Performance of Non-orthogonal Multiple Access (NOMA) with a Novel Asynchronous Interference Cancellation Technique

Huseyin Haci, Huiling Zhu, *Member, IEEE*, and Jiangzhou Wang, *Fellow, IEEE*

Abstract—The non-orthogonal multiple access (NOMA) allows one subcarrier to be allocated to more than one user at the same time in an orthogonal frequency division multiplexing (OFDM) system. NOMA is a promising technique to provide high throughput due to frequency reuse within a cell. In this paper, a novel interference cancellation (IC) technique is proposed for asynchronous NOMA systems. The proposed IC technique exploits a triangular pattern to perform the IC from all interfering users for the desired user. The bit error rate (BER) and capacity performance analysis of an uplink NOMA system with the proposed IC technique is presented, along with the comparison to orthogonal frequency division multiple access (OFDMA) systems. The numerical and simulation results show that the NOMA with the proposed asynchronous IC technique outperforms the OFDMA. It is also shown that employing iterative IC provides significant performance gain for NOMA and the number of required iterations depends on the modulation level and the detection method. With hard-decision, two iterations are sufficient, however with soft-decision, two iterations are enough only for low modulation level, and more iterations are desirable for high modulation level.

Index Terms – wireless communications, non-orthogonal multiple access, asynchronous interference cancellation.

I. Introduction

Orthogonal frequency division multiplexing (OFDM) has been widely adopted in wireless communications [1]. Because of the advantage of transforming a frequency selective fading channel into a number of narrowband flat fading subchannels, wireless multiple access techniques based on OFDM can be realized in two ways – orthogonal frequency division multiple access (OFDMA) [2], [3] and non-orthogonal multiple access (NOMA) [4]. In the OFDMA systems, the system throughput can be maximized by exploiting multiuser diversity gain [2] [3]. That is, based on received channel state information (CSI) of all subchannels, allocating a subcarrier (or a chunk of subcarriers) only to one user with the best CSI. However OFDMA does not allow frequency reuse within one cell, since a subcarrier is allocated only to one user, so that cell throughput is limited. Unlike OFDMA, the NOMA technique can allocate a subcarrier to more than one user at the same time within one cell, so that higher throughput is envisioned

H. Haci is with the Department of Electrical and Electronic Engineering, Near East University, Nicosia, Cyprus (e-mail: huseyin.haci@neu.edu.tr).

H. Zhu and J. Wang are with the School of Engineering and Digital Arts, University of Kent, Canterbury, CT2 7NT United Kingdom (e-mails: {h.zhu, j.z.wang}@kent.ac.uk.)

due to frequency reuse within a cell. Therefore, NOMA is a promising technique for future mobile communications.

Superposition coding (SC) is an effective technique to increase capacity in the NOMA system [5]. When the SC is applied, multiple users' signals are multiplexed over the same subcarrier with different received power at the base station (BS) for uplink transmissions. Then, in the BS, a superimposed signal is received for each subcarrier. The BS detects users' received signals, starting with the user having the strongest signal to noise ratio (SNR), in a descending order of users' SNRs. Once the strongest signal is detected, the detected data is passed through to the successive interference cancellation (SIC) algorithm. The SIC reconstructs the strongest signal by using its CSI, and subtracts it from the received superimposed signal. This suppresses co-channel interference from strong (earlier detected) signals for relatively weak (yet to be detected) signals. [6] investigated the impact of imperfect interference cancellation (IC) on the SC with successive interference cancellation. [7] proved that among all possible signaling methods, superposition coded modulation maximizes the output signal to interference plus noise ratio (SINR) in the linear minimum mean square error (MMSE) estimation. [8] studied the optimal number of users to be superposition-coded on a subcarrier, and concluded that at optimality there is a high probability that only a few (two to three) users are coded on a subcarrier. [9] applied NOMA to cooperative transmissions and showed that cooperative NOMA can achieve the maximum diversity gain for all users at a cooperative transmission. [10] applied MIMO to NOMA and showed that the use of MIMO can significantly improve the capacity of NOMA. IC should perform better when channel coding/decoding is involved since channel decoding can improve the accuracy of interference regeneration. The group decoding scheme [11] may be applied to NOMA.

So far, the researches on NOMA only consider time-synchronous transmissions. However, synchronous transmission is impractical especially for uplink transmissions since users are geographically distributed and the mobile environment is dynamic. Moreover, signals from different users will propagate via different paths and encounter various channel effects. These result in different time offsets when signals from different users arrive at the BS. Although a closed-loop system with feedback channels from the BS to users may help compensate for part of these time offsets, perfect synchronization is hard to achieve in practice. In asynchronous case, symbols in a subcarrier from the different users are time misaligned, which can cause a symbol of a user overlaping with two symbols of each of the other users. The asynchronous communication requires information from multiple symbols when SIC is exploited. Detection and IC performances are significantly

degraded if complete information of signals (desired and interfering) are not known in asynchronous communications [12], [13]. Therefore it is important to investigate NOMA in asynchronous communications.

The objective of this paper is to investigate asynchronous NOMA transmissions and its performance. The contributions of this paper are summarized as follows:

1) A novel technique called "Triangular SIC (T-SIC)" is proposed to perform asynchronous SIC, which uses multiple symbols from each interfering user to carry out IC in a triangular pattern. With the multiple symbol information from each interfering user the SIC performance can be improved substantially. Also the triangular pattern allows the use of minimum number of symbols at the signal processing to detect the desired symbol while achieving low BER performance for all users.

2) Bit error rate (BER) analysis is presented for the T-SIC and conventional SIC (Conv-SIC) techniques. The analysis shows that BER performance of the asynchronous NOMA is a function of the iteration number and received power ratio and time offset between users. The required power ratio of users and the modulation level are studied. The relationship among the number of iterations, detection method (including hard-decision and soft-decision), modulation level and time offset difference among users is investigated.

3) Capacity analysis is given, where performance comparison between NOMA and OFDMA is presented. The analysis show that asynchronous NOMA capacity performance is a function of the relative received power ratio and time offset between users. The effects of the power ratio and time offset difference of users on the capacity performance are presented. It is shown that the proposed T-SIC based NOMA significantly outperforms OFDMA.

The paper is organized as follows. In Section II system model is introduced including the received signal structure and the iterative signal processing at the receiver. T-SIC technique is presented in Section III and the performance analysis is given in Section IV. In Section V representative numerical results are shown. Finally, conclusions are drawn with remarks in Section VI.

II. SYSTEM MODEL

A. Uplink NOMA System and Signal Model

A single cell uplink NOMA system is considered with a BS serving multiple geographically distributed users, as shown in Fig. 1 (a). It is assumed that there are N subcarriers and each subcarrier is shared by K NOMA users, $K \geq 1$, within a small cell environment. It is envisioned that NOMA is used in conjuction with beamforming technology, where there are a small number of users (e.g. $K < 4$) within each beam [14]. Due to different distances and dynamic channels from the users to the BS, asynchronous transmissions are assumed among users. Legacy wireless communication systems employ timing adjustment mechanisms to achieve synchronization among uplink users' signals at the BS. Timing advance (TA)

[15] is the mechanism adopted in current long term evolution (LTE) systems, where the BS measures delay of uplink signal from each user and adjusts the uplink transmission timing by sending the value of advance time to the respective user. Due to the required signalling between the BS and users, updating advance time value frequently to achieve very highly accurate synchronization may introduce much overhead. Moreover, for future ultra high data rate mobile systems, the symbol duration will be extremely small so that perfect synchronization among uplink user transmissions is impractical. Thus it is important to study asynchronous transmissions in the uplink for future wireless communications as an alternative approach.

(a) Uplink NOMA system.
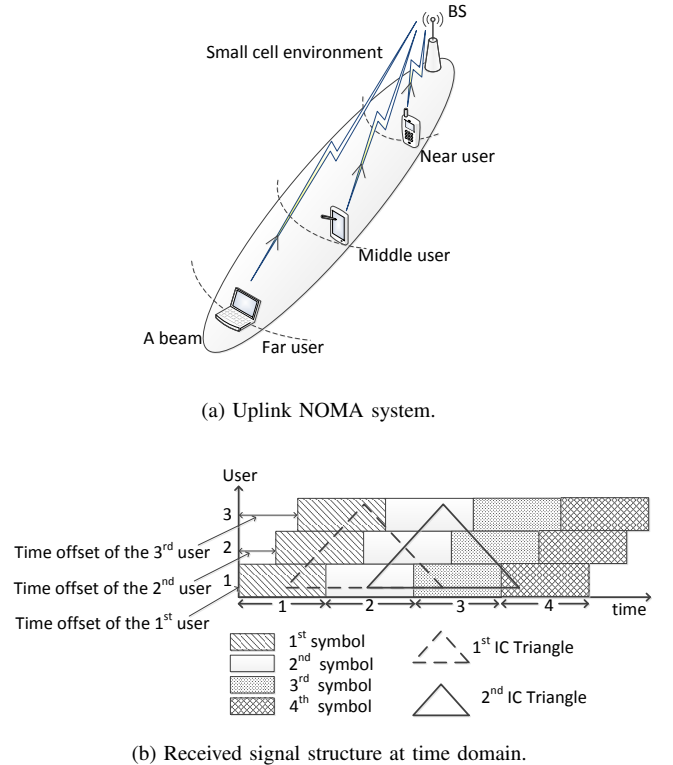

(b) Received signal structure at time domain.

Fig. 1: Illustration of uplink NOMA system and received signal structure.

At the BS receiver, time-to-frequency domain conversion of the received signal is done by fast Fourier transform (FFT) operation. Due to the timing offset between NOMA users, inter-carrier interference (ICI) may occur during this conversion and distort the resultant frequency component, i.e. the OFDM symbol. The ICI at an N subcarrier OFDM system can be expressed as follows. Let an OFDM signal at time t be $x(t) = \sum_{n=1}^{N} X[n] e^{j2\pi f_n t}$, $0 \leq t \leq T_N$, where f_n is the frequency of the nth subcarrier, j represents the complex number, T_N is the symbol time and $X[n]$ is the signal transmitted over the nth subcarrier. The frequency offset due to the asynchronism will introduce a multiplicative time-varying distortion, denoted as $\beta(t) = e^{j2\pi\rho\Delta f t}$, where $\rho = \delta f/\Delta f$ is the ratio of frequency offset δf to subcarrier spacing Δf, to the transmitted signal and the ICI on the mth subcarrier will be [16] $ICI_m = \sum_{\substack{n=1 \\ n \neq m}}^{N} \frac{1}{T_N} \int_0^{T_N} X[n] e^{j2\pi\rho\Delta f t} e^{-j2\pi\Lambda\Delta f t} \, \mathrm{d}t$, where $\Lambda = m - n$ represents the distance (index) of the

interfering subcarrier to the desired subcarrier. On the other hand, at NOMA systems due to frequency reuse the multiple access interference (MAI) is another source of interference that can significantly distort the OFDM symbol of the desired user. Expressions for MAI to the desired user's symbol on a subcarrier are given by (2-4). Fig. 2 shows the ratio of MAI to ICI, $|\text{MAI/ICI}|$, for $K = 3$, where the time offset between users are changed from $[1-15]\%$ of the symbol time. Average received SNR of the users are assumed to be 40, 34 and 28 dB for the 1st, 2nd and 3rd user respectively (see Scenario C in Table I). It can be seen from Fig. 2 that at asynchronous NOMA, MAI dominates ICI. Therefore, in order not to complicate the analysis by considering multiple problems and to provide useful and clear insights about frequency reuse (i.e. MAI) at NOMA, we take MAI as the main/dominant source of interference and focused on *one subcarrier* for the analysis. Further, due to very short symbol duration in high data rate transmissions it will be inefficient to use the cyclic prefix (CP) to compensate for asynchronism between users, since the asynchronism may be a relatively large percentage of the symbol duration in nanoseconds. Therefore it is assumed that CP is used only to mitigate multipath signals but not used to compensate for asynchronism between users.
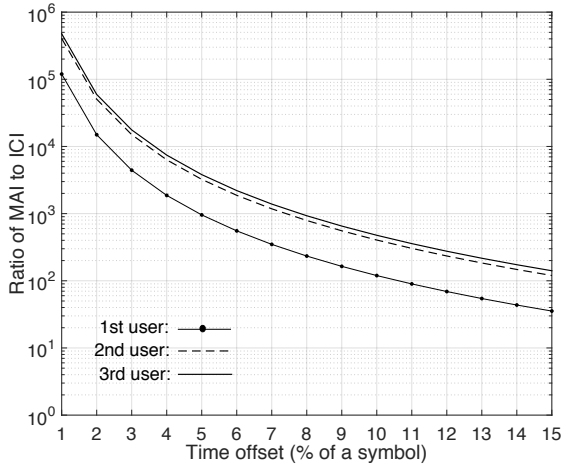

Fig. 2: $|\text{MAI/ICI}|$ versus different levels of asynchronism of users. $\Delta f = 15$ kHz, $N = 1024$.

Received signal structure in one subcarrier is illustrated in Fig. 1 (b), where four adjacent OFDM symbols of three users are shown in time domain. Although we consider three users as an example, the analytic approach is valid for the case of more than three users. For asynchronous transmissions, the arrival times at the BS of the users' signals are not aligned. A symbol from one user overlaps with two adjacent symbols from each of the other two users. In order to carry out IC, the information of two adjacent symbols of other users must be exploited for the reference user (which is denoted as the kth user in the rest of the paper).

Let $X_k[s]$ denote the sth symbol of the kth user, which is a complex symbol and output from a multi-level symbol mapper, such as quadrature amplitude modulation (QAM) [3]. Let $p_k^{\text{tx}}[s]$ be the transmit power allocated to $X_k[s]$ for power domain multiplexing of users' signals. Also let the same level

of power be applied for all symbols during a scheduling period, where the symbol index "s" can be dropped out and p_k^{tx} can be used to denote the transmit power. Then, the signal transmitted from the kth user at the sth symbol can be represented by $X_k[s] \cdot \sqrt{p_k^{\text{tx}}}$.

