The Necessity of Scheduling in Compute-and-Forward

Ori Shmuel
Department of Communication
System Engineering
Ben-Gurion University of the Negev
Email: shmuelor@bgu.ac.il

Asaf Cohen
Department of Communication
System Engineering
Ben-Gurion University of the Negev
Email: coasaf@bgu.ac.il

Omer Gurewitz
Department of Communication
System Engineering
Ben-Gurion University of the Negev
Email: gurewitz@bgu.ac.il

Abstract

Compute and Forward (CF) is a promising relaying scheme which, instead of decoding single messages or forwarding/amplifying information at the relay, decodes linear combinations of the simultaneously transmitted messages. The current literature includes several coding schemes and results on the degrees of freedom in CF, yet for systems with a fixed number of transmitters and receivers.

It is unclear, however, how CF behaves at the limit of a large number of transmitters. In this paper, we investigate the performance of CF in that regime. Specifically, 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 sum-rate tends to zero as well. This makes scheduling necessary in order to maintain the superior abilities CF provides. Indeed, under scheduling, we show that non-trivial linear combinations are chosen, and the sum-rate does not decay, even without state information at the transmitters and without interference alignment.

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 relay 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 received messages. Specifically, together with the use of lattice coding, the obtained signal, after decoding, can be considered as a linear combination of the transmitted messages. This is due to an important characteristic of lattice codes - every linear combination of codewords is a codeword itself.

However, since the wireless channel suffers from fading, the received signals are attenuated by real attenuations factors, hence the received linear combination is “noisy”. The receiver (e.g., a relay) then seeks a set of integer coefficients, denoted by a vector \( \mathbf{a} \), to be as close as possible to the true channel coefficients.

This problem was elegantly associated with Diophantine Approximation Theory in [2], and was compared to a similar problem, that of finding a co-linear vector for the true channel coefficients vector (between the receiver and the transmitters). In addition, the co-linear vector must be an integer valued vector, due to the fact that it should represent the coefficients of an integer linear combination of codewords. Based on this theory, if one wishes to find an integer vector \( \mathbf{a} \) that is close (in terms of co-linearity) to a real vector \( \mathbf{h} \), then one must increase \( ||\mathbf{a}|| \) in order to have a small approximation error between them. The increase in the norm value leads to a significant penalty in the achievable rate at the receiver and thus results in a tradeoff between the goodness of the approximation and the maximization of the rate.

The CF scheme was extended in many directions, such as MIMO CF [3], linear receivers (Integer Forcing) [4] [5], integration with interference alignment [2], scheduling [6] and more [7], [8]. All the mentioned works considered a general setting, where the number of transmitters is a parameter for the system and all transmitters are active at all times. That is, the receiver is able to decode a linear combination of signals from a large number of transmitters as long as the transmitters comply with the achievable rates at the receiver, and still promise, to some extent, an acceptable performance.

However, in this work, we show that the number of simultaneous transmitters is of great importance when the number of relays is fixed. In fact, this number cannot be considered solely as a parameter but as a restriction which forces the use of users scheduling to maintain the superior abilities CF provide.

As an intuition, assume the number of transmitters grows. As a result, the dimension of the channel vector grows, and thus, accordingly, the norm value of \( \mathbf{a} \) should also grow in order to keep the goodness of the approximation. We will show that in this case, 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 it would be better to use other coding schemes for multiple access and interference channels. Furthermore, we show that as the number of transmitters grow the scheme’s sum-rate goes to zero as well.

1One can define different criteria for the goodness of the approximation, for example, minimum difference between the vectors elements.
We conclude this paper with an optimistic view, that user scheduling can improve the CF gain. Using simple Round Robin scheduling and results for CF in fixed size systems, we lower bound the sum-rate. We thus show that even for a simple scheduling policy the system sum-rate does not decay to zero.

The paper is organized as follows. In Section III the system model is described. In Section III we derive an analytical expression for the probability of choosing a unit vector by the relay, as the number of users grows. Section IV depicts the scheduling policy the system sum-rate does not decay to zero. In CF each relay decodes a linear combination \( u \) of the users’ signals through the channel, after which the relay selects a scale coefficient \( \alpha \) to its codeword. Each transmitter then broadcasts it’s codeword to the channel. Hence, each relay \( m \in \{1...M\} \), observes a noisy linear combination of the transmitted signals through the channel,

\[
y_m = \sum_{l=1}^{L} h_{ml}x_l + z_m \quad m = 1, 2, ..., M,
\]

where \( h_{ml} \sim \mathcal{N}(0,1) \) are the real Rayleigh channel coefficients and \( z \) is an i.i.d., Gaussian noise, \( z \sim \mathcal{N}(0, I^{L \times L}) \). Let \( h_m = [h_{m1}, h_{m2}, ..., h_{mL}]^T \) denote the vector of channel coefficients at relay \( m \). We assume that each relay knows its own channel vector. After receiving the noisy linear combination, each relay selects a scale coefficient \( \alpha_m \in \mathbb{R} \), an integer coefficient vector \( a_m = (a_{m1}, a_{m2}, ..., a_{mL})^T \in \mathbb{Z}^L \), and attempts to decode the lattice point \( \sum_{l=1}^{L} a_{ml}x_l \) from \( \alpha_m y_m \).

In CF each relay decodes a linear combination \( u_m \) of the original messages, and forward it to the destination. With enough linear combinations, the destination is able to recover the desired (original) messages from all sources.

The main results in CF are the following.

**Theorem 1** (Theorem 1). For real-valued AWGN networks with channel coefficient vectors \( h_m \in \mathbb{R}^L \) and coefficients vector \( a_m \in \mathbb{Z}^L \), the following computation rate region is achievable:

\[
R(h_m, a_m) = \max_{\alpha_m \in \mathbb{R}} \frac{1}{2} \log^+ \left( \frac{P}{\alpha_m^2 + P \|\alpha_m h_m - a_m\|^2} \right),
\]

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

**Theorem 2** (Theorem 2). The computation rate given in Theorem 1 is uniquely maximized by choosing \( \alpha_m \) to be the MMSE coefficient

\[
\alpha_{MMSE} = \frac{P h_m^T a_m}{1 + P \|h_m\|^2},
\]

\footnote{Note that messages with different length can be allowed with zero padding to attain a length-\( k \) message which will result in different rates for the transmitters.}
which results in a computation rate region:

\[ \mathcal{R}(h_m, a_m) = \max_{a_m \in \mathbb{Z}^L \setminus \{0\}} \frac{1}{2} \log^+ \left( \| a_m \|^2 - \frac{P(h_m^T a_m)^2}{1 + P\|h_m\|^2} \right)^{-1}. \]  

(4)

Note that the above theorems are for real channels and the rate expressions for the complex channel are twice the above (\cite{1}, Theorems 3 and 4).

Since the relay can decide which linear combination to decode (i.e., the coefficients vector \( a \)), an optimal choice will be one that maximizes the achievable rate. That is,

\[ a_{\text{opt}}^m = \arg \max_{a_m \in \mathbb{Z}^L \setminus \{0\}} \frac{1}{2} \log^+ \left( \| a_m \|^2 - \frac{P(h_m^T a_m)^2}{1 + P\|h_m\|^2} \right)^{-1}. \]  

(5)

Remark 1 (The coefficients vector). The coefficients vector \( a \) plays a significant role in the CF scheme. It dictates which linear combination of the transmitted codewords the relay wishes to decode. That is, each non-zero element signifies the fact that the relay is interested in its corresponding codeword. If, starting from a certain number of simultaneously transmitting users, the coefficients vector the relay chooses is always (or with high probability) a unit vector, this means that essentially we treat all other users as noise and loose the promised gain of CF.

The following Lemma bounds the search domain for the maximization problem in (5).

