On Computing the Capacity of Relay Networks in Polynomial Time

Farzad Parvaresh and Raúl Etkin

***Abstract*—The capacity or approximations to capacity of various single-source single-destination relay network models has been characterized in terms of the cut-set upper bound. In principle, a direct computation of this bound requires evaluating the cut capacity over exponentially many cuts. We show that the minimum cut capacity of a relay network under some special assumptions can be cast as a minimization of a submodular function, and as a result, can be computed efficiently. We use this result to show that the capacity, or an approximation to the capacity within a constant gap for the Gaussian, wireless erasure, and Avestimehr-Diggavi-Tse deterministic relay network models can be computed in polynomial time. We present some empirical results showing that computing constant-gap approximations to the capacity of Gaussian relay networks with around 300 nodes can be done in order of minutes.**

I. Introduction

Relay networks, where one or more source nodes send information to one or more destination nodes with the help of intermediate nodes acting as relays, are often used to model communication in wireless sensor networks. In sensor networks, sensor nodes have limited power sources and often require multi-hop communication with the help of intermediate nodes to reach the data aggregation centers. To guide the design of these networks it is of interest to characterize fundamental communication limits such as the capacity, which represents the maximum reliable communication rate.

Various communication models for relay networks capture in an abstract setting different aspects of practical systems. The wireless erasure network model of [7] captures the effect of packet losses in the wireless setting. The deterministic network model of Avestimehr, Diggavi and Tse (ADT) [3] incorporates broadcast and interference and can be used to gain insights about communication in more complex models that incorporate noise. Among these, of special importance is the Gaussian relay network, which models power limited transmitters and received signals corrupted by additive white Gaussian noise.

While the capacity of some network models (e.g. wireless erasure and ADT) is well characterized, the capacity of the Gaussian relay network, even in its simplest form with one transmitter, one relay, and one receiver, is in general unknown. The best known capacity upper bound is the so-called *cut-set bound*. A cut Ω of a network can be considered as a subset of nodes which includes the source node and excludes the destination node. For this cut, the capacity $F(\Omega)$ is defined as the maximum rate that information can be transferred form the nodes in Ω to the nodes that are not in Ω conditioned on the fact the information on Ω^c (the nodes that are not in Ω) is known. The cut-set upper bound is the *minimum* cut capacity over all the possible cuts.

F. Parvaresh and R. Etkin are with Hewlett-Packard Laboratories, Palo Alto, CA 94304, USA. (emails: {parvaresh, raul.etkin}@hp.com)

In the Gaussian setting, there are several capacity lower bounds based on different communication schemes, such as amplify-and-forward, decode-and-forward, compress-and-forward, quantize-and-forward, etc. [4], [6], [16]. Recently, Avestimehr, et al. [2] made significant progress in the capacity characterization of Gaussian relay networks by showing that a quantization and coding communication scheme can achieve a communication rate within a constant gap of the cut-set upper bound, where the gap only depends on the number of nodes in the network (i.e. it is independent of the channel gains and power levels). However, the evaluation of the achievable communication rate, which is necessary to implement the scheme, requires the computation of the cut-set bound for the network. Assuming that for a given cut the cut capacity is easy to compute, finding the cut-set upper bound can be a challenging problem. For a network with n relay nodes there are up to 2^n many different cuts and a greedy algorithm needs exponential time in the number of relay nodes to compute the cut-set bound.

In this work we show that the achievable rate of the scheme of [2] for the Gaussian relay network can be computed in polynomial time, and as a result, can be computed efficiently. This result is obtained by showing that the cut capacity of a fairly large class of networks under the assumption of independent encoding at the nodes in Ω is a submodular function. Existing results on minimization of submodular functions provide algorithms with polynomial running time $O(n^5\alpha + n^6)$, where α is the time that it takes to compute $F(\Omega)$ and n is the number of nodes in the network [18]. In addition, there exist possibly faster algorithms without polynomial time performance guarantees based on Wolfe's minimization norm algorithm [11]. In Section V, by simulations, we show that the cut-set bound for a Gaussian relay network with around 300 nodes can be computed on a laptop computer in about a minute using a MATLAB package for submodular minimization provided in [17].

