

Bounds on the Benefit of Network Coding  
for Multicast and Unicast Sessions  
in Wireless Networks

Alireza Keshavarz-Haddad<sup>†</sup>

Rudolf Riedi<sup>†‡</sup>

Department of Electrical and Computer Engineering<sup>†</sup> and Department of Statistics<sup>‡</sup>  
Rice University, 6100 Main Street, Houston, TX 77005, USA

**Technical Report TR2007-09**  
Department of Statistics, Rice University

## Abstract

In this technical report we establish fundamental limitations to the benefit of network coding in terms of energy and throughput in multihop wireless networks. Thereby we adopt two well accepted scenarios in the field: single multicast session and multiple unicast sessions. Most of our results apply to arbitrary wireless network and are, in particular, not asymptotic in kind.

We prove that the gain of network coding in terms of *throughput and energy saving* of a single multicast session is at most a constant factor in wireless networks. This result is in contrast to wired case where the gain of network coding can be an unboundedly large factor [13].

Also, we present a lower bound on the expected number of transmissions of multiple unicast sessions under an arbitrary network coding. We identify scenarios for which the network coding gain for energy saving becomes surprisingly close to 1, in some cases even exactly 1, corresponding to no benefit at all.

Moreover, we prove that the gain of network coding in terms of *transport capacity* is bounded by a constant factor  $\pi$  in any arbitrary wireless network and for all traditional channel models. This shows that the traditional bounds on the transport capacity [1, 9, 10, 15] do not change more than constant factor  $\pi$  if we employ network coding. As a corollary, we find that the gain of network coding on the *throughput* of large scale homogeneous wireless networks is asymptotically bounded by a constant. Note that our result is more general than the previous work [25] and it is obtained by a different technique.

Furthermore, we establish important theorems on employing network coding for multiple unicast sessions. These theorems relate a coding scheme to a simple routing scheme and they can be used as criteria for evaluating the potential gain of network coding in a given wired/wireless network. Then, we identify more scenarios where network coding has no gain on throughput or energy saving of the network.

In conclusion, we show that in contrast to wired networks, the network coding gain in wireless networks is constraint by fundamental limitations. It is also important to mention that wireless channel model is different than wired network with bidirectional links.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Related Work</b>	<b>4</b>
<b>3</b>	<b>Model Assumptions</b>	<b>6</b>
3.1	Wireless Channel Models . . . . .	6
3.1.1	Protocol Model . . . . .	7
3.1.2	Physical Model . . . . .	7
3.2	Connectivity Graph and Traffic Pattern . . . . .	7
3.3	Transport Capacity . . . . .	8
3.4	Flow Schemes and Coding Schemes . . . . .	9
3.5	Data Streams and Energy Consumption Model . . . . .	9
<b>4</b>	<b>Bounds on the Gain of Network Coding for a Single Multicast Session</b>	<b>11</b>
4.1	Energy Gain of Network Coding in Wireless Networks . . . . .	11
4.2	Throughput Gain of Network Coding in Wireless Networks . . . . .	16
<b>5</b>	<b>Bounds on the Gain of Network Coding for Multiple Unicast Sessions</b>	<b>18</b>
5.1	Energy Gain of Network Coding . . . . .	18
5.2	Throughput Gain of Network Coding in Wireless Networks . .	21
5.3	Energy and Throughput Gain of Info-independent Coding Schemes . . . . .	26
<b>6</b>	<b>Conclusion</b>	<b>33</b>

# Chapter 1

## Introduction

In recent years, network coding has become an important research topic in network information theory. It has been shown that network coding can help improve the throughput and energy consumption of communication networks. Theoretical studies of network coding provide guidelines for designing or improving efficient and high performance wired or wireless networks. In this technical report, we present theory leading to fundamental bounds on the gain network coding in arbitrary wireless networks. In particular, our work reveals that the benefit of network coding is limited by constants depending only on the dimension of the underlying space.

Today, studies on network coding have grown to large numbers, mostly focussing on benefits in terms of throughput or energy saving. In this paper we study the fundamental limitations of network coding in the same terms and identify scenarios where the network coding gain on the performance is noticeably small. To the best of our knowledge, this is the first paper which studies the bounds on the gain of network coding in terms of energy saving in wireless networks.

As the first contribution, we prove that the *energy gain* of network coding for a single multicast session is bounded by a small constant factor in an arbitrary wireless network. The constant factor is 17 for planar wireless networks. This lies in stark contrast to wired network where the network coding gain in terms of energy consumption and throughput can be an unboundedly large factor [13].

As the second contribution, we bound the *throughput gain* of network coding for a single multicast session in an arbitrary wireless network. We show that network coding can improve the throughput in wireless networks

by at most a constant factor which is determined by the parameters of the underlying wireless channel model.

As the third contribution, we study the *energy gain* for multiple unicast sessions in an arbitrary wireless network. We compute a novel lower bound on the expected number of transmissions for transporting unicast bits. Assuming that each transmission carries the same size packet and that energy consumption is proportional to the number of transmissions, we find thus an upper bound on the energy gain of network coding. In addition, we identify important scenarios where network coding benefits in terms of energy consumption constitute a relatively small amount, in some cases *none at all*, e.g., for certain traffic patterns in wireless sensor and mesh networks.

As the fourth contribution, we study the benefit of network coding in terms of the *throughput* of wireless networks for multiple unicast sessions. We prove that network coding can increase the maximum *transport capacity* by at most a constant factor  $\pi$  in an arbitrary wireless network. From this, we conclude that the traditional bounds on the transport capacity [1, 9, 10, 15] would increase by at most a factor of  $\pi$  if we employed network coding. We also prove that the gain of network coding on the throughput of large homogeneous wireless networks is most a constant factor. This result is more general than the result of previous work [25] which has been proved for only a particular topology and wireless channel model.

As the fifth contribution, we establish theorems which relate a coding scheme on multiple unicast sessions to a simple routing scheme (without coding). The theorems can provide criteria for evaluating the potential gain of network coding in a given wired/wireless network. Then, using these theorems, we identify several scenarios where a simple routing scheme is the most efficient scheme in terms of throughput or energy saving among all coding schemes in the network.

In summary, we provide several important novel insights, especially on sensor and mesh networks, and generalize to arbitrary networks several results that had been known only for special cases.

The report is organized as follows. In Section 2, we review some related work. We introduce the network model and notations in Section 3. We bound the network coding gain in terms of energy and throughput for a single multicast in Section 4. Next, in Section 5, we study the benefit of network coding in wireless networks for multiple unicast sessions. Finally, we conclude the paper in Section 6.

# Chapter 2

## Related Work

Network coding was first introduced in the seminal paper by Ahlswede et al. [2] in which it was proven that the maximum flow capacity of a single multicast session can be achieved using network coding in an arbitrary network with directional links. Later, [19] and [17] show constructively that the linear network codes can achieve the capacity of a single multicast session as well. Since then, a large body of work has explored the construction of efficient network coding algorithms, e.g., [3, 5, 12, 13, 17, 19, 28].

For the networks with bidirectional links, [20, 23] show that network coding improves the throughput by at most a constant factor 2 for a single multicast. The constant factor turns out to be equal to one (no benefit) in the case of a single unicast or a broadcast. Note that these results do not extend for wireless networks where the channel is considered bidirectional. Because, network coding combined with wireless broadcasting can potentially improve the performance in terms of throughput, energy efficiency, and congestion control in wireless networks [26, 27].

The potential of network coding for energy savings in broadcasting in ad hoc networks was established in [6, 7]. Here, we complement this valuable insight by show that such a gain is in fact bounded by a small constant for any wireless multicast. Moreover, we clearly strengthen the results of [30] for the maximum flow achievable in random wired and wireless networks (modeled as geometric random graphs) for a single multicast session by taking wireless interference into account.

The case of multiple unicast sessions in wireless networks is studied in [21, 22]. Those papers studied cases where network coding would provide only marginal benefits. Most recent work [25] shows that the unicast capacity

gain of network coding in ad hoc networks is asymptotically bounded by a constant factor, as the network is scaled to large size. We highlight that in this paper we provide new results which are non-asymptotical and valid for any network.

Some papers assume correlation between the generated information from different sources in the network. They combine network coding with Distributed Source Coding (DSC) techniques to optimize the coding gain in the network [11, 24, 29]. However, in this paper, we focus on pure network coding gain. In other words, we assume that the sources generate independent information.

# Chapter 3

## Model Assumptions

In this section, we describe the models and notions used in this paper. A communication network is a collection of directed (or undirected) links connecting communication devices (nodes). The links can be established through actual wired or wireless transmissions. Here, we consider an arbitrary wireless networks. We assume that the channel is bidirectional and multiple-access in wireless network.

We emphasize that we consider an *arbitrary* topology for the wireless network. However, we assume that the topology has the *connectivity* which is needed for establishing the demanded sessions. We assume that the nodes are distributed in  $d$ -dimensional Euclidean space  $\mathbb{R}^d$  and denote the set of nodes by  $V$ .

### 3.1 Wireless Channel Models

We employ the Protocol and the Physical Model [9, 10] for modeling the wireless channel. We denote the set of transmitter-receiver pairs of *simultaneous direct* transmissions active at a given time by  $\mathcal{SD} := \{(S_1, D_1), (S_2, D_2), \dots, (S_m, D_m)\}$ . Also, we denote the set of transmitters by  $\mathcal{S} := \{S_1, \dots, S_m\}$ . Note that these sets vary over time; if not otherwise indicated, however, we will consider one fixed but arbitrary time instant. For simplicity of notation, the node symbols are used also to represent their locations. For example,  $|S_i - D_i|$  is the *Euclidean distance* between the nodes  $S_i$  and  $D_i$  in  $\mathbb{R}^d$ .

