Advantages of NOMA for Multi-User BackCom Networks

Zhiguo Ding, *Fellow, IEEE* and H. Vincent Poor, *Life Fellow, IEEE*

Abstract—Ambient backscatter communication (BackCom) is faced with the challenge that a single BackCom device can occupy multiple orthogonal resource blocks unintentionally. As a result, in order to avoid co-channel interference, a conventional approach is to serve multiple BackCom devices in different time slots, which reduces both spectral efficiency and connectivity. This letter demonstrates that the use of non-orthogonal multiple access (NOMA) can efficiently improve the system throughput and support massive connectivity in ambient BackCom networks. In particular, two transceiver design approaches are developed in the letter to realize different tradeoffs between system performance and complexity.

Index Terms—Non-orthogonal multiple access, backscatter communications, space division multiple access, orthogonal frequency-division multiple access.

I. Introduction

Recently, various novel backscatter communication (Back-Com) approaches have been developed to support the envisioned ultra-massive machine type communications (umMTC) [1]–[5]. The key idea of these BackCom approaches is to use the signals sent by users in a legacy system for exciting the circuits of BackCom devices. This type of BackCom is featured by the challenge that a single BackCom device can occupy multiple orthogonal resource blocks unintentionally. For example, if the legacy system is based on orthogonal frequency-division multiple access (OFDMA), a BackCom device needs to reflect all the legacy signals sent at different subcarriers, which means that a signal sent by the device appears at multiple subcarriers. A similar challenge also presents if space division multiple access (SDMA) is used in the legacy system. In order to avoid the co-channel interference between BackCom devices, orthogonal multiple access (OMA) approaches have been conventionally used, e.g., BackCom devices are served in different time slots [2], [3], which reduces both spectral efficiency and connectivity.

This letter is to demonstrate the advantage of using non-orthogonal multiple access (NOMA) to improve the system throughput and connectivity in BackCom networks. In particular, this letter considers a full-duplex (FD) network, where downlink users are viewed as legacy users and multiple uplink BackCom devices are served simultaneously in the downlink users' channels by using BackCom assisted NOMA (BAC-NOMA) [5]. Unlike [5], multiple legacy users are considered in this letter, where the use of both SDMA and OFDMA in the legacy system is investigated. Two transceiver design approaches are proposed in the letter, where one approach can realize the sum capacity of the multiple access channel (MAC) and the other one can be implemented in a low complexity manner. The provided simulation results demonstrate that

Z. Ding and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA. Z. Ding is also with the School of Electrical and Electronic Engineering, the University of Manchester, Manchester, UK (email: zhiguo.ding@manchester.ac.uk, poor@princeton.edu).

BAC-NOMA realizes a larger throughput and support more devices than OMA, even if random resource allocation is used.

II. System Model

Consider an ambient BackCom network, where SDMA is used in the legacy system. In particular, in the legacy system, a base station serves K downlink users, denoted by U_k, via K spatial beamforming vectors, denoted by $\mathbf{w}_k$, $1 \leq k \leq K$. The application of BAC-NOMA can ensure that additional M uplink BackCom devices, denoted by BD_m, $1 \leq m \leq M$, are admitted to the bandwidth used by the legacy system [5]. It is assumed that each user/device is equipped with a single antenna, the base station has N antennas with the full-duplex capability, and each uplink device is equipped with a BackCom circuit. We note that the proposed BAC-NOMA scheme can also been applied to the case with OFDMA based legacy systems, as shown in Section IV.

Denote x_k by the signal sent by downlink user U_k. Based on the principle of SDMA, the base station broadcasts the following superimposed signals: $\mathbf{s}_0 = \sqrt{P_0} \sum_{k=1}^{K} \mathbf{w}_k x_k$, where P_0 denotes the transmit power of the base station [6]. Optimizing the transmission strategy of the legacy system, i.e., optimizing $\mathbf{w}_k$ and P_0, is beyond the scope of this letter, since it should happen before the uplink BackCom devices are admitted via BAC-NOMA.

