Exploiting Time Asynchrony in Multi-user Transmit Beamforming

Mehdi Ganji, *Student Member, IEEE*, Xun Zou, *Student Member, IEEE*, and Hamid Jafarkhani, *Fellow, IEEE*

Abstract

In this paper, we analyze the benefits of intentionally adding timing mismatch in the downlink transmit beamforming for wireless transmission. Transmit beamforming enables the so-called space-division multiple access (SDMA), where multiple spatially separated users are served simultaneously. The optimal beamforming vectors can be found to minimize the average transmit power under each user's Quality-of-Service (QoS) constraint. We show that intentionally adding timing offsets between the transmitted signals can significantly reduce the average transmission power compared with the conventional optimal beamforming method while providing the same QoSs for users. We provide three different methods exploiting the time asynchrony which improve the performance with a computational complexity similar to that of the optimal synchronous beamforming. We derive the expressions for the achievable rates using the proposed methods and then provide efficient algorithms to solve the minimum power optimization. We show analytically and numerically that our proposed methods outperform the conventional optimal transmit beamforming.

Index Terms

Beamforming, Asynchronous Transmission, Interference Management, Power Minimization, Over-sampling, Timing Mismatch, Multi-user, SDMA.

I. Introduction

Transmit beamforming is a versatile technique for signal transmission to serve multiple users simultaneously in multiple-antenna systems [1]. With multiple antennas, the idea of transmit

This work was supported in part by the NSF Award CCF-1526780. The authors are with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, 92697 USA (email: {mganji, xzou4, hamidj}@uci.edu).

beamforming is to transmit directional beams to reduce the co-channel interference and thus, enables serving several users using the same resource slot (time/ frequency) which is called spatial division multiple access (SDMA). In contrast to the space-time coding methods [2] which can be designed in time-space domain without the aid of channel state information (CSI), transmit beamforming exploits the CSI to combat the channel fading [3–5]. In theory, dirty paper coding (DPC), a multi-user encoding strategy based on the interference pre-subtraction [6], is the optimal (capacity achieving) strategy in multiple input multiple output (MIMO) downlink channels from the base station to mobile users [7]. However, DPC is difficult to implement in practical systems due to the high computational burden caused by successive encoding and decoding, especially when the number of users is large.

Thus, in the practical scenarios where users have limited computational capabilities and only employ single antennas, the beamforming is proposed as an alternative solution of DPC [8–13]. Beamforming has been shown to achieve a fairly large fraction of DPC capacity with a lower computational complexity when the base station has multiple antennas and each user has a single antenna [14]. Moreover, it has been shown that if the beamforming vectors are chosen optimally, the sum rate of BF approaches that of DPC as the number of users goes to infinity [15, 16]. As an example, in multibeam satellite communications [17] the burden of interference cancellation is carried by the transmission side at the gateway instead of the user terminals equipped with single antennas and limited computational capabilities [18]. Therefore, beamforming techniques with low complexity are more favorable compared with more complex DPC or Superposition/ successive interference cancellation (SIC) methods.

Many different techniques are proposed in the literature for the multiple input single output (MISO) minimum power beamforming problem [19–22] which generally assume symbol-level synchronization. However, in this work, we incorporate the idea of introducing timing offsets at the transmitter side to improve the performance. The idea of intentionally introducing timing offset is investigated in other contexts. It is shown in [23] that with time asynchrony, the achievable rate region can be improved for multiple access channels. By intentionally introducing symbol asynchrony in the transmitted signal, a higher diversity gain can be achieved by zero-forcing detection in spatial multiplexing. [24–26]. The benefits of asynchronism in CDMA systems with random spreading is analyzed in [27] and it is shown that asynchronous transmission can indeed enhance the spectral efficiency. In addition, asynchronous NOMA (ANOMA) systems can achieve a better throughput performance compared with the conventional (synchronous)

NOMA systems [28–31]. Orthogonal differential decoding can be improved by utilizing the oversampling technique [32, 33] to achieve the sampling diversity gain. Asynchronous cognitive radio framework is studied in [34], where the primary user and the secondary user are not aligned in their timing which results in interference reduction and power saving. An asynchronous network coding (ANC) transmission strategy for multiuser cooperative networks is investigated in [35], where the received signals from multiple sources are asynchronous to each other. The proposed scheme achieves full diversity and outperforms the complex field network coding in terms of decoding complexity and bit error rate. To the best of our knowledge, the benefits of asynchronous transmission has never been studied in the downlink transmit beamforming and little is known about the possible performance improvement.

A. Contributions and Organizations

In this paper, we comprehensively investigate the benefits of time asynchrony in downlink beamforming. The primary contributions of the paper are summarized as follows:

- We derive the asynchronous beamforming system model and propose three methods to exploit the asynchronous beamforming. Each proposed method provides different complexity and performance. All of them outperform the conventional synchronous methods as verified by analytical and numerical comparisons.
- We derive the achievable rate expressions for the proposed asynchronous methods with "no time-domain precoding", "uncorrelated time-domain precoding" and "correlated time-domain precoding" which are named as "*Method A*","*Method B*" and "*Method C*", respectively.
- We analyze the effect of the time delays and the pulse shape in the performance of Method *A* and prove its superiority compared with the synchronous method due to the "*reduced InterUser Interference (IUI)*".
- We analyze the effect of the time delays and the pulse shape in the performance of Method *B* and propose a novel algorithm to optimize the beamforming vectors and power distribution functions simultaneously to exploit the "*frequency-selectivity*" imposed by the time asynchrony.
- We explain the oversampling technique in Method *C* which results in a signal space with a higher dimension involving higher rank matrices. We, then, propose a novel algorithm which includes a user scheduling step, a semi-definite relaxation (SDR) step and a power-

control step in order to exploit the "*additional available rank*" provided by asynchronous transmission and oversampling.

- We compare the proposed methods by considering various characteristics, including the complexity and detection delay, and finally present numerical results to compare their performances.

The remainder of the paper is organized as follows. In Section II, the general system model and some introductory notes are presented to set the scene for introducing the proposed methods. In Section III, Method *A*, its corresponding rate expressions and its performance are analyzed. In Section IV, Method *B* and its corresponding rate expressions are presented, and an efficient algorithm to solve the corresponding power minimization is provided. In Section V, Method *C* with user scheduling and oversampling technique is presented. Next, the rate expressions with simplifying sub-channel/power assignments are provided to enable SDR with a simple power-control step. In Sections VI and VII, the numerical results and final remarks are presented, respectively.

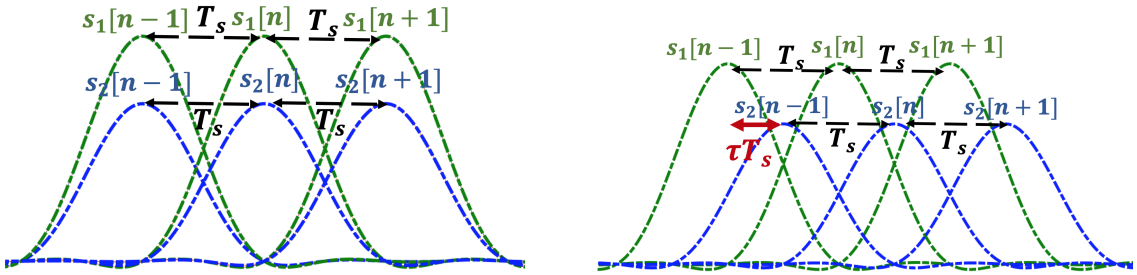
II. General System Model and preliminaries

We consider a downlink wireless communication system consisting of one transmitter equipped with M transmit antennas and K single-antenna users where the environment has limited frequency selectivity, or more specifically is frequency flat. Note that in frequency selective applications, the effectiveness of proposed method which is adding artificial frequency selectivity is not significant. The system operates over a multi-user fading channel and the transmitter sends a block of N complex symbols to each user. Channel coefficients change independently from one block to another and are perfectly known at the base station (BS) [20].

Let $\boldsymbol{d_k} = (d_k[1], \cdots, d_k[N])^T \in \mathbb{C}^{N\times 1}$ denote the random vector of complex constellation symbols, where $d_k[n]$ is the symbol intended for User k at time instant n with average power of $\mathbb{E}[|d_k[n]|^2] = 1$. These constellation symbols are assumed to be i.i.d. random variables selected from an arbitrary constellation. Thus, we have the covariance matrix $\mathbb{E}[\boldsymbol{d_k}\boldsymbol{d_k^H}] = \boldsymbol{I_N}$, where $\boldsymbol{I_N}$ is an $N \times N$ identity matrix. The constellation symbols can be time-precoded properly to yield precoded symbols of $\boldsymbol{s_k} = (s_k[1], \cdots, s_k[N])^T \in \mathbb{C}^{N\times 1}, k = 1, \cdots, K$. Then, the precoded symbols are linearly modulated by a unit-energy real-valued square-root Nyquist pulse as:

$$s_k(t) = \sum_{n=1}^{N} s_k[n]p(t - nT_s), \tag{1}$$

where T_s is the symbol interval and $p(t)$ is the pulse shaping filter which can be assumed to be a rectangular pulse shape (rect.) (a theoretical pulse shape most common in the literature) or a root-raised cosine (r.r.c.) pulse with roll-off factor of β (a common pulse shape in many communication standards like DVB-S2X standard [36] for high throughput satellite systems).



(a) Synchronous transmission

(b) Asynchronous transmission

Fig. 1: Demonstration of synchronous and asynchronous transmission.

To exploit the benefits of time asynchrony, the time delay of $\tau_k T_s$ is intentionally inserted between the transmitted streams where τ_k is the normalized timing offset intended for User k. Thus, the transmitted signal can be denoted as $\boldsymbol{s}(t) = \sum_{k=1}^{K} \boldsymbol{w_k} s_k(t - \tau_k T_s)$, where $\boldsymbol{w_k} \in \mathbb{C}^{M\times 1}$ denotes the beamforming vector applied to the transmitting antenna elements to generate the spatial channel for transmission to User k. Putting $\tau_k = 0, k = 1, \cdots, K$ will result in the synchronous system model. For the asynchronous system model, various works in the literature have shown that the uniform time delays, i.e., $\tau_k = (k-1)/K$, result in optimal performance in different settings [26, 29]. Thus, in this work, we use the uniform time delays although any other choice of time delays is still applicable. The difference between synchronous and asynchronous transmission is demonstrated in Fig. 1. The received signal at User k can be denoted as $y_k(t) = \sum_{l=1}^{K} \boldsymbol{h_k^H w_l} s_l(t-\tau_l T_s) + n_k(t)$ where $\boldsymbol{h_k} \in \mathbb{C}^{M\times 1}$ denotes the channel vector for User k and $n_k(t)$ is the additive white Gaussian noise (AWGN) at the kth user with variance of σ_k^2. Applying the received signal at the kth user to a matched filter with the impulse response $p(t)$ and sampling the filter output at time instants $t_n = nT_s + \tau_k T_s$, $n = 1, 2, \cdots, N$ yields a set of statistics, $y_k[n]$, for detecting the transmitted symbol vectors. Denoting $\boldsymbol{y_k} = (y_k[1], \cdots, y_k[N])^T$, we will have $\boldsymbol{y_k} = \sum_{l=1}^{K} \boldsymbol{G_{kl} h_k^H w_l s_l} + \boldsymbol{n_k}$. The matrix $\boldsymbol{G_{kl}}$, called the "*offset matrix*", is an $N \times N$ Toeplitz matrix whose elements depend on the pulse shape and the corresponding time

delay and are denoted as:

$$[\boldsymbol{G_{kl}}]_{m,n} = g(\tau_{kl}T_s + (m-n)T_s) \triangleq g_{\tau_{kl}}(m-n), \quad m,n = 1,\cdots,N, \tag{2}$$