In both, the Protocol and the Physical Model the assigned transmission

rate from node  $S_i \in \mathcal{S}$  to node  $D_i$  is  $W_i = W$  for a successful transmission where  $W$  is the *channel capacity*; for unsuccessful transmissions we set  $W_i = 0$ . Note that we assume a broadcast channel for wireless networks, so a transmission will typically be received by several nodes simultaneously. On one hand, broadcasting data to all neighbors can help to increase the throughput, on the other hand, simultaneous reception from different nodes is not feasible because of interference.

### 3.1.1 Protocol Model

Under the Protocol model a transmission is modeled as successful if  $|S_k - D_i| \geq (1 + \Delta)r$  for all  $S_k \in \mathcal{S} \setminus \{S_i\}$ , and  $|S_i - D_i| \leq r$  where  $\Delta > 0$  is the *interference parameter* and  $r > 0$  the *transmission range*.

### 3.1.2 Physical Model

Under the Physical model a transmission is modeled as successful if

$$\text{SINR} = \frac{PG_{ii}}{N_o + \sum_{k \neq i, k \in \mathcal{S}} PG_{ki}} \geq \beta \quad (3.1)$$

Here,  $\beta$  is the SINR-threshold,  $N_o$  represents the ambient noise, and  $G_{ki}$  denotes the signal loss, meaning that  $PG_{ki}$  is the receiving power at node  $D_i$  from transmitter  $S_k$ . We assume a low power decay for the signal loss of the form  $G_{ki} = |S_k - D_i|^{-\alpha}$ , where  $\alpha > d$  is the signal loss exponent. It is easy to show that  $r_{\max} = (\frac{P}{N_o\beta})^{1/\alpha}$  is the maximum transmission range for this model.

## 3.2 Connectivity Graph and Traffic Pattern

Wireless networks are usually modeled by *geometric graphs*. The nodes of network are the vertices  $V$  of geometric graphs. Two nodes are considered adjacent and connected by an edge  $e \in E$ , if the distance between them is less than a certain value  $r_G$ . We build the *connectivity graph* of a given wireless network by setting  $r_G = r$  for the Protocol Model and  $r_G = r_{\max}$  for the Physical Model in the geometric graph model. We represent the connectivity graph by  $G = (V, E)$ . It is easy to show that in order to establish

some sessions in the wireless network there must exist a path in  $G$  between the source and the terminals of each session.

We represent multiple unicast sessions by the set of source-terminal pairs  $\mathcal{AB} := \{(A_1, B_1), \dots, (A_k, B_k)\}$  where  $\mathcal{A} := \{A_1, \dots, A_k\}$  and  $\mathcal{B} := \{B_1, \dots, B_k\}$  are the sets of sources and terminals. We refer to the *hop-count distance* of two nodes  $A_i$  from  $B_j$  by  $\ell(A_i, B_j)$ . For the networks with directed links,  $\ell(A_i, B_j)$  is the hop-count length of shortest directed path from  $A_i$  to  $B_j$ . While the computation of the shortest paths in real-world scenarios are often hampered by incomplete information or other restrictions, using shortest paths is certainly appropriate when deriving theoretical bounds, as they constitute the best of worlds in terms of routing. Also, we define the distance of  $B_j$  from a set of nodes  $\mathcal{A}$  as usual as  $\ell(\mathcal{A}, B_j) = \min\{\ell(A_i, B_j) : A_i \in \mathcal{A}\}$ .

### 3.3 Transport Capacity

The *transport capacity* is an important parameter of a wireless networks, which reflects the maximum sum of the rates and the number of transmissions of unicast sessions. By computing this parameter in wireless network one can estimate the average rate. The transport capacity of a set of source-terminal pairs  $\mathcal{AB}$ , is defined as:

$$C_T(\mathcal{AB}) := \max_{\text{multi-hop paths}} \sum_k |A_k - B_k| R_k \quad (3.2)$$

where  $R_k$  is the average rate of unicast session between of  $A_k$  and  $B_k$  over a given multi-hop path. The maximum is taken over all possible multi-hop routes establishing the required connections between the sources and terminals. A simple upper bound which actually does not depend on the set  $\mathcal{AB}$  is found by noting that for the simultaneous routes achieving  $C_T$  there must be a time instance where the simultaneous direct hop-forwarding transmission reach at least  $C_T$  [10]:

$$C_T(\mathcal{AB}) \leq \max_{\mathcal{SD}} \sum_{(S_i, D_i) \in \mathcal{SD}} |S_i - D_i| W_i \quad (3.3)$$

where the maximum is over all possible sets of simultaneous, direct transmissions  $\mathcal{SD}$ , also  $W_i$  is the transmission rate of  $(S_i, D_i)$  transmitter-receiver pair. The transmission rate is computed based on the channel model.

### 3.4 Flow Schemes and Coding Schemes

We will use in our arguments a simple non-coding scheme which routes as commodity flows (replication, forwarding), a scheme we call the *flow scheme*. In contrast, we denote by a *coding scheme* any scheme using all of the operations of a flow scheme and in addition allowing the packets to be decoded or recoded at each node where they are received. In addition, in a coding scheme, intermediate nodes can send the results obtained from applying arbitrary functions to all previously received data and their own source data such with the only restriction that each destination node is able to decode the data intended for it from all of its received bits and local data.

Also, we call a coding scheme *info-independent* if the scheme routes the information bits independent of the context of the bits, in other words, the set of transmissions used by the coding scheme does not depend on the information bits. Note that many proposed coding scheme such as as random codes [2] and linear codes [17, 19] are info-independent.

### 3.5 Data Streams and Energy Consumption Model

We call a single multicast session or multiple unicast sessions *optimally source coded*, if the data streams generated by the source nodes are in optimally compressed format and independent from each other. More precisely, let the random process  $X(A_i)$  denote the binary data stream generated by a source node  $A_i$ . We assume that  $X(A_i)$  is in some optimal data compression format. Short,  $X(A_i)$  has the maximum integral entropy rate.

In addition, we assume that the random processes  $X(A_i)$  for different  $A_i$  are independent of each other, meaning that different sources generate independent information. Agreeably, this manifests a simplifying yet appropriate assumption. Indeed, assuming some dependence among the generated information leads naturally to Distributed Source Coding (DSC) techniques for maximizing the throughput [24]. However, the focus of this paper lies in pure *network coding gain* and not in (single nor multiple) source coding gain.

We quantize the *consumed energy* for transporting information by assuming that every transmission sends a constant amount of information, i.e. 1 bit. Consequently, the consumed energy for transporting the information becomes proportional to the *number of transmissions*. Note that we assume

that each transmission consumes the same amount of energy, independently of distance.

As a particular consequence of our setting, we assume that the *header* is not used as information, so the only 1 bit of information that is transmitted is the content of each transmission. Consequently, we do not consider the overhead provided by routing, nor potential ways to exploit routing information. In addition, we assume that the information cannot be communicated between the nodes without transmission, in other words, timing or omission of transmissions does not provide any new information to the receiver nodes.

From these assumptions, we can show that the expected number of transmissions for sending  $k$  specific bits of the data stream of some sources between two part of the network is at least  $k$ . Because the values of the bits are independent and each bit is equal to 0 or 1 with probability  $1/2$  and also each transmission delivers 1 bit. Note that the expectation is computed over different realizations of data streams of the sources.

In summary, the following assumptions are used for our analysis.

- A1 (Optimality source-coded data): Each source generates a data stream in an optimally compressed format.
- A2 (Independence of information for different sources): Different sources generate independent information.
- A3 (Energy consumption): The consumed energy for transporting the information is proportional to the number of transmissions and each transmission delivers 1 bit.

# Chapter 4

## Bounds on the Gain of Network Coding for a Single Multicast Session

In this section we bound the benefit of network coding in terms of energy savings and throughput for a single multicast session in wireless networks.

### 4.1 Energy Gain of Network Coding in Wireless Networks

Here, we study the energy consumption of a single multicast session in an arbitrary wireless networks. Examples for wireless networks are provided in [6, 7] in which the benefit of network coding in terms of reducing the number of transmissions (energy saving) of a broadcast session achieve the factors of 2 and 4/3, respectively. In [13], an example for a wired network with directed links is depicted where the gain of network coding on the throughput and energy is proportional to  $\log(\#V)$ , where  $\#V$  is the number of nodes.

In Theorem 4.4 we prove that under assumption A1 and A3 the expected number of transmissions can be reduced by at most a constant factor for a single multicast session. Certainly, a coding scheme can reduce the number of transmissions of a multicast session. Nevertheless, as we will argue, the mutual location of source and terminals within the network topology together with the geometric properties of the wireless channel models enforce a minimal number of transmissions for transporting the bits under any arbitrary

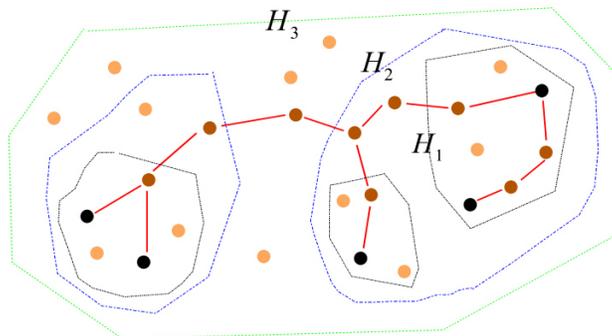


Figure 4.1: The nodes of  $\mathcal{M}$  have been colored black. The figure shows  $U$  which is built in 3 steps by considering the components of  $\mathcal{H}_1$ ,  $\mathcal{H}_2$ , and  $\mathcal{H}_3$ .

network coding. These minimal transmissions establish certain communication paths which can be exploited by a flow scheme to deliver the same information to the same terminals with just a constant factor more transmissions. As we show, this factor is surprisingly small.

