

Rate-Splitting for Overloaded Multigroup Multicast: Error Performance Evaluation

Hongzhi Chen*, De Mi*, Zilong Liu[†], Pei Xiao*, Rahim Tafazolli*

*Institute for Communication Systems, University of Surrey, United Kingdom

[†]School of Computer Science and Electronic Engineering, University of Essex, United Kingdom

Email:{hongzhi.chen, d.mi, p.xiao, r.tafazolli}@surrey.ac.uk, zilong.liu@essex.ac.uk

Abstract—In this work, we provide the first attempt to evaluate error performance of Rate-Splitting (RS) based transmission strategies with constellation-constrained coding/modulation. The consider scenario is an overloaded multigroup multicast, where RS can mitigate the inter-group interference thus achieve a better max-min fair group rate over conventional transmission strategies. We bridge the RS-based rate optimization with modulation-coding scheme selection, and implement them in a developed transceiver framework with either linear or non-linear receiver, where the latter equips with a generalized sphere decoder. Simulation results of a coded bit error rate demonstrate that, while the conventional strategies suffer from the error floor in the considered scenario, the RS-based strategy delivers a superior performance even with low complexity receiver techniques. The proposed analysis, transceiver framework and evaluation methodology provide a generic baseline solution to validate the effectiveness of the RS-based system design in practice.

Index Terms—Rate-splitting, overloaded system, multigroup multicast, rank-deficient, generalized sphere decoder, coded bit error rate.

I. INTRODUCTION

The pioneering work in [1]–[4] revealed the huge potential of multicast/broadcast as a critical technology element in 5G (5th Generation) systems in addition and as a complement to unicast, which sparks a new wave in both research community [5]–[7] and industry [8], [9] of studying next-generation multicast and broadcast technologies, e.g., multigroup multicast [10]. Rate-Splitting (RS) has been proposed to better neutralize the inter-group interference under an overloaded multigroup multicast scenario [10]–[12], leading to a significantly improved performance in maximizing the minimum group rate over classic transmission strategies such as semi-definite relaxation (SDR) based precoding.

Despite the promising performance as shown in [10], [11], it remains untapped to implement RS and validate its effectiveness in practice. One thing hindering the practical application of RS is that in the literature, e.g., for RS-based precoding, it is often assumed Gaussian signalling input, which may reach the optimum theoretically but is rarely feasible in realistic communication systems that usually with constellation constraints and

fixed-length coding schemes. Different solution approaches that may help implementing the RS-based precoding include the optimization of precoders based on finite-alphabet inputs [13], or the optimization of power allocation but with predetermined precoders and constellations [14]. However, the challenge to apply these approaches in the overloaded scenario comes from untraceable optimal precoders. To fill the research gap, we provide the first attempt to bridge the RS-based precoding optimization in the overloaded scenario and modulation-coding scheme (MCS) selection, as an alternative approach, towards the implementation of RS in practice.

In this work, we first reformulate the RS-based max-min fair group rate optimization in [11] with a weighted minimum mean square error (WMMSE) technique. Focusing on a specific achievable rate as target, we break this rate down to group-specific degraded and designated rates, and calculate modulation orders and coding rates that can be used for corresponding degraded and designated streams under a given channel. Having a matching MCS enables us to apply different evaluation methodologies to validate the performance gain of the RS-based strategy over the conventional strategies and to demonstrate the practicality of the strategy itself. Therein has a link level simulation which is widely used for the software implementation of radio access technologies. Metrics output from the computationally intensive link level simulations, such as bit error rate (BER) or block error rate, provide insightful guidelines to both software and hardware implementations of the testing technology. Leveraged on our project work [1], [2], a link level simulator compatible to the latest Long Term Evolution (LTE) and 5G New Radio (NR) standardizations has been developed. To introduce the optimized RS-based transmission strategy into the simulator and initialize its software implementation, we adapt the transceiver processing chain for the considered system, where in the proposed transceiver design we are particularly interested on the effect of having different decoding algorithms. Two types of receivers are investigated including the MMSE-based linear receiver (which is optimized as part of the rate optimization) and non-linear receiver, e.g., the generalized sphere decoder (GSD) which yields a magnificent decoding ability when facing a rank-deficient channel matrix as in overloaded scenarios [15]. The coded BER performance of the RS-based system is evaluated and compared with the conventional precoding system without using RS (NoRS), by using the developed simulator. Two key

We would like to acknowledge the support of the University of Surrey 5GIC (www.surrey.ac.uk/5gic) members for this work. This work was also supported in part by the European Commission under the 5GPPP project 5G-Xcast (H2020-ICT-2016-2 call, grant number 761498). The views expressed in this contribution are those of the authors and do not necessarily represent the project.

findings based on the simulation results are:

- The error performance of the RS strategy outperforms the NoRS strategies, which verifies the rate improvement of the former.
- The NoRS strategy suffers from a high error floor occurring with linear MMSE receiver, which can be overcome by applying the non-linear GSD receiver. However, with RS, both linear and non-linear receivers result in the near-identical error performance, which shows advantages of using RS in the overloaded scenario in further pursuit of lower complexity.