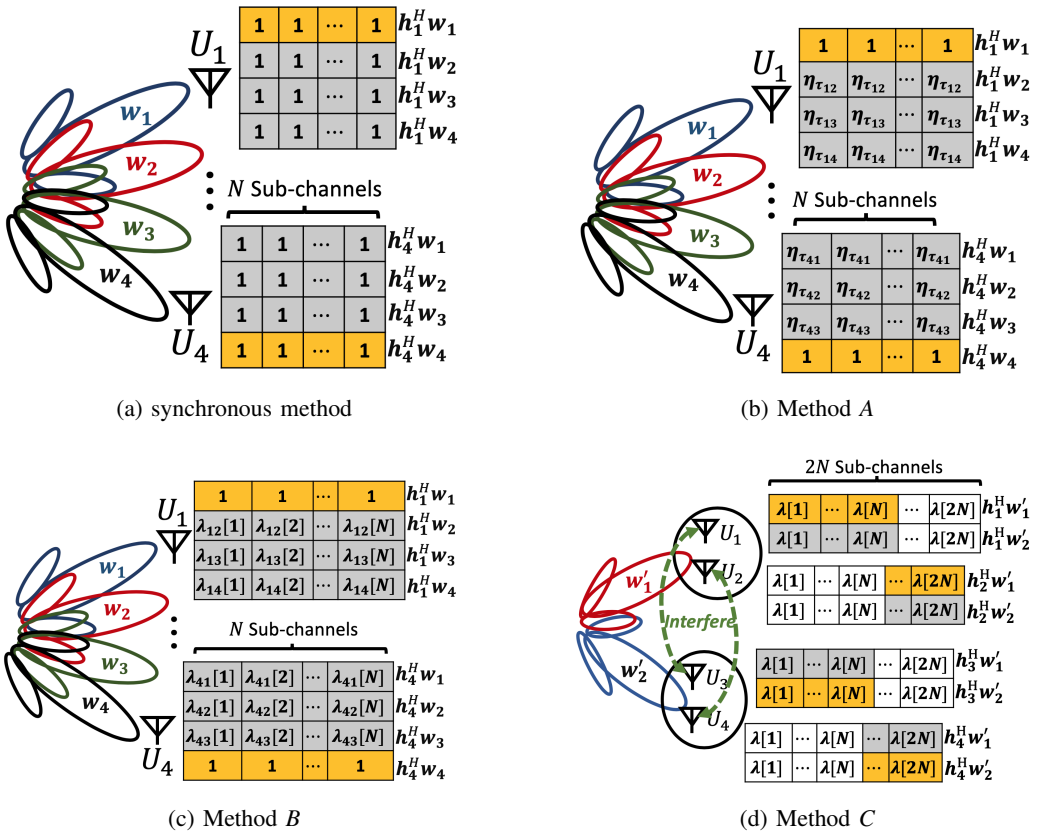
where $\tau_{kl} = \tau_k - \tau_l$ and $g(t) = p(t) * p(t)$. Denoting u as the number of significant (truncated) side-lobes in the Nyquist pulse shape, the offset matrix is a u-banded Toeplitz matrix. The vector $\boldsymbol{n_k}$ represents the noise vector at User k whose covariance matrix is $\boldsymbol{Q_k} = \mathbb{E}[\boldsymbol{n_k}\boldsymbol{n_k}^H] = \sigma_k^2\boldsymbol{I_N}$. For any square-root Nyquist pulse, e.g., r.r.c., $\boldsymbol{G_{kk}} = \boldsymbol{I_N}$ and $\boldsymbol{G_{kl}^T} = \boldsymbol{G_{lk}}$. Also note that for the synchronous transmission, i.e., $\tau_k = 0, \forall k$, all the timing offset matrices become an identity matrix. The introduced timing offsets at the transmit side is known by the BS and can be sent to the users. The channel can introduce additional time delay at the user side which can be estimated and compensated by the synchronization methods. The effect of residual timing synchronization error is analyzed in []; however, in this work, we assume that channel-imposed time delays are sufficiently compensated by the users.

In this paper, we will present three asynchronous beamforming methods as shown in Fig. 2. The details of these methods will be presented in different sections as follows.

III. Spatial Beamforming Design with no Time-Domain Precoding

In this section, we assume that no time-domain precoding is employed, i.e., $\boldsymbol{s_k} = \boldsymbol{d_k}$. The received samples at User k can be written as $y_k[n] = \boldsymbol{h_k^H w_k} d_k[n] + \sum_{\substack{l=1 \\ l\neq k}}^{K} \boldsymbol{h_k^H w_l} \sum_{m=1}^{N} [\boldsymbol{G_{kl}}]_{n,m} d_l[m] + n_k[n], n = 1,\cdots,N$ where $\boldsymbol{G_{kl}}$ becomes an identity matrix for the synchronous transmission. Assuming Gaussian-distributed symbols, the variance of the effective noise, i.e., $\tilde{n}_k[n] = \sum_{\substack{l=1 \\ l\neq k}}^{K} \boldsymbol{h_k^H w_l} \sum_{m=1}^{N} [\boldsymbol{G_{kl}}]_{n,m} d_l[m] + n_k[n]$, can be calculated as $\tilde{\sigma}_k^2[n] = \sum_{\substack{l=1 \\ l\neq k}}^{K} |\boldsymbol{h_k^H w_l}|^2 \sum_{m=1}^{N} [\boldsymbol{G_{kl}}]_{n,m}^2 + \sigma_k^2$. As a result, the achievable rate at User k can be written as $r_k = \lim_{N\to\infty} \frac{1}{N} \sum_n \log_2(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\tilde{\sigma}_k^2[n]})$. Note that r_k provides a lower-bound for the achievable rate of non-Gaussian symbols. Due to the Toeplitz structure of the offset matrices, the expression for the achievable rate can be further simplified. For non-boundary sub-channels, i.e., $n = u,\cdots,N-u$, the variance of effective noise is independent of the sub-channel index, i.e., $\tilde{\sigma}_k^2 = \sum_{\substack{l=1 \\ l\neq k}}^{K} |\boldsymbol{h_k^H w_l}|^2 \sum_{n=-u}^{u} g_{\tau_{kl}}^2(n) + \sigma_k^2$. Therefore, as $N \to \infty$, the effect of boundary sub-channels vanishes and the achievable rate can be written as:

$$r_k^A = \log_2\left(1 + \frac{|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1, l\neq k}^{K} \eta_{\tau_{kl}} |\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right), \tag{3}$$



(a) synchronous method

(b) Method *A*

(c) Method *B*

(d) Method *C*

Fig. 2: Simple illustration of the proposed methods: yellow, grey and white colors indicate the desired signal, the IUI and no-interference, respectively.

where r_k^A denotes the achievable rate by asynchronous transmission and symbol-by-symbol detection at User k, $\eta_{\tau_{kl}} = \sum_{n=-u}^{u} g_{\tau_{kl}}^2(n)$ represents the "*offset coefficient*", and u is the number of significant (truncated) side-lobes in the Nyquist pulse shape. Note that, although $N \to \infty$ is required for rigorous derivation of the achievable rate, r_k^A approximates the achievable rate very well even for a moderate block length.

The offset coefficients $\eta_{\tau_{kl}}$ are the key factors to analyze the achievable rates of Method *A*. With synchronous transmission, i.e., $\tau_k = 0, \forall k$, the offset coefficients are equal to one and as a result, the achievable rates simplify to the conventional synchronous rate expressions $r_k^{synch} = \log_2\left(1 + \frac{|\boldsymbol{h}_k^H \boldsymbol{w}_k|^2}{\sum_{l=1,l\neq k}^{K} |\boldsymbol{h}_k^H \boldsymbol{w}_l|^2 + \sigma_k^2}\right)$ [20]. In fact, in the next lemma, we show that for any other values of timing offsets $0 < \tau_k < 1$, the offset coefficients are less than one which results in the reduction of the IUI and hence, increase of the achievable rate.

Lemma 1: For any Nyquist filter denoted as $g(t)$, the offset coefficient defined as $\eta_\tau =$

$\sum_{n=-u}^{u} g_\tau^2(n)$, for a given τ, has the following property:

$$\eta_\tau \leq \eta_0 \quad \forall \tau \in (0,1). \tag{4}$$

In addition, η_τ can be calculated for the rectangular pulse and root raised cosine pulse with roll-off factor of β as:

$$\eta_\tau^{rect.} = \tau^2 + (1-\tau)^2, \quad \eta_\tau^{r.r.c.} \approx 1 - \beta/4 + \beta\cos(2\pi\tau)/4, \tag{5}$$

respectively.

Proof: The proof is presented in Appendix A. ■

Note that $\beta = 0$ results in $\eta_\tau^{r.r.c.} = 1$ for all values of τ. In other words, timing offset has no effect on interference when $\beta = 0$. In addition, $\tau = 0$ results in the maximum of $\eta_\tau = 1$ for all Nyquist pulse shapes and $\tau = 1/2$ results in the minimum of $1/2$ and $1 - \beta/2$ for the rect. and the r.r.c. pulse shapes, respectively. In Fig. 3, the behaviour of the offset coefficient with respect to the normalized timing offsets is shown for a rect. pulse and truncated r.r.c. pulses with the roll-off factor $\beta = 0.1, 0.5, 1$. Note that $\tau = 0.5$ results in the smallest offset coefficient.

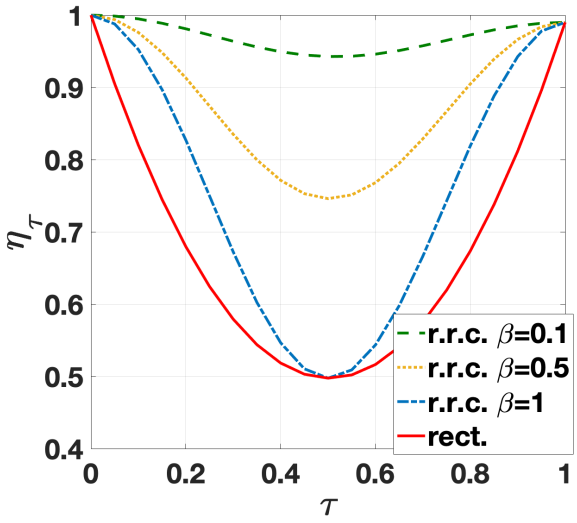


Fig. 3: Values of η_τ for different timing offsets.

The average transmit power equals $\sum_{k=1}^{K} ||\boldsymbol{w_k}||^2$, thus, the power minimization problem can be expressed as:

$$\min_{\{\boldsymbol{w_k}\}_{k=1}^{K}} p_{avg} = \sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \tag{6}$$

$$s.t.\ \frac{|\boldsymbol{h_k}^H \boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^{K} \eta_{\tau_{kl}} |\boldsymbol{h_k}^H \boldsymbol{w_l}|^2 + \sigma_k^2} \geq \gamma_k^*, \quad k = 1, \cdots, K.$$

The parameters $\gamma_k^* = 2^{r_k^*} - 1$ describe the signal to interference plus noise ratios (SINRs) required by each user. There are different methods proposed in the literature to solve this optimization problem for a feasible set of SINR requirements [22]. For example, using the uplink-downlink duality and modifying the optimal algorithm presented in [8, 22], we can solve the optimization problem. For completeness, the highlight of the algorithm is presented in Appendix B. It can be easily shown that the asynchronous method outperforms the synchronous method because each user suffers from less inter-user interference in the asynchronous systems.

Proposition 1: The average transmission power obtained from the asynchronous transmission in (6) is less than that of the synchronous method, i.e., $p_{avg,opt}^{A} < p_{avg,opt}^{synch}$.

Proof: The proof is trivial as the minimization problem for both the synchronous and asynchronous methods include the same objective function while having looser constraints for the latter due to less interuser interference. As a result, the average transmit power is reduced by using Method *A*. ■

IV. Spatial Beamforming design with uncorrelated time-domain precoding

In this section, we present Method *B* in Fig. 2 that uses uncorrelated time-domain precoding. Recalling the received samples at User k, $\boldsymbol{y_k} = \sum_{l=1}^{K} \boldsymbol{G_{kl}} \boldsymbol{h_k}^H \boldsymbol{w_l} \boldsymbol{s_l} + \boldsymbol{n_k}$, note that the offset matrices are banded Toeplitz matrices. Banded Toeplitz matrices are asymptotically equivalent to circulant matrices as the matrix dimension goes to infinity [37, 38]. The first implication of the asymptotic equivalence of Banded Toeplitz matrices with circulate matrices is that a banded Toeplitz matrix can be diagonalized by DFT matrices as its size grows large. In other words, as $N \to \infty$, matrix $\boldsymbol{G_{kl}}$ can be denoted as $\boldsymbol{U_N} \boldsymbol{\Lambda_{kl}} \boldsymbol{U_N}^H$ where $\boldsymbol{U_N}$ denotes the $N \times N$ DFT matrix and $\boldsymbol{\Lambda_{kl}}$ is a diagonal matrix whose nth diagonal element is denoted as $[\boldsymbol{\Lambda_{kl}}]_{nn} = \lambda_{kl}[n]$. Diagonal structure of $\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{kl}} = \boldsymbol{U}_{\boldsymbol{N}}^H \boldsymbol{G_{kl}} \boldsymbol{U_N}$ is also verified in Fig. 4 for a moderate number of block length. The off-diagonal residual, defined as sum of squared of off-diagonal elements is shown for the r.r.c. pulse and various values of N. It can be seen that, as N increases, the off diagonal elements converge to zero and, thus, the offset matrices can be diagonalized by DFT matrix.