The rigorous argument of Theorem 4.4 uses the following topological parameters. Let  $\mathcal{M}$  be the set of all terminals together with the source of the single multicast session. Next, let  $\mathcal{H}_0$  be a Maximal Independent Set (MIS) of  $\mathcal{M}$ , i.e.  $\mathcal{H}_0$  is a set with maximal size such that no two nodes of  $\mathcal{H}_0$  are neighbors in the connectivity graph  $G = (V, E)$ . By definition, we can show that every node of  $\mathcal{M}$  is either in  $\mathcal{H}_0$  or has at least one neighbor in  $\mathcal{H}_0$ . In wireless networks this means that every two nodes of  $\mathcal{H}_0$  are in distance larger than radio range ( $r$  for the Protocol Model and  $r_{\max}$  for the Physical Model) from each other. However, each node of  $\mathcal{M}$  is located inside of the radio range of at least one node of  $\mathcal{H}_0$ . Iteratively, let  $\mathcal{H}_i = \{u \in V : \ell(u, \mathcal{H}_{i-1}) \leq 1\}$  for  $i \geq 1$ . Clearly,  $\mathcal{H}_0 \subseteq \mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq \dots$  (see Fig. 4.1). We also define  $\mathcal{G}_0 = \mathcal{H}_0$  and  $\mathcal{G}_i = \mathcal{H}_i \setminus \mathcal{H}_{i-1}$ .

The following notation will come in handy: Let  $\kappa_d$  be the maximum number of nodes that can be placed in the unit  $d$ -dimensional sphere, such that every two of them have a distance strictly larger than 1. We find quickly that

$$\kappa_d := \begin{cases} 2 & \text{if } d = 1 \\ 5 & \text{if } d = 2 \\ 13 & \text{if } d = 3 \end{cases} \quad (4.1)$$

Further let

$$k := \min\{j : \text{the nodes of } \mathcal{H}_j \text{ are connected}\} \quad (4.2)$$

$$n_i := \#\{\text{connectivity components of } \mathcal{H}_i\} \quad (4.3)$$

Clearly,  $n_0 \geq n_1 \geq \dots \geq n_k = 1$ .

Next, we establish three lemmas which will be instrumental for the proof of Theorem 4.4.

**Lemma 4.1** *Denote the number of transmissions for sending a multicast bit under an arbitrary coding scheme by  $N$ . Then under assumptions A1 and A3,*

$$\mathbb{E}[N] \geq n_0/\kappa_d \quad (4.4)$$

Proof of Lemma 4.1: We compute a lower bound using some geometric properties of wireless networks and the channel model. By the geometric property of  $\kappa_d$  we find that every transmission can cover at most  $\kappa_d$  nodes of  $\mathcal{H}_0$ .

Next, we prove (4.4) using a contradiction argument. If (4.4) is not satisfied, then there exists a terminal in  $\mathcal{H}_0$  where the expected number of transmissions which it receives per multicast bit is less than 1. That node cannot decode all bits sent by the source correctly, because the source sends optimally compressed information.  $\diamond$

**Lemma 4.2** *Denote the number of transmissions for sending a multicast bit under an arbitrary coding scheme by  $N$ . Then under assumptions A1 and A3,*

$$\mathbb{E}[N] \geq n_1 + \dots + n_k \quad (4.5)$$

Proof of Lemma 4.2: Here, we derive the lower bound by considering the graph topology of the network introduced above. We consider the cutset between  $\mathcal{H}_{i-1}$  and  $V \setminus \mathcal{H}_{i-1}$  (the set of edges between  $\mathcal{G}_{i-1}$  and  $\mathcal{G}_i$ ). We claim that the expected number transmissions which is needed in the cutset for disseminating a multicast bit through  $\mathcal{H}_{i-1}$  components is at least  $n_i$ . Again, we prove the claim by a contradiction argument. If the claim is not true, then for at least one of the connectivity components of  $\mathcal{H}_{i-1}$  the expected number of transmissions in the cutset per multicast bit is less than 1. So, the terminals of that component cannot decode all generated bits correctly.

Note that we consider the direction of the transmission for a component of  $\mathcal{H}_{i-1}$  from  $\mathcal{G}_{i-1}$  to  $\mathcal{G}_i$  if the component contains the source node, and from

$\mathcal{G}_i$  to  $\mathcal{G}_{i-1}$  if it does not contain the source. Since,  $\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_k$  represent a partitioning over the set of nodes. Therefore, the expected number of transmissions for a multicast bit is at least  $n_1 + \dots + n_k$ .  $\diamond$

**Lemma 4.3** *There exists a flow scheme for the multicast session which transports every bit using  $N'$  transmissions, where*

$$N' \leq 3n_0 + 2(n_1 + \dots + n_k) - 2k - 1 \quad (4.6)$$

Proof of Lemma 4.3: For the proof it is enough to show that there exists a set of nodes  $U \subseteq V$ , such that the size of  $U$  is equal to  $N'$  and the induced graph by  $U$  is connected, contains the source node, and covers all terminals.

We Build  $U$  in  $k + 1$  steps. First, we set  $U_0 = \mathcal{H}_0$ . Clearly,  $U_0$  covers all nodes of  $\mathcal{M}$ . For  $1 \leq i \leq k$ , we let  $U_i$  be the set obtained by adding the minimal number of nodes  $U_{i-1}$  in order to reduce the number of connectivity components from  $n_{i-1}$  to  $n_i$ . The set  $U_k$  is our target set  $U$ .

Next, we bound the number of nodes which are added to  $U_{i-1}$  at step  $i$ . Without loss of generality, we may assume that  $n_{i-1} > n_i$ . This means  $n_{i-1} - n_i$  connections are created among the components of  $\mathcal{H}_{i-1}$  after building  $\mathcal{H}_i$ . We show that at most  $2i(n_{i-1} - n_i)$  nodes are need to create those connections among the corresponding components.

As an illustrative example, consider  $\mathcal{H}_{i-1}^{(1)}$  and  $\mathcal{H}_{i-1}^{(2)}$  as two components of  $\mathcal{H}_{i-1}$  which get connected after building  $\mathcal{H}_i$ . It follows from the definition of  $\mathcal{H}_i$  that there exists a path of length  $2i$  or  $2i + 1$  between two nodes of  $\mathcal{H}_0$  which belong to  $\mathcal{H}_{i-1}^{(1)}$  and  $\mathcal{H}_{i-1}^{(2)}$ . Because, from the definition, the nodes of  $\mathcal{H}_i$  are at most  $i$  hop away from at least one node in  $\mathcal{H}_0$ . So, two components are connected as we build  $\mathcal{H}_i$  from  $\mathcal{H}_{i-1}$  only if one of the depicted scenarios in Fig. 4.2 occurs. Then, we add  $2i - 1$  or  $2i$  nodes to the path which connect the terminals to  $U$ . This process is repeated  $n_{i-1} - n_i$  times to connect all the corresponding components of  $\mathcal{H}_{i-1}$ .

So, by continuing this algorithm and increasing  $i$ , the set  $U_k$  gets connected in the  $k^{\text{th}}$  step. Finally, we add the source node to  $U_k$  if it does not belong to it already. Clearly, the size of  $U$  is bounded by  $1 + n_0 + \sum_{i=1}^k 2i(n_{i-1} - n_i) = 3n_0 + 2(n_1 + n_2 + \dots + n_{k-1} + n_k) - 2k - 1$ .  $\diamond$

**Theorem 4.4** *Consider a single multicast session which is optimally source coded in an arbitrary wireless network. The gain of network coding in terms of reducing the expected number of transmissions is less than a factor of  $3\kappa_d + 2$ , where where  $\kappa_d$  is defined in (4.1).*

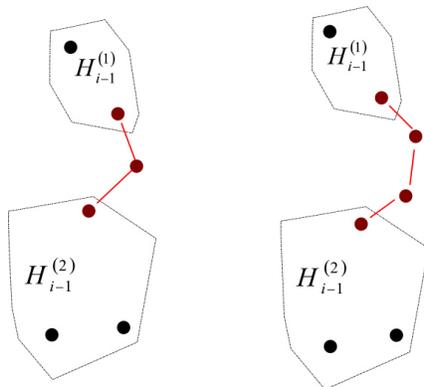


Figure 4.2: Two components  $\mathcal{H}_{i-1}^{(1)}$  and  $\mathcal{H}_{i-1}^{(2)}$  are connected after building  $\mathcal{H}_i$  if and only if one of these two scenarios occurs.

Proof of Theorem 4.4: Lemmas 4.1 and 4.2 give us lower bounds on the expected number of transmissions ( $\mathbb{E}[N]$ ) under any arbitrary network coding scheme. Also, Lemma 4.3 shows that there exists a flow scheme which uses  $N'$  transmissions per multicast bit. It follows that the benefit of network coding in terms of reducing the expected number of transmissions is bounded by  $N'/\mathbb{E}[N]$ . From (4.4), (4.5) and (4.6), it is easy to show that  $N'/\mathbb{E}[N] < 3\kappa_d + 2$ . ■

**Remark:** The proof of Theorem 4.4 also shows that the presented algorithm for building  $U$  in Lemma 4.3 can estimate the minimum multicast tree of a multicast session in a geometric graph up to the constant factor  $3\kappa_d + 2$ . To the best of our knowledge this, per se, constitutes a novel result in graph theory.