Each downlink user receives the following:

$$y_k = \mathbf{g}_k^T \mathbf{s}_0 + \sum_{m=1}^{M} \sqrt{\eta_m} g_{m,k} \mathbf{h}_m^T \mathbf{s}_0 s_m + n_k, \qquad (1)$$

where $\mathbf{g}_k$ denotes the channel vector between the base station and downlink user U_k, $\mathbf{h}_m$ denotes the channel vector between the base station and uplink device BD_m, $g_{m,k}$ denotes the channel gain between U_k and BD_m, η_m denotes the BackCom reflection coefficient of BD_m [7], s_m denotes the signal sent by BD_m, and n_k denotes the receiver noise. For notational convenience, it is assumed that the noise terms at different receivers have the same power, denoted by σ^2.

Because s_m is unknown to the downlink users, the term, $I \triangleq \sum_{m=1}^{M} \sqrt{\eta_m} g_{m,k} \mathbf{h}_m^T \mathbf{s}_0 s_m$, can be treated as interference, whose power is given by $\mathcal{E}_{x_k, s_m} \{II^*\} = P_0 \sum_{m=1}^{M} \eta_m |g_{m,k}|^2 |\mathbf{h}_m^T \mathbf{W}|^2$, where $\mathcal{E}\{\cdot\}$ denotes the expectation operation, and $\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_K \end{bmatrix}$. Therefore, the data rate of downlink user U_k is given by

$$R_k^D = \log \left(1 + \frac{P_0 |\mathbf{g}_k^T \mathbf{w}_k|^2}{P_0 \sum_{i=1, i \neq k}^{K} |\mathbf{g}_k^T \mathbf{w}_i|^2 + \mathcal{E}_{x_k, s_m} \{II^*\} + \sigma^2} \right).$$

The base station receives the following observation:

$$\mathbf{y}_{BS} = \sum_{m=1}^{M} \sqrt{\eta_m} \mathbf{h}_m \mathbf{h}_m^T \mathbf{s}_0 s_m + \mathbf{s}_{SI} + \mathbf{n}_{BS} \qquad (2)$$

where $\mathbf{s}_{SI}$ is assumed to be complex Gaussian distributed, i.e., $\mathbf{s}_{SI} \sim CN(0, \alpha P_0 \mathbf{C}_{SI})$, $\mathbf{C}_{SI}$ denotes covariance matrix of

the self-interference channels and α, $0 \le \alpha \le 1$, indicates the amount of the residual self-interference [8].

Depending on how successive interference cancellation (SIC) is carried out, different sum rates can be realized for the BackCom devices, as shown in the following section.

III. TWO APPROACHES WITH DIFFERENT TRADEOFFS BETWEEN PERFORMANCE AND COMPLEXITY

A. Approach I - A Sum-Capacity Approaching Design

By applying the noise pre-whitening process [9], i.e., applying a detection matrix $\left(\sigma^2\mathbf{I}_N + \alpha P_0 \mathbf{C}_{\mathrm{SI}}\right)^{-\frac{1}{2}}$ to $\mathbf{y}_{\mathrm{BS}}$, the model in (2) can be rewritten as follows:

$$\tilde{\mathbf{y}}_{\mathrm{BS}} = \sum_{m=1}^{M} \sqrt{\eta_m} \mathbf{h}_m^T \mathbf{s}_0 \tilde{\mathbf{h}}_m s_m + \tilde{\mathbf{n}}_{\mathrm{BS}} = \tilde{\mathbf{H}}\mathbf{D}\mathbf{s} + \tilde{\mathbf{n}}_{\mathrm{BS}} \quad (3)$$