Therefore, to diagonalize the offset matrices, each data stream can be precoded as, $\boldsymbol{s_k} = \boldsymbol{U_N} \boldsymbol{P_k}^{1/2} \boldsymbol{d_k}$ where $\boldsymbol{U_N}$ is the $N \times N$ DFT matrix and $\boldsymbol{P_k}$ is a diagonal matrix whose nth

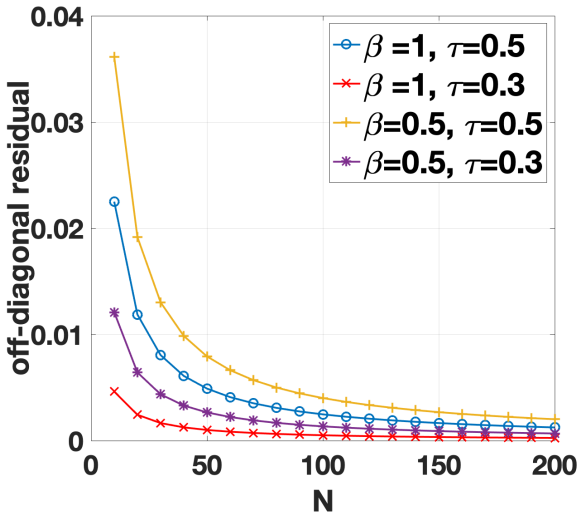


Fig. 4: Verification of diagonal structure of $\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{kl}} = \boldsymbol{U}_{\boldsymbol{N}}^{H}\boldsymbol{G}_{\boldsymbol{kl}}\boldsymbol{U}_{\boldsymbol{N}}$ with r.r.c. pulse.

diagonal element is the power coefficient of User k's nth sub-channel denoted by $P_k[n]$. At User k, the received samples are multiplied by $\boldsymbol{U}_{\boldsymbol{N}}^{H}$ to get:

$$\hat{\boldsymbol{y}}_{\boldsymbol{k}} = \boldsymbol{U}_{\boldsymbol{N}}^{H}\boldsymbol{y}_{\boldsymbol{k}} = \boldsymbol{h}_{\boldsymbol{k}}^{H}\boldsymbol{w}_{\boldsymbol{k}}\boldsymbol{P}_{\boldsymbol{k}}^{1/2}\boldsymbol{d}_{\boldsymbol{k}} + \sum_{\substack{l=1\\ l\neq k}}^{K}\boldsymbol{\Lambda}_{\boldsymbol{kl}}\boldsymbol{h}_{\boldsymbol{k}}^{H}\boldsymbol{w}_{\boldsymbol{l}}\boldsymbol{P}_{\boldsymbol{l}}^{1/2}\boldsymbol{d}_{\boldsymbol{l}} + \hat{\boldsymbol{n}}_{\boldsymbol{k}}, \tag{7}$$

where the covariance matrix of $\hat{\boldsymbol{n}}_{\boldsymbol{k}}$ is equal to $\sigma_k^2\boldsymbol{I}_{\boldsymbol{N}}$. Then, assuming Gaussian signaling, the achievable rate at User k can be written as:

$$r_k^B = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\log_2\left(1+\frac{P_k[n]|\boldsymbol{h}_{\boldsymbol{k}}^{H}\boldsymbol{w}_{\boldsymbol{k}}|^2}{\sum_{l=1,l\neq k}^{K}|\lambda_{kl}[n]|^2P_l[n]|\boldsymbol{h}_{\boldsymbol{k}}^{H}\boldsymbol{w}_{\boldsymbol{l}}|^2+\sigma_k^2}\right). \tag{8}$$

The next implication of the asymptotic equivalence of banded Toeplitz matrices with circular matrices is that the diagonal elements of $\boldsymbol{\Lambda}_{\boldsymbol{kl}}$ are samples of the generating function of matrix $\boldsymbol{G}_{\boldsymbol{kl}}$. In more details, defining the generating function of $\boldsymbol{G}_{\boldsymbol{kl}}$ as $G_{\tau_{kl}}(f) = \sum_{n=-\infty}^{\infty} g_{\tau_{kl}}(n)e^{-j2\pi fn}$, $f \in [0,1]$, we have $\lambda_{kl}[n] = G_{\tau_{kl}}(n/N), n = 1,\cdots,N$ [39]. For example, for the offset matrix $\boldsymbol{G}_{\boldsymbol{kl}}$ with r.r.c. pulse and $N = 100$, the absolute values of $\lambda_{kl}[n]$ and the function $|G_{\tau_{kl}}(f)|$ (denoted as $\lambda[n]$ and $G_\tau(f)$ in the legend for better presentation) are shown in Fig. 5.

Defining $f_n = n/N$, $df_N = 1/N$ and $P_k(f_n) = P_k[n]$, we can rewrite the achievable rate as $r_k^B = \lim_{N\to\infty}\sum_{n=1}^{N} C(f_n)df_N$ where $C(f_n) = \log_2\left(1+\frac{P_k(f_n)|\boldsymbol{h}_{\boldsymbol{k}}^{H}\boldsymbol{w}_{\boldsymbol{k}}|^2}{\sum_{l=1,l\neq k}^{K}|G_{\tau_{kl}}(f_n)|^2P_l(f_n)|\boldsymbol{h}_{\boldsymbol{k}}^{H}\boldsymbol{w}_{\boldsymbol{l}}|^2+\sigma_k^2}\right)$. Because $C(f_n)$ is bounded and almost everywhere continuous on the interval $[0,1]$, then it is Reimann integrable on the interval [40], and we get:

$$r_k^B = \int_0^1 \log_2\left(1+\frac{P_k(f)|\boldsymbol{h}_{\boldsymbol{k}}^{H}\boldsymbol{w}_{\boldsymbol{k}}|^2}{\sum_{l=1,l\neq k}^{K}\lambda_{kl}(f)P_l(f)|\boldsymbol{h}_{\boldsymbol{k}}^{H}\boldsymbol{w}_{\boldsymbol{l}}|^2+\sigma_k^2}\right)df, \tag{9}$$

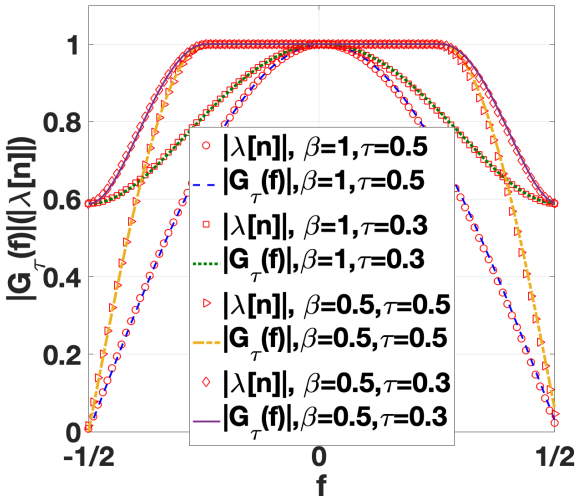


Fig. 5: Comparison of $|G_{\tau_{kl}}(f)|$ and $|\lambda_{kl}[n]|$ with r.r.c. pulse and $N = 100$.

where $\lambda_{kl}(f) = |G_{\tau_{kl}}(f)|^2$, called the "*offset function*", depends on the pulse shape and the corresponding time delay τ_{kl}. The power distribution function of User k is denoted as $P_k(f), f \in [0, 1]$. A similar approach is also used in [41] to show the capacity region of broadcast channels with inter-symbol interference (ISI) and colored Gaussian noise.

The offset function $\lambda_{kl}(f)$ is the deciding factor on the achievable rates of Method *B*. As shown in Appendix A, $G_{\tau_{kl}}(f)$ is indeed the sum of the shifted versions of the frequency spectrum of the pulse shape, while each shifted version is multiplied by an exponent of the corresponding time delay (refer to Appendix A, Eq. (22)). Therefore, for the synchronous beamforming and Nyquist pulse shape, the folded spectrum of the pulse shape adds up to one, which then results in $\lambda_{kl}(f) = 1$. Thus, the achievable rate expression simplifies to the conventional rate expression. However, exploiting time asynchrony will alter the function $\lambda_{kl}(f)$ which is demonstrated in Fig. 6 for r.r.c. pulse shape and different time delays. It can be interpreted that, the addition of time delays effectively transforms the flat fading channel into a "*frequency selective channel*" which can be exploited by proper power spectrum management.

The average transmission power can be calculated as $\sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \int_0^1 P_k(f)df$, hence, the power minimization can be written as:

$$\min_{\{P_k(f),\boldsymbol{w_k}\}_{k=1}^{K}} p_{avg} = \sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \int_0^1 P_k(f)df \tag{10}$$

$$s.t. \quad \int_0^1 \log_2\left(1 + \frac{P_k(f)|\boldsymbol{h_k^H w_k}|^2}{\sum_{l=1,l\neq k}^{K} \lambda_{kl}(f)P_l(f)|\boldsymbol{h_k^H w_l}|^2 + \sigma_k^2}\right)df \geq r_k^*, \quad k = 1, \cdots, K.$$

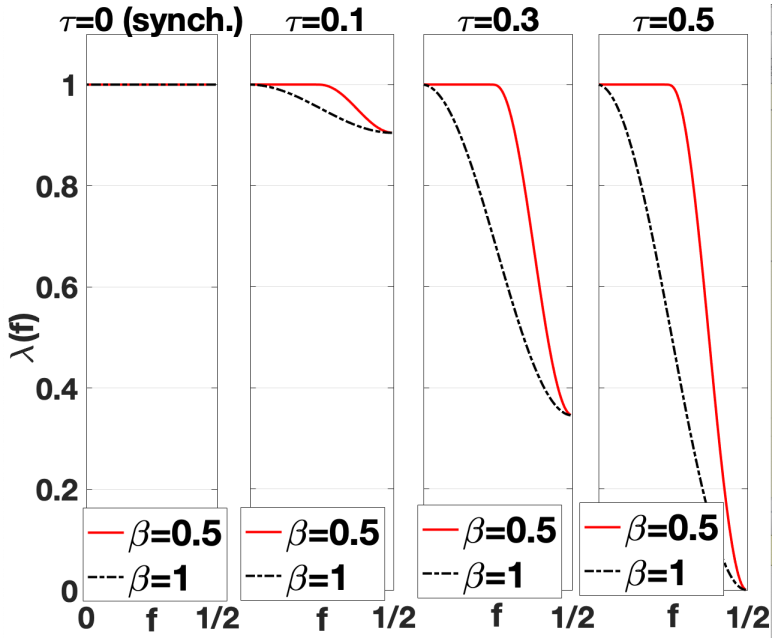


Fig. 6: Demonstration of $\lambda_{kl}(f)$ for r.r.c. pulse with $\beta = 0.5, 1$ and different values of timing offsets.

Proposition 2: The optimal transmission power obtained from the power minimization in (10) is less than Method *A*, i.e., $p^{B}_{avg,opt} < p^{A}_{avg,opt}$.