Next, we point out two particularly interesting scenarios, where concrete and simple bounds on the network coding gain can be given (see corollary 4.5). In the first case, the *long path scenario*, the source and the terminals are located far from each other (i.e.  $\gamma_1 \ll 1$ ). Then, the gain of network coding is bounded by a constant close to 2. The intuitive reason behind this result lies in the fact that any network coding scheme still needs to establish some disjoint paths to the terminals which are located far from each other. In the second case, explicit knowledge of the multicast tree leads to tighter bounds on the network coding gain.

**Corollary 4.5** *Assume that  $n_0, n_1, \dots, n_k$  are defined as (4.3) for an optimally source coded single multicast session. Then the gain of network coding*

on the expected number of transmissions is bounded as the following

(i) If  $n_0 = \gamma_1(n_1 + \dots + n_k)$ , the gain is bounded by a factor  $3\gamma_1 + 2$  for  $\gamma_1 < \kappa_d$  and  $\kappa_d(3 + 2/\gamma_1)$  for  $\gamma_1 > \kappa_d$ .

(ii) If there exists a multicast tree with size of  $N_m$  such that  $N_m \leq \gamma_2 \max(n_0/\kappa_d, n_1 + \dots + n_k)$ , the gain is bounded by a factor  $\gamma_2$ .

Proof of Corollary 4.5:

(i) The bound on  $N'/\mathbb{E}[N]$  can be easily computed using Lemmas 4.1, 4.2, and 4.3.

(ii) Obviously, we have  $N_m/\mathbb{E}[N] \leq \gamma_2$  using Lemmas 4.1 and 4.2. ■

## 4.2 Throughput Gain of Network Coding in Wireless Networks

The throughput of a single multicast sessions on a directed graph has been a popular case of study in network coding papers [2, 13, 17, 19]. Here we prove that network coding can increase the throughput of a single multicast session in wireless networks by at most a constant factor, in contrast to wired network where the throughput can be increased unboundedly large using network codes [13].

We emphasize that the results of [20, 23] which are obtained for wired network with bidirectional are not applicable for wireless network model. There are two main differences between the wired and wireless network models which make the network coding gain very different in these networks.

First, in the wireless networks we assume that a node receives one signal at a time (simple point-to-point coding for wireless communication). So, the receiving rate of every node is bounded by the wireless channel capacity. However, in the wired network model a node can receive several different bits over several links simultaneously, and there is no such bound for the receiving rate.

Second, in wireless network model, noise and interference constitute additional limiting factors for the network capacity. The SINR model for noise and interference which form a part of one of our channel models lead to geometric constraints which in turn imply the fundamental bounds we derive. However, the wired network model is not limited by such geometric constraints and can be in the form of any arbitrary directed or undirected graph.

Theorem 4.6 shows that network coding can increase the throughput of a single multicast session by at most a constant factor.

**Theorem 4.6** *Assume Protocol or Physical Model for the wireless channel. Consider a single multicast session which is optimally source coded in an arbitrary wireless network. Then, the network coding gain on the throughput of the multicast session is at most a constant factor  $c$ , where*

$$c := \begin{cases} [2 + (1 + \Delta)\sqrt{d+3}]^d \\ (5^d - 2^d) \left[ \sqrt{d} \left( 2 + \left( \frac{\beta \sum_{J \in \mathbb{Z}^d} |J|^{-\alpha}}{1 - \rho^{-\alpha}} \right)^{\frac{1}{\alpha}} \right) \right]^d \end{cases} \quad (4.7)$$

*under Protocol Model and Physical Model respectively ( $\rho > 1$  is a constant).*

Proof of Theorem 4.6: Consider an arbitrary terminal of the multicast session. The maximum rate at which data can be received by the terminal is equal to the wireless channel capacity  $W$ . By considering the cutset which separates the terminal from other nodes, we conclude that  $W$  is an upper bound on the throughput of a multicast session under any arbitrary network coding.

On the other hand, in [14, 16] flow scheme for the Protocol Model and the Physical Model are constructed with a broadcast throughput of  $W/c$  for any connected wireless network, where  $c$  is some constant determined by the channel model as (4.7). Such flow schemes can be used to send the data from the source to all terminals with rate  $W/c$ . This shows that network coding can improve the throughput of a single multicast session by at most a factor  $c$ . ■

Note that despite its apparent similarity to the proofs of previous work [14, 16], our argument is in fact different since network coding has not been taken into account in the mentioned existing work. Here, we study the gain of network coding on the throughput a multicast session.

# Chapter 5

## Bounds on the Gain of Network Coding for Multiple Unicast Sessions

In this section we study the benefit of network coding for multiple unicast sessions in terms of energy and throughput. By assumption, energy savings are proportional to savings in the number of transmissions needed. We first bound the gain of network coding in terms of the number of transmissions, then in terms of the throughput. Furthermore, we present novel criteria for evaluating the potential gain of info-independent coding scheme in wired/wireless networks.

### 5.1 Energy Gain of Network Coding

Here, we investigate the benefit of network coding in terms of energy saving for multiple unicast sessions. Note that in the single multicast scenario, network coding benefits by distributing the same information on different links and by efficiently using the links for sending the information toward several terminals. However, in multiple unicasts scenario, the links carry some independent information of different unicast sessions (assumption A2, Section 3.5). Intuitively, there is less chance for network coding to benefit in terms of the number of transmissions for unicast sessions. We will explore this fact more precisely later.

Theorem 5.1 provides a lower bound in terms of the expected number of

transmissions for multiple unicast sessions. The lower bound is determined simply by considering the hop-count distance of the source and terminal nodes. Interestingly, the theorem is also valid for wired networks under assumptions A1, A2, and A3 (Section 3.5).

Notably, the bounds of Theorem 5.1 apply to any set of optimally coded bits with independent information (assumptions A1 and A2), sent in unicast from a set of senders to their terminals. If the bits considered are such that the set of sources and the set of terminals are distant from each other, the lower bound becomes tighter.

**Theorem 5.1** *Assume that  $b_1, b_2, \dots, b_k$  are optimally source coded bits of some simultaneous unicast sessions. Denote  $N$  as the number of transmissions for transporting these bits in a given wired or wireless network. Also, denote the source and terminal of bit  $b_j$  by  $A_j$  and  $B_j$  for  $j = 1, \dots, k$ . Then, under any arbitrary coding scheme*

$$\mathbb{E}[N] \geq \max\left[\sum_{j=1}^k \ell(A_j, \mathcal{B}), \sum_{j=1}^k \ell(\mathcal{A}, B_j)\right] \quad (5.1)$$

We should point out that the source and the terminal of a unicast session are repeated equal to the number of bits of that the session has among  $b_1, b_2, \dots, b_k$  in the sum of equation (5.1). If the average rate of unicast sessions is given, we can choose the number of bits of each session proportional to its average rate in order to compute a lower bound for the rate of energy consumption.

Proof of Theorem 5.1: We group the nodes of the network in terms of their distances from the set of terminals ( $\mathcal{B}$ ). We define  $\tilde{\mathcal{H}}_0 = \mathcal{B}$  and  $\tilde{\mathcal{H}}_i = \{u \in V : \ell(u, \mathcal{B}) \leq i\}$ . Also, we define  $\tilde{\mathcal{G}}_0 = \tilde{\mathcal{H}}_0$  and  $\tilde{\mathcal{G}}_i = \tilde{\mathcal{H}}_i \setminus \tilde{\mathcal{H}}_{i-1}$ . It is easy to show that the nodes of  $\tilde{\mathcal{G}}_i$  are connected to the nodes of  $\tilde{\mathcal{G}}_{i-1}$ ,  $\tilde{\mathcal{G}}_i$  and  $\tilde{\mathcal{G}}_{i+1}$ . Fig. 5.1 shows the transmissions which carry the bits among different groups.

Denote the number of sources which are not in  $\tilde{\mathcal{H}}_i$  by  $q_i$ . Clearly,  $q_i = \sum_j \mathbb{I}_{[\ell(A_j, \mathcal{B}) > i]}$ . Note that the cutset between  $V \setminus \tilde{\mathcal{H}}_i$  to  $\tilde{\mathcal{H}}_i$  is the set of edges between  $\tilde{\mathcal{G}}_{i+1}$  and  $\tilde{\mathcal{G}}_i$  (in directed graph, consider the edges directed from  $\tilde{\mathcal{G}}_{i+1}$  to  $\tilde{\mathcal{G}}_i$ ).

From the assumptions  $b_1, b_2, \dots, b_k$  are independent optimally compressed bits. So, for transmitting them from  $V \setminus \tilde{\mathcal{H}}_i$  to  $\tilde{\mathcal{H}}_i$  the expected number of transmissions is at least  $q_i$ .

