Wireless Personal Communications (2022) 123:3755–3770
https://doi.org/10.1007/s11277-021-09312-3

Hybrid NOMA for Future Radio Access: Design, Potentials and Limitations

Kuntal Deka[1] · **Sanjeev Sharma**[2]

Accepted: 20 October 2021 / Published online: 8 November 2021

Abstract
Next-generation internet of things (IoT) applications need numerous low-powered wireless mobile devices to connect with each other, having ultra-reliability and low latency. Non-orthogonal multiple access (NOMA) is a promising technology to address massive connectivity for 5G and beyond by accom- modating several users within the same orthogonal resource block. Therefore, this article explores hybrid NOMA (HNOMA) for massive multiple access in the uplink scenarios due to its higher spectral efficiency. The HNOMA includes both power domain and code domain NOMA method due to diverse channel conditions in practice. We highlight that polar coded based data transmission can achieve higher reliability and lower latency in HNOMA-based wireless networks. Further, at the base station (BS), channel state information (CSI) of each link is not perfectly available or very complex to estimate due to non-orthogonal connections. Therefore, we analyze and review the performance of uplink based system involving HNOMA transmission in imperfect CSI. Furthermore, we summarize some key technical challenges and their potential solutions in futuristic IoT applications using HNOMA transmission. Finally, we offer some design guidelines for HNOMA-based systems using a deep learning approach to implement adaptive and efficient wireless networks.

Keywords Hybrid NOMA · SCMA · DNN · Polar Codes · Imperfect Channel Effect

1 Introduction

Forthcoming society will get highly benefited from using data-driven solutions and methods. In the data-driven environment, numerous of low-power mobile devices will connect to communicate and share their decisions. However, connecting the ultra-large number of devices through a central base station (BS) will be challenging due to limited radio resources with low-latency wireless links. Therefore, non-orthogonal multiple

✉ Kuntal Deka
kuntal@iitgoa.ac.in

✉ Sanjeev Sharma
tc2.sharma@gmail.com

1 IIT Goa, Goa, India

2 IIT BHU: Indian Institute of Technology BHU Varanasi, Varanasi, India

access (NOMA) and hybrid NOMA (HNOMA) transmission techniques can be suitable methods as compared to orthogonal multiple access (OMA) techniques to connect large devices in the internet of things (IoT) scenarios [1–6]. The non-orthogonal transmission techniques enhance spectral efficiency and lower the latency of a wireless system [1, 2, 7]. The HNOMA uses both power domain (PD) and code domain (CD) NOMA for multiple access. Further, HNOMA provides higher spectral efficiency as compared to NOMA and OMA techniques [2] and analyze its potential applications and challenges for 6G networks.

IoT applications can be divided into two main categories as massive IoT in which low power devices continuously send their observation to the cloud, and critical IoT, which includes control signaling, healthcare, and remote manufacturing. Some upcoming applications, wireless sensing, and signaling is essential, like in traffic management, self-driving vehicles. Therefore, in this article, we discuss the massive connectivity of low power devices with ultra-reliability and low-latency based IoT.

1.1 Smart IoT Network

In upcoming 5G and beyond IoT networks, various devices will connect using the virtual cloud networks. For example, smart farming can benefit from the data-driven design by analyzing the soil and atmospheric conditions of the environment for a particular type of crop. A similar technique can be adopted for IoT enabled intra-vehicular network (IVN), where a large number of sensors are connected with each other for sharing the vehicle's status information to develop an innovative vehicular system. Further, traffic management can be significantly improved by using sensing and wireless communication methods to build smart cities. Therefore, the integration of various areas like industries, offices, power grids, farming, healthcare, etc. using IoT platform will make society more conformable, flexible, and sustainable.

A smart IoT network, heterogeneous infrastructures are connected to a central cloud network, as shown in Fig. 1. Therefore, each type of infrastructure like smart grid can be optimized by accessing available data at the central cloud/node. Hence, smart infrastructure will give rise to dense/ultra-dense deployment of access points (AP) in 5G and beyond network.

