# Compute-and-Forward in Large Relaying Systems: Limitations and Asymptotically Optimal Scheduling 

Ori Shmuel, Asaf Cohen, Omer Gurewitz


#### Abstract

Compute and Forward (CF) is a coding scheme which enables receivers to decode linear combinations of simultaneously transmitted messages while exploiting the linear properties of lattice codes and the additive nature of a shared medium. The scheme was originally designed for relay networks, yet, it was found useful in other communication problems, such as MIMO communication. Works in the current literature assume a fixed number of transmitters and receivers in the system. However, following the increase in communication networks density, it is interesting to investigate the performance of CF when the number of transmitters is large.

In this work, we show that as the number of transmitters grows, CF becomes degenerated, in the sense that a relay prefers to decode only one (strongest) user instead of any other linear combination of the transmitted codewords, treating the other users as noise. Moreover, the system's sum-rate tends to zero as well. This makes scheduling necessary in order to maintain the superior abilities CF provides. We thus examine the problem of scheduling for CF. We start with insights on why good scheduling opportunities can be found. Then, we provide an asymptotically optimal, polynomial-time scheduling algorithm and analyze its performance. We conclude that with proper scheduling, CF is not merely non-degenerated, but, in fact, provides a gain for the system sum-rate, up to the optimal scaling law of $O(\log \log L)$.


## Index Terms

Relay Networks, User-Scheduling, Compute and Forward, lattice codes.

[^0]
## I. Introduction

Compute and Forward (CF) [1] is a coding scheme which enables receivers to decode linear combinations of transmitted messages, exploiting the broadcast nature of wireless networks. CF utilizes the shared medium and the fact that a receiver, which received multiple transmissions simultaneously, can treat them as a superposition of signals, and decode linear combinations of the transmitted messages. Specifically, with the use of lattice coding, the obtained signal after decoding can be considered as a linear combination of the transmitted messages. This is since for lattice codes, as a special case of linear codes, every linear combination of codewords is a codeword itself. However, since the wireless channel is subject to fading, the received signals are attenuated by real (and not integer) attenuations, hence the received linear combination is "noisy". In CF, the receiver (e.g., a relay) then seeks a set of integer coefficients, denoted by a vector a, to be as close as possible ${ }^{1}$ to the true channel coefficients and to serve as the coefficients for the linear combination it wishes to decode. The choice of coefficients, as well as the original channel gains, affect the resulting achievable rate.

Although CF was first introduced as a practical solution to mitigate users' interference in relays networks, the coding scheme and its properties were found to be an extremely useful tool in other types of communication channels and models. For example, linear MIMO receivers [2], decoding for the symmetric Gaussian K-user Interference Channel [3], cloud radio access networks [4] and even in physical-layer security [5]. In parallel, the CF scheme was extended to various settings which helped gain a deeper understanding of its abilities, advantages, and disadvantages. For example, MIMO CF [6], integration with interference alignment [7], scheduling in cellular networks [8], MAC (Multiple Access Channel) [9] and more [10]-[12].

In this work, we continue to shed light on the CF scheme and bring new results for a regime which, to the best of our knowledge, was not thoroughly discussed yet. That is, in contrast to the mentioned works, which assume that the number of simultaneously transmitting users is a fixed parameter of the system, in this work, we examine the effect of this number on the performance of the CF scheme. Specifically, we consider a general system of $L$ users and $M$ relays, and investigate the system sum-rate as $L$ increases. This asymptotic analysis, unlike the more common power asymptotic, takes a very realistic approach, which better fits current and future communication systems [13], [14]. Such systems tend to have a dense topology, with low

[^1]power and complexity devices, connected to a single or several access points. The ability of the CF scheme to handle simultaneous transmissions and use the users' interference constructively makes it a good candidate for such networks [4]. However, as we will show in the sequel, CF fails to perform well in such a regime, resulting in an asymptotically negligible rate. Yet, with proper scheduling and choices of coefficients, CF can retain its original benefits, and, in fact, distributively achieve the asymptotic upper bounds.

Numerous works discussed the capacity increase with the number of users, e.g., [15]-[17], and several have also showed how to gain multi-user diversity with scheduling [18]-[21]. However, we show that this is not necessarily the case with CF. In fact, a large number of simultaneous transmitters is not always a blessing when the number of relays is fixed, as when the former grows, the receiver will prefer to decode only the strongest user over all possible linear combinations. This will make the CF scheme degenerated, in the sense that a relay will choose a vector a which is actually a unit vector (a line in the identity matrix), thus treating all other signals as noise. In other words, the linear combination chosen will be trivial. Furthermore, we show that as the number of transmitters grows, the scheme's sum-rate goes to zero as well. Thus, one is forced to restrict the number of transmitting users, i.e., use scheduling, in order to maintain the superior abilities CF provides. We note that the idea of scheduling for CF networks was also considered in [22]. Therein, the authors showed by simulation that even a simple scheduling scheme can be useful. However, no performance guarantees or analysis for the optimal schedule were carried out.

Accordingly, we present a scheduling scheme for a large-scale relaying system that employs CF. The scheme is based on finding a group of users whose channel coefficients form a good coding opportunity for CF. Specifically, we present insight and analysis on what is considered as a good schedule for CF in terms of the channel's original coefficients and the coefficients of the approximated linear combination which is chosen to be decoded. We start our analysis from the point of view of a single relay, where we provide a lower bound on the expected system's sum-rate that can be achieved with our suggested scheduling scheme. Furthermore, we show that this bound is asymptotically optimal as the number of users grows. Then, we extend this scheduling policy for the case of a general system with multiple relays, for which we provide lower bounds on the expected system's sum-rate, prove the existence of good schedules for all relays simultaneously, and give heuristics for fast completion time.


Figure 1: Compute and Forward system model. $L$ transmitters communicate through a shared medium to $M$ relays.

## A. Paper Outline

The paper is organized as follows. In Section $\Pi$, the system model is described. In Section III, we present the main results, including both the results on the necessity of scheduling and results on the rates that can be achieved if scheduling is applied. Section IV brings the analytical derivation for the probability of choosing a unit vector by a relay, as the number of users grows. This is the essence of the proof for the necessity of scheduling. Then, in section V , we present our user scheduling algorithm and its analysis for a system of a single relay. Section VI extends the results to the case of multiple relays.

## II. System Model and Problem Statement

## A. Notational Conventions

Throughout the paper, we will use boldface lowercase to refer to vectors, e.g., $\mathbf{h} \in \mathbb{R}^{L}$, and boldface uppercase to refer to matrices, e.g., $\mathbf{H} \in \mathbb{R}^{M \times L}$. For a vector $\mathbf{h}$, we write $\|\mathbf{h}\|$ for its Euclidean norm, i.e. $\|\mathbf{h}\| \triangleq \sqrt{\sum_{i} h_{i}^{2}}$. We denote by $\mathbf{e}_{i}$ the unit vector with 1 at the $i$-th entry and zero elsewhere.

## B. System Model

Consider a multi-user multi-relay network, where $L$ transmitters (users) are communicating to a single destination $D$ via $M$ relays. All relays form a layer between the transmitters and the destination such that each transmitter can communicate with all the relays. Each transmitter has a length- $b$ message, assumed to be drawn $i . i . d$. and with uniform probability over a prime size finite field, that is, $\mathbf{w}_{l} \in \mathbb{F}_{p}^{b}, l=1,2, \ldots, L$, where $\mathbb{F}_{p}$ denotes the finite field with a set of $p$ elements.

This message is then encoded by an encoder $\mathcal{E}_{l}: \mathbb{F}_{p}^{b} \rightarrow \mathbb{R}^{n}$, which maps the messages to length- $n$ real-valued codewords, $\mathbf{x}_{l}=\mathcal{E}_{l}\left(\mathbf{w}_{l}\right)$. Each codeword is subject to a power constraint, $\left\|\mathbf{x}_{l}\right\|^{2} \leq n \mathrm{P}$. The message rate of each transmitter is defined by $R=\frac{b}{n} \log p$, measured in bits per channel use, and is equal for each transmitter ${ }^{2}$

Accordingly, each relay $m \in\{1, \ldots, M\}$ observes a noisy linear combination of the transmitted signals through the channel,

$$
\begin{equation*}
\mathbf{y}_{m}=\sum_{l=1}^{L} h_{m l} \mathbf{x}_{l}+\mathbf{z}_{m} \quad m=1,2, \ldots, M \tag{1}
\end{equation*}
$$

where $h_{m l} \sim \mathcal{N}(0,1)$ are the real ${ }^{3}$ channel coefficients and $\mathbf{z}$ is an i.i.d., Gaussian noise, $\mathbf{z} \sim$ $\mathcal{N}\left(0, \mathbf{I}^{n \times n}\right)$. Let $\mathbf{h}_{m}=\left(h_{m 1}, h_{m 2}, \ldots, h_{m L}\right)^{T}$ denote the vector of channel coefficients at relay $m$ and let $\mathbf{H}=\left(\mathbf{h}_{1}, \mathbf{h}_{2}, \ldots, \mathbf{h}_{M}\right)^{T}$ be the channel matrix, i.e., the m-th' row of $\mathbf{H}$ is $\mathbf{h}_{m}^{T}$. We assume a memoryless block-fading channel model, i.e., the channel remains constant over each slot (block) period, and at the beginning of each slot independent realizations of $\mathbf{H}$ are drawn; we assume that in each slot each relay knows its channel vector and that the channel vectors are independent of each other. This general model is illustrated in Figure 1.

## C. Compute and forward

Nazer and Gastpar [1] provided an achievable scheme which uses nested lattice codes for the computation of the linear equation of the transmitted signals over the channel (1). After receiving the noisy linear combination, each relay selects a scale coefficient $\alpha_{m} \in \mathbb{R}$, an integer coefficient vector $\mathbf{a}_{m}=\left(a_{m 1}, a_{m 2}, \ldots, a_{m L}\right)^{T} \in \mathbb{Z}^{L}$, and attempts to decode the lattice point $\sum_{l=1}^{L} a_{m l} \mathbf{x}_{l}$ from $\alpha_{m} \mathbf{y}_{m}$. Formally, the decoder has

$$
\begin{equation*}
\alpha_{m} \mathbf{y}_{m}=\sum_{l=1}^{L} \alpha_{m} h_{m l} \mathbf{x}_{l}+\alpha_{m} \mathbf{z}_{m}=\sum_{l=1}^{L} a_{m l} \mathbf{x}_{l}+\sum_{l=1}^{L}\left(\alpha_{m} h_{m l}-a_{m l}\right) \mathbf{x}_{l}+\alpha_{m} \mathbf{z}_{m} \tag{2}
\end{equation*}
$$

Due to the lattice algebraic structure, the relay decodes $\sum_{l=1}^{L} a_{m l} \mathbf{x}_{l}$ as a codeword while enduring the noise $\sum_{l=1}^{L}\left(\alpha_{m} h_{m l}-a_{m l}\right) \mathbf{x}_{l}+\alpha_{m} \mathbf{z}_{m}$, namely, an effective noise resulting from the true noise and the quantization error. The rate of the decoded codeword, i.e., the achievable

[^2]rate, defines a rate all transmitters must comply with to correctly decode the specific linear combination. The achievable rate and the optimal scale coefficient are given in the following two theorems.

Theorem 1 ([1], Theorem 1]). For real-valued AWGN networks with channel coefficient vectors $\mathbf{h}_{m} \in \mathbb{R}^{L}$ and coefficient vector $\mathbf{a}_{m} \in \mathbb{Z}^{L}$, the following computation rate region is achievable:

$$
\begin{equation*}
\mathcal{R}\left(\mathbf{h}_{m}, \mathbf{a}_{m}\right)=\max _{\alpha_{m} \in \mathbb{R}} \frac{1}{2} \log ^{+}\left(\frac{P}{\alpha_{m}^{2}+P\left\|\alpha_{m} \mathbf{h}_{m}-\mathbf{a}_{m}\right\|^{2}}\right), \tag{3}
\end{equation*}
$$

where $\log ^{+}(x) \triangleq \max \{\log (x), 0\}$.

Theorem 2 ([1, Theorem 2]). The computation rate given in Theorem 1$]$ is uniquely maximized by choosing $\alpha_{m}$ to be the MMSE coefficient

$$
\begin{equation*}
\alpha_{M M S E}=\frac{P \mathbf{h}_{m}^{T} \mathbf{a}_{m}}{1+P\left\|\mathbf{h}_{m}\right\|^{2}} \tag{4}
\end{equation*}
$$

which results in a computation rate region of

$$
\begin{equation*}
\mathcal{R}\left(\mathbf{h}_{m}, \mathbf{a}_{m}\right)=\frac{1}{2} \log ^{+}\left(\left\|\mathbf{a}_{m}\right\|^{2}-\frac{P\left(\mathbf{h}_{m}^{T} \mathbf{a}_{m}\right)^{2}}{1+P\left\|\mathbf{h}_{m}\right\|^{2}}\right)^{-1} . \tag{5}
\end{equation*}
$$

Note that the above theorems are for real channels and the rate expressions for the complex channel are twice the above [1, Theorems 3 and 4]. In addition, one should note that the coefficient vector $\mathbf{a}_{m}$ must satisfy,

$$
\begin{equation*}
\left\|\mathbf{a}_{m}\right\|^{2} \leq 1+\mathrm{P}\left\|\mathbf{h}_{m}\right\|^{2}, \tag{6}
\end{equation*}
$$

so that computation rate in (5) would not be zero [1, Lemma 1].
As mentioned, in order for relay $m$ to be able to decode a linear combination with coefficient vector $\mathbf{a}_{m}$, the rate of the messages which have a non zero entry in $\mathbf{a}_{m}$ must comply with the computation rate region [1]. From transmitter's $l$ viewpoint, this means,

$$
\begin{equation*}
R_{l}<\min _{m: a_{m l} \neq 0} \mathcal{R}\left(\mathbf{h}_{m}, \mathbf{a}_{m}\right) . \tag{7}
\end{equation*}
$$

However, since we assume all transmission rates are equal, we have, $R=\min _{m} \mathcal{R}\left(\mathbf{h}_{m}, \mathbf{a}_{m}\right)$ for all transmitters.

Recall that this work deals with the regime of large $L$ and a fixed $M$. Under this regime, the vectors (the channel and the coefficient vectors) in the rate expression in (5) have $L$ entries for any relay $m$. In the first part of this work, we analyze this rate and show that it tends to zero as $L$ grows, for all choices of $\mathbf{a}_{m}$. This will lead to the conclusion that a restriction
on the number of actually simultaneously transmitting users is required. That is, one must schedule users for transmission to achieve a non-zero rate. On the other hand, we will show that intelligent scheduling is indeed possible, resulting in strictly positive rates, and, in fact, scale-optimal performance.

## D. Compute and forward - Complete Decoding

The previous subsection provided the restriction on the achievable rate such that a single relay will be able to decode a single linear combination. For the complete decoding of all messages, the destination must acquire at least $L$ independent linear combinations on the $L$ messages and solve the linear system. The collecting process of the linear combinations can be done by multiple relays, multiple transmission slots or by a combination of the two. Thus, the ratio between $L$ and $M$ has a crucial effect on the time the destination will be able to decode all messages, i.e., the completion time. That is, it may happen that a single transmission of the messages would not suffice to decode all messages. Clearly, for $L \gg M$, we expect to have several transmission slots until complete decoding is possible. In each such slot, the users repeat their transmissions at the same rates and the destination collects $M$ coefficient vectors and their corresponding decoded codewords. We define by $\mathbf{A}_{n}=\left(\mathbf{a}_{1}(1), \ldots, \mathbf{a}_{M}(1), \ldots, \mathbf{a}_{1}(n), \ldots, \mathbf{a}_{M}(n)\right)^{T}$ the decoding matrix which is formed by the coefficient vectors collected in $n$ transmissions as it's rows.

Remark 1 (Number of transmission slots). We note here that the number of transmission slots may be reduced if a relay decodes several equations from the same transmission [3], [23]. However, this ability restricts the achievable rate to the lowest equation decoded. Moreover, the results in the first part of this work show that without proper scheduling the achievable rate goes to zero for any coefficient vector $\mathbf{a}$, in particular, the one which maximizes $\mathcal{R}(\mathbf{h}, \mathbf{a})$; thus, any other linear equation will result in an even lower rate.

We point out the following observations. First, note that since the channel changes between the slots, the transmission rates should also be constrained by the weakest time slot. Moreover, the throughput should be normalized by the number of slots used since a transmitter transmits the same message in each slot (until successful decoding). Second, since there is no coordination between the relays, the $M$ coefficient vectors decoded in each slot by the relays may be linearly dependent, either within a slot or across slots; thus, the number of Degrees of Freedom (DoF) may be less than $M$, increasing the number of slots needed for complete decoding. Accordingly,
in order to have a successful decoding of all messages at transmission slot $N$, we require that $\operatorname{rank}\left(\mathbf{A}_{N}\right)=L$ and that

$$
\begin{equation*}
R<\min _{n=1, \ldots, N} \min _{m} \mathcal{R}\left(\mathbf{h}_{m}(n), \mathbf{a}_{m}(n)\right) \tag{8}
\end{equation*}
$$

This assures that the linear system of equations can be solved and that each equation has the same rate and was decoded correctly by the relays. Note that $N$ is a r.v. since it depends implicitly on the channel vectors. Note also that the all linear operation are taken modulo $p$ [1] Theorem 7 and Remark 9]

We may now express the system sum-rate when successful decoding is possible, which is the sum of rates of the $L$ original messages (in our case, each of rate $R$ ) divided by the number of slots until complete decoding.

Definition 1 (The system sum-rate). The system sum-rate is,

$$
\begin{equation*}
C_{S R} \triangleq \frac{R L}{N} \tag{9}
\end{equation*}
$$

were $R$ is subject to (8) and $N$ is the first slot for which $\operatorname{rank}\left(\mathbf{A}_{N}\right)=L$.

## E. Problem statement

The performance of such a system depends on the proper choice of the coefficient vectors by the relays in each transmission slot. This choice rules the transmission rate of the messages and the completion time for successful decoding. Essentially, one would want to maximize the rate in each transmission slot, on one hand, i.e., find coefficient vectors that are the best fit for the real channel vectors. On the other hand, one would want to minimize the time by making sure these vectors contribute additional DoF in each slot. As mentioned, there is no coordination between the relays when it comes to the selection of these a vectors; therefore, for most of this work, we focus on the considerations for this selection from the perspective of a single relay, i.e., we consider a single relay model for which $M=1$. By this choice, one can remove, in general, the minimum constraint on the transmission rates as in (7), since there is only a single achievable rate, and analyze this system more easily. Hence, from this point and on, we omit the index $m$ and consider a single relay model. In Section VI we return to the case of multiple relays and elaborate on its implications.

