Capacity Scaling of (Long) Non-Regenerative MIMO Multi-Hop Channels

Jörg Wagner and Armin Wittneben
Communication Technology Laboratory, ETH Zurich
CH-8092 Zurich, Switzerland
{wagner, wittneben}@nari.ee.ethz.ch

***Abstract*— We consider coherent wireless MIMO multi-hop communication where $n_{\mathcal{S}}$ non-cooperating source antennas communicate to $n_{\mathcal{D}}$ fully cooperating destination antennas via L clusters of $n_{\mathcal{R}}$ non-cooperating relay antennas each. Assuming i.i.d. zero-mean (slow) fading coefficients of finite variance, the sum-capacity of this channel is proven to scale linearly in $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ in the limit $L \to \infty$, when the receive SNR is kept constant and $n_{\mathcal{R}} \in \Omega(\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\})$. This capacity scaling is achieved by a quantization based forwarding strategy. Moreover, we show that Slepian & Wolf compression allows to reduce the growth rate of the transmit power per cluster that is necessary to sustain a constant receive SNR at the destination antennas from exponential to linear in L.**

I. Introduction

Consider an i.i.d. (slow and flat) Rayleigh fading MIMO multi-hop channel with $n_{\mathcal{S}}$ non-cooperating source antennas and $n_{\mathcal{D}}$ fully cooperating destination antennas, as well as L clusters of $n_{\mathcal{R}}$ non-cooperating relay antennas each. The source signal traverses all L clusters of relay antennas, before it reaches the destination cluster after $L+1$ hops/time slots. This setup has been studied with respect to sum-capacity scaling in $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ in several works in recent time [1], [2], [3]. These references have in common that they assume an "amplify & forward" operation in the relay nodes.

Reference [1] shows that the amplify & forward sum-capacity scales linearly in $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ for a fixed L and $n_{\mathcal{R}} \in \Omega(\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\})$[1]. However, it is an immediate consequence of a result in [4] that for $n_{\mathcal{R}} \in \mathcal{O}(\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\})$ the amplify & forward sum-capacity cannot scale linearly in $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ in the limit $L \to \infty$. Motivated by this fact the works [2] and [3] search for conditions under which linear sum-capacity scaling in the amplify & forward channel is sustained even for $L \to \infty$. This is achieved by either increasing the signal-to-noise ratio (SNR) at the destination or the ratio $n_{\mathcal{R}}/\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ as a function of L. In particular, linear sum-capacity scaling can be enforced in the amplify & forward channel by

- [2] an exponential growth of the SNR at the destination antennas in L,
- [3] a linear growth of the ratio $n_{\mathcal{R}}/\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ in L.

The fundamental effect that harms the amplify & forward sum-capacity scaling for increasing L is an unfavorable distortion in the singular value spectrum of the product of single hop channel matrices [4], which is inherent in the amplify & forward strategy.

In search of better relaying strategies, we therefore resort to quantization based forwarding strategies. In particular, we distinguish between the "quantize & forward" and the "compress & forward" strategy, which differ in an additional Slepian & Wolf compression in the latter strategy. That is quantize & forward is suboptimal in the sense that it does not exploit the correlation between signals received by relay nodes within the same cluster. However, it does not require any channel state information of other relay nodes as opposed to compress & forward.

For decoding the destination cluster successively traces back the observations of the relay nodes in a cluster by cluster fashion. It decodes the channel codewords of the relay nodes in the cluster preceding it, in order to extract the quantized observations of this cluster. Based on these observations it decodes the channel codewords of the next cluster in turn, which allows for extracting the corresponding observations again. The destination cluster proceeds like this, until it decodes the source node channel codewords based on the quantized observations of the relay nodes in the cluster succeeding it.

In this paper we show that both the quantize & forward and the compress & forward strategy enable linear sum-capacity scaling in $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ in the limit $L \to \infty$, when the SNR at the destination antennas is kept constant and $n_{\mathcal{R}} \in \Omega(\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\})$. Note in particular, that a linear growth of $n_{\mathcal{R}}$ with $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ is sufficient, which constitutes a grave advantage over amplify & forward based relaying. Moreover, we show that the additional Slepian & Wolf compression step in the compress & forward strategy allows to reduce the required transmit power scaling in source and relay clusters from exponential to linear in L.

It should be emphasized that unlike references [1], [2], [3] this work is limited to slow fading, although the results are likely to carry over to the fast fading/ergodic case.

II. Notation

The superscripts ^{H} and ^{T} stand for conjugate transpose and transpose, respectively. E_A denotes the expectation operator

[1] We use the Landau notation for characterizing the asymptotic behavior of some function $f(\cdot)$ according to

$$f(n) \in \mathcal{O}(g(n)) \text{ if } \exists M, n_0 > 0 : M|g(n)| > |f(n)|, \forall n \geq n_0,$$
$$f(n) \in \Omega(g(n)) \text{ if } \exists M, n_0 > 0 : M|g(n)| < |f(n)|, \forall n > n_0.$$

with respect to the random variable A. $|\mathbf{A}|$, $\text{Tr}[\mathbf{A}]$ and $\lambda_i\{\mathbf{A}\}$ stand for determinant, trace and the ith eigenvalue of the matrix $\mathbf{A}$. $\|\mathbf{a}\|$ denotes the Euclidean norm of the vector $\mathbf{a}$. By $\mathsf{Pr}[A]$ we denote the probability of the event A. Throughout the paper all logarithms are to the base 2.