1.2 Wireless Multiple Access Techniques

To seamlessly connect numerous low-powered IoT devices, we need higher spectral efficiency and availability of a wide frequency spectrum. Currently, millimeter-wave (mmWave) range of the spectrum is explored for wireless communications [8]. High bandwidth at mmWave will help connect many devices; however, it alone can not be sufficient for IoT applications. Currently, NOMA techniques are preferred over OMA to enhance spectral efficiency [9–11]. In general, either PD or CD NOMA methods are considered in the literature [2, 12]. In CD-based NOMA technique includes interleave-division multiple access (IDMA), multi-user shared access (MUSA) technique, pattern division multiple access (PDMA), and sparse code multiple access (SCMA) [2, 5]. In PD NOMA, superposition coding techniques at the transmitter and successive interference cancellation (SIC) at the receiver are used, unlike the OMA. Further, HNOMA,

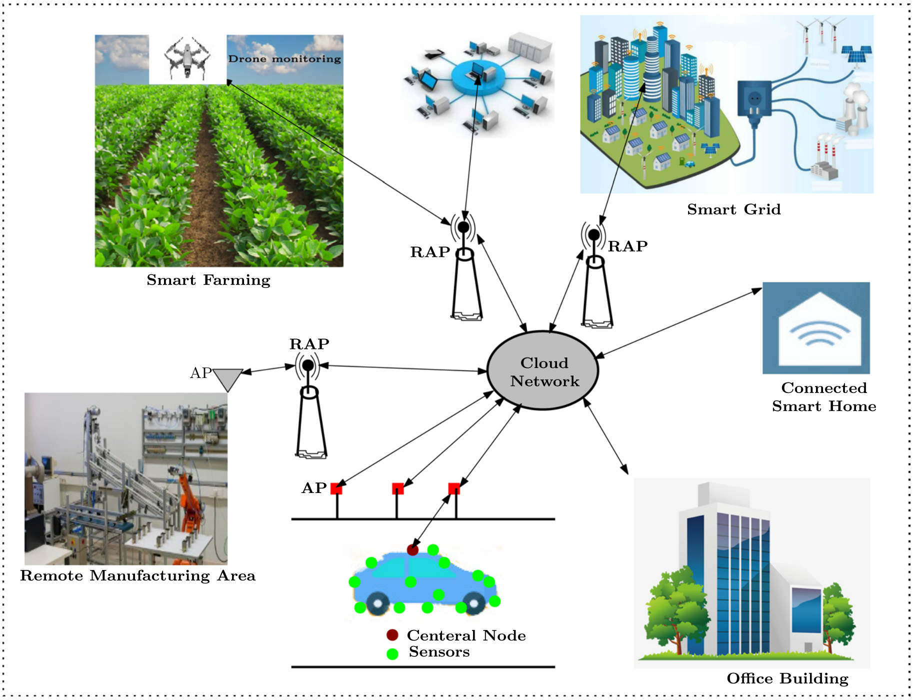


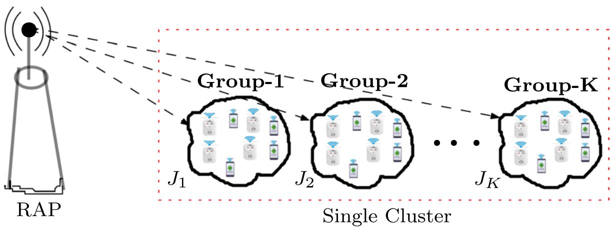
Fig. 1 A view of smart IoT Network. RAP: remote access point, AP: access point

which is an integrated method of PD and CD NOMA, is more suitable and has higher spectral efficiency than conventional PD or CD NOMA. Therefore, in this article, we discuss HNOMA for 5G and beyond wireless networks.