Lemma 1 (\cite{1}, lemma 1). For a given channel vector \( h \), the computation rate \( \mathcal{R}(h_m, a_m) \) in Theorem 2 is zero if the coefficient vector \( a \) satisfies

\[ \| a_m \|^2 \geq 1 + P\|h_m\|^2. \]  

(6)

The problem of finding the optimal \( a \) 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 (5) as an equivalent minimization problem:

\[ a_{\text{opt}}^m = \arg \min_{a_m \in \mathbb{Z}^L \setminus \{0\}} f(a_m) = a_m^T G_m a_m, \]  

(7)

where \( G_m = (1 + P\|h_m\|^2)I - Ph_m h_m^T \). \( G_m \) can be regarded as the Gram matrix of a certain lattice and \( a_m \) will be the shortest basis vector and the one which minimize \( f \). This problem is also known as the shortest lattice vector problem (SLV), which has known approximation algorithms due to its hardness \cite{9, 10}. The most notable of them is the LLL algorithm \cite{11}, which has an exponential approximation factor which grows with the size of the dimension. However, for special lattices, efficient algorithms exist \cite{13}. In \cite{14}, a polynomial complexity algorithm was introduced for the special case of finding the best coefficient vector in CF.

### III. Probability of a Unit Vector

In this section, we examine the coefficient vector at a single relay, hence, we omit the \( m \)-th index in the expressions.

#### A. The Matrix \( G \)

Examine the matrix \( G \), one can notice that as \( L \), the number of transmitters, grows, the diagonal elements grow very fast relatively to the off-diagonal elements. Specifically, each diagonal element is a random variable, which is a \( \chi^2_L \) r.v. minus a multiplication of two Gaussian r.v.s., whereas the off-diagonal elements are only a multiplication of two Gaussian r.v.s.. Of course, as \( L \) grows, the former has much higher expectation value compared to later. Examples of \( G \) are presented in Figure 2 for different dimension. It is clear that, even for moderate number of transmitters, the differences in values between the diagonal and off-diagonal elements are significant.

Consider now the quadric form (7) we wish to minimize. Any choice of \( a \) that is not a unit vector, will add more than one element from the diagonal of \( G \) to it. When \( L \) is large, the off-diagonal elements have little effect on the function value compared to the diagonal elements. Therefore, intuitively, one would prefer to have as little as possible elements from the diagonal although the off-diagonal elements can reduce the function value. This will happen if we choose \( a \) to be a unit vector.
above expression where $a$ is any integer vector that is not a unit vector. For mathematical simplification and ease of notation we rearrange the above expression

$$\text{Pr}(f(e_i) < f(a)) = \left(1 + P(||h||^2 - h_i^2)\right) - 2 \sum_{i=1}^{L-1} P h_i h_j a_i a_j$$

That is, we wish to understand when will a relay prefer a unit vector $a$ over any other non-trivial vector. Specifically, since $a$ is a function of the random channel $h$, we will compute the probability of having a unit vector as the minimizer of $f$. That is, we wish to find the probability

$$\text{Pr}(f(e_i) < f(a)),$$  \hspace{1cm} (8)

where $e_i$ is a 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. It is important to emphasize that the definition of any integer vector $a$ includes the search domain such that $||a|| \leq \sqrt{1 + P||h||^2}$ (Lemma 1)).

C. The Optimality of $e_i$ VS. a Certain Vector $a$

**Theorem 3.** Under the CF scheme, the probability that a unit vector $e_i$ will be the coefficient vector $a^{\text{opt}}$ which maximize the achievable rate $R(h, a^{\text{opt}})$ or alternatively will minimize $f(a^{\text{opt}})$, comparing with an integer vector $a$ that is not a unit vector, is given by

$$\text{Pr}(X \leq Y) = \int_0^\infty \int_{-||a||^2}^{y} \frac{(x - \frac{1-||a||^2}{P})}{\Gamma (\frac{L-1}{2})} \frac{y^{L-1}}{2^{L-1}} e^{-\frac{x}{2} - \frac{y}{2}} \frac{2^{L-1}}{\Gamma (\frac{L-1}{2})(2||a||^2)^{\frac{L-1}{2}}} e^{-\frac{x+y}{2||a||^2}} dx dy$$  \hspace{1cm} (9)

The proof of Theorem 3 is deferred to the end of this subsection. First, we wish to give the main consequence of Theorem 1 in the context of this work.

**Corollary 1.** As the number of simultaneously transmitting users grows, the probability of having a unit vector as the maximizer for the achievable rate goes to one. Specifically,

$$\lim_{L \to \infty} \text{Pr}(f(e_i) < f(a)) = 1,$$  \hspace{1cm} (10)

where $a$ is any integer vector that is not a unit vector.

**Proof.** Equation (8) can be written as,

$$\text{Pr}(f(e_i) < f(a)) = \text{Pr} \left(1 + P(||h||^2 - h_i^2) \leq ||a||^2 + P \sum_{i=1}^{L-1} (h_i a_j - h_j a_i)^2 \right)$$

where $a$ is any integer vector that is not a unit vector. For mathematical simplification and ease of notation we rearrange the above expression

![Figure 2: Example for the elements’ magnitude of the $G$ matrix for different dimensions (i.e., different values of $L$) for $P=10$ and for some realization of a channel vector respectively. The graphs were interpolated for ease of visualization.](image-url)
The rank of the covariance matrix

Let us again examine the random variable $Y = z^T A z$, where $z$ is a Gaussian vector of size $L(L-1)/2$ such that each element $z_i \sim \mathcal{N}(0, a_i^2 + a_j^2)$ and $A$ is the identity matrix with dimension $L(L-1)/2$, which satisfies $A = A^T$. That is, $z$ is an $L(L-1)/2$-variate singular normal with $E[z] = 0$ and $\text{Cov}(z) = \Sigma \geq 0$, such that $\Sigma = BB^T$, $B$ is an $L(L-1)/2 \times r$ matrix of rank $r \leq L(L-1)/2$. For the case of $r$-variate singular normal vectors, $Y$ can be represented as 

$$Y = z^T A z = \sum_{i=1}^{r} \lambda_i N_i^2,$$

where $N \sim \mathcal{N}(0, I_r)$ and $\{\lambda_1, ..., \lambda_r\}$ are the eigenvalues of $B^T A B$.

**Lemma 2.** The rank of the covariance matrix $\Sigma$ is $r = L - 1$.

**Proof.** Let us again examine the random variable $Y$. If we consider a linear transformation on the channel vector $h$, $Y$ can be written as $\|h^T D\|^2$ where $D$ is an $L \times L(L-1)/2$ matrix which has the following block structure $(D_1 \ D_2 \ ... \ D_{L-1})$ where each block $i$ has dimension of $(L-1 \times i)$ and a structure of first $i$ lines of zeros followed by a line of the form $(-a_{i+1}, -a_{i+1}, ... -a_L)$ and the rest is a diagonal matrix $i \times i$ with $a_i$ as the diagonal value. In Figure 3 an example for $L = 5$ is presented. It suffices to show that the rank of $D$ is $L - 1$. Multiplying each row $j$ in $D$ with $a_j$ and adding all other rows to row 1, results in a row echelon form with $L - 1$ non zero rows.

**Lemma 3.** All the non-zero eigenvalues of $B^T A B$ are equal, that is, $\lambda_i = \|a\|^2 \quad \forall i$.

**Proof.** The proof outline is as follows. The matrix $B^T A B$ essentially is $\Sigma$ since $A$ is the identity matrix. We show that $\Sigma = \|a\|^2 \Sigma$, which would lead to the fact that the matrix $\frac{\Sigma}{\|a\|^2}$ is an idempotent matrix. An idempotent matrix is always diagonalizable and its eigenvalues are either 0 or 1. Hence, all non-zero eigenvalues would be equal to $\|a\|^2$. Since $\Sigma$ is real symmetric matrix with rank $L - 1$ (lemma 2), it has $L - 1$ such eigenvalues. A more detailed proof is presented in appendix B.