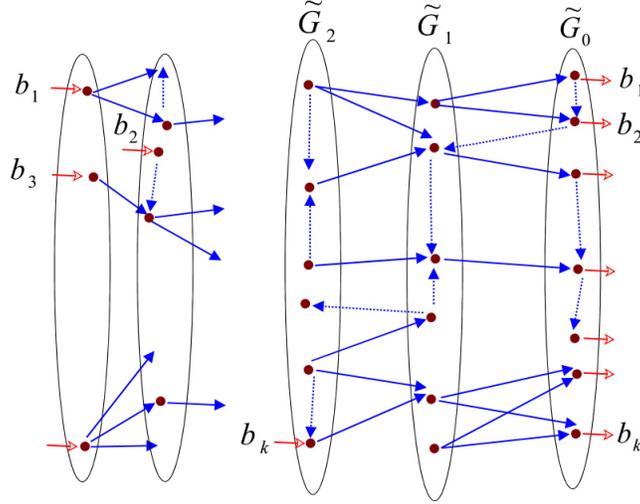


Figure 5.1: Grouping the nodes in terms of their distances from the set of terminals.

Therefore, the expected number of transmissions for transporting the bits all unicast sessions under any arbitrary network coding is at least  $\sum_{i=0}^{\infty} q_i$ . Then we have

$$\begin{aligned} \mathbb{E}[N] &\geq \sum_{i=0}^{\infty} q_i = \sum_{i=0}^{\infty} \sum_{j=1}^k \mathbb{I}_{[\ell(A_j, \mathcal{B}) > i]} \\ &= \sum_{j=1}^k \sum_{i=0}^{\infty} \mathbb{I}_{[\ell(A_j, \mathcal{B}) > i]} = \sum_{j=1}^k \ell(A_j, \mathcal{B}) \end{aligned}$$

Similarly, by grouping the nodes in terms of distance from the set of sources ( $\mathcal{A}$ ) and applying the same method, we can show that  $\mathbb{E}[N] \geq \sum_{j=1}^k \ell(\mathcal{A}, B_j)$ . ■

Two symmetrical cases of particular interest lead to a simple yet striking conclusion of when network coding should not be employed. Namely the case when independent unicast connections have *either the same source or the same terminal*. Surprisingly, network coding provides *no benefit* at all in terms of the number of transmissions in these cases (see Corollary 5.2).

Examples of such scenarios include *sensor networks* where the sensors send independent sensing information to a set of sinks. Note that every

sensor needs to send its information to at least one of the sinks, but it does not matter which one. If we consider a virtual node (say a super-sink) which receives the information of the sinks by direct (wired) links, then the traffic becomes similar to case (i) of Corollary 5.2 below.

Another concrete example concerns a certain part of traffic in mesh networks with an *internet gateway*, namely independent uplink traffic toward gateway (case (i)) and independent downlink traffic from the gateway (case (ii)). So, in mesh networks, network coding does not benefit the energy consumption for uplink and downlink traffic separately. However, note that network coding would still benefit by combining the uplink and downlink packets or for transporting multicast and broadcast packets.

Note that [24] studies case (i) more generally by considering a cost function in a single sink sensor networks. The paper shows that when the sources are independent and the cost is proportional to the number of transmissions, then shortest path routing protocol has the minimum cost. This agrees with our result which is proved in by a different technique.

**Corollary 5.2** *Consider multiple unicast sessions with optimally source coded data stream. For the following two scenarios, there is no network coding benefit in terms of the number of transmissions (energy).*

- (i) *A set of sources send their independent information to a single sink.*
- (ii) *A single source sends independent information to each of a set of terminals.*

Proof of Corollary 5.2: We establish the claim by showing that the shortest-path flow scheme is the most efficient scheme among all flow and coding schemes. If we use the flow scheme which transports the bits over the shortest path, then the number of transmissions is equal to  $\sum_{i=1}^k \ell(A_i, \mathcal{B})$  in case (i) and equal to  $\sum_{i=1}^k \ell(\mathcal{A}, B_i)$  in case (ii) of Theorem 5.1. So, we conclude that network coding does not benefit in terms of reducing the number of transmissions for these scenarios. ■

## 5.2 Throughput Gain of Network Coding in Wireless Networks

Here, we bound the gain of network coding on the throughput of multiple unicast sessions in wireless networks. We study the transport capacity and the throughput of the sessions under an arbitrary coding scheme.

Theorem 5.3 shows that the maximum transport capacity of an arbitrary network under an arbitrary coding scheme is bounded by constant factor  $\pi$  of the maximum transport capacity computed under the flow scheme. Therefore, the gain of network coding on the maximum transport capacity of arbitrary wireless network is bounded by a factor of  $\pi$ .

Moreover, the theorem shows that the computed upper bounds on the transport capacity in previous works [1, 9, 10, 15] increase by at most factor of  $\pi$  when we employ network coding in an arbitrary wireless network.

**Theorem 5.3** *Denote the maximum transport capacity of an arbitrary wireless network by using the flow scheme and coding scheme by  $C_T^f$  and  $C_T^{nc}$  respectively. Then,*

$$C_T^{nc} \leq K_d \cdot C_T^f \quad (5.2)$$

where  $K_d = 2$  if  $d = 1$  and  $K_d = \pi$  if  $d = 2, 3$  ( $d$  is the dimension).

Proof of Theorem 5.3: We prove the theorem in  $d = 2$  dimensional space. The proof can be easily extended for  $d = 1, 3$  dimensional space using the same method.

First, we prove the following lemma to capture geometric properties of wireless networks.

**Lemma 5.4** *Consider an arbitrary set of vectors  $Q = \{\vec{a}_1, \dots, \vec{a}_k\}$  in  $d$ -dimensional space. Then, there exists a unit vector  $\vec{i}$  and a set  $\underline{Q} \subseteq Q$  such that*

$$\sum_{\vec{a}_j \in Q} |\vec{a}_j| \leq K_d \sum_{\vec{a}_j \in \underline{Q}} \vec{i} \cdot \vec{a}_j \quad (5.3)$$

( $K_d$  is the same as Theorem 5.3).

Proof of Lemma 5.4: First we prove for  $d = 2$  dimensional space. Denote  $\vec{i}_\theta$  as the unit vector which has angle  $\theta$  from the horizontal axis. Also, denote the angles of vectors  $\vec{a}_1, \dots, \vec{a}_k$  from the horizontal axis by  $\varphi_1, \dots, \varphi_k$ .

By summing over  $\vec{a}_j$  and taking integral over  $\theta$ , we have

$$\begin{aligned} \int_0^{2\pi} \sum_{\vec{a}_j \in Q} |\vec{i}_\theta \cdot \vec{a}_j| d\theta &= \sum_{\vec{a}_j \in Q} \int_0^{2\pi} |\vec{a}_j| \cdot |\cos(\theta - \varphi_j)| d\theta \\ &= \sum_{\vec{a}_j \in Q} |\vec{a}_j| \int_0^{2\pi} |\cos \theta| d\theta = 4 \sum_{\vec{a}_j \in Q} |\vec{a}_j| \end{aligned}$$

By *mean value theorem*, there exists  $\theta_0$  such that

$$2\pi \sum_{\vec{a}_j \in Q} |\vec{i}_{\theta_0} \cdot \vec{a}_j| = 4 \sum_{\vec{a}_j \in Q} |\vec{a}_j| \quad (5.4)$$

Next, we partition  $Q$  into two sets:  $Q_1 = \{\vec{a}_j \in Q : \vec{i}_{\theta_0} \cdot \vec{a}_j \geq 0\}$  and  $Q_2 = V \setminus Q_1$ . Then,

$$\sum_{\vec{a}_j \in Q} |\vec{i}_{\theta_0} \cdot \vec{a}_j| = \sum_{\vec{a}_j \in Q_1} \vec{i}_{\theta_0} \cdot \vec{a}_j + \sum_{\vec{a}_j \in Q_2} (-\vec{i}_{\theta_0}) \cdot \vec{a}_j \quad (5.5)$$

By (5.4) and (5.5), we conclude one of the pairs  $(\vec{i}_{\theta_0}, Q_1)$  and  $(-\vec{i}_{\theta_0}, Q_2)$  can be considered as  $(\vec{i}, Q)$  to satisfy (5.3).

For  $d = 1$ , we set  $\theta_0 = 0$  and define  $Q_1$  and  $Q_2$  similarly. Then, it follows (5.3) for  $K_1 = 2$ .

For  $d = 3$ , we apply spherical coordinates. We write an unit vector  $\vec{k}$  as the sum two orthogonal vectors that one of them is parallel to  $\vec{a}_j$ . Then, we have

$$\int_{|\vec{k}|=1} |\vec{k} \cdot \vec{a}_j| = |\vec{a}_j| \int_0^\pi \int_0^{2\pi} |\cos(\theta)| d\theta \sin(\phi) d\phi = 8|\vec{a}_j| \quad (5.6)$$

On the other hand,  $\int_{|\vec{k}|=1} 1 = 4\pi$ . Similarly, there exists an unit vector  $\vec{k}_0$  such that  $4\pi \sum_{\vec{a}_j \in Q} |\vec{k}_0 \cdot \vec{a}_j| = 8 \sum_{\vec{a}_j \in Q} |\vec{a}_j|$ . For the rest, we build  $(\vec{k}_0, Q_1)$  and  $(-\vec{k}_0, Q_2)$  similarly.  $\diamond$

Now, consider an arbitrary set of unicast sessions  $\mathcal{AB} = \{(A_1, B_1), \dots, (A_k, B_k)\}$  with average rates  $R_1, \dots, R_k$ . We define  $\vec{a}_j = R_j \cdot \overrightarrow{A_j B_j}$ . By Lemma 5.4, there exists an unit vector  $\vec{i}$  and  $\underline{\mathcal{AB}} \subseteq \mathcal{AB}$  such that

$$\sum_{(A_j, B_j) \in \mathcal{AB}} R_j |\overrightarrow{A_j B_j}| \leq \pi \sum_{(A_j, B_j) \in \underline{\mathcal{AB}}} R_j \overrightarrow{A_j B_j} \cdot \vec{i} \quad (5.7)$$

Next, we rotate the Cartesian axes such that the axis  $X$  is aligned on the direction of unit vector  $\vec{i}$ . We denote the orthogonal line which crosses axis  $X$  at point  $x$  by  $l_x$  (see Fig. 5.2).

