Modular CSI Quantization for Massive MIMO Communication

Jialing Liao, *Member, IEEE*, Roope Vehkalahti, *Member, IEEE*, Tefjol Pllaha, Wei Han, Olav Tirkkonen, *Senior Member, IEEE*

***Abstract*—We consider high-dimensional MIMO transmissions in frequency division duplexing (FDD) systems. For precoding, the frequency selective channel has to be measured, quantized and fed back to the base station by the users. When the number of antennas is very high this typically leads to prohibitively high quantization complexity and large feedback. In 5G New Radio (NR), a modular quantization approach has been applied for this, where first a low-dimensional subspace is identified for the whole frequency selective channel, and then subband channels are linearly mapped to this subspace and quantized. We analyze how the components in such a modular scheme contribute to the overall quantization distortion. Based on this analysis we improve the technology components in the modular approach and propose an orthonormalized wideband precoding scheme and a sequential wideband feedback approach which provide considerable gains over the conventional method. Efficient subband quantization approaches are discussed as well. We compare the developed quantization schemes to the 5G NR standardized version by simulations in terms of the projection distance, overall distortion and spectral efficiency, respectively, in a scenario with a realistic spatial channel model.**

***Index Terms*—Massive MIMO, FDD, CSI quantization**

I. Introduction

USING very large antenna arrays at the base station (BS) is highly beneficial when serving multiple users in downlink communication [2], and massive MIMO communication is one of the key components in the 5G New Radio (NR). However, the performance of Massive MIMO downlink depends heavily on high quality channel state information (CSI) at the transmitter. In a Frequency Division Duplex (FDD) system the channel has to be measured by the users, quantized, and then fed back to the BS. FDD MIMO finite feedback has been studied intensively both in the single user [3], [4] and multiuser [5], [6] context. When the number of transmit antennas N_t at the BS is high, the complexity of the quantization and the amount of feedback needed can be prohibitively high. This problem becomes particularly difficult in a multiuser-MIMO setting, where the amount of feedback should scale with the Signal-to-Noise Ratio (SNR) [5].

J. Liao, and O. Tirkkonen are with the Department of Communications and Networking (Comnet), Aalto University, Finland (e-mail: jialing.liao@ieee.org, olav.tirkkonen@aalto.fi).

Roope Vehkalahti is with the Department of Mathematics and Statistics, University of Jyväskylä, Finland (e-mail: roope.i.vehkalahti@jyu.fi).

Tefjol Pllaha is with the Department of Mathematics, University of Nebraska–Lincoln, United States (e-mail: tefjol.pllaha@unl.edu).

Wei Han is with Huawei Technologies Co., Ltd., Shanghai, P. R. China, (e-mail: wayne.hanwei@huawei.com).

This work was funded in part by Huawei Technologies Co., Ltd., and the Academy of Finland (grant 319484). Part of the paper has been published in VTC Spring 2021 [1].

Back to the traditional channel feedback techniques, which utilize pre-defined codebooks directly to quantize and feedback the channel vector [7], [8], the size of the codebook increases with the number of antennas that prevents their application in massive MIMO networks [5]. To this end, the dimensionality of the effective channel for CSI feedback in massive MIMO systems needs to be reduced. The authors in [9] studied improved feedback schemes for FDD massive MIMO utilizing source coding techniques with only a small codebook. Besides, compressive sensing was introduced to this field in [10] utilizing the sparsity of the massive MIMO channel. Moreover, authors in [11] addressed the problem from a signal processing perspective by utilizing rate splitting encoding strategies at the transmitter to guarantee the robustness under limited CSI feedback. In [12], the authors proposed a cooperative feedback approach for device-to-device (D2D) networks by enabling CSI exchange among users. Later on, a modular CSI feedback framework was utilized in [6], [13]–[15] to compress the dimension of CSI feedback utilizing the low-rank property of the channel covariance matrices in Massive MIMO scenarios, which share some similarities to our work so that will be discussed later.

More recently, there is a significant tendency for applying deep learning (DL) in CSI feedback for massive MIMO networks [16]–[18]. Compared to the conventional vector quantized CSI feedback which imposes an overhead that grows linearly with system dimensions, and also depends on channel sparsity and iterative solutions, the DL-based CSI feedback overcomes the drawbacks of high complexity, latency and limited accuracy in capturing the channel structure. [16] introduced a DL-based CSI compression framework while [17] proposed a DL-based CSI reduction scheme with convolutional layers followed by quantization and entropy coding blocks. [18] considered CSI quantization for a neural network using temporal correlations in time-varying channels. This shall be a future direction of CSI quantization to benefit from the intelligence of DL techniques. However, the application of DL in CSI feedback is more likely to be on the implementations so that it is less measurable. To make full usage of the techniques, theoretical research on the feedback framework, quantization objectives and distortion analysis, which motivates our work, is needed as a boundary for exploiting DL-based CSI feedback.

A. Related Work

Of relevance to our work are [6], [13]–[15], [19]–[22] where they investigate CSI feedback schemes for wireless

cellular networks either utilizing a similar approach to our work, or focusing on the same topic of covariance eigenspace quantization of massive MIMO networks.

In [6] a modular approach was suggested to this problem, which benefited from the fact that while the channel can vary fast, the correlation between the antennas can stay stable for relatively many samples. If an individual $N_t \times 1$ channel vector $\mathbf{h}$ is correlated, the majority of channel energy lives in a low-dimensional subspace of $\mathbb{C}^{N_t}$ and therefore it is sufficient to feedback the coordinates of the signal in this subspace. With K the subspace dimensionality, this can be done by selecting an $N_t \times K$ unitary matrix $\mathbf{U}_K$ based on the covariance matrix of $\mathbf{h}$, and creating a K-dimensional *effective channel* vector by $\mathbf{c} = \mathbf{U}_K^{\mathrm{H}}(\mathbf{h})$. If $\mathbf{U}_K$ is known to the BS, the user can simply quantize $\mathbf{c}$ and feed back this data. The BS can then approximate the channel vector as $\mathbf{U}_K\mathbf{c}$. If $K \ll N_t$, this considerably reduces both quantization complexity and feedback rate. This work was then extend to deal with large-system regime where both antennas and users growing large [13]. Users are grouped so that the eigenspaces of different groups are near-orthogonal. Two-stage downlink precoding was used by designing pre-beamforming and multiuser MIMO linear precoding matrices. The same idea was also applied to time varying channels and interference networks [14], [15].

To use the modular approach in [6], the user has to quantize and feed back the basis matrix $\mathbf{U}_K$ using some codebook $\mathcal{C}_{\mathbf{W}}$. MIMO covariance matrix, and covariance eigenspace quantization has been considered in [19]–[22]. In [19], it was shown that preserving orthogonality after feedback quantization is optimal. Accordingly, matrix codebooks consisting of a collection unitary matrices is considered. In contrast, [20]–[22] consider independent vector quantization of the columns of $\mathbf{U}_K$, i.e. codebooks of the form $\mathcal{C}_{\mathbf{W}} = \mathcal{C}_{\mathbf{w}}^K$, where $\mathcal{C}_{\mathbf{w}}$ is a codebook of $N_t \times 1$ vectors. For a given quantization accuracy, the size of a vector codebook is only $\sqrt{K}$ times the size of a matrix codebook, which significantly reduces quantization complexity. However, when a vector codebook is used, orthogonality of the matrices cannot be guaranteed. In the approach of [19], orthogonality of the K columns in $\mathbf{U}_K$ is always guaranteed. In [20], orthogonality is guaranteed by sequence design, limiting the vector codebook size to N_t.

In order to have high descriptive power with manageable complexity, high-resolution FDD feedback in 5G NR relies on overcomplete vector quantization codebooks [21], [22]. This enables high precision of describing the basis $\mathbf{U}_K$, at the cost of a potential loss of orthogonality. This is true for example with the overcomplete codebooks used in 5G NR type 2 wideband feedback [21]. With a non-orthogonal basis, good subband effective channel quantization might not result in good overall quantization.

In this paper, we first provide an implicit method for feeding back unitary matrices despite using high precision vector codebooks of the form $\mathcal{C}_{\mathbf{w}}^K$. Thus, we may use precisely the same vector codebook and the same number of feedback bits as [21], [22], but guarantee orthogonality of the basis in the fed back matrices. We then prove how the overall quantization distortion of the channel essentially decomposes into two mostly independent parts. One describes the error of quantizing $\mathbf{U}_K$, while the other part is essentially the effective channel distortion. This analysis also provides a new criterion for quantizing $\mathbf{U}_K$ in an optimal way. Moreover in order to improve the accuracy of projecting subband channel vectors into the space generated by wideband fed back matrix, we develop a orthonormalized wideband precoding scheme and a sequential wideband quantization scheme which guarantee improved projection distance and overall distortion.

B. Contributions

In this paper, our aim is to provide a framework of CSI quantization for massive MIMO networks that provides improved quantization distortion with low complexity. In summary, this paper has made the following major contributions:

- We develop a modular CSI quantization framework for massive MIMO networks. The design objectives for both wideband quantization and subband quantization are studied, as well as the typical codebooks and the detailed steps in quantization and reconstruction.
- The quantization distortion is partitioned into two parts, the wideband quantization error (projection distance), and the subband quantization error (effective channel quantization distortion). Analysis states that the impact of wideband quantization quality on the overall distortion surpass that of effective channel quantization, which highlights the importance of improving wideband quantization.
- For wideband quantization, two schemes are proposed and proved to provide considerable gains over the standardized scheme used in [21], i.e., orthonormalized wideband precoding (OWP) scheme and sequential wideband feedback (SWF) scheme. The OWP scheme makes it possible to further increase quantization accuracy by improving subband quantization accuracy, while the sequential approach brings further gain over OWP.
- We discussed the feasibility of utilizing i.i.d. vector quantization for quantizing the subband effective channel. Besides the subband quantization option suggested for the standardized scheme [21], the best options for OWP and SWF are designed specifically.
- Simulations are done in a MIMO scenario with a high number of antennas, when these principles are applied, from different perspectives, i.e., wideband quantization under different polarization structures, overall quantization, and the performance in multiuser networks. The SWF scheme outperforms the OWP scheme while the standardized scheme is the worst.

C. Organization

The organization of the rest of the paper is as follows. Section II introduces the system model used in the paper and Section III presents a modular method combining wideband quantization and subband quantization, with a discussion on the quantization distortion partitioning. Section IV introduces two wideband quantization schemes, i.e. OWP and SWF, that are proved to improve the accuracy of wideband quantization and further provide a better projection basis for subband channels. Section V discusses the intuition and suggested

schemes for subband effective channel quantization. Section VI presents the simulation results of the proposed CSI quantization schemes with a comparison with the standardized method. Section VII summarizes the paper with a conclusion and discussion of main contributions.

D. Notation

$\mathbf{X}^{\mathrm{T}}$ and $\mathbf{X}^{\mathrm{H}}$ denote the transpose and the conjugate transpose (or Hermitian) of matrix $\mathbf{X}$, respectively. The notation $\|\mathbf{x}\|$ refers to the Euclidean norm of the vector $\mathbf{x}$. $\mathrm{Tr}(\mathbf{X})$ denotes the trace of matrix $\mathbf{X}$. The matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]$ is also expressed as $\{\mathbf{x}_k\}_{k=1}^{K}$ for briefness.

II. System Model

A. Network Model

We consider a frequency selective massive MIMO channel in which the transmitter has N_t antennas, and the receiver has a single antenna. We consider Orthogonal Frequency domain Multiplexing (OFDM), and model frequency selectivity on a subband basis, such that there are S subbands for which CSI is gathered. The channel on a subcarrier belonging to subband s is mathematically described as:

$$\mathbf{y}_s = \mathbf{h}_s^T \mathbf{x}_s + \mathbf{n}_s. \tag{1}$$

where $\mathbf{h}_s \in \mathbb{C}^{N_t \times 1}$ is the channel vector of the user. The vector $\mathbf{x}_s \in \mathbb{C}^{N_t \times 1}$ is the transmitted multiantenna signal on the subcarrier, and $\mathbf{n}_s$ is independent complex Gaussian noise with unit variance. The transmit power constraint is $E[\|\mathbf{x}\|^2] \leq P$ with transmit power threshold P.

It is assumed that the channel across subbands is block Rayleigh-faded, so that it is distributed as a complex Gaussian, with a joint distribution arising from the multipath structure of the environment. We assume that the user has perfect knowledge of its own channel. The joint distribution across subbands gives rise to an estimated channel covariance matrix

$$\mathbf{R} = E_{\mathbf{h}}[\mathbf{h}\mathbf{h}^{\mathrm{H}}] = \sum_{s=1}^{S} (\mathbf{h}_s \mathbf{h}_s^{\mathrm{H}}). \tag{2}$$

Thus the CSI can be completely captured by the covariance matrix $\mathbf{R}$ which describes the *wideband characteristics* of the channel, and the sub-band specific coordinates describing the sub-band channels in the basis given by $\mathbf{R}$.

In this paper, we are interested in the problem of quantizing the aggregate CSI $\{\mathbf{h}_s\}_{s=1}^{S}$, for feeding it back to the transmitter. We shall exploit the division to wideband and sub-band channels for efficient feedback.

B. Quantization Quality Measurement

In general, the *chordal distance* between two vectors $\mathbf{x}$ and $\mathbf{y}$ given by

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{1 - \frac{|\mathbf{x}^{\mathrm{H}}\mathbf{y}|^2}{\|x\|^2 \|y\|^2}} \tag{3}$$

can be used to measure the quantization quality.

Accordingly a single-user codebook $\hat{\mathcal{H}}$ for quantizing the channel vector $\mathbf{h}_s$ would be good if the minimum squared chordal distance between the channel vector and any codeword in $\hat{\mathcal{H}}$, i.e. $\min_{\hat{\mathbf{h}} \in \hat{\mathcal{H}}} \left\{ d(\mathbf{h}_s, \hat{\mathbf{h}}_s)^2 \right\}$, is small. The expected value of this is nothing but the quantization distortion with respect to the chordal distance given by

$$D(\mathbf{h}_s, \hat{\mathbf{h}}_s) = E_{\mathbf{h}} \left[\min_{\hat{\mathbf{h}} \in \hat{\mathcal{H}}} \left\{ d(\mathbf{h}_s, \hat{\mathbf{h}}_s)^2 \right\} \right]. \tag{4}$$

Our goal is now to develop quantization schemes that would minimize this distortion.

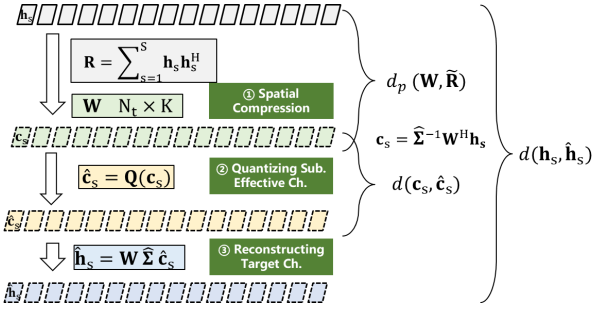

Fig. 1: Flowchart for general process of CSI quantization.