In one symbol period, the frequency response on one subcarrier is considered to be flat and depend on path loss, flat fading and phase, and is given by $h_k[s] = \sqrt{(d_k)^{-\lambda}}\alpha_k[s]e^{j\theta_k[s]}$, where $(d_k)^{-\lambda/2}$ is the path loss with the distance d_k between the BS and the kth user and propagation exponent λ, and $\alpha_k[s]$ denotes magnitude of fading for the sth symbol of the kth user and is assumed to follow Rayleigh distribution independently and identically (i.i.d.) for different users, with $\mathbb{E}[\alpha_k^2[s]] = 1$, where $\mathbb{E}[\cdot]$ is the statistical expectation. It is also assumed that the received signal for the sth symbol of the kth user has random phase $\theta_k[s]$, uniformly distributed over $[0, 2\pi]$. The channel coherence time is assumed to be much larger than the symbol time and $h_k[s]$ is fixed for block of symbols, during a scheduling period. Therefore index "s" is omitted and h_k is used to denote the CSI. Also, additive white Gaussian noise (AWGN) with double-side power spectral density, $N_0/2$, is assumed on the receiver side.

B. Iterative Signal Processing at the BS Receiver

At the BS receiver, iterative signal processing with maximum L iterations is carried out to improve system performance. At each iteration $l\,(1 \leq l \leq L)$, symbol detection is applied to each user in descending SNR order to estimate symbols of K users. For one user, other co-channel users are sources of interferences which reduce accuracy of symbol detection and cause high BER. Hence the first operation at the receiver is to apply SIC to suppress interference among K users, which requires a priori information (estimate) of symbols from co-channel users. By denoting $\hat{X}_k^l[s]$ as the estimate of the sth symbol of the kth user at the lth iteration, SIC reconstructs interfering signals by multiplying $\hat{X}_k^l[s]$ with the CSI of the kth user, h_k. Reconstructed signals are then subtracted from the received signal to suppress the interference. Due to asynchronism, to reconstruct interference accurately for the sth symbol of kth user, the information of symbols $\{s-1, s, s+1\}$ of the co-channel users is required. In our proposed technique, this information is provided through a vector that stores priori estimated symbols of the kth user at the lth iteration, $\hat{\mathbf{X}}_k^l = \{\hat{X}_k^l[\varsigma], \varsigma \in \{s, s+1, \cdots, s+K-1\}\}, \forall k$, denoted as the "priori symbol vector". Note that the vector contains consecutive symbols up to index $s + K - 1$. This is because, in order to achieve low BER performance for all K users, all the overlapping symbols of a stronger user need to be a priori detected before detecting a symbol of a weak user. For example, in the received signal structure shown in Fig. 1 (b) the order of users' SNR is assumed to be 1st user $>$ 2nd user $>$ 3rd user. The 1st symbol of the 3rd user overlaps with two symbols of the 2nd user and these two symbols overlap with three symbols of the 1st user. Therefore in order to achieve low BER for all three users, it is required to detect three consecutive symbols of the 1st user, then two consecutive symbols of the 2nd user and then the symbol of the 3rd user.

This is so called the 1st IC Triangle as shown in Fig. 1 (b). However, Conv-SIC detects only the 1st symbol of the three users in an iterative manner without constructing IC Triangles.

Let $\mathbf{Y} = \{Y[\varsigma], \varsigma \in \{s, s+1, \cdots, s+K-1\}\}$ denote the vector of the received signal, where $Y[\varsigma]$ is the received signal for the ςth symbol. In Fig. 3 (a), the 1st and 2nd iterations of the signal processing for the 1st IC Triangle are illustrated. At the 1st iteration, since the 1st user has the highest SNR, its three consecutive symbols are detected first, with no information available at the priori symbol vector. Then, the two consecutive symbols of the 2nd user is detected with a priori information of the 1st user detected, but no a priori information for the 3rd user. A priori information obtained from the detection of the 1st and 2nd users are used at the detection for the 1st symbol of the 3rd user. As the priori information of the interference is not achieved for detecting 1st and 2nd users' symbols, the results of these detections are less accurate. Since the SIC and symbol detection performance depends on the accuracy of $\hat{\mathbf{X}}_k^l, \forall k$, there is much room for performance improvement. This is done by employing multiple iterations of symbol detection which is based on exploiting a priori estimate obtained from the previous iteration. For example, at the 2nd iteration, the priori information of the 2nd and 3rd users obtained from the 1st iteration is used at the detection for the 1st user and the detection accuracy of the 2nd iteration is considerably improved. Employing more iterations can further improve system performance, but will increase the processing delay. Thus, for delay sensitive applications, L cannot be too large. After L iterations on the 1st IC Triangle, the signal processing advances in time (i.e. shifted by one symbol) to perform the detection for the next symbol of users, by using the 2nd IC Triangle at Fig. 1 (b).

The procedure of iterative processing for the 2nd IC Triangle is illustrated in Fig. 3 (b). When performing the IC for the 2nd IC Triangle, the priori information obtained from performing the 1st IC Triangle is exploited as initialization and additional information for the IC. As shown in Fig. 1 (b), the 1st and 2nd IC Triangles are overlapped on some symbols, specifically the 2nd and 3rd symbols of the 1st user and the 2nd symbol of the 2nd user. At the 1st iteration of the 2nd IC Triangle, the symbols overlapped by IC Triangles, i.e. 2nd and 3rd symbols, are initialized by values detected the latest at the 1st IC Triangle processing. Then, the order of detection for the 1st iteration of the 2nd IC Triangle is the detection of the 4th symbol of the 1st user, the 3rd symbol of the 2nd user and then the 2nd symbol of the 3rd user. At the 2nd and further detection iterations for the 2nd IC Triangle the estimates of all symbols covered by the 2nd IC Triangle are updated following the similar signal processing to the 2nd iteration of the 1st IC Triangle, with all additional information of interference achieved in the previous IC Triangle. For example, at the 2nd iteration of processing the 2nd IC Triangle, when detecting the 2nd symbol of the 1st user, a priori information of the 1st symbol of the 2nd and 3rd users are adopted from the estimates provided by the 1st IC Triangle processing. Repetitively, processing the following IC Triangles in time sequence is similar to processing the 2nd IC Triangle.

III. PROPOSED APPROACH

This section presents the approach of the T-SIC technique for asynchronous NOMA transmissions. T-SIC constructs IC Triangles by exploiting the triangular pattern of detection and IC, explained as follows. The triangular pattern starts to form the IC Triangle with the weakest user's symbol and sets it as the last symbol of the IC Triangle to be detected. Next, all the symbols of the second weakest user that overlap with the weakest user's symbol are added to the IC Triangle as the symbols to be detected just before the weakest user's symbol. Then, it is moved to the next weakest user and all symbols overlapping with weaker users' symbols included in the triangular pattern are added to the IC Triangle. This procedure is repeated until the strongest user's symbols are added to the IC Triangle. That is, the IC Triangle is formed such that all the symbols of the strongest user which interfere to weaker users' symbols are detected at the beginning, then all the interfering symbols of the next strongest user are detected, and so on. Note that, forming the IC Triangle by starting from the weakest user's symbol and only including the interfering symbols of stronger users, the minimum number of symbols are included at the signal processing of IC Triangles that are necessary to achieve low BER performance to all users. While constructing IC Triangles, i.e. following triangular pattern, T-SIC considers relative time offset between users in order to correctly determine the interfering symbols. Let τ_k represent the time offset of the kth user to a reference time. Fig. 1 (b) shows the received signal structure for $K = 3$ with $\tau_1 < \tau_2 < \tau_3$. Taking the 1st user as the reference user, its sth symbol overlaps with the $(s-1)$th and sth symbols of the 2nd and 3rd users. Thus, T-SIC uses the time offset information of users to correctly determine the symbols of co-channel users that interfere with the symbol to be detected. Then, it searches the priori symbol vector to obtain the latest estimates for these interfering symbols, and uses this information to reconstruct and subtract (i.e. IC) the interfering signals from the received signal.

In Fig. 4, symbols marked with numbers 1-6 illustrates the constructed 1st IC Triangle, where from small to big numbers on the symbols show the order of detection. After completing L iterations on the 1st IC Triangle, the information bits for the 1st symbol of users are output, and the processing advances in time by shifting the IC Triangle by one symbol to the 2nd IC Triangle. The 2nd IC Triangle is illustrated by the symbols marked with $\{2, 3, 5, 7, 8, 9\}$. At the 1st iteration on the 2nd IC Triangle, a priori information (latest) obtained from the 1st IC Triangle is used to initialize values of the common symbols to both IC Triangles, i.e. symbols $\{2, 3, 5\}$, and only the other symbols marked with $\{7, 8, 9\}$ are respectively detected. Then, at the lth iteration, $l \geq 2$, the values of all the symbols included under the 2nd IC Triangle are updated (detected) by exploiting both the information provided from the 1st IC Triangle symbols, i.e. $\{4, 6\}$, and the most recent iteration of the 2nd IC Triangle. The processing for the 3rd and later IC Triangle is similar to the 2nd IC Triangle's.
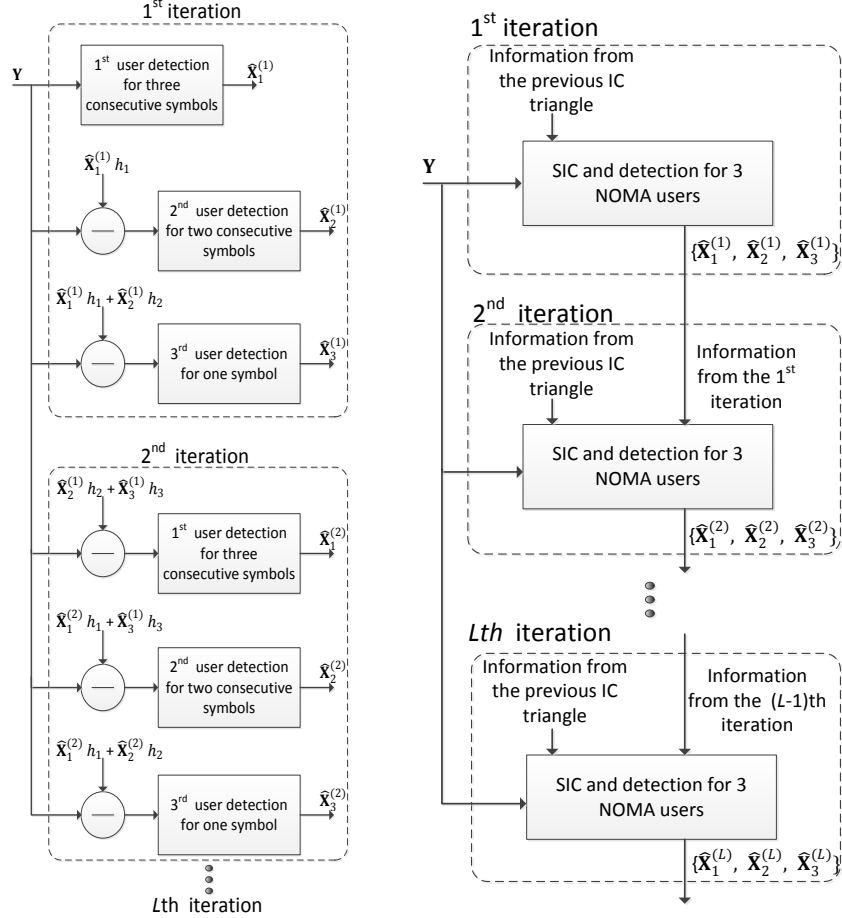
(a) Processing for the 1st IC Triangle.   (b) Processing for the 2nd and further IC Triangles.

Fig. 3: Iterative SIC and symbol detection processing for IC Triangles.

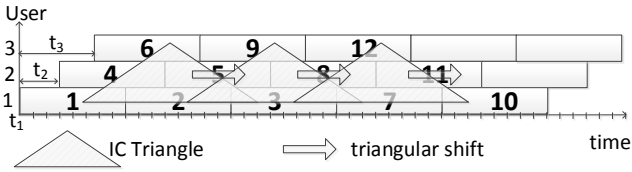

Fig. 4: T-SIC received signal structure.

IV. PERFORMANCE ANALYSIS

In the first subsection the received signal and average BER analysis are presented and the Shannon capacity analysis is given in the second subsection.

A. Received Signal and Average BER Analysis

Let the sth symbol of the k^*th user be the desired symbol to be detected. At the BS receiver, the received signal for the k^*th user at the sth symbol is given by

$$Y_{k^*}[s] = X_{k^*}[s] \cdot \sqrt{p_{k^*}} \cdot \alpha_{k^*} + \eta_{k^*}[s] + N_n. \quad (1)$$

On the right hand side (RHS) of (1), the first term is the desired signal and $p_{k^*} = p_{k^*}^{\text{tx}} \cdot d_{k^*}^{-\lambda/2}$ represents the average received

power of the signal of the k^*th user. N_n is the AWGN. $\eta_{k^*}[s]$ represents the total interference to the desired signal, given by

$$\eta_{k^*}[s] = \sum_{\substack{k \in \Omega \\ k \neq k^*}} \sum_{\varsigma=s-1}^{s+1} \eta_{k^*,k}[s,\varsigma], \quad (2)$$

where Ω is the set of users accessing the subcarrier and $\eta_{k^*,k}[s,\varsigma]$ is the interference from the $\varsigma \in \{(s-1), s, (s+1)\}$th symbol of the kth user to the desired symbol, given by

$$\eta_{k^*,k}[s,\varsigma] = \Delta_{k^*,k}[s,\varsigma] \cdot X_k[\varsigma] \cdot \sqrt{p_k} \cdot \alpha_k \cdot e^{j\theta_{k^*,k}}, \quad (3)$$

where $e^{j\theta_{k^*,k}}$ represents the phase mismatch of the kth user's signal to the k^*th user and $\Delta_{k^*,k}[s,\varsigma]$ represents percentage of the symbol time that the ςth symbol of the kth user overlaps with the desired symbol, given by

$$\Delta_{k^*,k}[s,\varsigma] = \frac{1}{T} \max \Big\{ 0, \Big(\delta(\varsigma - s) \cdot T - |\tau_{k^*} - \tau_k| - 1 \Big) , \Big((1 - \delta(\varsigma - s)) \cdot (\tau_{k^*} - \tau_k) \cdot (\varsigma - s) - 1 \Big) \Big\}, \quad (4)$$

where T is a symbol period. $\delta(\cdot)$ is the Dirac delta function. $\delta(\varsigma - s) = 1$ if $\varsigma = s$ and otherwise zero.