III. Signal Model & Protocol

A cluster of $n_\mathcal{S}$ non-cooperating single antenna source nodes, $\mathcal{S}$, aims to transmit data to a fully cooperating cluster of destination nodes $\mathcal{D}$ equipped with $n_\mathcal{D}$ antennas in total. Communication is assumed to occur via L clusters of $n_\mathcal{R}$ non-cooperating single antenna relay nodes each. That is the signal arrives at the destination via $L+1$ hops in just as many time slots. The individual clusters of relay nodes are labeled by $\mathcal{R}_1, \ldots, \mathcal{R}_L$. Cluster $\mathcal{R}_1$ denotes the one next to the destination cluster, $\mathcal{R}_L$ the one next to the sources (cf. Fig. 1). Moreover, the kth antenna in a cluster is labeled by $\mathcal{S}^{(k)}$, $\mathcal{R}_l^{(k)}$ or $\mathcal{D}^{(k)}$. We assume the $L+1$ single hop channels between source, relay and and destination clusters to be frequency-flat fading over the bandwidth of interest. Transmission is divided into $L+1$ time slots of n symbol durations each. The transmission of the ith symbol within the jth time slot, $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, L+1\}$, is described in terms of

- the source transmit vector
$$\mathbf{x}_\mathcal{S}^{(i)} = \left[x_{\mathcal{S}^{(1)}}^{(i)}, \ldots, x_{\mathcal{S}^{(n_\mathcal{S})}}^{(i)}\right]^T,$$
if $j = 1$, or the lth relay cluster transmit vector
$$\mathbf{x}_{\mathcal{R}_l}^{(i)} = \left[x_{\mathcal{R}_l^{(1)}}^{(i)}, \ldots, x_{\mathcal{R}_l^{(n_\mathcal{R})}}^{(i)}\right]^T,$$
if $j = L - l + 2$ and $1 \leq l < L$,
- the lth relay cluster receive vector
$$\mathbf{y}_{\mathcal{R}_l}^{(i)} = \left[y_{\mathcal{R}_l^{(1)}}^{(i)}, \ldots, y_{\mathcal{R}_l^{(n_\mathcal{R})}}^{(i)}\right]^T,$$
if $j = L - l + 1$ and $1 \leq l \leq L$, or the destination receive vector
$$\mathbf{y}_\mathcal{D}^{(i)} = \left[y_{\mathcal{D}^{(1)}}^{(i)}, \ldots, y_{\mathcal{D}_l^{(n_\mathcal{D})}}^{(i)}\right]^T,$$
if $j = L + 1$,
- the lth relay cluster noise vector
$$\mathbf{w}_{\mathcal{R}_l}^{(i)} = \left[w_{\mathcal{R}_l^{(1)}}^{(i)}, \ldots, w_{\mathcal{R}_l^{(n_\mathcal{R})}}^{(i)}\right]^T,$$
if $j = L - l + 1$ and $1 \leq l \leq L$, or the destination cluster noise vector
$$\mathbf{w}_\mathcal{D}^{(i)} = \left[w_{\mathcal{D}^{(1)}}^{(i)}, \ldots, w_{\mathcal{D}_l^{(n_\mathcal{D})}}^{(i)}\right]^T,$$
if $j = L + 1$,
- the matrix $\mathbf{H}_{\mathcal{S}\mathcal{R}_L} \in \mathbb{C}^{n_\mathcal{R} \times n_\mathcal{S}}$ containing the fading coefficients between $\mathcal{S}$ and $\mathcal{R}_L$, for $j = 1$, the matrix $\mathbf{H}_{\mathcal{R}_l\mathcal{R}_{l-1}} \in \mathbb{C}^{n_\mathcal{R} \times n_\mathcal{R}}$ containing the fading coefficient between $\mathcal{R}_l$ and $\mathcal{R}_{l-1}$, for $j = L - l + 2$ and $1 \leq l < L$, or the matrix $\mathbf{H}_{\mathcal{R}_L\mathcal{D}} \in \mathbb{C}^{n_\mathcal{D} \times n_\mathcal{R}}$ containing the channel coefficients between $\mathcal{R}_L$ and $\mathcal{D}$ for $j = L + 1$,