II. SYSTEM MODEL AND OPTIMIZATION PROBLEM FORMULATION

A. System Model

We consider a downlink system where a base station equipped with N_t transmit antennas serves M groups of single-antenna users. The total number of users is K and each group has K/M user(s). Here we assume an overloaded scenario with $1 < N_t < K$, as it falls into the following categories: **1.** From the transmit-side point of view, the total number of transmitter antennas is less than the total number of receiver antennas. **2.** From the receive-side point view, the total number of transmitter antennas is larger than the number of receiver antennas of each individual user. We define the set of users as $\mathbb{K} = \{1, \dots, K\}$, and the group index m where $m \in \{1, \dots, M\}$. A baseband model for input-output relationship at the t -th instant of the considered system, for the k -th user, can be given in the vector format as

$$y_k(t) = \mathbf{h}_k^H \mathbf{x}(t) + n_k(t), \quad (1)$$

where $\mathbf{x}(t) \in \mathbb{C}^{N_t}$ is the transmit vector after applying the precoding, which satisfies a power constraint of $\mathbf{E}\{\mathbf{x}^H(t)\mathbf{x}(t)\} \leq P_t$, $y_k(t)$ is the received signal. $\mathbf{h}_k^H \in \mathbb{C}^{1 \times N_t}$ represents the channel vector between the base station/transmitter and the k -th user/receiver. It is assumed that the channel knowledge is available at both transmitter and receiver, which remains constant over one or more symbol periods. We also consider the receiver noise as the additive white Gaussian noise (AWGN) with variance σ_n^2 , denoted by $n_k(t)$.

We briefly review the concept of RS and its generation procedure of $\mathbf{x}(t)$. Originally, the transmitter intends to pass the message W_m to the m -th group, $m = \{1, \dots, M\}$. With RS, W_m is divided into degraded and designated parts, denoted by $\{W_{m0}, W_{m1}\}$. The degraded part is intended to be decoded by all the groups while the designated part is decoded by a specific group. The degraded part for all are combined, coded and modulated into a single stream, namely degraded stream, e.g., $\{W_{10}, \dots, W_{M0}\} \rightarrow W_c \rightarrow \mathbf{s}_c$. The designated part is individually coded and modulated into M streams, namely designated stream, e.g., $W_{m1} \rightarrow \mathbf{s}_m$. Different streams are then mapped onto the transmit antenna through precoding. The transmit signal can be expressed as following:

$$\mathbf{x}(t) = \mathbf{p}_c \mathbf{s}_c(t) + \sum_{m=1}^M \mathbf{p}_m \mathbf{s}_m(t), \quad (2)$$

where $\mathbf{p}_c \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{p}_m \in \mathbb{C}^{N_t \times 1}$ are the precoding vector for the degraded and designated streams, respectively. Define $\mathbf{S}(t) \triangleq [\mathbf{s}_c(t), \mathbf{s}_1(t) \dots \mathbf{s}_M(t)]^T$ as the concatenated RS stream, where all the entries have zero mean and unit variance.

B. Problem Formulation

Due to the constant channel over symbols, the received power remains the same for the duration of the channel. Hence, we can drop the transmission instant index t , and express the average received power for a given channel at the k -th receiver as:

$$T_{c,k} = \underbrace{|\mathbf{h}_k^H \mathbf{p}_c|^2}_{S_{c,k}} + \underbrace{|\mathbf{h}_k^H \mathbf{p}_{g(k)}|^2}_{S_k} + \underbrace{\sum_{j=1, j \neq g(k)}^M |\mathbf{h}_k^H \mathbf{p}_{g(j)}|^2}_{I_k} + \sigma_n^2, \quad (3)$$