Referring to the iterative SIC and detection structure of T-SIC, shown at Fig. 3, at the lth iteration the reconstructed

interference signal from the ςth symbol of the kth user is given by,

$$\hat{\eta}_{k^*,k}^{(l)}[s,\varsigma] = \Delta_{k^*,k}[s,\varsigma] \cdot \hat{X}_k^{(\mathfrak{L})}[\varsigma] \cdot \sqrt{p_k} \cdot \alpha_k \cdot e^{j\theta_{k^*,k}}, \quad (5)$$

where $\mathfrak{L} \in \{l-1, l\}$ is the latest version of estimate available at the priori symbol vector for the ςth symbol of the kth user. For interferers with higher SNR, i.e. $k < k^*$, $\mathfrak{L} = l$ since symbols of these stronger interferers are most recently detected at the lth iteration before the k^*th user, but for interferers with smaller SNR, i.e. $k > k^*$, $\mathfrak{L} = l-1$ since symbols of these weaker interferers are detected after the k^*th user and only symbols at the $(l-1)$th iteration are estimated.

Then, subtracting the reconstructed interference of all the overlapping symbols of the interferers, the interference cancelled (ICed) signal for the desired symbol can be expressed in terms of the desired signal and residual interference plus noise and given by

$$\tilde{Y}_{k^*}^{(l)}[s] = Y_{k^*}[s] - \sum_{\substack{k\in\Omega \\ k\neq k^*}} \sum_{\varsigma=s-1}^{s+1} \hat{\eta}_{k^*,k}^{(l)}[s,\varsigma]$$

$$= X_{k^*}[s] \cdot \sqrt{p_{k^*}} \cdot \alpha_{k^*} + \tilde{\eta}_{k^*}^{(l)}[s] + N_n, \quad (6)$$

where $\tilde{\eta}_{k^*}^{(l)}[s]$ represents the residual interference at the lth iteration to the desired symbol, given by

$$\tilde{\eta}_{k^*}^{(l)}[s] = \sum_{\substack{k\in\Omega \\ k\neq k^*}} \sum_{\varsigma=s-1}^{s+1} \Delta_{k^*,k}[s,\varsigma] \cdot \left(X_k[\varsigma] - \hat{X}_k^{(\mathfrak{L})}[\varsigma] \right)$$

$$\cdot \sqrt{p_k} \cdot \alpha_k \cdot e^{j\theta_{k^*,k}}. \quad (7)$$

It is clear from (7) that the magnitude of residual interference at asynchronous NOMA is a function of the iteration number and it depends on the accuracy of the most recent detection of interferers' symbols.

The ICed signal, according to (6), is the decision parameter at the lth iteration's symbol detection and with hard-decision the data of the desired symbol is recovered by the minimum distance criteria, given by

$$\hat{X}_{k^*}^{(l)}[s] = \arg\min_{m_i\in\mathfrak{M},\forall i} \left| \tilde{Y}_{k^*}^{(l)}[s] - m_i \right|^2, \quad (8)$$

where $\arg\min$ denotes the argument of the minimum. For QAM, m_i is the ith constellation point of the constellation

$$\mathfrak{M} = \Big\{ m_i = (2i_I - 1 - \sqrt{M}) + j(2i_Q - 1 - \sqrt{M})$$
$$, i_I, i_Q \in \{1, 2, \cdots, \sqrt{M}\} \Big\}. \quad (9)$$

The error statistics at this detection is obtained as follows. The residual interference and noise at the ICed signal distort the desired signal and can cause detection errors. Based on (7), the magnitude of residual interference depends on the accuracy of detection at the $\mathfrak{L}$th iteration for symbols of the interferers (see term $(X_k[\varsigma] - \hat{X}_k^{(\mathfrak{L})}[\varsigma])$ at (7)). Therefore, the error statistics at the lth iteration are conditioned on the detection accuracy of symbols of the interferers at the $(l-1)$th (for weaker interferers) and lth (for stronger interferers) iterations. To obtain these conditional statistics, let $e_k^{(\mathfrak{L})}[\varsigma] = (X_k[\varsigma] \neq \hat{X}_k^{(\mathfrak{L})}[\varsigma])$ denote the event that at the latest

iteration the symbol of the interferer was detected in error and $c_k^{(\mathfrak{L})}[\varsigma] = (X_k[\varsigma] = \hat{X}_k^{(\mathfrak{L})}[\varsigma])$ denote the event that at the latest iteration the symbol of the interferer was correctly detected. Also, $z_k^{(\mathfrak{L})}[\varsigma] \in \{e_k^{(\mathfrak{L})}[\varsigma], c_k^{(\mathfrak{L})}[\varsigma]\}$ denotes the status of the latest detection for the symbol of the interferer, i.e. in error or correct, and $\mathbf{z}^{(\mathfrak{L})} = \{z_k^{(\mathfrak{L})}[\varsigma], \forall k, \forall \varsigma\}$ be the vector form of the latest indicator parameters.

Then, from (7), at the lth iteration the variance of the residual interference to the desired symbol is given by

$$\text{Var}(\tilde{\eta}_{k^*}^{(l)}[s] \mid \mathbf{z}^{(\mathfrak{L})}, \mathbf{\Delta}_{k^*}[s]) = E\left[(\tilde{\eta}_{k^*}^{(l)}[s] \mid \mathbf{z}^{(\mathfrak{L})}, \mathbf{\Delta}_{k^*}[s])^2 \right]$$

$$= \sum_{\substack{k\in\Omega \\ k\neq k^*}} \sum_{\varsigma=s-1}^{s+1} \Delta_{k^*,k}[s,\varsigma] \left(D_k^{(\mathfrak{L})}[\varsigma] \mid z_k^{(\mathfrak{L})}[\varsigma] \right) \frac{p_k}{2}, \quad (10)$$

where $(A \mid B)$ represents A conditioned on B and $\mathbf{\Delta}_{k^*}[s] = \{\Delta_{k^*,k}[s,\varsigma], \forall k, \forall \varsigma\}$ is the vector of interferers' time offset to the desired user. It is seen from (10) that the magnitude of residual interference at asynchronous NOMA is also a function of the relative time offset between users. The distribution of the time offset between users is assumed to be uniform $U[\tau_{min}, \tau_{max}]$, where τ_{min} and τ_{max} are the minimum and maximum time offset, respectively. $\left(D_k^{(\mathfrak{L})}[\varsigma] \mid z_k^{(\mathfrak{L})}[\varsigma] \right) = E[(X_k[\varsigma] - \hat{X}_k^{(\mathfrak{L})}[\varsigma])^2]$ represents the mean square error (MSE) of the detection at the $\mathfrak{L}$th iteration for the ςth symbol of the kth user, given by (11) [17]. Based on (6), at the lth iteration the SINR for the desired symbol is given by

$$\left(\tilde{\gamma}_{k^*}^{(l)}[s] \mid \mathbf{z}^{(\mathfrak{L})}, \alpha_{k^*}^2, \mathbf{\Delta}_{k^*}[s] \right) =$$

$$\frac{p_{k^*} \cdot \alpha_{k^*}^2}{\text{Var}(\tilde{\eta}_{k^*}^{(l)}[s] \mid \mathbf{z}^{(\mathfrak{L})}, \mathbf{\Delta}_{k^*}[s]) + \text{Var}(N_n)}, \quad (12)$$

where $\text{Var}(N_n)$ is the variance of the noise. (12) shows that the SINR at asynchronous NOMA is a function of the accuracy of the most recent detection of interferers' symbols and the relative time offset between users. The desired symbol is detected in error when the power of residual interference plus noise exceeds half of the distance between two nearest constellation points. Since the distance between constellation points depend on the received power of the desired user, the error statistics can be represented in terms of the SINR as follows. The conditional error probability at the lth iteration for the desired symbol is given by

$$P(e_{k^*}^{(l)}[s] \mid \mathbf{z}^{(\mathfrak{L})}, \alpha_{k^*}^2, \mathbf{\Delta}_{k^*}[s]) = 1 -$$

$$\left(1 - P\left(\text{Var}(\tilde{\eta}_{k^*}^{(l)}[s] \mid \mathbf{z}^{(\mathfrak{L})}, \mathbf{\Delta}_{k^*}[s]) + \text{Var}(N_n) > de_{k^*} \right) \right)^2, \quad (13)$$

where de_{k^*} is half of the distance between two nearest constellation points for the desired user, given by [18] (Section 6.1.4), $de_{k^*} = \sqrt{3 \cdot p_{k^*} \cdot \alpha_{k^*}^2 / 2 \cdot (M-1)}$ and $P\left(\text{Var}(\tilde{\eta}_{k^*}^{(l)}[s] \mid \mathbf{z}^{(\mathfrak{L})}, \mathbf{\Delta}_{k^*}[s]) + \text{Var}(N_n) > de_{k^*} \right)$ is the error probability for one branch of QAM. Note that at (13), it is assumed that conditional parameters $\mathbf{z}^{(\mathfrak{L})}, \alpha_{k^*}^2, \mathbf{\Delta}_{k^*}[s]$ are independent of each other and the error probability can be averaged over them one by one. The distribution of the power of residual interference plus noise is required to find out the error probability for one branch of QAM. Based on (10), it is clear that there are two symbols from each co-channel user

$$\left(\mathrm{D}_k^{(\mathfrak{L})}[\varsigma] \,\middle|\, \mathrm{z}_k^{(\mathfrak{L})}[\varsigma]\right) = \begin{cases} 1, & \text{for } \mathfrak{L} = 0, \text{i.e. no priori detection was done for the symbol} \\ 6/(M-1), & \text{for } \mathfrak{L} \geq 1 \text{ and } \mathrm{z}_k^{(\mathfrak{L})}[\varsigma] = \mathrm{e}_k^{(\mathfrak{L})}[\varsigma], \text{i.e. priori detection was in error} \\ 0, & \text{for } \mathfrak{L} \geq 1 \text{ and } \mathrm{z}_k^{(\mathfrak{L})}[\varsigma] = \mathrm{c}_k^{(\mathfrak{L})}[\varsigma], \text{i.e. priori detection was correct} \end{cases} \tag{11}$$

that interfers the desired symbol. Transmitted symbols independently and identically follow discrete uniform distribution over the constellation points. It is shown in Section 4.9 in [19] that the convolution of i.i.d. uniform random variables (r.v.s) converge to the Gaussian distribution very fast. Already for four convoluted r.v.s, i.e. for two asynchronous interferers, the difference between the Gaussian distribution and the exact distribution is negligible ([19] Section 4.9). Therefore for the considered case of $K \geq 2$ the residual interference can be approximated to be Gaussian distributed. Also, the distribution of the power of residual interference plus noise follows Gaussian distribution with zero mean and variance $\mathrm{Var}(\tilde{\eta}_{k^*}^{(l)}[s] \,|\, \mathbf{z}^{(\mathfrak{L})}, \boldsymbol{\Delta}_{k^*}[s]) + \mathrm{Var}(N_n)$. Then, one obtains ([18] Section 6.1.4)

$$\mathrm{P}\left(\mathrm{Var}(\tilde{\eta}_{k^*}^{(l)}[s] \,|\, \mathbf{z}^{(\mathfrak{L})}, \boldsymbol{\Delta}_{k^*}[s]) + \mathrm{Var}(N_n) > \mathrm{de}_{k^*}\right) =$$
$$\mathrm{Q}\left(\sqrt{\frac{3 \cdot \left(\tilde{\gamma}_{k^*}^{(l)}[s] \,|\, \mathbf{z}^{(\mathfrak{L})}, \alpha_{k^*}^2, \boldsymbol{\Delta}_{k^*}[s]\right)}{2 \cdot (M-1)}}\right), \tag{14}$$

where $\mathrm{Q}(\cdot)$ is the Q function. Substituting (14) into (13) the conditional error probability is obtained as

$$\mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s] \,|\, \mathbf{z}^{(\mathfrak{L})}, \alpha_{k^*}^2, \boldsymbol{\Delta}_{k^*}[s]) =$$
$$1 - \left(1 - \mathrm{Q}\left(\sqrt{\frac{3 \cdot \left(\tilde{\gamma}_{k^*}^{(l)}[s] \,|\, \mathbf{z}^{(\mathfrak{L})}, \alpha_{k^*}^2, \boldsymbol{\Delta}_{k^*}[s]\right)}{2 \cdot (M-1)}}\right)\right)^2. \tag{15}$$