III. Modular Single User CSI Quantization

In this section, the overall process of the developed modular CSI quantization approach will be presented followed by a discussion on the quantization distortion partitioning.

A. Overall Process of Modular CSI Quantization

In order to quantize the CSI for massive MIMO networks equipped with large scale antennas, the *modular channel quantization* [22] process can be performed. To proceed, there are generally three steps, i.e., compression, quantization, and reconstruction, the flowchart of which is presented in Fig. 1. While spatial compression provides a low complexity quantization capturing the wideband information from channel covariance matrix $\mathbf{R}$, the subband channel information $\mathbf{h}_s$ is derived through the subband effective channel quantization followed by a reconstruction.

Firstly, spatial compression, which is refereed to as the wideband quantization, is operated by reducing the quantization of wideband channel information conveyed in $\mathbf{R}$ from $N_t \times N_t$ dimension to $N_t \times K$ dimension (considering only the K strongest eigenvalues and eigenvectors). We thus have the singular value decomposition (SVD) of covariance matrix

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}}, \tag{5}$$

where $\mathbf{\Lambda}$ is an $r \times r$ diagonal matrix with its diagonal elements comprised of r non-zero eigenvalues of $\mathbf{R}$ (arranged in descending order), and $\mathbf{U}$ is the tall unitary $N_t \times r$ matrix each column denoting the eigenvector corresponding to the very eigenvalue. To be exact, we let $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]$ with any $\mathbf{u}_j$ denotes the jth strongest eigenvector. The rank of $\mathbf{U}$ is denoted by r. It is defined that $\mathbf{\Lambda} = \mathbf{\Sigma}^2, \mathbf{\Sigma} = \mathrm{diag}(\{\sigma_j\})$ where σ_j is the jth largest singular-value correspondingly. The singular values $\{\sigma_j\}$ are seen as the *wideband amplitudes*.

While conventionally wideband feedback is based on quantizing the covariance matrix $\mathbf{R} = E_{\mathbf{h}}[\mathbf{h}\mathbf{h}^{\mathrm{H}}]$ [20], or a normalized version $E_{\mathbf{h}}[\mathbf{h}\mathbf{h}^{\mathrm{H}}]/E_{\mathbf{h}}[\|\mathbf{h}\|^2]$ [22], we instead quantize

$$\widetilde{\mathbf{R}} \triangleq E_{\mathbf{h}}\left[\mathbf{h}\mathbf{h}^{\mathrm{H}}/\|\mathbf{h}\|^2\right] = \mathbf{U}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{U}, \tag{6}$$

and its singular value partition. The motivation for this will be given in Section III-B.

Now we can quantize the K strongest eigenvectors in $\mathbf{U}_K$ and eigenvalues in $\mathbf{\Lambda}_K$. After that, we derive a fed back matrix $\mathbf{W}_K = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K]$ of which each column $\mathbf{w}_j$ denotes the quantized version of the corresponding eigenvector $\mathbf{u}_j$ of $\mathbf{U}_K$. Similarly, the BS has also be fed back a matrix $\hat{\mathbf{\Lambda}}_K = \hat{\mathbf{\Sigma}}_K^2, \hat{\mathbf{\Sigma}}_K = \mathrm{diag}(\{\hat{\sigma}_j\})$, where $\hat{\mathbf{\Lambda}}_K, \hat{\mathbf{\Sigma}}_K$ and $\hat{\sigma}_j$ are the quantized versions of $\mathbf{\Lambda}_K, \mathbf{\Sigma}_K$ and σ_j, respectively. Note that unless otherwise specified, we omit the subscript K in $\mathbf{U}_K, \mathbf{\Lambda}_K, \mathbf{\Sigma}_K$ and the quantized versions for simplification.

The $N_t \times K$ dimensional wideband fed back matrix $\mathbf{W}$ is obtained by minimizing the quantization distortion, i.e., *projection distance* defined as

$$d_p(\mathbf{W}, \tilde{\mathbf{R}}) \triangleq 1 - \mathrm{Tr}\left(\mathbf{W}_O^{\mathrm{H}}\tilde{\mathbf{R}}\mathbf{W}_O\right) \tag{7}$$

which measures the ratio of the energy of channel covariance matrix outside of the space spanned by the orthonormalized basis. Here normalized covariance with $\mathrm{Tr}(\tilde{\mathbf{R}}) = 1$ is assumed. The intuition for this criterion shall be provided in Section III-B. In (7), $\mathbf{W}_O$ is an $N_t \times K$ matrix of which the columns form an orthogonal basis of the space spanned by the columns of $\mathbf{W}$ satisfying $\mathbf{W}_O^{\mathrm{H}}\mathbf{W}_O = \mathbf{I}$. The unitary matrix $\mathbf{W}_O$ captures the K-dimensional subspace resulting from spatial compression. The eigenvalues $\{\sigma_j\}$ can be easily quantized with a scalar codebook.

Next, subband channel quantization can now be performed as follows. The user first use the fed back wideband matrix $\mathbf{U}_K$ as a basis, and feeds back information about subband channels expanded in this basis. The derived sub-band effective channels are then quantized utilizing some Gaussian vector quantization approaches. The procedure is given bellow.

The subband channel vectors can be theoretically written as

$$\mathbf{h}_s = \mathbf{U}\mathbf{b}_s, s = 1, 2, \ldots, S. \tag{8}$$

where $\mathbf{b}_s$ is an $r \times 1$ vector with independently but non-identically distributed (i.n.i.d) coordinates, and the unit norm vector $\mathbf{b}_s$ is referred to as the *sub-band effective channel*. The subband effective channels are coordinates of the sub-band channel vectors in terms of the extract or approximate wideband covariance information. In *ideal* case with perfect wideband information, the $r \times 1$ dimension sub-band effective channels $\{\mathbf{b}_s\}$ can be obtained as

$$\mathbf{b}_s = \mathbf{U}^{\mathrm{H}}\mathbf{h}_s. \tag{9}$$

Assuming generic wideband quantization, where $\mathbf{W}$ is the $N_t \times K$ quantized wideband beams, the user can estimate the $K \times 1$ subband effective channels as

$$\mathbf{b}_s = \left(\mathbf{W}^{\mathrm{H}}\mathbf{W}\right)^{-1}\mathbf{W}^{\mathrm{H}}\mathbf{h}_s. \tag{10}$$

Note that in (10) pseudo-inversion is utilized because a generic $\mathbf{W}$ is assumed which may not be guaranteed to be orthogonal.

It is apparent that $\mathbf{b}_s$ in (10) satisfies $\sum_s \mathbf{b}_s^{\mathrm{H}}\mathbf{b}_s \approx \mathbf{\Lambda}$, so that it is not a i.i.d. vector. However, the difference from identity is rather mild. The distributions of the effective channel components $\mathbf{b}_s$ are not *identical*, but they are independent. The phase variables of the effective channel components are indeed i.i.d. where each phase component is independently distributed. The amplitudes are also independently distributed, *but not identically*. Therefore, $\mathbf{b}_s$ is characterised as *independently but non-identically distributed (i.ni.d.)* vector. Compared with i.ni.d.vector, i.i.d. vector is easier to be quantized.

We thus focus on the normalized version of subband effective channel instead based on the inversion of singular values:

$$\mathbf{c}_s = \hat{\mathbf{\Sigma}}^{-1}\mathbf{b}_s. \tag{11}$$

In perfect wideband feedback, (11) becomes $\mathbf{c}_s = \mathbf{\Sigma}^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{h}_s$.

Based on (11), it is indicated that $\sum_s \mathbf{c}_s^{\mathrm{H}}\mathbf{c}_s \approx \mathbf{I}$, so that $\mathbf{c}_s$ can be approximated as i.i.d. vector which can be quantized using existing i.i.d. quantization schemes. In the standard [21], it is suggested to rely on i.i.d. quantization approach to quantize $\mathbf{c}_s$. Though $\mathbf{c}_s$ is approximately i.i.d. with ideal feedback, it is *suboptimal* to use i.i.d. feedback principles to quantize $\mathbf{c}_s$ in general cases. In Section V, we will further illustrate our solution to resort the gap between i.ni.d. quantization regarding $\mathbf{b}_s$ and i.i.d. quantization regarding $\mathbf{c}_s$.

We define the quantized version of $\mathbf{c}_s$ as $\hat{\mathbf{c}}_s$. Finally, the user can reconstruct the subband channel vector $\mathbf{h}_s$ utilizing the wideband fed back matrix $\mathbf{W}$, quantized wideband amplitudes from $\hat{\mathbf{\Sigma}}$, and the quantized subband effective channels $\hat{\mathbf{c}}_s$ using

$$\hat{\mathbf{h}}_s = \mathbf{W}\hat{\mathbf{\Sigma}}\hat{\mathbf{c}}_s, \tag{12}$$

which ends the modular channel quantization for single user massive MIMO networks. The overall quantization quality can then be measured by the distortion given in (4).

To conclude, the problem we are now considering is the following. Let us assume that the covariance matrix $\mathbf{R}$ (or $\tilde{\mathbf{R}}$) is fixed and we have a budget of bits for feeding back wideband statistical data of the channel pertaining to $\mathbf{R}$, and a separate budget for feeding back subband specific CSI pertaining to the coordinates of $\mathbf{c}_s$. While this section summarizes the general procedure of CSI quantization, more information about the quantization schemes from principles to steps for wideband feedback and sub-band feedback will be illustrated in Sections IV and V. Note that in the following we shall omit the the subscript s for the sake of briefness.

B. Quantization Distortion Partitioning

The motivation of quantization distortion partitioning is that overall quality is given by chordal distance of the channel vector, but in modular quantization, we should understand how this can be partitioned between a quality measure of the wideband channel, and a measure on the subband channel.

To proceed, we assume a generic $\mathbf{W}$ for any wideband quantization, which may not be orthogonal. We can then obtain

$$\mathbf{W} = \mathbf{W}_O\mathbf{C}^{-1}, \tag{13}$$

where $\mathbf{W}_O$ is the corresponding unitary matrix with orthogonal columns satisfying $\mathbf{W}_O^{\mathrm{H}}\mathbf{W}_O = \mathbf{I}$. $\mathbf{C}$ is the Cholesky factor

of the basis correlation matrix $\mathbf{W}^{\mathrm{H}}\mathbf{W}$. If $\mathbf{W}$ is orthogonal, we have $\mathbf{C} = \mathbf{I}_K$. The subband channel can be rewritten as

$$\mathbf{h} = \mathbf{W}\mathbf{b} = \mathbf{W}_O\tilde{\mathbf{b}}, \tag{14}$$

where $\tilde{\mathbf{b}}$ is the subband effective channel corresponding to the unitary matrix $\mathbf{W}_O$. This means that for any generic wideband fed back matrix $\mathbf{W}$ and the subband effective channel $\mathbf{b}$, one is able to derive the orthogonal version of $\mathbf{W}$, i.e. $\mathbf{W}_O$, and a corresponding effective channel $\tilde{\mathbf{b}}$. Therefore, there is no appropriate reason not to assume that $\mathbf{W}$ is orthogonal. In the following we concentrate on analyzing the overall subband channel distortion in terms of the orthogonal subspace $\mathbf{W}$, and the associated subspace coordinates, without loss of generality. The *objective* to illustrate that overall distortion can be divided to a projection distance, describing wideband quantization, and a chordal distance of the subband effective vector.

We start with analytical derivations for the intuitive notions used. We consider modular quantization of a single user channel $\mathbf{h}$ using the quantization codebook $\hat{\mathcal{C}}$ for the $K \times 1$ effective channels, a *unitary* covariance fed back matrix $\mathbf{W}$ satisfying $\mathbf{W}^{\mathrm{H}}\mathbf{W} = \mathbf{I}$, and a quantized $K \times K$ wideband amplitude matrix $\hat{\mathbf{\Sigma}}$. We denote the induced quantization codebook for $\mathbf{h}$ by

$$\hat{\mathcal{H}} \triangleq \mathbf{W}\hat{\mathbf{\Sigma}}\hat{\mathcal{C}} \triangleq \{\mathbf{W}\hat{\mathbf{\Sigma}}\hat{\mathbf{c}} \mid \hat{\mathbf{c}} \in \hat{\mathcal{C}}\}. \tag{15}$$

According to (3)-(4) we are interested in how well the elements of $\hat{\mathcal{H}}$ quantize $\mathbf{h}$ in terms of chordal distance. Hence, here we assume that all the codewords $\hat{\mathbf{h}} \in \hat{\hat{\mathcal{H}}}$ satisfy $\|\hat{\mathbf{h}}\| = 1$.

Consider the projector map $\mathbf{\Pi}_W = \mathbf{W}\mathbf{W}^{\mathrm{H}}$ that maps the elements of $\mathbb{C}^{N_t}$ to the space spanned by the columns of $\mathbf{W}$. We can now decompose $\mathbf{h}$ to a component lying in the $\mathbf{W}$-subspace, and to a component in the perpendicular subspace;

$$\mathbf{h} = \mathbf{\Pi}_W\mathbf{h} + (\mathbf{I} - \mathbf{\Pi}_W)\mathbf{h} \equiv \mathbf{h}_{\parallel} + \mathbf{h}_{\perp} \ . \tag{16}$$

We use shorthand $\mathbf{h}_{\perp}/\|\mathbf{h}\| = \tilde{\mathbf{h}}_{\perp}$, $\mathbf{h}_{\parallel}/\|\mathbf{h}\| = \tilde{\mathbf{h}}_{\parallel}$ and $\mathbf{h}/\|\mathbf{h}\| = \tilde{\mathbf{h}}$. With this we have that $\frac{\mathbf{h}}{||\mathbf{h}||} = \tilde{\mathbf{h}}_{\parallel} + \tilde{\mathbf{h}}_{\perp}$ and

$$\left\langle \tilde{\mathbf{h}}_{\parallel} + \tilde{\mathbf{h}}_{\perp}, \tilde{\mathbf{h}}_{\parallel} + \tilde{\mathbf{h}}_{\perp} \right\rangle = \|\tilde{\mathbf{h}}_{\parallel}\|^2 + \|\tilde{\mathbf{h}}_{\perp}\|^2 = 1. \tag{17}$$

The proof of the following result is omitted due to space constraints.