Our results, extend and generalize previous results for the ADT model. This model can be seen as a high signal-to-noise-ratio (SNR) approximation of the Gaussian model, incorporating the effects of broadcasting and superposition of signals while de-emphasizing the effects of noise. Amaudruz

978-1-4577-0594-6/11/$26.00 ©2011 IEEE

et al. [1] showed that the cut-set bound for a *layered*[1] ADT model can be computed efficiently. They have extended graph flow algorithms such as Ford-Fulkerson's in a nontrivial way to find the maximum possible *linearly independent* (LI) paths in the network. They showed that the capacity of the network is equal to the maximum number of (LI) paths and can be computed in time $O(M \cdot |E| \cdot C^5)$, where M is the maximum number of nodes per layer, $|E|$ is the total number of edges and C is the capacity of the network. Moreover, they showed that the capacity can be achieved by using a very simple one-bit processing at the relay nodes. Later Goemans et al. [12] showed that the deterministic model is a special case of a flow model based on linking systems, a combinatorial structure with a tight connection to matroids. As a by-product, they obtained the submodularity of the cut capacity for layered ADT networks. Using this observation they provided various algorithms related to matroid theory to compute the cut capacity of the layered deterministic model based on finding intersection or partition of matroids. These results led to faster algorithms to compute the capacity of large layered ADT networks. In addition, there has been other extensions on improving the running time of the current algorithms for computing the capacity of ADT networks [8], [9], [20], [21].

In addition to showing that the capacity within a constant gap of the Gaussian relay network can be computed in polynomial time, our results allow us to compute in polynomial time the capacity of the wireless erasure network. Furthermore, we provide a simple proof for the computability in polynomial time of the capacity of the layered and non-layered ADT networks.

We obtain these results by considering a general framework to compute the cut-set bound. We assign transmit signal random variable X_i to node $i \in \mathcal{V}$ and we assume the probability distribution over the signals $X_1, X_2, \ldots, X_n$ to be independent, i.e $p(X_1, X_2, \ldots, X_n) = p_1(X_1)p_2(X_2)\cdots p_n(X_n)$. We also assign received signal random variables Y_i's to each node. The network is defined by the transition probability function $f(Y_1, Y_2, \ldots, Y_n|X_1, X_2, \ldots, X_n)$. We further assume that the transition probability function is of the form $f_1(Y_1|X_1, \ldots, X_n)\cdots f_n(Y_n|X_1, \ldots, X_n)$, meaning that the received signals are independent conditioned on the transmitted signals in the network. For such networks we show that $F(\Omega) = I(\mathbf{Y}_{\Omega^c}; \mathbf{X}_\Omega|\mathbf{X}_{\Omega^c})$[2] is submodular with respect to Ω. Later we show that for ADT networks, the Gaussian relay network and the wireless erasure network, we can find $p_1(X_1)\cdots p_n(X_n)$ such that $\min_\Omega F(\Omega)$ becomes equal to the capacity or the capacity within a constant gap. In other words, the min-cut problem for these networks can be cast as a minimization of a submodular function.

The paper is organized as follows. In Section III we show that for specific type of networks the cut value, $F(\Omega)$, is a submodular function. We then show in Section IV that for many wireless network models such as the ADT deterministic network, Gaussian relay network and wireless erasure network the capacity or an approximation to the capacity can be cast as a minimization of $F(\Omega)$. Finally, in Section V we describe results related to solving optimization problems involving submodular functions. We start by introducing the notation used in the rest of the paper.

[1]In a layered network, the nodes in one layer are only connected to the nodes in the next adjacent layer. In particular, there is no direct connection from source to destination.

[2]See Section II for a definition of the notation $\mathbf{X}_\Omega$, $\mathbf{Y}_{\Omega^c}$, etc.