where $\tilde{\mathbf{y}}_{\mathrm{BS}} = \left(\sigma^2\mathbf{I}_N + \alpha P_0 \mathbf{C}_{\mathrm{SI}}\right)^{-\frac{1}{2}} \mathbf{y}_{\mathrm{BS}}$, $\tilde{\mathbf{h}}_m = \left(\sigma^2\mathbf{I}_N + \alpha P_0 \mathbf{C}_{\mathrm{SI}}\right)^{-\frac{1}{2}} \mathbf{h}_m$, $\tilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{h}}_1 & \cdots & \tilde{\mathbf{h}}_M \end{bmatrix}$, $\tilde{\mathbf{n}}_{\mathrm{BS}} = \left(\sigma^2\mathbf{I}_N + \alpha P_0 \mathbf{C}_{\mathrm{SI}}\right)^{-\frac{1}{2}} (\mathbf{s}_{\mathrm{SI}} + \mathbf{n}_{\mathrm{BS}})$, $\mathbf{D}$ is a diagonal matrix with the elements on its main diagonal as $D_{m,m} = \sqrt{\eta_m} \mathbf{h}_m^T \mathbf{s}_0$, and $\mathbf{s} = \begin{bmatrix} s_1 & \cdots & s_M \end{bmatrix}^T$.

By treating $\mathbf{D}$ as a power allocation matrix, the system model in (3) can be viewed as a special case of a conventional MAC, whose sum capacity can be realized as follows [9]. Without loss of generality, assume that BD_m' signal is decoded at the m-th stage of SIC, where it is straightforward to show that different SIC decoding orders lead to the same sum capacity. Further assume that prior to decoding BD_m' signal, the signals from BD_i, $1 \le i \le m - 1$, have been decoded correctly. By applying $\sqrt{\eta_m} \mathbf{h}_m^H \mathbf{s}_0^* \tilde{\mathbf{h}}_m^H \left(\mathbf{I}_N + \sum_{i=m+1}^{M} \eta_i |\mathbf{h}_i^T \mathbf{s}_0|^2 \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H\right)^{-1}$ as the detector, the following data rate is achievable to BD_m:

$$R_m^{\mathrm{MAC}} = \log\det\left(\mathbf{I}_N + \left(\mathbf{I}_N + \sum_{i=m+1}^{M} \eta_i |\mathbf{h}_i^T \mathbf{s}_0|^2 \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H\right)^{-1}\right.$$
$$\left. \times \eta_m |\mathbf{h}_m^T \mathbf{s}_0|^2 \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^H\right), \quad (4)$$

for $1 \le m \le M - 1$, and $R_M^{\mathrm{MAC}} = \log\det\left(\mathbf{I}_N + \eta_M |\mathbf{h}_M^T \mathbf{s}_0|^2 \tilde{\mathbf{h}}_M \tilde{\mathbf{h}}_m^H\right)$. Therefore, the sum capacity of the uplink BackCom devices achieved by this type of SIC is given by

$$R_{\mathrm{sum}}^{\mathrm{MAC}} = \log\det\left(\mathbf{I}_N + \sum_{m=1}^{M} \eta_m \left|\mathbf{h}_m^T \mathbf{s}_0\right|^2 \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^H\right). \quad (5)$$

This letter considers a problem of throughput maximization for the BackCom devices, which can be formulated as follows:

$$\max_{\eta_m} \quad R_{\mathrm{sum}}^{\mathrm{MAC}} \quad (\text{P1a})$$

$$\text{s.t.} \quad \sum_{m=1}^{M} \eta_m |g_{m,k}|^2 |\mathbf{h}_m^T \mathbf{W}|^2 \le \tau_k, \quad 1 \le k \le K \quad (\text{P1b})$$

$$0 \le \eta_m \le 1, \quad 1 \le m \le M, \quad (\text{P1c})$$

where τ_k indicates the tolerable interference experienced by downlink user U_k. If U_k has a target data rate, denoted by R_k, one choice of τ_k is given by $\tau_k = \frac{|\mathbf{g}_k^T \mathbf{w}_k|^2 - \epsilon_k \sum_{i=1, i \ne k}^{K} |\mathbf{g}_k^T \mathbf{w}_i|^2}{\epsilon_k} - \frac{\sigma^2}{P_0}$, and $\epsilon_k = 2^{R_k} - 1$.

Remark 1: Problem P1 is a concave optimization problem since its objective function is in a log-det form and its constraints are affine [10]. Various optimization solvers, such as Matlab fmincon, can be straightforwardly applied to find the optimal solution of Problem P1.