Lemma 1: Assume that $\mathbf{h}$ is a channel realization and that $\hat{\mathbf{h}}$ is a quantized version selected from the codebook $\hat{\mathcal{H}}$. Then

$$1 - \left|\left\langle \tilde{\mathbf{h}}, \hat{\mathbf{h}} \right\rangle\right|^2 = \|\tilde{\mathbf{h}}_{\parallel}\|^2 \left(1 - \left|\left\langle \frac{\mathbf{h}_{\parallel}}{\|\mathbf{h}_{\parallel}\|}, \hat{\mathbf{h}} \right\rangle\right|^2\right) + \|\tilde{\mathbf{h}}_{\perp}\|^2$$

Proof: The vector $\hat{\mathbf{h}}$ belongs to the span of $\mathbf{W}$, hence $\left\langle \tilde{\mathbf{h}}, \hat{\mathbf{h}} \right\rangle = \left\langle \tilde{\mathbf{h}}_{\parallel}, \hat{\mathbf{h}} \right\rangle$. We then get that

$$\begin{aligned} 1 - \left|\left\langle \tilde{\mathbf{h}}, \hat{\mathbf{h}} \right\rangle\right|^2 &= \|\tilde{\mathbf{h}}_{\parallel}\|^2 + \|\tilde{\mathbf{h}}_{\perp}\|^2 - \left|\left\langle \tilde{\mathbf{h}}_{\parallel}, \hat{\mathbf{h}} \right\rangle\right|^2 \\ &= \|\tilde{\mathbf{h}}_{\parallel}\|^2 + \|\tilde{\mathbf{h}}_{\perp}\|^2 - \|\tilde{\mathbf{h}}_{\parallel}\|^2 \cdot \left|\left\langle \frac{\mathbf{h}_{\parallel}}{\|\mathbf{h}_{\parallel}\|}, \hat{\mathbf{h}} \right\rangle\right|^2 \\ &= \|\tilde{\mathbf{h}}_{\parallel}\|^2 \left(1 - \left|\left\langle \frac{\mathbf{h}_{\parallel}}{\|\mathbf{h}_{\parallel}\|}, \hat{\mathbf{h}} \right\rangle\right|^2\right) + \|\tilde{\mathbf{h}}_{\perp}\|^2 \end{aligned}$$

■

Proposition 1: Given a random vector $\mathbf{h}$ and corresponding quantization codebook $\hat{\hat{\mathcal{H}}}$ we get

$$\begin{aligned} &E_{\mathbf{h}}\left[\min_{\hat{\mathbf{h}} \in \hat{H}} \left\{1 - \left|\left\langle \tilde{\mathbf{h}}, \hat{\mathbf{h}} \right\rangle\right|^2\right\}\right] = \\ &E_{\mathbf{h}}\left[\min_{\hat{\mathbf{h}} \in \hat{H}} \left\{\|\tilde{\mathbf{h}}_{\parallel}\|^2 \left(1 - \left|\left\langle \frac{\mathbf{h}_{\parallel}}{\|\mathbf{h}_{\parallel}\|}, \hat{\mathbf{h}} \right\rangle\right|^2\right)\right\}\right] + d_p(\mathbf{W}, \widetilde{\mathbf{R}}). \end{aligned}$$

Proof: For simplicity, we will disregard the minimum term. By Lemma 1

$$\begin{aligned} E_{\mathbf{h}}\left[1 - \left|\left\langle \tilde{\mathbf{h}}, \hat{\mathbf{h}} \right\rangle\right|^2\right] &= E_{\mathbf{h}}\left[\|\tilde{\mathbf{h}}_{\parallel}\|^2 \left(1 - \left|\left\langle \frac{\mathbf{h}_{\parallel}}{\|\mathbf{h}_{\parallel}\|}, \hat{\mathbf{h}} \right\rangle\right|^2\right)\right] \\ &\quad + E_{\mathbf{h}}\left[\|\tilde{\mathbf{h}}_{\perp}\|^2\right]. \end{aligned}$$

It follows from (17) that

$$\begin{aligned} E_{\mathbf{h}}\left[\|\tilde{\mathbf{h}}_{\perp}\|^2\right] &= 1 - E_{\mathbf{h}}\left[\|\tilde{\mathbf{h}}_{\parallel}\|^2\right] = 1 - E_{\mathbf{h}}\left[\mathrm{Tr}\ \tilde{\mathbf{h}}\tilde{\mathbf{h}}^{\mathrm{H}}\mathbf{\Pi}_W\right] \\ &= 1 - E_{\mathbf{h}}\left[\mathrm{Tr}\left(\mathbf{W}^{\mathrm{H}}\frac{\mathbf{h}\mathbf{h}^{\mathrm{H}}}{\|\mathbf{h}\|^2}\mathbf{W}\right)\right] \end{aligned}$$

Since the expected value commutes with trace and multiplication with constant matrices, we get

$$E_{\mathbf{h}}\left[\mathrm{Tr}\left(\mathbf{W}^{\mathrm{H}}\frac{\mathbf{h}\mathbf{h}^{\mathrm{H}}}{\|\mathbf{h}\|^2}\mathbf{W}\right)\right] = \mathrm{Tr}\left(\mathbf{W}^{\mathrm{H}}\left(E_{\mathbf{h}}\left[\frac{\mathbf{h}\mathbf{h}^{\mathrm{H}}}{\|\mathbf{h}\|^2}\right]\right)\mathbf{W}\right).$$

The final result then follows straightforwardly. ■

Corollary 1: We have the following upper and lower bounds for the overall distortion

$$\begin{aligned} d_p(\mathbf{W}, \widetilde{\mathbf{R}}) &\leq E_{\mathbf{h}}[\min_{\hat{\mathbf{h}}}\{d(\hat{\mathbf{h}}, \mathbf{h})^2\}] \\ &\leq E_{\mathbf{b}}[\min_{\mathbf{b}^{\mathrm{H}}}\{d(\mathbf{b}^{\mathrm{H}}, \mathbf{b})^2\}] + d_p(\mathbf{W}, \widetilde{\mathbf{R}}), \end{aligned}$$

where $\mathbf{b}^{\mathrm{H}} \in \hat{\mathbf{\Sigma}}\hat{\mathcal{C}}$ and $\mathbf{b} = \mathbf{W}^{\mathrm{H}}\mathbf{h}$.

Proof: The first inequality is a direct corollary of Proposition 1. Since $\|\tilde{\mathbf{h}}_{\parallel}\|^2 \leq 1$ we have

$$\begin{aligned} &E_{\mathbf{h}}\left[\min_{\hat{\mathbf{h}} \in \hat{H}} \left\{\|\tilde{\mathbf{h}}_{\parallel}\|^2 \left(1 - \left|\left\langle \frac{\mathbf{h}_{\parallel}}{\|\mathbf{h}_{\parallel}\|}, \hat{\mathbf{h}} \right\rangle\right|^2\right)\right\}\right] \\ &\qquad \leq E_{\mathbf{h}}\left[\min_{\hat{\mathbf{h}} \in \hat{H}} \left\{1 - \left|\left\langle \frac{\mathbf{h}_{\parallel}}{\|\mathbf{h}_{\parallel}\|}, \hat{\mathbf{h}} \right\rangle\right|^2\right\}\right] \end{aligned} \tag{18}$$

Recall that $\{\hat{\mathbf{h}} \in \hat{H}\} = \{\mathbf{W}\mathbf{b}^{\mathrm{H}} \mid \mathbf{b}^{\mathrm{H}} \in \hat{\mathbf{\Sigma}}\hat{\mathcal{C}}\}$ and that $\mathbf{W}$ is as an isometry. Then

$$\begin{aligned} &\min_{\hat{\mathbf{h}} \in \hat{H}} \left\{1 - \left|\left\langle \frac{\mathbf{h}_{\parallel}}{\|\mathbf{h}_{\parallel}\|}\hat{\mathbf{h}} \right\rangle\right|^2\right\} \\ &\qquad = \min_{\mathbf{b}^{\mathrm{H}}} \left\{1 - \left|\left\langle \frac{\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{h}}{\|\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{h}\|}, \mathbf{W}\mathbf{b}^{\mathrm{H}} \right\rangle\right|^2\right\} \\ &\qquad = \min_{\mathbf{b}^{\mathrm{H}}} \left\{1 - \left|\left\langle \frac{\mathbf{W}^{\mathrm{H}}\mathbf{h}}{\|\mathbf{W}^{\mathrm{H}}\mathbf{h}\|}, \mathbf{b}^{\mathrm{H}} \right\rangle\right|^2\right\}. \end{aligned} \tag{19}$$

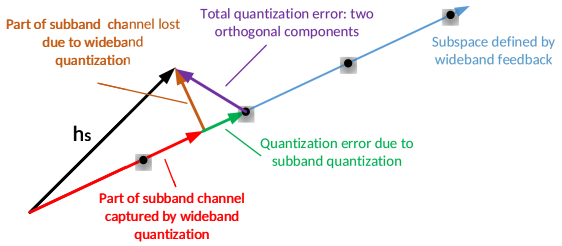

Fig. 2: Quantization error partition.

Since $\mathbf{W}^{\mathrm{H}}\mathbf{h} = \mathbf{b}$, the above yields

$$E_{\mathbf{h}}\left[\min_{\hat{\mathbf{h}}}\left\{1 - \left|\left\langle \frac{\mathbf{h}_{\|}}{\|\mathbf{h}_{\|}\|}, \hat{\mathbf{h}}\right\rangle\right|^2\right\}\right] = E_{\mathbf{b}}\left[\min_{\mathbf{b}^{\mathrm{H}}}\left\{1 - \left|\left\langle \frac{\mathbf{b}}{\|\mathbf{b}\|}, \mathbf{b}^{\mathrm{H}}\right\rangle\right|^2\right\}\right]. \quad (20)$$

This ends the proof of Corollary 1. ■

This result proves that the size of $d_p(\mathbf{W}, \widetilde{\mathbf{R}})$ provides an absolute lower bound for quantization distortion, when using the wideband fed back matrix $\mathbf{W}$. It also suggests that minimizing it is a good criterion for selecting the matrix $\mathbf{W}$. Furthermore we see that if we assume that $\mathbf{W}$ is an isometry (the columns are orthonormal) and $d_p(\mathbf{W}, \widetilde{\mathbf{R}})$ is small, then good effective channel quantization results into good quantization in the actual channel. Due to Corollary 1, the wideband amplitude quantization principle can be seen as a part of the subband quantization problem, it does not affect the projection metric d_p.

We summarize the quantization distortion partition in Fig. 2, based on what the following conclusions can be drawn as the theoretical base for our research on CSI quantization.

- To minimize the overall distortion, one should find a wideband codebook minimizing $d_p(\mathbf{W}, \tilde{\mathbf{R}})$ and subband effective channel quantization scheme minimizing $E_{\mathbf{b}}[d(\mathbf{b}, \hat{\mathbf{b}})^2]$.
- The quantization performance depends on two independent parts: effective subband quantization and how well the space generated by vectors in $\mathbf{W}$ captures $\mathbf{h}$, so that improving either part improves the overall quantization.
- If $\mathbf{W}$ is fixed, even perfect subband quantization does not provide perfect overall quantization due to $d_p(\mathbf{W}, \mathbf{R})$.
- If $\mathbf{W}$ does not consist of orthonormal columns, good quantization in the effective subband does not guarantee good overall quantization even if $\mathbf{W}$ captures $\mathbf{h}$ well.

IV. Wideband Quantization

If the single user channels $\mathbf{h}$ come from an arbitrary distribution, their quantization and feedback can be highly complex. However, if it is assumed that the channels from different antennas are correlated, it is possible to apply a modular quantization approach and reduce complexity considerably. The feedback method for 5G NR [21], [22] is based on such an approach. As in 5G NR, we consider a situation where multiple MIMO channels are operated in parallel on subbands in the frequency domain, and the frequency-selective channels are correlated between subbands as is presented in Section II.

The target of wideband quantization is to utilize a budget of bits for feeding back wideband statistical data pertaining to the channel covariance matrix $\tilde{\mathbf{R}}$ as precisely as possible.

To minimize the distortion, we should find an $N_t \times K$ dimensional matrix $\mathbf{W}$ that would minimize *projection distance* given in (7). The lower the projection distance is, the better the the orthonormalized basis $\mathbf{W}_O$ characterizes $\tilde{\mathbf{R}}$.

Motivated by the results of overall quantization distortion partitioning, we develop two types of wideband quantization approaches, namely *orthonormalized wideband precoding (OWP)* and *sequential wideband feedback (SWF)*, that improve not only the quality of wideband feedback but also the accuracy of subband channel projection.

A. Orthonormalized Wideband Precoding (OWP)

Assuming any codebook $\hat{\mathcal{C}}_{\mathbf{w}}$ for wideband feedback (e.g. oversampled DFT), the quantization process now proceeds as a series of parallel exhaust searches for the optimal codewords each of which quantizes one column of $\mathbf{U}_K$ given by

$$\min_{\mathbf{w}_k \in \hat{\mathcal{C}}_{\mathbf{w}}} d(\mathbf{u}_k, \mathbf{w}_k)^2 \quad \text{s.t. } k = 1, 2, \ldots, K. \quad (21)$$

Here codebook $\hat{\mathcal{C}}_{\mathbf{w}}$ consists of 2^B N_t-dimensional unit norm vectors, i.e. $\hat{\mathcal{C}}_{\mathbf{w}} \triangleq \{\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \ldots, \hat{\mathbf{w}}_{2^B}\}$. B is its cardinality. Hence, the wideband fed back matrix becomes $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K]$ with the kth column $\mathbf{w}_k$ being the optimal codeword quantizing eigenvector $\mathbf{u}_k$ for $k = 1, 2, \ldots, K$.

However, if the codebook $\hat{\mathcal{C}}_{\mathbf{w}}$ is overcomplete, as for example in the high resolution alternative of 5G NR [21], there is a high probability that $\mathbf{W}\mathbf{W}^{\mathrm{H}} \neq \mathbf{I}_{N_t}$, and the connection between the channel and effective channel quantization is partially broken. Which is to say, the basis spanned by $\mathbf{W}$ is probably independent but not orthogonal, so that the angles between any pairs of vectors may not always been preserved which shall cause errors when feeding back the sub-band channel utilizing the basis in $\mathbf{W}$. In the following we will show how we can avoid this problem by normalizing the channel covariance matrix and orthogonalizing the basis.

Let us now assume that we are aiming at $N_t \times K$ dimensional wideband feedback. After deriving $\mathbf{W}$, we orthogonalize the basis using Gram-Schmidt process, so that we have

$$\mathbf{W}_O = \mathrm{orth}(\mathbf{W}) \quad (22)$$

where $\mathbf{W}_O$ is a matrix with columns forming orthogonal basis of the space spanned by the columns of $\mathbf{W}$ as is defined in (13). As such, this matrix now satisfies $\mathbf{W}_O^{\mathrm{H}}\mathbf{W}_O = \mathbf{I}$.

It is worth notifying that using matrix quantization codebooks of $N_t \times K$ matrices $\mathbf{W}$ that would minimize the projection distance (7) leads to typically high complexity quantization. Instead we proceed as previously by quantizing $\mathbf{U}_K$ column by column using some vector quantization codebook $\hat{\mathcal{C}}_{\mathbf{w}}$. As a result, the user has now an $N_t \times K$ matrix $\mathbf{W}$ consisting of quantized norm 1 vectors. These vectors are not necessarily orthonormal. We also assume that the data of $\mathbf{W}$ will be fed back to the BS. However, now we assume that both the users and the BS *have agreed on a method to orthonormalize* the vectors $\mathbf{W}$. This will produce a set of

orthonormalized vectors which span the same space as the columns in $\mathbf{W}$. This method is independent of the structure of $\mathbf{W}$ and is assumed to be shared at the same time as the codebook $\hat{\mathcal{C}}_{\mathbf{w}}$. Now the user and BS both perform this operation, after which they both have matrices $\mathbf{W}_O$. Exactly the same number of bits can be used to feed back this matrix as feeding back the original $\mathbf{W}$. The matrix $\mathbf{W}_O$ can now be used in place of $\mathbf{W}$ for the rest of the quantization.