Also, we denote  $\mathbb{I}_{[A_j B_j \cap l_x]}$  as the indicator function that the line segment  $A_j B_j$  has been intersected by line  $l_x$ . Note that if  $(A_j, B_j) \in \underline{\mathcal{AB}}$  and  $\mathbb{I}_{[A_j B_j \cap l_x]} = 1$ , then we can show that  $A_j$  is located in the left side and  $B_j$  is located in the right side of  $l_x$  because  $\overrightarrow{A_j B_j} \cdot \vec{i} > 0$  (see the proof of

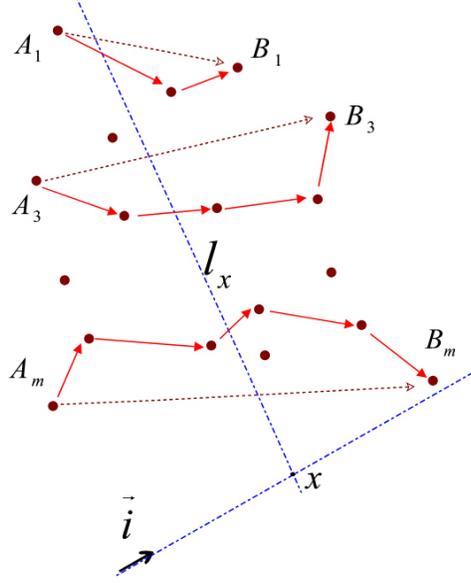


Figure 5.2:  $l_x$  is an geometric cutset for some source-terminal pairs of  $\underline{\mathcal{AB}}$ .

Lemma 5.4). Therefore,  $\sum_{(A_j, B_j) \in \underline{\mathcal{AB}}} R_j \mathbb{I}_{[A_j B_j \leftrightarrow l_x]}$  is the average rate of information for the unicast sessions of  $\underline{\mathcal{AB}}$  which go from the left to the right side of line  $l_x$ .

Denote the set of simultaneous transmissions at time instant  $\tau$  by  $\mathcal{SD}_\tau = \{(S_1, D_1), \dots, (S_m, D_m)\}$ . Then, the rate of bits are transmitted across  $l_x$  at this time instant is  $\sum_j W_j(\tau) \mathbb{I}_{[S_j D_j \leftrightarrow l_x]}$ . Therefore, from the definitions we have

$$\sum_{(A_j, B_j) \in \underline{\mathcal{AB}}} R_j \mathbb{I}_{[A_j B_j \leftrightarrow l_x]} \leq \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sum_{(S_j, D_j) \in \mathcal{SD}_\tau} W_j(\tau) \mathbb{I}_{[S_j D_j \leftrightarrow l_x]} d\tau \quad (5.8)$$

Now, we take integral by moving  $l_x$  from  $x = -\infty$  to  $\infty$ ,

$$\begin{aligned} & \sum_{(A_j, B_j) \in \underline{\mathcal{AB}}} R_j \overrightarrow{A_j B_j} \cdot \vec{i} \\ &= \sum_{(A_j, B_j) \in \underline{\mathcal{AB}}} R_j \int_{-\infty}^{\infty} \mathbb{I}_{[(A_j, B_j) \leftrightarrow l_x]} dx \\ &= \int_{-\infty}^{\infty} \sum_{(A_j, B_j) \in \underline{\mathcal{AB}}} R_j \mathbb{I}_{[(A_j, B_j) \leftrightarrow l_x]} dx \end{aligned}$$

$$\begin{aligned}
&\leq \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sum_{(S_j, D_j) \in \mathcal{SD}_\tau} W_j(\tau) \mathbb{I}_{[S_j D_j \rightsquigarrow l_x]} d\tau \, dx \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sum_{(S_j, D_j) \in \mathcal{SD}_\tau} W_j(\tau) \int_{-\infty}^{\infty} \mathbb{I}_{[S_j D_j \rightsquigarrow l_x]} dx \, d\tau \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sum_{(S_j, D_j) \in \mathcal{SD}_\tau} W_j(\tau) |\overrightarrow{S_j D_j} \cdot \vec{i}| \, d\tau \\
&\leq \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sum_{(S_j, D_j) \in \mathcal{SD}_\tau} W_j(\tau) |\overrightarrow{S_j D_j}| \, d\tau \\
&\leq \max_{\mathcal{SD}, \tau} \sum_{(S_j, D_j) \in \mathcal{SD}_\tau} W_j(\tau) |\overrightarrow{S_j D_j}| = C_T^f
\end{aligned}$$

Finally, by (5.7) we conclude that

$$C_T^{\text{nc}} = \max_{\mathcal{AB}} \left( \sum_{(A_j, B_j) \in \mathcal{AB}} R_j |\overrightarrow{A_j B_j}| \right) \leq \pi C_T^f \quad (5.9)$$

■

Next, we study the gain of network coding on the throughput of large homogeneous networks which is a popular model in network capacity papers [1, 8–10, 18]. Corollary 5.5 proves that network coding does not change the asymptotic behavior of the throughput of large homogeneous wireless networks.

Prior work (see [25]) shows a similar result under the Protocol Model for the wireless channel. Here, we establish the bound also under other channel models, i.e., the non-symmetric Protocol Model [9, 10], the Physical Model, and the Generalized Physical Models [1]). Further, our technique of proof is definitely different from [25] since we base our argument on the transport capacity.

**Corollary 5.5** *Network coding gain on the throughput of large homogeneous wireless networks is bounded by a constant factor.*

Proof of Corollary 5.5: In previous works [1, 9, 10] provide some asymptotic upper bounds on the throughput (sum of the rates of unicast sessions) of large homogeneous networks under various channel models. They do so in

two steps. First, an upper bound on the transport capacity ( $C_T^f$ ) is computed. Second, it is divided by the average distance of source-terminal pairs ( $L$ ). On the other hand, [8–10, 18] provide some flow schemes to achieve a throughput within a constant factor of the computed upper bounds. The throughput of these schemes represent tight lower bounds on the throughput capacity of the network.

Now, by Theorem 5.3, the gain of network coding on the transport capacity is bounded by a factor of  $K_d$ . Therefore, if we employ network coding then the throughput will be smaller than  $K_d$  multiply by the traditional upper bounds [1, 9, 10] (which have been computed for flow schemes). We also note the flow schemes can achieve the traditional upper bounds up to a constant factor [8–10, 18]. From these, we conclude that the network coding gain on the throughput is bounded by a constant. ■

### 5.3 Energy and Throughput Gain of Info-independent Coding Schemes

In this section, we study the potential gain of *info-independent* (see Section 3.4) coding schemes. Note that most proposed network coding schemes such as random codes [2] and linear codes [17, 19] are info-independent, thereby, the results of this section are valid for those schemes as well.

Our main tools, Theorems 5.6 and 5.10 relate an arbitrary info-independent coding scheme to a flow scheme. These theorems can be used as criteria for evaluating the potential gain of network coding on multiple unicast sessions in a given wired/wireless network. Putting these theorems to work, we identify several scenarios in Corollary 5.7 to 5.11 where network coding cannot improve the network in terms of throughput or energy savings.

**Theorem 5.6** *Under assumptions A1, A2, and A3 (see Section 3.5), for any info-independent coding scheme there exist a permutation of the destinations and a flow scheme that transports the same bits from the sources to the permutation of destinations using only a subset of the transmissions of the coding scheme.*

Proof of Theorem 5.6: Let us start the proof with a more detailed statement.

*Assume that an info-independent coding scheme transports optimally source coded bits  $b_1, \dots, b_k$ . Denote the source and destination of bit  $b_j$  by  $A_j$*

and  $B_j$  for  $j = 1, 2, \dots, k$  (a node can be the source or destination for several bits). Then, there exists a permutation  $\sigma(\cdot)$  and a flow scheme from  $A_j$  to  $B_{\sigma(j)}$  for  $j = 1, 2, \dots, k$  that can transport the same bits using a subset of transmissions used by the coding scheme; equivalently, the flow scheme can transport the same bits from  $A_{\sigma'(j)}$  to  $B_j$  for  $j = 1, 2, \dots, k$  where  $\sigma'(\cdot)$  is the inverse of  $\sigma(\cdot)$ .

We prove the statement for wired and wireless networks separately, since the graph models of these two are different.

(i) *Wired Networks:* We add a virtual *super-source* node  $A_0$  to the set of nodes. We assume that  $A_0$  generates the bits  $b_1, \dots, b_k$  and send them through  $k$  different links to  $A_1, \dots, A_k$  (each link transports one bit). Similarly, we add a virtual *super-destination* node  $B_0$  and assume that it receives the bits from  $B_1, \dots, B_k$  through different links. In other words, we assume that the bits are transported from  $A_0$  to  $B_0$ .

We construct a directed graph  $\bar{G} = (\bar{V}, \bar{E})$  by considering the set of transmissions which transport the bits in the wired network. We set  $\bar{V} = V \cup \{A_0, B_0\}$ . For every transmission (a transmission transports one bit), we consider a directed edge between its sender and its receiver. Note that there would exist more than one edge between two nodes of  $\bar{G}$ . The graph  $\bar{G}$  depicts how the bits are generated, combined, and transported at different stages from  $A_0$  to  $B_0$  in the network.