Then $\mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s])$ denotes the error probability and $\mathrm{P}(\mathrm{c}_{k^*}^{(l)}[s]) = (1 - \mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s]))$ is the probability of correct detection. The error probability at asynchronous NOMA is a function of the accuracy of the most recent detection of interferers' symbols and the relative time offset between users. In order to obtain $\mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s])$, we first derive the conditional probability in (15) further over the permutations of $\mathbf{z}^{(\mathfrak{L})}$. Then a conditional error probability can be obtained as (16), where $_i\mathrm{P}_{\mathbf{z}^{(\mathfrak{L})}}$ denote the ith permutation of $\mathbf{z}^{(\mathfrak{L})}$, $1 \leq i \leq 2^{2(K-1)}$. $\mathrm{Pr}_{i,k}^{(\mathfrak{L})}[\varsigma]$ is the probability of the detection status, i.e. being in error or correct, of the latest detection $\mathfrak{L}$, i.e. $\mathfrak{L} = l$ for stronger interferers and $\mathfrak{L} = (l-1)$ for weaker interferers, for the ςth symbol of the kth (interfering) user, considered at the ith permutation of $\mathbf{z}^{(\mathfrak{L})}$. Specifically, at the ith permutation of $\mathbf{z}^{(\mathfrak{L})}$ in case the status of the latest detection of the ςth symbol of the kth user is considered to be a correct detection $\mathrm{Pr}_{i,k}^{(\mathfrak{L})}[\varsigma] = 1 - \mathrm{P}(\mathrm{e}_k^{(\mathfrak{L})}[\varsigma])$. Otherwise, $\mathrm{Pr}_{i,k}^{(\mathfrak{L})}[\varsigma] = \mathrm{P}(\mathrm{e}_k^{(\mathfrak{L})}[\varsigma])$. Thus, it is seen from (16) that the error probability of the desired symbol at the lth iteration depends on the error probabilities of interferers' symbols at the most recent detection (current and previous) iteration. Then, $\mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s])$ is obtained by averaging (16) over $\alpha_{k^*}^2$ and $\boldsymbol{\Delta}_{k^*}[s]$, given by

$$\mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s]) = \frac{1}{\tau_{max} - \tau_{min}}$$
$$\cdot \int_{\tau_{min}}^{\tau_{max}} \int_0^\infty \mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s] \,|\, a, \boldsymbol{\Delta}) \cdot e^{-a} \, \mathrm{d}a \, \mathrm{d}\boldsymbol{\Delta}, \tag{17}$$

where $a = \alpha_{k^*}^2$ and $\boldsymbol{\Delta}$ represents $\boldsymbol{\Delta}_{k^*}[s]$. Then, assuming the symbol energy is divided equally among all bits and the Gray coding is used for mapping bits to symbols, the average BER performance is given by ([18] Section 6.1.1),

$$\mathrm{P}_{\mathrm{bit},k^*}^{(l)}[s] = \mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s]) / \log_2(M). \tag{18}$$

A demonstration to obtain $\mathrm{P}(\mathrm{e}_{k^*}^{(l)}[s])$ for T-SIC technique is provided as follows. For easy understanding, suppose $K = 2$ and T-SIC is applied to the 1st IC Triangle (see Fig. 1 (b) for case $K = 2$).

1) The 1st iteration:
Refering to the proposed T-SIC procedure, the 1st step is to detect the 1st symbol of the 1st user. Since in this step there is no a priori detection done for the 1st symbol of the 2nd user (i.e. the interfering symbol), (10)-(15) are conditioned on $\mathbf{z}^{(0)}$. Then based on (11), $\left(\mathrm{D}_2^{(0)}[1] \,\middle|\, \mathfrak{L} = 0\right) = 1$. Note that to avoid notational abundance we show $\mathfrak{L} = 0$ as the conditional parameter at (11), instead of $\mathrm{z}_2^{(0)}[1]$ since the MSE will always be equal to one when there is no a priori detection done for the interfering symbol. And then substituting the MSE into (10), the variance of residual interference is

$$\mathrm{Var}(\tilde{\eta}_1^{(1)}[1] \,|\, \mathfrak{L} = 0, \boldsymbol{\Delta}_1[1]) = \Delta_{1,2}[1,1] \cdot p_2/2. \tag{19}$$

Substituting (19) into (12) and (12) into (15), the conditional error probability is

$$\mathrm{P}(\mathrm{e}_1^{(1)}[1] \,|\, \alpha_1^2, \boldsymbol{\Delta}_1[1]) =$$
$$1 - \left(1 - \mathrm{Q}\left(\sqrt{\frac{3 \cdot p_1 \cdot \alpha_1^2}{2 \cdot (M-1) \cdot (\Delta_{1,2}[1,1] \cdot p_2/2 + N_0)}}\right)\right)^2. \tag{20}$$

Note that at (20) there is no need to condition on $\mathbf{z}^{(0)}$ since MSE will always be equal to one for $\mathfrak{L} = 0$. Then, substituting (20) into (17) we obtain $\mathrm{P}(\mathrm{e}_1^{(1)}[1])$ that is the error probabiliy at the 1st iteration for the 1st symbol of the 1st user. T-SIC will exploit $\mathrm{P}(\mathrm{e}_1^{(1)}[1])$ at later detection steps to obtain conditional error probabilities, e.g. see the 3rd step.

The 2nd step is to detect the 2nd symbol of the 1st user. Similar to the 1st step, the interfering symbols are not a priori detected. Therefore based on (11), $\left(\mathrm{D}_2^{(0)}[1] \,\middle|\, \mathfrak{L} = 0\right) = 1$ and $\left(\mathrm{D}_2^{(0)}[2] \,\middle|\, \mathfrak{L} = 0\right) = 1$, and substituting them into (10), the variance of residual interference is

$$\mathrm{Var}(\tilde{\eta}_1^{(1)}[2] \,|\, \{\mathfrak{L} = 0, \mathfrak{L} = 0\}, \boldsymbol{\Delta}_1[2]) =$$
$$(\Delta_{1,2}[2,1] + \Delta_{1,2}[2,2]) \cdot p_2/2. \tag{21}$$

Substituting (21) into (12) and (12) into (15), the conditional error probability is obtained as (22). Then, substituting (22) into (17) we obtain $\mathrm{P}(\mathrm{e}_1^{(1)}[2])$. Similar to the 1st step, T-SIC will exploit $\mathrm{P}(\mathrm{e}_1^{(1)}[2])$ at later detection steps to obtain conditional error probabilities, e.g. see the 3rd step.

$$P(e_{k^*}^{(l)}[s] \,|\, \alpha_{k^*}^2, \mathbf{\Delta}_{k^*}[s]) = \sum_{i=1}^{2^{2(K-1)}} P(e_{k^*}^{(l)}[s] \,|\, {}_i\mathbf{P}_{\mathbf{z}(\mathfrak{L})}, \alpha_{k^*}^2, \mathbf{\Delta}_{k^*}[s]) \prod_{\substack{k \in \Omega, \varsigma \in \{s-1, s, s+1\} \\ k \neq k^*}} \Pr_{i,k}^{(\mathfrak{L})}[\varsigma] \tag{16}$$

$$P(e_1^{(1)}[2] \,|\, \alpha_1^2, \mathbf{\Delta}_1[2]) = 1 - \left(1 - Q\left(\sqrt{\frac{3 \cdot p_1 \cdot \alpha_1^2}{2 \cdot (M-1) \cdot ((\Delta_{1,2}[2,1] + \Delta_{1,2}[2,2]) \cdot p_2/2 + N_0)}}\right)\right)^2 \tag{22}$$

The 3rd step is to detect the 1st symbol of the 2nd user. At previous steps of this iteration the interfering symbols, i.e. the 1st and 2nd symbols of the 1th user, are detected. Therefore following the T-SIC procedure, $P(e_2^{(1)}[1])$ is conditioned on the permutation of these latest detection results. Then, let $\{e_1^{(1)}[1], e_1^{(1)}[2]\}$ be the 1st permutation that detection results of both symbols of the 1st user was in error. This permutation occurs with probability $(P(e_1^{(1)}[1]) \cdot P(e_1^{(1)}[2]))$. Then based on (11), $\left(D_1^{(1)}[1] \,|\, e_1^{(1)}[1]\right) = (^6/_{(M-1)})$ and $\left(D_1^{(1)}[2] \,|\, e_1^{(1)}[2]\right) = (^6/_{(M-1)})$ and substituting them into (10), the variance of residual interference is

$$\text{Var}(\tilde{\eta}_2^{(1)}[1] \,|\, \{e_1^{(1)}[1], e_1^{(1)}[2]\}, \mathbf{\Delta}_2[1]) = (\Delta_{2,1}[1,1] + \Delta_{2,1}[1,2]) \cdot (^6/_{(M-1)}) \cdot p_1/2. \tag{23}$$

Substituting (23) into (12) and (12) into (15) we obtain $P(e_2^{(1)}[1] \,|\, \{e_1^{(1)}[1], e_1^{(1)}[2]\}, \alpha_2^2, \mathbf{\Delta}_2[1])$. Then, let $\{e_1^{(1)}[1], c_1^{(1)}[2]\}$ be the 2nd permutation that detection result for the 1st symbol was in error but the 2nd symbol was correct for the 1st user. This permutation occurs with probability $(P(e_1^{(1)}[1]) \cdot (1 - P(e_1^{(1)}[2])))$. In this case $\left(D_1^{(1)}[1] \,|\, e_1^{(1)}[1]\right) = (^6/_{(M-1)})$ and $\left(D_1^{(1)}[2] \,|\, c_1^{(1)}[2]\right) = 0$ and substituting them into (10), the variance of residual interference is

$$\text{Var}(\tilde{\eta}_2^{(1)}[1] \,|\, \{e_1^{(1)}[1], c_1^{(1)}[2]\}, \mathbf{\Delta}_2[1]) = \Delta_{2,1}[1,1] \cdot (^6/_{(M-1)}) \cdot p_1/2. \tag{24}$$

Substituting (24) into (12) and (12) into (15) we obtain $P(e_2^{(1)}[1] \,|\, \{e_1^{(1)}[1], c_1^{(1)}[2]\}, \alpha_2^2, \mathbf{\Delta}_2[1])$. Let the 3rd permutation be $\{c_1^{(1)}[1], e_1^{(1)}[2]\}$ which occurs with probability $((1 - P(e_1^{(1)}[1])) \cdot P(e_1^{(1)}[2]))$ and the 4th permutation be $\{c_1^{(1)}[1], c_1^{(1)}[2]\}$ which occurs with probability $((1 - P(e_1^{(1)}[1])) \cdot (1 - P(e_1^{(1)}[2])))$. Then, $P(e_2^{(1)}[1] \,|\, \{c_1^{(1)}[1], e_1^{(1)}[2]\}, \alpha_2^2, \mathbf{\Delta}_2[1])$ and $P(e_2^{(1)}[1] \,|\, \{c_1^{(1)}[1], c_1^{(1)}[2]\}, \alpha_2^2, \mathbf{\Delta}_2[1])$ are obtained in the similar way as the previous permutations. Once the conditional error probability for each permutation is obtained, the error probability that is averaged over the permutations is obtained by (25). Then, substituting (25) into (17) we obtain $P(e_2^{(1)}[1])$. T-SIC will exploit $P(e_2^{(1)}[1])$ at detection steps of later iteration to obtain conditional probabilities, e.g. see the 1st step of the 2nd iteration.

2) The 2nd iteration:
The 1st step is to detect the 1st symbol of the 1st user. Since the 1st symbol of the 2nd user was detected at the 1st iteration, with

respect to (w.r.t.) the T-SIC procedure $P(e_1^{(2)}[1])$ is conditioned on the permutation of that detection result. Then, let the 1st permutation be $\{e_2^{(1)}[1]\}$, which occurs with probability $P(e_2^{(1)}[1])$. From (11), $\left(D_2^{(1)}[1] \,|\, e_2^{(1)}[1]\right) = (^6/_{(M-1)})$ and substitute it into (10), the variance of residual interference is

$$\text{Var}(\tilde{\eta}_1^{(2)}[1] \,|\, e_2^{(1)}[1], \mathbf{\Delta}_1[1]) = \Delta_{1,2}[1,1] \cdot (^6/_{(M-1)}) \cdot p_2/2. \tag{26}$$

Substituting (26) into (12) and (12) into (15) we obtain $P(e_1^{(2)}[1] \,|\, e_2^{(1)}[1], \alpha_1^2, \mathbf{\Delta}_1[1])$. Then, the 2nd permutation is $\{c_2^{(1)}[1]\}$, which occurs with probability $(1 - P(e_2^{(1)}[1]))$. From (11), $\left(D_2^{(1)}[1] \,|\, c_2^{(1)}[1]\right) = 0$ and from (10), $\text{Var}(\tilde{\eta}_1^{(2)}[1] \,|\, c_2^{(1)}[1], \mathbf{\Delta}_1[1]) = 0$. Then by substituting $\text{Var}(\tilde{\eta}_1^{(2)}[1] \,|\, c_2^{(1)}[1], \mathbf{\Delta}_1[1])$ into (12) and (12) into (15) we obtain $P(e_1^{(2)}[1] \,|\, c_2^{(1)}[1], \alpha_1^2, \mathbf{\Delta}_1[1])$. The error probability that is averaged w.r.t. the permutations is given by

$$P(e_1^{(2)}[1] \,|\, \alpha_1^2, \mathbf{\Delta}_1[1]) = \left(P(e_1^{(2)}[1] \,|\, e_2^{(1)}[1], \alpha_1^2, \mathbf{\Delta}_1[1]) \cdot P(e_2^{(1)}[1])\right) + \left(P(e_1^{(2)}[1] \,|\, c_2^{(1)}[1], \alpha_1^2, \mathbf{\Delta}_1[1]) \cdot (1 - P(e_2^{(1)}[1]))\right). \tag{27}$$

Then, substituting (27) into (17) we obtain $P(e_1^{(2)}[1])$. T-SIC will exploit $P(e_1^{(2)}[1])$ at later detection steps.