1.2.1 A Simple View of Wireless Connectivity Using HNOMA

Figure 2 shows a pictorial representation of the connectivity of low powered devices to a remote access point (RAP). For easy processing and low latency, the large number of devices are divided into total multiple clusters in a RAP area. Further, each cluster has K multiple groups based on their distance from the RAP, as shown in Fig. 2. Therefore, a device $\mathcal{D}$ in a cluster is accessed using device and group indices, as $\mathcal{D}_{j,k}$, $j = 1, \cdots J_k, k = 1, \cdots, K$, where J_k denotes the total number of users in the kth group. Each cluster has $\mathcal{K}$ orthogonal (time/frequency/code) resources for communications. Further, these $\mathcal{K}$ orthogonal resources may be reused in other clusters in the system. Therefore,

Fig. 2 A sketch for devices connectivity to remote access point using HNOMA in a cluster



all the groups in a cluster use the same $\mathcal{K}$ orthogonal resources for communications, as shown in Fig. 2. However, K groups in a cluster are distinguished using their channel gains at the RAP. Further, devices in each group may have similar or correlated channel impulse response. Therefore, a total of $\sum_{k=1}^{K} J_k$ devices are connected to the RAP using $\mathcal{K}$ orthogonal resources, as depicted in Fig. 2. In the absence of strict power control and sharp frequency cut-off, each cluster may get some interference from other adjacent clusters. However, for simplicity, we have not considered interference from other adjacent clusters in the system.

1.2.2 Example: HNOMA Transmission in V2X Communications

Automotive industries are adopting connected and autonomous vehicles to improve traffic congestion, fuel efficiency, safety, etc. [13]. Therefore, a large number of sensor nodes are required to exchange information. In vehicle-to-everything (V2X) communication, cloud network gather information through several wireless access points such as local area central node, base station, etc. Therefore, to connect numerous low-powered devices, the high spectral efficiency of the network is essential with low latency and efficient signal processing methods. IEEE 802.11p standard is proposed for V2X Communications to support intelligent transportation system (ITS) applications. However, still, technologies for ITS will have to improve for bounded access channel delay and V2X wireless connections. In V2X, grant free HNOMA with polar coding can be used to achieve low latency and high reliability with high spectral efficiency. Further, we summarize the main contributions of the paper as follows:

- We highlight that HNOMA techniques can support a higher number of users as compared to conventional PD-NOMA, SCMA, and OMA-based techniques.
- Data detection of multiple users is carried out using MPA-based SIC and deep learning-based method in an uplink HNOMA-based system.
- The polar-coded HNOMA system's performance is analyzed using various parameters to improve the symbol error rate (SER) performance.
- The effect of imperfect channel estimation on data symbol detection is also highlighted.
- Further, various challenges and their possible solutions in an HNOMA system are also mentioned in the paper.
- We also highlight importance of smart IoT networks and multiple access techniques for 5G and beyond wireless networks.

The rest of the article is organized as follows. Given the potential advantages of HNOMA system, in Sect. 2, we illustrate HNOMA based uplink wireless system for massive connectivity in IoT. Section 3 briefly describes a polar-coded HNOMA-based system. The impact of channel estimation error on HNOMA is explained in Sect. 4. An example of HNOMA system in the uplink scenario is presented in Sect. 5. After that, we highlight a few research challenges in upcoming IoT networks for heterogeneous applications and HNOMA techniques in Sect. 6. Then, we discuss and propose a deep-learning-based detection and estimation technique for HNOMA in Sect. 7. Section 8 concludes this article.