From lemmas 2 and 3 we have that $Y = z^T A z = \|a\|^2 \sum_{i=1}^{L-1} N_i^2$ and since the $N_i$’s are mutually independent standard normal r.v.s., $Y \sim \text{Gamma}(k = \frac{L-1}{2}, \theta = 2 \|a\|^2)$. 

$$Pr \left(1 - \frac{\|a\|^2}{p} + (\|h\|^2 - h_i^2) \leq \sum_{i=1}^{L-1} \sum_{j=1}^{L-1} (h_i a_j - h_j a_i)^2 \right),$$

and define both left and right terms as

$$X = 1 - \frac{\|a\|^2}{p} + (\|h\|^2 - h_i^2)$$

$$Y = \sum_{i=1}^{L-1} \sum_{j=1}^{L-1} (h_i a_j - h_j a_i)^2.$$

Since $h_i$ is an element of $h$ together with the addition of $1-\frac{\|a\|^2}{p}$, we get a gamma r.v., specifically $X \sim \text{Gamma}(k = \frac{L-1}{2}, \theta = 2, \gamma = 1-\frac{\|a\|^2}{p})$. On the other hand, for each pair of $(i, j)$ of indices, the term $(h_i a_j - h_j a_i)^2 \sim \text{Gamma}(1/2, (a_i^2 + a_j^2)/2)$ since $h_i$ and $h_j$ are i.i.d. Gaussian r.v.s.. Hence, $Y$ is the squared norm of a dependent Gamma random vector with $\frac{L(L-1)}{2}$ elements.

The r.v. $Y$ can also be represented as the quadratic form, $Y = z^T A z$, where $z$ is a Gaussian vector of size $L(L-1)/2$ such that each element $z_i \sim \mathcal{N}(0, a_i^2 + a_j^2)$ and $A$ is the identity matrix with dimension $L(L-1)/2$, which satisfies $A = A^T$. That is, $z$ is an $L(L-1)/2$-variate singular normal with $E[z] = 0$ and $\text{Cov}(z) = \Sigma \geq 0$, such that $\Sigma = BB^T$, $B$ is an $L(L-1)/2 \times r$ matrix of rank $r \leq L(L-1)/2$. For the case of $r$-variate singular normal vectors, $Y$ can be represented as $\sum_{i=1}^{r} \lambda_i N_i^2$.
Figure 4: The lower bound given in (15) (solid lines) and simulation results (dashed lines) for the probability that a unit vector will minimize \( f \) compared to various values of \( \|a\|^2 \) as a function of simultaneously transmitting users.

According to (9), we have

\[
\Pr(X \leq Y) = \int_0^\infty \int_{-\|a\|^2}^y \frac{(x - 1 - \|a\|^2)}{\Gamma \left( \frac{L-1}{2} \right) 2^{\frac{L-1}{2}} e^{-\frac{y}{2}} \left( s - \frac{1 - \|a\|^2}{2} \right)} dx \, ds
\]

\[
\geq 1 - \frac{1}{\Gamma \left( \frac{L-1}{2} \right) 2^{\frac{L-1}{2}}} \int_0^y s^{\frac{L-1}{2} - 1} \Gamma \left( \frac{L-1}{2} \left( s - \frac{1 - \|a\|^2}{2} \right) \right) e^{-\frac{y}{2} \left( s - \frac{1 - \|a\|^2}{2} \right)} ds
\]

Using the d’Alembert’s ratio test one can check that the hypergeometric function \( \text{hypergeometric}(\cdot) \) presented above converges to some constant for the case of \( L \to \infty \), and

\[
\lim_{L \to \infty} \frac{\Gamma \left( \frac{L-1}{2} \right) \left( \|a\|^2 \right)^{\frac{L-1}{2}}}{\Gamma \left( \frac{L}{2} \right) \left( 1 + \|a\|^2 \right)^{\frac{L}{2}}} \text{hypergeometric}(\cdot) \to 0.
\]

Hence, we conclude that the probability goes to one. A more detailed proof is presented in appendix A.

Corollary 1 clarifies that for every \( P \), as the number of users grows, the probability of having a unit vector as the maximizer of the achievable rate, compared to any other coefficient vector with a certain norm \( \|a\|^2 \) is going to 1.

Figure 4 depicts the probability in (15) and simulation results, which clearly above the analytic bound, for various values of \( \|a\|^2 \).

**Proof of Theorem 3** The theorem is a direct consequence of Lemmas 2 and 3 which established the Gamma distribution of the r.v.s \( X \) and \( Y \) which we define in the sequel. Hence, we have a double integral which describes the probability of having a r.v. greater than another r.v.

\[ \Pr \left( \|a\|^2 \leq \|b\|^2 \right) \]

**D. The Optimality of \( e_i \) VS. All Possible Vectors \( a \)**

Corollary 1 refers to the probability that a unit vector will minimize \( f \) for a fixed \( a \). Next we wish to explore this probability for any possible \( a \). For the purpose of clarity, (15) gives a lower bound on the probability that a unit vector will minimize \( f \) compared to a certain possible integer coefficients vectors with a certain \( \|a\|^2 \). Where the probability of having an optimal vector which is not a unit vector will be the union of all probabilities for each a vector which satisfies \( \|a\|^2 < 1 + P \|h\|^2 \).

Let us define \( P(e_i) \) as the probability that a relay picked a unit vector as the coefficient vector, and \( P(e_i) \) as the probability which any other vector was chosen.

In (14), a polynomial time algorithm for finding the optimal coefficients vector \( a \) was given. The complexity result derives from the fact that the \( \text{cardinality of the set of all } a \text{ vectors} \) (denoted as \( \Phi \)) which are considered is upper bounded by

\[ \text{cardinality of } \Phi \text{ equals to the cardinality of } \Phi. \]

\[ \text{cardinality of } \Phi \text{ equals to the cardinality of } \Phi. \]
2L([\sqrt{1+P}\|h\|^2] + 1). That is, any vector which does not exist in this set has zero probability to be the one which maximize the rate. We shall note this set here as \( A \). Thus, we wish to compute
\[
P(\tau_i) = \bigcup_{\mathbf{a}} P( f(\mathbf{a}) \geq f(e_i) ) ,
\]
where \( A = \{ \mathbf{a} \in \mathbb{Z}^L : \mathbf{a} \in \Phi, \mathbf{a} \neq e_i \ \forall i \} \). Note that the cardinality of \( A \) grows with the dimension of \( h \), i.e., with \( L \) and can be easily upper bounded as follows,
\[
|A| \leq 2L([\sqrt{1+P}\|h\|^2] + 1) \leq 2L([1+P]\|h\|^2] + 1) \leq 2L(P\|h\|^2 + 3) .
\]

**Theorem 4.** Under the CF scheme, the probability which any other coefficients vector \( \mathbf{a} \) will be chosen to maximize the achievable rate \( R(\mathbf{h}, \mathbf{a}_{opt}) \) compared with a unit vector \( e_i \), as the number of simultaneously transmitting users grows, is zero. That is,
\[
\lim_{L \to \infty} P(\tau_i) = 0 .
\]

**Proof.** We have,
\[
P(\tau_i) = \bigcup_{\mathbf{a}} P( f(\mathbf{a}) \geq f(e_i) )
\leq \sum_{\mathbf{a}} P( f(\mathbf{a}) \geq f(e_i) ) = \sum_{\mathbf{a}} (1 - P( f(\mathbf{a}) < f(e_i) ) )
\leq (a) \sum_{\mathbf{a}} \frac{\Gamma(L - 1)}{\Gamma (L - 2)} (\frac{\|\mathbf{a}\|^2}{L - 2})^{\frac{L-1}{L-2}} 2F_1(1, L - 1; \frac{L - 1}{2}; 1, \frac{1}{1 + \|\mathbf{a}\|^2})
\leq (b) \frac{|A| \Gamma(L - 1)}{\Gamma (L - 2)} \frac{2^{\frac{L-1}{L-2}} L^{\frac{L-1}{L-2}} 3^{L-1}}{L^{\frac{L-1}{L-2}}} 2F_1(1, L - 1; \frac{L + 1}{2}; 1, \frac{1}{3})
\]
where (a) is due to the probability bound from Corollary 1 (b) is true since the term inside the sum is maximized with \( \|\mathbf{a}\|^2 = 2 \). The function \( 2F_1(\cdot) \) converges to a constant, thus, we have
\[
\lim_{L \to \infty} |A| \frac{\Gamma(L - 1)}{\Gamma (L - 2)} \left( \frac{2}{9} \right)^{\frac{L-1}{L-2}} \leq \lim_{L \to \infty} 4PL^2 e^{-LE(L)} = 0
\]
where \( E(L) = \frac{L-1}{2L} \log \frac{9}{8} \). To obtain this we used the strong law of large numbers to deal with the cardinality of \( A \) and the Stirling’s approximation. A more detailed proof is presented in appendix C.