The 2nd step is to detect the 2nd symbol of the 1st user. From the interfering symbols, only the 1st symbol of the 2nd user was a priori detected at the previous iteration. Therefore according to T-SIC procedure, $P(e_1^{(2)}[2])$ is conditioned on the permutations w.r.t. only the detection result for the 1st symbol of the 2nd user. Then let the 1st permutation be $\{e_2^{(1)}[1]\}$, which occurs with the probability $P(e_2^{(1)}[1])$. From (11), $\left(D_2^{(1)}[1] \,|\, e_2^{(1)}[1]\right) = (^6/_{(M-1)})$ and $\left(D_2^{(0)}[2] \,|\, \mathfrak{L} = 0\right) = 1$ and substituting them into (10), the variance of residual interference is

$$\text{Var}(\tilde{\eta}_1^{(2)}[2] \,|\, \{e_1^{(1)}, \mathfrak{L} = 0\}, \mathbf{\Delta}_1[2]) = \Delta_{1,2}[2,1] \cdot (^6/_{(M-1)}) \cdot p_2/2 + \Delta_{1,2}[2,2] \cdot p_2/2. \tag{28}$$

Substituting (28) into (12) and (12) into (15) we obtain $P(e_1^{(2)}[2] \,|\, \{e_1^{(1)}, \mathfrak{L} = 0\}, \alpha_1^2, \mathbf{\Delta}_1[2])$. Then the 2nd permutation is $\{c_2^{(1)}[1]\}$, which occurs with the probability $(1 - P(e_2^{(1)}[1]))$. From (11), $\left(D_2^{(1)}[1] \,|\, c_2^{(1)}[1]\right) = 0$ and $\left(D_2^{(0)}[2] \,|\, \mathfrak{L} = 0\right) = 1$ and substituting them into (10), the variance of residual interference is

$$\text{Var}(\tilde{\eta}_1^{(2)}[2] \,|\, \{c_1^{(1)}, \mathfrak{L} = 0\}, \mathbf{\Delta}_1[2]) = \Delta_{1,2}[2,2] \cdot p_2/2. \tag{29}$$

$$P(e_2^{(1)}[1]\,|\,\alpha_2^2,\boldsymbol{\Delta}_2[1]) = \Big(P(e_2^{(1)}[1]\,|\,\{e_1^{(1)}[1],e_1^{(1)}[2]\},\alpha_2^2,\boldsymbol{\Delta}_2[1])\cdot P(e_1^{(1)}[1])\cdot P(e_1^{(1)}[2])\Big)$$

$$+ \Big(P(e_2^{(1)}[1]\,|\,\{e_1^{(1)}[1],c_1^{(1)}[2]\},\alpha_2^2,\boldsymbol{\Delta}_2[1])\cdot P(e_1^{(1)}[1])\cdot (1-P(e_1^{(1)}[2]))\Big)$$

$$+ \Big(P(e_2^{(1)}[1]\,|\,\{c_1^{(1)}[1],e_1^{(1)}[2]\},\alpha_2^2,\boldsymbol{\Delta}_2[1])\cdot (1-P(e_1^{(1)}[1]))\cdot P(e_1^{(1)}[2])\Big)$$

$$+ \Big(P(e_2^{(1)}[1]\,|\,\{c_1^{(1)}[1],c_1^{(1)}[2]\},\alpha_2^2,\boldsymbol{\Delta}_2[1])\cdot (1-P(e_1^{(1)}[1]))\cdot (1-P(e_1^{(1)}[2]))\Big) \tag{25}$$

Substituting (29) into (12) and (12) into (15) we obtain $P(e_1^{(2)}[2]\,|\,\{c_1^{(1)},\mathfrak{L}=0\},\alpha_1^2,\boldsymbol{\Delta}_1[2])$. Then the error probability that is averaged w.r.t the permutations is given by

$$P(e_1^{(2)}[2]\,|\,\alpha_1^2,\boldsymbol{\Delta}_1[2]) =$$
$$\Big(P(e_1^{(2)}[2]\,|\,\{e_1^{(1)},\mathfrak{L}=0\},\alpha_1^2,\boldsymbol{\Delta}_1[2])\cdot P(e_2^{(1)}[1])\Big)$$
$$+ \Big(P(e_1^{(2)}[2]\,|\,\{c_1^{(1)},\mathfrak{L}=0\},\alpha_1^2,\boldsymbol{\Delta}_1[2])\cdot (1-P(e_2^{(1)}[1]))\Big). \tag{30}$$

Then, substituting (30) into (17) we obtain $P(e_1^{(2)}[2])$. T-SIC will exploit $P(e_1^{(2)}[2])$ at later detection steps.

The 3rd step is to detect the 1st symbol of the 2nd user. W.r.t. the T-SIC procedure, $P(e_2^{(2)}[1])$ is obtained by following similar calculations at the 3rd step of the 1st iteration but using the most recent information.

The error statistics for further iterations of T-SIC are obtained in the similar way to calculations provided at the 2nd iteration. And obtaining BER performance for T-SIC is straightforward by substituting $P(e_{k^*}^{(l)}[s])$ obtained by above calculations into (18). For Conv-SIC, BER performance is obtained by following above calculations only for detecting sth symbol of users and setting $\big(D_k^{(0)}[\varsigma]\,|\,\mathfrak{L}=0\big)=1$ for adjacent symbols $\varsigma \in \{(s-1),(s+1)\}$ of interferers.

B. Shannon Capacity based Performance Comparison

In this subsection theoretical analysis is provided to compare capacity between Conv-SIC and T-SIC based NOMA systems. Also an uplink OFDMA system is considered for comparison. The theoretical analysis of SC-SIC based NOMA systems is based on so called "onion peeling" or "stripping aided detection" [20], where it is assumed that the interference from a priori detected symbols is perfectly cancelled. Thus, a single iteration of signal processing is applied on the sth IC Triangle and the following capacity expressions are not a function of iterations. Also through the theoretical analysis, it is considered that the transmitted signals and residual interference signal are Gaussian signals ([13], Equation (2) in [5]). The analysis of the spectral efficiency and output SINR of system is given as follows.

1) NOMA Spectral Efficiency: Given the ICed signal by (6), the theoretical model of the desired symbol in terms of the desired signal and interference plus noise is given by

$$\tilde{Y}_{k^*}^{\text{Th}}[s] = X_{k^*}[s]\cdot\sqrt{p_{k^*}}\cdot\alpha_{k^*} + \tilde{\eta}_{k^*}^{\text{Th}}[s] + N_n, \tag{31}$$

where superscript "Th" represents the theoretical signal model and $\tilde{\eta}_{k^*}^{\text{Th}}[s]$ is the theoretical residual interference to the desired

symbol, given by

$$\tilde{\eta}_{k^*}^{\text{Th}}[s] = \sum_{\substack{k\in\Omega\\k<k^*}}\sum_{\varsigma=s-1}^{s+1}(1-I_{\text{asyn}})\cdot\Delta_{k^*,k}[s,\varsigma]\cdot\big(1-\delta(\varsigma-s)\big)$$
$$\cdot\big|X_k[\varsigma]-X_k[s]\big|\cdot\sqrt{p_k}\cdot\alpha_k\cdot e^{j\theta_{k^*,k}}$$
$$+ \sum_{\substack{k\in\Omega\\k>k^*}}\sum_{\varsigma=s-1}^{s+1}\Big(1-\delta\big((\varsigma-s)+1+(1-I_{\text{asyn}})\big)\Big)$$
$$\cdot\Delta_{k^*,k}[s,\varsigma]\cdot X_k[\varsigma]\cdot\sqrt{p_k}\cdot\alpha_k\cdot e^{j\theta_{k^*,k}}, \tag{32}$$

where I_{asyn} is the indicator parameter that shows if employed SIC technique uses a priori information from adjacent symbols $\{(s-1),(s+1)\}$ of interferers, given by $I_{\text{asyn}}\in\{0,1\}$, $I_{\text{asyn}}=1$ for T-SIC and $I_{\text{asyn}}=0$ for Conv-SIC. (32) provides the following insights about capacity performance of IC techniques at asynchronous NOMA. On the RHS of (32), the first term shows that T-SIC technique completely removes interference from stronger users. But, in case of Conv-SIC technique there is residual interference from adjacent symbols $\{(s-1),(s+1)\}$ of stronger users. Residual interference is due to using inaccurate information of the sth symbol to suppress these signals. The second term on the RHS of (32) shows that the T-SIC technique removes interference from the $(s-1)$th symbol of weaker users which are late interferers, i.e. interferers whose time offset is larger than the desired user's. However, Conv-SIC does not exploit the information. Thus, T-SIC significantly improves the SINR and spectral efficiency performances at asynchronous NOMA compared to Conv-SIC techniques [13].

From (31), the SINR of the desired symbol at the NOMA system is given by

$$\gamma_{k^*}^{\mathfrak{N}}[s] = \frac{p_{k^*}\cdot\alpha_{k^*}^2}{\text{Var}(\tilde{\eta}_{k^*}^{\text{Th}}[s]) + \text{Var}(N_n)}, \tag{33}$$

where superscript "$\mathfrak{N}$" denotes NOMA-based system and $\text{Var}(\tilde{\eta}_{k^*}^{\text{Th}}[s])$ is the variance of theoretical residual interference to the desired symbol, given by

$$\text{Var}(\tilde{\eta}_{k^*}^{\text{Th}}[s]) =$$
$$\sum_{\substack{k\in\Omega\\k<k^*}}\sum_{\varsigma=s-1}^{s+1}(1-I_{\text{asyn}})\cdot\Delta_{k^*,k}[s,\varsigma]\cdot\big(1-\delta(\varsigma-s)\big)\cdot p_k$$
$$+ \sum_{\substack{k\in\Omega\\k>k^*}}\sum_{\varsigma=s-1}^{s+1}\Big(1-\delta\big((\varsigma-s)+1+(1-I_{\text{asyn}})\big)\Big)\cdot\Delta_{k^*,k}[s,\varsigma]$$
$$\cdot p_k/2. \tag{34}$$

Based on (34), for T-SIC the variance is caused only by the sth and $(s+1)$th symbols of the weaker users. But for Conv-SIC,

the variance is caused by adjacent symbols $\{(s-1),(s+1)\}$ of stronger users and all overlapping symbols of weaker users. Thus, T-SIC significantly reduces the variance of the residual interference, compared to Conv-SIC.

The spectral efficiency for the desired symbol is given by

$$\zeta_{k^*}^{\mathfrak{N}}[s] = \int_0^\infty \log_2(1+\gamma) \cdot f_\gamma^{\mathfrak{N}}(\gamma)\,d\gamma, \tag{35}$$

where $\gamma = \gamma_{k^*}^{\mathfrak{N}}[s]$ represents the instantaneous SINR for the desired symbol and $f_\gamma^{\mathfrak{N}}(\gamma)$ represents probability density function (pdf) of γ. Defining

$$c_{k^*} = \frac{p_{k^*}}{\mathrm{Var}(\tilde{\eta}_{k^*}^{\mathrm{Th}}[s]) + \mathrm{Var}(N_n)}, \tag{36}$$

that is c_{k^*} is the average SINR, and substituting (36) into (33), $\gamma = c_{k^*} \cdot \alpha_{k^*}^2$. Since $\alpha_{k^*}^2$ follows exponential distribution with unit mean, γ follows an exponential distribution with mean equal to c_{k^*}, given by

$$f_\gamma^{\mathfrak{N}}(\gamma) = \left({}^1/_{c_{k^*}}\right) \cdot e^{-\frac{\gamma}{c_{k^*}}}. \tag{37}$$

Replacing $f_\gamma^{\mathfrak{N}}(\gamma)$ in (35) by (37), the spectral efficiency of the desired symbol at the NOMA system can then be obtained as $\zeta_{k^*}^{\mathfrak{N}}[s] = {}^1/_{c_{k^*}} \int_0^\infty \log_2(1+\gamma) \cdot e^{-\gamma/c_{k^*}}\,d\gamma$. From the expression, it is seen that the spectral efficiency is directly proportional to c_{k^*}. Thus, by providing a higher c_{k^*}, T-SIC can provide superior capacity performance over Conv-SIC. Also note that the spectral efficiency is not a function of iterations due to the assumption of perfect cancellation at Shannon capacity analysis.

2) OFDMA Spectral Efficiency: In OFDMA-based systems, users contend to access a subcarrier. This paper considers proportional fairness scheduler (PFS) for allocation of subcarriers to users in order to guarantee a fair data rate is achieved by each user that is proportional to its transmission distance. Let $\mathrm{Pr}_{\mathrm{access}}(k)$ denote the probability of the subcarrier to be allocated to the kth user and $\alpha_{\mathrm{access}}^2$ denote the squared magnitude of channel fading for the user accessing the subcarrier. It has been shown that with i.i.d. fading channel users, PFS provides the same opportunity for being allocated with a subcarrier to users regardless of their location (i.e. average channel condition) [21]. Therefore, with K users contending for the subcarrier, $\mathrm{Pr}_{\mathrm{access}}(k) = 1/K, \forall k$. Further, the subcarrier allocation only depends on the fading of the channel and the user with the best channel condition will be allocated for transmission over the subcarrier, i.e. user that wins the contention is $k^* = \arg\max_{k\in\Omega}\{\alpha_k^2\}$ and $\alpha_{\mathrm{access}}^2 = \max_{k\in\Omega}\{\alpha_k^2\}$, where $\arg\max$ denotes the argument of maximum. Let the superscript "$\mathfrak{O}$" represent OFDMA-based system and $\gamma_{k^*}^{\mathfrak{O}}[s]$ denote the instantaneous SINR for the sth symbol of the k^*th user when allocated with the subcarrier, given by

$$\gamma_{k^*}^{\mathfrak{O}}[s] = c \cdot \alpha_{\mathrm{access}}^2, \tag{38}$$

where $c = \left({}^{p_{k^*}}/_{\mathrm{Var}(N_n)}\right)$ is the average SNR. The pdf of $\gamma_{k^*}^{\mathfrak{O}}[s]$ is given by,

$$f_\gamma^{\mathfrak{O}}(\gamma) = \frac{K}{c}\left(1 - e^{-\frac{\gamma}{c}}\right)^{K-1} e^{-\frac{\gamma}{c}}. \tag{39}$$

(39) is from the order statistics of maximum of K i.i.d. exponentially distributed r.v.s ([22] Section 8.1). Then, by considering the percentage of time the k^*th user can access the channel and $f_\gamma^{\mathfrak{O}}(\gamma)$, the spectral efficiency for the sth symbol of the k^*th user is given by

$$\zeta_{k^*}^{\mathfrak{O}}[s] = \frac{1}{K}\int_0^\infty \log_2(1+\gamma) \cdot f_\gamma^{\mathfrak{O}}(\gamma)\,d\gamma. \tag{40}$$

Comparing SINR expression of NOMA, given by (33), to OFDMA, given by (38), there is $\mathrm{Var}(\tilde{\eta}_{k^*}^{\mathrm{Th}}[s])$ contributing as an additional noise term that degrade SINR performance of a NOMA user compared to an OFDMA user. The impact of SINR loss on spectral efficiency is in *logarithmic scale*. However, by comparing spectral efficiency expression of NOMA, given by (35), to OFDMA, given by (40), it is clear that NOMA has frequency reuse gain of $(1/K)^{-1}$ times at *linear scale*. The rate of increase of *linear scale* is superior to *logarithmic scale*. Thus, for acceptable amount of residual interference, NOMA can provide much higher spectral efficiency than OFDMA.