where a mapping function $g(k)$ executes the user-to-group mapping, and $S_{c,k}$, S_k , I_k denote the received power for the degraded stream, for the desired designated stream and for all the unwanted components, respectively. Hence, T_k represents the interference for the degraded stream reception. At the receive side, the degraded stream is first decoded by treating all the interference T_k as noise. The k -th user estimates the degraded stream as: $\hat{\mathbf{s}}_{c,k} = \mu_{c,k} \mathbf{y}_k$ where $\mu_{c,k}$ represents the equalizer for the degraded stream. After decoding the degraded part, successive interference cancellation (SIC) is applied to remove the degraded stream from \mathbf{y}_k before decoding the designated part. The estimation of $\mathbf{s}_{g(k)}$ is provided as: $\hat{\mathbf{s}}_{g(k)} = \mu_k (\mathbf{y}_k - \mathbf{h}_k^H \mathbf{p}_c \hat{\mathbf{s}}_{c,k})$, where μ_k represents the equalizer for the k -th designated stream. The mean square error from the k -th user's perspective for its degraded and designated stream are defined as: $\varepsilon_{c,k} \triangleq \mathbf{E}\{|\hat{\mathbf{s}}_{c,k} - \mathbf{s}_c|^2\}$ and $\varepsilon_k \triangleq \mathbf{E}\{|\hat{\mathbf{s}}_{g(k)} - \mathbf{s}_{g(k)}|^2\}$, respectively. The optimum equalizer can be calculated following the MMSE algorithm, as $\mu_{c,k}^{\text{MMSE}} = \mathbf{p}_c^H \mathbf{h}_k(T_{c,k})^{-1}$ and $\mu_k^{\text{MMSE}} = \mathbf{p}_{g(k)}^H \mathbf{h}_k(T_k)^{-1}$. Therefore, $\varepsilon_{c,k}^{\text{MMSE}} = (T_{c,k})^{-1} T_k$ and $\varepsilon_k^{\text{MMSE}} = (T_k)^{-1} I_k$, respectively. The k -th user experiences an average signal-to-interference-plus-noise (SINR) as: $\gamma_{c,k} = S_{c,k}/T_k = (\varepsilon_{c,k}^{\text{MMSE}})^{-1} - 1$ and $\gamma_k = S_k/I_k = (\varepsilon_k^{\text{MMSE}})^{-1} - 1$. Then we can formulate the corresponding achievable rate as: $R_{c,k} = \log_2(1 + \gamma_{c,k}) = -\log_2(\varepsilon_{c,k}^{\text{MMSE}})$ and $R_k = \log_2(1 + \gamma_k) = -\log_2(\varepsilon_k^{\text{MMSE}})$. The transmission rate of the degraded stream is defined as $R_c = \sum_{m=1}^M C_m$ where C_m denotes the corresponding transmission rate that is contributed from the m -th group. Considering that the degraded stream should be successfully decoded by all users, it should satisfy the condition that $R_c = \min_{k \in \mathbb{K}} R_{c,k}$. Same applies to the transmission rate of the designated stream for each group, i.e., $\min_{m=g(k)} R_k$, for all the users (e.g., the k -th user here) belong to the m -th group. Therefore, the group rate of the m -th group can be expressed as: $R_m = C_m + \min_{m=g(k)} R_k$. In this paper, we consider the max-min fairness (MMF) optimization. The optimal precoder is obtained by solving the following

MMF problem:

$$\begin{aligned} \max_{\mathbf{C}, \mathbf{P}} \quad & \min(C_m + \min_{m=g(k)} R_k) \\ \text{s.t.} \quad & R_{c,k} \geq \sum_{m=1}^M C_m, \end{aligned} \quad (4a)$$

$$C_m \geq 0, \quad \forall m \quad (4b)$$

$$\|\mathbf{p}_c\|^2 + \sum_{m=1}^M \|\mathbf{p}_m\|^2 \leq P_t \quad (4c)$$

where, $\mathbf{C} \triangleq [C_1, \dots, C_M]$ is the concatenated vector for $C_m, \forall m$. $\mathbf{P} \triangleq [\mathbf{p}_c, \mathbf{p}_1, \dots, \mathbf{p}_M]$ is the concatenated matrix for \mathbf{p}_c and \mathbf{p}_m . The constraints shown in (4a) and (4b) guarantee the decodability and non-negative rate of the degraded stream for each group/user.

III. FROM RATE OPTIMIZATION TO MODULATION-CODING SELECTION

In this section, we first provide the method of acquiring the optimized RS precoder and equalizer that giving the best MMF rate. Then we select the modulation order and coding rates of each stream that giving the best-possible match with the optimization results. The selected MCS will be later used in the error performance evaluation.

A. Rate optimization and precoder/equalizer acquisition

The rate-to-weighted MMSE (Rate-WMMSE) relationship is considered to reformulate the problem [16], for acquiring the optimum precoder and equalizer. We define the augmented weighted-MSE for the k -th user as:

$$\zeta_{c,k} = u_{c,k} \varepsilon_{c,k} - \log_2(u_{c,k}) \quad (5)$$

$$\zeta_k = u_k \varepsilon_k - \log_2(u_k) \quad (6)$$

where $u_{c,k}$ and u_k are the non-negative valued weights for the degraded and designated streams. With a given set of precoders, the MSEs $\varepsilon_{c,k}$ and ε_k become functions of $g_{c,k}$ and g_k , respectively. The Rate-WMMSE relationship can be established with the optimized equalizer and weights, such that:

$$\zeta_{c,k}^{opt} = \min_{u_{c,k}, g_{c,k}} \zeta_{c,k} = 1 - R_{c,k} \quad (7)$$