Our goal is to maximize the system's sum-rate given in (9). Thus, since the rate and the completion time are both a function of the coefficient vectors selections made by the relay in each slot, we may write this maximization as follows,

$$
\begin{equation*}
\max _{\mathbf{A}_{N} \in \mathbb{Z}^{N \times L}} \frac{R\left(\mathbf{A}_{N}\right) L}{N\left(\mathbf{A}_{N}\right)}, \tag{10}
\end{equation*}
$$

where now $\mathbf{A}_{N}$ includes the coefficients vectors of the single relay. We note that one may suggest a sub-optimal solution for this problem by letting the relay choose, in each slot, the coefficient vector a that maximizes the achievable rate regardless the completion time $N{ }^{4}$. That is,

$$
\begin{equation*}
\mathbf{a}^{\max }=\underset{\mathbf{a} \in \mathbb{Z}^{L} \backslash\{\mathbf{0}\}}{\arg \max } \frac{1}{2} \log ^{+}\left(\|\mathbf{a}\|^{2}-\frac{\mathrm{P}\left(\mathbf{h}^{T} \mathbf{a}\right)^{2}}{1+\mathbf{P}\|\mathbf{h}\|^{2}}\right)^{-1} \tag{11}
\end{equation*}
$$

Accordingly, the destination would have to wait until collecting enough independent linear combinations to attain a full rank.

The problem of finding the maximizing a can be done by exhaustive search for small values of $L$. However, as $L$ grows, the problem becomes prohibitively complex quickly. In fact, it becomes a special case of the lattice reduction problem, which has been proved to be NP-complete. This can be seen if we write the maximization problem of (11) as an equivalent minimization problem [24]:

$$
\begin{equation*}
\mathbf{a}^{\max }=\underset{\mathbf{a} \in \mathbb{Z}^{L} \backslash\{\mathbf{0}\}}{\arg \min } f(\mathbf{a})=\mathbf{a}^{T} \mathbf{G a}, \tag{12}
\end{equation*}
$$

where $\mathbf{G}=\left(1+\mathrm{P}\|\mathbf{h}\|^{2}\right) \mathbf{I}-\mathbf{P h h}^{T} . \mathbf{G}$ can be regarded as the Gram matrix of a certain lattice and a will be the shortest basis vector and the one which minimizes $f$. This problem is also known as the shortest lattice vector problem (SLV), which has known approximation algorithms due to its hardness [25], [26]. The most notable of them is the LLL algorithm [27], [28] which has an exponential approximation factor that grows with the size of the dimension. However, for special lattices, efficient algorithms exist [29]. In [24], a polynomial complexity algorithm was introduced for the special case of finding the maximizing coefficient vector in CF .

In what follows, we show that regardless of complexity considerations, when $L$ grows the optimal solution for (10) tends to zero, and in order to promise a positive system sum-rate one must require that not all users transmit simultaneously. This extends the maximization problem in another dimension - where scheduling is also allowed. While the problem is even more complicated with scheduling, in this paper we show that simple scheduling algorithms exist, they result in non-negligible rates, hence overcome the limitations of a large $L$, and, moreover,

[^3]we present a low-complexity, asymptotically optimal algorithm for joint scheduling for CF which achieves the best possible scaling law.

## III. Main Results

The main results of this work can be divided into two main threads. The first thread shows that the optimal solution for (10) results in negligible rate if indeed all users transmit together. Hence, one must use user scheduling to avoid this pitfall. The second thread shows that under scheduling, not only strictly positive rates are possible, but actually such schedules can be easy to implement, distributed, and asymptotically achieve the optimal scaling laws for the expected system sum-rate. We summarize these in the following subsections. The proofs are given in Sections IV, V and VI.

## A. Necessity of Scheduling

When the number of relays is fixed and the number of transmitters is large, scheduling a small number of users is necessary for the CF scheme not to degenerate. Specifically, we first show that letting all users transmit will cause a relay, with a probability that goes to one with the number of transmitters, to choose a vector $\mathbf{e}_{i}$ as the coefficient vector which maximizes the achievable rate. This will result in decoding only the strongest user and not a non-trivial linear combination.

Theorem 3. Under the CF scheme, the probability that a non-trivial vector a will be the coefficient vector which maximize the achievable rate $\mathcal{R}(\mathbf{h}, \mathbf{a})$, i.e., minimize $f(\mathbf{a})$ in (12), compared to the best unit vector $\mathbf{e}_{i}$, i.e., $i=\arg \min _{i} f\left(\mathbf{e}_{i}\right)$, is upper bounded by

$$
\begin{equation*}
P_{r}\left(f(\mathbf{a}) \leq \min _{i} f\left(\mathbf{e}_{i}\right)\right) \leq 1-I_{\Phi(\mathbf{a})}\left(\frac{1}{2}, \frac{L-1}{2}\right) \tag{13}
\end{equation*}
$$

where $I_{x}(a, b)$ is the CDF of the Beta distribution with parameters $a$ and $b$, and $\Phi(\mathbf{a})=1-\frac{1}{\|\mathbf{a}\|^{2}}$.
Note that $\frac{1}{2} \leq \Phi(\mathbf{a}) \leq 1$ for any a which is not a unit vector.
The main consequence of Theorem 3 is the following.

Corollary 1. As the number of simultaneously transmitting users grows, the probability that a non-trivial a will be the maximizer for the achievable rate goes to zero. Specifically,

$$
\begin{equation*}
P_{r}\left(f(\mathbf{a}) \leq \min _{i} f\left(\mathbf{e}_{i}\right)\right) \leq e^{-L E_{1}(L)} \tag{14}
\end{equation*}
$$

where $\mathbf{a}$ is any integer vector that is not a unit vector, $\mathbf{e}_{i}$ is a unit vector and $E_{1}(L)=(1-$ $\left.\frac{3}{L}\right) \log \|\mathbf{a}\|$.

Corollary 1 clarifies that for every power P , as the number of users grows, the probability of having a non-trivial vector a as the maximizer of the achievable rate tends to 0 . Note that the assumption of $L>3$, which arises naturally from this paper's regime, along with the fact that $\|\mathbf{a}\| \geq 2$, guarantees that $E_{1}(L)$ is positive. Figure 2 depicts the probability in (13), it's upper bound given in (14) and simulation results. From the analytic results as well as the simulations on the rate of decay, one can deduce that even for relatively small values of simultaneously transmitting users ( $L>20$ ), a relay will prefer to choose a unit vector. Also, one can observe from the results and from the analytic bound that as the norm of a grows, the rate of decay increases. This faster decay reflects the increased penalty of approximating a real vector using an integer-valued vector.

Note that Corollary 1 refers to the probability that a non-trivial fixed a will be the maximizer of $f$. Next, we wish to explore this probability for any possible a that is not a unit vector, yet satisfies $\|\mathbf{a}\|^{2}<1+\mathrm{P}\|\mathbf{h}\|^{2}$. Define by $P_{r}(\mathbf{e})$ the probability that a relay picked any unit vector as the coefficient vector, and by $P_{r}(\overline{\mathbf{e}})$ the probability that a non-trivial vector was chosen.

Theorem 4. Under the CF scheme, the probability that any other non-trivial coefficient vector a will be chosen to maximize the achievable rate $\mathcal{R}(\mathbf{h}, \mathbf{a})$ compared with any unit vector $\mathbf{e}_{i}$, as the number of simultaneously transmitting users grows, is zero. That is,

$$
\begin{equation*}
\lim _{L \rightarrow \infty} P_{r}(\overline{\mathbf{e}})=0 \tag{15}
\end{equation*}
$$

The main consequences of Theorem 4 are the following results, that apply for the general model with $M$ relays. The results describe the behavior of the achievable rate of any relay and the system's sum-rate at the limit of a large number of simultaneously transmitting users.

Theorem 5. As L grows, the achievable rate of a relay converges to zero in probability, that is,

$$
\begin{equation*}
\lim _{L \rightarrow \infty} P_{r}(\mathcal{R}(\mathbf{h}, \mathbf{a})>\epsilon)=0 \tag{16}
\end{equation*}
$$

for all $\epsilon>0$.


Figure 2: The upper bounds on the probability of not having a unit vector as the minimizer of $f$ ,given in (14) (solid lines) and (13) (dashed lines) compared to simulation results (dotted lines) for various values of $\|\mathbf{a}\|^{2}$. The $x$ axis is the number of simultaneously transmitting users.

Corollary 2. As L grows, the sum-rate of CF for a general system with $M$ relays converges to zero in probability, that is,

$$
\begin{equation*}
\lim _{L \rightarrow \infty} P_{r}\left(\frac{R L}{N}>\epsilon\right)=0 \tag{17}
\end{equation*}
$$

for all $\epsilon>0$, where $R<\min _{n=1, \ldots, N} \min _{m} \mathcal{R}\left(\mathbf{h}_{m}(n), \mathbf{a}_{m}(n)\right)$.
Theorem 5 and corollary 2 show that the sum-rate of all decoded linear combinations tends to zero as the number of simultaneously transmitting users grows. Thus, applying the CF coding scheme in its vanilla settings, where all transmitters transmit and the relays decode linear combinations of all transmitted messages, is futile. A possible course of action, while still applying CF , is to restrict the number of simultaneously transmitting users in each transmission slot. In the sequel, we show that by performing this restriction, not only a strictly positive sum-rate is achievable, smart scheduling can provide an overall gain to the system's sum-rate.

## B. Scheduling in $C F$

In this part of the work, we will suggest a scheduling policy, and prove its asymptotic optimality by meeting a global upper bound on the system's sum-rate. The scheduling policy is based on setting the number of scheduled users in each slot properly, as well as identifying the specific subset and a specific set of coefficient vectors from which the relay chooses the linear combination for the scheduled users. We note that for the case of a general system, with $M>1$
relays, our suggested scheduling policy should be further adapted to the restrictions which arise from the presence of multiple relays; we discuss this in Section VI.

We first present the scheduling problem under the model of $L$ users and a single relay. We first assume that there is a scheduler (e.g., the relay itself) that handles the scheduling process. In Section V-F we discuss the implementation of the scheduling process in a distributed manner, i.e., without having a centralized entity. We assume that in each slot $i$ a subset of $k$ users is chosen by the scheduler. This subset is denoted by $\mathcal{S}_{k}(i)$. The value of $k$ is fixed and will be dealt with in the sequel. Accordingly, the total number of subsets is $\binom{L}{k}$, each having a channel vector which we denote by $\mathbf{h}\left(\mathcal{S}_{k}(i)\right)$, and a corresponding coefficient vector a $\left(\mathcal{S}_{k}(i)\right)$. In addition, we denote by $\mathbf{h}_{L}(i)$ the channel vector of all users, to create a distinction with the channel vector of the scheduled users. As mentioned above, the scheduling problem consists of two highly connected optimization problems. The first can be viewed as finding the proper subset of users $\mathcal{S}_{k}(i)$, and the second is finding the proper $\mathbf{a}\left(\mathcal{S}_{k}(i)\right)$ for the $\mathbf{h}\left(\mathcal{S}_{k}(i)\right)$ of the selected users, to maximize the system's sum-rate. Let us denote by $\mathcal{S}_{k}^{N}=\left(\mathcal{S}_{k}(1), \mathcal{S}_{k}(2), \ldots, \mathcal{S}_{k}(N)\right)$ the sequence of all subsets of users which where scheduled until complete decoding. In addition, denote by $\mathcal{A}_{k}^{N}=\left(\mathbf{a}\left(\mathcal{S}_{k}(1)\right), \mathbf{a}\left(\mathcal{S}_{k}(2)\right), \ldots, \mathbf{a}\left(\mathcal{S}_{k}(N)\right)\right)$ the coefficients matrix with the coefficient vectors in all slots. Note that since $\mathcal{A}_{k}^{N} \in \mathbb{Z}^{N \times k}$, yet decoding is preformed over dimension $L$, we must map $\mathcal{A}_{k}^{N}$ to $\mathbf{A}_{N}$ by setting each row of $\mathbf{A}_{N}$ to have the entries of $\mathbf{a}\left(\mathcal{S}_{k}(i)\right)$ for the scheduled users and 0 otherwise. We denote this map as $(\cdot)_{\Uparrow}^{L}: \mathbf{a}\left(\mathcal{S}_{k}(i)\right) \mapsto \mathbf{a}(i)$. For example, let $\mathcal{S}_{3}(1)=\{1,2,5\}$ be the 3 users that where scheduled, with coefficient vector $\mathbf{a}\left(\mathcal{S}_{3}(1)\right)=(1,1,2)$ in the first slot. Then, for $L=6$, we have $\left(\mathbf{a}\left(\mathcal{S}_{3}(1)\right)\right)_{\Uparrow}^{6}=\mathbf{a}(1)=(1,1,0,0,2,0)$. Furthermore, it will be useful to define the set of vectors of length $L$, which have $k$ non-zero values taken from a certain set $\Omega$ as $\mathbb{S}_{L, k}^{\Omega}$. For example, for $\Omega=\{1\}$, the set $\mathbb{S}_{L, 2}^{\{1\}}$ is all binary vectors of length $L$ with exactly two ones.

We thus modify the maximization in (10) to include our scheduling problem,

$$
\begin{equation*}
\max _{\mathcal{S}_{k}^{N}, \mathcal{A}_{k}^{N} \in \mathbb{Z}^{N \times k}} \frac{R\left(\mathcal{S}_{k}^{N}, \mathcal{A}_{k}^{N}\right) L}{N\left(\left(\mathcal{A}_{k}^{N}\right)_{\Uparrow}^{L}\right)} . \tag{18}
\end{equation*}
$$

In this work, we provide a polynomial-time (in both $k$ and $L$ ) scheduling algorithm, described in Algorithm 1, Section V. Algorithm 1 finds the asymptotically (with $L$ ) optimal schedule for the maximization problem in (18) for all transmission slots as well as the coefficient vectors which ensure complete decoding. Its asymptotic guarantees are bellow.

Theorem 6. The asymptotic expected achievable rate Algorithm 1 achieves in each slot, for $L \geq 4$, is lower bounded by the following,

$$
\mathbb{E}\left[\mathcal{R}_{a c h}^{s c h}\right] \geq \frac{1}{2} \log ^{+}\left(k\left(1-\frac{P k u^{4}}{(u+\delta)^{2}\left(1+P k u^{2}\right)}(1-o(1))\right)\right)^{-1},
$$

where $u=\sqrt{2 \ln \frac{\delta \sqrt{L}}{\sqrt{2 \pi}}}-\delta, \delta=\frac{1}{\ln L}$ and $o(1) \rightarrow 0$ as $L \rightarrow \infty$. Thus, the expected achievable rate for Algorithm 1 in each slot scales at least as $O\left(\frac{1}{4} \log \log L\right)$.

The values for $u$ and $\delta$ relate to the characteristics of the scheduled users' channel, and determine the search domain for the optimal schedule, as will be explained in the sequel. They are chosen to provide asymptotically optimal result. The suggested scheduling scheme, however, is independent of $u$ and $\delta$ and works well even for moderate number of transmitters, as Figure 8 depicts. In addition, although it is left as a parameter, the number of scheduled users $k$ in each slot, under our scheduling paradigm, should be $O(\log L)$; we refer to this choice in the sequel as well. We thus have the following corollary.

Corollary 3. The expected system's sum-rate that Algorithm 1 can achieve for a single relay system is lower bounded by,

$$
\begin{equation*}
\mathbb{E}\left[C_{S R}\right] \geq\left(\frac{1}{4}-\epsilon\right) \log \log L(1-o(1)) \tag{19}
\end{equation*}
$$

where $\epsilon$ is a small positive constant and $o(1) \rightarrow 0$ as $L \rightarrow \infty$.

Theorem 6 and corollary 3 indicate that indeed, as the number of users grows, the system's sum-rate grows as well, making scheduling not only mandatory but worthwhile.

In order to have a comparison with the best scheduling policy one can attain, we present an upper bound on the expected performance of any scheduling algorithm and its scaling law, at the limit of large $L$.

Theorem 7. The expected system's sum-rate of any scheduling algorithm designed for a CF system with a single relay, is upper bounded by the following,

$$
\mathbb{E}\left[C_{S R}\right] \leq \frac{1}{2} \log \left(1+P\left(2 \ln L-\ln \ln L-2 \ln \Gamma\left(\frac{1}{2}\right)+\frac{\gamma}{2}+o(1)\right)\right)
$$

where $\gamma$ is the Euler-Mascheroni constant. Thus, the expected sum-rate of any scheduling algorithm scales at most as $\frac{1}{2} \log \log L$.

Theorem 7 and corollary 3 show that Algorithm 1 is asymptotically optimal, and we have

Corollary 4. Algorithm 1 attains the optimal scaling law of the expected sum-rate, which is $O(\log \log L)$.

The above results suggest that the scaling law of the CF sum-rate is similar to the scaling law which multiuser diversity schemes achieve [20], [30],[21]. In a way, this can be expected since in schemes exploiting multiuser diversity, the scheduler seeks a user (or users) which has the highest channel gain in each transmission slot, where, as we will see in the sequel, our scheduling paradigm also searches for a group of users which have high channel gains. However, these users are not necessarily the ones with the highest gain among all users. Moreover, we point out that, in terms of fairness and delay considerations, our suggested scheduling algorithm for CF promises, with very high probability, the shortest completion time for decoding all users' messages by letting more than one user to use the channel simultaneously. This is with contrast with basic multiuser diversity schemes which may results with longer periods of completion time. We elaborate on this is Section V-D.

Finally, when extending the above results to multiple relays, a major challenge is finding a subset of transmitters that will allow all relays to successfully decode non-trivial linear combinations. In Section VI, we prove that indeed good schedules exist, satisfying all relays simultaneously, and give heuristic schedules to allow for fast completion time with multiple relays.

## IV. CF With a Large Number of Users

In this section, we analyze the system's behavior, under the CF scheme, for the regime of a large number of users and a single relay. This analysis will be the foundation for the theorems given in Section $I I I$ and the motivation for the scheduling given in section $\square \mathrm{We}$ assume that the relay seeks a coefficient vector which maximizes the achievable rate (minimize $f$ as described in (12)) in a given slot. Though this may not be the optimal choice that maximizes (10), it gives an upper bound on the achievable rate of the relay in that slot.

## A. Minimization of the quadratic form $f$

Examining the matrix $G$ in the minimization problem (12), one can notice that as $L$, the number of transmitters, grows, the positive diagonal elements grow very fast relative to the offdiagonal elements. Specifically, each diagonal element is a $\chi_{L}^{2}$ r.v. minus a $\chi_{1}^{2}$ r.v., whereas the


Figure 3: An example for the magnitude of the elements of $G$ for different dimensions (i.e., different values of $L$ ), for $\mathrm{P}=10$. The graphs depict a single realization for each $L$, and were interpolated for ease of visualization.
off-diagonal elements are only a multiplication of two Gaussian r.vs. As $L$ grows, the former has much higher expected value compared to the later and in the limit of large $L$, the matrix $\mathbf{G}$ will tend to a diagonal matrix if properly normalized. Examples of a specific G, for a certain realization of the channel vector $\mathbf{h}$, are presented in Figure 3, for different dimensions.

Consider now the quadric form (12) we wish to minimize. Roughly speaking, any choice of a that is not a unit vector will add more than one (probably) positive element from the diagonal of $\mathbf{G}$ to it together with possibly negative off-diagonal elements. However, when $L$ is large, the off-diagonal elements have lesser effect compared to the diagonal ones. Therefore, intuitively, one would prefer to choose a to be a unit vector. In the remainder of this section, we make this argument formal.