V. NUMERICAL RESULTS

In this section, representative numerical results are provided for evaluating BER and capacity performances of the proposed T-SIC based NOMA technology, compared to Conv-SIC based NOMA and OFDMA. The power allocation to users' uplink transmission is considered as a given system parameter. As the initial work for asynchronous NOMA system, we firstly were focused on a single-cell model in order to make analysis simple and do not consider an uplink power control mechanism for the moment. Thus, a single-cell model is considered in figures unless stated otherwise. It is straightforward to apply well-known power allocation techniques [18] and uplink power control mechanisms [15] to the T-SIC technique.

A. BER Results

Figs. 5 - 8 show the BER performance of T-SIC and Conv-SIC techniques for different QAM modulation levels and iterative signal processing with $L = 3$ iterations. Low level, $\{4, 16\}$, to high level, $\{64, 256\}$, QAM are considered. Two users are considered to share a subcarrier at a NOMA based system. Average received power and time offset of users are varied to investigate the performance. The ratio of the average received power of the 1st user to that of the 2nd user, called received power ratio, is changed from 0 to 40 dB.

In Figs. 5 and 6, the time offset difference of users are assumed to be uniformly distributed between $[1-50]\%$ of the symbol time. Fig. 5 shows the comparison of BER at the 3rd iteration for the analysis and systematic simulations, when 16-QAM and hard-decision are employed. It can be seen from the figure that the analysis and systematic simulations have good agreement for both T-SIC and Conv-SIC techniques.

Figs. 6 (a)-(d) compare the BER results obtained from the analyis of T-SIC and Conv-SIC, presented in Section IV-A, where hard-decision and $\{4, 16, 64, 256\}$-QAM are taken, respectively. It can be seen that when Conv-SIC is employed, only the strong user can achieve low BER but the 2nd user
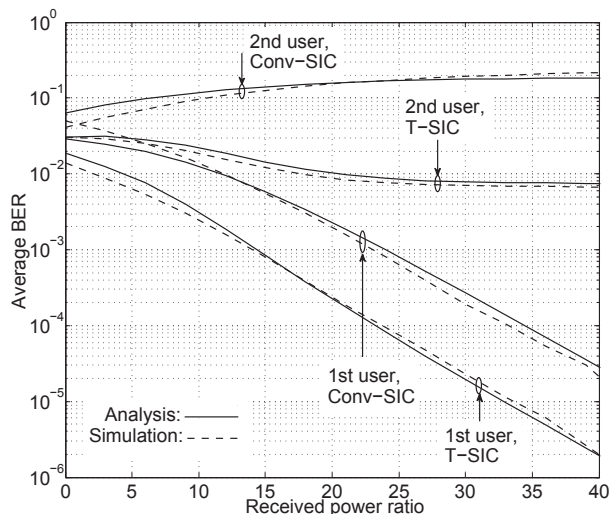
Fig. 5: Analysis versus simulation BER results. 3rd iteration. 16-QAM, 2nd user's received SNR=25dB.

suffers from high BER and cannot have reliable communication even with the lowest modulation. This is due to strong interference from the adjacent symbol of the 1st user. For reliable communications, a NOMA system needs to satisfy BER constraints of all users. It is seen that by suppressing the interference from all overlapping symbols of the co-channel user, T-SIC technique can provide low BER to both users for all considered modulation levels when received power ratio of users becomes large. Thus for reliable NOMA communications, it is crucial to handle the asynchronous overlapping symbols, and users need to have large received power ratio [12]. When the power ratio between users increases, the 1st user's signal become much stronger than the 2nd user's signal and many of the 1st user's symbol estimates are correct with high probability. SIC can suppress interference from these correctly estimated symbols and many of the 2nd user's symbols can be correctly estimated with high probability. It can be also seen that as the modulation level increases, the received power ratio required to achieve a given BER increases. In order to achieve BER $\leq 10^{-3}$ for the 1st user and BER $\leq 10^{-2}$ for the 2nd user, 4-QAM requires 15 dB but 256-QAM requires 28 dB received power ratio between users. This is because energy per bit required for a given BER is high when the modulation level is high, and the 1st user's signal should be much more stronger than the 2nd user's.

Another important observation from Figs. 6 (a)-(d) is the performance gain by iterative signal processing when hard-decision is employed. Figs. 6 (a) and (b) show that for low level modulation, there is no significant performance gain after the 2nd iteration. And in Figs. 6 (c) and (d) for high modulation level, the 3rd iteration provides approximately 2 dB gain when the received power ratio between users is small, but as the received power ratio between users increases the gain of the 3rd iteration decreases. Thus, when hard-decision is employed it is enough to have two iterations to obtain most of the performance gain when the received power ratio of users is relatively large, e.g. larger than 20 dB.

Figs. 7 and 8 show BER performance obtained from systematic simulations of T-SIC technique, when MMSE equalization and soft-decision are employed. Time offset (TO) difference of users, called level of asynchronism, is used as a system parameter to investigate the relationship to BER performance. TO is defined as percentage of the symbol time. For example, TO $= 10\%$ means the TO of the users are different by 10% of the symbol time. Two scenarios of TO between users are considered – little asynchronism, TO $= 10\%$, illustrated in Fig. 7 and high asynchronism, TO $= 35\%$, illustrated in Fig. 8.

From Figs. 7 and 8, it can be seen that when asynchronism is higher the iterative signal processing provides higher performance gain and a lower BER is experienced. This is explained as follows. Referring to the received signal structure in Fig. 1 (b) for the two-user case; with the increase of asynchronism from 10% to 35%, at the 1st IC Triangle there is larger portion of the 1st symbol of the 1st user that is interference free from the 2nd user. This results in a more accurate detection of the symbol. Therefore more interference can be suppressed when detecting the 1st symbol of the 2nd user and this detection is more accurate. Performing multiple iterations on the 1st IC Triangle also improves the accuracy of detection of other symbols. When the processing is shifted to the $(s \geq 2)$th IC Triangle, the sth symbol of the 1st user is initialized to a more accurate value and also its overlapping with the $(s-1)$th symbol of the 2nd user is larger, which is more accurate. These result in symbols of the $(s \geq 2)$th IC Triangle to have more accurate detection. In summary, large asynchronism leads to a start with more accurate detection and then better performances of the subsequent SIC and detection, which results in BER improvement for both users.

The other effect seen from Figs. 8 (a) and (b) is the performance gain by iterative signal processing when soft-decision and MMSE equalization are employed. It is seen from Fig. 8 (a) that when a low modulation level is used, most of the gain is obtained at first two iterations. Specifically the 2nd iteration provides about 16 dB gain over the 1st iteration, but the 3rd iteration provides less than 1 dB gain over the 2nd iteration. This is similar to the hard-decision case shown in Fig. 6 (b), since soft-decision has small performance gain when the modulation level is low. On the other hand when the modulation level increases the gain provided by the 3rd iteration increases. For 256-QAM in Fig. 8 (b), the 2nd iteration provides about 13 dB gain over the 1st iteration and the 3rd iteration provides about 5 dB gain over the 2nd iteration. Thus, soft-detection provides more gain as the modulation level increases and can benefit from a number of iterations that is larger than two.

Fig. 9 shows the effect of varying transmit SNR of the users on the BER performance when the power ratio between two users is fixed. The BER performances of T-SIC at the 3rd iteration for three different power ratios – 15, 25 and 35 dB – are shown in the figure. Assuming $\lambda = 3$, the power ratios correspond to having the path loss for about 3, 7 and 15 meters distance between the two users. Also, the distance between the BS and the 1st user (near user) is assumed to be 5 meters corresponding to about 21 dB path loss. The transmit SNR

(a) 4-QAM. 2nd user's average received SNR=17dB.

(b) 16-QAM. 2nd user's average received SNR=25dB.

(c) 64-QAM. 2nd user's average received SNR=30dB.
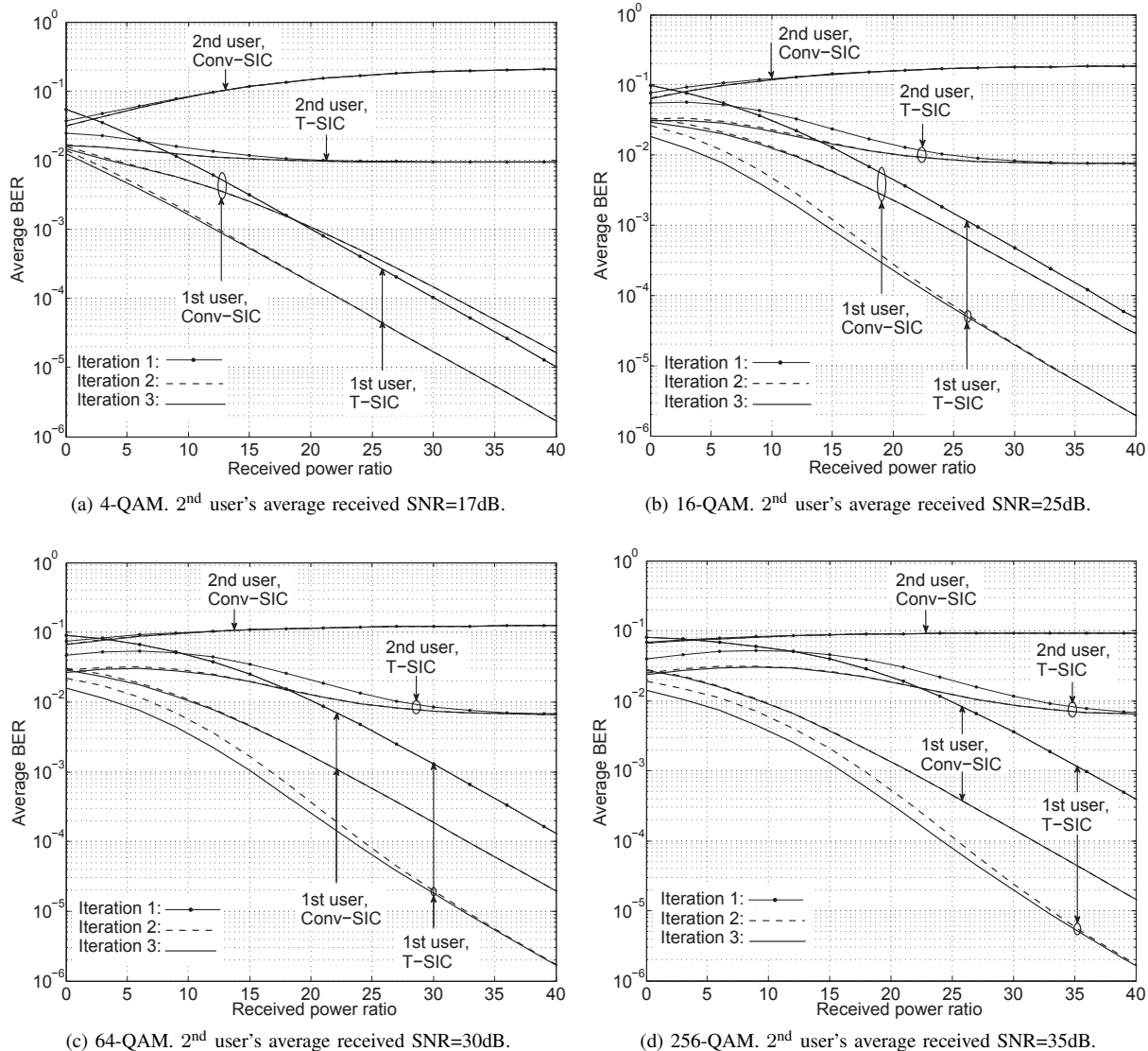
(d) 256-QAM. 2nd user's average received SNR=35dB.

Fig. 6: Average BER performance.

of the two users are assumed to be equal and varied from $60 - 100$ dB to investigate the effect on BER performance. It is seen from the figure that, when the power ratio is 15 dB, both users' BER performance reaches an interference-limited floor at about 75 dB transmit SNR and any further increase in the transmit SNR does not improve the users' BER performance. This is due to the relatively high interference among the co-channel users and the interference dominates the BER performance rather than the noise at this region. In other words, the BER performance becomes interference-limited rather than noise-limited. By increasing the power ratio between the two users, it is seen that the interference-limited floors occur at lower BER values; such as for 25 dB power ratio the floor goes down to about 10^{-3} and 10^{-5} for the 2nd and 1st users, respectively. The interference-limited floor for 35 dB power ratio does not occur within the 100 dB transmit SNR range and the system can achieve BER less than 10^{-3} and 10^{-6} for the 2nd and 1st users, respectively. Therefore, in order

to achieve desired BER constraint of users, it is important to consider the power ratio between users (i.e. interference-limited BER region), as well as the noise-limit.