$$\zeta_k^{opt} = \min_{u_k, g_k} \zeta_k = 1 - R_k \quad (8)$$

The optimum equalizer and weights are obtained by taking account the first order optimality conditions. More specifically, $g_{c,k}^{opt} = g_{c,k}^{\text{MMSE}}$, $g_k^{opt} = g_k^{\text{MMSE}}$, $u_{c,k}^{opt} = (\varepsilon_{c,k}^{\text{MMSE}})^{-1}$, and $u_k^{opt} = (\varepsilon_k^{\text{MMSE}})^{-1}$. Therefore, the problem (4) is equivalent to:

$$\begin{aligned} \max_{\mathbf{C}, \mathbf{P}, \mathbf{g}, \mathbf{u}, r_g} \quad & r_g \\ \text{s.t.} \quad & (4b), (4c) \end{aligned} \quad (9a)$$

$$C_m + R_m \geq r_g, \quad \forall m, \forall k \quad (9a)$$

$$1 - \zeta_{c,k} \geq \sum_{m=1}^M C_m, \quad \forall m, \forall k \quad (9b)$$

$$1 - \zeta_{g(k)} \geq R_m, \quad \forall k, \forall m, \forall k \quad (9c)$$

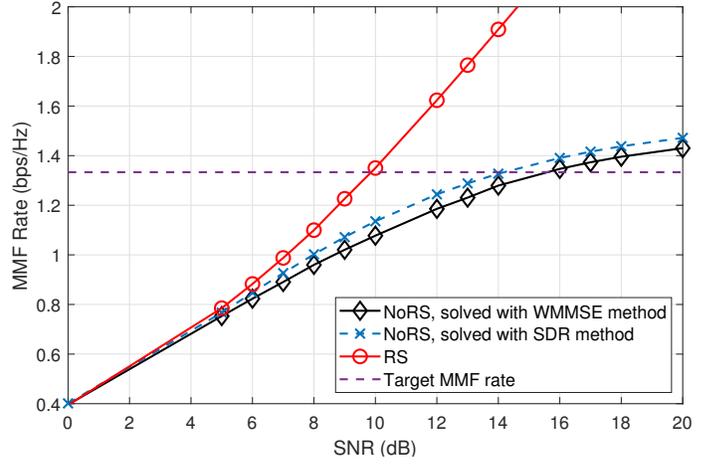


Fig. 1: MMF rate performances. $N_t = 4$ antennas, $M = 3$ groups, $K = 9$ users.

where r_g is an auxiliary variable, denoting the lowest user experienced rate to be maximized. $\mathbf{g} \triangleq \{g_{c,k}, g_k | k \in \mathbb{K}\}$ and $\mathbf{u} \triangleq \{u_{c,k}, u_k | k \in \mathbb{K}\}$ are the sets of equalizer and weights. Then the optimum precoders and equalizers can be obtained via, e.g., the alternating optimization algorithm 1 in [16].

B. How to decide MCS based on optimization results

1) *Rate Results*: We emulate an overloaded 3-group multi-group multicast system with four transmit antennas, i.e., $N_t = 4$ and the number of users for each group is 3, which yields a total number of $K = 9$ users. Due to the fact that the optimization outcomes, e.g., precoders and MMF rates, are heavily dependent on the random channel matrix, and will affect the selection of modulation coding schemes. Therefore, to best illustrate the superiority of RS, we assume the channel for the three group of users is a static non-fading channel. We also investigate the rate results for benchmarks, i.e., the classic linear precoding [17], denoted as NoRS and solved by both SDR and WMMSE method (by changing the constraint (4b) to $C_m = 0, \forall m$ downgrade the RS system into a NoRS system). The MMF rate for the RS and NoRS after optimization is presented in Fig. 1, in which, a target MMF rate is selected as 4/3 bps/Hz (shown as the purple dot line in Fig. 1). We highlight the following:

- SDR-based NoRS method achieves an upper bound MMF rate, and requires 14dB total SNR to achieve 4/3 bps/Hz, for which the WMMSE-based method requires 16dB. Note that the MMSE method provides the exact precoder matrix, while the SDR method provides the covariance matrix of the precoder matrix.
- RS clearly outperforms the NoRS strategies, and it only requires 10dB SNR to achieve the same MMF rate.