Proof: The proof is presented in Appendix C. ■

Next, a two-step sub-optimal algorithm is proposed to find the beamforming vectors and power distribution functions. In the first step, the beamforming vectors are obtained by assuming uniform power distribution functions which reduces the optimization problem to

$$\min_{\{\boldsymbol{w_k}\}_{k=1}^{K}} p_{avg} = \sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \tag{11}$$
$$s.t. \quad \int_0^1 \log_2\left(1 + \frac{|\boldsymbol{h_k}^H \boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^{K} \lambda_{kl}(f)|\boldsymbol{h_k}^H \boldsymbol{w_l}|^2 + \sigma_k^2}\right) df \geq r_k^*, \quad k = 1, \cdots, K.$$

To efficiently find the beamforming vectors under the rate constraints, the beamforming vectors can be found using the lower bounds of the actual rate expressions (refer to Appendix C for more details). The optimization problem can be rewritten as

$$\min_{\{\boldsymbol{w_k}\}_{k=1}^{K}} p_{avg} = \sum_{k=1}^{K} ||\boldsymbol{w_k}||^2 \tag{12}$$
$$s.t. \quad log_2\left(1 + \frac{|\boldsymbol{h_k}^H \boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^{K} \int_0^1 \lambda_{kl}(f) df |\boldsymbol{h_k}^H \boldsymbol{w_l}|^2 + \sigma_k^2}\right) \geq r_k^*, \quad k = 1, \cdots, K.$$

Since the rate constraints in (12) are tighter versions compared with those in (11), the solution of (12) is also a suboptimal solution for (11). The optimization problem in (12) is equivalent to the optimization problem in (6) because $\int_0^1 \lambda_{kl}(f)df = \eta_{\tau_{kl}}$. Hence, the optimal algorithm presented in Appendix B can be used to find the optimal beamforming vectors for the above problem as well. However, the average transmit power can be further reduced by proper spectrum management as discussed in the next step.

In the second step, the power distribution functions are found. Power spectrum management or Dynamic Spectrum Management (DSM) is a well-known and an effective method for reducing the effect of crosstalk in Digital Subscriber Line (DSL) systems [42]. Various DSM algorithms are proposed in the literature including the Optimal Spectrum Balancing (OSB), Iterative Spectrum Balancing (ISB), Iterative Water-Filling (IWF) and Successive Convex Approximation for Low complExity (SCALE) [43]. The SCALE optimization has shown the lowest complexity with fastest convergence properties and closest solutions to the optimal one [44]. Let denote $||\boldsymbol{w_k^*}||^2$ and $\lambda_{kl}(f)|\boldsymbol{h_k^H w_l^*}|^2$ as ρ_k and $\hat{\lambda}_{kl}(f)$, respectively, where $\boldsymbol{w_k^*}$ is the beamforming vectors obtained from the first step. Then, the DSM problem can be formulated as:

$$\min_{\{P_k(f)\}_{k=1}^K} p_{avg} = \sum_{k=1}^{K} \rho_k \int_0^1 P_k(f)df \tag{13}$$

$$s.t. \quad \int_0^1 \log_2\left(1 + \frac{P_k(f)\hat{\lambda}_{kk}(f)}{\sum_{l=1,l\neq k}^{K} \hat{\lambda}_{kl}(f)P_l(f) + \sigma_k^2}\right)df \geq r_k^*, \quad k = 1, \cdots, K,$$

which can be solved efficiently with SCALE algorithm [44]. For example, the power distribution functions for the case of $M = 4$, $K = 4$ using r.r.c. pulse shape with $\beta = 0.5$ and a random channel realization is shown in Fig. 7. It can be observed that the frequency selectivity that is imposed by intentional time delays is exploited by the proposed algorithm. As a result, further reduction in the average transmit power can be achieved as shown in the legend of Fig. 7 and also in the simulation section.

V. Spatial Beamforming design with correlated time-domain precoding

In this section, we present Method *C* that exploits Oversampling and timing asynchrony. Oversampling technique provides independent set of sufficient statistics for the asynchronous transmission which can be exploited to cancel the IUI interference [23].

To obtain the set of sufficient statistics, each user samples the received signal K times more than the previous methods, i.e., at time instants $t_n^j = nT_s + \tau_j T_s$, for $n = 1, 2, \cdots, N, j =$

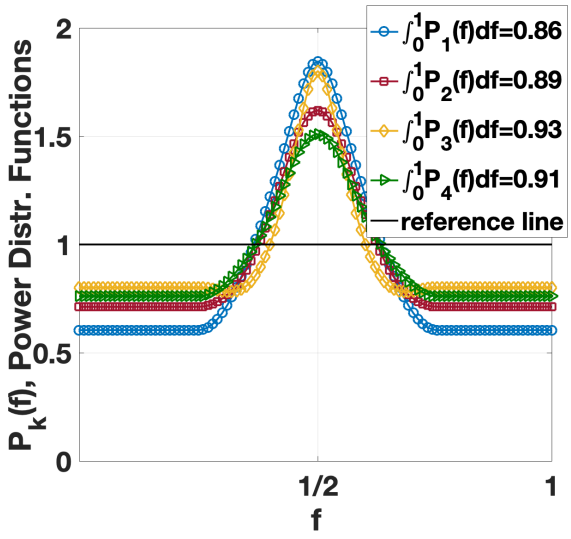
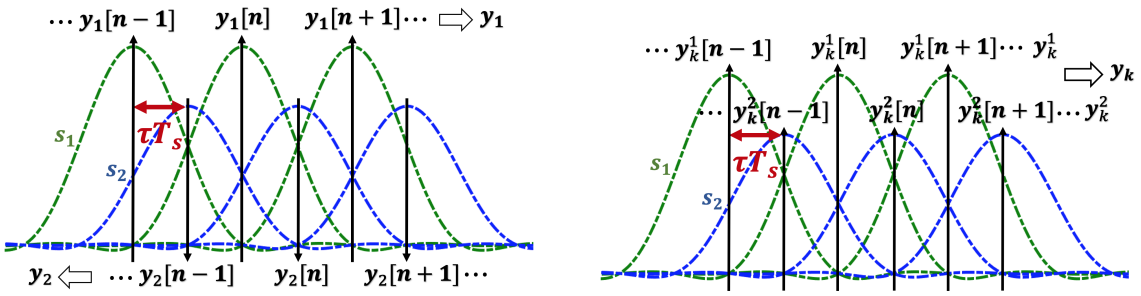


Fig. 7: Demonstration of power distribution functions obtained by the proposed two-step algorithm for r.r.c. pulse with $\beta = 0.5$.



(a) Sampling at generic users 1 and 2

(b) Oversampling at User k

Fig. 8: Demonstration of oversampling method.

$1, \cdots, K$ which yields a set of sufficient statistics $y_k^j[n]$ for detecting the transmitted symbol vectors [23]. The oversampling method is demonstrated in Fig. 8. Denoting $\boldsymbol{y_k^j} = (y_k^j[n], \cdots, y_k^j[N])^T$, we will have $\boldsymbol{y_k^j} = \sum_{l=1}^K \boldsymbol{h_k^H w_l G_{jl} s_l} + \boldsymbol{n_k^j}$ where $\boldsymbol{n_k^j}$ represents the noise vector of the jth set of samples at User k whose covariance matrix is $\boldsymbol{Q_k^j} = \mathbb{E}[\boldsymbol{n_k^j n_k^{j}}^H] = \sigma_k^2 \boldsymbol{I_N}$. Putting all the samples together as $\boldsymbol{y_k} = (\boldsymbol{y_k^1}^T, \cdots, \boldsymbol{y_k^K}^T)^T$ results in:

$$\boldsymbol{y_k} = \begin{pmatrix} \boldsymbol{G_{11}} & \boldsymbol{G_{12}} & \cdots & \boldsymbol{G_{1K}} \\ \boldsymbol{G_{12}^T} & \boldsymbol{G_{22}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{G_{K-1,K}} \\ \boldsymbol{G_{1K}^T} & \cdots & \boldsymbol{G_{K-1,K}^T} & \boldsymbol{G_{KK}} \end{pmatrix} \begin{pmatrix} \boldsymbol{h_k^H w_1 I_N} & \boldsymbol{0_N} & \cdots & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \boldsymbol{h_k^H w_2 I_N} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \cdots & \boldsymbol{0_N} & \boldsymbol{h_k^H w_K I_N} \end{pmatrix} \begin{pmatrix} \boldsymbol{s_1} \\ \boldsymbol{s_2} \\ \vdots \\ \boldsymbol{s_K} \end{pmatrix} + \begin{pmatrix} \boldsymbol{n_k^1} \\ \boldsymbol{n_k^2} \\ \vdots \\ \boldsymbol{n_k^K} \end{pmatrix}$$

$$= \boldsymbol{G H_k s} + \boldsymbol{n_k}, \tag{14}$$

where $\mathbf{0_N}$ represents the $N \times N$ all-zero matrix. Matrix $\boldsymbol{G}$ contains the offset matrices as its constructive blocks. Matrix $\boldsymbol{H_k}$ represents the effective channel coefficients at User k which depends on the actual channel and the choice of beamforming vectors at the antenna elements. The vectors $\boldsymbol{s}$ and $\boldsymbol{n_k}$ denote the precoded symbols and the noise vector at User k whose covariance matrix equals $\boldsymbol{G}\sigma_k^2$ due to oversampling. It is shown that Matrix $\boldsymbol{G}$ with distinctive time delays is positive definite for any *time-limited* pulse shape (which encompasses all the pulse shapes in practice) [45]. Therefore, unlike the synchronous transmission where matrix $\boldsymbol{G}$ becomes rank-deficient, in the asynchronous transmission with distinctive time delays, matrix $\boldsymbol{G}$ is full-rank [46]. Because matrix $\boldsymbol{G}$ is full-rank, NK sub-channels are available to be exploited. In addition, matrix $\boldsymbol{G}$ is a block Toeplitz matrix which with uniform time delays and proper ordering in the received samples can become banded Toeplitz [47] and, thus can be diagonalized by DFT matrix, as explained in Section IV.

In order to use the "*additional available rank*", and be able to cancel the co-channel interference, the users should be divided into groups with a common beamforming vector assigned to each group. In this way, the effective channel matrix will be the same for all the users in a group and the offset matrix can be diagonalized by DFT matrix to remove the co-channel interference. The information symbols of users within each group is precoded together and transmitted by the same beamforming vector. Therefore, the concept of multicast multi-group beamforming is utilized as shown in Fig. 9, where spatial-domain beamfroming is used to avoid inter-group interference and time-domain precoding is used to avoid intra-group interference.

Assume K/q groups, $\{\mathcal{G}_1, \cdots, \mathcal{G}_{K/q}\}$, where each of them includes q users, i.e., $|\mathcal{G}_g| = q,\ g = 1, \cdots, K/q$ and define a user-grouping function that assigns each user to a group, i.e., $\pi : \mathcal{K} \rightarrow \mathcal{G}$, where $\mathcal{K} = \{1, \cdots, K\}$ and $\mathcal{G} = \{1, \cdots, K/q\}$ are the set of user and group indices, respectively. Each user, k, is assigned to a group, g, $\pi(k) = g$, and the user-grouping policy will be discussed later. Unlike Method *B* with uncorrelated precoding, where each user's data is precoded separately, here, the intended symbols of all users in a group are precoded together. The intended symbols for the users in Group g are precoded as $\left(\boldsymbol{s}^T_{\boldsymbol{q(g-1)+1}}, \cdots, \boldsymbol{s}^T_{\boldsymbol{qg}}\right)^T = \boldsymbol{U_{Nq}}\boldsymbol{P'_g}^{1/2}\left(\boldsymbol{d}^T_{\boldsymbol{q(g-1)+1}}, \cdots, \boldsymbol{d}^T_{\boldsymbol{qg}}\right)^T$ where $\boldsymbol{U_{Nq}}$ and $\boldsymbol{P'_g}$ are the $Nq \times Nq$ DFT matrix and $Nq \times Nq$ diagonal power allocation matrix, respectively. After time-precoding, power allocation and pulse shaping, the signal for Group g can be written as: $S_g(t) = \sum_{i=1}^{q}\sum_{n=1}^{N} s_{q(g-1)+i}[n]p(t - nT_s - \tau_i T_s)$. Then, the signal for each group is beamformed by $\boldsymbol{w'_g} \in \mathbb{C}^{M\times 1}$, and hence, the transmitted signal can be written as $\boldsymbol{S}(t) = \sum_{g=1}^{K/q} \boldsymbol{w'_g} S_g(t)$.