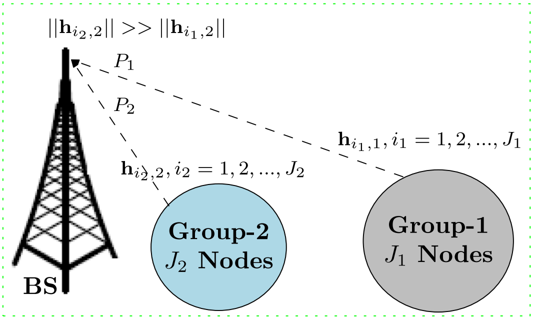


Fig. 3 HNOMA-based system by considering only two near and far groups

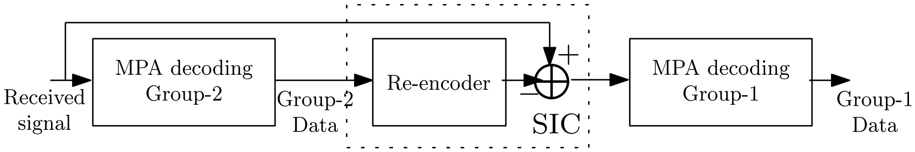


Fig. 4 Detection procedure in HNOMA

2 HNOMA

Recently, HNOMA scheme is proposed for multi-user scenario [2, 14]. The HNOMA scheme using either PD or CD NOMA is more spectrally efficient than either of the conventional PD-NOMA or the OMA-based system [2]. Therefore, in this paper, we consider a joint system comprising of PD and SCMA-based NOMA for HNOMA. HNOMA exploits both power difference and sparsity among the users and orthogonal resources, respectively for transmission. Consider the uplink scenario for HNOMA system, as shown in Fig. 3, where Group-1 and Group-2 have J_1 and J_2 communicating nodes, respectively. By considering Group-1 and Group-2 as the far and the near groups, respectively, we can get the desired variation in strengths of the received signals at the BS. Group-2 (near group) receives higher signal power than Group-1 in the HNOMA system (in Fig. 3), i.e., $P_2 >> P_1$. Therefore, the data for Group-2 is decoded first by considering the Group-1 signal as an interference using massage passing algorithm (MPA). After decoding Group-2, the signal corresponding to Group-2 is subtracted from the received signal at the BS. Then, from the interference-canceled signal, Group-1 symbols are decoded using MPA. Therefore, Group-1 has full diversity, while Group-2 has interference-limited scenarios. Hence, at the BS, both MPA and successive interference cancellation (SIC) based detection is used in HNOMA, as shown in Fig. 4. In general, K groups use K-1 and K times SIC and MPA, respectively to successfully decode all users' data in HNOMA system. The performance of HNOMA depends on the channel gain difference and the power allocation amongst the groups. In HNOMA, multiple groups can communicate by optimizing transmit power amongst the different groups and codebook design within a group. Therefore, K groups can communicate to BS in HNOMA, where each group has $J_k, k = 1, 2, ..., K$ user nodes. The overloading factor λ can be defined as $\lambda = (\sum_{k=1}^{K} J_k)/\mathcal{K}$, where $\mathcal{K}$ is the orthogonal resources in the system.

The HNOMA-based system can achieve higher spectral efficiency in the IoT network by marginally increasing the detection complexity. Since the BS or the central node can support a higher complexity, HNOMA is an attractive approach for uplink scenarios. Further, deep-learning-based information detection can also be used in HNOMA. Deep neural network (DNN) can be trained off-line and is currently being considered for online data detection in wireless communication systems [15]. Furthermore, hybrid non-orthogonal transmission can also include orthogonal and non-orthogonal multiple access schemes for next-generation adaptive networks depending on the nodes' data rate constraints.

The impact of overloading in HNOMA is shown in Fig. 5. In each group, the numbers of user nodes are $J_1 = J_2 = 6, 8$ and orthogonal resources are four, i.e., $\mathcal{K} = 4$ with power allocation $P_1 = 0.2$ and $P_2 = 0.8$. The transmitted signals of the user nodes experience independent flat Rayleigh fading. The codebooks in [2] are used for the user nodes. Observe from Fig. 5 that the performance degrades around 4 dB at SER $= 10^{-3}$ in high overloading scenario ($2 \times 8/4 = 400\%$) as compared to 300% ($2 \times 6/4 = 300\%$) overloading due to higher interference level. However, this range of reliability is acceptable in most IoT applications. Further, the detection performance of the HNOMA-based system can be improved using polar coding techniques.

The HNOMA-based system employs the MPA-based SIC detection, whose complexity increases as the number of groups in a cluster, and the overloading factor increases. To limit the complexity, the devices in the IoT networks can be partitioned into small groups and clusters based on their channel gains, as mentioned in Fig. 2. In HNOMA, the SIC ordering needs to be carefully designed to reduce the error propagation effect [16]. We consider SCMA-based NOMA scheme with polar codes to improve the reliability of a wireless network in the next section.