2) *MCS determination*: As we know, Gaussian input based rate optimization is capacity-achieving but rarely realized in practical communication systems, where symbols are taken from finite-size constellations (such as phase-shift keying - PSK, quadrature amplitude modulation - QAM) and coding rate is specified before precoding. Therefore, we map the

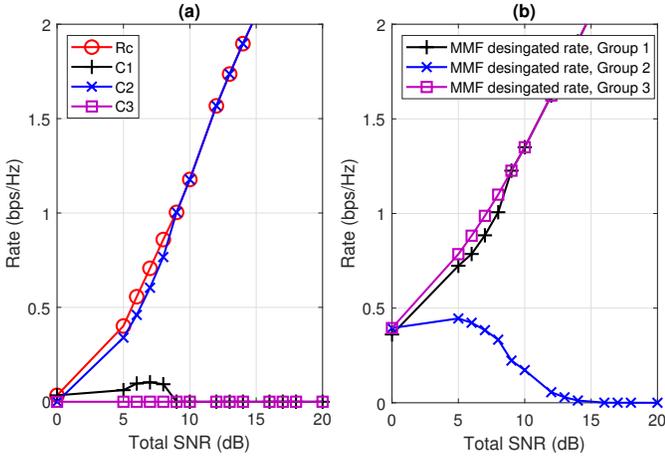


Fig. 2: RS rate contributions. $N_t = 4$ antennas, $M = 3$ groups, $K = 9$ users.

obtained rate results to MCSs for NoRS and RS separately, as follows:

- *NoRS*: The target rate $4/3$ bps/Hz can be achieved by different modulation orders and coding rates. With BPSK, the highest rate it can achieve is 1bps/Hz. Therefore, we use QPSK that provides a maximum rate of 2 bps/Hz. Then any coding scheme with a code rate of $2/3$ can result in an effective coding rate of $4/3$ bps/Hz. Notice that it is possible to use a higher modulation order than QPSK with a lower coding rate, but in this paper we only consider the QPSK case. As a result, for the NoRS scheme, all three streams should be QPSK modulated and coded with a $2/3$ coding rate.
- *RS*: It is different for the RS-based strategy as the MMF rate is comprised by the (part of) degraded rate and the designated rate. In Fig. 2 (a), we present the components i.e., C_1, C_2, C_3 that comprise the degraded rate R_c . Fig. 2 (b) shows the corresponding designated rate of each group. In the table below we summarize the degraded and designated rates of each group to achieve the MMF rate of $4/3$ bps/Hz.

TABLE I: Possible combinations of degraded and designated rates to reach $4/3$ bps/Hz

	Degraded rate	Designated rate
Group 1	0 bps/Hz	1.33 bps/Hz
Group 2	1.21 bps/Hz	0.12 bps/Hz
Group 3	0 bps/Hz	1.33 bps/Hz

It is observed that the first and third group's rates are solely contributed by the designated stream. Therefore, the modulation and coding for the designated stream of Group 1 and 3 is given as: QPSK with a coding rate of $2/3$. For the second group, we can see that over 90% of its group rate is contributed by the degraded stream. Hence, we have two options for the Group 2 MCS setup: (1): The degraded stream is QPSK modulated with a code rate of $3/5$ which results in 1.2bps/Hz rate. Since it is less

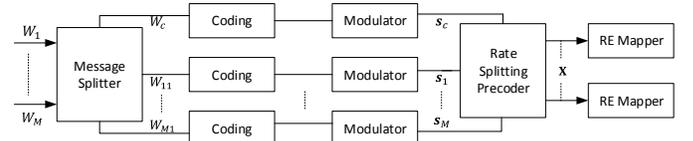


Fig. 3: Adapted transmitter side diagram of the RS-based precoding system

than 1.21bps/Hz, the degraded part can be successfully decoded. The designated stream can be BPSK modulated and use a $1/10$ code rate. (2): A simplified/bias setup, the degraded stream modulated with QPSK and a code rate of $2/3$. No restriction on the modulation and coding rate for the designated stream. This setting reduces the decodability of the degraded stream since the potential transmission rate slightly exceeds the achievable rate.

IV. TRANSCIVER DESIGN

A. RS adapted transmission processing chain

We adapt the transmission processing chain to introduce the RS functionality. Based on the developed link-level simulator equipping the LTE/NR compatible transmission chain [2], [4], we propose a modified transmission chain as shown in Fig. 3. The modifications include:

- Additional block: Message Splitter transforms the original message $\{W_1, \dots, W_M\}$ to the RS-based message $\{W_c, W_1, \dots, W_M\}$.
- RS precoder are placed between the modulator and resource element (RE) mapping.

B. Receiver Design

In the rate optimization, the impact of different receiver techniques is not considered. However, this is crucial when examining the error performance, thus we consider both linear and non-linear receivers.

1) *Linear receiver - MMSE*: The received signal at the k -th user y_k is equalized symbol-by-symbol by the obtained MMSE equalizer $g_{c,k}^{\text{MMSE}}$, then passed through a soft demodulator to estimate the entire degraded stream i.e., $\hat{w}_{c,k}$. The corresponding degraded part $\hat{w}_{g^{(k)}0}$ for the user k can be extracted from $\hat{w}_{c,k}$. For the designated stream, after substituting the estimated degraded stream $\hat{w}_{c,g^{(k)}}$ via SIC, the estimated designated stream $\hat{w}_{g^{(k)}}$ is obtained by applying equalizer g_k^{MMSE} . Finally, the degraded and designated parts are combined to form the estimated message $\hat{w}_{g^{(k)}}$.