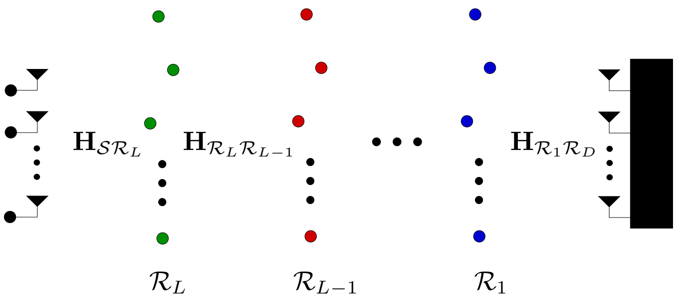

Fig. 1. $n_\mathcal{S}$ non-cooperating source antennas transmit to a fully cooperating destination cluster with $n_\mathcal{D}$ antennas via L clusters of $n_\mathcal{R}$ non-cooperating relay antennas.

and the affine mappings

- $f_1 : \mathbb{C}^{n_\mathcal{S}} \longrightarrow \mathbb{C}^{n_\mathcal{R}}$
$$\mathbf{y}_{\mathcal{R}_L}^{(i)} = \mathbf{H}_{\mathcal{S}\mathcal{R}_L}\mathbf{x}_\mathcal{S}^{(i)} + \mathbf{w}_{\mathcal{R}_L}^{(i)},$$
if $j = 1$,
- $f_l : \mathbb{C}^{n_\mathcal{R}} \longrightarrow \mathbb{C}^{n_\mathcal{R}}$
$$\mathbf{y}_{\mathcal{R}_l}^{(i)} = \mathbf{H}_{\mathcal{R}_{l+1}\mathcal{R}_l}\mathbf{x}_{\mathcal{R}_{l+1}}^{(i)} + \mathbf{w}_{\mathcal{R}_l}^{(i)},$$
if $j = L - l + 1$ and $1 \leq l < L$,
- $f_{L+1} : \mathbb{C}^{n_\mathcal{R}} \longrightarrow \mathbb{C}^{n_\mathcal{D}}$
$$\mathbf{y}_\mathcal{D}^{(i)} = \mathbf{H}_{\mathcal{R}_1\mathcal{D}}\mathbf{x}_{\mathcal{R}_1}^{(i)} + \mathbf{w}_\mathcal{D}^{(i)}, \tag{1}$$
if $j = L + 1$.

Note that nodes within the same cluster are assumed to be symbol synchronized.

IV. Effective I-O Relation

The sequence of source transmit vectors $(\mathbf{x}_\mathcal{S}^{(1)}, \ldots, \mathbf{x}_\mathcal{S}^{(n)})$ is determined by the $n_\mathcal{S}$ messages the source nodes aim to convey to the destination cluster. More specifically, the sequence of kth elements $(x_{\mathcal{S}^{(k)}}^{(1)}, \ldots, x_{\mathcal{S}^{(k)}}^{(n)})$ corresponds to the codeword representing the message of the kth source node. The messages of the $n_\mathcal{S}$ source nodes are assumed to be statistically independent.

The transmit vector sequence of the lth relay cluster $(\mathbf{x}_{\mathcal{R}_l}^{(1)}, \ldots, \mathbf{x}_{\mathcal{R}_l}^{(n)})$ is generated upon reception of all n symbol transmissions from the preceding cluster. Each relay node maps its receive sequence onto a transmit sequence that is chosen from its channel codebook $\mathcal{C}_l^{(k)}$. For the kth relay in $\mathcal{R}_l$ that is the transmit sequence is generated by some relay node specific function $g_l^{(k)} : \mathbb{C}^n \longrightarrow \mathcal{C}_l^{(k)}$
$$\left(x_{\mathcal{R}_l^{(k)}}^{(1)}, \ldots, x_{\mathcal{R}_l^{(k)}}^{(n)}\right) = g_l^{(k)}\left(\left(y_{\mathcal{R}_l^{(k)}}^{(1)}, \ldots, y_{\mathcal{R}_l^{(k)}}^{(n)}\right)\right).$$

The decoding in the destination cluster is performed in a successive fashion. In a first step, the messages sent by the relay nodes in $\mathcal{R}_1$ are decoded based on the receive vector sequence $(\mathbf{y}_\mathcal{D}^{(1)}, \ldots, \mathbf{y}_\mathcal{D}^{(n)})$ and the knowledge of $\mathbf{H}_{\mathcal{R}_1\mathcal{D}}$, which is required for standard coherent MIMO decoding according to (1).