The minimization function $f(\mathbf{a})=\mathbf{a}^{T} \mathbf{G a}$ can be written as

$$
\begin{aligned}
\mathbf{a}^{\mathbf{T}} \mathbf{G} \mathbf{a} & =\sum_{i=1}^{L}\left(1-\mathbf{P}\left(\|\mathbf{h}\|^{2}-h_{i}^{2}\right)\right) a_{i}^{2}-2 \sum_{i=1}^{L} \sum_{j=1}^{i-1} \mathrm{P} h_{i} h_{j} a_{i} a_{j} \\
& =\|\mathbf{a}\|^{2}+\mathbf{P} \sum_{i=1}^{L} \sum_{j=1}^{i-1}\left(h_{i} a_{j}-h_{j} a_{i}\right)^{2} \\
& =\|\mathbf{a}\|^{2}+\mathbf{P}\left(\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}-\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}\right) .
\end{aligned}
$$

We wish to understand when will a relay prefer a unit vector over any other non-trivial vector a. Specifically, since a is a function of the random channel $h$, we will compute the probability that a certain non-trivial a will minimize $f$ compared to a unit vector. In particular, the best unit vector among all other unit vectors. We thus wish to find the probability

$$
\begin{equation*}
P_{r}\left(f(\mathbf{a}) \leq \min _{i} f\left(\mathbf{e}_{i}\right)\right)=P_{r}\left(\|\mathbf{a}\|^{2}+\mathbf{P}\left(\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}-\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}\right) \leq \min _{i}\left\{1+\mathbf{P}\left(\|\mathbf{h}\|^{2}-h_{i}^{2}\right)\right\}\right), \tag{20}
\end{equation*}
$$

where $\mathbf{e}_{i}$ is any unit vector of size $L$ with 1 at the $i$-th entry and zero elsewhere, and a is any integer valued vector that is not a unit vector. Note that the right and left-hand sides of the inequality in (20) are dependent, hence direct computation of this probability is not trivial. Still, this probability can be evaluated exactly noting that the angle between a and $\mathbf{h}$ is what mainly affects it. Formally, we give the following Lemma.

Lemma 1. The distribution of $\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}}$, which is the squared cosine of the angle between an integer vector $\mathbf{a}$ and a standard normal vector $\mathbf{h}$, both of dimension L, is Beta $\left(\frac{1}{2}, \frac{L-1}{2}\right)$.

Proof. Let $\mathbf{Q}$ be an orthogonal rotation matrix such that $\mathbf{Q a}=\mathbf{a}^{\prime}$, where $\mathbf{a}^{\prime}$ is co-linear with the basis vector $e_{1}$. That is, $\mathbf{a}^{\prime}=(\|\mathbf{a}\|, 0, \ldots, 0)$. Define $\mathbf{h}^{\prime}=\mathbf{Q h}$. Note that $\mathbf{h}^{\prime}$ is a standard normal vector since $E\left[\mathbf{h}^{\prime}\right]=E[\mathbf{Q h}]=0$, and $\mathbf{Q I Q}^{T}=\mathbf{Q Q}^{T}=\mathbf{I}$. We have

$$
\begin{align*}
\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}} & =\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\left(\mathbf{a}^{T} \mathbf{a}\right)\left(\mathbf{h}^{T} \mathbf{h}\right)}=\frac{\left(\mathbf{a}^{T} \mathbf{Q}^{T} \mathbf{Q h}\right)^{2}}{\left(\mathbf{a}^{T} \mathbf{Q}^{T} \mathbf{Q a}\right)\left(\mathbf{h}^{T} \mathbf{Q}^{T} \mathbf{Q h}\right)} \\
& =\frac{\left((\mathbf{Q a})^{T} \mathbf{Q h}\right)^{2}}{\left((\mathbf{Q a})^{T} \mathbf{Q a}\right)\left((\mathbf{Q h})^{T} \mathbf{Q h}\right)}=\frac{\left((\mathbf{Q a})^{T} \mathbf{Q h}\right)^{2}}{\|\mathbf{Q a}\|^{2}\|\mathbf{Q h}\|^{2}}  \tag{21}\\
& =\frac{\left\|\mathbf{a}^{\prime}\right\|^{2}\left(\mathbf{e}_{\mathbf{1}}{ }^{T} \mathbf{h}^{\prime}\right)^{2}}{\left\|\mathbf{a}^{\prime}\right\|^{2}\left\|\mathbf{h}^{\prime}\right\|^{2}}=\frac{\left(\mathbf{e}^{T} \mathbf{h}^{\prime}\right)^{2}}{\left\|\mathbf{h}^{\prime}\right\|^{2}}=\frac{h_{1}^{\prime 2}}{\left\|\mathbf{h}^{\prime}\right\|^{2}} .
\end{align*}
$$

Therefore,

$$
\begin{equation*}
\cos ^{2} \theta=\frac{h_{1}^{\prime 2}}{h_{1}^{\prime 2}+h_{2}^{2}+\ldots+h_{L}^{\prime 2}} . \tag{22}
\end{equation*}
$$

This expression can be represented as $\frac{W}{W+V}$, where $W=h_{1}^{\prime 2}$ is a $\chi_{1}^{2}$ r.v. and $V=\sum_{i=2}^{L} h_{i}^{\prime 2}$ is a $\chi_{L-1}^{2}$ r.v. independent of $W$. This ratio has a $\operatorname{Beta}(a, b)$ distribution, with $a=\frac{1}{2}$ and $b=\frac{L-1}{2}$ [31]. Note that $a$ and $b$ correspond to the degrees of freedom of $W$ and $V$.

We can now give the proof for Theorem 3 .
Proof of Theorem 3 According to equation (20), we have,

$$
\begin{aligned}
P_{r}\left(f(\mathbf{a}) \leq \min _{i} f\left(\mathbf{e}_{i}\right)\right) & =P_{r}\left(\|\mathbf{a}\|^{2}+\mathbf{P}\left(\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}-\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}\right) \leq \min _{i}\left\{1+\mathbf{P}\left(\|\mathbf{h}\|^{2}-h_{i}^{2}\right)\right\}\right) \\
& =P_{r}\left(\frac{1-\|\mathbf{a}\|^{2}}{\mathbf{P}}+\|\mathbf{h}\|^{2}-\max _{i} h_{i}^{2}-\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}+\left(\mathbf{a}^{T} \mathbf{h}\right)^{2} \geq 0\right) \\
& \stackrel{(a)}{\leq} P_{r}\left(\|\mathbf{h}\|^{2}-\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}+\left(\mathbf{a}^{T} \mathbf{h}\right)^{2} \geq 0\right) \\
& =P_{r}\left(\frac{1}{\|\mathbf{a}\|^{2}}-1+\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}} \geq 0\right) \\
& =P_{r}\left(\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}} \geq 1-\frac{1}{\|\mathbf{a}\|^{2}}\right) \\
& \stackrel{(b)}{=} 1-I_{\Phi(\mathbf{a})}\left(\frac{1}{2}, \frac{L-1}{2}\right),
\end{aligned}
$$

where (a) follows from removing the negative terms $\left(\|\mathbf{a}\|^{2}>1\right)$ and $(b)$ follows from Lemma 1 with $\Phi(\mathbf{a})=1-\frac{1}{\|\mathbf{a}\|^{2}}$.

The bound on the probability given in Theorem 3 consists of a complicated analytic function $I_{\Phi(\mathbf{a})}(\cdot)$. Hence, Corollary 1 includes a simplified bound which avoids the use of $I_{\Phi(\mathbf{a})}(\cdot)$, yet keeps the nature of the result in Theorem 3. The proof of Corollary 1 is based on the following lemma.

Lemma 2. The CDF of $\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}}$ can be lower bounded by the CDF of the minimum of $\left(\left\lfloor\frac{L}{2}\right\rfloor-1\right)$ i.i.d. uniform random variables in $[0,1]$.

Proof. We start by assuming that $L$ is even. The case of odd $L$ will be dealt with later. From Lemma 1, the r.v. $\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}}$ has the same distribution as $\frac{h_{1}^{2}}{\|\mathbf{h}\|^{2}}$, that is, for any $0 \leq \alpha \leq 1$,

$$
\begin{aligned}
P_{r}\left(\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}} \leq \alpha\right) & =P_{r}\left(\frac{h_{1}^{2}}{\|\mathbf{h}\|^{2}} \leq \alpha\right) \\
& \geq P_{r}\left(\frac{h_{1}^{2}+h_{2}^{2}}{\|\mathbf{h}\|^{2}} \leq \alpha\right) \\
& =P_{r}\left(\frac{h_{1}^{2}+h_{2}^{2}}{\left(h_{1}^{2}+h_{2}^{2}\right)+\ldots+\left(h_{L-1}^{2}+h_{L}^{2}\right)} \leq \alpha\right) \\
& =1-(1-\alpha)^{\frac{L}{2}-1},
\end{aligned}
$$

where the last line is due to the observation that $\frac{h_{1}^{2}+h_{2}^{2}}{\|\mathbf{h}\|^{2}}$ can be represented as $\frac{W}{W+V}$, where $W=h_{1}^{2}+h_{2}^{2}$ and $V=\sum_{i=3}^{L} h_{i}^{2}$ are independent exponential r.vs. Note that $V$ is essentially a sum of $\frac{L}{2}-1$ independent pairs. This ratio is distributed as the minimum of $\left(\frac{L}{2}-1\right)$ i.i.d. uniform $[0,1]$ random variables [32, Lemma 3.2], [33]. This is since the ratio can be interpreted as the proportion of the waiting time from the first arrival to the $\frac{L}{2}$ arrival of a Poisson process.

In case $L$ is an odd number, we can increase the term in the proof by replacing it with $\frac{h_{1}^{2}+h_{2}^{2}}{\|\mathbf{h}\|^{2}-h_{L}^{2}}$, resulting with a ratio which is distributed as the minimum of $\left(\frac{L-1}{2}-1\right)$ i.i.d. uniform random variables in the same manner.

Using Lemma 2 we now give the proof for Corollarly 1.

Proof of Corollary 1 I

$$
\begin{align*}
P_{r}\left(f(\mathbf{a}) \leq \min _{i} f\left(\mathbf{e}_{i}\right)\right) & \stackrel{(a)}{\leq} P_{r}\left(\frac{\left(\mathbf{a}^{T} \mathbf{h}\right)^{2}}{\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}} \geq 1-\frac{1}{\|\mathbf{a}\|^{2}}\right) \\
& \stackrel{(b)}{\leq}\left(1-\left(1-\frac{1}{\|\mathbf{a}\|^{2}}\right)\right)^{\left\lfloor\frac{L}{2}\right\rfloor-1}  \tag{23}\\
& \leq\left(\frac{1}{\|\mathbf{a}\|^{2}}\right)^{\frac{L-1}{2}-1} \\
& =e^{-L E_{1}(L)},
\end{align*}
$$

where $(a)$ and $(b)$ follow from Lemmas 1 and 2, respectively, $E_{1}(L)=\left(1-\frac{3}{L}\right) \log \|\mathbf{a}\|$ and $\|\mathbf{a}\|^{2}>1$.

As mentioned, Corollary 1 refers to the probability that a non-trivial fixed a will be the maximizer of $f$. Thus, to consider all possible coefficient vector and to analyze $P_{r}(\overline{\mathbf{e}})$, the probability which any other vector that is not a unit vector was chosen, at the limit of large $L$, one needs to consider the set of all a vectors in the search domain.

In [24], a polynomial-time algorithm for finding the optimal coefficient vector a was given. The complexity result derives from the fact that the cardinality of the set of all a vectors which are possible is upper bounded by $2 L\left(\left\lceil\sqrt{1+\mathbf{P}\|\mathbf{h}\|^{2}}\right\rceil+1\right)$. That is, any vector which is not in this set has zero probability to be the one that maximizes the rate. Note that all unit vectors are included in this set. We define a new set without the unit vectors by $\mathcal{A}_{\mathbf{h}}=\left\{\mathbf{a} \in \mathbb{Z}^{L}\right.$ : $\mathbf{a}$ is possible, $\left.\mathbf{a} \neq \mathbf{e}_{i} \forall i\right\}$. Thus, we wish to compute

$$
\begin{equation*}
P_{r}(\overline{\mathbf{e}})=P_{r}\left(\bigcup_{\mathbf{a} \in \mathcal{A}_{\mathbf{h}}}\left\{f(\mathbf{a}) \leq f\left(\mathbf{e}_{i}\right) \text { for all } i\right\}\right) . \tag{24}
\end{equation*}
$$

Note that the probability that another unit vector $\mathbf{e}_{j}$ will have a better metric than $\mathbf{e}_{i}$ is not negligible. In fact, as $L$ grows, the unit vector with the best metric is the one corresponding to the strongest user. Note also that the cardinality of $\mathcal{A}_{\mathbf{h}}$ grows with the dimension of $\mathbf{h}$, i.e., with $L$, and can be easily upper bounded as follows,

$$
\begin{equation*}
\left|\mathcal{A}_{\mathbf{h}}\right| \leq 2 L\left(\left[\sqrt{1+\mathrm{P}\|\mathbf{h}\|^{2}}\right\rceil+1\right)-L \leq 2 L\left(\mathrm{P}\|\mathbf{h}\|^{2}+2.5\right) \tag{25}
\end{equation*}
$$

With the above definition and the upper bound on the cardinality of the search domain, we can now give the proof for Theorem 4

Proof of Theorem 4

$$
\begin{aligned}
\lim _{L \rightarrow \infty} P_{r}(\overline{\mathbf{e}}) & =\lim _{L \rightarrow \infty} P_{r}\left(\bigcup_{\mathbf{a} \in \mathcal{A}_{\mathbf{h}}}\left\{f(\mathbf{a}) \leq f\left(\mathbf{e}_{i}\right) \text { for all } i\right\}\right) \\
& \leq \lim _{L \rightarrow \infty} \sum_{\mathcal{A}_{\mathbf{h}}} P_{r}\left(f(\mathbf{a}) \leq f\left(\mathbf{e}_{i}\right) \text { for all } i\right) \\
& =\lim _{L \rightarrow \infty} \sum_{\mathcal{A}_{\mathbf{h}}} P_{r}\left(f(\mathbf{a}) \leq \min _{i} f\left(\mathbf{e}_{i}\right)\right) \\
& \stackrel{(a)}{\leq} \lim _{L \rightarrow \infty} \sum_{\mathcal{A}_{\mathbf{h}}}\left(\frac{1}{\|\mathbf{a}\|^{2}}\right)^{\frac{L-1}{2}-1} \\
& \stackrel{(b)}{\leq} \lim _{L \rightarrow \infty}\left|\mathcal{A}_{\mathbf{h}}\right|\left(\frac{1}{2}\right)^{\frac{L-1}{2}-1} \\
& \stackrel{(c)}{\leq} \lim _{L \rightarrow \infty} 2 L\left(\mathrm{P}\|\mathbf{h}\|^{2}+2.5\right)\left(\frac{1}{2}\right)^{\frac{L-1}{2}-1} \\
& =\lim _{L \rightarrow \infty} 2 L\left(\mathrm{P} \sum_{i=1}^{L} h_{i}^{2}+2.5\right)\left(\frac{1}{2}\right)^{\frac{L-1}{2}-1} \\
& =\lim _{L \rightarrow \infty} 2 L^{2} \mathrm{P}\left(\frac{1}{2}\right)^{\frac{L-1}{2}-1} \frac{1}{L} \sum_{i=1}^{L} h_{i}^{2} \\
& \stackrel{(d)}{=} \lim _{L \rightarrow \infty} 2 L^{2} \mathrm{P}\left(\frac{1}{2}\right)^{\frac{L-1}{2}-1} \\
& =4 \mathrm{P} \lim _{L \rightarrow \infty} L^{2} 2^{-\frac{L-1}{2}} \\
& =4 \mathrm{P} \lim _{L \rightarrow \infty} L^{2} e^{-L E_{2}(L)}=0
\end{aligned}
$$

where (a) follows from Corollary 1 and $(b)$ is true since a is an integer vector which is not the unit vector. (c) is due to (25). ( $d$ ) follows from the strong law of large numbers - the normalized sum converges with probability one to the expected value of $\chi_{1}^{2}$ r.v. which is one. Lastly we define $E_{2}(L)=\frac{1}{2}\left(1-\frac{1}{L}\right) \log 2$.

This result implies that the probability of having any non-unit vector as the rate maximizer is decreasing exponentially to zero as the number of users grows. The direct consequence is that the relay will try to decode a single message when all other transmitters are treated as interference. Theorem 5 shows that in this case the rate of that single message has to go to zero as the number of transmitters grows. The proof is presented in Appendix A. Moreover, Corollary 2 which relates to a general system consisting of $M$ relays, shows that as the number of users


Figure 4: Simulation results for the achievable rate at a relay, as given in Theorem 2. The rate was plotted as a function of the number of simultaneously transmitting users $L$, for different values of P . In each sample, the relay chose the coefficient vector $\mathbf{a}^{\max }$.
grows, the system's sum-rate decreases to zero as well. That is, without scheduling users, not only each individual rate is negligible; this is true for the sum-rate of the entire system as well. The proof is presented in Appendix B.

Simulation results for the achievable rate at a relay, as a function of $L$ and for different values of $P$, are depicted in Figure 4. Here, in each sample, the relay chose the rate maximizing coefficient vector $\mathbf{a}^{\max }$. One can observe that for large $L$, the achievable rate decreases to zero as Theorem 5 suggests. Accordingly, following corollary 2, Figure 5 depicts simulation results for the behavior of the sum-rate, for a system with $M$ relays, as a function of $L$. In fact, the curves constitute an upper bound since in the simulation we considered the achievable rates in each slot rather than the minimum rate among all slots. In addition, for decoding, we considered the $L$ linearly independent coefficient vectors (the rows of $\mathbf{A}$ ) with the highest rates. As the plot shows, similar to the achievable rate at each relay, the sum-rate goes to zero as well. On the other hand, Figure 6 presents simulation results where $k=3$ users were scheduled for transmission in each slot in a uniform manner. It is clear that even a simple scheduling policy can guarantee a non-zero rate. Note that the decrease in the sum-rate is a result of the constant number of users who were scheduled regardless of $L$. As we will see in the sequel, $k$ should also grow with $L$, or else the completion time will grow and thus the sum-rate will decreases as the figure depicts.

The results of this section show that the increase in the number of users can dramatically degrade the performance of a system that applies the CF coding scheme. This is in contrast to the common knowledge for the capacity of MAC which is increasing with the number of simultaneously transmitting users. The decrease in the system performance results from the


Figure 5: Simulation results for the sum-rate of a system with $M=2,6$ relays, for different values of P , as a function of $L$. The sum-rate decreases since each relay prefers to decode a single massage and treat all other massages as noise. Here, all the users are transmitting.


Figure 6: Simulation results for the sum-rate with a random scheduling policy. In each slot $k=3$ users were scheduled for transmission while each relay chose the rate maximizing coefficient vector $\mathbf{a}^{\text {max }}$.
approximation error of the large real channel vector by the vector of integer coefficients of the linear combination of the messages.