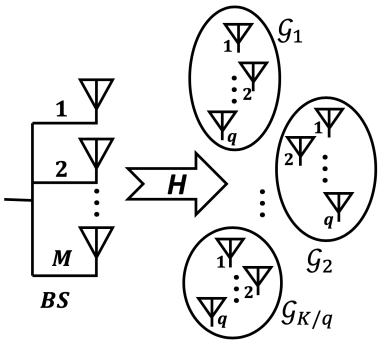


Fig. 9: Method C's System Model.

The received signal at the kth user can be denoted as $y_k(t) = \sum_{g=1}^{K/q} \boldsymbol{h_k^H w'_g} S_g(t) + n_k(t)$. By employing the oversampling technique, explained before, we can have:

$$\begin{aligned}\boldsymbol{y'_k} &= \begin{pmatrix} \boldsymbol{G_{11}} & \boldsymbol{G_{12}} & \cdots & \boldsymbol{G_{1,q}} \\ \boldsymbol{G_{12}^T} & \boldsymbol{G_{22}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{G_{q-1,q}} \\ \boldsymbol{G_{1,q}^T} & \cdots & \boldsymbol{G_{q-1,q}^T} & \boldsymbol{G_{q,q}} \end{pmatrix} \sum_{g=1}^{K/q} \boldsymbol{h_k^H w'_g U_{Nq} P'_g}^{1/2} \begin{pmatrix} \boldsymbol{d_{q(g-1)+1}} \\ \boldsymbol{d_{q(g-1)+2}} \\ \vdots \\ \boldsymbol{d_{qg}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{n_k^1} \\ \boldsymbol{n_k^2} \\ \vdots \\ \boldsymbol{n_k^q} \end{pmatrix} \\ &= \boldsymbol{G'} \sum_{g=1}^{K/q} \boldsymbol{h_k^H w'_g U_{Nq} P'_g}^{1/2} \boldsymbol{d'_g} + \boldsymbol{n'_k}. \end{aligned} \tag{15}$$

As explained before, the banded Toeplitz matrix $\boldsymbol{G'}$ can be diagonalized by DFT matrix, i.e., $\boldsymbol{G'} = \boldsymbol{U_{Nq}} \boldsymbol{\Lambda'} \boldsymbol{U_{Nq}^H}$, as $N \to \infty$ [38]. After multiplying $\boldsymbol{y'_k}$ with $\boldsymbol{U_{Nq}^H}$, the resulting samples can be written as $\hat{\boldsymbol{y}}'_{\boldsymbol{k}} = \boldsymbol{\Lambda'} \sum_{g=1}^{K/q} \boldsymbol{h_k^T w'_g P'_g}^{1/2} \boldsymbol{d'_g} + \hat{\boldsymbol{n}}'_{\boldsymbol{k}}$ where the covariance matrix of the effective noise vector equals $\boldsymbol{\Lambda'} \sigma_k^2$. Denoting the nth diagonal elements of the $\boldsymbol{\Lambda'}$ and $\boldsymbol{P'_g}$, as $\lambda'[n]$ and $P'_g[n]$, respectively, and assuming Gaussian signaling, the rate at the kth user can be written as $r_k^C = \lim_{N\to\infty} \frac{1}{N} \sum_{n\in\mathcal{I}_k} \log_2 \left(1 + \frac{P'_{\pi(k)}[n]\lambda'[n]|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} P'_j[n]\lambda'[n]|\boldsymbol{h_k^H w'_j}|^2 + \sigma_k^2}\right)$ where $\mathcal{I}_k$ represents the set of sub-channels indices which are assigned to User k. The average transmit power equals $p_{avg} = \lim_{N\to\infty} \frac{1}{N} \sum_{g=1}^{K/q} \mathrm{tr}(\boldsymbol{\Lambda' P'_g}) ||\boldsymbol{w'_g}||^2$, thus, the optimization problem can be formulated as

$$\min_{\{\boldsymbol{w'_g}, \boldsymbol{P'_g}\}_{g=1}^{K/q}, \{\mathcal{I}_k\}_{k=1}^{K}, \pi(.)} p_{avg} = \lim_{N\to\infty} \frac{1}{N} \sum_{g=1}^{K/q} \mathrm{tr}(\boldsymbol{\Lambda' P'_g}) ||\boldsymbol{w'_g}||^2 \tag{16}$$

$$s.t. \quad \lim_{N\to\infty} \frac{1}{N} \sum_{n\in\mathcal{I}_k} \log_2 \left(1 + \frac{P'_{\pi(k)}[n]\lambda'[n]|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} P'_j[n]\lambda'[n]|\boldsymbol{h_k^H w'_j}|^2 + \sigma_k^2}\right) \geq r_k^*, \quad k = 1, \cdots, K.$$

The sub-channel and group assignment polices make the above optimization problem overly complicated and intractable. To solve the optimization problem efficiently, we use a sub-optimal method of sub-channel and power assignment rule to simplify the optimization problem. We

assume the following simplifying assumptions: (I) a simple power assignment rule for sub-channels which is, indeed, optimal for an AWGN channel, and (II) the same q available sub-channel configurations for each group. These assumptions are explained in Appendix D in more details. After considering the above simplifying assumptions, the power minimization problem can be formulated as:

$$\min_{\{\boldsymbol{w'_g}\}_{g=1}^{K/q},\{P_k\}_{k=1}^{K},(\pi(.),\phi(.))} p_{avg} = \sum_{k=1}^{K} P_k ||\boldsymbol{w'_{\pi(k)}}||^2 \tag{17}$$
$$s.t. \quad \frac{P_k |\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} P_{c(k,j)} |\boldsymbol{h_k^H w'_j}|^2 + \sigma_k^2} \geq \gamma_k^*, \quad k = 1, \cdots, K,$$

where $\pi(.), \phi(.)$ are the group assignment and sub-channel assignment functions, respectively, assigning each user to a group index, i.e., ($\pi : \mathcal{K} \rightarrow \mathcal{G} = \{1, \cdots, K/q\}$) and a sub-channel assignment index ($\phi : \mathcal{K} \rightarrow \mathcal{S} = \{1, \cdots, q\}$). The index $c(k, j)$ denotes the index of the user in Group j which has the same sub-channel assignment as the kth user, i.e., $c(k, j) = \{l \in \mathcal{G}_j | \phi(l) = \phi(k)\}$. After simplification, the problem is similar to the well-known multigroup multicast QoS problem in the literature [18, 48–50] and similar approaches can be used to solve the problem.

To this end, we first solve the power optimization problem given a set of group and sub-channel assignments $\pi(.)$ and $\phi(.)$, then, consider a heuristic group/sub-channel assignment policy to further improve the objective function. The well-known semi-definite relaxation [51] technique can be applied by assuming all the power adjustment coefficients being equal to one. Thus, the optimization problem reduces to:

$$\min_{\{\boldsymbol{w'_g}\}_{g:1}^{K/q}} p_{avg} = \sum_{g=1}^{K/q} ||\boldsymbol{w'_g}||^2 \tag{18}$$
$$s.t. \ \frac{|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} |\boldsymbol{h_k^H w'_j}|^2 + \sigma_k^2} \geq \gamma_k^*, k = 1, \cdots, K.$$

The above problem is a quadratically constrained quadratic programming (QCQP) problem with non-convex constraints. By some manipulations, we can have the following standard semi-definite programming (SDP).

$$\min_{\{\boldsymbol{W'_g}\}_{g:1}^{K/q}} p_{avg} = \sum_{g=1}^{K/q} \mathrm{tr}(\boldsymbol{W'_g}) \tag{19}$$
$$s.t. \ \gamma_k^* \sum_{j=1,j\neq\pi(k)}^{K/q} \mathrm{tr}(\boldsymbol{H_k W'_j}) + \gamma_k^* \sigma_k^2 - \mathrm{tr}(\boldsymbol{H_k W'_{\pi(k)}}) \leq 0$$

$$\boldsymbol{W_g'} \geq 0, \ \forall g \in \{1, \cdots, K/q\}, \tag{20}$$

where $\boldsymbol{H_k} = \boldsymbol{h_k h_k^H}$, $\boldsymbol{W_g'} = \boldsymbol{w_g' w_g'}^H$. The non-convex rank-one condition is dropped and thus, the solution can be found by SDP solvers like CVX [52]. However, due to the relaxation, the obtained solution will not, in general, consist of rank-one matrices. Hence, an approximate solution to the original problem can be found using a randomization technique like Gaussian randomization method [53]. Then, similar to the multigroup multicast power control (MMPC) [48] step which converts the candidate beamforming vectors to a candidate solution, we apply the following step to find the power adjustment coefficients.

Let denote $||\boldsymbol{w_j'^*}||^2$ and $|\boldsymbol{h_k^H w_j'^*}|^2$ as ρ_j' and α_{kj}, respectively, where $\hat{\boldsymbol{w}}_{\boldsymbol{j}}'^*$ is a beamforming vector candidate obtained from the Gaussian randomization method. Then, to find the power adjustment coefficients, we solve the following optimization problem:

$$\begin{aligned} &\min_{\{P_k\}_{k=1}^K} \ p_{avg} = \sum_{k=1}^{K} P_k \rho'_{\pi(k)} \\ &s.t. \ \ \frac{P_k \alpha_{k,\pi(k)}}{\sum_{j=1, j\neq\pi(k)}^{K/q} P_{c(k,j)} \alpha_{kj} + \sigma_k^2} \geq \gamma_k^*, \quad k = 1, \cdots, K, \end{aligned} \tag{21}$$

which is a linear program (LP) and can be solved easily by matrix inversion as the inequalities are active at the optimal solution (see Appendix B for more details). After feeding N_{rand} beamforming vector candidates to the power control step, the one with lowest objective function value is chosen as the final solution. In summary, the proposed algorithm includes solving the SDP problem once and solving the LP problem N_{rand} times. The choice of N_{rand} is a trade-off between the extent of sub-optimality of the final solution and the overall complexity of the algorithm [48].

A key step in the proposed method lies in sub-channel/user-grouping. User scheduling and sub-channel assignment is studied in various context in the literature including the Non-Orthogonal Multiple Access (NOMA) and Multicast Mutigroup beamforming [18, 48, 54, 55]. In Multicast transmission, same information is transmitted to all users in a group and the interference is only caused by the users in other groups [56]. However, in NOMA systems, users assigned to the same group also cause interference to each other which is removed by ordered SIC based on the channel strengths [57]. Therefore, because of the ordered SIC, the user-grouping algorithms in NOMA settings are more involved, compared with similar algorithms for multicast multigroup transmission. Method *C* is similar to multicast transmission as the same information is sent to all users in a group with no intra-group interference which is eliminated by the oversampling

technique. The underlying intuition for user-grouping in multicast transmission is that users assigned to the same group should have co-linear (i.e., similar) channels since they need to use the same beamforming vector. On the contrary, interfering users, assigned to other groups, should be orthogonal to minimize the interference [16].

Therefore, inspired by the multigroup multicast nature of Method *C*, we use the following low-complexity user-grouping method, detailed in [18]:

- One user per group is allocated according to the semi-orthogonality criteria originally proposed in [16]. The goal is to allocate non-interfering users in different groups.
- For each of the groups, the most parallel users to the previously selected user are assigned to the same group. As a result, the similarity of the co-group channels is maximized.

Note that the focus of this work is analyzing the benefits of time asynchrony and not the user-grouping algorithm. Thus, any other user grouping algorithm that fits our system model can be employed. The effect of the user-grouping method is analyzed in Section VI.