Since the functions $g_1^{(k)}$ are not injective, and thus not invertible in general, it is impossible to obtain a perfect reconstruction of the receive vector sequence $(\mathbf{y}_{\mathcal{R}_1}^{(1)}, \dots, \mathbf{y}_{\mathcal{R}_1}^{(n)})$. Due to this ambiguity an additive quantization noise vector $(\mathbf{q}_{\mathcal{R}_1}^{(1)}, \dots, \mathbf{q}_{\mathcal{R}_1}^{(n)})$ is added, such that the ith element in the estimated receive vector sequence of $\mathcal{R}_1$ is written in terms of the ith element in the corresponding transmit vector sequence $\mathbf{x}_{\mathcal{R}_2}^{(i)}$ as

$$\hat{\mathbf{y}}_{\mathcal{R}_1}^{(i)} = \mathbf{H}_{\mathcal{R}_2\mathcal{R}_1}\mathbf{x}_{\mathcal{R}_2}^{(i)} + \mathbf{w}_{\mathcal{R}_1}^{(i)} + \mathbf{q}_{\mathcal{R}_1}^{(i)}.$$

The decoder continues with decoding the messages sent by $\mathcal{R}_2$ based on this estimated receive vector sequence and the knowledge of $\mathbf{H}_{\mathcal{R}_2\mathcal{R}_1}$. Proceeding like this allows to trace back through the relay chain cluster by cluster based on the estimated receive vector sequences and the knowledge of the respective channel matrix. In the lth iteration, the decoder obtains an estimate of the receive vector sequence $(\mathbf{y}_{\mathcal{R}_l}^{(1)}, \dots, \mathbf{y}_{\mathcal{R}_l}^{(n)})$ by decoding the messages of $\mathcal{R}_l$. The effective input-output relation between $\mathcal{R}_{l+1}$ and $\mathcal{D}$ can thus be written as

$$\hat{\mathbf{y}}_{\mathcal{R}_l}^{(i)} = \mathbf{H}_{\mathcal{R}_{l+1}\mathcal{R}_l}\mathbf{x}_{\mathcal{R}_{l+1}}^{(i)} + \mathbf{w}_{\mathcal{R}_l}^{(i)} + \mathbf{q}_{\mathcal{R}_l}^{(i)}.$$

where $\mathbf{q}_{\mathcal{R}_l}^{(i)}$ is the ith element of the respective quantization noise vector sequence. In the $L+1$st iteration the decoder finally arrives at the source cluster, whose messages are decoded based on the effective input-output relation

$$\hat{\mathbf{y}}_{\mathcal{R}_L}^{(i)} = \mathbf{H}_{\mathcal{S}\mathcal{R}_L}\mathbf{x}_{\mathcal{S}}^{(i)} + \mathbf{w}_{\mathcal{R}_L}^{(i)} + \mathbf{q}_{\mathcal{R}_L}^{(i)},$$

which is equivalent to the input-output relation of a point-to-point MIMO channel with an additive noise term $\mathbf{w}_{\mathcal{R}_L}^{(i)} + \mathbf{q}_{\mathcal{R}_L}^{(i)}$.

V. Power Constraints, Fading & Channel State Information

We assign an average transmit power P to each transmitting cluster. Inside the cluster this transmit power is uniformly distributed over all nodes. That is, we impose per node power constraints on source and relays nodes as follows:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left|x_{\mathcal{S}^{(k)}}^{(i)}\right|^2 \le P/n_{\mathcal{S}},$$
$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left|x_{\mathcal{R}_l^{(k)}}^{(i)}\right|^2 \le P/n_{\mathcal{R}}.$$

The parameter P is a function of L and chosen such that the SNR at the destination antennas remains constant with respect to L. It is not crucial for our result that all clusters transmit at the same power P. It holds true for any cluster power allocation that ensures the required SNR at the destination antennas. However, it is crucial that all nodes within the same cluster transmit at the same average power.

The elements of the vectors $\mathbf{w}_{\mathcal{R}_l}^{(i)}$ and $\mathbf{w}_{\mathcal{D}}^{(i)}$ representing the thermal noise introduced at the receiver front-ends are assumed to be i.i.d. (both in space and time) zero-mean circularly symmetric complex Gaussian random variables of variance σ_w^2. The elements of the channel matrices $\mathbf{H}_{\mathcal{S}\mathcal{R}_L}$, $\mathbf{H}_{\mathcal{R}_l\mathcal{R}_{l-1}}$ and $\mathbf{H}_{\mathcal{R}_1\mathcal{D}}$ are assumed to be i.i.d. according to an arbitrary distribution with zero-mean and unit variance. We consider a block fading model, i.e. the channel realization remains constant over one block of several symbol durations. This block length is assumed to be larger than the codeword length (slow fading).

The destination cluster is assumed to know all realizations of these random matrices in order to be able to perform coherent decoding. For the relay clusters we distinguish two cases. In the first case, the relay nodes do not have any channel state information, in the second case each node in relay cluster $\mathcal{R}_l$ is assumed to know the full channel matrix of its preceding hop. The source node cluster is assumed not to possess any channel state information.