However, as $\mathbf{W}$ is transformed to $\mathbf{W}_O$ it does not make sense to use quantized singular values as wideband amplitudes. Instead we calculate wideband amplitudes as

$$\sigma_j = \sqrt{\mathbf{w}_j^{\mathrm{H}} \widetilde{\mathbf{R}} \mathbf{w}_j} \tag{23}$$

for each column $\mathbf{w}_j$ in $\mathbf{W}_O$, and then feed back quantized versions of these elements $\{\hat{\sigma}_j\}$ by using the same quantization scheme designated for the original singular values. For each wideband amplitude which is a complex scalar, any scalar codebook with several bits shall work well. After this operation, BS and user share the matrix $\mathbf{W}_O$ and the quantized wideband amplitude matrix $\hat{\mathbf{\Sigma}} = \mathrm{diag}(\{\hat{\sigma}_j\})$. Using these matrices, the subband feedback can be performed as in Equations (10) and (11), replacing $\mathbf{W}$ and $\hat{\mathbf{\Sigma}}$ with $\mathbf{W}_O$ and the corresponding quantized wideband amplitudes using (23).

B. Sequential Wideband Feedback (SWF)

In this section, a sequential wideband feedback approach is presented which is proved to provide a lower projection distance, so that it further reduces the overall quantization distortion. The intuition is to provide a self-cancellation gain to the quantization distortion by quantizing the wideband beams in an iterative process instead of the normal independently parallel manner.

Define a new matrix $\mathbf{\Pi}^{\perp}$ initialized at unit matrix, i.e. $\mathbf{\Pi}_0^{\perp} = \mathbf{I}_{N_t}$. $\mathbf{\Pi}^{\perp}$ is defined as the perpendicular space of previous projections, such that it is updated in each iteration with $\mathbf{\Pi}_j^{\perp}, j = 1, \ldots, K$ denoting the updated version of $\mathbf{\Pi}^{\perp}$ in the jth iteration. In this section, we always use $\tilde{\mathbf{R}}$ of (6), but for brevity, just call it $\mathbf{R}$. To proceed, we shall sequentially calculate the projection of the covariance matrix $\mathbf{R}$ to the current perpendicular space generated by $\mathbf{\Pi}_{j-1}^{\perp}$ using

$$\mathbf{R}_{p,j} = \mathbf{\Pi}_{j-1}^{\perp,\mathrm{H}} \, \mathbf{R} \, \mathbf{\Pi}_{j-1}^{\perp}, \; j = 1, \ldots, K. \tag{24}$$

Then $\mathbf{R}_{p,j}$ is used for singular value partition instead of $\mathbf{R}$ as

$$\mathbf{R}_{p,j} = \mathbf{E}_j \mathbf{\Xi}_j \mathbf{E}_j^{\mathrm{H}}, \tag{25}$$

where $\mathbf{E}_j = [\mathbf{e}_{j,1}, \mathbf{e}_{j,2}, \ldots, \mathbf{e}_{j,r}]$ with each column denoting one eigenvector and r being the rank. The eigenvalues in $\mathbf{\Xi}_j$ are arranged in descending order. Hence the strongest eigenvector in the jth iteration becomes $\mathbf{e}_{j,1}$. Utilizing wideband quantization codebook (e.g. oversampled DFT), we quantize the strongest eigenvector $\mathbf{e}_{j,1}$ to $\mathbf{w}_j$ in the j-th iteration.

Next, we shall project the quantized strongest eigenvector $\mathbf{w}_j$ to the perpendicular space of previous codewords. The orthogonalized version of $\mathbf{w}_j$ is obtained as

$$\tilde{\mathbf{w}}_{j,\mathrm{q}} = \mathbf{\Pi}_{j-1}^{\perp} \mathbf{w}_j, \tag{26}$$

with the orthonormalized version being $\mathbf{w}_{j,\mathrm{q}} = \tilde{\mathbf{w}}_{j,\mathrm{q}} / \|\tilde{\mathbf{w}}_{j,\mathrm{q}}\|$.

The corresponding wideband amplitude is obtained as

$$\sigma_j = \mathbf{w}_{j,\mathrm{q}}^{\mathrm{H}} \, \mathbf{R}_{p,j} \, \mathbf{w}_{j,\mathrm{q}}. \tag{27}$$

Then quantize the wideband amplitude σ_j into $\hat{\sigma}_j$ using some scalar codebook. Finally $\mathbf{w}_{j,\mathrm{q}}$ and the current perpendicular space $\mathbf{\Pi}_{j-1}^{\perp}$ is utilized to update $\mathbf{\Pi}^{\perp}$ for next iteration as

$$\mathbf{\Pi}_j^{\perp} = \mathbf{\Pi}_{j-1}^{\perp} - \mathbf{w}_{j,\mathrm{q}} \mathbf{w}_{j,\mathrm{q}}^{\mathrm{H}}. \tag{28}$$

In this manner, we obtain a wideband fed back matrix after K iterations as $\mathbf{W}_{\mathrm{Q}} = [\mathbf{w}_{1,\mathrm{q}}, \mathbf{w}_{2,\mathrm{q}}, \ldots, \mathbf{w}_{K,\mathrm{q}}]$. The projection distance defined in (7) becomes

$$d_p(\mathbf{W}_{\mathrm{Q}}, \mathbf{R}) = 1 - \mathrm{Tr}\left(\mathbf{W}_{\mathrm{Q}}^{\mathrm{H}} \mathbf{R} \mathbf{W}_{\mathrm{Q}}\right). \tag{29}$$

The procedure of the sequential wideband quantization method is summarized in Alg.1 below. We now have

Algorithm 1 Sequential Wideband Feedback (SWF)

procedure Initialization
  Define $\mathbf{\Pi}_j^{\perp}$ for $j = 0, 1, \ldots, K$
  $\mathbf{\Pi}_0^{\perp} = \mathbf{I}_{N_t}$
end procedure
procedure Wideband Quantization
  for $j = 1, 2, \ldots, K$ do
    Project $\mathbf{R}$ to $\mathbf{\Pi}_{j-1}^{\perp}$ using (24): $\mathbf{R}_{p,j}$
    Conduct SVD to $\mathbf{R}_{p,j}$ using (25): $\mathbf{E}_j, \mathbf{\Xi}_j$
    Derive the strongest eigenvector $\mathbf{e}_{j,1}$
    Quantize $\mathbf{e}_{j,1}$ using wideband codebook to $\mathbf{w}_j$
    Project $\mathbf{w}_j$ to $\mathbf{\Pi}_{j-1}^{\perp}$ with (26): $\mathbf{w}_{j,\mathrm{q}}$ (normalized)
    Obtain wideband amplitude σ_j using (27) with $\mathbf{w}_{j,\mathrm{q}}$
    Quantize σ_j into $\hat{\sigma}_j$
    Update $\mathbf{\Pi}^{\perp}$ with $\mathbf{\Pi}_{j-1}^{\perp}$ and $\mathbf{w}_{j,\mathrm{q}}$ using (28): $\mathbf{\Pi}_j^{\perp}$
  end for
end procedure
procedure Output
  Obtain fed back matrix $\mathbf{W}_{\mathrm{Q}} = [\mathbf{w}_{1,\mathrm{q}}, \mathbf{w}_{2,\mathrm{q}}, \ldots, \mathbf{w}_{K,\mathrm{q}}]$
  Calculate projection distance using (29)
end procedure

Proposition 2: The sequential wideband feedback of Algorithm 1 provides a smaller projection distance than utilizing separate wideband quantization of Section IV-A, i.e.,

$$d_p(\mathbf{W}_{\mathrm{Q}}, \widetilde{\mathbf{R}}) < d_p(\mathbf{W}, \widetilde{\mathbf{R}}). \tag{30}$$

Proof: See Appendix A. ■

V. Subband Effective Channel Quantization

This section starts with a justification of the feasibility of using i.i.d. quantization for $\mathbf{c}$ with a careful bit allocation among the coordinates in subband quantization, then moves to the discussion of the suggested subband quantization schemes.

A. Effective Channel Quantization i.i.d. versus i.ni.d.

Note that even though $\mathbf{c}$ is i.i.d., using i.i.d. quantization principles directly to $\mathbf{c}$ is suboptimal in terms of subband quantization. We shall illustrate the intuition below. [1]

[1]This derivation is based on orthogonal matrix $\mathbf{W}_O$. The rationality has been clarified previously without loss of generality.

According to (10) and (11), the normalized version of effective channel $\mathbf{c}$ can be written as

$$\mathbf{c} = \hat{\mathbf{\Sigma}}^{-1}\mathbf{W}_O^{\mathrm{H}}\mathbf{h} \ . \tag{31}$$

As we are not interested in the overall phase of $\mathbf{h}$, the same holds for $\hat{\mathbf{c}}$. Also, we have a power constraint such that we are only interested in quantizing normalized channels. Given an effective channel $\mathbf{c}$ and quantized version $\hat{\mathbf{c}}$, the normalized channel vector and the quantized version are

$$\mathbf{h} = \frac{\mathbf{W}_O\,\hat{\mathbf{\Sigma}}\,\mathbf{c}}{\sqrt{\mathbf{c}^{\mathrm{H}}\hat{\mathbf{\Lambda}}\mathbf{c}}}, \qquad \hat{\mathbf{h}} = \frac{\mathbf{W}_O\,\hat{\mathbf{\Sigma}}\,\hat{\mathbf{c}}}{\sqrt{\hat{\mathbf{c}}^{\mathrm{H}}\hat{\mathbf{\Lambda}}\hat{\mathbf{c}}}}. \tag{32}$$

The squared chordal distance for the quantization is then

$$d^2(\mathbf{h},\hat{\mathbf{h}}) = 1 - \left|\mathbf{h}^{\mathrm{H}}\hat{\mathbf{h}}\right|^2 = 1 - \frac{\left|\mathbf{c}^{\mathrm{H}}\hat{\mathbf{\Lambda}}\mathbf{c}\right|^2}{\mathbf{c}^{\mathrm{H}}\hat{\mathbf{\Lambda}}\mathbf{c}\ \hat{\mathbf{c}}^{\mathrm{H}}\hat{\mathbf{\Lambda}}\hat{\mathbf{c}}} \ . \tag{33}$$

This is generically *not* the chordal distance of Grassmannian vectors in K dimensions, which is

$$d_K^2(\mathbf{c},\hat{\mathbf{c}}) = 1 - \left|\mathbf{c}^{\mathrm{H}}\hat{\mathbf{c}}\right|^2 \ . \tag{34}$$

Thus by performing the inversion with respect to $\hat{\mathbf{\Sigma}}$ when defining effective K-dimensional channels according to (31), we have indeed rendered the effective channel to be approximately i.i.d. However, we have *deformed* the quantization measure. The mapping (31) is not an isometry between the space of Grassmannian vectors in N_t and K dimensions.

In contrast, a K-dimensional effective channel without the singular value inversion is

$$\mathbf{b} = \mathbf{W}_O^{\mathrm{H}}\,\mathbf{h} \ . \tag{35}$$

As is discussed previously, the effective channel $\mathbf{b}$ is not i.i.d. With this mapping, the normalized channel and quantized channel vectors are

$$\mathbf{h} = \frac{\mathbf{W}_O\,\mathbf{b}}{\sqrt{\mathbf{b}^{\mathrm{H}}\mathbf{b}}} \ , \qquad \hat{\mathbf{h}} = \frac{\mathbf{W}_O\,\hat{\mathbf{b}}}{\sqrt{\hat{\mathbf{b}}^{\mathrm{H}}\hat{\mathbf{b}}}} \ . \tag{36}$$

Thus if $\mathbf{b}$ and $\hat{\mathbf{b}}$ are normalized, so are automatically $\mathbf{h}$ and $\hat{\mathbf{h}}$. The chordal distance becomes

$$d^2(\mathbf{h},\hat{\mathbf{h}}) = 1 - \left|\mathbf{h}^{\mathrm{H}}\hat{\mathbf{h}}\right|^2 \equiv d^2(\mathbf{b},\hat{\mathbf{b}}) \tag{37}$$

Thus the mapping (35) does not produce i.i.d. effective channels, but the mapping is an isometry.

As the objective of quantization is ultimately to represent the sub-band channel $\mathbf{h}$ with as small distortion as possible, which is directly related to the chordal distance metric, the argument above indicates that subband quantization has to be performed with the isometric mapping (35) rather than with the non-isometry (31). Thus *instead of an i.i.d. sub-band quantization problem we have an i.ni.d. (independently but non-identically distributed) quantization problem.*

Conclusions of the performed analysis are presented below:

- It makes little sense not to have amplitude information of the strongest eigenvector effective channel. With mapping (31), amplitude quantization is about the relative strength of the amplitude as compared to its typical value. The effective channels are fading channels, and the effective channel of the strongest eigenvalue as well.
- If one has amplitude information about all effective channel elements, the phase reference beam should be *the beam with the largest expected amplitude*. When assessing this, both the quantized singular values $\hat{\mathbf{\Sigma}}$, and the subband amplitude feedback should be considered.
- Since we are ultimately interested in the N_t-dimensional chordal distance (33), more feedback bits should be allocated to the stronger eigenvalues, less to the smaller. An extreme version is that no feedback at all is allocated to some of the weakest eigenvalues. This shall hold both for amplitude and phase bits, and be allocated based on eigenvector power $\hat{\mathbf{\Lambda}}$.
- Phase bits can then be allocated according to expected realized channel amplitudes, where both $\hat{\mathbf{\Lambda}}$ and possible amplitude feedback bits are taken into account. The larger the expected amplitude, the more phase bits it is worth allocating. Consequently, the largest expected amplitude should be the phase reference, the second largest expected amplitude should have most phase bits, and so on.

B. The Subband Quantization Schemes

According to Equation (8)-(11), the *normalized* subband effective channel $\mathbf{c}$ is a $K\times 1$ i.i.d. $\mathcal{CN}(0,1)$ vector refereed to as $\mathbf{c} = [c_1, c_2, \ldots, c_K]^{\mathrm{T}}$. The user quantizes $\mathbf{c}$ to $\hat{\mathbf{c}}$ using some Grassmannian quantization codebook $\hat{\mathcal{C}}$ developed for K-dimensional i.i.d. Gaussian vectors modulo norm and phase, and then sends this information to the BS. The BS can then construct an estimate of $\mathbf{h}$ according to (12).

As is defined in (4), the goal of the whole quantization process is to find a quantization scheme that minimizes the distortion, i.e. the average squared chordal distance between the subband channel vector and the quantized version over the subbands defined as $E_{\mathbf{h}}[d(\mathbf{h},\hat{\mathbf{h}})^2]$. In Section V-A, it is agreed to utilize i.i.d. quantization for $\mathbf{c}$ with reasonable bit allocation as the suboptimal solution. The objective of subband effective channel quantization is thus given by

$$\min_{\hat{\mathbf{c}}\in\hat{C}} \ E_{\mathbf{c}}[d(\hat{\mathbf{\Sigma}}\mathbf{c},\hat{\mathbf{\Sigma}}\hat{\mathbf{c}})^2], \tag{38}$$

as an approximation of the quantization of i.ni.d. vector $\mathbf{b}$.