Remark 2: The solution of Problem P1 is based on an instantaneous realization of x_k. Therefore, significant system overhead can be consumed for the base station to inform the BackCom devices about the optimal choices of η_m^*. A low-complexity alternative is to use random choices of η_m which satisfy (P1b) and (P1c). The simulation results provided in Section V show that Approach I with random η_m can still significantly outperform OMA.

Remark 3: Another type of complexity introduced by Approach I is explained in the following. During each SIC step, $\sqrt{\eta_m} \mathbf{h}_m^H \mathbf{s}_0^* \tilde{\mathbf{h}}_m^H \left(\mathbf{I}_N + \sum_{i=m+1}^{M} \eta_i |\mathbf{h}_i^T \mathbf{s}_0|^2 \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H\right)^{-1}$ needs to be computed, where the inverse of the matrix requires a computational complexity of $\mathcal{O}(N^3)$. In the case of the number of antennas, N, is large, the computational complexity for generating this detector can be significant, which motivates the low-complexity approach introduced in the next section.

B. Approach II - A Low-complexity QR Based Deign

Assume that the composite channel matrix $\tilde{\mathbf{H}}$ in (3) can be decomposed via QR decomposition as follows: $\tilde{\mathbf{H}} = \mathbf{Q}\mathbf{R}$, where $\mathbf{Q}$ is an $N \times N$ unitary matrix, and $\mathbf{R}$ is an $N \times M$ upper triangular matrix [11]. The base station can use $\mathbf{Q}^H$ as a detection matrix, which simplifies the system model in (3) as follows:

$$\mathbf{Q}^H \tilde{\mathbf{y}}_{\mathrm{BS}} = \mathbf{R}\mathbf{D}\mathbf{s} + \mathbf{Q}\tilde{\mathbf{n}}_{\mathrm{BS}}. \quad (6)$$

By using the upper triangular structure of $\mathbf{R}$, SIC can be implemented in a low-complexity manner [11]. In particular, during the $(M - m + 1)$-th step, the signal from BD_m can be decoded with the following data rate:

$$R_m = \log\left(1 + R_{m,m}^2 \eta_m |\mathbf{h}_m^T \mathbf{s}_0|^2\right), \quad (7)$$

where $R_{m,m}$ is defined similar to $D_{m,m}$. To further reduce system overhead, it is ideal to formulate the resource allocation problem based on the following average sum rate:

$$\bar{R}_{\mathrm{sum}} = \sum_{m=1}^{M} \mathcal{E}_{x_k} \left\{\log\left(1 + R_{m,m}^2 \eta_m |\mathbf{h}_m^T \mathbf{s}_0|^2\right)\right\}. \quad (8)$$

Therefore, the considered long-term throughput maximization problem can be formulated as follows:

$$\max_{\eta_m} \quad \bar{R}_{\mathrm{sum}} \quad \text{s.t.} \quad (\text{P1b}), (\text{P1b}). \quad (\text{P2a})$$

We note that there is an analogy between Problem P2 and the one developed for cognitive MAC [12]; however, the explicit expression of $\bar{R}_{\mathrm{sum}}$ can be obtained, as shown in the following. First define $\bar{R}_m$ as follows:

$$\bar{R}_m = \mathcal{E}_{x_k} \left\{\log\left(1 + R_{m,m}^2 \eta_m \mathbf{s}_0^H \mathbf{h}_m^* \mathbf{h}_m^T \mathbf{s}_0\right)\right\}. \quad (9)$$

Recall that $\mathbf{s}_0 = \sqrt{P_0}\sum_{k=1}^{K}\mathbf{w}_k x_k$, which means that $\bar{R}_m$ can be rewritten as follows:

$$\bar{R}_m = \mathcal{E}_{x_k}\left\{\log\left(1 + P_0 R_{m,m}^2 \eta_m \mathbf{x}^H \mathbf{W}^H \mathbf{h}_m^* \mathbf{h}_m^T \mathbf{W}\mathbf{x}\right)\right\}. \tag{10}$$