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.

**IV. Compute and Forward Sum-Rate**

In order for relay \( m \) to be able to decode a linear combination with coefficients vector \( \alpha_m \), all messages’ rates which are involved in the linear combination must be within the computation rate region \( \Pi \). I.e., all the messages for which the corresponding entry in the coefficient vector is non zero. That is,
\[
R_l < \min_{\alpha_m \neq 0} R(\mathbf{h}_m, \alpha_m)
\]

Hence, the sum-rate of the system is defined as the sum of messages’ rates, i.e.,
\[
\sum_{l=1}^{L} \min_{m: \alpha_m \neq 0} R(\mathbf{h}_m, \alpha_m).
\]

Following the results from previous subsections, we would like to show that as the number of users 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 as well. This will strengthen the necessity to schedule users in CF.

**Theorem 5.** As \( L \) grows, the sum-rate of CF tends to zero, that is,
\[
\lim_{L \to \infty} \sum_{l=1}^{L} \min_{m: \alpha_m \neq 0} R(\mathbf{h}_m, \alpha_m) = 0 .
\]
users may transmit simultaneously. The value of $k$ should be chosen in order to apply the CF scheme for systems with a large number of sources, scheduling a smaller number of users should take place. In fact, even higher sum-rates can be obtained if the number of scheduled users is higher than the number of relays. Figure 6 depicts such a scenario.

Theorem 4 and 5 suggest that a restriction on the number of simultaneously transmitting users should be made. That is, in order to apply the CF scheme for systems with a large number of sources, scheduling a smaller number of users should take place.

The most simple scheduling scheme is to schedule users in a Round Robin (RR) manner, where in each transmission only $k$ users transmit and the relay use CF. However, the simulations suggest a peak at a small number of users, for different values of $P$.

V. SCHEDULING IN COMPUTE AND FORWARD

Figure 5: The sum-rate as give in (22) for the case of 4 relays as a function of the number of simultaneously transmitting users, for different values of $P$.

Figure 6: Simulation results for the average sum-rate per transmission. Here, the number of relays is 3 and scheduling was performed in a Round Robin manner, where in every phase 3 sources were scheduled among the transmitting users.

The proof outline is as follows. The sum-rate expression is divided into two parts, which describe two scenarios. The first is for the case where a relay chooses a unit vector as the coefficients vector and the second is for the case where any other vector is chosen. The probabilities for that are $P(e_i)$ and $P(\tau_i)$, respectively. Then, we show that each part goes to zero by upper bounding the corresponding expressions. The complete proof is given in appendix D.

Simulation for the sum-rate for different values of $P$ can be found in Figure 5. It is obvious that for large $L$ the sum-rate decreases, hence, for a fixed number of relays there is no use in scheduling a large number of users, as CF degenerates to choosing unit vectors and treating other users as noise. However, the simulations suggests a peak at a small number of transmitters. We explore this in the next section.
Proof. Let us recall that $Pr(f(e_i) < f(a)) = Pr(X \leq Y)$ and according to (9) we have
\[ Pr(X \leq Y) = \int_0^\infty \int_{-\infty}^{y} \frac{(x - \frac{1}{2}||a||^2)}{\Gamma \left( \frac{L-1}{2} \right) 2^{\frac{L-1}{2}}} e^{-\left(\frac{x - \frac{1}{2}||a||^2}{2}\right)} \frac{y^{L-1}}{\Gamma \left( \frac{L-1}{2} \right) (2||a||^2)^{\frac{L-1}{2}}} e^{-\frac{y}{2||a||^2}} dxdy. \]  

(26)

Let us define \( l = \frac{L-1}{2} \) and \( t = \left(\frac{x - \frac{1}{2}||a||^2}{2}\right) \) hence,

\[
\begin{align*}
&= \frac{1}{\Gamma \left( l \right)^2 2^l ||a||^{2l}} \int_0^\infty y^{l-1} e^{-\frac{y}{2||a||^2}} \int_0^y t^{l-1} e^{-t} dt \, dy \\
&\quad \left( a \right) = \frac{1}{\Gamma \left( l \right)^2 2^l ||a||^{2l}} \int_0^\infty y^{l-1} e^{-\frac{y}{2||a||^2}} \left[ \Gamma \left( l \right) - \Gamma \left( l, \frac{1}{2} \right) \right] \, dy \\
&\quad \left( b \right) = \frac{1}{\Gamma \left( l \right)^2 ||a||^{2l}} \int_0^\infty y^{l-1} e^{-\frac{y}{2||a||^2}} \left[ l, s - \frac{1}{2} ||a||^2 \right] e^{-\frac{y}{2||a||^2}} ds,
\end{align*}
\]

(27)

where (a) is due to the result of the lower incomplete gamma function and in (b) another variable change of \( y = 2s \) was made along with the fact that the first integral reduces to one due to the Gamma function definition.

Note that \( \Gamma \left( l, s + \alpha \right) \) for \( \alpha > 0 \), and since \( ||a||^2 > 1 \) the term \( s - \frac{1}{2} ||a||^2 > 0 \) for the integration range. Therefore one can bound the above integral as follows (using the integral solution given in (16))

\[
\begin{align*}
&\geq 1 - \frac{1}{\Gamma \left( l \right)^2 ||a||^{2l}} \int_0^\infty y^{l-1} \Gamma \left( l, s - \frac{1}{2} ||a||^2 \right) e^{-\frac{y}{2||a||^2}} ds \\
&\geq 1 - \frac{1}{\Gamma \left( l \right)^2 ||a||^{2l}} \int_0^\infty y^{l-1} \Gamma \left( l, s \right) e^{-\frac{y}{2||a||^2}} ds \\
&\geq 1 - \frac{1}{\Gamma \left( l \right)^2 ||a||^{2l}} \frac{\Gamma \left( 2l \right)}{l \left( 1 + \frac{1}{||a||^2} \right)^l} 2 F_1 \left( 1, 2l; l + 1; \frac{1}{1 + ||a||^2} \right) \\
&\geq 1 - \frac{\Gamma \left( 2l \right)}{\Gamma \left( l \right)^2 l \left( 1 + ||a||^2 \right)^l} 2 F_1 \left( 1, 2l; l + 1; \frac{1}{1 + ||a||^2} \right) \\
&\geq 1 - \frac{\Gamma \left( l - 1 \right)}{\Gamma \left( \frac{L-1}{2} \right)^2 \frac{L-1}{2} \left( 1 + ||a||^2 \right)^{L-1}} 2 F_1 \left( 1, L - 1; \frac{L-1}{2}; 1 + \frac{1}{1 + ||a||^2} \right).
\end{align*}
\]

(28)

Where in the last line we replaced \( l \) for its previous definition of \( \frac{L-1}{2} \).