Now, we claim that the size of a minimum cut-set directed from  $A_0$  to  $B_0$  in graph  $\bar{G}$  is larger than or equal to  $k$ . For proof, we consider an arbitrary cut-set between  $A_0$  and  $B_0$ . Note that each edge of the cut-set corresponds to transmission of one bit of combined information. From the assumption, the bits  $b_1, \dots, b_k$  are optimally source coded and the coding scheme is info-independent (i.e. the transmissions have no side information about the transported bits), therefore, at least  $k$  transmissions (directed edges on the cut-set) are needed in order to losslessly transport the bits from the part which contains  $A_0$  to the parts which contains  $B_0$ .

Next, we apply Menger's Theorem for edge-disjoint paths (or Min-cut Max-flow Theorem) [4]. By using our claim, the theorem proves that there exist at least  $k$  edge-disjoint paths from  $A_0$  and  $B_0$  in  $\bar{G}$ . From the construction of  $\bar{G}$ , we conclude that these  $k$  disjoint paths must connect  $A_j$  to  $B_{\sigma(j)}$  for  $j = 1, 2, \dots, k$  where  $\sigma(\cdot)$  is some permutation of  $(1, \dots, k)$ . We construct the flow scheme by considering the these paths which can be used for transporting the bits from  $A_j$  to  $B_{\sigma(j)}$  ( $j = 1, 2, \dots, k$ ).

(ii) *Wireless Networks:* In wireless networks, the broadcast nature of wireless channel possesses a different graph model. Here, when a node transmits several neighbor nodes can receive the packet simultaneously. We construct a directed graph  $\hat{G} = (\hat{V}, \hat{E})$  based on this property of the network.

First, we consider two nodes  $A_0$  and  $B_0$  in  $\hat{V}$  similar to the proof of wired network case. Next, we repeat every node of  $V$  in  $\hat{V}$  equal to the number that it transmits. For example, assume that  $v \in V$  transmits  $s$  times, then we consider  $v^{(1)}, \dots, v^{(s)}$  in  $\hat{V}$  which correspond to the transmissions of  $v$  ordered by the transmitting times. To be more clear on drawing the graph  $\hat{G}$ , we arrange the nodes of  $\hat{V}$  from left to right based on their transmitting times. We consider an edge from  $v^{(i)}$  to  $u^{(j)}$  if for some  $i' \geq i$ , the transmission of  $v^{(i')}$  is received by the node  $u$  before the time that the transmission of  $u^{(j)}$  occurs. Also, we consider edges from  $v^{(i)}$  to  $v^{(j)}$  for all  $v \in V$  and  $i < j$ . The graph  $\hat{G}$  depicts how the bits are combined and transported from the far left node  $A_0$  to the far right node  $B_0$  by different transmissions over the time.

Now, we claim that the size of a minimum disconnecting-vertex-set between  $A_0$  and  $B_0$  in  $\hat{G}$  is larger than or equal to  $k$ . The disconnecting-vertex-set is a set of nodes that if they are removed (with their edges), then there will exist no path from  $A_0$  to  $B_0$  in  $\hat{G}$ . The proof of the claim is similar to our proof for the wired network case. We note that a disconnecting-vertex-set forwards the optimally source codes bits from the left part (which contains  $A_0$ ) to the right part (which contains  $B_0$ ) in  $\hat{G}$ , also each node in  $\hat{V}$  corresponds to one transmission (carries one bit) and in the info-independent coding scheme there is no side information about the bits. Therefore, at least  $k$  transmissions are needed for losslessly transporting the bits from the part which contains  $A_0$  to the part which contains  $B_0$ . This clearly show that the size of disconnecting-vertex-set is at least  $k$ .

Then, we apply Menger's Theorem for vertex-disjoint paths [4]. By using our claim, the theorem proves that there exist at least  $k$  vertex-disjoint paths from  $A_0$  and  $B_0$  in  $\hat{G}$ . From the construction, we conclude that a set of transmissions which correspond the internal vertices of these paths can transport the bits from  $A_j$  to  $B_{\sigma(j)}$  for  $(1, 2, \dots, k)$  where  $\sigma(\cdot)$  is a permutation. ■

In Corollary 5.7, we compute a lower-bound for the energy consumption of info-independent coding schemes that is tighter than the given lower bound in Theorem 5.1. Next, in Corollary 5.8, we identify new scenarios where network coding does not have any gain on the energy savings.

**Corollary 5.7** *In the notation of Theorem 5.6, denote the number of transmissions used in an arbitrary info-independent coding scheme for transporting the bits by  $N$ . Then we have*

$$N \geq \min_{\sigma} \left( \sum_{j=1}^k \ell(A_j, B_{\sigma(j)}) \right) \quad (5.10)$$

Proof of Corollary 5.7: We consider an arbitrary info-independent coding scheme and apply Theorem 5.6. We find the corresponding permutation  $\sigma(\cdot)$  and flow scheme. The number of transmissions used by the flow scheme is at least  $\sum_{j=1}^k \ell(A_j, B_{\sigma(j)})$ . This is also a lower bound on  $N$ . Note that (5.10) takes the minimum of  $\sum_{j=1}^k \ell(A_j, B_{\sigma(j)})$  over all permutations  $\sigma(\cdot)$ , therefore it gives us a general lower bound on the number of transmissions of any info-independent coding scheme. ■

**Corollary 5.8** *In the notation of Theorem 5.6, assume that we have  $\sum_{j=1}^k \ell(A_j, B_j) = \min_{\sigma} \left( \sum_{j=1}^k \ell(A_j, B_{\sigma(j)}) \right)$ . Then, the flow scheme via shortest-path uses less transmissions than any info-independent coding scheme.*

Proof of Corollary 5.8: From Corollary 5.7, the number of transmissions used by any info-independent coding scheme is at least  $\min_{\sigma} \left( \sum_{j=1}^k \ell(A_j, B_{\sigma(j)}) \right)$ . Since, the flow scheme achieves this lower bound, so no coding scheme can be more energy efficient. ■

In Corollary 5.9, we introduce some network scenarios in wired/wireless networks where network coding has no benefit in terms of either throughput or energy saving in the network. Note that Corollary 5.9 covers the scenarios which discussed in Corollary 5.2.

Here, we demonstrate that for the scenarios where the sources or destinations of independent unicast sessions can be rearranged arbitrarily, then any arbitrary info-independent coding scheme can be replaced by a flow scheme with the same or a better performance. Below, we explain two scenarios with such properties.

We call the first scenario *multi-sink reporting*. In this scenario, a set of sources send (report) their independent information to a set of sinks, however the information can be sent to any sink (no matter which one). This scenario

often occurs in sensor networks, where the sensors act as sources sending independent information to the sinks.

We call the second scenario *multi-server downloading*. In this scenario, a set of terminals receive (download) independent unicast information from a set of sources (servers), however every source has access to the entire information which is received by the terminals. Such scenario occurs for the Internet sites which provides downloading services to independent clients by using multiple data servers.

**Corollary 5.9 [Scenarios of no network coding gain]** *There is no network-coding gain on multiple unicast sessions for “multi-sink reporting” or “multi-server downloading” scenarios.*

Proof of Corollary 5.9: Similar to the notation of Theorem 5.6, denote the source and destination of the transported bits using the coding scheme by  $A_1, \dots, A_k$  and  $B_1, \dots, B_k$ . Theorem 5.6 proves that there exists a flow scheme which transport the bits from  $A_1, \dots, A_k$  to  $B_{\sigma(1)}, \dots, B_{\sigma(k)}$ . Since, for “multi-sink reporting” scenario the destinations can be rearranged in arbitrary way, so the flow scheme can be employed for transporting the bits.

For “multi-server downloading” scenario, we consider the flow scheme which transports the bits from  $A_{\sigma'(1)}, \dots, A_{\sigma'(k)}$  to  $B_1, \dots, B_k$ . Since, the information bits are available in all sources, so we can use the flow scheme for transporting the bits to their corresponding terminals. ■

In Theorem 5.10, we relate an info-independent coding scheme to a flow scheme when the average rates of the unicast sessions are given in the network. In Corollary 5.11, we identify some scenarios where network coding does not have any gain on the throughput. Note that the throughput of a multiple unicast sessions is defined by a vector  $\mathbf{R} = (R_1, \dots, R_k)$  which represents the average rates of the sessions. The corollary generally considers an arbitrary partial ordering relation ( $\leq^o$ ) defined for  $\mathbf{R} \in \mathbb{R}^k$ .

**Theorem 5.10** *Under assumptions A1, A2, and A3 (see Section 3.5), for any info-independent coding scheme there exists a flow scheme that transports information between the set of sources and the set of destinations at the same average rates for each session and that uses only a subset of the transmissions used by the coding scheme.*

**Remark:** Note that the flow scheme guarantees the transport of the bits between *sets* of sources and destinations but will not necessarily match the sources and destinations of the coding scheme. Still, such an application might become useful in certain scenarios.

Proof of Theorem 5.10: Let us start the proof with a more detailed statement.