A closed-form expression of $\bar{R}_m$ can be found by using the method developed in [13]:

$$\bar{R}_m = \log(e)\mathcal{E}_{x_k}\left\{\int_0^\infty \left(\frac{e^{-t}}{t}\right.\right. \tag{11}$$
$$\left.\left. - \frac{1}{t}e^{-t\left(1 + P_0 R_{m,m}^2 \eta_m \mathbf{x}^H \mathbf{W}^H \mathbf{h}_m^* \mathbf{h}_m^T \mathbf{W}\mathbf{x}\right)}\right)dt\right\},$$

which can be simplified as follows:

$$\bar{R}_m = \log(e)\int_0^\infty \left(\frac{e^{-t}}{t}\right. \tag{12}$$
$$\left. - \frac{e^{-t}}{t}\mathcal{E}_{x_k}\left\{e^{-tP_0 R_{m,m}^2 \eta_m \mathbf{x}^H \mathbf{W}^H \mathbf{h}_m^* \mathbf{h}_m^T \mathbf{W}\mathbf{x}}\right\}\right)dt.$$

Assuming that x_k's are independent and identically distributed (i.i.d.) complex Gaussian variables with mean zero and unit variance, $\bar{R}_m$ can be evaluated as follows:

$$\bar{R}_m = \log(e)\int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-t}}{t\left(1 + tP_0 R_{m,m}^2 \eta_m |\mathbf{h}_m^T \mathbf{W}|^2\right)}\right)dt$$
$$= \log(e)f\left(P_0 R_{m,m}^2 \eta_m |\mathbf{h}_m^T \mathbf{W}|^2\right), \tag{13}$$

where $f(x) \triangleq -e^{\frac{1}{x}}E_i\left(-\frac{1}{x}\right)$ and the last step follows from [14, 3.352.6] and $E_i(\cdot)$ denotes the exponential integral function. We note that an alternative way to obtain (13) is to treat $\mathbf{h}_m^T \mathbf{W}\mathbf{x}$ in (10) as a complex Gaussian variable with mean zero and variance $|\mathbf{h}_m^T \mathbf{W}|^2$.

So Problem P2 can be recast in the following equivalent form:

$$\max_{\eta_m} \quad \sum_{m=1}^{M} -e^{\frac{1}{P_0 R_{m,m}^2 \eta_m |\mathbf{h}_m^T \mathbf{W}|^2}} E_i\left(-\frac{1}{P_0 R_{m,m}^2 \eta_m |\mathbf{h}_m^T \mathbf{W}|^2}\right) \tag{P3a}$$

$$\text{s.t.} \quad \text{(P1b), (P1b).} \tag{P3b}$$

Although Problem P3 contains the exponential integral function, it is still concave as shown in the following lemma.

Lemma 1. *Problem P3 is a concave optimization problem.*

Proof. See Appendix A. $\qquad\square$

Remark 4: Because Problem P3 is concave, it can be straightforwardly solved by using various optimization solvers. We note that the computation of the function $f(x)$ in (13) can be difficult even for a moderately small x. For example, for $x = 0.0013$, $\frac{1}{x} = 750$, and Matlab returns $f(750) =$ Inf. To overcome this computational issue, the following approximation of $f(x)$ is used for small x. Recall that $-E_i\left(-\frac{1}{x}\right) = \Gamma\left(0, \frac{1}{x}\right)$, where $\Gamma(\beta, y)$ denotes the incomplete gamma function and can be approximated as follows [14]:

$$\Gamma(\beta, y) \approx y^{\beta-1}e^{-y}\sum_{m=0}^{L-1}\frac{(-1)^m \Gamma(1-\beta+m)}{y^m \Gamma(1-\beta)}, \tag{14}$$

for $y \to \infty$. By letting $L = 1$ and $\beta = 0$, $f(x)$ can be approximated as follows:

$$f(x) = -e^{\frac{1}{x}}\Gamma\left(0, \frac{1}{x}\right) \approx x, \quad x \to 0, \tag{15}$$

which can be used to approximate (P3a).