## V. Scheduling in CF

The scheduling problem presented in (18) requires maximizing the transmission rates and minimizing the number of slots, $N$, simultaneously. This should be done via jointly selecting users and the appropriate coefficient vectors. In the following subsections, we show that asymptotically with $L$, a computationally efficient solution for this problem exists. Specifically, we will show that a policy that maximizes the achievable rate in each slot will also minimize $N$ to its (asymptotic) minimum value as a direct outcome, as long as the number of scheduled users, $k$, in each slot, is $O(\log L)$. Essentially, this will mean that the scheduling problem can be reduced to


Figure 7: The achievable rate as a function of $\theta$, i.e., the angle between $\mathbf{a}$ and $\mathbf{h}$, and the squared norm of $\mathbf{a}$. The discrete points are simulation results for the rate of each subset of users of size $k=3$ for a specific realization of the channel vector $\mathbf{h}_{L}$ with $(a) L=15,(b) L=45$, and $\mathrm{P}=1000$. The transparent curved plane describes an upper bound on the rate as in (27), where, for ease of visualization, a continues function was plotted.
finding the subset of $k=O(\log L)$ users $\mathcal{S}_{k}^{*}(i)$ which yields the highest $\mathcal{R}\left(\mathbf{h}\left(\mathcal{S}_{k}^{*}(i)\right), \mathbf{a}\left(\mathcal{S}_{k}^{*}(i)\right)\right)$ in each slot $i$ separately. That is,

$$
\begin{equation*}
\mathcal{S}_{k}^{*}=\underset{\mathcal{S}_{k}}{\arg \max }\left\{\max _{\mathbf{a}\left(\mathcal{S}_{k}\right) \in \mathbb{Z}^{k} \backslash\{\mathbf{0}\}} \mathcal{R}\left(\mathbf{h}\left(\mathcal{S}_{k}\right), \mathbf{a}\left(\mathcal{S}_{k}\right)\right)\right\} . \tag{26}
\end{equation*}
$$

Thus, we start the analysis by examining the achievable rate as a function of $\mathbf{h}$ and $\mathbf{a}$ and discuss the properties of the optimal schedule. Based on this analysis, we then provide a scheduling algorithm for the complete transmission of all messages, with its complexity and asymptotic guarantees. Note that since the slots are assumed to be memoryless, we omit the time index in $\mathcal{R}(\mathbf{h}, \mathbf{a})$. Furthermore, occasionally, we also omit the users' subset index when it is obvious from the context and write $h$ and a as a general writing for the sake of notation simplicity.

## A. Achievable Rate Under Scheduling

The main challenge in low complexity scheduling for CF is identifying the characteristics of the channel values of a certain subset of users such that this subset would be classified as a good option for CF. Remembering the rate expression in (5), the challenge is to identify a subset of the channel values that can be well approximated by an integer vector.

We start by exploring an upper bound on the achievable rate. This bound motivates the suggested algorithm, which will be given in the sequel. We have

$$
\begin{align*}
\mathcal{R}(\mathbf{h}, \mathbf{a}) & =\frac{1}{2} \log ^{+}\left(\|\mathbf{a}\|^{2}-\frac{\mathrm{P}\left(\mathbf{h}^{T} \mathbf{a}\right)^{2}}{1+\mathrm{P}\|\mathbf{h}\|^{2}}\right)^{-1} \\
& \leq \frac{1}{2} \log ^{+}\left(\|\mathbf{a}\|^{2}-\frac{\left(\mathbf{h}^{T} \mathbf{a}\right)^{2}}{\|\mathbf{h}\|^{2}}\right)^{-1} \\
& =\frac{1}{2} \log ^{+}\left(\|\mathbf{a}\|^{2}-\|\mathbf{a}\|^{2} \cos ^{2}(\theta)\right)^{-1} \\
& =\frac{1}{2} \log ^{+}\left(\|\mathbf{a}\|^{2} \sin ^{2}(\theta)\right)^{-1}, \tag{27}
\end{align*}
$$

where $\theta$ is the angle between $\mathbf{h}$ and the chosen coefficient vector a.
The behavior of the achievable rate as a function of $\theta$ and $\|\mathbf{a}\|^{2}$ is depicted in Figure 7. The discrete lines represent simulation results for $\mathcal{R}\left(\mathbf{h}\left(\mathcal{S}_{k}\right), \mathbf{a}\left(\mathcal{S}_{k}\right)\right)$ for each subset of size $k=3$, out of a realization of the channel vector $\mathbf{h}_{L}$. That is, for each subset $\mathcal{S}_{k}$, the optimal a $\left(\mathcal{S}_{k}\right)$ was chosen according to (11). $\mathrm{P}=1000$. The continuous curve is a smoothed representation of (27). The smoothing is since, for one, a is an integer vector, hence, its squared norm takes only integer values. Second, for a certain $\|\mathbf{a}\|^{2}$, there are only finitely many possible choices of $\mathbf{a}$ and thus a finite number of angles with $\mathbf{h}$. For example, for $\|\mathbf{a}\|^{2}=5$ and dimension 2 the possible vectors are only $(1,2),(-1,2),(1,-2)$ and $(-1,-2)$. That is, in this case, there are 4 possible angles with a given $\mathbf{h}$.

One can infer from the upper bound (27) and Figure 7 that the subsets of users that attain the highest rates are those with low values of $\|\mathbf{a}\|^{2}$ and $\theta$. Moreover, we note that the slope of the rate as a function of $\theta$ is sharper than the slope as a function of $\|\mathbf{a}\|^{2}$, with an exception for the smallest values of $\|\mathbf{a}\|^{2}$. Accordingly, for any given $\mathbf{h}_{L}$, we expect to use only a fixed set of coefficient vectors with low norm values as the set the relay will choose from. Then, try to find a subset of transmitters $\mathcal{S}_{k}$ with the smallest angle between $\mathbf{h}\left(\mathcal{S}_{k}\right)$ and one vector out of this set.

Considering the above, define the set of coefficient vectors:

$$
\begin{equation*}
\mathbf{a}^{\{\mathbf{1}, \mathbf{k}\}}=\left\{\mathbf{a}\left|\mathbf{a} \in \mathbb{Z}^{k},\left|a_{i}\right|=1, i=1, \ldots, k\right\} .\right. \tag{28}
\end{equation*}
$$

Note that $\forall \mathbf{a} \in \mathbf{a}^{\{\mathbf{1}, \mathbf{k}\}},\|\mathbf{a}\|^{2}=k$, and that the cardinality of $\mathbf{a}^{\{\mathbf{1}, \mathbf{k}\}}$ is $2^{k}$. In what follows, we suggest that the scheduled subset of users would be a subset for which the relay will be able
to choose the coefficient vector out of $\mathbf{a}^{\{1, \mathbf{k}\}}$ as its rate maximizer. Thus, Equation (26) can be written as,

$$
\begin{equation*}
\mathcal{S}_{k}^{\mathbf{a}^{\{1, \mathbf{k}\}}}=\underset{\mathcal{S}_{k}}{\arg \max }\left\{\max _{\mathbf{a}\left(\mathcal{S}_{k}\right) \in \mathbf{a}\{1, \mathbf{k}\}}\left\{\mathcal{R}\left(\mathbf{h}\left(\mathcal{S}_{k}\right), \mathbf{a}\left(\mathcal{S}_{k}\right)\right)\right\}\right\} . \tag{29}
\end{equation*}
$$

Note that since $\mathbf{a} \in \mathbf{a}^{\{1, \mathbf{k}\}}$ refers to the coefficient vector of scheduled users, it has no zero entries.

The following Lemma shows an important property of the optimal coefficient vector a which maximizes $\mathcal{R}(\mathbf{h}, \mathbf{a})$.

Lemma 3. The optimal vector a satisfies either, $\operatorname{sign}\left(h_{i}\right)=\operatorname{sign}\left(a_{i}\right)$ for all $i$ or $\operatorname{sign}\left(h_{i}\right) \neq$ $\operatorname{sign}\left(a_{i}\right)$ for all $i$.

Proof: Considering the rate expression (5), since $\|\mathbf{a}\|^{2}$ does not depend on the signs, the optimal signs must maximize the inner product $\left(\mathbf{h}^{T} \mathbf{a}\right)^{2}$. Obviously, all signs must match (or be oposite) in order to have only positive (or negative) elements in the summation of the inner product.

The usefulness of Lemma 3 is in making the inner maximization in (29) trivial, since given a subset of users $\mathcal{S}_{k}$ with channel coefficients $\mathbf{h}\left(\mathcal{S}_{k}\right)$, the optimal $\mathbf{a}\left(\mathcal{S}_{k}\right) \in \mathbf{a}^{\{\mathbf{1}, \mathbf{k}\}}$ is clear - just set the signs according to those of $\mathbf{h}\left(\mathcal{S}_{k}\right)$. Consequently, the following procedure is optimal for solving (29): disregard the signs in $\mathbf{h}_{L}$; find the optimal subset $\mathcal{S}_{k}$ by considering only absolute values and finding the subset which best fits $\mathbf{a}=(1,1, \ldots, 1) \triangleq \mathbf{1}$; then simply set the signs of a from all positive to the original signs of $\mathbf{h}\left(\mathcal{S}_{k}\right)$. This reduces the double optimization in (29), with $2^{k}$ options in the inner one, to a much simpler optimization:

$$
\begin{equation*}
\mathcal{S}_{k}^{\mathbf{a}^{\{\mathbf{1}, \mathbf{k}\}}}=\underset{\mathcal{S}_{k}}{\arg \max }\left\{\mathcal{R}\left(\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|, \mathbf{1}\right)\right\}, \tag{30}
\end{equation*}
$$

where $\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|=\left(\left|h_{1}\right|,\left|h_{2}\right|, \ldots,\left|h_{k}\right|\right)$. Thus, when searching for the optimal schedule in each slot, we significantly relax the optimization: we fix a reasonably good a and search for the best $\mathbf{h}$. As it turns out, this will be asymptotically optimal.

## B. Best channel for a fixed a

Towards the solution of (30), note that

$$
\begin{align*}
\underset{\mathcal{S}_{k}}{\arg \max }\left\{\mathcal{R}\left(\mathbf{h}\left(\mathcal{S}_{k}\right), \mathbf{a}\right)\right\} & =\underset{\mathcal{S}_{k}}{\arg \max }\left\{\frac{1}{2} \log ^{+}\left(\|\mathbf{a}\|^{2}-\frac{\mathrm{P}\left(\mathbf{h}\left(\mathcal{S}_{k}\right)^{T} \mathbf{a}\right)^{2}}{1+\mathrm{P}\left\|\mathbf{h}\left(\mathcal{S}_{k}\right)\right\|^{2}}\right)^{-1}\right\} \\
& =\frac{1}{2} \log ^{+}\left(\|\mathbf{a}\|^{2}-\underset{\mathcal{S}_{k}}{\arg \max }\left\{\frac{\mathrm{P}\left(\mathbf{h}\left(\mathcal{S}_{k}\right)^{T} \mathbf{a}\right)^{2}}{1+\mathrm{P}\left\|\mathbf{h}\left(\mathcal{S}_{k}\right)\right\|^{2}}\right\}\right)^{-1} \\
& =\underset{\mathcal{S}_{k}}{\arg \max }\left\{\frac{\mathrm{P}\left(\mathbf{h}\left(\mathcal{S}_{k}\right)^{T} \mathbf{a}\right)^{2}}{1+\mathrm{P}\left\|\mathbf{h}\left(\mathcal{S}_{k}\right)\right\|^{2}}\right\}  \tag{31}\\
& =\underset{\mathcal{S}_{k}}{\arg \max }\left\{\frac{\mathrm{P}\left\|\mathbf{h}\left(\mathcal{S}_{k}\right)\right\|^{2}\|\mathbf{a}\|^{2} \cos ^{2}(\theta)}{1+\mathrm{P}\left\|\mathbf{h}\left(\mathcal{S}_{k}\right)\right\|^{2}}\right\} \\
& =\underset{\mathcal{S}_{k}}{\arg \max }\left\{\frac{\cos ^{2}(\theta)}{1+\frac{1}{\mathrm{P}\left\|\mathbf{h}\left(\mathcal{S}_{k}\right)\right\|^{2}}}\right\} .
\end{align*}
$$

Thus, generally speaking, an $h$ which maximizes the achievable rate should have a high norm and a small angle with a. Clearly, the highest norm vector may not be the one with the smallest angle to a. Thus, the scheduler should seek the optimal tradeoff point to maximize the achievable rate. Our suggested scheduling algorithm, given in the next subsection, searches for this optimal tradeoff point in polynomial-time. The search relies on the following lemma which shows that, for the case of all-ones coefficient vector, sorting the channel vector $\mathbf{h}_{L}$ by the elements' absolute value simplifies the search significantly. Thus, we define $\mathbf{h}_{L}^{s}$ to be an ascending ordered vector according to $\left|\mathbf{h}_{L}\right|$.

Lemma 4. The optimal subset $\mathcal{S}_{k}$ for the all-ones vector $\mathbf{1}$ is a subset for which $\mathbf{h}\left(\mathcal{S}_{k}\right)$ is $k$ consecutive elements in $\mathbf{h}_{L}^{s}$. That is,

$$
\max _{\mathcal{S}_{k}}\left\{\mathcal{R}\left(\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|, \mathbf{1}\right)\right\}=\max _{i}\left\{\mathcal{R}\left(\left|\mathbf{h}_{i}^{\prime}\right|, \mathbf{1}\right)\right\}
$$

where $\mathbf{h}_{i}^{\prime}=\left(h_{L, i}^{s}, \ldots, h_{L, i+k-1}^{s}\right)$ for $i \in[1, \ldots, L-k+1]$.

Proof: Consider the expression for the achievable rate in Theorem 1. We have,

$$
\begin{align*}
\max _{\mathcal{S}_{k}}\left\{\mathcal{R}\left(\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|, \mathbf{1}\right)\right\} & =\max _{\mathcal{S}_{k}}\left\{\frac{1}{2} \log ^{+}\left(\frac{\mathrm{P}}{\min _{\alpha \in \mathbb{R}}\left\{\alpha^{2}+\mathrm{P}\left\|\alpha\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|-\mathbf{1}\right\|^{2}\right\}}\right)\right\} \\
& =\frac{1}{2} \log ^{+}\left(\frac{\mathrm{P}}{\min _{\mathcal{S}_{k}}\left\{\min _{\alpha \in \mathbb{R}}\left\{\alpha^{2}+\mathrm{P}\left\|\alpha\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|-\mathbf{1}\right\|^{2}\right\}\right\}}\right)  \tag{32}\\
& =\frac{1}{2} \log ^{+}\left(\frac{\mathrm{P}}{\min _{\alpha>0}\left\{\min _{\mathcal{S}_{k}}\left\{\alpha^{2}+\mathrm{P}\left\|\alpha\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|-\mathbf{1}\right\|^{2}\right\}\right\}}\right)
\end{align*}
$$

where in the last line we can reduce the minimization to $\alpha>0$ since for $\alpha<0$ we would increase the term for all $\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|$. Therefore we need to show that for any $\alpha>0$

$$
\mathbf{h}\left(\underset{\mathcal{S}_{k}}{\arg \min }\left\{\left\|\alpha\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|-\mathbf{1}\right\|^{2}\right\}\right)=\mathbf{h}_{i}^{\prime},
$$

for some $i \in[1, \ldots, L-k+1]$.
Define the sequence $\Delta_{j}=\left(\alpha h_{L, j}^{s}-1\right)^{2}$, for $j=1, \ldots, L$. This sequence can be monotonic increasing, monotonic decreasing or monotonic decreasing and then monotonic increasing with $j$; it depends on the value of $\alpha h_{L, 1}^{s}$ and $\alpha h_{L, L}^{s}$ with respect to 1 . For example, if $\alpha h_{L, 1}^{s}>1$ then the sequence is monotonic increasing with $j$. Let us choose some $\mathcal{S}_{k}$ with $\mathbf{h}\left(\mathcal{S}_{k}\right)$ such that its corresponding elements in $\mathbf{h}_{L}^{s}$ are not consecutive. Hence, w.l.o.g. assume that two elements in $\mathbf{h}\left(\mathcal{S}_{k}\right)$ corresponds to two elements $h_{L, i}^{s}$ and $h_{L, j}^{s}$ such that $i+1 \neq j$. Accordingly, either the choices $\mathbf{h}_{j}^{\prime}$ or $\mathbf{h}_{j-k}^{\prime}$ will minimize $\left\{\left\|\alpha\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|-\mathbf{1}\right\|^{2}\right\}$ since in at least one of the choices we would decreased with the sequence $\Delta_{i}$. Note also that this is true for the choices $\mathbf{h}_{i}^{\prime}$ or $\mathbf{h}_{i-k}^{\prime}$

## C. Scheduling Algorithm

In this section, we present a polynomial-time scheduling algorithm presented as Algorithm 1 which is an asymptotically optimal solution for the scheduling problem as defined in (18). The algorithm relies on the properties suggested in the previous subsections, and therefore in each slot searches the subset of users which maximizes (30). Specifically, this search is done by the subroutine Algorithm 2 which relies on Lemmas 3 and 4. The output of Algorithm 1 is the set of subsets of users which should be scheduled in each slot along with the decoding matrix $\mathbf{A}$.

The complexity of Algorithm 2 is $O(L \log L+(L-k) k)$ due to the sorting of $\mathbf{h}_{L}$ and the scan of $L-k$ scheduling options for which it computes the achievable rate on vectors with length
$k$. Accordingly, the complexity of Algorithm 1 is $O((L-1)(L \log L+(L-k) k))$ which uses Algorithm $2 L-1$ slots. With $k=O(\log L)$, this results in $O\left(L^{2} \log L\right)$. Since $O\left(L^{2} \log L\right)$ is the required complexity to invert the coefficient matrix [34], the complexity of the scheduling algorithm is within this range and does not add any significant computations above the necessary order.

Simulation results of the system's expected sum-rate for Algorithm 1 is depicted in Figure 8 , The simulation results are compared with an upper bound on the expected system sum-rate which is calculated as the mean of the highest achievable rate in $L$ slots. That is, we ignore the rank restriction and take the highest rate in each transmission slot. However, since this calculation becomes prohibitively complex, the exhaustive search was done only up to $L=40$. For larger $L$ the plot merely an interpolation. The figure also includes the upper bound as given in Theorem 7 and $\log \log L$ to reflect the scaling law anticipated by Corollary 4. One can observe that, as $L$ grows, the expected sum-rate that Algorithm 1 provides coincides with the curve of the optimal schedule.

Remark 2 (Real-time algorithm). Algorithm 1 attains its optimality without requiring global CSI of all slots in advance. That is, the scheduler only needs the CSI in the beginning of each slot in order to determine which $k$ users to schedule.

Algorithm 1 completes, yet successful decoding can actually occur if both the rank of the decoding matrix $\mathbf{A}_{L}$ is $L$ and the transmission rate of all users were below the minimal achievable rate at the relay among all slots. Theorem 8 and Lemma 6 in the next subsection discuss these two critical conditions.