According to (37), the quantization distortion of $\mathbf{b}$ equals the overall distortion $E_{\mathbf{h}}[d(\mathbf{h},\hat{\mathbf{h}})^2]$. Thought this approximation does not exactly hold as some part of $\mathbf{h}$ may live outside the space spanned by $\mathbf{W}$, optimizing subband effective channel quantization is still a good goal for minimizing the overall subband distortion. The intuition is clarified in Section III-B. The benefit is the dimensionality compression of subband quantization from N_t to K dimensions.

Note that the BS does not need to know either the overall channel phase or norm. The former is irrelevant, and the later is subsumed by separately transmitted channel quality indication feedback. Accordingly, the estimate $\hat{\mathbf{h}}$ is up to phase and norm. Here we skip the details of subband effective channel quantization which is well under control i.i.d. vector quantization [3], [4]. In simulation, we shall use separate amplitude and phase quantization of the type used in the

standard [22] to quantize $\mathbf{c}$. The coordinates are divided into two levels, i.e. the stronger ones and the weaker ones, and use different numbers of bits to quantize them. The idea is to use more bits from the budget for the stronger beams.

For the effective channels, quantization codebooks $\hat{\mathcal{C}}$ that act on individual coordinates directly will be used. A general codeword of codebook $\hat{\mathcal{C}}$ can be written as $\hat{\mathbf{c}} = (\hat{c}_1, \ldots, \hat{c}_K)$, where $\hat{c}_i$ is freely selected from a coordinate codebook $\mathcal{C}_i$. The advantage is that we can use large codebooks, while quantization complexity stays limited as it can be performed coordinate by coordinate. Note that as overall phase is irrelevant, one may rotate the vector such that a chosen coordinate c_i of the sub-band effective channel $\mathbf{c}$ is real positive. A natural codebook for each coordinate is one where each $\mathcal{C}_i$ is quantizing a $\mathcal{CN}(0,1)$ random variable. However when the BS reconstructs the channel by $\mathbf{W}\hat{\mathbf{\Sigma}}\mathbf{c}$, coordinates corresponding to larger values of $\hat{\sigma_j}$ have a larger effect on the overall distortion. Hence we should use variable-size codebooks $\mathcal{C}_i$ for different coordinates. This leads to rather hard bit allocation problem. Irrespectively of the chosen method, the effective channel quantization then continues as follows.

1) Bit Allocation in Extrinsic Order:

Bit allocation for effective channel coordinate quantization is based on an *extrinsic order* of the coordinates, i.e., based on the size of the corresponding $\hat{\sigma_i}$. As we are not interested on feeding back constant multiplicative terms, we first divide all the other coordinates with the coordinate corresponding to the largest $\hat{\sigma_i}$. No bits are needed for quantizing this reference coordinate. Then for the next m coordinates we quantize the amplitude with codebook $\{1, \sqrt{0.5}\}$ under power ratio $\eta = 2$ for example and use α bit uniform phasing quantization. For the last $K - (m+1)$ coordinates, we allocate 0 amplitude bits and β ($\beta < \alpha$) bits for uniform phase quantization. The granularity in bits for the quantization of $\mathbf{c}$ is thus given by

$$L = m + \alpha m + \beta(K - (m+1)). \tag{39}$$

When using this extrinsic order, the user feeds back the amplitude information, so that the base station will know which quantized amplitudes a_i correspond to which vector of $\mathbf{W}_O$. Then phase information is transmitted in any agreed order. The basic idea underlying this bit allocation is that while typically the coordinates with largest $\hat{\sigma}_i$ have the largest true amplitude $\sigma_i|c_i|$, this does not happen always. For contrast, we shall use an *intrinsic order* in the quantization. With the intrinsic order we allocate more phase bits for the coordinates having the largest impact on overall quantization distortion.

2) Bit Allocation in Intrinsic Order:

First we normalize $\mathbf{c}$, and then quantize all coordinates with one bit amplitude quantization using codebook $\{0.208, 0.462\}$ under power ratio $\eta = 5$ for instance. Assuming perfect phase quantization, these provide optimal amplitude quantization for Rayleigh fading variables with average energy $1/\sqrt{K}$ when $K = 8$, see [23]. We denote the quantized amplitude value of the ith coordinate with a_i. The user then compares $\hat{\sigma}_i a_i$ to each other and divides all the coordinates with the phase of the coordinate corresponding to the largest value of $\hat{\sigma}_i a_i$.

After this the user allocates phase bits for phase quantization based on the sizes of $\hat{\sigma}_i a_i$, not just based on $\hat{\sigma}_i$ as when using extrinsic order. No bits are needed for quantizing the phase of the coordinate with largest $\hat{\sigma}_i a_i$. Then α phase bits are allocated to the m next largest coordinates, and β phase bits to the last $K - (m+1)$ coordinates, where the total number of subband quantization bits required is thus given by

$$L = K + \alpha m + \beta(K - (m+1)). \tag{40}$$

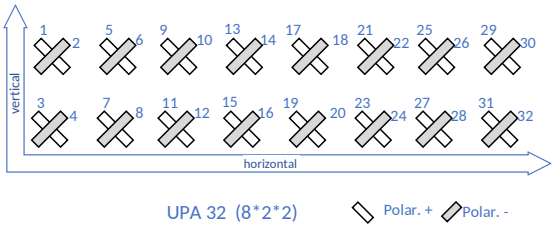

Fig. 3: $(8 \times 2 \times 2)$-element UPA diagram illustration.

VI. Simulations

A. Settings and Schemes

For simulating the modular CSI quantization scheme, we need a model of the channel which introduces a realistic combination of channel directivity, and correlation in the frequency domain, as well as a model of the antenna array, including the polarization of the antenna elements. In addition, we describe the wideband and subband codebooks used.

1) Channel Model:

QuaDRiGa V2.0.0 [24] was used to generate MIMO channel correlations with 3GPP 38.901 UMa NLOS settings. The BS was assumed having Uniform Planar Array (UPA) with $N_t = 32$ antennas with the array $8 \times 2 \times 2$ ($horizontal \times vertical \times polarization$), of which the diagram illustration is given in Fig. 3. The network layout is created by setting up one cell with 3 sectors (120 degree) and a centrally located base station. 9000 single-antenna users are dropped uniformly at random within 250m from the BS. Both the BS and users use omni-directional antennas. We consider 3000 channel covariance matrices of the users in one of the sectors. The center frequency is set to be 1.84 GHz with the Bandwidth of 18 MHz where there are in total 1200 subcarriers divided into $S = 30$ sub-bands each with 40 subcarriers. 8 clusters are considered with indoor ratio fixed at 0.8.

2) Polarization Decomposition of Channel:

If we recall the antenna arrangement of UPA32 given in Fig. 3 which matches the structure of Kronecker tensor product in MATLAB, it turns out that the subband channel vector $\mathbf{h} = [h_1, h_2, \ldots, h_{32}]^{\mathrm{T}}$ can be seen as a Kronecker tensor product of the decomposition from the three dimensions. i.e., $\mathbf{h} = \mathbf{h}_h \otimes \mathbf{h}_v \otimes \mathbf{h}_p$, where $\mathbf{h}_h, \mathbf{h}_v, \mathbf{h}_p$ denote the horizontal, vertical and polarization parts with $8, 2, 2$ elements, respectively. For instance, the $N_t/2$ odd elements belong to the positive polarization while the even elements form the negative polarization. This motivates the decomposition of the channel covariance matrix $\mathbf{R}$ (or $\widetilde{\mathbf{R}}$) and wideband codebook for $\mathbf{W}$.

The covariance matrix $\mathbf{R}$ is found by calculating the instantaneous covariance matrix and averaging over subbands. To proceed, we now consider the structure of an instantaneous

$\mathbf{h}\mathbf{h}^{\mathrm{H}} = [h_j h_k^{\mathrm{H}}]_{j=1,k=1}^{N_t,N_t}$ before performing averaging. Recalling the independence and orthogonality between the two polarization denoted by $+/-$, i.e. $\mathbf{h}_p = 1/\sqrt{2}[i, 1], i^2 = -1$, we have that $h_j h_k^{\mathrm{H}} \approx 0$ if j, k come from different polarizations. We rearrange $\mathbf{h}$ into

$$\mathbf{h} = [\underbrace{h_1, h_3, \ldots, h_{31}}_{polar.+}, \underbrace{h_2, h_4, \ldots, h_{32}}_{polar.-}]^{\mathrm{T}}$$

according to polarization. To reduce feedback we may assume an approximate polarization structure for $\mathbf{R}$: $\mathbf{R}_{B+B-} = [\mathbf{B}_+ \ \mathbf{0}; \mathbf{0} \ \mathbf{B}_-]$. In this case, $\mathbf{B}_+$ and $\mathbf{B}_-$ are $N_t/2 \times K/2$ matrices. The wideband feedback simplifies correspondingly.

Furthermore, we may reduce complexity into quantizing one $N_t/2 \times K/2$ matrix by approximating the covariance in terms of the polarization structure $\mathbf{R}_{B00B} = [\mathbf{B} \ \mathbf{0}; \mathbf{0} \ \mathbf{B}]$ based on a single matrix $\mathbf{B}$ which is an average over $\mathbf{B}_+$ and $\mathbf{B}_-$.

Finally, the covariance may be quantized directly, *without* polarization decomposition, which we refer to as the *fully-occupied* case. Then, one needs to quantize a $N_t \times K$ matrix. The utilization of polarization structures can reduce the number of bits needed for wideband feedback.

3) Quantization Codebooks: In simulations, we quatnize the covariance matrix with $K = 8$ wideband beams. For the UPA antenna, wideband feedback codebooks are usually of type $\mathcal{C}_h \otimes \mathcal{C}_v \otimes \mathcal{C}_p$, where $\mathcal{C}_h, \mathcal{C}_v$ and $\mathcal{C}_p$ quantize the horizontal, vertical and polarization dimensions, respectively. Good candidates for $\mathcal{C}_h$ and $\mathcal{C}_v$ are oversampled DFT codebooks in respective dimensions. $\mathcal{C}_p$ can use some simple i.i.d. codebook. Since here the polarization decomposition has been performed, only $\mathcal{C}_h$ and $\mathcal{C}_v$ are needed. This type of wideband codebook is denoted by Tensor oversampled DFT (TSODFT), which gives us $N_t \times 2L$ matrix $\mathbf{W}$.

Next, we present the quantization for the original singular values or new wideband amplitudes, i.e. $\{\sigma_j\}$. Strongest beam is indicated and the amplitudes of the other $K-1$ beams are quantized by 3 bits each by using the scalar codebook [21]:

$$\{1, \sqrt{0.5}, \sqrt{0.25}, \sqrt{0.125}, \sqrt{0.0625}, \sqrt{0.0313}, \sqrt{0.0156}, 0\}.$$

For subband feedback, we use the options below with the detailed steps given in Section V.

- Extrinsic order with power ratio $\eta = 2$ denoted by EXT2
- Intrinsic order with power ratio $\eta = 5$ denoted by INT5

Note that the two options are selected as the best combinations based on the performance comparison among all the 4 possible combinations. We consider different bits ranging from 24 to 49 bits for quantizing the subband effective channel $\mathbf{c}$.

4) The Wideband Quantization Schemes:

Standarddized feedback (Standard) [21]: The user derives $\mathbf{W}$ and $\hat{\mathbf{\Sigma}}$ by quantizing $\mathbf{U}_8$ and $\mathbf{\Sigma}_8$ from (5). The original singular values are used as unquantized wideband amplitudes, while $\mathbf{c}$ is obtained using $\mathbf{c} = \hat{\mathbf{\Sigma}}^{-1}\mathbf{W}^{\mathrm{H}}\mathbf{h}$ directly, or $\mathbf{c} = \hat{\mathbf{\Sigma}}^{-1}(\mathbf{W}^{\mathrm{H}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{H}}\mathbf{h}$ with pseudoinversion.

Orthonormalized wideband precoding (OWP): We start from $\widetilde{\mathbf{R}}$ and find fed back matrix $\mathbf{W}$. We then orthogonalize $\mathbf{W}$ to get $\mathbf{W}_O$, and find new wideband amplitudes using (23). The effective channel is derived as $\mathbf{c} = \hat{\mathbf{\Sigma}}^{-1}\mathbf{W}_O^{\mathrm{H}}\mathbf{h}$.

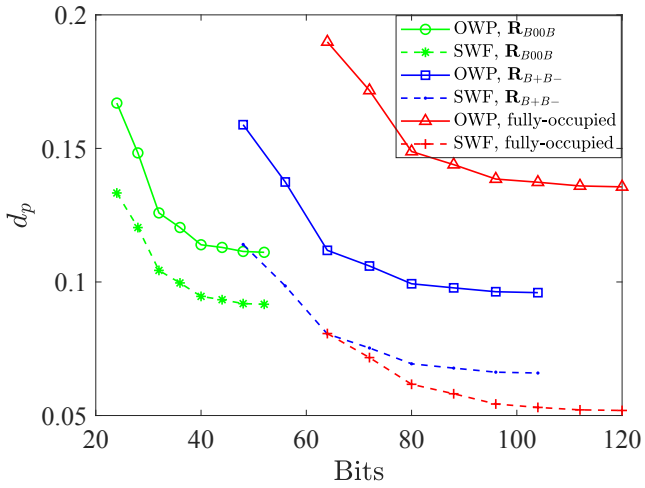

Fig. 4: Comparison between wideband feedback methods in terms of projection distance.

Sequential wideband feedback (SWF): According to Alg.1, we quantize the wideband beams iteratively based on the quantized beams from previous iterations. Other settings are similar to those used in OWP.

B. Projection Distance Comparison in Wideband Quantization

Firstly, we focus on wideband feedback, and compare the performance of the OWP scheme and the SWF scheme regarding the projection distance defined in (7) under three polarization structures, $\mathbf{R}_{B00B}$, $\mathbf{R}_{B+B-}$, and the fully-occupied $\mathbf{R}$.[2] In Fig. 4, a combination of a given polarization structure and wideband feedback scheme is simulated for 8 different codebooks; the over-sampling ratio for the TSODFT codebook in horizontal and vertical dimensions is given by $\eta = 2^i, i \in [2, 9]$ for all cases. The best division of oversampling between horizontal and vertical is applied. In addition, a 2-bit binary chirps (BC) codebook [25] is used to quantize the polarization dimension in full-occupied case. The total number of bits used for feeding back $\mathbf{W}$ are given by $K/2 \times \log 2(N_t/2 \times \eta)$ for $\mathbf{R}_{B00B}$, $K \times \log 2(N_t/2 \times \eta)$ for $\mathbf{R}_{B+B-}$ and $K \times \log 2(N_t \times \eta \times 2)$ for the fully-occupied case.

As can be observed from the figure, SWF is considerably better than OWP in all cases which agrees with Proposition 2. In this case, the gain of SWF over OWP in terms of projection distance is around $0.02, 0.03, 0.08$ for the three types of poloarization structures accordingly, which results from the different numbers and sizes of the wideband beams to be quantized. When comparing the polarization structures, $\mathbf{R}_{B00B}$ performs the best in terms of a trade-off between performance and code-book cardinality. Although the SWF scheme in full-rank case yields the lowest projection distance, the bits needed are around 2 to 3 times more than those used

[2]Note that the projection distances provided with the standardized wideband feedback method would correspond to the ones of OWP with polarization structure $\mathbf{R}_{B00B}$. In the standard, the fed back $\mathbf{W}$ is not orthogonalized, but to compute the projection distance, orthogonalization is assumed.