Remark 5: Compared to Approach I, the QR based design can be implemented with low computational complexity, as explained in the following. First, there is no need to calculate the inverse of a matrix with size of N at each SIC step. Second, the resource allocation solution is not based on the instantaneous realizations of x_k, which reduces the system complexity. However, it is worth to point out that, unlike Approach I, Approach II cannot achieve the sum capacity of MAC, or support the overloading case, i.e., $M > N$.

IV. EXTENSION TO OFDMA-BASED LEGACY SYSTEMS

The aforementioned BAC-NOMA scheme can also be applied to the case, where OFDMA is used in the legacy system. In particular, in the considered legacy system, a base station serves K downlink users, each denoted by U_k, $1 \le k \le K$, via K orthogonal OFDMA subcarriers. If OMA is used, all the K subcarriers will be occupied by a single BackCom device, since a signal reflected by one BackCom device can block all subcarriers. The application of BAC-NOMA can ensure that M uplink BackCom devices are simultaneously admitted to share the subcarriers. For the purpose of illustration, it is assumed that each node is equipped with a single antenna.

Without loss of generality, assume that downlink user U_k is served at the k-th subcarrier. Following the ambient BackCom model in [2], [3], at subcarrier k, the frequency-domain baseband signal received by downlink user U_k is given by

$$y_k^{\mathrm{D}} = \sqrt{P_0}G_k x_k + \sqrt{P_0}\sum_{m=1}^{M}\sqrt{\eta_m}G_{m,k}H_{m,k}x_k s_m + n_k^{\mathrm{D}},$$

where G_k denotes the channel gain between the base station and U_k at subcarrier k, $H_{m,k}$ denotes the forward channel gain from the base station to uplink device BD_m at subcarrier k, $G_{m,k}$ denotes the channel gain between U_k and BD_m at subcarrier k, and n_k^D denotes the receiver noise.

Because s_m is unknown to the downlink users, the term, $I_o \triangleq \sqrt{P_0}\sum_{m=1}^{M}\sqrt{\eta_m}G_{m,k}^k H_{m,k}x_k s_m$, is again treated as interference. The power of this interference term is given by $\mathcal{E}_{x_k,s_m}\{I_o I_o^*\} = P_0 \sum_{m=1}^{M}\eta_m |G_{m,k}|^2 |H_{m,k}|^2$. Therefore, the data rate of downlink user U_k is given by

$$R_k^{\mathrm{D}} = \log\left(1 + \frac{P_0|G_k|^2}{P_0 \sum_{m=1}^{M}\eta_m |G_{m,k}|^2 |H_{m,k}|^2 + \sigma^2}\right). \tag{16}$$

At the base station, the frequency-domain baseband signal at the k-th subcarrier is given by

$$y_k^{\mathrm{BS}} = \sqrt{P_0}\sum_{m=1}^{M}\sqrt{\eta_m}F_{m,k}H_{m,k}x_k s_m + s_k^{\mathrm{SI}} + n_k^{\mathrm{BS}}, \tag{17}$$

where $F_{m,k}$ denotes BD_m's backward channel gain at subcarrier k, s_k^{SI} denotes the self-interference and n_k^{BS} denote the noise. As in the previous section, it is assumed that $s_k^{\mathrm{SI}} \sim$

$CN(0, \alpha P_0 |h_{\mathrm{SI}}^k|^2)$, where h_{SI}^k denotes the self-interference channel. Furthermore, it is assumed that self-interference at different subcarriers is independent.

By applying the pre-whitening process, the system model at the base station can be expressed as follows:

$$\breve{y}_k^{\mathrm{BS}} = \sqrt{P_0} x_k \breve{\mathbf{h}}_k^H \boldsymbol{\eta}^{\frac{1}{2}} \mathbf{s} + \breve{n}_k^{\mathrm{BS}}, \qquad (18)$$