II. Notation

Let $\mathcal{V}$ denote the set of nodes in the network and $|\mathcal{V}|$ its cardinality. For any subset A of nodes we denote by $\mathcal{V}\backslash A$ or A^c the set of nodes in $\mathcal{V}$ that are not in A. We assume $\mathcal{V}\backslash\mathcal{A} \cup \mathcal{B} = \mathcal{V}\backslash(\mathcal{A} \cup \mathcal{B})$. A cut Ω is defined as a subset of nodes in $\mathcal{V}$. A cut splits the nodes in the network into two groups, the nodes that are in Ω and the ones that belong to $\mathcal{V}\backslash\Omega$. Random variables are shown in capital letters such as X_i and Y_i. We use boldface letter for vectors, e.g. $\mathbf{x}$ is a constant vector and $\mathbf{X}$ is a random vector. We use $\mathbf{X}_\Omega$ to denote $(X_{v_1}, X_{v_2}, \ldots, X_{v_{|\Omega|}})$ with $v_i \in \Omega$. The function $I(X; Y|Z)$ is the mutual information between random variables X and Y conditioned on random variable Z. With a slight abuse of notation we use $H(X)$ to denote either the entropy or differential entropy of the discrete or continuous random variable X [5]. By $\mathbb{F}_p$ we denote a finite field with p elements. Finally, all the $\log(\cdot)$ functions are in base two.

III. Submodularity of cut-set function

Submodularity arises in many combinatorial optimization problems and large body of research has been developed on minimizing or maximizing submodular functions under various constraints.

A submodular function $f : 2^{\mathcal{V}} \to \mathbb{R}$ is defined as a function over subsets of $\mathcal{V}$ with *diminishing marginal returns*, i.e. if $A, B \subseteq \mathcal{V}$ with $A \subseteq B$ and any $v \in \mathcal{V}\backslash B$,

$$f(A \cup v) - f(A) \geqslant f(B \cup v) - f(B).$$

The theorem below establishes the submodularity of the cut capacity function of a general relay network under some special assumptions. This theorem will be used in Section IV to prove that the capacity or an approximation to the capacity of various specific relay network models can be computed by minimizing a submodular function.

Theorem 1. *Consider a network consisting of nodes in $\mathcal{V}$. Each node sends a message $X_i, i \in \mathcal{V}$ and receives $Y_i, i \in \mathcal{V}$. If the messages are independent $p(X_1, X_2, \ldots, X_{|\mathcal{V}|}) = p_1(X_1)p_2(X_2)\cdots p_{|\mathcal{V}|}(X_{|\mathcal{V}|})$ and conditioned on the sent messages the received messages are independent, then the function*

$$F(A) \;=\; I(\mathbf{X}_A; \mathbf{Y}_{\mathcal{V}\backslash A}|\mathbf{X}_{\mathcal{V}\backslash A}) \;\;,\; A \subseteq \mathcal{V}$$

is submodular.

Proof: To show that $F(A)$ is submodular we show that $F(A \cup a) - F(A)$ is monotonically non-increasing in A for

$a \notin A$.

$$\begin{aligned}
F(A\cup a) &= I(\mathbf{X}_{A\cup a};\mathbf{Y}_{\mathcal{V}\setminus A\cup a}|\mathbf{X}_{\mathcal{V}\setminus A\cup a})\\
&\overset{(a)}{=} H(\mathbf{X}_{A\cup a}|\mathbf{X}_{\mathcal{V}\setminus A\cup a}) - H(\mathbf{X}_{A\cup a}|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A\cup a})\\
&\overset{(b)}{=} H(\mathbf{X}_A) + H(X_a|\mathbf{X}_A) - H(X_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A\cup a})\\
&\quad - H(\mathbf{X}_A|X_a,\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A\cup a})\\
&= H(\mathbf{X}_A) + H(X_a|\mathbf{X}_A) - H(X_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A\cup a})\\
&\quad - H(\mathbf{X}_A|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A})
\end{aligned}$$

where (a) is the definition of mutual information and (b) is from the chain rule for the entropy function. Therefore,