It can be checked out using the d’Alembert’s ratio test that the hypergeometric function \( 2 F_1 \left( \cdot \right) \) presented above converges to some constant for the case of \( L \to \infty \), so we are left to check the behaviour of the rest of the expression as \( L \) grows. Thus, we have

\[
\begin{align*}
&\lim_{L \to \infty} \frac{\Gamma \left( L - 1 \right)}{\Gamma \left( \frac{L-1}{2} \right)^2 \frac{L-1}{2} \left( 1 + ||a||^2 \right)^{L-1}} = \lim_{L \to \infty} \frac{\Gamma \left( L - 1 \right)}{\Gamma \left( \frac{L-1}{2} \right)^2 \frac{L-1}{2} \left( 1 + \frac{1}{||a||^2} \right)^2} \\
&\quad \left( a \right) = \lim_{L \to \infty} \frac{\Gamma \left( L - 1 \right)}{\Gamma \left( \frac{L-1}{2} \right)^2 \frac{L-1}{2} \left( 1 + \frac{1}{||a||^2} \right)^2} = \lim_{L \to \infty} \frac{\Gamma \left( L - 1 \right)}{\Gamma \left( \frac{L-1}{2} \right)^2 \frac{L-1}{2} \left( 1 + \frac{1}{||a||^2} \right)^2} \\
&\quad = \frac{\left( L - 1 \right)!}{\left( L - 3 \right)!^2 \pi 2^{L-1}} = \frac{\left( L - 2 \right)!}{\left( L - 3 \right)!^2 \pi 2^{L-1}} = \frac{\left( L - 2 \right)!}{\left( L - 3 \right)!^2 \pi 2^{L-1}}.
\end{align*}
\]

(29)

where (a) is due to the fact that \( ||a||^2 > 1 \). Note here that for even values of \( L \) we would get half integer inside the Gamma function in the denominator which need to be considered with care. Hence, we use the identity \( \Gamma \left( \frac{1}{2} \right) = \sqrt{\pi} \frac{(n-2)!}{2^{n-1}n!} \) and the fact that \( \Gamma(n) = (n - 1)! \),

\[
\begin{align*}
&\lim_{L \to \infty} \frac{\Gamma \left( L - 1 \right)}{\Gamma \left( \frac{L-1}{2} \right)^2 \frac{L-1}{2} \left( 1 + \frac{1}{||a||^2} \right)^2} = \lim_{L \to \infty} \frac{\Gamma \left( L - 1 \right)}{\Gamma \left( \frac{L-1}{2} \right)^2 \frac{L-1}{2} \pi 2^{L-1}} = \lim_{L \to \infty} \frac{\Gamma \left( L - 2 \right)!}{\Gamma \left( L - 3 \right)!^2 \pi 2^{L-1}}.
\end{align*}
\]

(30)
The double factorial has the following relation to factorial, \( n!! = 2^k \cdot k! \) for even positive \( n = 2k \), \( k \geq 0 \), and \( n!! = \frac{(2k-1)!}{2^{k-1}} \) for odd \( n = 2k - 1 \), \( k \geq 1 \).

Therefore we have \((L-3)!! = \frac{(L-2)!!}{2^{(L-2)/2}}\) and \((L-3)!! = 2(L-3)/2(\frac{L-3}{2})!!\) for even and odd \( L \) respectively. Considering even \( L \) first, we have,

\[
\lim_{L \to \infty} \frac{1}{\pi} \frac{(L-2)!}{(L-3)!!(L-1)} = \lim_{L \to \infty} \frac{1}{\pi} \frac{(L-2)!}{2^{(L-2)/2} (\frac{L-2}{2})!!} (L-1) = \lim_{L \to \infty} \frac{1}{\pi} \frac{2^{L-2}}{(\frac{L-2}{2})!!(\frac{L-2}{2})} (L-1)
\]

\[
= \lim_{L \to \infty} \frac{1}{\pi} \frac{2^{L-2}}{(\frac{L-2}{2})^2} (L-1) \leq \lim_{L \to \infty} \frac{1}{\pi} \frac{2}\left(\frac{L-2}{2}\right) (L-1) = 0,
\]

where \((b)\) is due to the Stirling’s approximation. For the case of odd \( L \) we have,

\[
\lim_{L \to \infty} \frac{1}{\pi} \frac{(L-2)!}{(L-3)!!(L-1)} = \lim_{L \to \infty} \frac{1}{\pi} \frac{(L-2)!}{2^{(L-3)/2} (\frac{L-3}{2})!!} (L-1) = \lim_{L \to \infty} \frac{1}{\pi} \frac{(L-3)!}{2^{L-3} ((\frac{L-3}{2})!!)^2} (L-1)
\]

\[
= \frac{1}{\pi} \cdot \lim_{L \to \infty} \frac{L-3}{\sum_{i=1}^{L-3} (\frac{L-3}{2})} = 0.
\]

Considering the above we have,

\[
\lim_{L \to \infty} 1 - \frac{\Gamma(L-1)}{\Gamma(\frac{L+1}{2})^2 \frac{L-1}{2} (1 + ||a||^2)^{L-1}} 2F_1(1, L-1; \frac{L-1}{2}; 1 + \frac{1}{1+||a||^2}) = 1
\]

\[
(33)
\]

**APPENDIX B**

**PROOF FOR LEMMA 3**

**Proof.** Since \( A \) is the identity matrix then \( B^T \Sigma B = \Sigma \), which is symmetric with dimension of \( \frac{L(L-1)}{2} \), and as proved in lemma [it’s rank equals to \( L - 1 \). As can be seen in Figure 8a, \( \Sigma \) has a unique block structure. Each block \( B_{kl} \) has dimension of \( (L - k) \times (L - l) \), and since the matrix is symmetric we have \( B_{kl} = B_{lk}^T \). The diagonal blocks and off diagonal blocks are described by the following expressions and presented in figures 8b and 8c respectively.

\[
k = l:
\]

\[
b_{ij}^{(kl)} = \begin{cases} (a_k^2 + a_{k+i}^2) \delta_{i=j} + a_k a_{k+i} \delta_{i \neq j} & \text{if } k > l; \\ a_k a_l & \text{if } k = l; \\ a_k a_l \delta_{i=k-l+i} - a_k a_{l+i} \delta_{j=k-l} & \text{if } k < l; \\ a_k a_l \delta_{i=k-l+i} - a_k a_{l+i} \delta_{j=k-l} & \text{if } k < l;
\end{cases}
\]

\[
\text{(34)}
\]

where \( b_{ij}^{(kl)} \) is the element in the \((i, j)\) position of the block \( B_{kl} \).

Due to the finite-dimensional spectral theorem regarding real symmetric matrices (which is a special case of hermitian matrices), \( \Sigma \) has \( \frac{L(L-1)}{2} \) linearly independent eigenvectors. Therefore, the matrix has \( L - 1 \) non zero eigenvalues. We will show that \( \Sigma^2 = ||a||^2 \Sigma \), which would lead to the fact that the matrix \( \Sigma \) is an idempotent matrix. An idempotent matrix is always diagonalizable and its eigenvalues are either 0 or 1. Hence, all the \( L - 1 \) eigenvalues of \( \Sigma \) would be equal to \( ||a||^2 \).

Owing to the block structure of \( \Sigma \), in order to compute \( \Sigma^2 \), we can perform block multiplication which will simplify the analysis. Each block in \( \Sigma^2 \) is obtained by

\[
C_{kl} = \sum_{r=1}^{L-1} B_{kr} B_{rl}.
\]

\[
(35)
\]

We will describe the sum of multiplications in 37 result for the diagonal and off-diagonal blocks in \( \Sigma^2 \).
The diagonal blocks of $\Sigma^2$.

The diagonal blocks $C_{kk}$ consist of the following

$$C_{kk} = B_{k1}B_{1k} + B_{k2}B_{2k} + \ldots + B_{kk}B_{kk} + \ldots + B_{k(L-1)}B_{(L-1)k}$$

where we have 3 different block multiplications for the case of $l < k, l = k$ and $l > k$. We will note the element $b_{ij}^{kl}$ as the element in the $(i, j)$ position of the multiplication result of $B_{kl}B_{lk}$. Hence, the elements of the blocks $C_{kk}$ is obtained by

$$c_{ij}^{(kk)} = \sum_{1 \leq l < k} b_{ij}^{kl} + b_{ij}^{kk} + \sum_{k < l \leq L-1} b_{ij}^{lk}$$

specifically,

$$b_{ij}^{kk} = \begin{cases} (a_k^2 + a_{k+1}^2)a_{k+j}a_{k+i} + (a_k^2 + a_{k+1}^2)a_{k+j}a_{k+i} & \text{if } i \neq j \\ (a_k^2 + a_{k+1}^2) & \text{if } i = j \end{cases}$$

$$b_{ij}^{kl,(k>l)} = \begin{cases} (a_k^2a_{k+i} + a_{k+1}^2a_{k+j}) & \text{if } i \neq j \\ a_k^2a_{k+i}^2 + a_{k+1}^2a_{k+j}^2 & \text{if } i = j \end{cases}$$

$$b_{ij}^{kl,(k<l)} = \begin{cases} (a_k^2a_{k+i} + a_{k+1}^2a_{k+j}) & \text{if } i \neq j \\ a_k^2a_{k+i}^2 + a_{k+1}^2a_{k+j}^2 & \text{if } i = j \end{cases}$$

where, $\delta_{\{\}}$ is 1 when the condition satisfies and zero otherwise. The diagonal elements of the blocks $C_{kk}$ are then reduced to

$$c_{ii}^{(kk)} = b_{ii}^{kk} + \sum_{1 \leq l < k} b_{ii}^{lk} + \sum_{k < l \leq L-1} b_{ii}^{lk}$$

$$= (a_k^2 + a_{k+1}^2)^2 + \sum_{r=k+1}^L a_k^2a_r^2 - a_{k+i}^4 + \sum_{1 \leq l < k} (a_k^2a_l^2 + a_{k+1}^2a_{l+1}^2)$$

$$+ \sum_{k < l \leq L-1} \left( a_k^2a_l^2 + \left( a_k^2a_l^2 + \sum_{r=l+1}^L a_k^2a_r^2 \right) \delta_{i=l-k} \right) \delta_{i>l-k-1}$$