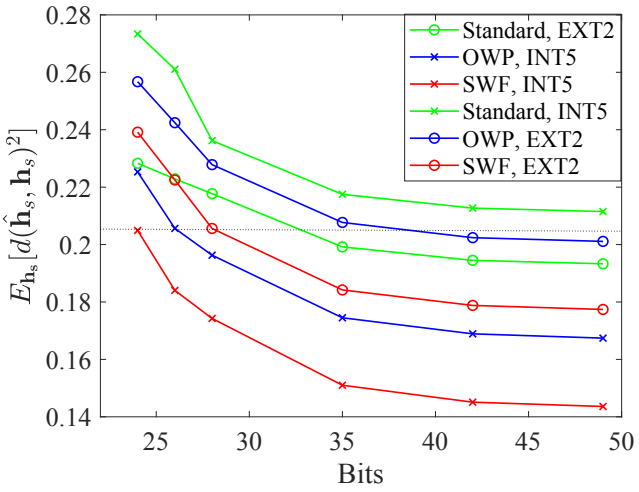

Fig. 5: Overall quantization Distortion. The first three points are with fixed phase bits $\alpha = 3, \beta = 2$ but varying the number of strong beams $m(m = 5, 6, 7$ for EXT2 and $m = 2, 4, 6$ for INT5). The rest are obtained with fixed m ($m = 7$ for EXT2 and $m = 6$ for INT5) while phase bits pair (α, β) ranges within $(4, 3), (5, 4), (6, 5)$. The corresponding number of bits can be calculated according to (39) and (40).

assuming $\mathbf{R}_{B00B}$, so that we shall focus on structure $\mathbf{R}_{B00B}$ in the following simulations.

C. Overall Distortion Comparison among Different Schemes

Here we investigate the impacts of different wideband feedback schemes with corresponding subband quantization options on the overall distortion $E_{\mathbf{h}_s}[d(\hat{\mathbf{h}}_s, \mathbf{h}_s)^2]$. Over-sampling ratio η for TSODFT codebook is given by $\eta = 16$ while the polarization structure $\mathbf{R}_{B00B}$ is used as is mentioned previously. For the standard scheme as is suggested in [21], the selected subband effective channel quantization approach is based on extrinsic order with power ratio $\eta = 2$ denoted by EXT2 [3] while for the OWP and SWF schemes, the best option turns to be INT5 using intrinsic order with power ratio $\eta = 5$. The number of bits counted here is the number of bits used for quantizing a $K \times 1$ dimension subband effective channel.

Fixing one type of wideband feedback approach, the distortion reflects the preference of the two subband quantization options in this case. As is observed in Fig. 5, there are considerable gains from using INT5 for both OWP and SWF schemes while EXT2 is the better option for the standardized scheme as is suggested in [21]. Moreover, SWF with INT5 performs the best followed by the OWP scheme with INT5 while Standard with EXT2 are the worst among the three better combinations. For instance, to reach a distortion of 0.205, $24, 26$ and 33 bits are needed for the three schemes accordingly (see the dash-line). For reference, the corresponding projection distances in wideband quantizion are $0.1222, 0.0997$ for OWP and SWF schemes while the distortions assuming infinite precision subband quantization (using perfect $\mathbf{c}$ as $\hat{\mathbf{c}}$) are given by $0.1297, 0.1050$ accordingly. With infinite precision subband quantization, the distortion is almost precisely the projection distance as is discussed in Section III-B.

D. Spectrum Efficiency Comparison for Multiuser Networks

Besides the CSI quantization for single user MIMO network averaging over thousands of independent user samples, we also evaluate performance in a multiuser network where an $N_t = 32$ antenna BS serves $M = 4$ single antenna users. The subcarriers are divided into $S = 25$ sub-bands each with 48 subcarriers. The other settings are the same as previously. Defining the channel between a user and the BS on subband s as $\mathbf{h}_s$, the user constructs the sample covariance $\mathbf{R}$ or $\widetilde{\mathbf{R}}$ by averaging over the subbands, quantizes this, and then quantizes $\mathbf{h}_s$ on each subband. Based on the feedback, the BS performs zero forcing (ZF) on each subband. The three schemes from Section VI-A4 are compared in terms of the spectral efficiency. After frequency selective channel generation, user channels were normalized to a SNR. The spectral efficiency is derived using $R_k = \log_2(1 + \gamma_k)$ where γ_k is the signal-to-interference-and-noise ratio (SINR) of user k given by with $\mathbf{v}_j$ referring to the ZF-beamforming vector (seeing Section II of [1]). The benchmark assuming single user SISO AWGN is given by replacing the SINR γ_k with the SNR. Here the three wideband schemes are operated with their best subband quantization options using two types of effective channel quantization granularity ($24/28$ bits). [4]

Average single user spectral efficiency given an SNR is plotted in Figure 6. Simulations corroborate the theoretical principles discussed above. OWP provides a gain of more than 25% against the standardized versions while SWF further brings a gain of more than 8% over OWP. Using pseudoinversion with the standardized scheme gives nominal gain. Increasing the effective channel quantization granularity does improve the performance of OWP and SWF considerably, while doing so with the standardized version provides little gain. While we here state the results only with $M = 4$ users, similar results can be observed with higher numbers of users.

VII. Conclusions

We have addressed the CSI quantization for massive MIMO systems in a modular quantization scheme. Analyzing the separation of feedback to wideband and subband parts, we found quantization objectives for optimal modular quantization. We show considerable performance improvement in a MIMO scenario with a high number of antennas, when these principles are applied. An orthonormalized wideband precoding (OWP) scheme and a sequential wideband feedback (SWF) scheme are developed with their best subband quantization options. Orthogonalization of wideband feedback makes it possible to further increase quantization accuracy by improving subband quantization accuracy, while if orthogonalization is omitted, improving subband quantization seems useless. The sequential wideband feedback scheme brings a further improvement to the quantization quality over the OWP scheme.

[3] Here standardized scheme uses pseudoinversion form for deriving effective channel which gives higher accuracy quantization.

[4] To quantize $\mathbf{c}$, the parameters are set as $(\alpha, \beta) = (3, 2)$ and $m = 5/7$ (Standard) or $m = 2/6$ (OWP and SWF) to get in total $24/28$ bits.

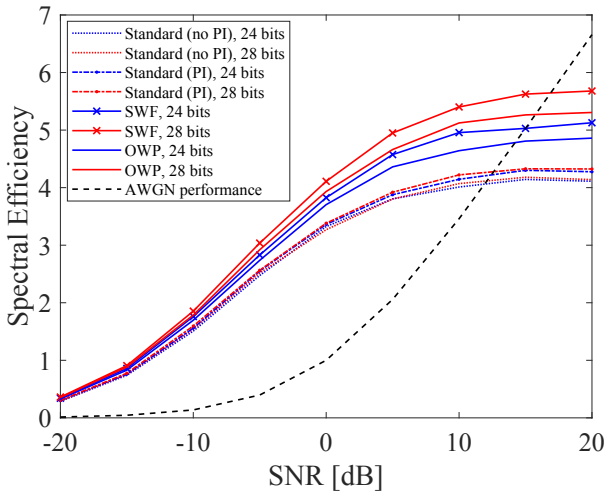

Fig. 6: Spectral efficiency with 4 users: SWF and OWP versus Standard with pseudoinversion (PI) and without pseudoinversion (no PI) compared to single user SISO AWGN scenario.

APPENDIX

To prove the gain of the sequential wideband feedback method over the normal approach where the eigenvectors are quantized separately in parallel, we first interpret the quantization processes into optimization problems. [5]

Taking any strong eigenvector $\mathbf{u}_k \in \mathbf{U}$ as an example, the original method is aimed to find out the optimal codeword $\mathbf{w}_k$ that quantizes it with lowest distortion (highest inner product). The optimization problem is thus written as

$$\text{P1:} \quad \max_{\mathbf{w}_k \in \mathcal{V}_K} \quad \mathbf{w}_k^{\mathrm{H}} \mathbf{u}_k \tag{41}$$

while the optimization problem for the sequential wideband quantization approach becomes

$$\text{P2:} \quad \max_{\mathbf{w}_k \in \mathcal{V}_K} \quad \mathbf{w}_k^{\mathrm{H}} \mathbf{\Pi}_{k-1}^{\perp \mathrm{H}} \mathbf{R} \mathbf{\Pi}_{k-1}^{\perp} \mathbf{w}_k, \tag{42}$$

where $\mathcal{V}_K$ refers to any vector codebook consisting of codewords in dimension $N_t \times 1$.

When all the k eigenvectors are taken into account sequentially, problem (42) becomes

$$\begin{aligned} \text{S1:} \quad \min_{\mathbf{w}_k \in \mathcal{V}_K} \quad & 1 - \mathrm{Tr}\left(\left[\mathbf{W}_{k-1}, \mathbf{w}_k\right]^{\mathrm{H}} \mathbf{R} \left[\mathbf{W}_{k-1}, \mathbf{w}_k\right] \right) \\ \text{s.t.} \quad & \mathbf{w}_k^{\mathrm{H}} \mathbf{W}_{k-1} = 0, \end{aligned} \tag{43}$$

where $\mathbf{W}_{k-1} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{k-1}]$ is derived in previous phase and hence is fixed. Minimizing the objective function of (43) equals to maximizing $\mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{H}} \mathbf{R} \mathbf{W}_k\right)$ given by

$$\mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{H}} \mathbf{R} \mathbf{W}_k\right) = \mathbf{w}_k^{\mathrm{H}} \mathbf{R} \mathbf{w}_k + \underbrace{\mathrm{Tr}\left(\mathbf{W}_{k-1}^{\mathrm{H}} \mathbf{R} \mathbf{W}_{k-1}\right)}_{\text{fixed}}. \tag{44}$$

Since the second term of (44) is fixed, the objective simplifies into $\mathbf{w}_k^{\mathrm{H}} \mathbf{R} \mathbf{w}_k$. Because $\mathbf{W}_{k-1}^{\mathrm{H}} \mathbf{w}_k = 0$, $\mathbf{w}_k$ becomes

$$\mathbf{w}_k = (\mathbf{I} - \mathbf{W}_{k-1} \mathbf{W}_{k-1}^{\mathrm{H}}) \mathbf{w}_k = \mathbf{\Pi}_{k-1}^{\perp} \mathbf{w}_k, \tag{45}$$

[5]For briefness, here $\mathbf{W}_{\mathrm{Q}}$ and $\tilde{\mathbf{R}}$ are simplified into $\mathbf{W}$ and $\mathbf{R}$.

where the projectors are defined in the manner that $\mathbf{\Pi}_{\mathbf{W}} = \mathbf{W}\mathbf{W}^{\mathrm{H}}$ and $\mathbf{\Pi}_{\mathbf{W}}^{\perp} = \mathbf{I} - \mathbf{W}\mathbf{W}^{\mathrm{H}}$. Substitute (45) into the simplified objective function[6], so that we obtain the optimization problem for the sequential wideband quantization approach as

$$\text{S2:} \quad \max_{\bar{\mathbf{w}}_k \in \mathcal{V}_K} \quad \frac{\bar{\mathbf{w}}_k^{\mathrm{H}} \mathbf{\Pi}_{k-1}^{\perp} \mathbf{R} \mathbf{\Pi}_{k-1}^{\perp} \bar{\mathbf{w}}_k}{\bar{\mathbf{w}}_k^{\mathrm{H}} \mathbf{\Pi}_{k-1}^{\perp} \bar{\mathbf{w}}_k}, \tag{46}$$

where $\mathbf{w}_k = \frac{\bar{\mathbf{w}}_k}{\|\mathbf{\Pi}_{k-1}^{\perp} \bar{\mathbf{w}}_k\|}$ is then the orthonormalized codeword.

Assuming a codeword $\mathbf{w}_j$ close to $\mathbf{u}_j$ with squared chordal distance give by $d_c^2(\mathbf{u}_j, \mathbf{w}_j) = 1 - |\mathbf{w}_j^{\mathrm{H}} \mathbf{u}_j| = \varepsilon_j \geq 0$, one can expand $\mathbf{w}_j$ in the eigen basis $\mathbf{U}$ as

$$\begin{aligned} \mathbf{w}_j &= \sqrt{1-\varepsilon_j}\mathbf{u}_j + \sqrt{\varepsilon_j} \sum_{k \neq j} x_{j,k} \mathbf{u}_k \\ &= \sqrt{1-\varepsilon_j}\mathbf{u}_j + \sqrt{\varepsilon_j}\mathbf{U}_j^{\perp} \mathbf{x}_j = \mathbf{U}\mathbf{z}_j \end{aligned} \tag{47}$$

where $\mathbf{x}_j$ is a $(K-1) \times 1$ unit norm vector with $x_{j,k}$, if $k \neq j$ as elements. $\mathbf{U}_j^{\perp}$ is $\mathbf{U}$ with $\mathbf{u}_j$ removed. $\mathbf{z}_j$ is a coordinate vector of $\mathbf{w}_j$ in $\mathbf{U}$. The squared chordal distance is

$$d_c^2(\mathbf{w}_j, \mathbf{u}_k) = 1 - \varepsilon_j |x_{j,k}|^2, \text{ for } j \neq k. \tag{48}$$

Moreover the norm of the codeword $\mathbf{w}_j$ is given by

$$\|\mathbf{w}_j\|^2 = \mathbf{w}_j^{\mathrm{H}} \mathbf{w}_j = (1 - \varepsilon_j) + \varepsilon_j \mathbf{x}_j^{\mathrm{H}} \mathbf{x}_j = 1.$$

In terms of ε_j and $\mathbf{x}_j$, projection distance is given by

$$\begin{aligned} d_{\mathbf{\Pi}} &= 1 - \mathrm{Tr}\left(\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{R}\right) = 1 - \mathrm{Tr}\left(\mathbf{W}^{\mathrm{H}}\mathbf{R}\mathbf{W}\right) \\ &= 1 - \sum_j \mathbf{w}_j^{\mathrm{H}} \mathbf{R} \mathbf{w}_j = 1 - \sum_{j,k} \lambda_k |\mathbf{w}_j^{\mathrm{H}} \mathbf{u}_k|^2. \end{aligned} \tag{49}$$

Consequently, we derive that

$$\mathbf{w}_j^{\mathrm{H}} \mathbf{u}_k = \begin{cases} \sqrt{\varepsilon_j} x_{j,k}, & \text{if } j \neq k, \\ \sqrt{1-\varepsilon_j} x_{j,k}, & \text{if } j = k, \end{cases} \tag{50}$$

so that we derive that $\sum_j |\mathbf{w}_j^{\mathrm{H}} \mathbf{u}_k|^2 = 1 - \varepsilon_k + \sum_{j \neq k} \varepsilon_j |x_{j,k}|^2$. Hence, it holds true that

$$\begin{aligned} d_{\mathbf{\Pi}} &= 1 - \sum_k \lambda_k \left(1 - \varepsilon_k + \sum_{j \neq k} \varepsilon_j |x_{j,k}|^2 \right) \\ &= \underbrace{1 - \sum_k \lambda_k}_{=0} + \sum_k \lambda_k \left(\varepsilon_k - \sum_{j \neq k} \varepsilon_j |x_{j,k}|^2 \right) \leq \sum_k \lambda_k \varepsilon_k \end{aligned} \tag{51}$$

The upper bound on $d_{\mathbf{\Pi}}$ is thus minimized by minimizing ε_k which is the deviation of quantizing $\mathbf{u}_k$. Hence we derive

$$d_{\mathbf{\Pi}} = \sum_j \varepsilon_j (\lambda_j - \sum_{j \neq k} |\lambda_k x_{j,k}|^2) = \sum_j \varepsilon_j (\lambda_j - \mathbf{x}_j^{\mathrm{H}} \mathbf{\Lambda}_j^{\perp} \mathbf{x}_j), \tag{52}$$

where $\mathbf{x}_j$ is the $(K-1) \times 1$ vector defined previously, and $\mathbf{\Lambda}_j^{\perp}$ is the diagonal matrix of eigenvalues with λ_j removed. From (52), we see that to minimize contribution from $j > 1$, minimizing ε_j may not be the best. It shows that if $\varepsilon_j \neq 0$, it is possible to reduce $d_{\mathbf{\Pi}}$ via proper choice of $\mathbf{x}_j$. Also, for $j > 1$, it is possible that $\varepsilon_j = 0$ is not optimal, if $\varepsilon_1 \neq 0$.