3 Polar Codes

Erdal Arikan invented a new class of channel codes known as polar codes [17]. These are the first class of codes which can achieve the Shannon capacity in a mathematically provable fashion. These codes are based on the principle of *channel polarization*. Here, with

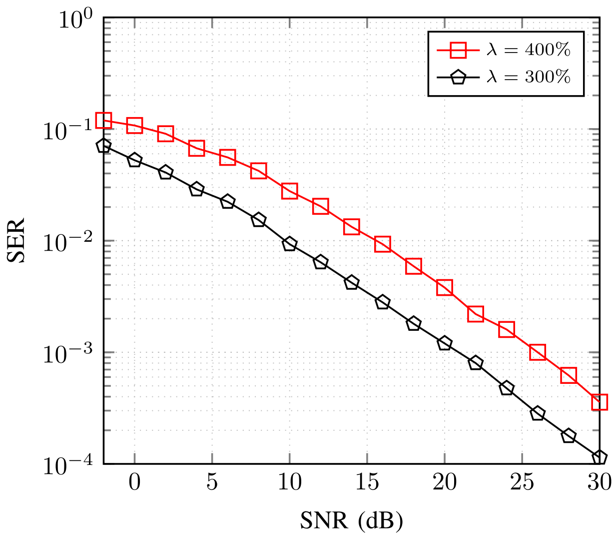


Fig. 5 Impact of overloading factor λ on the HNOMA-based system's SER performance over Rayleigh fading channel

the help of polar encoding, n synthetic bit channels are manufactured from a given set of n identical channels. The newly created bit-channels are polarized in the sense that a fraction of them become noise-free and others become completely noisy channel [17]. The key aspect here is that the fraction of the noiseless channels is equal to the capacity of the original underlying channel. Therefore, a capacity-achieving coding scheme can be devised where the information bits are sent over the noiseless channels and a set of frozen bits (known to the decoder) are sent over the completely noisy channels. The decoding is done with the help of successive cancellation (SC) algorithm. Unlike other modern codes like, Turbo and low-density parity-check (LDPC) codes, the decoding of polar codes is non-iterative. The performance of the polar codes with SC decoding was not impressive at practical block-lengths. Then, Tal and Vardy proposed successive cancellation list decoding for polar codes [18]. It has been found that aided by a small cyclic-redundancy check (CRC) code, the list decoding of polar codes can provide significantly better performance than LDPC and turbo codes when block-length is in the small to medium range.

The performance of an SCMA system can be significantly improved by employing an error correcting scheme. Here we consider the polar codes as they can provide the best performance amongst all the modern codes. The block diagram of a polar-coded SCMA system is shown in Fig. 6. The coding scheme is a concatenated one with a small CRC code of length n_c and polar code of length n. Observe from Fig. 6 that the jth user's data are divided into message frames $\mathbf{u}_j$ of size k. Here we have $k = n \times r - n_c$, where r is the rate of the coding scheme. The frame $\mathbf{c}_j$ at the output of the CRC encoder has length n_c. The frame $\mathbf{c}_j$ is sent to polar encoder as the input. The output of the polar encoder is $\mathbf{x}_j$. Suppose the SCMA mapper involves $M-$ary constellation. Then the bits in the frame $\mathbf{x}_j$ are converted into $n_s = \frac{n}{\log_2 M}$ symbols with each symbol containing $m = \log_2 M$ bits. We have $\mathbf{x}_j = \left(x_j^1, \cdots x_j^{n_s}\right)$. For every symbol slot $s \in \left\{1, 2, \dots, n_s\right\}$, the SCMA encoding (mapping) and detection is performed for all the J users. The SCMA detection is carried out by MPA. From the MPA detector, the soft values i.e. the probability of an user's data being a particular symbol are considered. These symbol-wise probability values are converted to bit-wise probability values which are fed to the individual polar decoder. Every polar decoder employs CRC-aided list decoding algorithm [18, 19]. This decoding scheme offers two improvisations: (1) at any stage of the decoding, L (known as the list size) best decoder estimates are maintained, and (2) at the last stage, out of the L best possible codewords, the codeword satisfying the CRC check condition is considered as the decoder output. If multiple codewords satisfy the CRC check, the decoder estimate is made randomly from them.