## D. Successful Decoding of Algorithm 1

We first link the number of transmitters, the number of scheduled users, and the coefficient vectors to the rank of $\mathbf{A}_{L}$. Recall that in each slot a coefficient vector $\mathbf{a}\left(\mathcal{S}_{k}\right) \in \mathbf{a}^{\{1, \mathbf{k}\}}$ is chosen by the relay for the scheduled subset of users $\mathcal{S}_{k}$. This vector is then mapped to an $L$-length vector using the function $(\cdot)_{\Uparrow}^{L}$. Thus, is each slot, a vector from the set $\mathbb{S}_{L, k}^{\{-1,1\}}$ is added to $\mathbf{A}_{L}$. Since the users' channel coefficients are i.i.d., the scheduled subset $\mathcal{S}_{k}$ in each slot is uniformly distributed among all subsets. In addition, so do the signs of the channel's coefficients. Accordingly, the distribution of the vectors from $\mathbb{S}_{L, k}^{\{-1,1\}}$ which are added to $\mathbf{A}_{L}$ is also uniform.

```
Algorithm 1 Scheduling for CF (single relay)
    Input: \(\mathbf{h}_{L}\)
    Output: \(\left(\mathcal{S}_{k}^{N}, \mathbf{A}\right)\) or error
```


## Initialization:

```
    \(\mathbf{A} \leftarrow \emptyset ; L \leftarrow \operatorname{length}\left(\mathbf{h}_{L}\right)\)
```

    \(R_{\text {min }} \leftarrow \infty ; R^{*} \leftarrow 0 ; R \leftarrow\) The transmission rate
    \(k \leftarrow\lceil\ln L\rceil+1\)
    \(\mathcal{S}_{k}^{N} \leftarrow \emptyset\)
    
## Main:

for $i=1 ; i \leq L-1 ; i++\mathbf{d o}$
$\left(\mathcal{S}_{k}^{*}, \mathbf{a}^{*}, R^{*}\right) \leftarrow$ Algorithm2 $\left(\mathbf{h}_{L}, \mathbf{1}^{k}\right)$
Schedule users according to $\mathcal{S}_{k}^{*}$ and $\mathcal{S}_{k}^{N} \leftarrow \mathcal{S}_{k}^{N} \cup \mathcal{S}_{k}^{*}$
$\mathbf{A} \leftarrow \mathbf{A} \cup\left(\mathbf{a}^{*}\left(\mathcal{S}_{k}^{*}\right)\right)_{\Uparrow}^{L}$
if $R^{*}<R_{\text {min }}$ then

$$
R_{\min }=\mathcal{R}\left(\mathbf{h}\left(\mathcal{S}_{k}^{*}\right), \mathbf{a}^{*}\left(\mathcal{S}_{k}^{*}\right)\right)
$$

end if
end for
Find $\mathbf{e}_{i}$ that is linearly independent with $\mathbf{A}$
$\mathbf{A} \leftarrow \mathbf{A} \cup \mathbf{e}_{i}$
Schedule user $i$
if $\operatorname{Rank}(\mathbf{A})==L \& \& R<R_{\text {min }}$ then
Solve linear system According to A
return $\left(\mathcal{S}_{k}^{N}, \mathbf{A}\right)$
else
return (error)
end if

```
Algorithm 2 Finding optimal schedule for \(\mathbf{a}=1\) per slot
    Input: \(\left(\mathbf{h}_{L}, \mathbf{a}\right)\)
    Output: \(\left(\mathcal{S}_{k}^{*}, \mathbf{a}^{*}, R^{*}\right)\)
```


## Initialization:

1: $\mathbf{h}_{L}^{s} \leftarrow$ sort according to $\operatorname{Abs}\left(\mathbf{h}_{L}\right)$
: $\mathbf{h}_{L}^{I} \leftarrow$ ordering of $\mathbf{h}_{L}^{s}$ in $\mathbf{h}_{L}$
$\triangleright$ indices vector
$\mathbf{h}_{L}^{\text {sign }} \leftarrow \mathbf{h}_{L}^{s} \cdot / \operatorname{Abs}\left(\mathbf{h}_{L}^{s}\right) \quad \triangleright$ element-wise devision
$: k \leftarrow \operatorname{length}(\mathbf{a})$
$: R^{*} \leftarrow 0$;
6: $i^{*} \leftarrow-1$
7: $\mathcal{S}_{k}^{*} \leftarrow \emptyset$
8: $\mathbf{a}^{*} \leftarrow \emptyset$

## Main:

: for $i=1 ; i \leq L-k ; i++\mathbf{d o}$
10:
$\mathbf{h} \leftarrow \mathbf{h}_{L}^{s}(i: i+k)$
if $\mathcal{R}(|\mathbf{h}|, \mathbf{a})>R^{*}$ then
$i^{*} \leftarrow i$
$R^{*} \leftarrow \mathcal{R}(|\mathbf{h}|, \mathbf{a})$
end if
end for
if $i^{*}=-1$ then return $(\{1: k\}, \mathbf{a})$
$\triangleright$ zero rate
else
$\mathcal{S}_{k}^{*} \leftarrow\left\{\mathbf{h}_{L}^{I}\left(i^{*}\right): \mathbf{h}_{L}^{I}\left(i^{*}\right)+k-1\right\}$
$\mathbf{a}^{*} \leftarrow \mathbf{a} . * \mathbf{h}_{L}^{\text {sign }}\left(i^{*}: i^{*}+k-1\right)$
end if
return $\left(\mathcal{S}_{k}^{*}, \mathbf{a}^{*}, R^{*}\right)$


Figure 8: The system's expected sum-rate of the Algorithm 1 compared with the expected sumrate for the optimal schedule as a function of $L$ with $\mathrm{P}=1000$. The asymptotic upper bound which is given in Theorem 7 along with the function $\log \log L$, which is the scaling law Algorithm ${ }_{1}$ achieves as given in Corollary 4, are also plotted.

Theorem 8 below shows that with $O(\log L)$ scheduled users in each slot, we can indeed achieve a full rank with high probability after $L$ slots.

Theorem 8. Assume $k=O(\log L)$ and the coefficient vectors for the scheduled subset $\mathcal{S}_{k}$ are drawn uniformly from $\mathbf{a}^{\{1, \mathbf{k}\}}$ for $L-1$ slots. Then, there exists a unit vector such that selecting it at the Lth slot results in $\operatorname{rank}\left(\mathbf{A}_{L}\right)=L$ with probability $1-o(1)$.

In other words, Theorem 8 asserts that by uniformly selecting vectors from $\mathbf{a}^{\{1, \mathbf{k}\}}$ for $L-$ 1 slots, then completing the matrix with a single unit vector, one has a full rank with high probability. To prove Theorem 8, we first give the lemma bellow.

Lemma 5. Let $\mathbf{A}_{n}$ be the decoding matrix at slot $n$, where each row $i \leq n$ is a vector from $\mathbb{S}_{L, k}^{\{-1,1\}}$ corresponding to $\left(\mathbf{a}\left(\mathcal{S}_{k}(i)\right)\right)_{\Uparrow}^{L}$, where $\mathbf{a}\left(\mathcal{S}_{k}(i)\right)$ was drawn uniformly from $\mathbf{a}^{\{\mathbf{1}, \mathbf{k}\}}$. If

$$
\begin{equation*}
n<\left(1-\frac{e^{-k}}{\ln 2}-o\left(e^{-k}\right)\right) L \tag{33}
\end{equation*}
$$

then $\operatorname{rank}\left(\mathbf{A}_{n}\right)=n$ with probability $1-o(1)$.
Proof: The proof follows from [35, Theorem 3], also appearing in [36] with a similar formulation. Specifically, the results in [35], [36] consider random vectors over the binary field $\mathbb{F}_{2}^{L}$ with exactly $k$ ones, i.e., $\mathbb{S}_{L, k}^{\{1\}}$.

Define the matrix $\mathbf{B}_{n}=\left(\mathbf{A}_{n} \bmod 2\right)$, where the modulo operation is element-wise. The modulo operation acts as a mapping between vectors in $\mathbb{S}_{L, k}^{\{-1,1\}}$, which have non-zero elements at certain positions, to vectors in $\mathbb{S}_{L, k}^{\{1\}}$ which have only ones at the same positions. Every -1 becomes 1 . Since the rows of $\mathbf{A}_{n}$ are uniformly distributed from $\mathbb{S}_{L, k}^{\{-1,1\}}$ and for every vector in $\mathbb{S}_{L, k}^{\{1\}}$ there are exactly $2^{k}$ corresponding vectors in $\mathbb{S}_{L, k}^{\{-1,1\}}$, the probability remains uniform on the elements of $\mathbb{S}_{L, k}^{\{1\}}$. Thus, the rows of $\mathbf{B}_{n}$ are uniformly distributed from $\mathbb{S}_{L, k}^{\{1\}}$ and according to [35], if $n<\left(1-\frac{e^{-k}}{\ln 2}-o\left(e^{-k}\right)\right) L$ then $\operatorname{rank}\left(\mathbf{B}_{n}\right)=n$ with probability $1-o(1)$.
Thus, since $\operatorname{rank}\left(\mathbf{A}_{n}\right) \geq \operatorname{rank}\left(\mathbf{B}_{n}\right)$, the result follows. Finally, since we will need a more precise expression for the $o\left(e^{-k}\right)$ term we note that this is in fact [35],
$\frac{1}{2 \log 2}\left(k^{2}-2 k+\frac{2 k}{\log 2}-1\right) e^{-2 k}-O\left(k^{4}\right) e^{-3 k}=o\left(e^{-k}\right)$.
Proof of Theorem 8: According to Lemma 5, a rank of $L-1$ can be achieved in $L-1$ slots with probability $1-o(1)$ as long as

$$
\begin{equation*}
L-1<\left(1-\frac{e^{-k}}{\ln 2}-o\left(e^{-k}\right)\right) L \tag{34}
\end{equation*}
$$

The above inequality reduces to

$$
\begin{equation*}
\frac{1}{L}>\frac{e^{-k}}{\ln 2}+o\left(e^{-k}\right) \tag{35}
\end{equation*}
$$

which is satisfied as long as

$$
\begin{equation*}
k>\ln L+\ln 2 . \tag{36}
\end{equation*}
$$

Note that the term $o\left(e^{-k}\right)$ stated in Lemma 5 decays very fast with $k$. Letting $k=\lceil\ln L\rceil+1$ meets the requirement above. In the $L$-th slot, the scheduler may schedule a single user (a unit vector) which will be the missing degree of freedom for achieving rank $L$. Such vector always exists since otherwise, this means that all the unit vectors are spanned by $\mathbf{A}_{L-1}$ which is a contradiction.

Theorem 8 shows that indeed the rank of $\mathbf{A}_{L}$ is $L$ with probability $1-o(1)$, which is asymptotically the shortest completion time for a single relay model.

We now show that the rate restriction (8) is also satisfied. That is, if $R$ is the transmission rate of all the users, then the following must be satisfied,

$$
\begin{equation*}
R<\min _{n: 1, \ldots, L-1} \mathcal{R}(\mathbf{h}(n), \mathbf{a}(n)) \tag{37}
\end{equation*}
$$

Accordingly, we define an outage scenario when the above condition is not satisfied. Formally,

$$
\begin{equation*}
P_{\text {out }}(R)=P_{r}\left(\min _{n: 1, \ldots, L-1}\left\{\mathcal{R}\left(\mathbf{h}^{*}(n), \mathbf{a}^{*}(n)\right)\right\}<R\right), \tag{38}
\end{equation*}
$$

where $\mathbf{h}^{*}(n)$ and $\mathbf{a}^{*}(n)$ are the channel and coefficient vectors of the scheduled subset $\mathcal{S}_{k}^{*}(n)$ in time slot $n$. Note that we assume that in the last slot a single user $i$ is scheduled; thus it's rate is restricted to a SISO Gaussian channel capacity, i.e., $R<\frac{1}{2} \log \left(1+h_{i}^{2} \mathrm{P}\right)$. The following lemma shows that using the lower bound on the achievable rate, as given in Theorem 6, allows a zero outage probability at the limit of large $L$.

Lemma 6. If the transmission rate is set to be $R=\left(\frac{1}{4}-\epsilon\right) \log \log L$ where $\epsilon$ can be chosen to be arbitrary small, the outage probability vanishes. That is,

$$
\begin{equation*}
\lim _{L \rightarrow \infty} P_{\text {out }}(R)=0 . \tag{39}
\end{equation*}
$$

The proof is given in Appendix C.

## E. Proofs for the Asymptotic Guarantees

We are now ready to present the proofs for Theorems 6 and 7 which rely on the previous subsections.

Proof of Theorem 6. The expected rate that can be achieved in each slot by Algorithm 1 is as follows. Note that the time index is omitted due to the independence between the slots.

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{R}_{a c h}^{s c h}\right] & =\mathbb{E}\left[\max _{\mathcal{S}_{k}}\left\{\max _{\mathbf{a}\left(\mathcal{S}_{k}\right) \in \mathbb{Z} k \backslash\{\mathbf{0}\}}\left\{\mathcal{R}\left(\mathbf{h}\left(\mathcal{S}_{k}\right), \mathbf{a}\left(\mathcal{S}_{k}\right)\right)\right\}\right\}\right] \\
& \stackrel{(a)}{\geq} \mathbb{E}\left[\max _{\mathcal{S}_{k}}\left\{\max _{\mathbf{a}\left(\mathcal{S}_{k}\right) \in \mathbf{a}\{1, \mathbf{k}\}}\left\{\mathcal{R}\left(\mathbf{h}\left(\mathcal{S}_{k}\right), \mathbf{a}\left(\mathcal{S}_{k}\right)\right)\right\}\right\}\right] \\
& \stackrel{(b)}{\geq} \mathbb{E}\left[\max _{\mathcal{S}_{k}}\left\{\mathcal{R}\left(\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|, \mathbf{1}\right)\right\}\right],
\end{aligned}
$$

(a) and (b) follow since we reduce the search domain as explained in (29) and (30), respectively. This enables the following steps,

$$
\begin{aligned}
\mathbb{E}\left[\max _{\mathcal{S}_{k}}\left\{\mathcal{R}\left(\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|, \mathbf{1}\right)\right\}\right] & =\mathbb{E}\left[\max _{\mathcal{S}_{k}}\left\{\frac{1}{2} \log ^{+}\left(k-\frac{\mathrm{P}\left(\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|^{T} \mathbf{1}\right)^{2}}{1+\mathrm{P}\left\|\mathbf{h}\left(\mathcal{S}_{k}\right)\right\|^{2}}\right)^{-1}\right\}\right] \\
& =\mathbb{E}\left[\frac{1}{2} \log ^{+}\left(k-\max _{\mathcal{S}_{k}}\left\{\frac{\mathrm{P}\left(\left|\mathbf{h}\left(\mathcal{S}_{k}\right)\right|^{T} \mathbf{1}\right)^{2}}{1+\mathrm{P}\left\|\mathbf{h}\left(\mathcal{S}_{k}\right)\right\|^{2}}\right\}\right)^{-1}\right] \\
& \stackrel{(a)}{\geq}\left[\frac{1}{2} \log ^{+}\left(k-\frac{\mathrm{P}\left(\mathbf{h}^{\prime T} \mathbf{1}\right)^{2}}{1+\mathrm{P}\left\|\mathbf{h}^{\prime}\right\|^{2}}\right)^{-1}\right]
\end{aligned}
$$

$$
\begin{align*}
& \stackrel{(b)}{\geq} \frac{1}{2} \log ^{+}\left(k-\mathbb{E}\left[\frac{\mathrm{P}\left(\mathbf{h}^{\prime T} \mathbf{1}\right)^{2}}{1+\mathrm{P}\left\|\mathbf{h}^{\prime}\right\|^{2}}\right]\right)^{-1} \\
& =\frac{1}{2} \log ^{+}\left(k-\mathbb{E}\left[\frac{\mathrm{P} k\left\|\mathbf{h}^{\prime}\right\|^{2} \cos ^{2}\left(\theta^{\prime}\right)}{1+\mathrm{P}\left\|\mathbf{h}^{\prime}\right\|^{2}}\right]\right)^{-1}, \tag{40}
\end{align*}
$$

where (a) is by choosing some specific $\mathbf{h}\left(\mathcal{S}_{k}^{\prime}\right)$, denoted in short by $\mathbf{h}^{\prime}$ and (b) follows from Jensen's inequality.

As section V-B suggests, the optimal schedule should be a subset of users with a high norm channel vector and a small angle between its channel vector and the corresponding coefficient vector. Thus, define the values $u(L)$ and $\delta(L)$ such that $\mathbf{h}^{\prime}$ maintains

$$
\begin{equation*}
u \leq\left|h_{i}^{\prime}\right| \leq u+\delta, \forall i \tag{41}
\end{equation*}
$$

With this definition, we are able to bound the parameters for a good schedule. The values of $u(L)$ and $\delta(L)$ can help tune the norm (by taking a high value of $u$ ) and the angle with 1 (by taking a small value of $\delta$ ) to attain a high rate (see Section V-B).

Let $\xi$ denote the event of having at least $k$ elements in $\mathbf{h}_{L}(n)$ with values in the interval $[-(u+\delta),-u] \cup[u, u+\delta]$ for all $n$, i.e., in each of the $L$ slots. We thus write the last equation in (40) as follows,

$$
\begin{align*}
& =\frac{1}{2} \log ^{+}\left(k-\left(\mathbb{E}\left[\left.\frac{\mathrm{P} k\left\|\mathbf{h}^{\prime}\right\|^{2} \cos ^{2}\left(\theta^{\prime}\right)}{1+\mathrm{P}\left\|\mathbf{h}^{\prime}\right\|^{2}} \right\rvert\, \xi\right] P_{r}(\xi)+\mathbb{E}\left[\left.\frac{\mathrm{P} k\left\|\mathbf{h}^{\prime}\right\|^{2} \cos ^{2}\left(\theta^{\prime}\right)}{1+\mathrm{P}\left\|\mathbf{h}^{\prime}\right\|^{2}} \right\rvert\, \bar{\xi}\right]\left(1-P_{r}(\xi)\right)\right)\right)^{-1}  \tag{42}\\
& \geq \frac{1}{2} \log ^{+}\left(k-\mathbb{E}\left[\left.\frac{\mathrm{P} k\left\|\mathbf{h}^{\prime}\right\|^{2} \cos ^{2}\left(\theta^{\prime}\right)}{1+\mathbf{P}\left\|\mathbf{h}^{\prime}\right\|^{2}} \right\rvert\, \xi\right] P_{r}(\xi)\right)^{-1} .
\end{align*}
$$

Under $\xi$, we can lower bound $\left\|\mathbf{h}^{\prime}\right\|^{2}$ and $\cos ^{2}\left(\theta^{\prime}\right)$ as follows,

$$
\begin{align*}
& \left\|\mathbf{h}^{\prime}\right\|^{2} \geq k u^{2} \\
& \cos ^{2}\left(\theta^{\prime}\right)=\frac{\left(\sum_{i=1}^{k} h_{i}^{\prime}\right)^{2}}{k\left\|\mathbf{h}^{\prime}\right\|^{2}} \geq \frac{k^{2} u^{2}}{k^{2}(u+\delta)^{2}}=\frac{u^{2}}{(u+\delta)^{2}} . \tag{43}
\end{align*}
$$

The probability $P_{r}(\xi)$ can be computed and lower bounded as follows. Consider the probability to find in a certain slot $k$ users satisfying (41). This probability follows a binomial distribution with probability of success $p(u, \delta)=2(\Phi(u+\delta)-\Phi(u))$ where $\Phi$ is the CDF of the normal distribution. Accordingly, since the $L$ time slots are independent we have,

$$
\begin{align*}
P_{r}(\xi) & =\left(\sum_{i=k}^{L}\binom{L}{i} p(u, \delta)^{i}(1-p(u, \delta))^{L-i}\right)^{L} \\
& =\left(1-\sum_{i=0}^{k-1}\binom{L}{i} p(u, \delta)^{i}(1-p(u, \delta))^{L-i}\right)^{L}  \tag{44}\\
& \geq\left(1-e^{-\frac{1}{2 p(u, \delta)} \frac{(L p(u, \delta)-(k-1))^{2}}{L}}\right)^{L}
\end{align*}
$$

where the last row follows from the Chernoff's bound for the lower tail, which requires that $L p(u, \delta) \geq k-1$. By setting

$$
\begin{equation*}
u(L)=\sqrt{2 \ln \frac{\delta \sqrt{L}}{\sqrt{2 \pi}}}-\delta \quad \text { and } \quad \delta(L)=\frac{1}{\ln L} \tag{45}
\end{equation*}
$$

the requirement is satisfied and the probability $P_{r}(\xi)$ goes to one with $L$. This is since for for $L \geq 4$ we have

$$
\begin{align*}
p(u, \delta) & =2(\Phi(u+\delta)-\Phi(u)) \\
& =\frac{2}{\sqrt{2 \pi}} \int_{u}^{u+\delta} e^{-\frac{t^{2}}{2}} d t \\
& \geq \delta \frac{2}{\sqrt{2 \pi}} e^{-\frac{(u+\delta)^{2}}{2}}  \tag{46}\\
& =\delta \frac{2}{\sqrt{2 \pi}} e^{-\frac{\left(\sqrt{\left.2 \ln \frac{\delta \sqrt{L}}{\sqrt{2 \pi}}-\delta+\delta\right)^{2}}\right.}{2}} \\
& =\frac{2}{\sqrt{L}},
\end{align*}
$$

and since, $\frac{2}{\sqrt{L}} \geq \frac{\lceil\ln L\rceil}{L}$ the Chernoff's bound requirement is satisfied when choosing $k$ according to (36). On the other hand,

$$
\begin{align*}
\lim _{L \rightarrow \infty} P_{r}(\xi) & \geq \lim _{L \rightarrow \infty}\left(1-e^{-\frac{1}{2 p(u, \delta)} \frac{(L p(u, \delta)-(k-1))^{2}}{L}}\right)^{L} \\
& \stackrel{(a)}{\geq} \lim _{L \rightarrow \infty} 1-L e^{-\frac{\sqrt{L}}{4} \frac{(2 \sqrt{L}-\lceil\ln L])^{2}}{L}}  \tag{47}\\
& =1,
\end{align*}
$$

where (a) is due to (46) and Bernoulli's inequality which state that, $(1-x)^{L} \geq 1-L x$ for $L \geq 1$ and $0 \leq x \leq 1$.