2) *Non-linear receiver - GSD*: The procedure of using generalized sphere decoder in the proposed RS framework is the same for decoding the degraded and designated streams. Due to the space limit, here we brief the process of decoding the degraded stream. Recall equation (1) and (2), we have:

$$y_k = \mathbf{h}_k^H \mathbf{P} \mathbf{S} + n_k = \tilde{\mathbf{H}}_k \mathbf{S} + n_k \quad (10)$$

where $\tilde{\mathbf{H}}_k \in \mathbb{C}^{1 \times (M+1)}$ is the effective channel matrix for the k -th user. With non-linear detectors like Maximum Likelihood (ML), for the considered multigroup system, it will decode the entire stream, i.e., \mathbf{S} . From the k -th user's perspective, ML

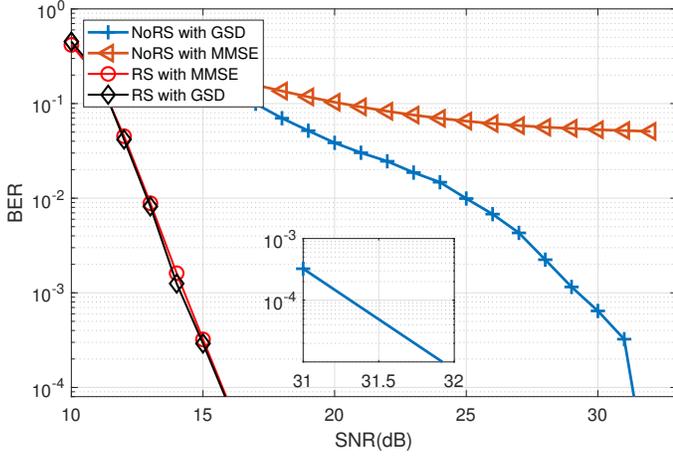


Fig. 4: Overall coded BER of 3-group Multigroup Multicast system with QPSK modulation

detector decodes the transmitted stream in (10) by solving the following equation:

$$\hat{\mathbf{S}}_k = \arg \min_{\mathbf{S}_{con} \in \Omega^{M+1}} \|\mathbf{y}_k - \tilde{\mathbf{H}}_k \mathbf{S}_{con}\|^2 \quad (11)$$

where Ω is the signal constellation set and \mathbf{S}_{con} denotes one possible combination of input symbols that picked from the constellation set. Hence, it is equivalent to solve:

$$\hat{\mathbf{S}}_k = \arg \min_{\mathbf{S}_{con} \in \Omega^{M+1}} \|\mathbf{R}(\tilde{\mathbf{H}}_k^H \mathbf{y}_k - \mathbf{S}_{con})\|^2 \quad (12)$$

where \mathbf{R} is an upper triangular matrix which satisfies $\mathbf{R}^H \mathbf{R} = \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k$. Since each user has one single receive antenna, $\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k$ is positive semidefinite. The Cholesky factor \mathbf{R} of $\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k$ is generally not full rank and only the first row of \mathbf{R} is non-zero. Such rank deficiency makes a normal sphere decoding (SD) not applicable. A generalized SD has been developed to address this issue [15]. Since the product $\mathbf{S}_{con}^H \mathbf{S}_{con}$ results in a constant value, i.e., $M+1$, we get an equivalent minimization problem of (11) as:

$$\hat{\mathbf{S}}_k = \arg \min_{\mathbf{S}_{con} \in \Omega^{M+1}} \|\mathbf{y}_k - \tilde{\mathbf{H}}_k \mathbf{S}_{con}\|^2 + \alpha \mathbf{S}_{con}^H \mathbf{S}_{con} \quad (13)$$

where, unlike (11), we have the term $\mathbf{S}_{con}^H (\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k + \alpha \mathbf{I}_{M+1}) \mathbf{S}_{con}$ after expanding (13). Let the positive definite matrix $\mathbf{G} = \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k + \alpha \mathbf{I}_{M+1}$ be Cholesky factorized as $\mathbf{G} = \mathbf{D}^H \mathbf{D}$, where \mathbf{D} is an upper triangular matrix. Therefore, the optimization problem can be rewritten as

$$\hat{\mathbf{S}}_k = \arg \min_{\mathbf{S}_{con} \in \Omega^{M+1}} \|\mathbf{D}(\beta - \mathbf{S}_{con})\|^2 \quad (14)$$

where $\beta = \mathbf{G}^{-1} \tilde{\mathbf{H}}_k^H \mathbf{y}_k$ and the diagonal elements of \mathbf{D} are all non-zero and the original SD can be applied. As the designed precoder is optimized for delivering the degraded stream, inside all $M+1$ decoded streams, every user only requires the first stream, i.e., the degraded stream. Same applies when decoding the designated stream, where users only extract useful stream.