VI. Comparison and Numerical Results

TABLE I: Comparison of the Proposed Methods

Methods	Complexity at Users	Detection Delay	Uplink	Knowledge of Delays	Oversampling	Algorithm
synch.	STE	SS	yes	own delay	no	optimal
Method *A*	STE	SS	yes	own delay	no	optimal
Method *B*	N-IFFT+STE	FF	yes	own delay	no	sub-optimal
Method *C*	Nq-IFFT+STE	FF	no	all delays in own group	yes	sub-optimal

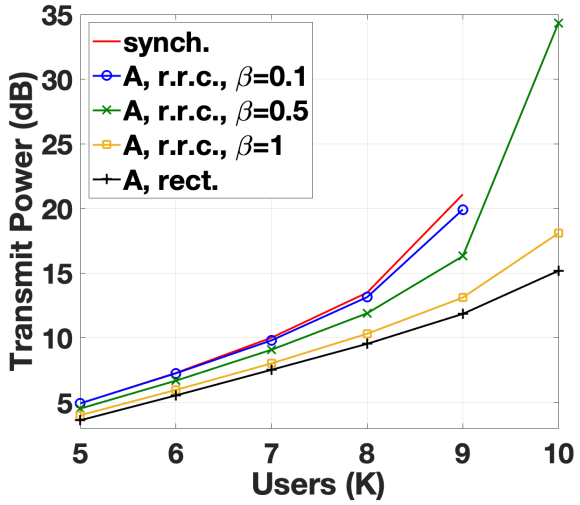
A. Comparison

In Table I, various properties of the proposed methods, including the complexity and delay, are compared. Due to the limited computational capabilities of the users in a beam-forming application, we only consider the complexity at the user side. The synchronous method and Method *A* enjoy symbol-by-symbol (SS) detection with a possible single-tap equalizer (STE) while Methods *B* and *C* perform detection on a frame-by-frame (FF) basis and have additional complexity of $N \times N$ IDFT matrix multiplication and $Nq \times Nq$ IDFT matrix multiplication, respectively. However, the $N \times N$ and $Nq \times Nq$ IDFT matrix multiplications can be effectively perforemd by N-point and Nq-point Inverse Fast Fourier Transform (IFFT). The synchronous

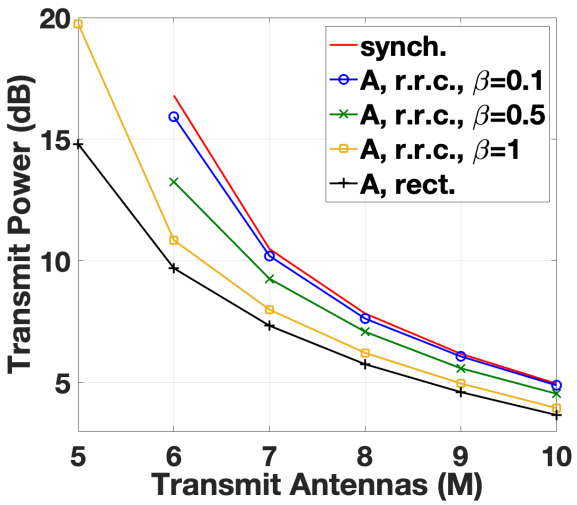
method, Methods *A* and *B* are applicable to uplink with minor modifications, however, Method *C* is only applicable to downlink due to the correlated precoding which requires the collective information of all users in a group. In addition, in Method *C*, the users require the knowledge of the user's time delays in their own group to perform oversampling, however, in other methods, each user only requires its own time delay. The proposed algorithms for the synchronous and Method *A* are optimal while the multi-step algorithms for Methods *B* and *C* are sub-optimal. Providing improved algorithms for Methods *B* and *C* is an interesting topic for future work.

B. Numerical Results

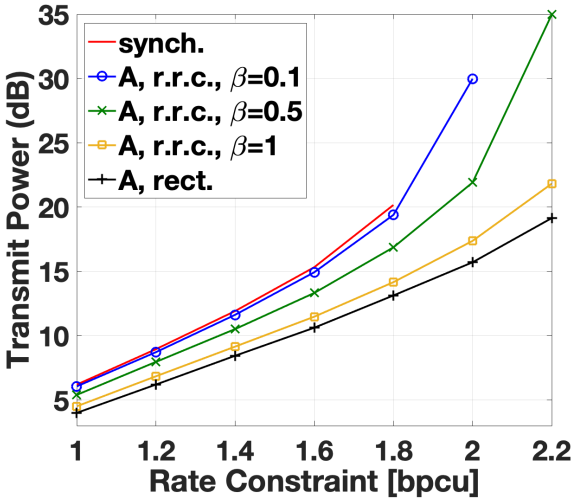
In this section, numerical simulations are performed to verify our presented results. The Monte Carlo simulations are performed over 1000 realizations of random channel coefficients where the channel coefficients follow a Rayliegh fading model $\mathcal{CN}(0,1)$. For each such configuration, the same rate constraint is assumed for all users and the noise variance is set to $\sigma_k^2 = 0.1$ for all users. Different configurations of transmit antennas (M), number of users (K) and rate constraints (r_k^*) are considered to show the advantages of the asynchronous methods. In Fig. 10a, the performance of the synchronous method is compared with Method *A* with $M = 6$ transmit antennas and the rate constraint of $r_k^* = 1.5(bpcu)$ for each user. Different pulse shapes are included to show the effect of pulse shape in the performance of beamforming methods. For fair comparison, the synchronous method is assumed to use the same pulse shape as its counterpart. As explained in Proposition 1, due to asynchrony, the choice of pulse shape has significant effect on the performance improvement of Method *A*. As the roll-off factor of the r.r.c. pulse shape increases, the decrease in the transmit power increases. $\beta = 1$ provides the highest improvement while $\beta \to 0$ provides no reduction in power as proved in Proposition 1. Beside decreasing the average transmit power, Method *A* can also support 10 users which is not possible with the synchronous beamforming. In Fig. 10b, the comparison is performed for $K = 10$ and $r_k^* = 1.2(bpcu)$ with respect to different number of transmit antennas. By increasing the number of transmit antennas, the average transmit power decreases and the performance of both synchronous and asynchronous methods converge. With large number of antennas, the IUI can be completely removed by spatial beamforming and asynchrony loses its benefits. However, as the number of transmit antennas decreases and the system becomes overloaded, the reduction in the average transmit power, obtained by asynchrony, increases. For example, with $M = 6$ and r.r.c. pulse with $\beta = 0.5$, around $3dB$ power reduction is achieved. In addition, using Method A,



(a) $M = 6$, $r_k^* = 1.5(bpcu)$



(b) $K = 10$, $r_k^* = 1.2(bpcu)$



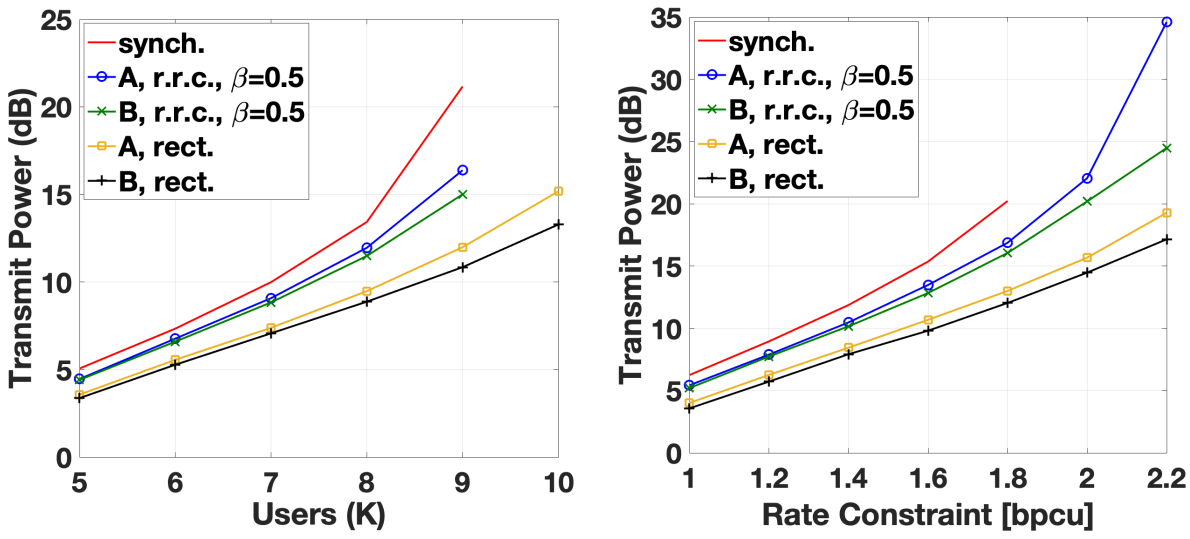
(c) $M = 6$, $K = 8$

Fig. 10: Performance of Method *A*

.

the required rate constraints can be supported by 5 transmit antennas which is impossible with the synchronous beamforming. In Fig. 10c, the comparison is presented for $M = 6$ and $K = 8$ with respect to various rate constraints. For small rate constraints the gain provided by Method *A* is not noticeable, however, as the rate constraints increase the reduction in the average transmit power increases. In addition, the largest rate constraint that can be provided by the synchronous method is $r_k^* = 1.8(bpcu)$ while Method *A* can support up to $r_k^* = 2.2(bpcu)$.

In Fig. 11, the performance of Method *B* using the 2-step algorithm proposed in Section IV

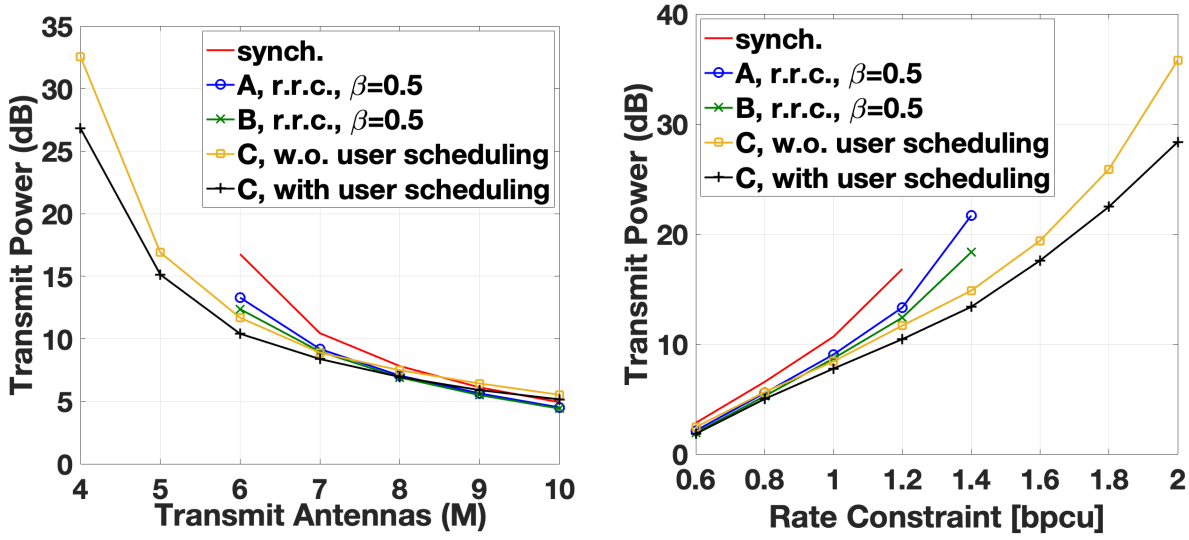


(a) $M = 6$, $r_k^* = 1.5(bpcu)$ (b) $M = 6$, $K = 8$

Fig. 11: Performance of Method *B*.

with SCALE DSM algorithm [44] and $N = 100$ is shown. In Fig. 11a, $M = 6$, $r_k^* = 1.5[bpcu]$ and different number of users (K) are considered. As the number of users increases, the reduction in the average transmit power increases. For example, for $K = 9$, around $1dB$ power reduction is achieved. In Fig. 11b, considering $M = 6$ and $K = 8$, the power reduction of around $10dB$ and $2dB$ is achieved at $r_k^* = 2.2$ compared with Method *A* by r.r.c. ($\beta = 0.5$) and rect. pulses, respectively.

In Fig. 12, the performance of Method C using the proposed algorithm in Section V with $N_{rand} = 300$, $q = 2$, and with/without user scheduling is shown. In Fig. 12a, $K = 10$, $r_k^* = 1.2(bpcu)$ and different number of transmit antennas (M) are considered. As the number of transmit antennas increases, the spatial domain becomes sufficient to cancel the IUI and Method *C* is not helpful. However, with low number of transmit antennas, the spatial domain is unable to effectively cancel the IUI and user-grouping and the precoding/oversampling technique greatly improves the performance. Furthermore, by using Method *C*, the required rate constraints can be satisfied by $M = 4$ transmit antennas. In Fig. 12b, with $M = 6$, $K = 10$, the power reduction of around $6dB$ and $4dB$ are achieved at $r_k^* = 1.4(bpcu)$ compared with Methods *A* and *B*, respectively. In addition, for the chosen parameters, Method *C* can support up to $r_k^* = 2(bpcu)$ while the synchronous method and Methods *A/B* can support up to $r_k^* = 1.2(bpcu)$



(a) $K = 10$, $r_k^* = 1.2(bpcu)$ (b) $M = 6$, $K = 10$

Fig. 12: Performance of Method *C*.

and $r_k^* = 1.4(bpcu)$, respectively. Note that the heuristic user scheduling method further reduces the average transmit power. While we have presented the results for a few sets of parameters, we have done extensive simulations with other choices of parameters and a similar trend has been observed.