As $L$ grows, the values of $u(L)$ and $\delta(L)$ are increasing and decreasing, respectively. Specifically, the slow increase of $u(L)$ promises that the channel gains of the scheduled users will increase, while the decrease of $\delta(L)$ improves the match to the all-ones coefficient vector.

Substituting (43) and (44) in (42) we have,

$$
\begin{align*}
& \frac{1}{2} \log ^{+}\left(k-\mathbb{E}\left[\left.\frac{\mathrm{P} k\left\|\mathbf{h}^{\prime}\right\|^{2} \cos ^{2}\left(\theta^{\prime}\right)}{1+\mathrm{P}\left\|\mathbf{h}^{\prime}\right\|^{2}} \right\rvert\, \xi\right] P_{r}(\xi)\right)^{-1} \\
& \quad \geq \frac{1}{2} \log ^{+}\left(k\left(1-\frac{\mathrm{P} k u^{4}}{(u+\delta)^{2}\left(1+\mathrm{P} k u^{2}\right)}\left(1-e^{-\frac{1}{2 p(u, \delta)}} \frac{\left(\frac{(L p(u, \delta)-(k-1))^{2}}{L}\right.}{L}\right)^{L}\right)\right)^{-1}  \tag{48}\\
& \quad \stackrel{(a)}{=} \frac{1}{2} \log ^{+}\left(k\left(1-\frac{\mathrm{P} k u^{4}}{(u+\delta)^{2}\left(1+\mathrm{P} k u^{2}\right)}(1-o(1))\right)\right)^{-1} .
\end{align*}
$$

It can be verified (see Appendix D) that the scaling law of (48) is indeed $\frac{1}{4} \log \log L$, which completes the proof.

Proof of Corollary 3: The system expected sum-rate can be lower bounded as follows,

$$
\begin{align*}
\mathbb{E}\left[C_{S R}\right] & =\mathbb{E}\left[\frac{R L}{N}\right] \\
& \geq \mathbb{E}\left[\left.\frac{R L}{N} \right\rvert\, N=L\right] P_{r}(N=L) \\
& \stackrel{(a)}{=}\left(\frac{R(L-1)}{L}+\frac{R^{S I S O}}{L}\right)(1-o(1))  \tag{49}\\
& \geq \frac{R(L-1)}{L}(1-o(1)) \\
& =\left(\frac{1}{4}-\epsilon\right) \log \log L(1-o(1))
\end{align*}
$$

(a) follows due to Theorem 8 and the fact that Algorithm 1 schedules in the last slot a single user with rate of the SISO channel. The other slots are with rate $R$. Finally, Lemma 6 guarantees that if one sets $R=\left(\frac{1}{4}-\epsilon\right) \log \log L$ for any small positive $\epsilon$, an innovative linear combination with rate $R$ would be successfully decoded in each of the first $L-1$ slots with probability that goes to 1 as $L$ grows.

Proof of Theorem 7; In [37], the following universal upper bound on the achievable rate was given

$$
\begin{equation*}
\mathcal{R}\left(\mathbf{h}, \mathbf{a}^{o p t}\right) \leq \frac{1}{2} \log \left(1+\mathrm{P} \max _{i}\left\{h_{i}^{2}\right\}\right), \tag{50}
\end{equation*}
$$

where $\mathbf{h}$ is any channel vector of dimension $k$ and $\mathbf{a}^{o p t}$ is the coefficient vector which maximizes the achievable rate. Considering (9), we have

$$
\begin{aligned}
\mathbb{E}\left[C_{S R}\right] & =\mathbb{E}\left[\frac{R L}{N}\right] \\
& \stackrel{(a)}{\leq} \mathbb{E}\left[\min _{n=1, \ldots, N} \mathcal{R}(\mathbf{h}(n), \mathbf{a}(n))\right] \\
& \leq \mathbb{E}\left[\mathcal{R}\left(\mathbf{h}, \mathbf{a}^{o p t}\right)\right] \\
& \stackrel{(b)}{\leq} \mathbb{E}\left[\frac{1}{2} \log \left(1+\mathrm{P} \max _{i: 1 . . k}\left\{h_{i}^{2}\right\}\right)\right] \\
& \stackrel{(c)}{\leq} \frac{1}{2} \log \left(1+\mathrm{P} \mathbb{E}\left[\max _{i: 1 . . L}\left\{h_{L i}^{2}\right\}\right]\right) \\
& \stackrel{(d)}{=} \frac{1}{2} \log \left(1+\mathrm{P}\left(2 \ln L-\ln \ln L-2 \ln \Gamma\left(\frac{1}{2}\right)+\frac{\gamma}{2}+o(1)\right)\right)
\end{aligned}
$$

Where in (a) we set $N=L$ which is the minimal number of transmission slots and in $(b)$ we used the universal upper bound in (50). (c) follows due to Jensen's inequality and the consideration of all values of $\mathbf{h}_{L} .(d)$ follows from the asymptotic results for the expectation of the maximum value in a $\chi^{2}$ i.i.d. random vector of dimension $L$ in the limit of large $L$ [38, Table 3.4.4]. It can be verified (see Appendix E) that the scaling law is $O\left(\frac{1}{2} \log \log L\right)$, which completes the proof.

Corollary 3 and Theorem 7 show that the upper and lower bounds on the expected sum-rate scale as $O(\log \log L)$, which results in Corollary 4 .

## F. Distributed Scheduling

The asymptotic guarantees presented in the previous subsection rely on the existence of $k$ users in a predefined interval of channel gain values. In fact, Algorithm 2 only presents an efficient way for the scheduler to find these channel gains and the corresponding users in each slot. Considering a single relay model, the obvious choice for the scheduler is the relay itself, since the relay has the channel vector from all users to it; in addition, the relay performs the decoding of the received messages. Nevertheless, one can devise a distributed threshold-based algorithm where only users with channel gain values that exist in a predefined interval can transmit in a certain slot. The interval upper and lower values should be computed in a way that promises, with very high probability, that there exist exactly $k$ users in it and that these users are a good choice for the all one coefficient vector. To analyze such a distributed scheme and compute the specific interval, one can emend the analysis from the proof of Theorem 6 to fit
these requirements. Alternately, one can use other statistical tools, such as extreme value theory and point process analysis, similar to other works that considered distributed threshold-based scheduling algorithms (e.g., [21], [30]). We note that the last schedule where the scheduler picks the user which completes the full rank of the decoding matrix can be replaced by a few more random schedules as done in the first $L-1$ steps.

## VI. Multiple Relays

In the general model of $M \geq 1$ relays and $L$ users, each relay $m$ sees a different channel vector $\mathbf{h}_{m}$ between itself and all the users. Hence, a certain schedule of users which is good for a certain relay may not be the right choice for the other relays. However, we show that, as $L$ grows, there exist $k$ users that are simultaneously good for all relays.

Recall the scheduling problem given in (18), where we wish to maximize $C_{S R}$. In each slot, $M$ coefficient vectors are added to the decoding matrix simultaneously. These vectors depend on the subset of users $\mathcal{S}_{k}$ which was scheduled, making the scheduling problem very hard to solve and analyze. It requires taking into consideration all the $M$ channel vectors a specific schedule compels and find the appropriate coefficient vectors while ensuring that the maximal DoF from these vectors is obtained. In addition, the rates must also satisfy (8), so that the decoding matrix can be solved. Nevertheless, we show that if one employs a scheduling scheme which follows Algorithm 1 guidelines, good schedules can be found in each slot. Moreover, assuming the number of relays $M=O(1)$, and using heuristics solutions for a fast completion time, simulations show that indeed a pre-log gain of $M$ to the expected system sum-rate can be attained. We conclude the discussion by showing why this is possible for small $M$, yet fails when $M$ is too large.

The suggested scheme searches for $k$ users, all having channel gains satisfying

$$
\begin{equation*}
u \leq\left|h_{m}\right| \leq u+\delta, \quad 1 \leq m \leq M \tag{51}
\end{equation*}
$$

We then use coefficient vectors from $\mathbf{a}^{\{1, \mathbf{k}\}}$. We do not present the actual scheme to find these simultaneously good $k$ users, however, Lemma 7 bellow asserts that such users can indeed be found ${ }^{5}$

[^4]We start with the following corollary, which states that one can always attain the performance of a single relay system.

Corollary 5. The asymptotic expected system sum-rate for a general system consisting $M \geq 1$ relays is lower bounded by the following,

$$
\begin{equation*}
\mathbb{E}\left[C_{S R}^{M}\right] \geq \mathbb{E}\left[C_{S R}\right] \tag{52}
\end{equation*}
$$

The above follows immediately if one employs Algorithm 1 on the general system while the scheduling decision is taken according to a single leading relay. Any contribution by the other relays can only reduce the completion time by adding innovative coefficient vectors to A. Note that the choice of the leading relay can be optimized.

Remark 3 (Scheduling according to a single relay). Since the scheduled users are expected to transmit with a relatively high rate, which is tailored to the channel vector the leading relay sees, relays $2, \ldots, M$ are not expected to contribute much. A possible adaptation that will enable the other relays to contribute is to employ superposition coding. In superposition coding, one can divide a message into several messages, each belonging to a different level. By doing so, the leading relay, which can decode equations with a high rate, will be able to decode all parts. while the other relays, which have a lower rate, would be able to decode only messages belonging to certain levels. In CF this may be done by superimposing lattice codes that are scaled according to the power constraints. Further explanation and results can be found in [1], [39].

To show that, in fact, good schedules can be found for all relays simultaneously, return to the analysis of the scheme suggested in Section V. Since the channels between the relays and the users are i.i.d., the probability of finding at least $k$ users which are good simultaneously for all relays is $P_{r}(\xi)^{M}$. This probability goes to 1 as $L$ grows with the same choice of $u(L)$ and $\delta(L)$ given in Theorem 6. Specifically, we have the following.

Lemma 7. The probability of finding $k$ users in each transmission slot, such that their channel gains for all relays are in $[-(u+\delta),-u] \cup[u, u+\delta]$, tends to 1 as $L$ grows. That is,

$$
\begin{equation*}
\lim _{L \rightarrow \infty} P_{r}(\xi)^{M}=1 \tag{53}
\end{equation*}
$$



Figure 9: The average rank of $\mathbf{A}_{n}$ as a function of the transmission slots $n$ for different numbers of relays $M=\{1,2,3,4\}$ and for $L=150$ transmitters.

Proof: A simple extension of (44) results in

$$
\begin{equation*}
\lim _{L \rightarrow \infty} P_{r}(\xi)^{M}=\lim _{L \rightarrow \infty}\left(\sum_{i=k}^{L}\binom{L}{i} p(u, \delta)^{i}(1-p(u, \delta))^{L-i}\right)^{L M}=1 \tag{54}
\end{equation*}
$$

We emphasize that the probability tends to one even for $M=O(L)$.
Lemma 7 promises that asymptotically with $L$, a favourable group of users exists, and therefore the expected achievable rate, in each slot and at each relay, can be lower bounded by the result in Theorem 6. This implies that the rate of each linear combination scales as $O(\log \log L)$ which is the optimal scaling law (Theorem 7).

On the other hand, the crux of the suggested scheme lies in the following. Since there are $M$ relays, in each slot $n, M$ coefficient vectors are added to $\mathbf{A}_{n}$. Denote by $\mathbf{A}_{n}^{\prime}$ the sub-matrix added to $\mathbf{A}_{n}$ at slot $n$. The rows of $\mathbf{A}_{n}^{\prime}$ are $\left(\mathbf{a}_{m}\left(\mathcal{S}_{k}(n)\right)\right)_{\Uparrow}^{L}$ for $m=1, \ldots, M$, respectively. That is, $\mathbf{A}_{n}=\left(\mathbf{A}_{1}^{\prime}, \mathbf{A}_{2}^{\prime}, \ldots, \mathbf{A}_{n}^{\prime}\right)$.

On average, the rank of each such $\mathbf{A}_{n}^{\prime}$ is at least $\min \{M, k\}-2$. This follows from a result given in [40, Theorem 4] which states that the expected number of random vectors that are needed to be drawn uniformly from $\mathbb{F}_{2}^{k}$ to attain $k$ DoF is at most $k+2$. Since the rows of $\mathbf{A}_{n}^{\prime}$ were chosen uniformly from $\mathbf{a}^{\{1, \mathbf{k}\}}$ before the expansion $(\cdot)_{\Uparrow}^{L}$, this result applies. However, the average rank of $\mathbf{A}_{n}$ does not grow linearly at rate $M$ with $n$ since, in each slot $n$, the relays choose the coefficient vectors with respect to $\mathcal{S}_{k}(n)$. Thus, every $M$ consecutive rows in the matrix $\mathbf{A}_{n}$ have the same $L-k$ columns with zeroes. That is, although at each slot at least $\min \{M, k\}-2$ independent vectors are added to $\mathbf{A}_{n}$, at a certain slot, the vectors of $\mathbf{A}_{n}^{\prime}$ can be linearly dependent with those collected thus far in other $\mathbf{A}_{j}^{\prime}$ for $j<n$. This behavior is attributed
to the "curse" of the coupon collector problem and is depicted in Figure 9 Specifically, Figure 9 plots the average rank of $\mathbf{A}_{n}$ as a function of $n$ for different values of $M$ and $L=150$. One can observe that the average rank of $\mathbf{A}_{n}$ start growing linearly with the straight line $f(n)=M n$. That is, at the beginning, in each slot $M$ DoF are added to $\mathbf{A}_{n}$. However, at a certain point, it becomes harder to find new innovative coefficient vectors, and the curves start to flatten. Fortunately, for $M \ll L$, the curves continue at the maximal rate almost up to a full rank. As $M$ grows, the curves flattens earlier.

Following the scheduling scheme suggested in Section V, the scheduler can thus randomly schedule users and collect the coefficient vectors up until recognizing a substantial decrease in the rate of the rank of $\mathbf{A}_{n}$. At this point, the scheduler lets the users which their messages complete a full rank of $\mathbf{A}_{n}$, transmit separately (similar to the final SISO step in Algorithm 1). As Figure 9 depicts, the number of such slots is small when $M=O(1)$. Thus, the completion time of the transmission of all messages is roughly $N=\frac{L}{M-o(1)}=\frac{L}{M}+o(L)$ which results with a pre-log gain to the sum-rate. That is,

$$
\begin{aligned}
C_{S R}^{M} & =\frac{R(L-o(L))+R^{S I S O} o(L)}{N} \\
& =M \frac{R(L-o(L))+R^{S I S O} o(L)}{L+o(L)} \\
& \geq \frac{L-o(L)}{L+o(L)} M R=M R(1-o(1))
\end{aligned}
$$

Note that the transmission rate $R$ is set such that a relay will be able to decode successfully any linear combination when $L$ is large. That is, a rate that will satisfy zero outage probability at the limit of large $L$. Thus, the probability for outage as given in (38) should be modified to the following,

$$
\begin{equation*}
P_{\text {out }}(R)=P_{r}\left(\min _{m: 1, \ldots, M}\left\{\min _{n: 1, \ldots, L}\left\{\mathcal{R}\left(\mathbf{h}_{m}^{*}(n), \mathbf{a}_{m}^{*}(n)\right)\right\}\right\}<R\right) . \tag{55}
\end{equation*}
$$

Where setting $R=\left(\frac{1}{4}-\epsilon\right) \log \log L$ for any small positive $\epsilon$ will satisfy

$$
\begin{equation*}
\lim _{L \rightarrow \infty} P_{\text {out }}(R)=0 \tag{56}
\end{equation*}
$$

This can be proved if one follows the proof's steps of Lemma 6 while remembering that the values of $u$ and $\delta$ were chosen such that in each time slot a simultaneously good $k$ users for all the relays can be found with probability that goes to one with $L$. Specifically, one can use the lower bounds given in (43) in the achievable rate of each relay and thus remove the minimum on $m$.

## VII. Conclusions

The CF coding scheme provides a well understood framework for handling multiple transmissions, and decode them as linear combinations of messages using lattice codes. Accordingly, this enables a good utilization of the shared medium and can be employed in various communication systems. In this work, we have studied the impact of the number of transmitters in a CF system which was not addressed up until today. Specifically, when the receiver decodes linear combinations of all transmitted messages, the number of transmitters heavily affects its ability to provide high computation rates. The analysis and results presented in the first part of this paper provide a good understanding on this effect and show that as the number of transmitters grows, CF becomes degenerated, in the sense that the receiver prefers to decode a single message instead of a non-trivial linear combination. As a direct outcome, the computation rate and the system's sum-rate tend to zero since all other users are treated as noise. Thus, one is forced to restrict the number of transmitting users, i.e., use scheduling, in order to maintain the superior abilities CF provides.

In the second part of this work, we formulated the scheduling problem for a large scale CF system and presented a scheduling paradigm. The paradigm is based on the fact that one can always (with very high probability) find users that, if scheduled together, achieve the scaling law of the optimal computation rate. Moreover, the search of this simultaneously good users can be done in polynomial time considering the properties of what is considered as a good linear combination for CF. The analysis and results used in this work can be further applied to communication problems for which there is freedom of choosing the transmitters and thus the decoded linear combinations.