$$\begin{aligned}
&F(A\cup a) - F(A)\\
&\quad = H(X_a|\mathbf{X}_A) - H(X_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A\cup a})\\
&\qquad\quad - H(\mathbf{X}_A|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A})\\
&\qquad\quad + H(\mathbf{X}_A|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},Y_a,\mathbf{X}_{\mathcal{V}\setminus A})\\
&\quad = H(X_a|\mathbf{X}_A) - H(X_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A\cup a})\\
&\qquad\quad - I(\mathbf{X}_A;Y_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A})\\
&\quad = H(X_a|\mathbf{X}_A) - H(X_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A\cup a})\\
&\qquad\quad - H(Y_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A})\\
&\qquad\quad + H(Y_a|\mathbf{X}_A,\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A})\\
&\quad = \underbrace{H(X_a|\mathbf{X}_A)}_{\text{non-increasing in } A} - \underbrace{H(X_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A})}_{\text{nondecreasing in } A}\\
&\qquad\quad \underbrace{- H(Y_a|\mathbf{Y}_{\mathcal{V}\setminus A\cup a},\mathbf{X}_{\mathcal{V}\setminus A})}_{\text{nondecreasing in } A} + H(Y_a|\mathbf{X}_{\mathcal{V}})
\end{aligned}$$

where the last equality follows because Y_a is independent of $\mathbf{Y}_{\mathcal{V}\setminus A\cup a}$ conditioned on $\mathbf{X}_{\mathcal{V}}$. So, $F(A\cup a) - F(A)$ is non-increasing in A and thus $F(A)$ is submodular. ∎

IV. Wireless network models

In this section, by applying the result of Theorem 1, we show that the capacity or an approximation to the capacity for the ADT deterministic network, Gaussian relay network, and wireless erasure network can be cast as a minimization of a submodular function.

A. Deterministic model (ADT)

We start by briefly describing the network model of [3][3]. In this model, each link from node i to node j has an associated non-negative integer gain n_{ij}. Each node $i \in \mathcal{V}$ transmits a signal $\mathbf{X}_i$ and receives a signal $\mathbf{Y}_i$, both in $\mathbb{F}_p^q$ where $q = \max_{i,j} n_{ij}$. At any given time, the received signal at node j is given by

$$\mathbf{Y}_j = \sum_{i\in\mathcal{V}\setminus\{d\}} \mathbf{S}^{q-n_{ij}}\mathbf{X}_i \qquad (1)$$

[3]Please, refer [3] for a more complete description of the model and its motivation.

where d is the destination node, the shifting matrix $\mathbf{S}$ is given by

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0\\ 1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

and the sums and products are in $\mathbb{F}_p$.

For a given cut Ω of the network, where Ω includes the source node and excludes the destination node, we can stack together the input vectors $\mathbf{X}_i, i \in \Omega$ and output vectors $\mathbf{Y}_i, i \in \Omega^c$, and define a transition matrix Λ_Ω that gives the input-output relationship of these vectors according to (1). It is shown in [2] that the capacity of the deterministic network is equal to $\min_\Omega \operatorname{rank}(\Lambda_\Omega)$. We show next in Theorem 2 that $\operatorname{rank}(\Lambda_\Omega)$ is submodular, and hence the capacity can be computed by minimizing a submodular function.

Proposition 1. *Assume an $m \times n$ matrix A over $\mathbb{F}_p$. Let $\mathcal{N}$ be the subspace $\mathcal{N} \overset{\text{def}}{=} \{\mathbf{x} \in \mathbb{F}_p^n \mid A\mathbf{x} = 0\}$, and let $\mathcal{G}$ be the set of cosets of $\mathcal{N}$ in $\mathbb{F}_p^n$. Pick $\hat{\mathbf{x}}_i$ to be an element in the ith coset of $\mathcal{N}$ for $i = 1, 2, \ldots, |\mathcal{G}|$, and set $\mathbf{y}_i = A\hat{\mathbf{x}}_i$. Notice that $\mathbf{y}_i \neq \mathbf{y}_j$ if $i \neq j$. Now, if we choose $\mathbf{x}$ uniformly at random from elements of $\mathbb{F}_p^n$ with probability $1/|\mathbb{F}_p^n|$, then the mapping $A\mathbf{x}$ maps $\mathbf{x}$ to $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{|\mathcal{G}|}\}$ uniformly at random with probability $1/|\mathcal{G}|$. In addition, the cosets of $\mathcal{N}$ form a partition of $\mathbb{F}_p^n$ into $p^n/|\mathcal{N}|$ sets. Also rank$(A) + \log_p(|\mathcal{N}|) = n$. Thus, $\log_p|\mathcal{G}| =$ rank(A).*