where $\breve{y}_k^{\mathrm{BS}} = \left(\alpha P_0 |h_{\mathrm{SI}}^k|^2 + \sigma^2\right)^{-\frac{1}{2}} y_k^{\mathrm{BS}}$, $\breve{\mathbf{h}}_k = \left(\alpha P_0 |h_{\mathrm{SI}}^k|^2 + \sigma^2\right)^{-\frac{1}{2}} \left[F_{1,k} H_{1,k} \quad \cdots \quad F_{M,k} H_{M,k}\right]^H$, $\mathbf{s} = \left[s_1 \quad \cdots \quad s_M\right]^T$, $\boldsymbol{\eta}$ is an $M \times M$ diagonal matrix, i.e., $\boldsymbol{\eta} = \mathrm{diag}\{\eta_1, \cdots, \eta_M\}$, and $\breve{n}_k^{\mathrm{BS}}$ is a complex Gaussian white noise with mean zero and unit variance. Stacking the K observations in one vector, the system model at the base station can be rewritten as follows:

$$\breve{\mathbf{y}}^{\mathrm{BS}} = \sqrt{P_0} \mathbf{D}_x \breve{\mathbf{H}}^H \boldsymbol{\eta}^{\frac{1}{2}} \mathbf{s} + \breve{\mathbf{n}}^{\mathrm{BS}}, \qquad (19)$$

where $\mathbf{D}_x$ is an $K \times K$ diagonal matrix, i.e., $\mathbf{D}_x = \mathrm{diag}\{x_1, \cdots, x_K\}$, $\breve{\mathbf{H}} = \left[\breve{\mathbf{h}}_1 \quad \cdots \quad \breve{\mathbf{h}}_K\right]$, $\breve{\mathbf{y}}^{\mathrm{BS}} = \left[\breve{y}_1^{\mathrm{BS}} \quad \cdots \quad \breve{y}_K^{\mathrm{BS}}\right]^T$, and $\breve{\mathbf{n}}^{\mathrm{BS}}$ is constructed similarly to $\breve{\mathbf{y}}^{\mathrm{BS}}$.

By defining $\bar{\mathbf{H}} = \sqrt{P_0} \mathbf{D}_x \breve{\mathbf{H}}^H$ and treating $\boldsymbol{\eta}^{\frac{1}{2}}$ as a power allocation matrix, one can view the system model in (19) as a special case of conventional MAC, whose sum capacity can be realized as follows. Without loss of generality, assume that BD_m' signal is decoded at the m-th stage of SIC, where it is noted that that different SIC decoding orders lead to the same sum capacity. Further assume that prior to decoding BD_m' signal, the signals from BD_i, $1 \leq i \leq m-1$, have been decoded correctly. By applying $\bar{\mathbf{h}}_m^H \left(\mathbf{I}_K + \sum_{i=m+1}^M \eta_i \bar{\mathbf{h}}_i \bar{\mathbf{h}}_i^H\right)^{-1}$ as the detector, the following data rate is achievable to BD_m:

$$R_m^{\mathrm{MAC}} =$$
$$\log \det \left(\mathbf{I}_K + \left(\mathbf{I}_K + \sum_{i=m+1}^M \eta_i \bar{\mathbf{h}}_i \bar{\mathbf{h}}_i^H\right)^{-1} \eta_m \bar{\mathbf{h}}_m \bar{\mathbf{h}}_m^H\right),$$

for $1 \leq m \leq M-1$, and $R_M^{\mathrm{MAC}} = \log \det \left(\mathbf{I}_K + \eta_M \bar{\mathbf{h}}_M \bar{\mathbf{h}}_m^H\right)$, where $\bar{\mathbf{h}}_i$ denotes the i-th column of $\bar{\mathbf{H}}$. Therefore, the sum capacity of the uplink BackCom devices is given by

$$R_{\mathrm{sum}}^{\mathrm{MAC}} = \log \det \left(\mathbf{I}_K + \sum_{m=1}^M \eta_m \bar{\mathbf{h}}_m \bar{\mathbf{h}}_m^H\right), \qquad (20)$$

which can be used to formulate the following throughput maximization problem:

$$\max_{\eta_m} \quad R_{\mathrm{sum}}^{\mathrm{MAC}} \qquad (\mathrm{P4a})$$
$$\mathrm{s.t.} \quad \sum_{m=1}^M \eta_m |G_{m,k}|^2 |H_{m,k}|^2 \leq \tau_k, \quad 1 \leq k \leq K \qquad (\mathrm{P4b})$$
$$(\mathrm{P1c}).$$