V. ERROR PERFORMANCE RESULTS

The error performance of the considered RS-based precoding strategy with different receiver configurations is evaluated.

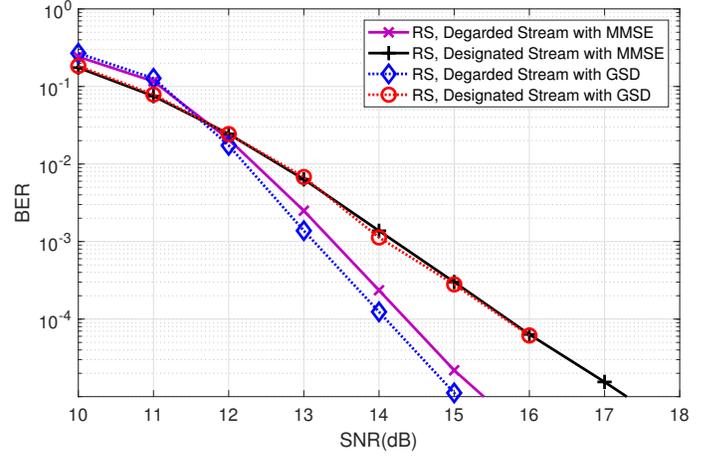


Fig. 5: Separated coded BER of RS scheme with QPSK modulation

Similar to the rate simulation, the classic linear precoding NoRS is also included as the benchmark. Simulation results are averaged over 1000 samples. For the evaluation criterion, we use a BER threshold $> 0.1\%$ [2]. System parameters, e.g., number of transmit antennas, groups and users remains the same as before. At the transmit side, the coding rate and modulation order for different streams of RS and NoRS follow the settings described in Section III-B (the simplified setup is used for RS), and convolutional encoder is used with the code length of 8400. In Fig. 4, the BER vs SNR results for the RS and NoRS systems are presented, including both MMSE and GSD based receivers. It is clear that the error performance of the RS strategy outperforms that of the NoRS strategies. We also observe the following:

- For the NoRS strategies, SNR of 16dB is required to achieve the target rate of 4/3 bps/Hz. However, if the targeted BER is 10^{-4} , the NoRS precoders combined with MMSE receiver are not able to reach this target in a reasonable SNR region, e.g., below 40dB. However, when combined with the non-linear GSD receiver, it requires about 31.3dB SNR to achieve the target BER which is still far from the optimal case, i.e., 16dB.
- For the RS strategy, SNR of 9.85dB is required to achieve the target rate of 4/3 bps/Hz. We can see that the RS system with MMSE or GSD receivers requires nearly the same amount of power to reach the target error performance, be they 15.7dB and 16dB for GSD and MMSE receivers, respectively. It still requires about 7dB more than the optimal case.
- It can be found that without using RS, a high complexity non-linear receiver is required to decode the signal in the overloaded scenario. However, with the aid of RS, linear techniques like MMSE receiver can provide a comparable result but with a largely reduced complexity compared to non-linear techniques.
- The mismatch in the rate and error performance results occurs for multiple reasons, and we name a few here:
 - a. Constellation position: the discrete finite-alphabet in-

puts can lead to performance significantly different from that with Gaussian inputs. **b.** Coding scheme and code length: in the rate optimization, it is assumed to have the optimal coding scheme with infinite code length, which is not feasible in the error performance evaluation (or in practice).

In order to gain more insights of error performance of the RS system with MMSE and GSD receivers, we simulate and show in Fig. 5 the separated BER vs SNR results for the degraded and designated streams of the RS system. It can be observed that the decodability of the designated stream is same for both linear MMSE and non-linear GSD receivers, that reaches 10^{-4} BER level at around SNR of 15.8dB. However, GSD receiver performs slightly better than MMSE when decoding the degraded stream, reaching 10^{-4} BER level at around 14.1dB that is 0.2dB better than that of the MMSE receiver. However, as discussed, the trade-off between complexity and performance enhancement suggests us that in the considered case it is possible for the RS-based strategy working with low-complexity receivers.