VII. Conclusion

In this paper, we investigated the benefits of adding intentional time delays in the performance of downlink transmit beamforming. We compared our proposed methods with the optimal transmit beamforming method and showed that our proposed methods improve the performance by decreasing the average transmit power. Method *A* which uses no time-domain precoding exploits the reduction in the co-channel interference and by using the optimal algorithm (which exploits the duality of uplink and downlink beamforming) can improve the performance. In Method *B*, which uses uncorrelated time-domain precoding, the offset matrices are diagonalized and thus, different sub-channels experience different qualities. The imposed frequency selectivity can be exploited by DSM algorithms, e.g., the SCALE algorithm, to further reduce the average transmit power. Note that the proposed two-step algorithm provides a sub-optimal solution. Method *C* employs correlated time-domain precoding and an oversampling technique to obtain a set of sufficient statistics which results in a system model with higher dimensional full-rank

matrices. The additional available rank helps to cancel the intra-group co-channel interference while the inter-group IUI is reduced by the spatial beamforming. The hueristic group/sub-channel assignment, in conjunction with the SD relaxtaion and the power control step provides a sub-optimal solution. In a nutshell, adding intentional timing offsets provide powerful tools to manage the co-channel interference in downlink beamforming scenarios.

APPENDIX A
PROOF OF LEMMA 1

Assuming a rectangular pulse shape, due to its finite time support, it can be easily shown that $\eta_\tau^{rect.} = \sum_{i=-\infty}^{\infty} g^2(\tau T_s + iT_s) = \tau^2 + (1-\tau)^2$. Therefore, we focus on deriving the results for the r.r.c. pulse shape which has infinite time support. Denoting $g(t)$ as a raised cosine pulse shape with symbol period of T_s and the roll-off factor of β, the goal is to find $\eta_\tau = \sum_{i=-u}^{u} g^2(\tau T_s + iT_s)$. Assuming truncation of the raised cosine pulse shape with large number of side lobes, we can approximate $\eta_\tau \approx \eta_\tau^{\infty} = \sum_{i=-\infty}^{\infty} g^2(\tau T_s + iT_s)$. The frequency spectrum of the discrete sequence of $\{g(\tau T_s + iT_s)\}_{i=-\infty}^{\infty}$ can be denoted as:

$$G_\tau(f) = \frac{1}{T_s} \sum_{i=-\infty}^{\infty} e^{-j2\pi\tau(f+i)} \hat{g}(\frac{f+i}{T_s}), \quad f \in [-1/2, 1/2], \tag{22}$$

where $\hat{g}(f)$ is the Fourier transform of the raised cosine pulse shape $g(t)$ and is denoted as:

$$\hat{g}(f) = \begin{cases} T_s & |f| \leq \frac{1-\beta}{2T_s} \\ \frac{T_s}{2}\left[1 + \cos\left(\frac{\pi T_s}{\beta}\left(|f| - \frac{1-\beta}{2T_s}\right)\right)\right] & \frac{1-\beta}{2T_s} < |f| \leq \frac{1+\beta}{2T_s} \\ 0 & o.w. \end{cases}. \tag{23}$$

The spectrum function $G_\tau(f)$ is periodic with period of 1 and based on the Parseval's theorem, we have $\eta_\tau^{\infty} = \int_{-1/2}^{1/2} |G_\tau(f)|^2 df$. Based on the definition of $\hat{g}(f)$, the spectrum function $G_\tau(f)$ for $f \in [-1/2, 1/2]$ can be derived as:

$$G_\tau(f) = \begin{cases} e^{-j2\pi\tau f} A(-f) + e^{-j2\pi\tau(f+1)} A(f+1) & \frac{-1}{2} < f \leq \frac{-1+\beta}{2} \\ e^{-j2\pi\tau f} & |f| \leq \frac{1-\beta}{2} \\ e^{-j2\pi\tau f} A(f) + e^{-j2\pi\tau(f-1)} A(-f+1) & \frac{1-\beta}{2} < f \leq \frac{1}{2} \end{cases}, \tag{24}$$

where $A(f) = \frac{1}{2}\left[1 + \cos\left(\frac{\pi T_s}{\beta}\left(\frac{f}{T_s} - \frac{1-\beta}{2T_s}\right)\right)\right]$. Thus, η_τ^{∞} can be calculated as follows:

$$\eta_\tau^{\infty} = \underbrace{\int_{\frac{-1}{2}}^{\frac{-1+\beta}{2}} \left|A(-f) + e^{-j2\pi\tau} A(f+1)\right|^2 df}_{b} + \underbrace{\int_{\frac{1-\beta}{2}}^{\frac{1}{2}} \left|A(f) + e^{-j2\pi\tau} A(-f+1)\right|^2 df}_{a} + (1-\beta).$$

In general, for the pulse shapes that satisfy the Nyquist no-ISI condition, the frequency-shifted replicas of the spectrum add up to a constant value, here, e.g., $A(f)+A(-f+1)=1,\ f\in[\frac{1-\beta}{2},\frac{1}{2}]$. Thus, putting $\tau=0$ yields $\eta_0^{\infty}=1$, which is in fact the maximum value of η_τ^{∞}. For non-zero values of timing offset, the phase rotation of the frequency replicas due to timing offset results in out-of-phase addition of replicas and hence, $\eta_\tau^{\infty}<\eta_0^{\infty}, \tau\in(0,1)$.

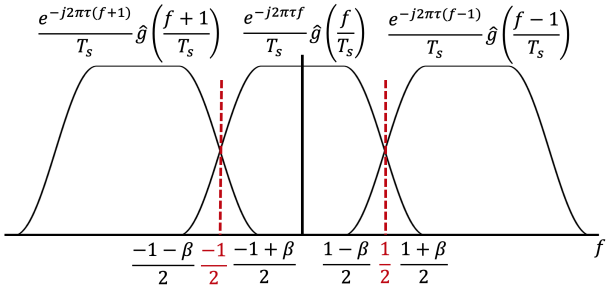


Fig. 13: Schematic illustration of the folded spectrum with phase rotation of the frequency replicas.

To further simplify η_τ^{∞}, note that $a=b$, and we calculate one of them as:

$$a=\int_{\frac{1-\beta}{2}}^{\frac{1}{2}} A^2(f)+A^2(-f+1)+2A(f)A(-f+1)\cos(2\pi\tau)df. \tag{25}$$

Because the function $A(f)$ is always positive, thus, $\tau=0$ maximizes a as explained before, and $\tau=1/2$ minimizes a. By some calculations, we can show that $a=3\beta/8+\beta\cos(2\pi\tau)/8$. As a result, η_τ^{∞} can be calculated as $\eta_\tau^{\infty}=1-\beta/4+\beta\cos(2\pi\tau)/4$ which concludes the proof.

APPENDIX B
SUMMARY OF THE OPTIMAL ALGORITHM FOR DOWNLINK BEAMFORMING

Denoting $\boldsymbol{w_k}=\sqrt{\rho_k}\boldsymbol{u_k}$ where $||\boldsymbol{u_k}||^2=1$, the power optimization problem can be written as:

$$\min_{\{\rho_k,\boldsymbol{u_k}\}_{k=1}^K} p_{avg}=\sum_{k=1}^{K}\rho_k \tag{26}$$
$$s.t.\ \frac{\rho_k\boldsymbol{u_k}^H\boldsymbol{h_k}\boldsymbol{h_k}^H\boldsymbol{u_k}}{\sum_{l=1,l\neq k}^{K}\eta_{\tau_{kl}}\rho_l\boldsymbol{u_l}^H\boldsymbol{h_k}\boldsymbol{h_k}^H\boldsymbol{u_l}+\sigma_k^2}\geq\gamma_k^*,\quad k=1,\cdots,K.$$

By exploiting the virtual uplink duality, the power minimization can be equivalently stated as [8]:

$$\min_{\{\rho_k,\boldsymbol{u_k}\}_{k=1}^K} p_{avg}=\sum_{k=1}^{K}\rho_k \tag{27}$$

$$s.t.\ \frac{\rho_k \boldsymbol{u_k}^H \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H \boldsymbol{u_k}}{\boldsymbol{u_l}^H \left(\sum_{l=1,l\neq k}^K \eta_{\tau_{kl}} \rho_l \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H + \boldsymbol{I_M}\right) \boldsymbol{u_l}} \geq \gamma_k^*,\ \ k = 1, \cdots, K.$$

where $\tilde{\boldsymbol{h}}_{\boldsymbol{k}} = \boldsymbol{h_k}/\sigma_k$. Then, the beamforming direction $\boldsymbol{u_k}$ and beamforming amplitude ρ_k can be recursively updated to find the optimal answer, as follows:

- Update beamforming direction: $\boldsymbol{u_k}(t+1) = \left(\sum_{l=1,l\neq k}^K \eta_{\tau_{kl}} \rho_l(t) \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H + \boldsymbol{I_M}\right)^{-1} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}$.
- Update beamforming amplitude: $\rho_k(t+1) = \frac{\gamma_k^*}{\mu_k(t)} \rho_k(t)$ where $\mu_k(t) = \frac{\rho_k(t) \boldsymbol{u}^H(t)_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H \boldsymbol{u_k}(t)}{\boldsymbol{u_l}^H(t) \left(\sum_{l=1,l\neq k}^K \eta_{\tau_{kl}} \rho_l(t) \tilde{\boldsymbol{h}}_{\boldsymbol{k}} \tilde{\boldsymbol{h}}_{\boldsymbol{k}}^H + \boldsymbol{I_M}\right) \boldsymbol{u_l}(t)}$.

The rate constraint inequalities are active at the optimal point, thus, to satisfy the rate constraint inequalities, the power control procedure is adopted as follows $\boldsymbol{\rho} = \boldsymbol{F}^{-1} \boldsymbol{\gamma^*}$ where $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_K)$, $\boldsymbol{\gamma^*} = (\gamma_1^*, \cdots, \gamma_K^*)$ and $[\boldsymbol{F}]_{i,j} = \begin{cases} \boldsymbol{u_i}^H \tilde{\boldsymbol{h}}_{\boldsymbol{i}} \tilde{\boldsymbol{h}}_{\boldsymbol{i}}^H \boldsymbol{u_i} & i = j \\ -\eta_{\tau_{ij}} \gamma_i^* \boldsymbol{u_j}^H \boldsymbol{h_i} \boldsymbol{h_i}^H \boldsymbol{u_j} & i \neq j \end{cases}$.

APPENDIX C
PROOF OF PROPOSITION 2

To show the superiority of Method *B* compared with Method *A*, we simply assume uniform power distribution function in Method *B* which simplifies the achievable rate expression for User k to $r_k^B = \int_0^1 \log_2 \left(1 + \frac{|\boldsymbol{h_k}^H \boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^K \lambda_{kl}(f) |\boldsymbol{h_k}^H \boldsymbol{w_l}|^2 + \sigma_k^2}\right) df$. Then, to prove that the asynchronous method with uncorrelated precoding can reduce the power transmission, we show that for the set of optimal beamforming vectors obtained for Method *A*, we have $r_k^B > r_k^A$. Then, the beamforming vectors' amplitudes can be accordingly reduced which results in power reduction. To show that, we use the Jensen's inequality for the convex function of $\log_2(1 + 1/x)$, which results in $\sum_i \log_2(1 + 1/x_i) \geq \log_2(1 + 1/\sum_i x_i)$. Therefore, we have:

$$r_k^B = \int_0^1 log_2 \left(1 + \frac{|\boldsymbol{h_k}^H \boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^K \lambda_{kl}(f) |\boldsymbol{h_k}^H \boldsymbol{w_l}|^2 + \sigma_k^2}\right) df \geq \log_2 \left(1 + \frac{|\boldsymbol{h_k}^H \boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^K \int_0^1 \lambda_{k,l}(f) df \boldsymbol{h_k}^H \boldsymbol{w_l}|^2 + \sigma_k^2}\right) = r_k^A. \quad (28)$$

Besides, by applying the Parseval's theorem, we can show that $\eta_{\tau_{kl}} = \sum_{n=-\infty}^{\infty} g_{\tau_{kl}}^2(n) = \int_0^1 |G_{\tau_{kl}}(f)|^2 df = \int_0^1 \lambda_{kl}(f) df$ (refer to Appendix A for more details). Thus, it can be seen that the right hand side of the inequality is equal to $r_k^A = \log_2 \left(1 + \frac{|\boldsymbol{h_k}^H \boldsymbol{w_k}|^2}{\sum_{l=1,l\neq k}^K \eta_{\tau_{k,l}} |\boldsymbol{h_k}^H \boldsymbol{w_l}|^2 + \sigma_k^2}\right)$ which verifies that $r_k^B > r_k^A$ and concludes the proof.