![Covariance matrix partitioned to blocks](image)

Figure 8: Illustration for the covariance matrix and it’s block structures.
$$
\begin{align*}
\text{(a)} \quad & (a_k^2 + a_{k+i}^2)^2 + a_{k+i}^2 \left( \sum_{r=k+1}^{L} a_r^2 - a_{k+i}^2 \right) + \left( a_k^2 + a_{k+i}^2 \right) \sum_{1 \leq l < k} a_l^2 \\
+ & \sum_{r=k+1}^{L} a_k^2 a_r^2 - a_k^2 a_{k+i}^2 \\
= & \left( a_k^2 + a_{k+i}^2 \right) \left[ a_k^2 + a_{k+i}^2 \sum_{r=k+1}^{L} a_r^2 - a_{k+i}^2 + \sum_{1 \leq l < k} a_l^2 \right] \\
\equiv & \text{(b)} \quad (a_k^2 + a_{k+i}^2) ||a||^2.
\end{align*}
$$

Where the last term in (a) can easily verified to be correct if one considers its associated expression as a function of $i$ and (b) is essentially the diagonal elements of the block $C_{kk}$ multiplied by $||a||^2$. Similarly the off-diagonal elements of the blocks $C_{lk}$ are reduced to

$$
\begin{align*}
C_{ij}^{(kk)} &= b_{ij}^{kk} + \sum_{1 \leq l < k} b_{ij}^{lk} + \sum_{k < l \leq L-1} b_{ij}^{lk} \\
&= (a_k^2 + a_{k+j}^2) a_{k+j} a_{k+i} + (a_k^2 + a_{k+i}^2) a_{k+j} a_{k+i} + \sum_{r=k+1}^{L} a_{k+j} a_{k+i} a_r^2 \delta_{r \neq j} + \delta_{r \neq i+k} \\
+ & \sum_{1 \leq l < k} a_l^2 a_{k+i} a_{k+j} \\
+ & \sum_{k < l \leq L-1} \left( -a_k^2 a_l a_{k+i} \delta_{j=l-k} - a_k^2 a_l a_{k+j} \delta_{j=l-k} \right) \delta_{i \neq l-k} \\
\equiv & \text{(a)} \quad a_{k+i} a_{k+j} \left( 2a_k^2 + a_{k+i}^2 + a_{k+j}^2 \right) + a_{k+i} a_{k+j} \left( \sum_{r=k+1}^{L} a_r^2 - a_{k+i}^2 - a_{k+j}^2 \right) \\
+ & a_{k+j} a_{k+i} \sum_{1 \leq l < k} a_l^2 - a_k^2 a_{k+i} a_{k+j} \\
= & a_{k+i} a_{k+j} \left[ 2a_k^2 + a_{k+i}^2 + a_{k+j}^2 + \sum_{r=k+1}^{L} a_r^2 - a_{k+i}^2 - a_{k+j}^2 + \sum_{1 \leq l < k} a_l^2 - a_k^2 \right] \\
\equiv & \text{(b)} \quad a_{k+i} a_{k+j} ||a||^2.
\end{align*}
$$

Where the second term in (a) can be written like this since $i \neq j$ and the last term in (a) can easily verified to be correct if one considers its associated expression as a function of $i$ and $j$. In (b) we see that each off-diagonal element of the block $C_{kk}$ is multiplied by $||a||^2$.

**The off-diagonal blocks of $\Sigma^2$:**

The off-diagonal blocks $C_{kl}$ of $\Sigma^2$ consist of the following

$$
C_{kl} = B_{k1} B_{1l} + B_{k2} B_{2l} + \ldots + B_{kr} B_{rl} + \ldots + B_{k(L-1)} B_{(L-1)l} \quad r = 1, \ldots, L-1
$$

(39)

where we have 5 different block multiplications for the case of $r = k$, $r = l$, $(r < k \cap r < l)$, $(r > k \cap r < l)$ and $(r > k \cap r > l)$. We will note the element $b_{ij}^{kk}$, $b_{ij}^{kl}$, $b_{ij}^{kl}$, $b_{ij}^{kkr}$, and $b_{ij}^{kkl}$ as the element in the $(i, j)$ position of each of the block multiplication cases respectively. Since $\Sigma$ is symmetric and it commutes with itself, $\Sigma^2$ is also symmetric, therefore it is sufficient to examine the off-diagonal elements of the blocks $C_{kl}$ such that $l > k$, herein,
\[
b_{ij}^{kkl} = \begin{cases} 
-ak_{a_i+j}a_k^2\delta_{i=l-k} + a_k a_t a_k^2\delta_{i=l-k+j} \\
-ak_{a_i+j} \sum_{m=l}^L a_m^2\delta_{i=l-k} + a_k a_t \left(a_t^2 + a_t^2 \right) \delta_{i=l-k+j} \\
+ak_{a_t}a_{k+a_t+j}\delta_{i \neq l-k+j}, \\
& i > l-k \\
\end{cases} 
\]

\[
b_{ij}^{kll} = \begin{cases} 
-ak_{a_i+j}a_k^2\delta_{i=l-k} + a_k a_t a_k^2\delta_{i=l-k+j} \\
-ak_{a_i+j} \sum_{m=l}^L a_m^2\delta_{i=l-k} + a_k a_t a_k^2\delta_{i \neq l-k+j}, \\
& i > l-k \\
\end{cases} 
\]

\[
b_{ij}^{rkl} = \begin{cases} 
a_k a_t \sum_{m=r+1}^L a_m^2\delta_{i=r-k} - ak_{a_t} a_t a_k^2\delta_{i=r-k} - ak_{a_t} a_k a_{a_t+j} \delta_{i=r-l}, \\
& i > r-k \\
\end{cases} 
\]

\[
b_{ij}^{ktr} = \begin{cases} 
a_k a_t \sum_{m=r+1}^L a_m^2\delta_{i=r-k} - ak_{a_t} a_k a_{a_t+j} \delta_{i=r-l}, \\
& i > r-k \\
\end{cases} 
\]

where, \( \delta_{\{i\}} \) is 1 when the condition satisfies and zero otherwise. The off-diagonal elements of the blocks \( C_{kl} \) are then reduced to

\[
c_{ij}^{(kl)} = b_{ij}^{kkl} + b_{ij}^{kll} + \sum_{r=1}^{k-1} b_{ij}^{rkl} + \sum_{r=k+1}^{l-1} b_{ij}^{rtr} + \sum_{r=l+1}^{L-1} b_{ij}^{ktr} 
\]

\[
= -ak_{a_t+j}a_k^2\delta_{i=l-k} + a_k a_t a_k^2\delta_{i=l-k+j} \\
-ak_{a_t+j} \sum_{m=l}^L a_m^2\delta_{i=l-k} + a_k a_t a_k^2\delta_{i \neq l-k+j}, \\
& i > l-k \\
+ \sum_{r=1}^{k-1} \left( a_k a_t a_k^2 \delta_{i=l-k+j} - a_k a_{a_t+j} a_k^2 \delta_{i=l-k} \right) \\
+ \sum_{r=k+1}^{l-1} \left( a_k a_t a_k^2 \delta_{i=l-k+j} - a_k a_{a_t+j} a_k^2 \delta_{i=l-k} \right) \\
+ \sum_{r=l+1}^{L-1} \left( a_k a_t \sum_{m=r+1}^L a_m^2 \delta_{i=r-l} - ak_{a_t} a_k a_{a_t+j} \delta_{i=r-l-j} \right), \\
(i \neq r-k, j \neq r-l) \\
\cup (j \neq r-l, i \neq r-k) \\
\]