VI. Capacity Scaling

A. Quantize & Forward

We derive achievable rates for the transmission from $\mathcal{S}$ or $\mathcal{R}_l$ to $\mathcal{D}$ by rendering the elements of the $\mathbf{q}_l^{(i)}$ i.i.d. (both in space and time) circularly symmetric complex Gaussian random variables with zero-mean and variance σ_l^2. We denote the set of nodes in $\mathcal{R}_l$ whose quantization noises follow such a distribution by $\mathcal{R}_l^*$. The cardinality of $\mathcal{R}_l^*$ is denoted by $n_{\mathcal{R}_l^*}$. According to [5] achievable rates[2] (normalized to the number of destination antennas) are then given by

$$R_{\mathcal{S}} = \frac{1}{n_{\mathcal{D}}}\cdot\log\left|\mathbf{I}_{n_{\mathcal{R}_L^*}} + \frac{P}{n_{\mathcal{S}}\cdot(\sigma_w^2+\sigma_L^2)}\cdot\mathbf{H}_{\mathcal{S}\mathcal{R}_L^*}\mathbf{H}_{\mathcal{S}\mathcal{R}_L^*}^H\right|$$

for the transmission from source to destination cluster and

$$R_l = \begin{cases} \frac{1}{n_{\mathcal{D}}}\cdot\log\left|\mathbf{I}_{n_{\mathcal{R}_{l-1}^*}} + \frac{P}{n_{\mathcal{R}_l^*}\cdot\left(\sigma_w^2+\sigma_{l-1}^2\right)}\cdot\mathbf{H}_{\mathcal{R}_l^*\mathcal{R}_{l-1}^*}\mathbf{H}_{\mathcal{R}_l^*\mathcal{R}_{l-1}^*}^H\right|, \\ \qquad\qquad \text{if } 1 < l \le L, \\ \frac{1}{n_{\mathcal{D}}}\cdot\log\left|\mathbf{I}_{n_{\mathcal{D}}} + \frac{P}{n_{\mathcal{R}_1^*}\cdot\sigma_w^2}\cdot\mathbf{H}_{\mathcal{R}_1^*\mathcal{D}}\mathbf{H}_{\mathcal{R}_1^*\mathcal{D}}^H\right|, \\ \qquad\qquad \text{if } l = 1, \end{cases} \tag{2}$$

for transmission from lth relay to destination cluster, where we discarded all $\mathcal{R}_l^{(k)} \notin \mathcal{R}_l^*$ each. In order to achieve these rates, the source nodes need to generate their channel codebooks by choosing their entries $x_{\mathcal{S}^{(k)}}^{(i)}$ as independent realizations of a zero-mean circularly symmetric complex Gaussian random variable $X_{\mathcal{S}}$. The variance of this random variable is chosen to fulfill the average power constraint, i.e. $\mathsf{E}_{X_{\mathcal{S}}}[|X_{\mathcal{S}}|^2] = P/n_{\mathcal{S}}$. Similarly, the relay nodes generate their channel codebooks by choosing their entries $x_{\mathcal{R}_l^{(k)}}^{(i)}$ as independent realizations of a zero-mean circularly symmetric complex Gaussian random variable $X_{\mathcal{R}}$. The variance of this random variable is again chosen to fulfill the average power constraint, i.e. $\mathsf{E}_{X_{\mathcal{R}}}[|X_{\mathcal{R}}|^2] = P/n_{\mathcal{R}}$.

In order to prove that the end-to-end sum-capacity scales linearly in $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ in the limit $L \to \infty$, it is thus

[2] We ignore the pre-log factor due to the use of several time slots here.

sufficient to show that for every L there exists P such that the SNR at the destination antennas, given by $\mathrm{snr}_{\mathcal{D}} = P/(\sigma_w^2 + \sigma_L^2)$, is kept constant, while the cardinality of the set $\mathcal{R}_L^*$ grows linearly with $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$. We prove this in the following for the special case

$$\frac{n_{\mathcal{R}}}{n_{\mathcal{S}}} \xrightarrow{n_{\mathcal{S}}\to\infty} 1 \text{ and } \frac{n_{\mathcal{D}}}{n_{\mathcal{S}}} \xrightarrow{n_{\mathcal{S}}\to\infty} 1.$$

This special case immediately implies that linear sum-capacity scaling in $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ is achieved for $n_{\mathcal{R}} \in \Omega(\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\})$.

By a "quantize & forward" strategy we mean that each relay node performs the quantization of its receive signal without exploiting the statistical correlation with the received signals of other relay nodes within the same cluster. More specifically, the mapping $g_l^{(k)}$ is specified as follows: Relay node $\mathcal{R}_l^{(k)}$ generates a quantization codebook containing 2^{nR_l} codewords. The entries in the quantization codebook are generated as independent realizations of a zero-mean circularly symmetric complex Gaussian random variable $\hat{Y}$ of variance $Q_l^{(k)} + \sigma_w^2 + \sigma_l^2$, where $Q_l^{(k)}$ denotes the received signal power of node $\mathcal{R}_l^{(k)}$. A codeword is declared to be the quantization of the observation sequence, if both are jointly typical. Under the constraint that a quantization codeword for the observation vector is found based on a joint typicality check with probability one as $n \to \infty$, the mutual information between original and quantized observation of node $\mathcal{R}_l^{(k)}$ must fulfill

$$R_l \geq I(y_{\mathcal{R}_l^{(k)}}; \hat{y}_{\mathcal{R}_l^{(k)}}) = \log\left(1 + \frac{Q_l^{(k)} + \sigma_w^2}{\sigma_l^2}\right).$$