Theorem 2. *For a deterministic model, given a cut Ω assume Λ_Ω is the transition matrix form nodes in Ω to nodes in Ω^c. Set $D(\Omega) = rank(\Lambda_\Omega)$, then $D(\Omega)$ is submodular.*

Remark 1. *A special case of Theorem 2 for* layered *ADT networks was proved in earlier works [12], [21].*

Proof: In the network, assume node i sends b_i symbols $x_{i,1}, x_{i,2}, \ldots, x_{i,b_i}$ with $x_{i,j} \in \mathbb{F}_p$. We assume $x_{i,j}$'s drawn i.i.d. with uniform probability distribution over $\mathbb{F}_p$, i.e. $p(x_{i,j} = q) = 1/|\mathbb{F}_p|$ for all $q \in \mathbb{F}_p$. From the definition of transition matrix, Λ_Ω, if we assume for the cut Ω, $\mathbf{s} = (s_1, s_2, \ldots, s_k)^t$ symbols are being sent from nodes in Ω and $\mathbf{r} = (r_1, r_2, \ldots, r_\ell)^t$ symbols are being received by nodes in Ω^c then $\mathbf{r} = \Lambda_\Omega\mathbf{s}$. Then we can write

$$\begin{aligned}
I(\mathbf{X}_\Omega;\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}) &= H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}) - H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_\Omega,\mathbf{X}_{\Omega^c})\\
&\overset{(a)}{=} H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c})\\
&= H(\Lambda_\Omega\mathbf{s}|\mathbf{X}_{\Omega^c})\\
&\overset{(b)}{=} \log_p|\mathcal{G}| = \operatorname{rank}(\Lambda_\Omega)
\end{aligned}$$

where $\mathcal{G}$ is the set of cosets of $\mathcal{N}$ where $\mathcal{N} = \{\mathbf{s} : \Lambda_\Omega\mathbf{s} = 0\}$. Equality (a) is because $\mathbf{Y}_{\Omega^c}$ is a deterministic function of $\mathbf{X}_\Omega$ and (b) is the result of Proposition 1 and the fact the $\mathbf{s}$ has uniform probability distribution.

Notice that for the independent probability distribution on the sources the received signals are independent con-

ditioned on transmitted signals so, based on Theorem 1, $I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c})$ which is equal to $D(\Omega)$ is submodular. ■

B. Gaussian relay network

The Gaussian network model captures the effects of broadcasting, superposition and noise of power constrained wireless networks. In this model, at any time index (which we omit) the received signal at node $j \in \mathcal{V}\backslash\{s\}$ is given by

$$Y_j = \sum_{i \in \mathcal{V}\backslash\{d\}} h_{ij} X_i + N_j \tag{2}$$

where $X_i \in \mathbb{C}$ is the transmitted signal at node i, subject to an average power constraint $E(|X_i|^2) \leqslant 1$, $h_{ij} \in \mathbb{C}$ is the channel gain from node i to node j, and $N_j \in \mathcal{CN}(0,1)$ is additive white circularly symmetric complex Gaussian noise, independent for different j.

It has been show in [19, Theorem 2.1] that using lattice codes for transmission and quantization at the relays, all rates R between source $\{s\}$ and destination $\{d\}$ satisfying

$$R \leqslant \min_{\Omega} I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}) - |\mathcal{V}| \tag{3}$$

can be achieved, where Ω is a source-destination cut of the network and $\mathbf{X}_\Omega = \{X_i, i \in \Omega\}$ are i.i.d. $\mathcal{CN}(0,1)$. In addition, the restriction to i.i.d. Gaussian input distributions is within $|\mathcal{V}|$ bits/s/Hz of the cut-set upper bound [2]. Therefore the rate achieved using lattice codes in the above result is within $2|\mathcal{V}|$ bits/s/Hz of the capacity of the network.

The following corollary is an immediate consequence of Theorem 1.

Corollary 1. *The function $F(\Omega) = I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c})$ with the elements of $\mathbf{X}_\Omega$ being i.i.d. $\mathcal{CN}(0,1)$ is submodular.*

Due to Corollary 1 the minimization in (3) is the minimization of a submodular function and the resulting optimal value is within $2|\mathcal{V}|$ of the capacity of the network.[4].