(a) \( \delta_{i=l-k} \)

(b) \( \delta_{i=l-k+j} \)

(c) \( \delta_{i=l-k} \)

\[
= \begin{cases} 
\delta_{i=l-k} \\
\delta_{i=l-k+j} \\
\delta_{i=l-k} \\
\end{cases} 
\]

where, \( \delta_{\{i\}} \) is 1 when the condition satisfies and zero otherwise. The off-diagonal elements of the blocks \( C_{kl} \) are then reduced to
\[ \delta_{i=l-k} \left( -a_k a_{l+j} \sum_{m=1}^{L} a_m^2 \right) + \delta_{i=l-k+j} \left( a_k a_l \left( a_{l+j}^2 + \sum_{m=1}^{l} a_m^2 \right) \right) \\
+ \delta_{i=l-k+j} a_k a_l \left( \sum_{r=l+1}^{L} \sum_{m=r+1}^{L} a_m^2 \delta_{j=r-l} + \sum_{r=l+1}^{L} a_r^2 \delta_{i>r-k} \right) \\
+ a_k a_l a_{l+j} \left( \sum_{i>l-k} \delta_{i=r-k+l} - \sum_{r=l+1}^{L} \delta_{i=r-k+j} + \sum_{j=r-l,i>r-k} \delta_{i+r-l} \right) \\
= \delta_{i=l-k} \left( -a_k a_{l+j} \sum_{m=1}^{L} a_m^2 \right) + \delta_{i=l-k+j} \left( a_k a_l \left( a_{l+j}^2 + \sum_{m=1}^{l} a_m^2 \right) \right) \\
+ \delta_{i=l-k+j} a_k a_l \left( \sum_{m=1}^{L} a_m^2 - a_{l+j}^2 \right) \\
= \delta_{i=l-k} \left( -a_k a_{l+j} \sum_{m=1}^{L} a_m^2 \right) + \delta_{i=l-k+j} \left( a_k a_l \sum_{m=1}^{l} a_m^2 \right) \\
= \delta_{i=l-k} \left( -a_k a_{l+j} \|a\|^2 \right) + \delta_{i=l-k+j} \left( a_k a_l \|a\|^2 \right), \]

where in (a) the expressions are arranged according to their delta function. In (c) the last term of (b) has been expanded. In (d) overlapping areas of the delta conditions was combined. The third term in (e) can be verified if one fix \( j = r - l \) then the only element in the first summation is the one which \( r = j + l \). On the other hand, since \( i = l - k + j \), then the second summation will go up until \( r = l + (j - 1) \). In (f) we see that each element of the block \( C_{kl} \) is multiplied by \( \|a\|^2 \). \( \square \)

**Appendix C**

**Proof for Theorem 4**

**Proof.** Using the union bound we have

\[
P(\bar{x}_i) = \bigcup_{A} P( f(a) \geq f(e_i)) \leq \sum_{A} P( f(a) \geq f(e_i)) = \sum_{A} (1 - P( f(a) < f(e_i)))
\]

\[
= \sum_{A} \frac{\Gamma((L-1)/2)}{\Gamma((L-1)/2) L^{-1}} \frac{L^{-1}}{L^{-1} (1 + \|a\|^2) L^{-1}} 2F_1(1, L - 1; \frac{L - 1}{2}; \frac{1}{1 + \|a\|^2})
\]

\[
\leq \sum_{A} \frac{\Gamma((L-1)/2)}{\Gamma((L-1)/2) L^{-1}} \frac{L^{-1}}{L^{-1} (1 + 2) L^{-1}} 2F_1(1, L - 1; \frac{L - 1}{2}; \frac{1}{1 + 2})
\]

\[
= |A| \frac{\Gamma((L-1)/2)}{\Gamma((L-1)/2) L^{-1}} \frac{L^{-1}}{L^{-1}} 2F_1(1, L - 1; \frac{L - 1}{2}; \frac{1}{3}),
\]

where (a) is true since both terms inside the sum are maximized for \( \|a\|^2 = 2 \) due to fact that both are decreasing functions with \( \|a\|^2 \). We will assume that \( L \) is odd for the analytical derivation that follows, however, it is easy to consider even \( L \) and perform similar calculation. This can be seen in the proof of Corollary 1 in appendix A. Remember that \( 2F_1(\cdot) \) converges to some constant, thus, we are left with

\[ (41) \]
\[
\lim_{L \to \infty} |A| \frac{\Gamma(L-1)}{\Gamma \left( \frac{L-1}{2} \right)^2} \left( \frac{2}{9} \right)^{\frac{L-1}{2}} \\
\leq \lim_{L \to \infty} 2L(P\|h\|^2 + 3) \frac{\Gamma(L-1)}{\Gamma \left( \frac{L-1}{2} \right)^2} \left( \frac{2}{9} \right)^{\frac{L-1}{2}} \\
= \lim_{L \to \infty} 2L(P \sum_{i=1}^L h_i^2 + 3) \frac{\Gamma(L-1)}{\Gamma \left( \frac{L-1}{2} \right)^2} \left( \frac{2}{9} \right)^{\frac{L-1}{2}} \\
\equiv \lim_{L \to \infty} 2L^2P \frac{\Gamma(L-1)}{\Gamma \left( \frac{L-1}{2} \right)^2} \left( \frac{2}{9} \right)^{\frac{L-1}{2}} \\
= \lim_{L \to \infty} 2L^2P \left( \frac{L-3}{L} \right) \left( \frac{L-2}{L} \right) \left( \frac{2}{9} \right)^{\frac{L-1}{2}} \\
\leq 4P \lim_{L \to \infty} L^2 \left( \frac{L-3}{L} \right) \left( \frac{2}{9} \right)^{\frac{L-1}{2}} \\
\equiv 4P \lim_{L \to \infty} L^2 \left( \frac{L-1}{2} \right) \left( \frac{2}{9} \right)^{\frac{L-1}{2}} \\
\equiv 4P \lim_{L \to \infty} L^2 \left( \frac{L-1}{2} \right) \left( \frac{2}{9} \right)^{\frac{L-1}{2}} \\
= 4P \lim_{L \to \infty} L^2 e^{-LE(L)} \\
= 0
\]

where \((b)\) is due to \((17)\), in \((c)\) we multiplied and divide with \(L\) and eliminate the limit term which is multiplied by \(3\) since it goes to zero. \((d)\) follows from the strong law of large numbers were the normalized sum converge with probability one to the expected value of \(\chi_1^2\) r.v. which is one. In \((e)\) we used the Stirling’s approximation. And lastly we define \(E(L) = \frac{L-1}{2L} \log \frac{9}{2} \).

**Appendix D**

**Proof for Theorem 5**

**Proof.** The probabilities \(P(e_i)\) and \(P(\overline{e_i})\) define a partition on the channel vectors a relays sees, specifically we define,

\[
H_e = \{ h \in \mathbb{R}^L | \arg \min_{a \in Z} f(a) = e_i \} \\
H_{\overline{e}} = \{ h \in \mathbb{R}^L | \arg \min_{a \in Z} f(a) \neq e_i \}.
\]

That is, with probability \(P(e_i)\) a relay sees a channel vector \(h \in H_e\) and with probability \(P(\overline{e_i})\) a relay sees a channel vector \(h \in H_{\overline{e}}\). We note \(h^e\) and \(h^\overline{e}\) as a channel vectors which belongs to \(H_e\) and \(H_{\overline{e}}\) respectively.

