate symmetric correlations. *Information and Control*, 36(2):133–156, February 1978.
6. Ailan Hu, Ioanis Nikolaidis, and Peter van Beek. On the design of efficient video-on-demand broadcast schedules. In *Proceedings of the 7th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, pages 262–269. College Park, Maryland, 24–28 October 1999.
7. Li-Shen Juhn and Li-Ming Tseng. Harmonic broadcasting for video-on-demand service. *IEEE Transactions on Broadcasting*, 43(3):268–271, September 1997.
8. Aleksandr Ia. Khinchin. *Mathematical Foundations of Information Theory.* Dover Publications, New York, 1957.
9. Ioanis Nikolaidis, Fulu Li, and Ailan Hu. An inherently loss-less and bandwidth efficient periodic broadcast scheme for VBR video. In *Proceedings of the International Conference on Measurement and Modeling of Computer Systems*, pages 116–117. Santa Clara, California, 17–21 June 2000.
10. Jehan-François Pâris, Steven W. Carter, and Darrell D. E. Long. Efficient broadcasting protocols for video on demand. In *Proceedings of the 6th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, pages 127–132. Montréal, July 1998.
11. Jehan-François Pâris, Steven W. Carter, and Darrell D. E. Long. A low bandwidth broadcasting protocol for video on demand. In *Proceedings of the 7th International Conference on Computer Communications and Networks*, pages 690–697. Lafayette, LA, October 1998.