C. Wireless erasure network

In [7] the authors introduce a special class of wireless networks, called wireless erasure networks. In these networks, a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ defines the interconnections between nodes. To model the broadcast effect of wireless networks, the signals on all outgoing arcs of any given node are equal to each other. There is no interference among multiple arcs arriving at a given node in this model, and the signals on the various arcs are erased independently of each other. We assume binary transmitted signals at each node, i.e. $X_i \in \{0,1\}, i \in \mathcal{V}\backslash\{d\}$, but all the results can be extended to models with larger input alphabets. It has been shown in [7] that the capacity of the network is

$$C = \min_{\Omega} F(\Omega) = \min_{\Omega} \sum_{i \in \Omega} \left(1 - \prod_{j \in \Omega^c} \epsilon_{ij}\right) \tag{4}$$

[4]Notice that $I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}) = \log\det(I + HH^\dagger)$ where H is the matrix of channel gains from nodes in Ω to nodes in Ω^c and $H^\dagger$ is the conjugate transpose of H. Therefore, it is easy to compute the capacity of each cut.

where ϵ_{ij} is the probability of erasure when node i is sending information to node j. We show in the following theorem that $F(\Omega)$ is submodular.

Theorem 3. *The function $F(\Omega) = \sum_{i\in\Omega}\left(1 - \prod_{j\in\Omega^c}\epsilon_{ij}\right)$ equals $I(\mathbf{X}_\Omega; \mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c})$ where X_i are i.i.d. $\sim$ Bernoulli$(1/2)$ for $i \in \Omega$. Therefore, $F(\Omega)$ is submodular.*

Proof: For i.i.d. $X_i \sim$ Bernoulli$(1/2)$, we can write

$$\begin{aligned}
&I(\mathbf{X}_\Omega;\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}) \\
&\quad= H(\mathbf{X}_\Omega|\mathbf{X}_{\Omega^c}) - H(\mathbf{X}_\Omega|\mathbf{Y}_{\Omega^c}, \mathbf{X}_{\Omega^c}) \\
&\quad\overset{(a)}{=} \sum_{i\in\Omega} \left(H(X_i) - H(X_i|\mathbf{Y}_{\Omega^c})\right) \\
&\quad\overset{(b)}{=} \sum_{i\in\Omega} \left(1 - H(X_i|\mathbf{Y}_{\Omega^c})\right) \\
&\quad= \sum_{i\in\Omega} \Big(1- \\
&\qquad\qquad \sum_{y_j\in\{1,0,e\}, j\in\Omega^c} H(X_i|Y_j = y_j, j\in\Omega^c) p(Y_j = y_j, j \in \Omega^c)\Big) \\
&\quad= \sum_{i\in\Omega} \Big(1 - H(X_i|Y_j = e, j\in\Omega^c) p(Y_j = e, j\in\Omega^c)\Big) \\
&\quad\overset{(c)}{=} \sum_{i\in\Omega} \left(1 - \prod_{j\in\Omega^c} \epsilon_{ij}\right).
\end{aligned}$$

We used in (a) the independence among X_i and the channel erasures, in (b) the fact that for $X_i \sim$ Bernoulli(1/2), $H(X_i) = 1$, and in (c) the fact that for $X_i \sim$ Bernoulli(1/2), $H(X_i|Y_j = e, j \in \Omega^c) = 1$ and for independent erasures we have $p(Y_j = e, j \in \Omega^c) = \prod_{j\in\Omega^c}\epsilon_{ij}$. Theorem 1 can be applied to conclude that $F(\Omega)$ is submodular. ■

V. Algorithms and simulations

One approach to solve the submodular minimization problem due to Lovász is based on *extension* of the set function $f : 2^{\mathcal{V}} \to \mathbb{R}$ to a convex function $g : [0,1]^{|\mathcal{V}|} \to \mathbb{R}$ that agrees with f on the vertices of the hypercube $[0,1]^{|\mathcal{V}|}$, with a guarantee that $\min_{A\subseteq\mathcal{V}} f(A)$ is equal to $\min_{\mathbf{x}} g(\mathbf{x})$ for $\mathbf{x} \in [0,1]^{|\mathcal{V}|}$. In this section we assume the normalization $f(\emptyset) = 0$.