Fig. 10 shows the effect of inter-cell interference (ICeI) on the BER performance for a NOMA system with two co-channel deployed cells. The interfering user is located at different distances to the cell-edge user of the desired cell. As an interesting 5G system model, it is assumed that all BSs are connected to a central control unit (CU) and fully cooperate to detect users' signals. Thus, a BS shares its knowledge (i.e. estimate) about each detected symbol with other BSs, through the CU. And while detecting a symbol, a BS will use a priori estimates of interfering symbols not only from the intra-cell users but also from inter-cell users. Due to high path loss in dense urban areas, it is assumed that the ICeI is dominated by the mobile user located at cell-edge of the interfering cell [18]. To evaluate the performance, it is assumed that at the desired cell the distance between the BS and 1st user (near user) is
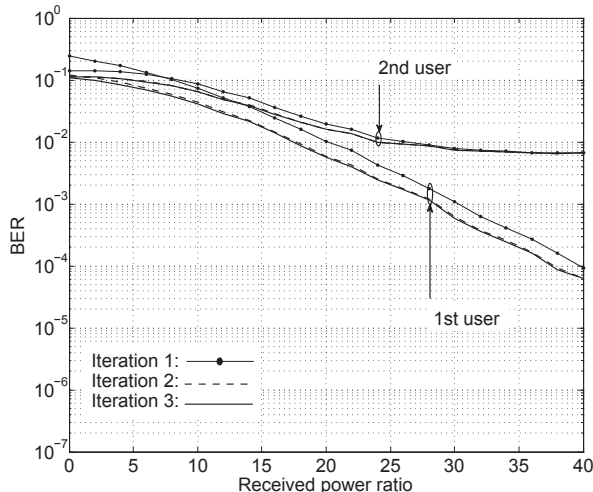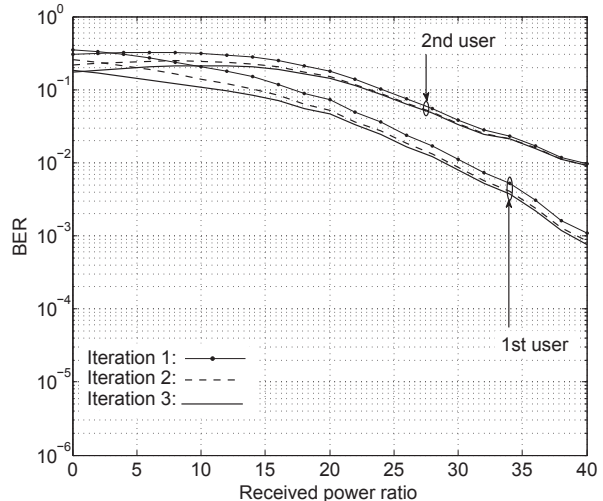
(a) 16-QAM. 2nd user's average received SNR=25dB.

(b) 256-QAM. 2nd user's average received SNR=35dB.

Fig. 7: BER performance for TO $=10\%$.
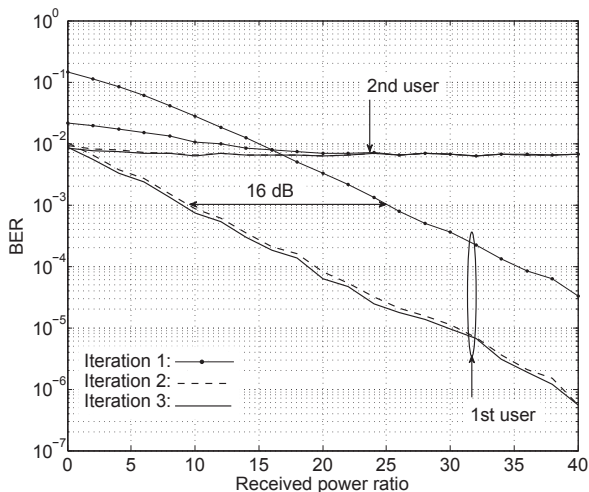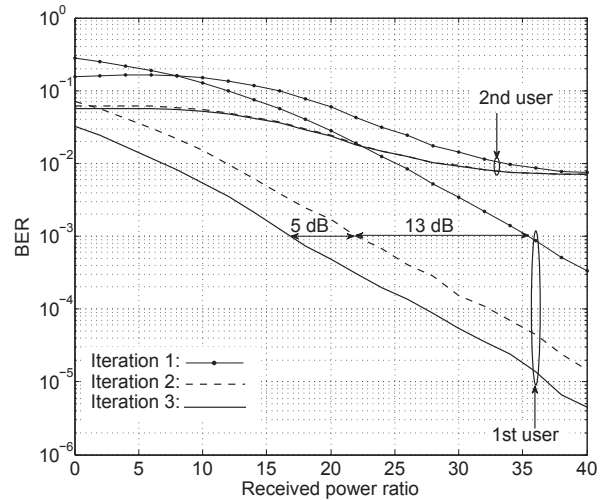

(a) 16-QAM. 2nd user's average received SNR=25dB.

(b) 256-QAM. 2nd user's average received SNR=35dB.

Fig. 8: BER performance for TO $=35\%$.

about 5 meters and the 2nd user (cell-edge user) is located about 15 meters from the near user. The BER performance of T-SIC at the 3rd iteration for five different power ratios of the desired cell's cell-edge user to interfering cell's cell-edge user - 0, 5, 10, 15, 20 dB - are shown in the figure. Also the performance without ICeI is shown as the benchmark. It is seen from the figure that, when the power ratio between two cells' users is small, i.e. cell-edge user of the interfering cell is located close to the desired cell (0, 5, 10 dB power ratio), there is significant performance degradation due to high ICeI. For power ratios greater than 20 dB, the performance degradation is less significant but still considerable. Therefore, it is important to address the ICeI problem, especially at future dense small cell networks, where cell-edge users are closely located.

B. Capacity Results

In this subsection the capacity performance is numerically evaluated based on analysis in Section IV-B. Three users accessing a subcarrier for uplink transmissions is considered. Transmit SNR of users is set to be 50 dB, where the variance of noise is normalized to be unity. Received SNR and time offsets of users are varied to investigate performance. Combinations of considered average received SNR of users are shown in Table I. Time offset triplets (ToT = (τ_1, τ_2, τ_3)) are used to represent the time offset of three users.

Fig. 11 shows the ratio of residual MAI to transmit power for the desired symbol, called ratio of MAI to transmit power, given by $|\mathrm{MAI}/p_{k^*}| = \mathrm{Var}(\tilde{\eta}_{k^*}^{\mathrm{Th}}[s])/p_{k^*}$, where $\mathrm{Var}(\tilde{\eta}_{k^*}^{\mathrm{Th}}[s])$ is given by (34), when Conv-SIC and T-SIC techniques are employed. The cases of little asynchronism, ToT = $(0\%, 5\%, 10\%)$, and high asynchronism, ToT = $(0\%, 35\%, 50\%)$, are considered. For T-SIC it is seen that
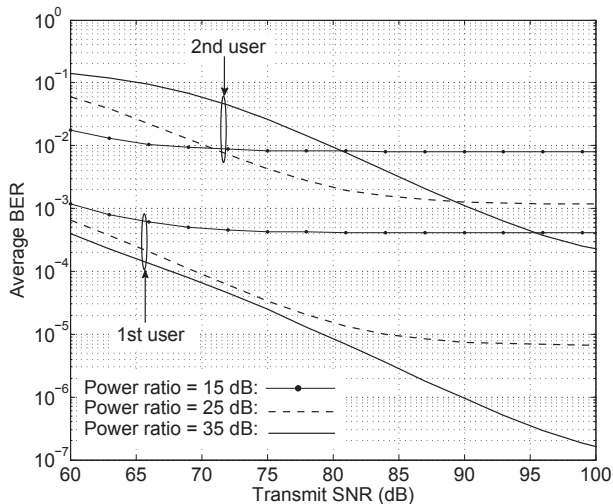
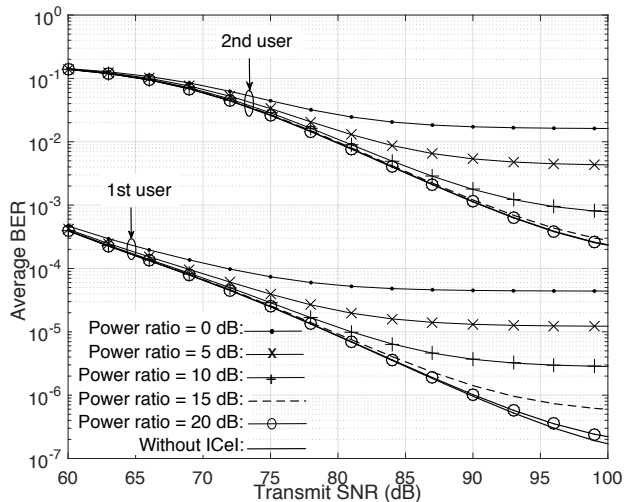Fig. 9: Average BER performance at the 3rd iteration versus transmit SNR. 16 QAM.

Fig. 10: Average BER performance with respect to different received power ratio of the cell-edge users.

$|\text{MAI}/p_{k*}|$ decreases as the power ratio of users increases, since interfering signals are getting weaker. This output result shows that by superimposing users from different distances and/or adjusting transmission power of users, $|\text{MAI}/p_{k*}|$ can be controlled. However for Conv-SIC the 2nd and 3rd users meet a high interference floor due to the residual interference from the adjacent symbol of the stronger users. Note that the value of interference floor depends on the time offset. With little asynchronism in Fig. 11 (a), the interference floor is
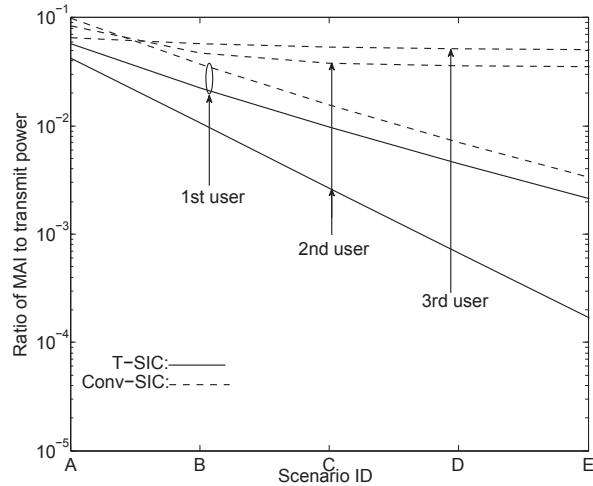
TABLE I: Average received SNR of users.

Scenario ID	Avg. received SNR (dB)			Power ratio (dB)
	1st user	2nd user	3rd user	
A	40	40	40	0
B	40	37	34	3
C	40	34	28	6
D	40	31	22	9
E	40	28	16	12

much lower than high asynchronism in Fig. 11 (b). This is because as the asynchronism increases the overlapping with (i.e. interference from) adjacent symbol of stronger users increases, see $\Delta_{k*,k}[s,\varsigma]$ term at (34). It is also seen that, for the 1st user T-SIC provides much significant performance gain over Conv-SIC at high asynchronism, compared to the performance gain at little asynchronism. This is because the overlapping with the $(s-1)$th symbol of weaker interferers increases. In summary, at asynchronous NOMA transmissions $|\text{MAI}/p_{k*}|$ performance do not only depend on the power ratio of users, but it also strongly depend on the time offset of users.
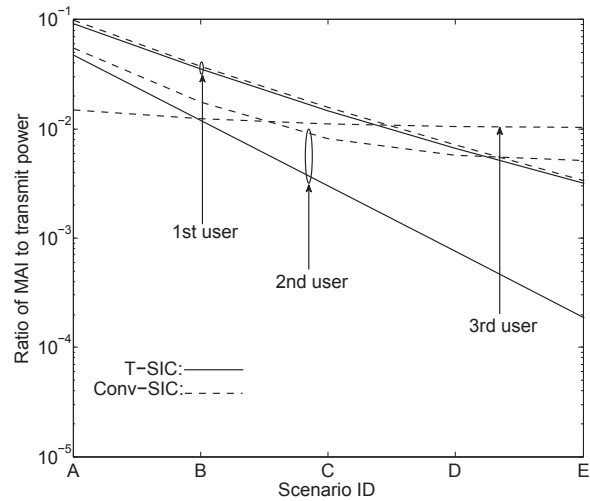
(a) Little asynchronism, ToT = $(0\%, 5\%, 10\%)$.

(b) High asynchronism, ToT = $(0\%, 35\%, 50\%)$.

Fig. 11: $|\text{MAI}/p_{k*}|$ versus different average received SNR ratios (see Table I).