## Appendix A <br> Proof for Theorem 5

Define,

$$
\begin{align*}
& H_{e}=\left\{\mathbf{h} \in \mathbb{R}^{L} \mid \underset{\mathbf{a} \in \mathbb{Z}^{L} \backslash\{\mathbf{0}\}}{\arg \min } f(\mathbf{a})=\mathbf{e}_{i}, \text { for some } i\right\}  \tag{57}\\
& H_{\bar{e}}=\left\{\mathbf{h} \in \mathbb{R}^{L} \underset{\mathbf{a} \in \mathbb{Z}^{L} \backslash\{\mathbf{0}\}}{\arg \min } f(\mathbf{a}) \neq \mathbf{e}_{i}, \text { for all } i\right\} .
\end{align*}
$$

That is, with probability $P_{r}(\mathbf{e})$ a relay sees a channel vector $\mathbf{h} \in H_{e}$ and with probability $P_{r}(\overline{\mathbf{e}})$ a relay sees a channel vector $\mathbf{h} \in H_{\bar{e}}$. Accordingly, the complementary CDF of the achievable
rate can be expressed as,

$$
\begin{equation*}
P_{r}(\mathcal{R}(\mathbf{h}, \mathbf{a})>\epsilon)=P_{r}(\mathbf{e}) P_{r}\left(\mathcal{R}(\mathbf{h}, \mathbf{e})>\epsilon \mid \mathbf{h} \in H_{e}\right)+P_{r}(\overline{\mathbf{e}}) P_{r}\left(\mathcal{R}(\mathbf{h}, \mathbf{a})>\epsilon \mid \mathbf{h} \in H_{\bar{e}}\right) . \tag{58}
\end{equation*}
$$

We treat the two terms above separately. The first term describes the case where the maximizing coefficient vector is some unit vector, while the second term describes the case where the maximizing coefficient vector may be any integer vector excluding the unit vectors. Note that we write $\mathbf{e}$ to express the event that some unit vector $\mathbf{e}_{i}$ was chosen. We show that both terms tends to zero. Starting with the second term,

$$
\begin{align*}
P_{r}(\overline{\mathbf{e}}) P_{r}\left(\mathcal{R}(\mathbf{h}, \mathbf{a})>\epsilon \mid \mathbf{h} \in H_{\bar{e}}\right) & =P_{r}(\overline{\mathbf{e}}) P_{r}\left(\left.\frac{1}{2} \log ^{+}\left(\frac{1+\mathbf{P}\|\mathbf{h}\|^{2}}{\|\mathbf{a}\|^{2}+\mathbf{P}\left(\|\mathbf{a}\|^{2}\|\mathbf{h}\|^{2}-\left(\mathbf{h}^{T} \mathbf{a}\right)^{2}\right)}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{\bar{e}}\right)  \tag{59}\\
& \leq P_{r}(\overline{\mathbf{e}}) P_{r}\left(\left.\frac{1}{2} \log ^{+}\left(1+\mathrm{P}\|\mathbf{h}\|^{2}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{\bar{e}}\right),
\end{align*}
$$

where the inequality is due to Cauchy-Schwarz and since $\|\mathbf{a}\|^{2}>1$.
We upper bound the above using the Markov and Jensen's inequalities as follows,

$$
\begin{align*}
P_{r}(\overline{\mathbf{e}}) P_{r}\left(\left.\frac{1}{2} \log ^{+}\left(1+\mathrm{P}\|\mathbf{h}\|^{2}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{\bar{e}}\right) & \leq P_{r}(\overline{\mathbf{e}}) \frac{1}{\epsilon} \mathbb{E}\left[\left.\frac{1}{2} \log ^{+}\left(1+\mathrm{P}\|\mathbf{h}\|^{2}\right) \right\rvert\, \mathbf{h} \in H_{\bar{e}}\right]  \tag{60}\\
& \leq P_{r}(\overline{\mathbf{e}}) \frac{1}{2 \epsilon} \log ^{+}\left(1+\mathrm{P} \mathbb{E}\left[\|\mathbf{h}\|^{2} \mid \mathbf{h} \in H_{\bar{e}}\right]\right) .
\end{align*}
$$

To further upper bound the above note that

$$
\begin{align*}
\mathbb{E}\left[\|\mathbf{h}\|^{2}\right] & =\mathbb{E}\left[\|\mathbf{h}\|^{2} \mid \mathbf{h} \in H_{\bar{e}}\right] P_{r}(\overline{\mathbf{e}})+\mathbb{E}\left[\|\mathbf{h}\|^{2} \mid \mathbf{h} \in H_{e}\right] P_{r}(\mathbf{e})  \tag{61}\\
& \geq \mathbb{E}\left[\|\mathbf{h}\|^{2} \mid \mathbf{h} \in H_{\bar{e}}\right] P_{r}(\overline{\mathbf{e}}),
\end{align*}
$$

which results with

$$
\begin{equation*}
\mathbb{E}\left[\|\mathbf{h}\|^{2} \mid \mathbf{h} \in H_{\bar{e}}\right] \leq \frac{L}{P_{r}(\overline{\mathbf{e}})} \tag{62}
\end{equation*}
$$

since the channel vector $\mathbf{h}$ is a Gaussian random vector and its squared norm follows the $\chi_{L}^{2}$ distribution.

Applying the expectation's upper bound in (60) we have,

$$
\begin{align*}
P_{r}(\overline{\mathbf{e}}) \frac{1}{2 \epsilon} \log ^{+}\left(1+\mathrm{P} \mathbb{E}\left[\left\|\mathbf{h}^{\bar{e}}\right\|^{2} \mid \overline{\mathbf{e}}\right]\right) & \leq P_{r}(\overline{\mathbf{e}}) \frac{1}{2 \epsilon} \log ^{+}\left(1+\frac{\mathrm{P} L}{P_{r}(\overline{\mathbf{e}})}\right) \\
& \stackrel{(a)}{\leq} P_{r}(\overline{\mathbf{e}}) \frac{1}{2 \epsilon} \sqrt{2 \frac{\mathrm{P} L}{P_{r}(\overline{\mathbf{e}})}} \\
& =\frac{1}{2 \epsilon} \sqrt{2 \mathrm{P} L P_{r}(\overline{\mathbf{e}})}  \tag{63}\\
& \stackrel{(b)}{\leq} \frac{1}{2 \epsilon} \sqrt{8 \mathrm{P}^{2} L^{3} e^{-L E_{2}(L)}} \\
& =\frac{1}{\epsilon} \mathrm{P} L \sqrt{2 L} e^{-L E_{3}(L)},
\end{align*}
$$

where $(a)$ is due to $\log (1+x) \leq \sqrt{2 x},(b)$ follows from Theorem 4 and $E_{3}(L)=\frac{1}{4}\left(1-\frac{1}{L}\right) \log 2$. Considering the above, as $L$ grows, the second summand in (58) tends to zero for all $\epsilon>0$.

We are left with the first term in (58) which can be upper bounded as follows

$$
\begin{align*}
P_{r}(\mathbf{e}) P_{r}\left(\mathcal{R}(\mathbf{h}, \mathbf{e})>\epsilon \mid \mathbf{h} \in H_{e}\right) & \leq P_{r}\left(\mathcal{R}(\mathbf{h}, \mathbf{e})>\epsilon \mid \mathbf{h} \in H_{e}\right) \\
& \stackrel{(a)}{=} P_{r}\left(\left.\frac{1}{2} \log ^{+}\left(\frac{1+\mathrm{P}\|\mathbf{h}\|^{2}}{1+\mathbf{P}\left(\|\mathbf{h}\|^{2}-h_{i}^{2}\right)}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{e}\right) \\
& =P_{r}\left(\left.\frac{1}{2} \log ^{+}\left(\frac{1}{1-\frac{\mathrm{P} h_{i}^{2}}{1+\mathrm{P}\|\mathbf{h}\|^{2}}}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{e}\right) \\
& \leq P_{r}\left(\left.\frac{1}{2} \log ^{+}\left(\frac{1}{1-\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}}}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{e}\right)  \tag{64}\\
& \stackrel{(b)}{=} P_{r}\left(\left.\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}}>1-\frac{1}{2^{2 \epsilon}} \right\rvert\, \mathbf{h} \in H_{e}\right) \\
& \stackrel{(c)}{=} \frac{1}{\epsilon^{\prime}} \mathbb{E}\left[\left.\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}} \right\rvert\, \mathbf{h} \in H_{e}\right] .
\end{align*}
$$

where in (a) we set the specific unit vector $\mathbf{e}_{i}$ the relay has chosen for the channel vector $\mathbf{h} \in H_{e}$. Note that the unit vector that maximizes the rate corresponds to the strongest transmitter, i.e., given $\mathbf{h}$, the relay chooses $i=\arg \max _{i} \mathbf{h}$. In (b), since the argument of the $\log$ is greater than one $\log ^{+}(\cdot)=\log (\cdot) .(c)$ follows from Markov's inequality where we denote $\epsilon^{\prime}=1-\frac{1}{2^{2 \epsilon}}$.

To further upper bound the above note that

$$
\begin{align*}
\mathbb{E}\left[\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}}\right] & =\mathbb{E}\left[\left.\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}} \right\rvert\, \mathbf{h} \in H_{\bar{e}}\right] P_{r}(\overline{\mathbf{e}}) \mathbb{E}\left[\left.\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}} \right\rvert\, \mathbf{h} \in H_{e}\right] P_{r}(\mathbf{e})  \tag{65}\\
& \geq \mathbb{E}\left[\left.\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}} \right\rvert\, \mathbf{h} \in H_{e}\right] P_{r}(\mathbf{e}),
\end{align*}
$$

which results with

$$
\begin{equation*}
\mathbb{E}\left[\left.\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}} \right\rvert\, \mathbf{h} \in H_{e}\right] \leq \frac{1}{P_{r}(\mathbf{e})} \mathbb{E}\left[\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}}\right] \tag{66}
\end{equation*}
$$

Theorem 4 implies that $P_{r}(\mathbf{e}) \rightarrow 1$ as $L \rightarrow \infty$. In addition the numerator scales like $O(\ln L)$ as was shown in the proof of Theorem 7 and the denominator scales as $O(L)$ since $\|\mathbf{h}\|^{2}$ follows the chi-squared distribution with $L$ DoF. Therefore, it is clear that as $L$ grows the first summand in (58) tends to zero for all $\epsilon^{\prime}>0$. For an exact bound recall that

$$
\mathbb{E}\left[h_{i}^{2} \frac{1}{\|\mathbf{h}\|^{2}}\right]=\mathbb{E}\left[h_{i}^{2}\right] \mathbb{E}\left[\frac{1}{\|\mathbf{h}\|^{2}}\right]+\operatorname{cov}\left(h_{i}^{2}, \frac{1}{\|\mathbf{h}\|^{2}}\right) .
$$

Since the covariance is negative we have

$$
\begin{align*}
\mathbb{E}\left[\frac{h_{i}^{2}}{\|\mathbf{h}\|^{2}}\right] & \leq \mathbb{E}\left[\left(h_{i}\right)^{2}\right] \mathbb{E}\left[\frac{1}{\|\mathbf{h}\|^{2}}\right]  \tag{67}\\
& \leq\left(2 \ln L+\frac{\gamma}{2}+o(1)\right) \frac{1}{L-2}
\end{align*}
$$

where in the last line the expectations are for the maximum of a chi-squared r.v. [38, Table 3.4.4] and an inverse chi-squared with $L$ DoF r.v., respectively. As $L$ grows, the above tends to zero.

## Appendix B

## Proof for Corollary 2

Considering Equation (58) in Appendix A, we can upper bound $P_{r}\left(C_{S R}>\epsilon\right)$ as follows,

$$
\begin{align*}
P_{r}\left(C_{S R}>\epsilon\right) & \triangleq P_{r}\left(\frac{R L}{N}>\epsilon\right) \\
& \leq P_{r}\left(\frac{L}{N} \min _{n=1, \ldots, N} \min _{m} \mathcal{R}\left(\mathbf{h}_{m}(n), \mathbf{a}_{m}(n)\right)>\epsilon\right) \\
& \leq P_{r}\left(\frac{L}{N} \mathcal{R}\left(\mathbf{h}_{m}(n), \mathbf{a}_{m}(n)\right)>\epsilon\right) \\
& =P_{r}(\mathbf{e}) P_{r}\left(\left.\frac{L}{N} \mathcal{R}\left(\mathbf{h}_{m}, \mathbf{e}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{e}\right)+P_{r}(\overline{\mathbf{e}}) P_{r}\left(\left.\frac{L}{N} \mathcal{R}\left(\mathbf{h}_{m}, \mathbf{a}_{m}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{\bar{e}}\right) . \tag{68}
\end{align*}
$$

Following the same steps as the proof for Theorem 5 and remembering that $\frac{L}{N}$ can be upper bounded by $M$, we have

$$
\begin{equation*}
\lim _{L \rightarrow \infty} P_{r}(\overline{\mathbf{e}}) P_{r}\left(\left.\frac{L}{N} \mathcal{R}\left(\mathbf{h}_{m}, \mathbf{a}_{m}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{\bar{e}}\right) \leq \lim _{L \rightarrow \infty} \frac{1 \mathrm{P} M L \sqrt{L}}{\epsilon} \frac{\sqrt{0.5}}{\sqrt{-L E_{3}(L)}}=0 \tag{69}
\end{equation*}
$$

for all $\epsilon>0$. Similarly, the first term is upper bounded by

$$
\begin{align*}
\lim _{L \rightarrow \infty} P_{r}(\mathbf{e}) P_{r}\left(\left.\frac{L}{N} \mathcal{R}\left(\mathbf{h}_{m}, \mathbf{e}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{e}\right) & \leq \lim _{L \rightarrow \infty} P_{r}\left(\left.\frac{1}{2} M \log ^{+}\left(\frac{1}{1-\frac{\left(h_{i}\right)^{2}}{\|\mathbf{h}\|^{2}}}\right)>\epsilon \right\rvert\, \mathbf{h} \in H_{e}\right) \\
& =0 \tag{70}
\end{align*}
$$

provided that $M$ is fixed.

## Appendix C

## Proof For LEMMA 6

$$
\begin{aligned}
& \lim _{L \rightarrow \infty} P_{\text {out }}(R)=\lim _{L \rightarrow \infty} P_{r}\left(\min _{n}\left\{\mathcal{R}\left(\mathbf{h}^{*}(n), \mathbf{a}^{*}(n)\right)\right\}<\left(\frac{1}{4}-\epsilon\right) \log \log L\right) \\
& =\lim _{L \rightarrow \infty} P_{r}\left(\min _{n}\left\{\frac{1}{2} \log ^{+}\left(k-\frac{\mathbf{P}\left(\mathbf{h}^{*}(n)^{T} \mathbf{1}\right)^{2}}{1+\mathbf{P}\left\|\mathbf{h}^{*}(n)\right\|^{2}}\right)^{-1}\right\}<\left(\frac{1}{4}-\epsilon\right) \log \log L\right) \\
& \stackrel{(a)}{\leq} \lim _{L \rightarrow \infty} P_{r}\left(\frac{1}{2} \log \left(k-\min _{n}\left\{\frac{\mathrm{P}\left(\mathbf{h}^{*}(n)^{T} \mathbf{1}\right)^{2}}{1+\mathbf{P}\left\|\mathbf{h}^{*}(n)\right\|^{2}}\right\}\right)^{-1}<\left(\frac{1}{4}-\epsilon\right) \log \log L\right) \\
& =\lim _{L \rightarrow \infty} P_{r}\left(\left(k-\min _{n}\left\{\frac{\mathrm{P}\left(\mathbf{h}^{*}(n)^{T} \mathbf{1}\right)^{2}}{1+\mathbf{P}\left\|\mathbf{h}^{*}(n)\right\|^{2}}\right\}\right)>(\log L)^{-\left(\frac{1}{2}-2 \epsilon\right)}\right) \\
& \stackrel{(b)}{\leq} \lim _{L \rightarrow \infty} \frac{\mathbb{E}\left[k-\min _{n}\left\{\frac{\mathrm{P}\left(\mathbf{h}^{*}(n)^{T} \mathbf{1}\right)^{2}}{1+\mathrm{P}\left\|\mathbf{h}^{*}(n)\right\|^{2}}\right\}\right]}{(\log L)^{-\left(\frac{1}{2}-2 \epsilon\right)}} \\
& =\lim _{L \rightarrow \infty} k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}\left(1-\mathbb{E}\left[\min _{n}\left\{\frac{\mathrm{P}\left\|\mathbf{h}^{*}(\mathbf{n})\right\|^{2} \cos ^{2}(\theta(n))}{1+\mathbf{P}\left\|\mathbf{h}^{*}(n)\right\|^{2}}\right\}\right]\right) \\
& \leq \lim _{L \rightarrow \infty} k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}\left(1-\mathbb{E}\left[\left.\min _{n}\left\{\frac{\cos ^{2}(\theta(n))}{\frac{1}{\mathrm{P}\left\|\mathbf{h}^{*}(\mathbf{n})\right\|^{2}}+1}\right\} \right\rvert\, \xi\right] P_{r}(\xi)\right) \\
& \stackrel{(c)}{\leq} \lim _{L \rightarrow \infty} k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}\left(1-\mathbb{E}\left[\left.\min _{n}\left\{\frac{\frac{u^{2}}{(u+\delta)^{2}}}{\frac{1}{\mathrm{P} k u^{2}}+1}\right\} \right\rvert\, \xi\right] P_{r}(\xi)\right) \\
& =\lim _{L \rightarrow \infty} k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}\left(1-\frac{\frac{u^{2}}{(u+\delta)^{2}}}{\frac{1}{\mathrm{P} k u^{2}}+1} P_{r}(\xi)\right) \\
& \leq \lim _{L \rightarrow \infty} k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}\left(\frac{1+\mathrm{P} k u^{2}-\frac{u^{2} P_{r}(\xi)}{(u+\delta)^{2}} \mathrm{P} k u^{2}}{\mathrm{P} k u^{2}}\right) \\
& =\lim _{L \rightarrow \infty} k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}\left(\frac{1}{\mathrm{P} k u^{2}}+1-\frac{u^{2} P_{r}(\xi)}{(u+\delta)^{2}}\right) \\
& =\lim _{L \rightarrow \infty} \frac{(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}}{\mathrm{P} u^{2}}+\lim _{L \rightarrow \infty} k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}\left(1-\frac{u^{2} P_{r}(\xi)}{(u+\delta)^{2}}\right) \\
& \stackrel{(d)}{=} \lim _{L \rightarrow \infty} k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}\left(1-\frac{u^{2} P_{r}(\xi)}{(u+\delta)^{2}}\right) \\
& =\lim _{L \rightarrow \infty} \frac{k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}}{(u+\delta)^{2}}\left(u^{2}\left(1-P_{r}(\xi)\right)+2 u \delta+\delta^{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
& \stackrel{(e)}{=} \lim _{L \rightarrow \infty} \frac{k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}}{(u+\delta)^{2}} \cdot \lim _{L \rightarrow \infty}\left(u^{2}\left(1-P_{r}(\xi)\right)\right) \\
& \stackrel{(f)}{\leq} \lim _{L \rightarrow \infty} \frac{k(\log L)^{\left(\frac{1}{2}-2 \epsilon\right)}}{(u+\delta)^{2}} \lim _{L \rightarrow \infty}\left(u^{2}\left(1-\left(1-L e^{-\sqrt{L}+(\ln (L)+1)}\right)\right)\right) \\
& \stackrel{(g)}{\leq} 0 \cdot \lim _{L \rightarrow \infty}\left(u^{2} L e^{-\sqrt{L}+(\ln (L)+1)}\right) \\
& =0
\end{aligned}
$$

(a) follows since $\log ^{+}(x) \geq \log (x)$. In (b) we used the Markov inequality and in (c) we bound the norm and the cosine as was done in the proof of Theorem 6. Note that the values of $u$ and $\delta$ were chosen to fit all slots and in particular to the slot with the minimal achievable rate. (d), $(e)$ and $(g)$ follows since $u=O(\sqrt{\log L})$ and $\delta=O\left(\frac{1}{\log L}\right)$. In ( $f$ ) we used 47).