The Lovász extension g of *any* set function f can be defined as follows. For a given $\mathbf{x} \in [0,1]^{|\mathcal{V}|}$ order the elements of $\mathcal{V}$ such that $x(v_1) \geqslant x(v_2) \geqslant \cdots \geqslant x(v_n)$, where $x(v_i)$ is the v_ith element of the vector $\mathbf{x}$. Set $\lambda_0 = 1 - x(v_1)$, $\lambda_i = x(v_i) - x(v_{i+1})$, $\lambda_n = x(v_n)$, and

$$g(\mathbf{x}) \overset{\text{def}}{=} \sum_{i=1}^{n} \lambda_i f(\{v_1, v_2, \ldots, v_i\}).$$

Define $\mathbf{1}_\emptyset = \mathbf{0} \in \mathbb{R}^n$ and $\mathbf{1}_{\{v_1,v_2,\ldots,v_i\}}$ as an n dimensional vector such that the coordinates $v_1, v_2, \ldots, v_i$ are equal to one and all the other coordinates are equal to zero. Then, it is easy to see that $\mathbf{x} = \sum_{i=0}^{n} \lambda_i \mathbf{1}_{\{v_1,v_2,\ldots,v_i\}}$, $\sum_{i=0}^{n}\lambda_i = 1$ and $\lambda_i \geqslant 0$. So, $\mathbf{x}$ is a unique linear convex combination

of some vertices of the hypercube and $g(\mathbf{x})$ is linear convex combination of values of f on those vertices.

A key result is that f is submodular if and only if its Lovász extension g is a convex function [13], [10]. In addition, finding the minimum of the submodular function f over subsets of $\mathcal{V}$ is equivalent to finding the minimum of the convex function g in the hypercube $[0,1]^{|\mathcal{V}|}$. The optimization can be done in polynomial time using Ellipsoid algorithm [13].

There are other algorithms with faster running time to solve the submodular minimization problem [14], [15], [18]. To the best of our knowledge, the running time of the fastest algorithm is in the order of $O(n^5\alpha + n^6)$, where α is the time that the algorithms takes to compute $f(A)$ for any subset $A \subseteq \mathcal{V}$ [18]. For ADT networks, Gaussian relay networks, and erasure networks, α is the time to compute: the rank of $n \times n$ matrices, the determinant of $n \times n$ matrices, and equation (4), respectively.

However, for networks of large size, a complexity of $O(n^5\alpha + n^6)$ may still be computationally cumbersome. As a result, in these cases it is desirable to have faster algorithms. Recently, Fujishing [10], [11] showed that the minimization of any submodular function can be cast as a minimum norm optimization over the base polytope of f, $B_f = P_f \cap \{\mathbf{x} \mid \sum_{i\in\mathcal{V}} x(i) = f(\mathcal{V})\}$, where

$$P_f \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \forall A \subseteq \mathcal{V} : \sum_{i\in A} x(i) \leqslant f(A) \right\}$$

and the corresponding minimum norm optimization is

$$\text{minimize} \;\; ||\mathbf{x}||_2, \qquad \text{subject to } \mathbf{x} \in B_f. \tag{5}$$

Letting $\mathbf{x}^*$ be the solution of this minimization, the set $A^* = \{v_i : x^*(v_i) < 0\}$ is the solution to $\min_A f(A)$. Whether the above optimization problem can be solved in polynomial time is an open problem. However empirical studies [11] have shown that this algorithm has comparable or even faster running times than the other algorithms with polynomial time performance guarantees.

In our specific setting, for layered Gaussian relay networks of size up to around 300 nodes with 4 nodes per layer, we were able to find the approximate capacity (cf. (3)) in order of minutes on a laptop computer with a 2.8 GHz AMD Dual-Core Processor and 4 GB of memory (see Figure 1). In order to solve the minimization (5) we used the Matlab package provided in [17].