Under the above definitions, the sum-rate can be written as follows,

\[
\sum_{l=1}^L \min_{m:a_m \neq 0} \mathcal{R}(h_m, a_m) \\
= \sum_{l=1}^L P(e_i) \min_{m:a_m \neq 0} \mathcal{R}(h^e_m, e_i) + P(\overline{e_i}) \min_{m:a_m \neq 0} \mathcal{R}(h^\overline{e}_m, a_m).
\]

We treat the two terms above separately where the second term represents the sum-rate for the case which the optimal coefficients vectors may be any integer vector excluding the unit vector \(e_i\). And the first term is for the case that the optimal coefficients vector is \(e_i\). We will show that both terms goes to zero while starting with the second term.
\[
\sum_{l=1}^{L} P(\pi_l) \min_{m, a_m \neq 0} \mathcal{R}(h_m^\pi, a_m) = P(\pi_l) \sum_{l=1}^{L} \min_{m, a_m \neq 0} \mathcal{R}(h_m^\pi, a_m) \\
\leq P(\pi_l) \max_{m} \mathcal{R}(h_m^\pi, a_m) = P(\pi_l) \max_{m} \frac{1}{2} \log^+ \left(1 + \|h_m^\pi\|^2 \right) \\
\leq P(\pi_l) \max_{m} \frac{1}{2} \log^+ \left(1 + P\|h_m^\pi\|^2\right) = \max_{m} \left\{ P(\pi_l) \frac{1}{2} \log^+ \left(1 + P\|h_m^\pi\|^2\right) \right\}.
\]

Define \( R_L^\pi = P(\pi_l) \frac{1}{2} \log^+ \left(1 + P\|h_m^\pi\|^2\right) \), we would like to show that \( R_L^\pi \to 0 \), that is,

\[
\lim_{L \to \infty} P(R_L^\pi \geq \epsilon) = 0.
\]

Using The Markov and Jensen’s inequalities we have,

\[
P(R_L^\pi \geq \epsilon) \leq \frac{1}{\epsilon} E \left[ R_L^\pi \right] = \frac{1}{\epsilon} E \left[ P(\pi_l) \frac{1}{2} \log^+ \left(1 + P\|h_m^\pi\|^2\right) \right] \\
\leq \frac{1}{\epsilon} P(\pi_l) \frac{L}{2} \log^+ \left(1 + P\|h_m^\pi\|^2\right),
\]

therefore, we are interested in analyzing the expectation of the squared norm values belonging to all channel vectors \( h^\pi \).

Remember that, without any constraints, the channel vector \( h \) is a Gaussian random vector which it’s squared norm follows the \( \chi^2 \) distribution, we shall note as \( f_{\chi^2}(x) \). A single squared norm value can belong to a several different Gaussian random vectors. Hence, we define \( H^\pi \text{norm} \) as the set of squared norm values which belongs to \( h^\pi \), formally,

\[
H^\pi \text{norm} = \{ \|h\|^2 \in \mathbb{R} | h \in H^\pi \},
\]

which means in words, all possible squared norm values which belong to all vectors \( h^\pi \). We define then, \( P(\xi) = P(\|h\|^2 \in H^\pi \text{norm}) \) as the probability to belong to \( H^\pi \text{norm} \). That is,

\[
\int_{H^\pi \text{norm}} f_{\chi^2}(x) dx = P(\xi).
\]

Returning to the expectation in (47) we have,

\[
E \left[ \|h_m^\pi\|^2 \right] = \int_{H^\pi \text{norm}} x f_{\chi^2}(x) \frac{dx}{P(\xi)} \leq \int_{\alpha}^\infty x f_{\chi^2}(x) \frac{dx}{P(\xi)} \\
\text{(a)} \leq \int_{\alpha}^\infty x f_{\chi^2}(x) \frac{dx}{P(\pi_l)} \leq \frac{1}{P(\pi_l)} \int_{0}^\infty x f_{\chi^2}(x) dx \\
= \frac{L}{P(\pi_l)}.
\]

Where, \( \alpha \) satisfies \( \int_{\alpha}^\infty f_{\chi^2}(x) dx = P(\xi) \) and (a) is due to the fact that \( P(\xi) \geq P(\pi_l) \) since it may happen that two vectors \( h^c \) and \( h^\pi \) would have the same squared norm value.

Applying the expectation’s upper bound in (47) we have,

\[
\frac{1}{\epsilon} P(\pi_l) \frac{L}{2} \log^+ \left(1 + P E \left[ \|h_m^\pi\|^2 \right] \right) \leq \frac{1}{\epsilon} P(\pi_l) \frac{L}{2} \log^+ \left(1 + P L \right) \\
\text{(a)} \leq \frac{1}{\epsilon} P(\pi_l) \frac{L}{2} \log^+ \left(1 + P L \right) \leq \frac{1}{\epsilon} P(\pi_l) \frac{L}{2} \sqrt{2PLP(\pi_l)} \\
\text{(b)} \leq \frac{1}{\epsilon} \frac{L}{2} \sqrt{2PL^2e^{-LE'(L)}} = \frac{1}{\epsilon} P L^2 \sqrt{2Le^{-LE'(L)}}
\]

where (a) is due the bound \( \log(1 + x) \leq \sqrt{2x} \), (b) following directly from Theorem 4 and \( E'(L) = \frac{L-1}{4L} \log \frac{9}{8} \). Considering the above, as \( L \) grows, the second term of (44) is going to zero, that is,

\[
\lim_{L \to \infty} \sum_{l=1}^{L} P(\pi_l) \min_{m, a_m \neq 0} \mathcal{R}(h_m^\pi, a_m) \leq \lim_{L \to \infty} \frac{1}{\epsilon} \frac{P L^2 \sqrt{L}}{\sqrt{0.5}} e^{-LE'(L)} = 0
\]
for all $\epsilon > 0$.

Thus, we are left with the first term in (44)

$$
\lim_{L \to \infty} \sum_{l=1}^{L} P(e_i) \min_{m:e_i \neq 0} \mathcal{R}(h_m, e_i) \leq \sum_{l=1}^{L} \min_{m:e_i \neq 0} \mathcal{R}(h_m, e_i)
$$

\begin{align}
&= \lim_{L \to \infty} \sum_{l=1}^{L} \min_{m:e_i \neq 0} \frac{1}{2} \log^+ \left( \frac{1 + P\|h_m\|^2}{1 + P(\|h_m\|^2 - h_i^2)} \right) \\
&\leq \lim_{L \to \infty} \sum_{m=1}^{M} \frac{1}{2} \log^+ \left( \frac{1 + P\|h_m\|^2}{1 + P(\|h_m\|^2 - h_i^2)} \right) \\
&= \sum_{m=1}^{M} \frac{1}{2} \log^+ \left( \lim_{L \to \infty} \frac{1 + P\|h_m\|^2}{1 + P(\|h_m\|^2 - h_i^2)} \right) = 0
\end{align}

where in (a) we set the unit vector $e_i$ in the rate expression $\mathcal{R}(h_m, a_m)$. The upper bound (b) is for the best case scenario for which each relay has different unit vector $e_i$. Finally, it is clear that as $L$ grows for each realization of $h_m$, the argument of the log is going to 1.

\[ \square \]

\section*{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] 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.

[3] 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.

[4] 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.

[5] A. Sakzad, J. Harshan, and E. Viterbo, “Integer-forcing mimo linear receivers based on lattice reduction,” Wireless Communications, IEEE Transactions on, vol. 12, no. 10, pp. 4905–4915, 2013.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] A. K. Lenstra, H. W. Lenstra, and L. Lovász, “Factoring polynomials with rational coefficients,” Mathematische Annalen, vol. 261, no. 4, pp. 515–534, 1982.

[12] 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.

[13] J. H. Conway and N. J. A. Sloane, Sphere packings, lattices and groups. Springer Science & Business Media, 2013, vol. 290.

[14] 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.

[15] A. M. Mathai and S. B. Provost, Quadratic forms in random variables: theory and applications. M. Dekker New York, 1992.

[16] I. S. Gradshteyn and I. M. Ryzhik, Table of integrals, series, and products. Academic press, 2014.