[6]Normalization is needed on top of (44). $\bar{\mathbf{w}}_k$ is a auxiliary variable.

Now we analyze the sequential selection approach in decreasing order of eigenvalues. Quantization of $\mathbf{u}_1$ becomes

$$\begin{aligned} \min \quad & d_{\mathbf{\Pi},1} = \varepsilon_1(\lambda_1 - \mathbf{x}_1^{\mathrm{H}}\mathbf{\Lambda}_1^{\perp}\mathbf{x}_1) \\ \text{s.t.} \quad & \mathbf{w}_1 \in \mathcal{V},\ \lambda_1 \geq \lambda_K,\ \text{if } K > 1. \end{aligned}$$

Minimizing ε_1 is not far from optimal. If we recall (46), finding $\mathbf{w}_k$ is then a search over $\mathbf{w}_k = \frac{\bar{\mathbf{w}}_k}{\|\mathbf{\Pi}_{k-1}^{\perp}\bar{\mathbf{w}}_k\|}$ with $\bar{\mathbf{w}}_k \in \mathcal{V}$, so that it equals to a search over $\mathbf{w}_k \in \mathcal{V}_K$ where the vectors in $\mathcal{V}_K$ are normalized as well as those in $\mathcal{V}$, i.e.,

$$\text{S2}': \quad \max_{\mathbf{w}_k \in \mathcal{V}_K} \quad \mathbf{w}_k \mathbf{\Pi}_{k-1}^{\perp}\mathbf{R}\mathbf{\Pi}_{k-1}^{\perp}\mathbf{w}_k. \tag{53}$$

A good approximation would then be to find the $\mathbf{w}_k \in \mathcal{V}_K$ closest to the largest eigenvector of $\mathbf{\Pi}_{k-1}^{\perp}\mathbf{R}\mathbf{\Pi}_{k-1}^{\perp}$.

Recall $\mathbf{w}_j = \mathbf{U}\mathbf{z}_j$ in (47), we then get $\mathbf{W}_{K-1} = \mathbf{U}\mathbf{Z}_{K-1}$ with $\mathbf{W}_{K-1}, \mathbf{Z}_{K-1}$ in $N_t \times (K-1)$. Moreover, we obtain

$$\mathbf{\Pi}_{K-1}^{\perp} = \mathbf{I} - \mathbf{W}_{K-1}\mathbf{W}_{K-1}^{\mathrm{H}} = \mathbf{U}\underbrace{\left(\mathbf{I} - \mathbf{Z}_{K-1}\mathbf{Z}_{K-1}^{\mathrm{H}}\right)}_{\triangleq \tilde{\mathbf{\Pi}}_{K-1}}\mathbf{U}^{\mathrm{H}}, \tag{54}$$

so that we get $\mathbf{\Pi}_{K-1}^{\perp}\mathbf{R}\mathbf{\Pi}_{K-1}^{\perp} = \mathbf{U}\tilde{\mathbf{\Pi}}_{K-1}\mathbf{\Lambda}\tilde{\mathbf{\Pi}}_{K-1}\mathbf{U}^{\mathrm{H}}$. Thus the problem is reformulated to

$$\text{S2}'': \quad \max_{\mathbf{w}_k \in \tilde{\mathcal{V}}_K} \quad \mathbf{w}_k^{\mathrm{H}}\tilde{\mathbf{\Pi}}_{k-1}^{\perp}\mathbf{\Lambda}\tilde{\mathbf{\Pi}}_{k-1}^{\perp}\mathbf{w}_k, \tag{55}$$

where $\tilde{\mathcal{V}}_K = \left\{ \frac{\mathbf{U}^{\mathrm{H}}\mathbf{W}}{\|\mathbf{\Pi}_{k-1}^{\perp}\mathbf{W}\|} \,\middle|\, \mathbf{W} \in \mathcal{V} \right\}$. Note that the eigenvalues of $\tilde{\mathbf{\Pi}}_{K-1}^{\perp}\mathbf{\Lambda}\tilde{\mathbf{\Pi}}_{K-1}^{\perp}$, $\tilde{\mathbf{\Pi}}_{K-1}^{\perp}\mathbf{\Lambda}$, $\mathbf{\Pi}^{\perp}\mathbf{R}\mathbf{\Pi}^{\perp}$ and $\mathbf{\Pi}^{\perp}\mathbf{R}$ are the same. The eigenvalues of the projectors are $(K-1)\times 0$ and $(N_t - K + 1)\times 1$.

Similarly, we extend the calculation to the case when $k = 2$. We have $\mathbf{w}_1 = \mathbf{U}\left[\sqrt{1-\varepsilon}, \sqrt{\varepsilon}\mathbf{x}\right]^{\mathrm{T}}$ with $\mathbf{x}^{\mathrm{H}}\mathbf{x} = 1$, so that $\mathbf{z}_1 = \left[\sqrt{1-\varepsilon}, \sqrt{\varepsilon}\mathbf{x}\right]^{\mathrm{T}}$ and $\tilde{\mathbf{\Pi}}_1^{\perp} = \mathbf{I} - \mathbf{z}_1\mathbf{z}_1^{\mathrm{H}}$. To find the eigen structure of the projector $\tilde{\mathbf{\Pi}}_1^{\perp}$, we first define a unitary matrix $\mathbf{V}$ subject to $\mathbf{V}\mathbf{z}_1 = \mathbf{e}_1 = (1, 0, \ldots, 0)^{\mathrm{T}}$. Such a diagonalizer is directly of the form $\mathbf{V} = [\mathbf{z}_1, \mathbf{z}^{\perp}]^{\mathrm{H}}$ where $\mathbf{z}_1^{\mathrm{H}}\mathbf{z}^{\perp} = 0$. Now one can find the representation of $\mathbf{z}^{\perp}$ directly given by $\mathbf{z}^{\perp} = \left[\sqrt{\varepsilon}\mathbf{x}^{\mathrm{H}}, \mathbf{M}^{\mathrm{H}}\right]^{\mathrm{T}}$, so that we derive

$$\mathbf{V} = \begin{bmatrix} \sqrt{1-\varepsilon} & \sqrt{\varepsilon}\mathbf{x}^{\mathrm{H}} \\ \sqrt{\varepsilon}\mathbf{x} & \mathbf{M} \end{bmatrix},$$

$$\mathbf{V}^{\mathrm{H}}\mathbf{V} = \begin{bmatrix} 1 & \sqrt{\varepsilon}\left(\sqrt{1-\varepsilon}\mathbf{x} + \mathbf{M}^{\mathrm{H}}\mathbf{x}\right)^{\mathrm{H}} \\ \sqrt{\varepsilon}\left(\sqrt{1-\varepsilon}\mathbf{x} + \mathbf{M}^{\mathrm{H}}\mathbf{x}\right) & \varepsilon\mathbf{x}\mathbf{x}^{\mathrm{H}} + \mathbf{M}^{\mathrm{H}}\mathbf{M} \end{bmatrix}.$$

In this case, we need to prove that $\left(\sqrt{1-\varepsilon}\ \mathbf{I} + \mathbf{M}^{\mathrm{H}}\right)\mathbf{x} = 0$. We can expand $\mathbf{M}^{\mathrm{H}}$ into two parts, i.e., $\mathbf{M}^{\mathrm{H}} = \mathbf{A}\mathbf{\Pi}_{\mathbf{x}} + \mathbf{B}\mathbf{\Pi}_{\mathbf{x}}^{\perp}$, where $\mathbf{\Pi}_{\mathbf{x}} = \mathbf{x}\mathbf{x}^{\mathrm{H}}$ and $\mathbf{\Pi}_{\mathbf{x}}^{\perp} = \mathbf{I} - \mathbf{\Pi}_{\mathbf{x}}$ in $N_t - 1$ dimensions. Recall that $\mathbf{x}^{\mathrm{H}}\mathbf{x} = 1$, we thus need $\sqrt{1-\varepsilon}\ \mathbf{I} + \mathbf{A} = \mathbf{0}$ on the part of $\mathbf{M}^{\mathrm{H}}$ that projects to $\mathbf{x}$, so that $\mathbf{A} = -\sqrt{\mathbf{1}-\varepsilon}\ \mathbf{I}$ and $\mathbf{M} = -\sqrt{1-\varepsilon}\ \mathbf{\Pi}_{\mathbf{x}} + \mathbf{\Pi}_{\mathbf{x}}^{\perp}\mathbf{B}^{\mathrm{H}}$.

Then one can analyze the unitary in lower right corner as

$$\begin{aligned} \varepsilon\mathbf{\Pi}_{\mathbf{x}} + \mathbf{M}^{\mathrm{H}}\mathbf{M} &= \varepsilon\mathbf{\Pi}_{\mathbf{x}} + \left(-\sqrt{1-\varepsilon}\ \mathbf{\Pi}_{\mathbf{x}} + \mathbf{B}\mathbf{\Pi}_{\mathbf{x}}^{\perp}\right) \\ &\quad \times \left(-\sqrt{1-\varepsilon}\ \mathbf{\Pi}_{\mathbf{x}} + \mathbf{\Pi}_{\mathbf{x}}^{\perp}\mathbf{B}^{\mathrm{H}}\right) \\ &\overset{(a)}{=} \varepsilon\mathbf{\Pi}_{\mathbf{x}} + \sqrt{1-\varepsilon}^2\ \mathbf{\Pi}_{\mathbf{x}} + \mathbf{B}\mathbf{\Pi}_{\mathbf{x}}^{\perp}\mathbf{B}^{\mathrm{H}} \\ &= \mathbf{\Pi}_{\mathbf{x}} + \mathbf{B}\mathbf{\Pi}_{\mathbf{x}}^{\perp}\mathbf{B}^{\mathrm{H}} = \mathbf{I}_{N_t-1} \overset{(b)}{=} \mathbf{\Pi}_{\mathbf{x}} + \mathbf{\Pi}_{\mathbf{x}}^{\perp}, \end{aligned} \tag{56}$$

where (a) is obtained by definition $\mathbf{\Pi}_{\mathbf{x}}\mathbf{\Pi}_{\mathbf{x}}^{\perp} = \mathbf{\Pi}_{\mathbf{x}}^{\perp}\mathbf{\Pi}_{\mathbf{x}} = 0$. Because of (b), we derive that $\mathbf{B} = \mathbf{I}$, so that

$$\mathbf{M} = -\sqrt{1-\varepsilon}\ \mathbf{\Pi}_{\mathbf{x}} + \mathbf{I} - \mathbf{\Pi}_{\mathbf{x}} = \mathbf{I}_{N_t-1} - \left(1 + \sqrt{1-\varepsilon}\right)\mathbf{\Pi}_{\mathbf{x}}. \tag{57}$$

Substituting (57) in to the structure of $\mathbf{V}$, $\mathbf{V}$ becomes

$$\mathbf{V} = \begin{bmatrix} \sqrt{1-\varepsilon} & \sqrt{\varepsilon}\mathbf{x}^{\mathrm{H}} \\ \sqrt{\varepsilon}\mathbf{x} & \mathbf{I}_{N_t-1} - (1+\sqrt{1-\varepsilon})\mathbf{\Pi}_{\mathbf{x}} \end{bmatrix} \tag{58}$$

which is unitary and Hermitian. Moreover, it follows that

$$\mathbf{V}\left[\sqrt{1-\varepsilon}, \sqrt{\varepsilon}\mathbf{x}\ \right]^{\mathrm{T}} = [1, 0, \ldots, 0]^{\mathrm{T}},$$

which agrees with the definition $\mathbf{V}\mathbf{z}_1 = \mathbf{e}_1$ previously. This $\mathbf{V}$ is not good, however, as $\mathbf{V} \overset{\varepsilon\to\mathbf{0}}{\longrightarrow} \mathrm{diag}(\{\mathbf{1}, \mathbf{I_{N_t-1}} - \mathbf{2\Pi_x}\}) \neq \mathbf{I_{N_t}}$. Set $\tilde{\mathbf{V}} \triangleq \mathbf{V}\,\mathrm{diag}(\{\mathbf{1}, \mathbf{I_{N_t-1}} - \mathbf{2\Pi_x}\})$, so that $\tilde{\mathbf{V}} \overset{\varepsilon\to 0}{\longrightarrow} \mathbf{I}_{N_t}$. Note that here $(\mathbf{I}_{N_t-1} - 2\mathbf{\Pi}_{\mathbf{x}})^2 = \mathbf{I} - 4\mathbf{\Pi}_{\mathbf{x}} + 4\mathbf{\Pi}_{\mathbf{x}} = \mathbf{I}$. We obtain that $\tilde{\mathbf{V}}^{\mathrm{H}}\tilde{\mathbf{V}} = \mathbf{I}$ and $\mathbf{z}_1 = \tilde{\mathbf{V}}\mathbf{e}_1 = \mathbf{V}\mathbf{e}_1$. What of interests to us are the eigenvalues of $\tilde{\mathbf{\Pi}}_1^{\perp}\mathbf{\Lambda}\tilde{\mathbf{\Pi}}_1^{\perp}$ where $\tilde{\mathbf{\Pi}}_1^{\perp} = \tilde{\mathbf{V}}(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^{\mathrm{H}})\tilde{\mathbf{V}}^{\mathrm{H}} \triangleq \tilde{\mathbf{\Pi}}_{1,D}^{\perp}$. The eigen structure becomes

$$\mathbf{L} = \tilde{\mathbf{\Pi}}_{\mathbf{1,D}}^{\perp}\tilde{\mathbf{V}}^{\mathrm{H}}\mathbf{\Lambda}\tilde{\mathbf{V}}\tilde{\mathbf{\Pi}}_{\mathbf{1,D}}^{\perp}. \tag{59}$$