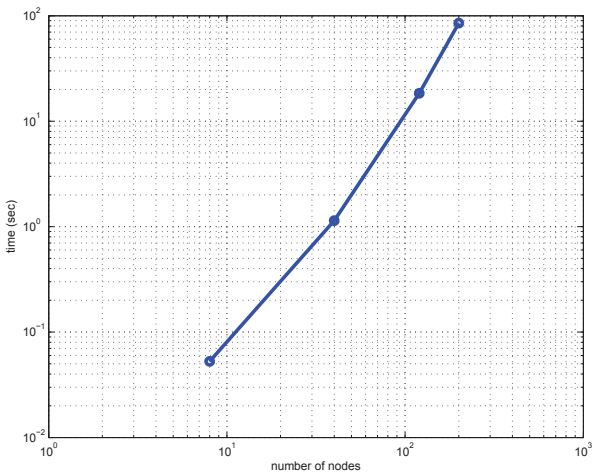

Fig. 1. Running time of minimum norm algorithm for a layered Gaussian relay network. Each layer has four nodes.

REFERENCES

[1] A. Amaudruz and C. Fragouli, "Combinatorial algorithms for wireless information flow," *SODA '09: Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2009.

[2] A. S. Avestimehr, S. N. Diggavi and D. N. C. Tse, "Approximate Capacity of Gaussian Relay Networks," *ISIT '08: IEEE international Symposium on Information Theory*, pp. 474–478, July 2008.

[3] S. Avestimehr, S.N. Diggavi and D. N C. Tse, "A deterministic approach to wireless relay networks," *Forty-Fifth Allerton Conference*, Illinois, September 2007.

[4] Brett E. Schein, *Distributed coordination in network information theory*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[6] T. M. Cover and A. El Gamal, "Capacity Theorems for the Relay Channel," *IEEE Trans. Info. Theory*, vol. 25, no. 5, Sept. 1979, pp. 572–584.

[7] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, M. Effros, "Capacity of wireless erasure networks," *IEEE Transactions on Information Theory,* vol.52, no.3, pp.789–804, March 2006

[8] J. Ebrahimi, C. Fragouli, "Combinatorial Algorithms for Wireless Information Flow," *under submission in ACM Transactions in Algorithms,* 2009. Also available at arXiv:0909.4808v1.

[9] E. Erez, Y. Xu and E. M. Yeh, "Coding for the Deterministic Network Model," *ITA workshop*, San Diego, 2010.

[10] S. Fujishige, *Submodular functions and optimization*,(Second Edition), Annals of Discrete Mathematics, Vol. 58, Elsevier, Piscataway, N.J., U.S.A., 2005.

[11] S. Fujishige, T. Hayashi, and S. Isotani, *The Minimum-Norm-Point Algorithm Applied to Submodular Function Minimization and Linear Programming,* Kyoto University, Kyoto, Japan, 2006.

[12] M. X. Goemans, S. Iwata and R. Zenklusen, "An Algorithmic Framework for Wireless Information Flow," *Forty–Seventh Annual Allerton Conference*, October 2009.

[13] M. Groetschel, L. Lovasz, and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," *In Combinatorica*, **1**:169–197, 1981.

[14] S. Iwata, "A faster scaling algorithm for minimizing submodular functions," *SIAM Journal on Computing*, 32(4):833–840, 2003.

[15] S. Iwata and J. B. Orlin, "A simple combinatorial algorithm for submodular function minimization," In *SODA '09* , pp 1230–1237, 2009.

[16] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory,* 51(9):3037–3063, September 2005.

[17] R. A. Krause, Matlab Toolbox for Submodular Function Optimization, *http://www.cs.caltech.edu/ krausea/sfo/*

[18] J. Orlin, "A faster strongly polynomial time algorithm for submodular function minimization," *Mathematical Programming*, 118(2): 237–251, 2009.

[19] A. Özgür, and S. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," *to be submitted to IEEE Journal on Selected Areas in Communications,* 2010. Available on–line at arXiv:1005.1284v1.

[20] C. Shi and A. Ramamoorthy, "Improved combinatorial algorithms for wireless information flow," *Forty–Eighth Annual Allerton Conference*, pp. 875–880, October 2010.

[21] S. M. Sadegh Tabatabaei Yazdi and S. A. Savari, "A combinatorial study of linear deterministic relay networks," *IEEE Information Theory Workshop (ITW),* pp. 1–5, Jan 6–8, 2010.