Solving this equation for σ_l^2 yields

$$\sigma_l^2 \geq \frac{Q_l^{(k)} + \sigma_w^2}{2^{R_l} - 1}. \tag{3}$$

Let us now fix the quantization noise variance according to

$$\sigma_l^2 = \frac{P + \varepsilon + \sigma_w^2}{2^{R_l} - 1}, \tag{4}$$

for some $\varepsilon > 0$. Then, the set $\mathcal{R}_l^*$ consists of all nodes in $\mathcal{R}_l$ that fulfill (3). Denoting the kth rows of $\mathbf{H}_{\mathcal{S}\mathcal{R}_L}$ and $\mathbf{H}_{\mathcal{R}_{l+1}\mathcal{R}_l}$ by $\mathbf{h}_{\mathcal{S}\mathcal{R}_L}^{(k)T}$ and $\mathbf{h}_{\mathcal{R}_{l+1}\mathcal{R}_l}^{(k)T}$, respectively, and explicitly writing down $Q_l^{(k)}$, that is

$$\mathcal{R}_l^*(\varepsilon) = \begin{cases} \left\{\mathcal{R}_l^{(k)} : \left|\frac{P}{n_{\mathcal{S}}} \cdot \|\mathbf{h}_{\mathcal{S}\mathcal{R}_L}^{(k)T}\|^2 - P\right| < \varepsilon\right\}, \\ \qquad\qquad\qquad \text{if } l = L, \\ \left\{\mathcal{R}_l^{(k)} : \left|\frac{P}{n_{\mathcal{R}_{l+1}^*}} \cdot \|\mathbf{h}_{\mathcal{R}_{l+1}^*\mathcal{R}_l}^{(k)T}\|^2 - P\right| < \varepsilon\right\}. \\ \qquad\qquad\qquad \text{if } 1 \leq l < L. \end{cases}$$

Next, we study the cardinality of the sets $\mathcal{R}_l^*$, $l \in \{1, \ldots, L\}$. Whether $\mathcal{R}_l^{(k)} \in \mathcal{R}_l^*(\varepsilon)$ or not, corresponds to a Bernoulli experiment whose success probability $p(\varepsilon)$ is lower bounded according to Chebychev's inequality:

$$p(\varepsilon) \triangleq \Pr\left(\mathcal{R}_l^{(k)} \in \mathcal{R}_l^*(\varepsilon)\right) \geq \begin{cases} 1 - \frac{P^2}{n_{\mathcal{S}}\varepsilon^2}, & \text{if } l = L, \\ 1 - \frac{P^2}{n_{\mathcal{R}_l^*}\varepsilon^2}, & \text{if } 1 \leq l < L. \end{cases}$$

The cardinality $n_{\mathcal{R}_L^*}$ of the set $\mathcal{R}_L^*(\varepsilon)$ is identified with the number of successes in the Bernoulli experiment and thus Binomial distributed. The probability that $\mathcal{R}_L^*(\varepsilon)$ has cardinality smaller k is therefore upper-bounded according to Hoeffding's inequality [6] by

$$\Pr\left(n_{\mathcal{R}_L^*} < k\right) \leq \exp\left(-\frac{2}{n_{\mathcal{S}}} \cdot (n_{\mathcal{S}} \cdot p(\varepsilon) - k)^2\right).$$

This upper-bound goes to zero as $n_{\mathcal{S}} \to \infty$ if we set $k = n_{\mathcal{S}} - n_{\mathcal{S}}^{\alpha}$ for some $\alpha \in (0,1)$. This shows that $n_{\mathcal{R}_L^*}/n_{\mathcal{R}} \xrightarrow{n_{\mathcal{S}}\to\infty} 1$ in probability. As the probability that $n_{\mathcal{R}_l^*} < n_{\mathcal{S}} - n_{\mathcal{S}}^{\alpha}$ decreases exponentially in $n_{\mathcal{S}}$, the Borel-Cantelli Lemma allows to conclude that this convergence also holds almost surely.

Tracing through the relay chain by repeatedly applying Hoeffding's inequality and the Borel-Cantelli Lemma yields that $n_{\mathcal{R}_l^*}/n_{\mathcal{R}} \xrightarrow{n_{\mathcal{S}}\to\infty} 1$ almost surely for all $l \in \{1, \ldots, L\}$. Consequently, for every arbitrarily large (but fixed) L, we have that

$$\Pr\left(\lim_{n_{\mathcal{S}}\to\infty} \max_{l\in\{1,\ldots,L\}} \left|\frac{n_{\mathcal{R}_l^*}}{n_{\mathcal{R}}} - 1\right| = 0\right) = 1,$$

which implies that the cardinalities of all sets $\mathcal{R}_l$ grow linearly in $n_{\mathcal{S}}$ almost surely.