VI. CONCLUSIONS

In this paper, we have mainly focused on evaluating the error performance of the RS based transmission strategies with constellation-constrained coding/modulation. First the rate optimization has been reformulated in the overloaded multigroup multicast scenario. The rate results have shown that the RS-based approach exhibits strictly higher and non-saturated MMF-rates compared to the NoRS strategies. Then based on a given achievable rate, we have introduced the approach that helps to determine the modulation order and coding rate according to the optimized rate results, for both NoRS and RS systems. Using the obtained optimum precoder and the matching MCS, we have then investigated the error performance of the RS and NoRS systems, as the first step for the RS software implementation. Both linear and non-linear receivers have been taken into account. From the coded BER performance, we have shown the superior performance of the RS-based strategy over the conventional precoding strategies. The latter enjoys no error floor with the help of the non-linear GSD receiver, but requires a much higher transmit power to reach the optimized rate performance. The former, on the other hand, achieves a comparable error performance by using the linear MMSE receiver or the non-linear GSD receiver, which indicates that it is very likely to realize the promising optimized rate performance of the RS-based strategy, even with a low complexity transceiver.

One can use the proposed analysis, transceiver framework and evaluation methodology in this work as a baseline solution to validate the effectiveness of the RS-based system design or to continue its software implementation in practice.

REFERENCES

[1] "5G-Xcast project website," <http://5g-xcast.eu/>, accessed: 2020-01-30.

- [2] M. Fuentes, D. Mi, H. Chen, E. Garro, J. L. Carcel, D. Vargas, B. Mouhouche, and D. Gomez-Barquero, "Physical Layer Performance Evaluation of LTE-Advanced Pro Broadcast and ATSC 3.0 Systems," *IEEE Transactions on Broadcasting*, vol. 65, no. 3, pp. 477–488, Sep. 2019.
- [3] D. Gomez-Barquero, D. Navratil, S. Appleby, and M. Stagg, "Point-to-multipoint communication enablers for the fifth generation of wireless systems," *IEEE Communications Standards Magazine*, vol. 2, no. 1, pp. 53–59, MARCH 2018.
- [4] H. Chen, D. Mi, M. Fuentes, D. Vargas, E. Garro, J. L. Carcel, B. Mouhouche, P. Xiao, and R. Tafazolli, "Pioneering Studies on LTE eMBMS: Towards 5G Point-to-Multipoint Transmissions," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2018, pp. 1–4.
- [5] Z. Chu, F. Zhou, P. Xiao, Z. Zhu, D. Mi, N. Al-Dhahir, and R. Tafazolli, "Resource Allocation for Secure Wireless Powered Integrated Multicast and Unicast Services With Full Duplex Self-Energy Recycling," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 620–636, Jan. 2019.
- [6] W. Hao, G. Sun, F. Zhou, D. Mi, J. Shi, P. Xiao, and V. C. M. Leung, "Energy-Efficient Hybrid Precoding Design for Integrated Multicast-Unicast Millimeter Wave Communications With SWIPT," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10956–10968, Nov. 2019.
- [7] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-Splitting for Multi-Antenna Non-Orthogonal Unicast and Multicast Transmission: Spectral and Energy Efficiency Analysis," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8754–8770, Dec. 2019.
- [8] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Study on LTE-based 5G terrestrial broadcast (Release 16)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 36.776, March 2019.
- [9] RP-193248, "New Work Item on NR support of Multicast and Broadcast Services," Huawei, 3GPP TSG RAN #86, Dec. 2009.
- [10] H. Joudeh and B. Clerckx, "Rate-Splitting for Max-Min Fair Multigroup Multicast Beamforming in Overloaded Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7276–7289, Nov. 2017.
- [11] H. Chen, D. Mi, B. Clerckx, Z. Chu, J. Shi, and P. Xiao, "Joint power and subcarrier allocation optimization for multigroup multicast systems with rate splitting," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2019.
- [12] H. Chen, D. Mi, Z. Chu, P. Xiao, and R. Tafazolli, "Rate-Splitting for Multigroup Multicast Beamforming in Multicarrier Systems," in *IEEE 19th International Workshop on SPAWC*, June 2018, pp. 1–5.
- [13] Y. Wu, M. Wang, C. Xiao, Z. Ding, and X. Gao, "Linear Precoding for MIMO Broadcast Channels With Finite-Alphabet Constraints," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2906–2920, Aug. 2012.
- [14] A. Salem, C. Masouros, and B. Clerckx, "Rate Splitting with Finite Constellations: The Benefits of Interference Exploitation vs Suppression," 2019. [Online]. Available: <https://arxiv.org/abs/1907.08457>
- [15] T. Cui and C. Tellambura, "An efficient generalized sphere decoder for rank-deficient MIMO systems," *IEEE Communications Letters*, vol. 9, no. 5, pp. 423–425, May 2005.
- [16] H. Joudeh and B. Clerckx, "A Rate-Splitting Strategy for Max-Min Fair Multigroup Multicasting," in *IEEE 17th International Workshop on SPAWC*, July 2016, pp. 1–5.
- [17] D. Mi, M. Dianati, L. Zhang, S. Muhaidat, and R. Tafazolli, "Massive MIMO Performance with Imperfect Channel Reciprocity and Channel Estimation Error," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3734–3749, Sep. 2017.