Fig. 12 shows spectral efficiency performance of users, given by (35), when Conv-SIC and T-SIC techniques are employed. Average received SNR of users are assumed to be 30 dB for the 1st user, 18 dB for the 2nd user and 6 dB for the 3rd user. That is, the power ratio between the 1st user and the 2nd user or the 2nd user and the 3rd user is 12 dB. Time offset between users are represented by the x-axis of Fig. 12. That is, for the x-axis value equal to 10, the time

offset difference between the 1st user and the 2nd user or the 2nd user and the 3rd user is 10% of the symbol time. It is seen from the figure that the spectral efficiency of the 1st user increases with increasing asynchronism, when T-SIC is used, since more interference is suppressed from the $(s-1)$th symbol of the 2nd and 3rd users. However, Conv-SIC does not exploit this a priori information and has worse performance than T-SIC, where the performance difference increase with increasing asynchronism. Spectral efficiency of the 2nd user is increasing slightly with increasing asynchronism, when T-SIC is employed, since more interference from the $(s-1)$th symbol of the 3rd user is suppressed. When Conv-SIC is employed, spectral efficiency of the 2nd user decreases rapidly as asynchronism increases. This is due to interference from the 1st user getting stronger. Spectral efficiency of the 3rd user is fixed at approx. 2 bits/symbol, when T-SIC is employed. But when Conv-SIC is employed the 3rd user is not able to have any successful data transmission. These observations show that spectral efficiency performance of NOMA users depends on the time offset of users. Therefore, performance analysis cannot be accurate without considering the time offset of users. Also it is seen that if asynchronism is not handled at NOMA uplink transmissions, only the strongest user can communicate its data and other users need to be turned off.
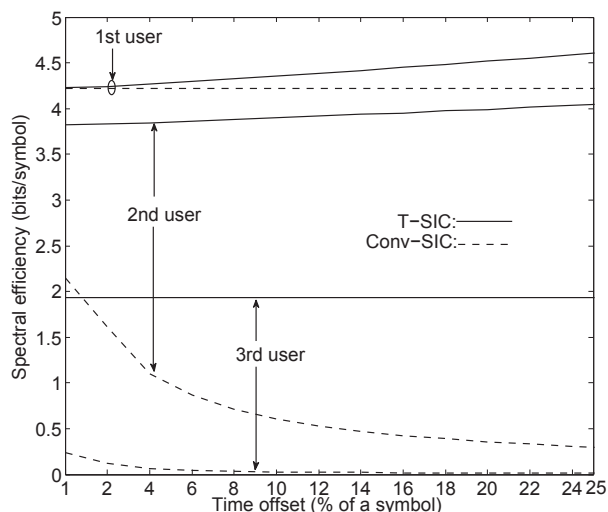

Fig. 12: Spectral efficiency versus different levels of asynchronism of users.

Fig. 13 shows sum spectral efficiency of a subcarrier when OFDMA and NOMA with Conv-SIC and T-SIC techniques are considered for multiple access. Sum spectral efficiency of a subcarrier is given by $\zeta = \sum_{k \in \Omega} \zeta_k$, where ζ_k is given by (40) and (35) for OFDMA and NOMA, respectively. Several scenarios from little asynchronism, ToT = $(0\%, 5\%, 10\%)$, to high asynchronism, ToT = $(0\%, 35\%, 50\%)$, with different power ratio of users, given by Table I, are considered. It is seen from Fig. 13 that performance of OFDMA decreases linearly as average power ratio increases. This is because the 2nd and 3rd users achieve less bits/symbol due to smaller average SNR. When Conv-SIC is employed for NOMA its performance is worse than the OFDMA even with little asynchronism, for all

considered power ratios. This is because of strong interference among SCed users. When the power ratio increases, the performance gradually improves, but still significantly worse than OFDMA. On the contrary when T-SIC is employed, NOMA has superior performance to OFDMA for all the considered scenarios. The performance gain over OFDMA increases with increasing the power ratio. This is because NOMA allows all users to access a subcarrier concurrently. While the 2nd and 3rd users' reduced average received SNR reduces their achievable bits/symbol, it also reduces co-channel interference to other users (as illustrated by Fig. 11) so that other users' achievable bits/symbol is increased. These two contrary effect prevent sum spectral efficiency of a subcarrier to degrade, unlike the case of OFDMA. Therefore NOMA can benefit from near-far effect in wireless communications. Further, performance gain over OFDMA increases with increasing asynchronism since T-SIC can suppress more interference. In summary, NOMA has performance gain over OFDMA when asynchronism is properly addressed.
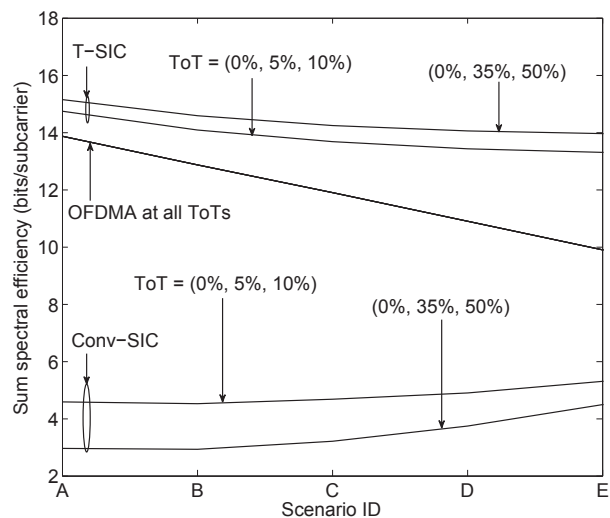

Fig. 13: Sum spectral efficiency verus different user power ratios (see Table I).

Fig. 14 shows the sum spectral efficiency of a subcarrier when T-SIC is employed and a two co-channel deployed cells environment is considered. It is assumed that there are three users with 40, 34 and 28 dB received power (see Scenario C from Table I) at the desired cell. Power ratio of the desired cell's cell-edge user to interfering cell's cell-edge user is set to be 0, 5, 10, 15, 20 and 25 dB to investigate the performance at various scenarios from high to low ICeI. Also the performance without ICeI is shown in the figure as the benchmark. It is seen from the figure that when the power ratio between two cells' users is small there is significant performance degradation due to high ICeI. For the worst case the performance for T-SIC with ICeI is only about 43% of T-SIC without ICeI. As the power ratio between cell-edge users increases, the performance degradation decreases. When the power ratio is 25 dB, the performance of T-SIC with ICeI is about 92% of T-SIC without ICeI. These results are in agreement with results shown in Fig. 10 and show that it is required to address the ICeI problem in

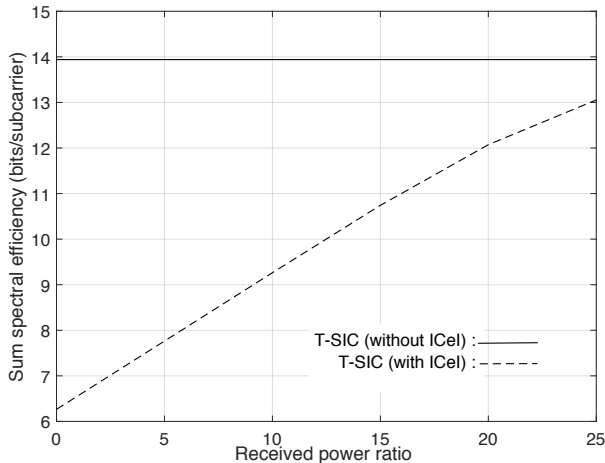order to achieve high capacity at future dense small networks.

Fig. 14: Sum spectral efficiency versus received power ratio of the cell-edge users.

VI. Conclusion and Remarks

NOMA can overcome a major problem of the OFDMA technique which does not allow frequency reuse within one cell. The NOMA is an advantageous technique for future wireless communications. By considering the effect of asynchronism in the NOMA transmissions a novel SIC technique, called T-SIC, was proposed that can benefit from frequency reuse as well as the near-far effect. The BER and capacity performances of the NOMA with Conv-SIC and T-SIC techniques were investigated. In the NOMA transmissions, the proposed T-SIC technique with iterative signal processing provides significant BER performance improvement. The capacity of NOMA and OFDMA were compared. The following conclusions are drawn:

1) Unlike synchronous communications, at uplink transmissions, users' BER and capacity performance strongly depends on the relative time offset between interfering users. If asynchronism is not considered, when the time offset difference between users increases, the interference increases and performance degrades. But when asynchronism is addressed and interference is suppressed by T-SIC, performance can be improved significantly.

2) At iterative signal processing for NOMA, the number of iterations to obtain most of the gain at BER performance depends on the modulation level and the detection method. With hard-decision it is sufficient to have two iterations, however with soft-decision, only for low modulation level, it is enough to have two iterations, and for high modulation level, a larger number of iterations is desirable.

3) NOMA with the proposed T-SIC technique significantly outperforms OFDMA due to frequency reuse.

The results presented for a two co-channel deployed cells environment showed that the ICeI problem need to be carefully addressed. For future multi-cell environments ICeI coordination techniques [4] can be employed together with T-SIC to mitigate ICeI and improve the system performances.

References

[1] N. Abu-Ali, A. Taha, M. Salah, and H. Hassanein, "Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework," *IEEE Communications Surveys and Tutorials*, vol. PP, no. 99, pp. 1–27, 2013.

[2] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - part I: chunk allocation," *IEEE Transactions on Communications*, vol. 57, pp. 2734–2744, September 2009.

[3] H. Zhu and J. Wang, "Chunk-Based Resource Allocation in OFDMA Systems Part II: Joint Chunk, Power and Bit Allocation," *IEEE Transactions on Communications*, vol. 60, pp. 499–509, February 2012.

[4] Y. Endo, Y. Kishiyama, and K. Higuchi, "Uplink non-orthogonal access with MMSE-SIC in the presence of inter-cell interference," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 261–265, Aug 2012.

[5] R. Zhang and L. Hanzo, "A Unified Treatment of Superposition Coding Aided Communications: Theory and Practice," *IEEE Communications Surveys and Tutorials*, vol. 13, pp. 503–520, Third 2011.

[6] J. Schaepperle, "Throughput of a wireless cell using superposition based multiple-access with optimized scheduling," in *2010 IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, pp. 212–217, Sept 2010.

[7] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear MMSE detection," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 995–1004, August 2009.

[8] A. Zafar, M. Shaqfeh, M.-S. Alouini, and H. Alnuweiri, "On Multiple Users Scheduling Using Superposition Coding over Rayleigh Fading Channels," *IEEE Communications Letters*, vol. 17, pp. 733–736, April 2013.

[9] Z. Ding, M. Peng, and H. V. Poor, "Cooperative Non-Orthogonal Multiple Access in 5G Systems," *arXiv:1410.5846s*, October 2014.

[10] Z. Ding, F. Adachi, and H. Poor, "The Application of MIMO to Non-Orthogonal Multiple Access," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2015.

[11] C. Gong, A. Tajer, and X. Wang, "Group Decoding for Multi-Relay Assisted Interference Channels," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 1489–1499, September 2012.

[12] H. Haci, H. Zhu, and J. Wang, "A Novel Interference Cancellation Technique for Non-Orthogonal Multiple Access (NOMA)," in *2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec 2015.

[13] H. Haci and H. Zhu, "Performance of non-orthogonal multiple access with a novel interference cancellation method," in *2015 IEEE International Conference on Communications (ICC)*, pp. 2912–2917, June 2015.

[14] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, June 2013.

[15] 3rd Generation Partnership Project, "TS 36.213 v10.1.0 Release 10: Physical layer procedures for Evolved Universal Terrestrial Radio Access (E-UTRA)," April 2011.

[16] Y. Li and L. G. Stuber, *Orthogonal Frequency Division Multiplexing for Wireless Communications*. New York, NY, USA: Springer, 2006.

[17] H. Haci, *Non-orthogonal Multiple Access (NOMA) with Asynchronous Interference Cancellation*. PhD thesis, School of Engineering and Digital Arts, University of Kent, March 2015.

[18] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

[19] A. Reyni, *Probability Theory*. Amsterdam: North-Holland, 1970.

[20] J. Tong, L. Ping, and X. Ma, "Superposition Coded Modulation With Peak-Power Limitation," *IEEE Transactions on Information Theory*, vol. 55, pp. 2562–2576, June 2009.

[21] E. Liu, Q. Zhang, and K. Leung, "Asymptotic Analysis of Proportionally Fair Scheduling in Rayleigh Fading," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 1764–1775, June 2011.

[22] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

Huseyin Haci holds a B.Sc. degree with High Honours and a M.Sc. degree both in Computer Engineering from Eastern Mediterranean University, Cyprus, and a Ph.D. degree in Electronic Engineering from University of Kent, UK. He is currently an Assistant Professor at the Department of Electrical and Electronic Engineering, Near East University, Cyprus. He received a best paper award and travel grant in IEEE GLOBECOM 2012, USA. He is actively involved in volunteer services at IEEE and other national and international societies. He is currently serving as a TPC member for VTC2017-Spring, Australia. He also served as the Web Chair of IEEE ICC 2015, UK, a TPC member for EUSPN 2015, Germany, and VTC2016-Spring, China, and was a Session Chair for IEEE WCNC 2013, China. His research interests are in the area of wireless communications, covering topics such as signal processing, problem optimization and performance analysis.

Huiling Zhu (M04) received the B.S degree from Xidian University, Xian, China, and the Ph.D. degree from Tsinghua University, Beijing, China. She is currently a Reader (Associate Professor) in the School of Engineering and Digital Arts, University of Kent, Canterbury, United Kingdom. Her research interests are in the area of broadband wireless mobile communications, covering topics such as radio resource management, distributed antenna systems, MIMO, cooperative communications, device to device communications, and small cells and heterogeneous networks. She received the best paper award from IEEE Globecom2011, Houston, USA. She was European Commission Marie Curie Fellowship from 2014 to 2016. She served as the Publication Chair for IEEE WCNC2013, Shanghai, Operation Chair for IEEE ICC2015, London, Symposium Co-Chair for IEEE Globecom2015, San Diego, and Track Co-Chair of IEEE VTC2016-Spring, Nanjing. Currently, she serves as an Editor for IEEE Transactions on Vehicular Technology.

Jiangzhou Wang is currently the Head of School of Engineering and Digital Arts and Professor of Telecommunications, University of Kent, United Kingdom. He has authored over 200 papers in international journals and conferences in the areas of wireless mobile communications and three books. Professor Wang was the Technical Program Chair of the 2013 IEEE WCNC in Shanghai and the Executive Chair of the 2015 IEEE ICC in London. He serves/served as an Editor for a number of international journals. For example, he was an Editor for IEEE Transactions on Communications from 1998 to 2013 and was a Guest Editor for IEEE Journal on Selected Areas in Communications and IEEE Communications Magazine. He is currently an Editor for Science China Information Science. He received the Best Paper Award from the 2012 IEEE GLOBECOM and was an IEEE Distinguished Lecturer from 2013 to 2014. He is an IEEE Fellow and IET Fellow.