It remains to verify that any SNR value, $\mathrm{snr}_{\mathcal{D}}$, can be sustained at the destination antennas for increasing L by increasing P appropriately. Under the assumption that $n_{\mathcal{R}}/n_{\mathcal{S}} \xrightarrow{n_{\mathcal{S}}\to\infty} 1$ and $n_{\mathcal{D}}/n_{\mathcal{S}} \xrightarrow{n_{\mathcal{S}}\to\infty} 1$, the rates (2) converge to [7]

$$R_l^{(\infty)} = \begin{cases} 2\log\left(1 + \frac{P}{\sigma_w^2+\sigma_{l-1}^2} - \frac{1}{4}\left(\sqrt{\frac{4P}{\sigma_w^2+\sigma_{l-1}^2} + 1} - 1\right)^2\right) \\ \quad - \frac{\log e}{\frac{4P}{\sigma_w^2+\sigma_{l-1}^2}}\left(\sqrt{\frac{4P}{\sigma_w^2+\sigma_{l-1}^2} + 1} - 1\right)^2, \\ \qquad\qquad\qquad \text{if } 1 < l \leq L, \\ 2\log\left(1 + \frac{P}{\sigma_w^2} - \frac{1}{4}\left(\sqrt{\frac{4P}{\sigma_w^2} + 1} - 1\right)^2\right) \\ \quad - \frac{\log e}{\frac{4P}{\sigma_w^2}}\left(\sqrt{\frac{4P}{\sigma_w^2} + 1} - 1\right)^2, \text{ if } l = 1, \end{cases} \tag{5}$$

almost surely in the limit $n_{\mathcal{S}} \to \infty$ and, moreover, for arbitrarily large, but fixed L, we have that

$$\Pr\left(\lim_{n_{\mathcal{S}}\to\infty} \max_{l\in\{1,\ldots,L\}} \left|R_l - R_l^{(\infty)}\right| = 0\right) = 1.$$

Substituting the asymptotic rates (5) into (4) yields a first order difference equation in σ_l^2. This difference equation simplifies substantially in the limit $P \to \infty$:

$$\lim_{P\to\infty} \sigma_l^2 = \begin{cases} e \cdot \left(\sigma_{l-1}^2 + \sigma_w^2 + \varepsilon\right), & \text{if } l > 1, \\ e \cdot \left(\sigma_w^2 + \varepsilon\right), & \text{if } l = 1. \end{cases}$$

which firstly reveals an exponential growth of σ_L^2 in L, and secondly proves that for L fixed σ_L^2 is bounded as $P \to \infty$. Thus arbitrary SNR values can be achieved at the destination antennas by increasing the transmit power per source and relay cluster.

B. Compress & Forward

By "compress & forward" we denote the forwarding strategy that performs a Slepian & Wolf compression upon the quantization of the receive signal, thus exploiting the correlation in the receive signals of relay nodes within the same cluster. The quantization is performed completely analogous as in the quantize & forward case. In particular, the quantization noise is again i.i.d. zero-mean circularly symmetric complex Gaussian. The compression problem at hand has been studied in [8] in the context of a two-hop setup with orthogonal second hop. The quantization vectors need to fulfill

$$n_{\tilde{\mathcal{R}}_l} R_l \geq I\left(\mathbf{y}_{\mathcal{R}_l}; \hat{\mathbf{y}}_{\tilde{\mathcal{R}}_l} \middle| \hat{\mathbf{y}}_{\tilde{\mathcal{R}}_l^C}\right) \ \forall \tilde{\mathcal{R}}_l \subseteq \mathcal{R}_l,$$

where the vector $\hat{\mathbf{y}}_{\tilde{\mathcal{R}}_l}$ contains the quantized observations of the relay nodes in $\tilde{\mathcal{R}}_l$ and $n_{\tilde{\mathcal{R}}_l}$ denotes the cardinality of $\tilde{\mathcal{R}}_l$. Due to symmetry and channel hardening (we skip a detailed proof), the dominating constraint for a common quantization noise power σ_l^2 in the limit $n_{\mathcal{S}} \to \infty$ is the one for $\tilde{\mathcal{R}}_l = \mathcal{R}_l$ for sufficiently large P (or L equivalently) with probability one:

$$\begin{aligned} R_l &\geq \frac{I\left(\mathbf{y}_{\mathcal{R}_l}; \hat{\mathbf{y}}_{\mathcal{R}_l}\right)}{n_{\mathcal{R}}} \\ &= \begin{cases} \frac{1}{n_{\mathcal{R}}} \cdot \log \left| \left(1 + \frac{\sigma_w^2}{\sigma_l^2}\right) \cdot \mathbf{I}_{n_{\mathcal{R}}} + \frac{P}{n_{\mathcal{R}} \sigma_l^2} \mathbf{H}_{\mathcal{R}_{l+1}\mathcal{R}_l} \mathbf{H}_{\mathcal{R}_{l+1}\mathcal{R}_l}^H \right|, & \text{if } l < L, \\ \frac{1}{n_{\mathcal{R}}} \cdot \log \left| \left(1 + \frac{\sigma_w^2}{\sigma_L^2}\right) \cdot \mathbf{I}_{n_{\mathcal{R}}} + \frac{P}{n_{\mathcal{S}} \sigma_L^2} \mathbf{H}_{\mathcal{S}\mathcal{R}_L} \mathbf{H}_{\mathcal{S}\mathcal{R}_L}^H \right|, & \text{if } l = L. \end{cases} \end{aligned}$$