*Assume that an info-independent coding scheme transports independent and optimally source coded information from source  $A_j$  to destination  $B_j$  at average rate  $R_j$  for  $j = 1, 2, \dots, k$ . Then, there exists a flow scheme which transports information from the set of sources  $\{A_1, \dots, A_k\}$  to the set of destinations  $\{B_1, \dots, B_k\}$  such that the average sending rate of  $A_j$  and receiving rate of  $B_j$  are equal to  $R_j$  for all  $j$ , and it uses a subset of transmissions used by the coding scheme.*

We consider the transported bits by the coding scheme in time period of  $[0, T]$  where  $T \rightarrow \infty$ . The scheme transports  $n_j \simeq R_j T$  bits from  $A_j$  to  $B_j$  for all  $j = 1, 2, \dots, k$ . We repeat  $A_j$  and  $B_j$  nodes  $n_j$  times for all  $j$ , so that a pair of source and destination is assigned to every transported bit. Then, we apply Theorem 5.6. The theorem proves there exists a flow scheme which transports  $\sum_{j=1}^k n_j$  bits from the (repeated) sources  $A_1, \dots, A_k$  to a permutation of the (repeated) destinations  $B_1, \dots, B_k$ . Since,  $A_j$  and  $B_j$  nodes are repeated  $n_j$  times as source and destination of the transported bits, we conclude that the average sending rate of  $A_j$  and receiving rate of  $B_j$  are equal to  $R_j \simeq \frac{n_j}{T}$  for all  $j$  in the flow scheme. ■

**Corollary 5.11** *Assume that a flow scheme  $\mathfrak{F}$  transports independent and optimally source coded information from source  $A_j$  to destination  $B_j$  at average rate  $R_j$  for  $j = 1, 2, \dots, k$ . Also, assume that for any other flow scheme  $\mathfrak{F}'$  which transports information from  $\{A_1, \dots, A_k\}$  to  $\{B_1, \dots, B_k\}$  such that the at average sending rate of  $A_j$  and receiving rate of  $B_j$  are  $R'_j$  ( $j = 1, 2, \dots, k$ ), we have  $\mathbf{R} \not\leq^o \mathbf{R}'$ , where  $\leq^o$  is an arbitrary partial ordering relation defined on  $\mathbb{R}^k$ . Then, there exists no info-independent coding scheme with a higher throughput (based on  $\leq^o$ ) than the flow scheme  $\mathfrak{F}$ .*

Proof of Corollary 5.11: From Theorem 5.10, for any info-independent coding scheme which transports information from  $A_j$  to  $B_j$  at rate  $R'_j$ , there exists a flow scheme  $\mathfrak{F}'$  which transports the  $\{A_1, \dots, A_k\}$  to  $\{B_1, \dots, B_k\}$  such that the average sending rate of  $A_j$  and receiving rate of  $B_j$  are  $R'_j$ . By the assumptions, we have  $\mathbf{R} \not\leq^o \mathbf{R}'$ . This shows that the rate vector of the coding

scheme  $(\mathbf{R}')$  is not higher than of the rate vector of  $\mathfrak{F}$  based on the partial ordering relation of  $\leq^o$ . ■

We can define  $\leq^o$  in several ways. For instance, we can consider the sum of the rates  $\|\mathbf{R}\|_{L1} = \sum_j R_j$  to define the partial ordering relation. Then, the corollary demonstrates that the throughput of a flow scheme which has the maximum sum of the rates among all permutations of source and destination pairs cannot be improved further using an info-independent coding scheme.

Also, we may find the following partial ordering relation useful for some scenarios:  $\mathbf{R} \leq^o \mathbf{R}'$  if  $R_j \leq R'_j$  for all  $j = 1, \dots, k$ . In this partial ordering relation, a scheme has a higher throughput than another scheme if and only if it provides a higher rate for all unicast sessions. For example, in the scenarios where the paths of unicast sessions intersect each other, increasing the rate of one session can decrease the rate of another session. Therefore, applying this partial ordering relation with Corollary 5.11 reveals more cases where the throughput cannot be improved further by employing info-independent coding schemes.

# Chapter 6

## Conclusion

In this work we studied fundamental limitations of the benefit of network coding for arbitrary wireless multihop networks. We focused on two popular network scenarios: single multicast session and multiple unicast sessions.

We proved the benefit of network coding in terms of throughput or energy saving is bounded by a constant factor for a single multicast session in wireless networks. This result was in contrast to wired case where the gain of network coding can be unboundedly large in some networks.

Also, we studied network coding on multiple unicast sessions in wireless networks. We computed bounds for the gain of network coding in terms of number of transmissions. Interestingly, we showed that network coding has no benefit in terms of energy in sensor networks where the sensors gather independent information for the sink or in mesh networks for unidirectional traffic from/towards the gateway. Notably, these bounds are not asymptotic and apply to any network.

Moreover, we proved that network coding can increase the transport capacity of an arbitrary wireless network by at most a factor of  $\pi$ . This result verifies that network coding does not change the throughput of large homogeneous networks more than a constant.

Furthermore, we studied the potential gain of info-independent coding schemes in wired/wireless networks. We presented theorems which relate an info-independent coding scheme to a flow scheme that transports the bits from the sources to a permutations of destinations. We showed that for several network scenarios which correspond some extreme cases of the theorems, network coding has no gain on throughput or energy saving.

Finally, from the established bounds, complemented with previous related

work on capacity bounds [1, 9, 10, 15, 25], we conclude that network coding plays valuable but not crucial role with regards to the performance of wireless networks. Rather, channel interference and topology of wireless networks seem to emerge as the determinant parameters on the throughput and energy consumption.

# Bibliography

- [1] A. Agarwal and P. R. Kumar. Capacity bounds for ad hoc and hybrid wireless networks. *Computer Communication Review*, 34(3):71–81, 2004.
- [2] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.
- [3] P. Chou, Y. Wu, and K. Jain. Practical network coding. In *Allerton*, 2003.
- [4] R. Diestel. *Graph Theory*. Springer, third edition, 2005.
- [5] R. Dougherty, C. F. Freiling, and K. Zeger. Insufficiency of linear coding in network information flow. *IEEE Transactions on Information Theory*, 51(8):2745–2759, 2005.
- [6] C. Fragouli, J. Widmer, and J. Y. L. Boudec. Energy efficient broadcasting in wireless ad-hoc networks using network coding. In *First Workshop on Network Coding*, 2005.
- [7] C. Fragouli, J. Widmer, and J. Y. L. Boudec. A network coding approach to energy efficient broadcasting: from theory to practice. In *INFOCOM*, 2006.
- [8] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran. On the throughput capacity of random wireless networks. *IEEE Transactions on Information Theory*, 2004.
- [9] P. Gupta and P. Kumar. Internets in the sky: The capacity of three dimensional wireless networks. *Communications in Information and Systems*, 1(1):33–50, 2001.

- [10] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, 2000.
- [11] T. Ho, M. Medard, M. Effros, and R. Koetter. Network coding for correlated sources. In *CISS*, 2004.
- [12] T. Ho, M. Medard, J. Shi, M. Effros, and D. R. Karger. On randomized network coding. In *Allerton*, 2003.
- [13] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. G. M. Tolhuizen. Polynomial time algorithms for multicast network code construction. *IEEE Transactions on Information Theory*, 51(6):1973–1982, 2005.
- [14] A. Keshavarz-Haddad, V. Ribeiro, and R. Riedi. Broadcast capacity in multihop wireless networks. In *MobiCom*, 2006.
- [15] A. Keshavarz-Haddad and R. Riedi. Bounds on the capacity of wireless networks based on topology and demands. In *MobiHoc*, 2007.
- [16] A. Keshavarz-Haddad and R. Riedi. On the broadcast capacity of multihop wireless networks: Interplay of power, density and interference. In *SECON*, 2007.
- [17] R. Koetter and M. Medard. An algebraic approach to network coding. *IEEE/ACM Transactions on Network*, 11(5):782–795, 2003.
- [18] S. R. Kulkarni and P. Viswanath. A deterministic approach to throughput scaling in wireless networks. *IEEE Transactions on Information Theory*, 50(6):1041–1049, 2004.
- [19] S. Y. R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49(2):371–381, 2003.
- [20] Z. Li and B. Li. Network coding in undirected networks. In *CISS*, 2004.
- [21] Z. Li and B. Li. Network coding: The case for multiple unicast sessions. In *Allerton*, 2004.
- [22] Z. Li and B. Li. On increasing end-to-end throughput in wireless ad hoc networks. In *Proceedings of the Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, 2005.

- [23] Z. Li, B. Li, and L. C. Lau. On achieving maximum multicast throughput in undirected networks. *IEEE/ACM Transactions on Networking*, 14(SI):2467–2485, 2006.
- [24] J. Liu, M. Adler, D. Towsley, and C. Zhang. On optimal communication cost for gathering correlated data through wireless sensor networks. In *MobiCom*, 2006.
- [25] J. Liu, D. Goeckel, and D. Towsley. Bounds on the gain of network coding and broadcasting in wireless networks. In *INFOCOM*, 2007.
- [26] D. S. Lun, M. Mard, and R. Koetter. Efficient operation of wireless packet networks using network coding. In *IWCT*, 2005.
- [27] D. S. Lun, M. Mard, and R. Koetter. Network coding for efficient wireless unicast. In *IEEE International Zurich Seminar on Communications*, 2006.
- [28] M. Mard, M. Effros, , D. Karge, and T. Ho. On coding for non-multicast networks. In *Allerton*, 2003.
- [29] A. Ramamoorthy, K. Jain, P. A. Chou, and M. Effros. Separating distributed source coding from network coding. *IEEE/ACM Transactions on Networking*, 14(SI), 2006.
- [30] A. Ramamoorthy, J. Shi, and R. D. Wesel. On the capacity of network coding for random networks. *IEEE Transactions on Information Theory*, 51(8):2878–2885, 2005.

## **Acknowledgment**

Financial support comes in part from NSF, grant number ANI-0338856.