APPENDIX D
DERIVATION OF THE RATE EXPRESSION FOR METHOD *C* WITH SIMPLIFYING ASSUMPTIONS

The term $\frac{1}{N} \mathrm{tr}(\boldsymbol{\Lambda' P'_g})$ in the average transmit power expression can be written as $\frac{1}{N} \sum_{n=1}^{Nq} \lambda'[n] P'_g[n] = \sum_{k \in \mathcal{G}_g} P_k$ where $P_k = \frac{1}{N} \sum_{n \in \mathcal{I}_k} \lambda'[n] P'_g[n]$. To simplify the optimization

problem, we use a sub-optimal power allocation which is, in fact, optimal for AWGN channels [47]. Assuming an AWGN channel, the optimal capacity-achieving power allocation simplifies to $\lambda'[n]P'_g[n] = P_k,\ \forall n \in \mathcal{I}_k$ which is followed by the concavity of $\log_2$ function and Jensen's inequality. Therefore, by substitution, the optimization problem simplifies to:

$$\min_{\{\boldsymbol{w'_g}\}_{g=1}^{K/q},\{\mathcal{I}_k,P_k\}_{k=1}^{K},\pi(.)} p_{avg} = \sum_{k=1}^{K} P_k||\boldsymbol{w'_{\pi(k)}}||^2 \tag{29}$$

$$s.t. \quad \lim_{N\to\infty} \frac{1}{N} \sum_{n\in\mathcal{I}_k} \log_2 \left(1 + \frac{P_k|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} P_{c(n,j)}|\boldsymbol{h_k^H w'_j}|^2 + \sigma_k^2}\right) \geq r_k^*, \quad k = 1, \cdots, K,$$

where $c(n,j) = \{l \in \mathcal{G}_j | n \in \mathcal{I}_l\}$. In simple words, $P_{c(n,j)}$ is the power adjustment coefficient of the user in Group j which Sub-channel n is assigned to it. In the above optimization, the summation is over assigned sub-channels, however, with a simple sub-channel assignment rule, we can get rid of the summation and the optimization problem can be further simplified. We

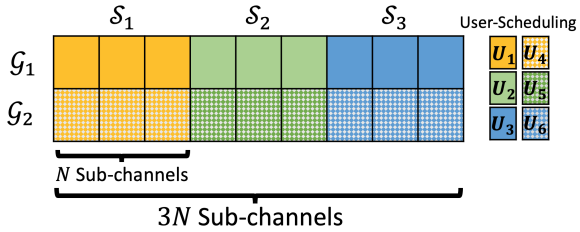


Fig. 14: Schematic representation of the simplified sub-channel assignment.

assume that there are q different sub-channel configurations, $\{\mathcal{S}_1, \cdots, \mathcal{S}_q\}$, where each of them includes K/q different users in K/q different groups, i.e., $|\mathcal{S}_s| = K/q,\ s = 1, \cdots, q$. Define a sub-channel-assignment function that assigns each user to a sub-channel configuration, i.e., $\phi : \mathcal{K} \to \mathcal{S}$, where $\mathcal{K} = \{1, \cdots, K\}$ and $\mathcal{S} = \{1, \cdots, q\}$ are the set of user and sub-channel configuration indices, respectively. Each user, k, is assigned to a sub-channel configuration, s, $\phi(k) = s$. An example of the simplified sub-channel assignment is shown in Fig. 14 for $K = 6$ and $q = 3$, where $\mathcal{G}_1 = \{1, 2, 3\}$, $\mathcal{G}_2 = \{4, 5, 6\}$, $\mathcal{S}_1 = \{1, 4\}$, $\mathcal{S}_2 = \{2, 5\}$ and $\mathcal{S}_3 = \{3, 6\}$. Note that the users with the same color (same sub-channel configuration) interfere with each other. Therefore, the optimization problem is simplified to

$$\min_{\{\boldsymbol{w'_g}\}_{g=1}^{K/q},\{P_k\}_{k=1}^{K},(\pi(.),\phi(.))} p_{avg} = \sum_{k=1}^{K} P_k||\boldsymbol{w'_{\pi(k)}}||^2 \tag{30}$$

$$s.t. \quad \frac{P_k|\boldsymbol{h_k^H w'_{\pi(k)}}|^2}{\sum_{j=1,j\neq\pi(k)}^{K/q} P_{c(k,j)}|\boldsymbol{h_k^H w'_j}|^2 + \sigma_k^2} \geq \gamma_k^*, \quad k = 1, \cdots, K.$$

where $c(k,j) = \{l \in \mathcal{G}_j | \phi(k) = \phi(l)\}$.

REFERENCES

[1] L. C. Godara, *Handbook of antennas in wireless communications*. CRC press, 2018, vol. 4.

[2] H. Jafarkhani, *Space-time coding: theory and practice*. Cambridge university press, 2005.

[3] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[4] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[5] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO wireless communications*. Cambridge university press, 2007.

[6] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[7] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. IEEE ISIT*, Jun. 2004, p. 174.

[8] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Oct. 1998.

[9] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[10] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[11] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Dec. 2005.

[12] L. Liu and H. Jafarkhani, "Novel transmit beamforming schemes for time-selective fading multiantenna systems," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4767–4781, Dec. 2006.

[13] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, May 2007.

[14] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 802–811, Jun. 2003.

[15] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 11–15, Jan. 2007.

[16] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[17] M. A. Vazquez, A. Perez-Neira, D. Christopoulos, S. Chatzinotas, B. Ottersten, P.-D. Arapoglou, A. Ginesi, and G. Tarocco, "Precoding in multibeam satellite communications: Present and future challenges," *IEEE Trans. Wireless Commun.*, vol. 23, no. 6, pp. 88–95, Dec. 2016.

[18] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup precoding and user scheduling for frame-based satellite communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4695–4707, Apr. 2015.

[19] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[20] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[21] E. Koyuncu and H. Jafarkhani, "Variable-length limited feedback beamforming in multiple-antenna fading channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7140–7165, Nov. 2014.

[22] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," *in Handbook of Antennas in Wireless Communications, L. C. Godara, volume 4, CRC press*, 2018.

[23] S. Verdu, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 4, pp. 733–751, Jul. 1989.

[24] S. Shao, Y. Tang, T. Kong, K. Deng, and Y. Shen, "Performance analysis of a modified V-BLAST system with delay offsets using zero-forcing detection," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3827–3837, 2007.

[25] A. Das and B. D. Rao, "MIMO systems with intentional timing offset," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–14, Dec. 2011.

[26] M. Ganji and H. Jafarkhani, "Interference mitigation using asynchronous transmission and sampling diversity," in *Proc. IEEE GLOBECOM*, Dec. 2016, pp. 1–6.

[27] L. Cottatellucci, R. R. Muller, and M. Debbah, "Asynchronous CDMA systems with random spreadingPart I: Fundamental limits," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1477–1497, Mar. 2010.

[28] J. Cui, G. Dong, S. Zhang, H. Li, and G. Feng, "Asynchronous NOMA for downlink transmissions," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 402–405, Oct. 2016.

[29] X. Zou, B. He, and H. Jafarkhani, "An analysis of two-user uplink asynchronous non-orthogonal multiple access systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1404–1418, Jan. 2019.

[30] M. Ganji and H. Jafarkhani, "Time asynchronous NOMA for downlink transmission," in *Proc. IEEE WCNC*, Apr. 2019, pp. 1–6.

[31] ——, "Improving NOMA multi-carrier systems with intentional frequency offsets," *IEEE Wireless Commun. Lett.*, Aug. 2019.

[32] M. Avendi and H. Jafarkhani, "Differential distributed space-time coding with imperfect synchronization in frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 1811–1822, Nov. 2014.

[33] S. Poorkasmaei and H. Jafarkhani, "Asynchronous orthogonal differential decoding for multiple access channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 481–493, Jan. 2015.

[34] S. Sodagari and H. Jafarkhani, "Enhanced spectrum sharing and cognitive radio using asynchronous primary and secondary users," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 832–835, Apr. 2018.

[35] X. Zhang, M. Ganji, and H. Jafarkhani, "Exploiting asynchronous signaling for multiuser cooperative networks with analog network coding," in *Proc. IEEE WCNC*, Mar. 2017, pp. 1–6.

[36] P.-D. Arapoglou, A. Ginesi, S. Cioni, S. Erl, F. Clazzer, S. Andrenacci, and A. Vanelli-Coralli, "DVB-S2X-enabled precoding for high throughput satellite systems," *International Journal of Satellite Communications and Networking*, vol. 34, no. 3, pp. 439–455, May 2016.

[37] E. E. Tyrtyshnikov, "A unifying approach to some old and new theorems on distribution and clustering," *Linear algebra and its applications*, vol. 232, pp. 1–43, 1996.

[38] Z. Zhu and M. B. Wakin, "On the asymptotic equivalence of circulant and Toeplitz matrices," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2975–2992, Mar. 2017.

[39] R. M. Gray *et al.*, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[40] W. Rudin *et al.*, *Principles of mathematical analysis*. McGraw-hill New York, 1964, vol. 3.

[41] A. J. Goldsmith and M. Effros, "The capacity region of broadcast channels with intersymbol interference and colored gaussian noise," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 219–240, Jan. 2001.

[42] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, Aug. 2002.

[43] S. Huberman, C. Leung, and T. Le-Ngoc, "Dynamic spectrum management (DSM) algorithms for multi-user xDSL," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 1, pp. 109–130, Oct. 2010.

[44] J. Papandriopoulos and J. S. Evans, "SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3711–3724, Jul. 2009.

[45] M. Torbatian, "Communication over asynchronous networks: Signaling and rate-reliability analysis," Ph.D. dissertation, University of Waterloo, 2011.

[46] K. Barman and O. Dabeer, "Capacity of MIMO systems with asynchronous PAM," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3366–3375, Nov. 2009.

[47] Y. J. Kim, "Faster than Nyquist transmission over continuous-time channels: Capacity analysis and coding," Ph.D. dissertation, McGill University, 2013.

[48] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, Feb. 2008.

[49] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, Mar. 2013.

[50] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5132–5142, Aug. 2014.

[51] W.-K. K. Ma, "Semidefinite relaxation of quadratic optimization problems and applications," *IEEE Signal Process. Mag.*, vol. 1053, no. 5888/10, May 2010.

[52] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming, version 2.1," 2014.

[53] S. Zhang and Y. Huang, "Complex quadratic optimization and semidefinite programming," *SIAM Journal on Optimization*, vol. 16, no. 3, pp. 871–890, Jul. 2006.

[54] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[55] J. Seo, Y. Sung, and H. Jafarkhani, "A High-Diversity Transceiver Design for MISO Broadcast Channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2591–2606, May 2019.

[56] H. Won, H. Cai, D. Y. Eun, K. Guo, A. Netravali, I. Rhee, and K. Sabnani, "Multicast scheduling in cellular data networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4540–4549, Sep. 2009.

[57] B. Di, L. Song, and Y. Li, "Sub-Channel Assignment, Power Allocation, and User Scheduling for Non-Orthogonal Multiple Access Networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7686–7698, Nov. 2016.