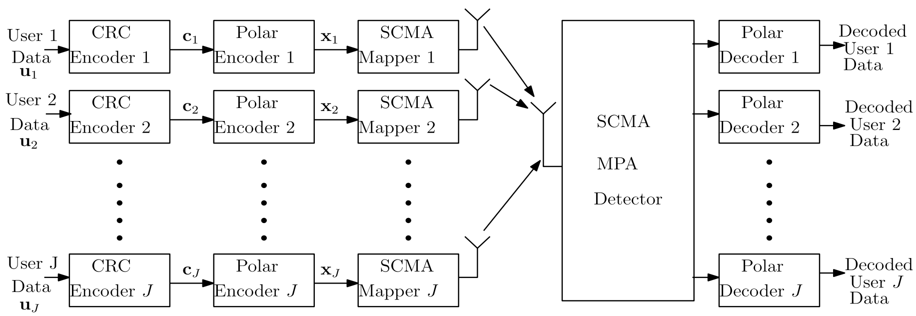


Fig. 6 Polar coded SCMA system for uplink

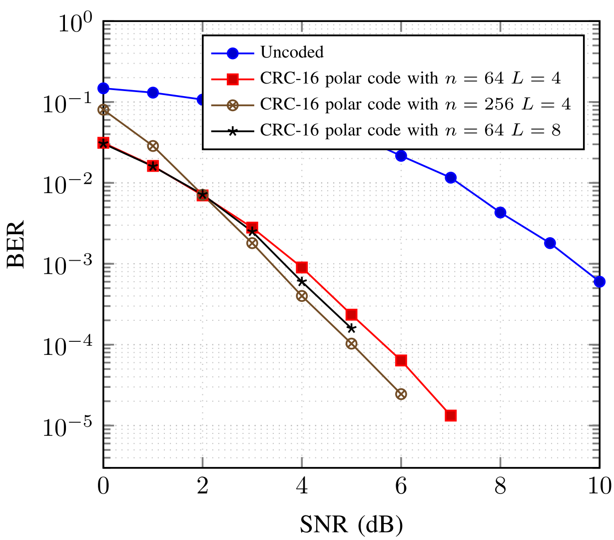
The estimated bits across all symbol duration are accumulated and finally the estimated frame $\hat{\mathbf{u}}_j$ is obtained.

The selection of the frozen bits, known as the construction process, is an essential step for polar coding [20]. Historically, polar codes have been constructed by Bhattacharyya-parameter-based algorithm designed for binary erasure channels (BP-BEC) [17]. The other popular methods are Monte-Carlo-based method, density-evolution-based method [21], Gaussian-approximation (GA) method [22], Tal and Vardy's method [20], etc. The features considered for BP-BEC, Monte-Carlo-based method, density-evolution-based method GA, and Tal and Vardy's method are Bhattacharyya parameter, bit-error-rate (BER) values, probability density function (PDF) of the LLRs, mean of the LLRs, and approximate PDF of the LLRs, respectively. Here, we consider the Monte-Carlo-based method to construct polar codes in the multiple-access channel or uplink system.

Figure 7 shows the BER performance of polar coded 6×4 SCMA system with different block-lengths and the list size L at rate $r = 0.5$. We consider AWGN channel for simulation. In recent communication standards, the short polar codes are considered for error correction [23]. Therefore, we have shown the results for polar codes of block-lengths $n = 64$ and $n = 256$. For decoding, CRC-aided list decoding is used. The CRC code employed is of length 16 with generator polynomial $g(x) = x^{16} + x^{12} + x^5 + 1$. This code is known as CRC-16-CCITT. The list size of $L = 4$ and $L = 8$ are considered. Observe from Fig. 7 that the with the help of short polar code of length $n = 64$, the performance can be drastically improved from that of the uncoded SCMA system. With a slightly higher value of block-length $n = 256$, the performance can be further enhanced as shown in Fig. 7. Moreover, observe that, as the list size increases from $L = 4$ to $L = 8$, the BER reduces marginally.