Similar to (P1a), (P4a) is also in the concave log-det form. Therefore, Problem P4 is also concave and hence can be straightforwardly solved.
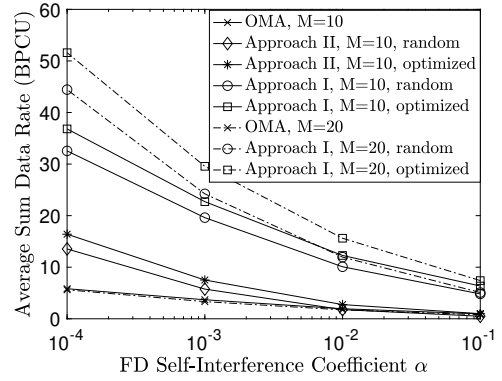
Fig. 1. The impact of FD self-interference on the performance of the considered transmission schemes. BPCU denotes bits per channel use. Case I is used for the locations of the downlink users and $K = N = 10$.

V. SIMULATION RESULTS

In this section, simulation results are presented to evaluate the performance of the proposed BAC-NOMA schemes. For the presented results, the path loss exponent is set as 3, $P_0 = 20$ dBm, $\tau_k = 0.01$, $\sigma_n^2 = -94$ dBm, and the base station is located at $(0,0)$ m. The uplink BackCom devices are randomly located inside a square of side 6 m whose center is located at $(0,0)$ m. Two cases, termed Cases I and II, are considered for the locations of the downlink users. For Case I, the downlink users are randomly located inside a square of side 6 m with its center located at $(0,0)$ m, whereas for case II, the users are also in the same size square with its center at $(3,0)$ m.

In Fig. 1, the impact of FD self-interference on the performance of the considered transmission schemes is studied. As can be seen from the figure, the NOMA schemes can outperform the OMA scheme, since multiple uplink Back-Com devices can be simultaneously supported by the NOMA schemes. Furthermore, the figure shows that Approach I outperforms Approach II and also is applicable to the overloading cases, i.e., $M > N$. However, it is worth to point out that the performance gain of Approach I is obtained at a price of more system complexity, as discussed in Remarks 2 and 3. In addition, Fig. 1 shows that increasing α decreases the performance of all schemes, which is due to the fact that a larger α means more residual FD self-interference and hence leads to more performance degradation.

In Fig. 2, the impact of the number of BackCom devices on the performance of the considered schemes is studied. With more BackCom devices participating in NOMA transmission, the performance of the BAC-NOMA schemes is improved, whereas the impact of M on the performance of OMA is insignificant. Another important observation from Fig. 2 is that the performance of the schemes for Case II is better than that of Case I, which is due to the fact that the interference between the uplink and downlink users is more severe in Case I. As discussed in Section IV, the concept of BAC-NOMA can be straightforwardly extended to the case with OFDMA based legacy systems, which is demonstrated in Fig. 3. Because K subcarriers are used in the considered legacy system, the normalized sum-rate, i.e., $\frac{R_{\mathrm{sum}}^{\mathrm{MAC}}}{K}$, is used as the metric for the performance evaluation. As can be observed from Fig. 3, the BAC-NOMA schemes can outperform OMA, and inviting more Ba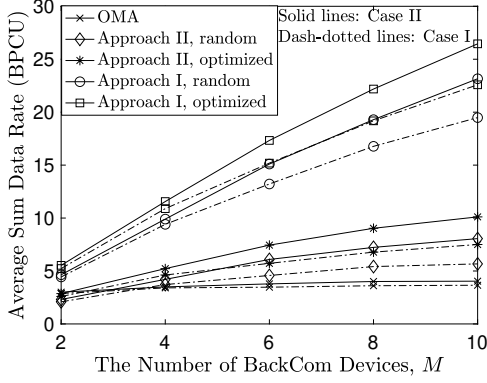ckCom devices to participate in NOMA