## Appendix D

## Proof for the scaling laws of Theorem 6

To prove that the scaling law is $\frac{1}{4} \log \log L$, we show that the limit of the division of the lower bound with $\frac{1}{4} \log \log L$ equals 1 as follows,

$$
\begin{equation*}
\lim _{L \rightarrow \infty} \frac{\frac{1}{2} \log ^{+}\left(k\left(1-\frac{\mathrm{P} k u^{4}}{(u+\delta)^{2}\left(1+\mathrm{P} k u^{2}\right)}(1-g(L))^{L}\right)\right)^{-1}}{\frac{1}{4} \log \log L}=\lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(1-\frac{\mathrm{P} k u^{4}(1-g(L))^{L}}{(u+\delta)^{2}\left(1+\mathrm{P} k u^{2}\right)}\right)\right)}{\log \log L} \tag{71}
\end{equation*}
$$

where $g(L)=e^{-\frac{1}{2 p(u, \delta)} \frac{(L \rho p(u, \delta)-(k-1))^{2}}{L}}$ which we expressed as $o(1)$ in the theorem. We start with an upper bound on the ratio.

$$
\begin{aligned}
\lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(1-\frac{\mathrm{P} k u^{4}(1-g(L) L}{(u+\delta)^{2}\left(1+\mathrm{Pk} u^{2}\right)}\right)\right)}{\log \log L} & \leq \lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(1-\frac{u^{2}}{(u+\delta)^{2}}\right)\right)}{\log \log L} \\
& \leq \lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(\frac{\delta}{(u+\delta)}\right)\right)}{\log \log L} \\
& \stackrel{(a)}{=} \lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(\frac{3}{(u+\delta)}\right)}{\log \log L} \\
& \stackrel{(b)}{=} \lim _{L \rightarrow \infty} \frac{2 \log ^{+}\left(\sqrt{2 \ln \frac{\delta \sqrt{L}}{\sqrt{2 \pi}}}\right)}{\log \log L} \\
& =\lim _{L \rightarrow \infty} \frac{\log ^{+}(2 \ln \delta+\ln L-\ln 2 \pi)}{\log \log L} \\
& \leq \lim _{L \rightarrow \infty} \frac{\log ^{+}(\ln L)}{\log \log L}=1 .
\end{aligned}
$$

In $(a)$ we used our choice of $\delta=\frac{1}{\ln L}$ and $k=\lceil\ln L\rceil+1$ to upper bound $k \delta<3$. In $(b)$ we set $u+\delta=\sqrt{2 \ln \frac{\delta \sqrt{L}}{\sqrt{2 \pi}}}$.
The lower bound is as follows,

$$
\begin{aligned}
& \lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(1-\frac{\mathrm{P} k u^{4}(1-g(L))^{L}}{(u+\delta)^{2}\left(1+\mathrm{P} k u^{2}\right)}\right)\right)}{\log \log L} \\
& \geq \lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(1-\frac{k u^{4}(1-g(L))^{L}}{(u+\delta)^{2}\left(1+k u^{2}\right)}\right)\right)}{\log \log L} \\
& =\lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(1-\frac{k u^{4}(1-g(L))^{L}}{u^{2}+2 u \delta+\delta^{2}+k u^{4}+2 k u^{3} \delta+k u^{2} \delta^{2}}\right)\right)}{\log \log L} \\
& \stackrel{(c)}{\geq} \lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(1-\frac{k u^{4}(1-g(L))^{L}}{u^{2}+2 u^{2} \delta k+u^{2} \delta^{2} k+k u^{4}+2 k u^{3} \delta+k u^{2} \delta^{2}}\right)\right)}{\log \log L} \\
& =\lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(1-\frac{k u^{2}(1-g(L))^{L}}{1+2 k \delta(1+\delta+u)+k u^{2}}\right)\right)}{\log \log L} \\
& =\lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(k\left(\frac{1+2 k \delta(1+\delta+u)+k u^{2}\left(1-(1-g(L))^{L}\right)}{1+2 k \delta(1+\delta+u)+k u^{2}}\right)\right)}{\log \log L} \\
& \stackrel{(d)}{\geq} \lim _{L \rightarrow \infty} \frac{-2 \log ^{+}\left(\frac{7(1+\delta+u)+k u^{2}\left(1-(1-g(L))^{L}\right)}{u^{2}}\right)}{\log \log L} \\
& =\lim _{L \rightarrow \infty} \frac{2 \log ^{+}\left(u^{2}\right)}{\log \log L}-\lim _{L \rightarrow \infty} \frac{2 \log ^{+}\left(7(1+\delta+u)+k u^{2}\left(1-(1-g(L))^{L}\right)\right)}{\log \log L} \\
& \stackrel{(e)}{=} 2-\lim _{L \rightarrow \infty} \frac{2 \log ^{+}\left(7(1+\delta+u)+k u^{2}\left(1-(1-g(L))^{L}\right)\right)}{\log \log L} \\
& =2-\lim _{L \rightarrow \infty} \frac{2 \log ^{+}(7(1+\delta+u))}{\log \log L}-\lim _{L \rightarrow \infty} \frac{2 \log ^{+}\left(1+\frac{k u^{2}}{7(1+\delta+u)}\left(1-(1-g(L))^{L}\right)\right)}{\log \log L} \\
& \stackrel{(f)}{\geq} 2-1-\lim _{L \rightarrow \infty} \frac{\frac{2 k u^{2}}{7(1+\delta+u)}\left(1-(1-g(L))^{L}\right)}{\log \log L} \\
& \geq 1-\lim _{L \rightarrow \infty} \frac{\frac{2 k(u+\delta)^{2}}{7(\delta+u)}\left(1-(1-g(L))^{L}\right)}{\log \log L} \\
& \geq 1-\lim _{L \rightarrow \infty} \frac{k(u+\delta)\left(1-(1-g(L))^{L}\right)}{\log \log L} \\
& \geq 1-\lim _{L \rightarrow \infty} \frac{4(\ln L)^{2}\left(1-(1-g(L))^{L}\right)}{\log \log L} \\
& \geq 1-\lim _{L \rightarrow \infty} 4(\ln L)^{2}\left(1-(1-g(L))^{L}\right) \\
& \stackrel{(g)}{\geq} 1-\lim _{L \rightarrow \infty} 4(\ln L)^{2}\left(1-\left(1-L e^{-\sqrt{L}+(\ln (L)+1)}\right)\right) \\
& =1-\lim _{L \rightarrow \infty} 4(\ln L)^{2} L e^{-\sqrt{L}+(\ln (L)+1)} \\
& =1-0=1
\end{aligned}
$$

(c) follows since, for large enough $L, u^{2} k>1$ and $u k>1$. In $(d)$, we used again the bound $k \delta<3$, the fact that $\delta>0$ and that for large enough $L, u>0$. (e) follows from the following,

$$
\begin{aligned}
\lim _{L \rightarrow \infty} \frac{2 \log ^{+}\left(u^{2}\right)}{\log \log L} & =\lim _{L \rightarrow \infty} \frac{4 \log ^{+}\left(\sqrt{2 \ln \frac{\delta \sqrt{L}}{\sqrt{2 \pi}}}-\delta\right)}{\log \log L} \\
& =\lim _{L \rightarrow \infty} \frac{4 \log ^{+}(\sqrt{\ln L+2 \ln \delta-\ln 2 \pi}-\delta)}{\log \log L} \\
& =2 .
\end{aligned}
$$

In $(f)$, the third term can be upper bounded using the relation of $\log (1+x) \leq x$ for $x>0$ and the second term follows from the following,

$$
\begin{aligned}
\lim _{L \rightarrow \infty} \frac{2 \log ^{+}(7(1+\delta+u))}{\log \log L} & =\lim _{L \rightarrow \infty} \frac{2 \log ^{+}(1+\delta+u)}{\log \log L} \\
& =\lim _{L \rightarrow \infty} \frac{2 \log ^{+}\left(1+\sqrt{2 \ln \frac{\delta \sqrt{L}}{\sqrt{2 \pi}}}\right)}{\log \log L} \\
& =1
\end{aligned}
$$

Finally, in ( $g$ ) we used equation (47). The correctness of these limits was verified also on Mathematica.

## Appendix E

## Proof for the scaling laws of Theorem 7

To prove that the scaling law is $\frac{1}{2} \log \log L$, we show that the limit of the division of the lower bound with $\frac{1}{2} \log \log L$ equals 1 as follows,

$$
\begin{equation*}
\lim _{L \rightarrow \infty} \frac{\frac{1}{2} \log \left(1+\mathrm{P}\left(2 \ln L-\ln \ln L-2 \ln \Gamma\left(\frac{1}{2}\right)+\frac{\gamma}{2}\right)\right)}{\frac{1}{2} \log \log L} . \tag{72}
\end{equation*}
$$

We start with an upper bound on the ratio.

$$
\begin{align*}
\lim _{L \rightarrow \infty} \frac{\log \left(1+\mathrm{P}\left(2 \ln L-\ln \ln L-2 \ln \Gamma\left(\frac{1}{2}\right)+\frac{\gamma}{2}\right)\right)}{\log \log L} & \leq \lim _{L \rightarrow \infty} \frac{\log \left(\mathrm{P}\left(1+2 \ln L-\ln \ln L-2 \ln \Gamma\left(\frac{1}{2}\right)+\frac{\gamma}{2}\right)\right)}{\log \log L} \\
& \leq \lim _{L \rightarrow \infty} \frac{\log \left(1+2 \ln L+\frac{\gamma}{2}\right)}{\log \log L}=1 . \tag{73}
\end{align*}
$$

The lower bound is as follows,

$$
\begin{align*}
\lim _{L \rightarrow \infty} \frac{\log \left(1+\mathrm{P}\left(2 \ln L-\ln \ln L-2 \ln \Gamma\left(\frac{1}{2}\right)+\frac{\gamma}{2}\right)\right)}{\log \log L} & \geq \lim _{L \rightarrow \infty} \frac{\log \left(2 \ln L-\ln \ln L-2 \ln \Gamma\left(\frac{1}{2}\right)+\frac{\gamma}{2}\right)}{\log \log L}  \tag{74}\\
& \geq \lim _{L \rightarrow \infty} \frac{\log \left(\ln \ln L-2 \ln \Gamma\left(\frac{1}{2}\right)\right)}{\log \log L}=1 .
\end{align*}
$$

## AcKNOWLEDGMENT

The authors would like to thank Or Ordentlich for his contribution in Lemma 4

## REFERENCES

[1] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," IEEE Transactions on Information Theory, vol. 57, no. 10, pp. 6463-6486, 2011.
[2] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," IEEE Transactions on Information Theory, vol. 60, no. 12, pp. 7661-7685, 2014.
[3] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric gaussian-user interference channel," IEEE Transactions on Information Theory, vol. 60, no. 6, pp. 3450-3482, 2014.
[4] S.-H. Park, O. Simeone, O. Sahin, and S. S. Shitz, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," IEEE Signal Processing Magazine, vol. 31, no. 6, pp. 69-79, 2014.
[5] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, "Semantically secure lattice codes for the gaussian wiretap channel," IEEE Transactions on Information Theory, vol. 60, no. 10, pp. 6399-6416, 2014.
[6] J. Zhan, B. Nazer, M. Gastpar, and U. Erez, "MIMO compute-and-forward," in 2009 IEEE International Symposium on Information Theory. IEEE, 2009, pp. 2848-2852.
[7] U. Niesen and P. Whiting, "The degrees of freedom of compute-and-forward," IEEE Transactions on Information Theory, vol. 58, no. 8, pp. 5214-5232, 2012.
[8] W. He, C. Feng, C. I. Ionita, and B. Nazer, "Collision scheduling for cellular networks," in Information Theory (ISIT), 2015 IEEE International Symposium on. IEEE, 2015, pp. 1417-1421.
[9] J. Zhu and M. Gastpar, "Gaussian multiple access via compute-and-forward," IEEE Transactions on Information Theory, vol. 63, no. 5, pp. 2678-2695, 2017.
[10] L. Wei and W. Chen, "Compute-and-forward network coding design over multi-source multi-relay channels," IEEE Transactions on Wireless Communications, vol. 11, no. 9, pp. 3348-3357, 2012.
[11] S.-N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," Information Theory, IEEE Transactions on, vol. 59, no. 9, pp. 5227-5243, 2013.
[12] S. H. Lim, C. Feng, A. Pastore, B. Nazer, and M. Gastpar, "A joint typicality approach to compute-forward," IEEE Transactions on Information Theory, 2018.
[13] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka et al., "Scenarios for 5 g mobile and wireless communications: the vision of the metis project," IEEE Communications Magazine, vol. 52, no. 5, pp. 26-35, 2014.
[14] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5g: solutions, challenges, opportunities, and future research trends," IEEE Communications Magazine, vol. 53, no. 9, pp. 74-81, 2015.
[15] P. Gupta and P. R. Kumar, "The capacity of wireless networks," IEEE Transactions on information theory, vol. 46, no. 2, pp. 388-404, 2000.
[16] P. Gupta and P. Kumar, "Towards an information theory of large networks: An achievable rate region," Information Theory, IEEE Transactions on, vol. 49, no. 8, pp. 1877-1894, 2003.
[17] M. Gastpar and M. Vetterli, "On the capacity of large gaussian relay networks," IEEE Transactions on Information Theory, vol. 51, no. 3, pp. 765-779, 2005.
[18] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in Communications, 1995. ICC'95 Seattle,'Gateway to Globalization', 1995 IEEE International Conference on, vol. 1. IEEE, 1995, pp. 331-335.
[19] X. Qin and R. Berry, "Exploiting multiuser diversity for medium access control in wireless networks," in INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, vol. 2. IEEE, 2003, pp. 1084-1094.
[20] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," IEEE Journal on Selected Areas in Communications, vol. 24, no. 3, pp. 528-541, 2006.
[21] O. Shmuel, A. Cohen, and O. Gurewitz, "Performance analysis of opportunistic distributed scheduling in multi-user systems," IEEE Transactions on Communications, 2018.
[22] D. Ramirez and B. Aazhang, "Scheduling for compute and forward networks," in 2015 49th Asilomar Conference on Signals, Systems and Computers. IEEE, 2015, pp. 57-58.
[23] B. Nazer, "Successive compute-and-forward," in 22th International Zurich Seminar on Communications (IZS). Eidgenössische Technische Hochschule Zürich, 2012.
[24] S. Sahraei and M. Gastpar, "Compute-and-forward: Finding the best equation," in Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on. IEEE, 2014, pp. 227-233.
[25] D. Dadush, C. Peikert, and S. Vempala, "Enumerative lattice algorithms in any norm via m-ellipsoid coverings," in Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on. IEEE, 2011, pp. 580-589.
[26] M. Alekhnovich, S. A. Khot, G. Kindler, and N. K. Vishnoi, "Hardness of approximating the closest vector problem with pre-processing," in 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05). IEEE, 2005, pp. 216-225.
[27] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," Mathematische Annalen, vol. 261, no. 4, pp. 515-534, 1982.
[28] N. Gama and P. Q. Nguyen, "Finding short lattice vectors within mordell's inequality," in Proceedings of the fortieth annual ACM symposium on Theory of computing. ACM, 2008, pp. 207-216.
[29] J. H. Conway and N. J. A. Sloane, Sphere packings, lattices and groups. Springer Science \& Business Media, 2013, vol. 290.
[30] J. Kampeas, A. Cohen, and O. Gurewitz, "Capacity of distributed opportunistic scheduling in nonhomogeneous networks," IEEE Transactions on Information Theory, vol. 60, no. 11, pp. 7231-7247, 2014.
[31] C. Walck, "Hand-book on statistical distributions for experimentalists," Tech. Rep., 1996.
[32] K. P. Jagannathan, S. Borst, P. Whiting, and E. Modiano, "Efficient scheduling of multi-user multi-antenna systems," in Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2006 4th International Symposium on. IEEE, 2006, pp. 1-8.
[33] J. Kampeas, A. Cohen, and O. Gurewitz, "The ergodic capacity of the multiple access channel under distributed schedulingorder optimality of linear receivers," IEEE Transactions on Information Theory, vol. 64, no. 8, pp. 5898-5919, 2018.
[34] D. Wiedemann, "Solving sparse linear equations over finite fields," IEEE transactions on information theory, vol. 32, no. 1, pp. 54-62, 1986.
[35] N. J. Calkin, "Dependent sets of constant weight binary vectors," Combinatorics, Probability and Computing, vol. 6, no. 3, pp. 263-271, 1997.
[36] A. Mazumdar, V. Chandar, and G. W. Wornell, "Update-efficiency and local repairability limits for capacity approaching codes," IEEE Journal on Selected Areas in Communications, vol. 32, no. 5, pp. 976-988, 2014.
[37] B. Nazer and O. Ordentlich, "Diophantine approximation for network information theory: A survey of old and new results," in Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on. IEEE, 2016, pp. 990-996.
[38] P. Embrechts, C. Klüppelberg, and T. Mikosch, Modelling extremal events: for insurance and finance. Springer Science \& Business Media, 2013, vol. 33.
[39] B. Nazer, A. Sanderovich, M. Gastpar, and S. Shamai, "Structured superposition for backhaul constrained cellular uplink," in Information Theory, 2009. ISIT 2009. IEEE International Symposium on. IEEE, 2009, pp. 1530-1534.
[40] D. E. Lucani, M. Médard, and M. Stojanovic, "Random linear network coding for time-division duplexing: Field size considerations," in Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE. IEEE, 2009, pp. 1-6.


[^0]:    This research was partially supported by European Unions Horizon 2020 Research and Innovation Program SUPERFLUIDITY, Grant Agreement 671566 and by MAFAT. Parts of this work appeared at the 2017 IEEE Information Theory Workshop (ITW) and at the 2018 IEEE Information Theory Workshop (ITW). The authors are from Ben-Gurion University of the Negev, Israel.

    Email: \{shmuelor,coasaf,gurewitz\} @bgu.ac.il

[^1]:    ${ }^{1}$ One can define different criteria for the goodness of the approximation.

[^2]:    ${ }^{2}$ As our main contributions focus on the necessity of user scheduling for CF networks with a large number of users and the asymptotical optimal scheduling policies for such networks, we consider only equal transmissions rates. One can extend this formulation and the results for messages with different lengths and thus different rates as was done in [1].
    ${ }^{3}$ We assume real channels to ease the analysis. The extension for complex channels is not critical for the results in this paper. As will be mentioned later, in the CF framework, complex channels change the achievable rates by a pre-log factor of 2.

[^3]:    ${ }^{4}$ Since the slots are independent and there is no consideration of the rank of $\mathbf{A}$, we may write the maximization on an arbitrary slot and omit the time index.

[^4]:    ${ }^{5}$ One can extend the polynomial-time algorithm presented in this work for a single relay, which scans the channel vector a single relay sees in ascending order, to a parallel scan of all channel vectors simultaneously, while searching the simultaneously good $k$ users. We omit the details.