In the limit $n_{\mathcal{S}} \to \infty$ both expressions on the right hand side converge almost surely to

$$\begin{aligned} \xi_l &= \int_0^\infty \log\left(1 + \frac{P \cdot x + \sigma_w^2}{\sigma_l^2}\right) f_\lambda(x) \cdot dx \\ &= \int_0^\infty \log\left(1 + \frac{P \cdot x}{\sigma_w^2 + \sigma_l^2}\right) f_\lambda(x) \cdot dx + \log\left(1 + \frac{\sigma_w^2}{\sigma_l^2}\right) \\ &= 2\log\left(1 + \frac{P}{\sigma_w^2 + \sigma_l^2} - \frac{1}{4}\left(\sqrt{\frac{4P}{\sigma_w^2 + \sigma_l^2} + 1} - 1\right)^2\right) \\ &\quad - \frac{\log e}{\frac{4P}{\sigma_w^2 + \sigma_l^2}}\left(\sqrt{\frac{4P}{\sigma_w^2 + \sigma_l^2} + 1} - 1\right)^2 + \log\left(1 + \frac{\sigma_w^2}{\sigma_l^2}\right), \end{aligned}$$

where $f_\lambda(\cdot)$ denotes the asymptotic eigenvalue distribution of the matrices $\mathbf{H}_{\mathcal{S}\mathcal{R}_L} \mathbf{H}_{\mathcal{S}\mathcal{R}_L}^H$ and $\mathbf{H}_{\mathcal{R}_{l+1}\mathcal{R}_l} \mathbf{H}_{\mathcal{R}_{l+1}\mathcal{R}_l}^{(i)H}$, which is given by the Marčenko-Pastur law [9]. The solution to the integral is found in [7]. Moreover, we have for arbitrary large, but fixed L, that

$$\Pr\left(\lim_{n_{\mathrm{S}} \to \infty} \max_{l \in \{1,\ldots,L\}} \left| \frac{I(\mathbf{y}_{\mathcal{R}_l}; \hat{\mathbf{y}}_{\mathcal{R}_l})}{n_{\mathcal{R}}} - \xi_l \right| = 0\right) = 1.$$

Again, we substitute the asymptotic expression (5) for R_l. Taking the limit $P \to \infty$ in the resulting (implicit) first order difference equation simplifies in the limit $P \to \infty$ to

$$\lim_{P \to \infty} \sigma_l^2 = \begin{cases} \sigma_{l-1}^2 + \sigma_w^2, & \text{if } l > 1, \\ \sigma_w^2, & \text{if } l = 1. \end{cases}$$

We observe that the Slepian & Wolf compression reduces the scaling of the cluster transmit power P with L from exponential to linear, when a fixed SNR at the destination antennas is to be sustained.

VII. CONCLUSIONS

We have shown that – unlike the amplify & forward strategy – quantization based relay strategies enable linear sum-capacity scaling in $\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}$ in the limit $L \to \infty$ for a constant SNR at the destination antennas and $n_{\mathcal{R}} \in \Omega\left(\min\{n_{\mathcal{S}}, n_{\mathcal{D}}\}\right)$. Moreover, Slepian & Wolf compression provides a drastic reduction of the transmit power that is required to sustain a constant SNR at the destination antennas.

REFERENCES

[1] S. Yeh and O. Lévêque, "Asymptotic capacity of multi-level amplify-and-forward relay networks," in *Proc. IEEE Int. Symposium on Inf. Theory*, Nice, France, June 2007.

[2] S. Borade, L. Zheng, and R. Gallager, "Amplify and forward in wireless relay networks: Rate, diversity and network size," *IEEE Trans. Inform. Theory, Special Issue on Relaying and Cooperation in Communication Networks*, vol. 53, no. 10, pp. 3302–3318, Oct. 2007.

[3] J. Wagner and A. Wittneben, "On the distortion of the eigenvalue spectrum in MIMO amplify-and-forward multihop channels," submitted to Asilomar Conf. Signals, Syst., Comp.

[4] R. R. Müller, "On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 2086–2091, July 2002.

[5] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[6] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *American Statistical Association Journal*, vol. 58, no. 301, pp. 13–30, 1963.

[7] S. Verdu and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3117–3128, Dec. 2002.

[8] A. Sanderovich, S. Shamai (Shitz), Y. Steinberg, and M. Peleg, "Decentralized receiver in a MIMO system," in *Proc. IEEE Int. Symposium on Inf. Theory*, Seattle, WA, July 2006, pp. 6–10.

[9] V. A. Marcenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.