Because only lower right corner survives in $\tilde{\mathbf{\Pi}}_{1,D}^{\perp}$, i.e., $\tilde{\mathbf{\Pi}}_{1,D}^{\perp} = \mathrm{diag}(\{0, 1\})$, we obtain

$$\mathbf{L} = \mathrm{diag}(\{0,\ (\mathbf{I} - 2\mathbf{\Pi}_{\mathbf{x}})[\varepsilon\lambda_1\mathbf{\Pi}_{\mathbf{x}} + \mathbf{M}\mathbf{\Lambda}^{\perp}\mathbf{M}](\mathbf{I} - 2\mathbf{\Pi}_{\mathbf{x}})\}),$$

where $\mathbf{\Lambda}^{\perp} = \mathrm{diag}(\{\lambda_2, \ldots, \lambda_{N_t}\})$ is a $(N_t - 1)\times(N_t - 1)$ matrix. Define the lower right corner of $\mathbf{L}$ as $\mathbf{L}_{er}$, and obtain

$$\mathbf{M}(\mathbf{I} - 2\mathbf{\Pi}_{\mathbf{x}}) = (\mathbf{I} - 2\mathbf{\Pi}_{\mathbf{x}})\mathbf{M} = \mathbf{I} + (a-2)\mathbf{\Pi}_{\mathbf{x}} \triangleq \mathbf{I} + b\mathbf{\Pi}_{\mathbf{x}},$$

where $a = 1 + \sqrt{1-\varepsilon}$ and $b = \sqrt{1-\varepsilon} - 1$. Thus we get

$$\mathbf{L}_{er} = \varepsilon\lambda_1\mathbf{\Pi}_{\mathbf{x}} + (\mathbf{I} + b\mathbf{\Pi}_{\mathbf{x}})\mathbf{\Lambda}^{\perp}(\mathbf{I} + b\mathbf{\Pi}_{\mathbf{x}}). \tag{60}$$

Now $b = \sqrt{1-\varepsilon} - 1 \approx -\frac{1}{2}\varepsilon + \varphi(\varepsilon^2)$ subject to the constraint that $(\mathbf{I}_{N_t-1} + b\mathbf{\Pi}_{\mathbf{x}}) \overset{\varepsilon\to 0}{\longrightarrow} \mathbf{I}_{N_t-1} - \frac{1}{2}\varepsilon\mathbf{\Pi}_{\mathbf{x}}$. $\mathbf{L}_{er}$ becomes

$$\mathbf{L}_{er} = \mathbf{\Lambda}^{\perp} + \varepsilon(\lambda_1\mathbf{\Pi}_{\mathbf{x}} - \frac{1}{2}\mathbf{\Pi}_{\mathbf{x}}\mathbf{\Lambda}^{\perp} - \frac{1}{2}\mathbf{\Lambda}^{\perp}\mathbf{\Pi}_{\mathbf{x}}), \tag{61}$$

what implies that

- if $\varepsilon \to 0$ (perfect $\mathbf{U}$ quantization), eigen structure of the remaining part is $\mathbf{\Lambda}^{\perp}$.
- if $\varepsilon > 0$, there is a component of λ_1.
- if $\varepsilon < 0$, $\mathbf{U}_1$ quantization has *canceled* a part of $\mathbf{\Lambda}^{\perp}$. [7]

Next, we approximate the eigenvalues and eigenvectors via computation utilizing *perturbation* theory.

Denote $\mathbf{H} = \lambda_1\mathbf{\Pi}_{\mathbf{x}} - \frac{1}{2}\mathbf{\Pi}_{\mathbf{x}}\mathbf{\Lambda}^{\perp} - \frac{1}{2}\mathbf{\Lambda}^{\perp}\mathbf{\Pi}_{\mathbf{x}}$. The unperturbed system has eigenvalues $\mathbf{\Lambda}^{\perp}$ and eigen vectors $\tilde{\mathbf{e}}_j, j = 2, \ldots, N_t$ which are unit vectors in $N_t - 1$ dimensions. The largest eigen values is λ_2, and $\tilde{\mathbf{e}}_2 = [1, \mathbf{0}]^{\mathbf{T}} \in \mathbb{C}^{\mathbf{N_t-1}}$. The perturbed eigenvalues for $j = 2, \ldots, N_t$ become

$$\tilde{\lambda}_j = \lambda_j + \varepsilon\tilde{\mathbf{e}}_j^{\mathrm{H}}\mathbf{H}\mathbf{e}_j = \lambda_j + \varepsilon(\lambda_1 - \lambda_j)(\mathbf{\Pi}_{\mathbf{x}})_{j,j} \tag{62}$$

Here $(\mathbf{\Pi}_{\mathbf{x}})_{j,j}$ is the jth diagonal element of $\mathbf{x}\mathbf{x}^{\mathrm{H}}$ where $\mathbf{x}_{k,j} = \mathbf{w}_k^{\mathrm{H}}\mathbf{u}_j/\sqrt{\varepsilon_j}$, and $(\mathbf{\Pi}_{\mathbf{x}})_{j,j} = |\mathbf{x}_{1,j}|^2$. We thus have

$$\tilde{\lambda}_j = \lambda_j + \varepsilon(\lambda_1 - \lambda_j)|\mathbf{x}_{1,j}|^2 = \lambda_j + (\lambda_1 - \lambda_j)\mathbf{w}_1^{\mathrm{H}}\mu_j \geq \lambda_j.$$

[7] The corresponding part is removed, i.e. $-\mathbf{\Pi}_x\mathbf{\Lambda}^{\perp} - \mathbf{\Lambda}^{\perp}\mathbf{\Pi}_x$.

The first form gives the perturbed eigenvalues in term of quantization error ε. $|\mathbf{x}_{1,j}|^2$ can be assumed uniformly distributed across the other eigenvalues. Since $\mathbb{E}\{|\mathbf{x}_{1,j}|^2\} = 1/(N_t-1)$, we obtain $\sum_j |\mathbf{x}_{1,j}|^2 = 1$ with $\mathbf{x}$ characterizing the distribution of ε over the other eigenbeams. The second form gives a precise expectation when $\mathbf{w}_1$ and $\mathbf{U}$ are known. Note that $\mathbf{I} = \mathbf{\Pi}_1 + \mathbf{\Pi}_1^{\perp}$, thus it holds true

$$\mathrm{Tr}(\mathbf{R}) = \mathrm{Tr}(\mathbf{\Lambda}) = \mathrm{Tr}\left(\mathbf{\Pi}_1\mathbf{R}\mathbf{\Pi}_1 + \mathbf{\Pi}_1^{\perp}\mathbf{R}\mathbf{\Pi}_1^{\perp}\right).$$

Let $\tilde{\lambda}_1 = \mathrm{Tr}\left(\mathbf{\Pi}_1\mathbf{R}\mathbf{\Pi}_1\right) = \mathbf{w}_1\mathbf{R}\mathbf{w}_1 = \mathbf{z}_1^{\mathrm{H}}\mathbf{\Lambda}\mathbf{z}_1 = (1-\varepsilon)\lambda_1 + \varepsilon\mathbf{x}^{\mathrm{H}}\mathbf{\Lambda}^{\perp}\mathbf{x}$. The second part becomes

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{\Pi}_1^{\perp}\mathbf{R}\mathbf{\Pi}_1^{\perp}\right) = \mathrm{Tr}\left(\tilde{\mathbf{\Pi}}_1^{\perp}\mathbf{\Lambda}\tilde{\mathbf{\Pi}}_1\right) &\approx \sum_{j=2}^{N_t}\tilde{\lambda}_j \\ &= \sum_{j=2}^{N_t}\lambda_j + \varepsilon\lambda_1\underbrace{\sum_{j=2}^{N_t}|\mathbf{x}_{1,j}|^2}_{=1} + \varepsilon\mathbf{x}^{\mathrm{H}}\mathbf{\Lambda}^{\perp}\mathbf{x}. \quad (63)\end{aligned}$$

Thus $\tilde{\lambda}_1 + \sum_{j=2}^{N_t}\tilde{\lambda}_j = \sum_{j=1}^{N_t}\lambda_j$. Trace is preserved in this first order approximation. The perturbation could be find without diagonalization, so that $\mathbf{\Pi}_1^{\perp} = \mathbf{U}(\mathbf{I} - \mathbf{z}_1\mathbf{z}_1^{\mathrm{H}})\mathbf{U}^{\mathrm{H}}$ with $\mathbf{z} = \left[\sqrt{1-\varepsilon}, \sqrt{\varepsilon}\mathbf{x}\right]^{\mathrm{T}}$. Hence, it follows that

$$\mathbf{z}\mathbf{z}^{\mathrm{H}} = \begin{bmatrix} 1-\varepsilon & \sqrt{\varepsilon(1-\varepsilon)}\mathbf{x}^{\mathrm{H}} \\ \sqrt{\varepsilon(1-\varepsilon)}\mathbf{x} & \varepsilon\mathbf{x}\mathbf{x}^{\mathrm{H}} \end{bmatrix}. \quad (64)$$

Moreover, we obtain

$$\tilde{\mathbf{\Pi}}_1^{\perp} \approx \begin{bmatrix} 0 & \\ & \mathbf{1} \end{bmatrix} - \sqrt{\varepsilon}\begin{bmatrix} 0 & \mathbf{x}^{\mathrm{H}} \\ \mathbf{x} & 0 \end{bmatrix} - \varepsilon\begin{bmatrix} -1 & \\ & \mathbf{x}\mathbf{x}^{\mathrm{H}} \end{bmatrix}, \quad (65)$$

so that it follows

$$\begin{aligned}\tilde{\mathbf{\Lambda}} = \tilde{\mathbf{\Pi}}_1^{\perp}\mathbf{\Lambda}\tilde{\mathbf{\Pi}}_1^{\perp} &= \begin{bmatrix} 0 & \\ & \mathbf{\Lambda}^{\perp} \end{bmatrix} - \sqrt{\varepsilon}\begin{bmatrix} & \mathbf{x}^{\mathrm{H}}\mathbf{\Lambda}^{\perp} \\ \mathbf{\Lambda}^{\perp}\mathbf{x} & \end{bmatrix} \\ &+\varepsilon\begin{bmatrix} \mathbf{x}^{\mathrm{H}}\mathbf{\Lambda}^{\perp}\mathbf{x} & \\ & \lambda_1\mathbf{x}\mathbf{x}^{\mathrm{H}} \end{bmatrix} + \varepsilon\begin{bmatrix} 0 & \\ & \mathbf{\Pi}_{\mathbf{x}}\mathbf{\Lambda}^{\perp} + \mathbf{\Lambda}^{\perp}\mathbf{\Pi}_{\mathbf{x}} \end{bmatrix}. \quad (66)\end{aligned}$$

According to (66), the second term does not affect the eigen values, but does change the eigenvectors. When building the codewords for minimizing $\mathbf{w}_2$, it holds true in normalization

$$\|\mathbf{\Pi}_1^{\perp}\mathbf{w}\|^2 = \mathbf{w}^{\mathrm{H}}\mathbf{\Pi}_1^{\perp}\mathbf{w} = \mathbf{w}^{\mathrm{H}}\mathbf{w} - |\mathbf{w}_1^{\mathrm{H}}\mathbf{w}|^2 = d_c^2(\mathbf{w}_1, \mathbf{w}). \quad (67)$$

As is proved in (67), the sequential method achieves error self-cancellation by quantizing eigenvectors iteratively, so that it avoids accumulating errors for different eigenvectors which happens when quantizing the eigenvectors separately in parallel in normal wideband quantization. This ends the proof of performance gain of the sequential scheme over the parallel scheme.

REFERENCES

[1] V. Roope, J. Liao, T. Pllaha, W. Han, and O. Tirkkonen, "CSI quantization for fdd massive mimo communication," in *Proc. IEEE VTC-Spring*, 2021, pp. 1–5.

[2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE T. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[3] J. Hamalainen and R. Wichman, "Closed-loop transmit diversity for FDD WCDMA systems," in *Proc. Asilomar Conf.*, vol. 1, 2000, pp. 111–115.

[4] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE T. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.

[5] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE T. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.

[6] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE T. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[7] D. J. Love and R. W. Heath, "Limited feedback diversity techniques for correlated channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 718–722, 2006.

[8] V. Raghavan and R. W. Heath and A. M. Sayeed, "Systematic codebook designs for quantized beamforming in correlated MIMO channels," *IEEE J. Sel. Areas Commun*, vol. 25, no. 7, pp. 1298–1310, 2007.

[9] J. Choi and D. J. Love and T. Kim, "Trellis-extended codebooks and successive phase adjustment: A path from LTE-advanced to FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2007–2016, 2015.

[10] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, 2015.

[11] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4611–4624, 2016.

[12] J. Chen, H. Yin, L. Cottatellucci, and D. Gesbert, "Feedback mechanisms for FDD massive MIMO with D2D-Based limited CSI sharing," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5162–5175, 2017.

[13] J. Nam, A. Adhikary, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[14] J. Chen and V. K. N. Lau, "Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1230–1238, 2014.

[15] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2200–2211, 2015.

[16] D. Gunduz and P. Kerret and & al., "Machine learning in the air," *IEEE J. Selected Areas in Commun.*, vol. 37, no. 10, pp. 2184–2199, 2019.

[17] M. Mashhadi and Q. Yang and D. Gunduz, "Distributed deep convolutional compression for massive MIMO CSI feedback," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2621–2633, 2021.

[18] T. Wang, C. K. Wen, J. Shi, and G. Y. Li, "Deep Learning-Based CSI Feedback Approach for Time-Varying Massive MIMO Channels," *IEEE Wireless Commun. L.*, vol. 8, no. 2, pp. 416–419, 2019.

[19] S. Ghosh, B. D. Rao, and J. R. Zeidler, "Techniques for MIMO channel covariance matrix quantization," *IEEE T. Sign. Proc.*, vol. 60, no. 6, pp. 3340–3345, 2012.

[20] Y. Liu, G. Y. Li, and W. Han, "Quantization and feedback of spatial covariance matrix for massive MIMO systems with cascaded precoding," *IEEE T. Commun.*, vol. 65, no. 4, pp. 1623–1634, 2017.

[21] Samsung et. al, "WF on Type I and II CSI codebooks," 3GPP, Tech. Rep. R1-1709232, 2017.

[22] E. Onggosanusi & al., "Modular and high-resolution channel state information and beam management for 5G New Radio," *IEEE Comm. Mag.*, vol. 56, no. 3, pp. 48–55, 2018.

[23] W. Pearlman, "Polar quantization of a complex Gaussian random variable," *IEEE T. Commun.*, vol. 27, no. 6, pp. 892–899, 1979.

[24] Fraunhofer Heinrich Hertz Institute, "The implementation of quasi deterministic radio channel generator (QuaDRiGa) v2.0.0," https://quadriga-channel-model.de/.

[25] S. D. Howard, A. R. Calderbank, and S. J. Searle, "A fast reconstruction algorithm for deterministic compressive sensing using second order Reed-Muller codes," in *Conf. Information Sciences and Systems*, Mar. 2008, pp. 11–15.