Figure 8 presents the BER performance of polar coded 6×4 SCMA system for various rates and CRC codes with blocklength $n = 128$ and list size $L = 4$. We consider three rates: $r = \frac{3}{8}, 0.5,$ and $\frac{5}{8}$. Observe from Fig. 8 that as the rate of the polar code increases, the BER increases. Apart from the CRC-16-CCITT code, we also consider a CRC code of length 24 with generator polynomial $g(x) = x^{24} + x^{23} + x^{14} + x^{12} + x^8 + 1$. It can be seen from Fig. 8

Fig. 7 BER performance of various polar coded 6×4 SCMA systems at rate $r = 0.5$



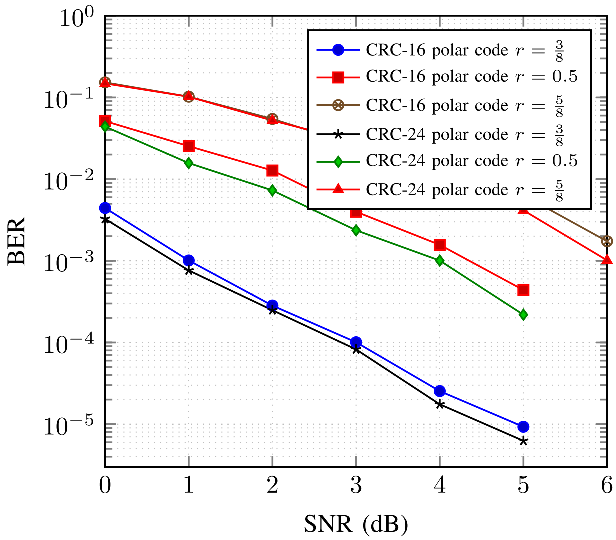


Fig. 8 BER performance of various $n = 128$ polar coded 6×4 SCMA systems with $L = 4$

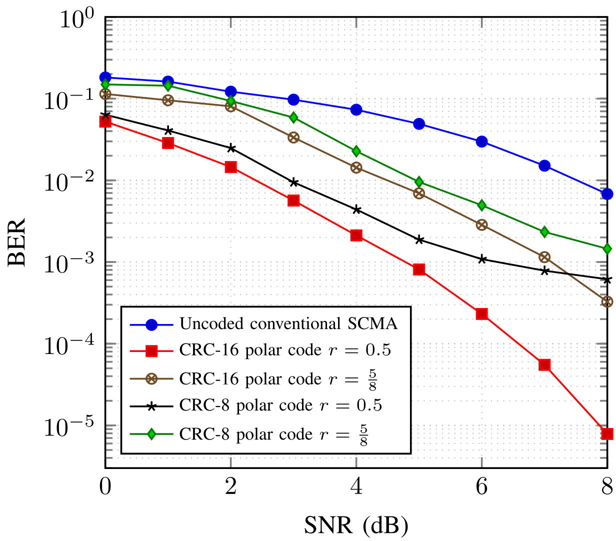


Fig. 9 BER performance of various $n = 64$ polar coded 12×6 SCMA systems with $L = 4$ and overloading factor $\lambda = 200\%$

that CRC-24 provides slightly better performance than CRC-16 due to the increase of the block length of the CRC code.

Finally, Fig. 9 shows the BER performance of the polar coded 12×6 SCMA system for various rates and CRC codes with block length $n = 64$ and list size $L = 4$.

Note that the overloading factor of this system is $\lambda = 200\%$. The SCMA codebooks are designed using the method proposed in [24]. We consider two rates: $r = 0.5$ and $\frac{5}{8}$. Observe from Fig. 9 that, in this case also, the BER increases as the rate increases. The performance of the polar-coded system is impressive in spite of the increase in the overloading factor. We also consider CRC code of length 8 with generator polynomial $g(x) = x^8 + x^7 + x^6 + x^4 + x^2 + 1$. This CRC-8 code is recommended in DVB-S2 standard [25]. The performance of the system with CRC-8 is poorer than that with CRC-16. An early error floor is observed for CRC-8 at the rates considered in the simulation.

4 Imperfect Channel Estimation

The symbol detection at the BS requires perfect channel impulse response (CIR) of each communicating user node. For two groups (near and far), the received signal **r**, without the noise, can be written as

$$\mathbf{r} = \underbrace{\sum_{i=1}^{J_1} \sqrt{P_i^n}\mathrm{diag}\left(\mathbf{h}_i^n\right)\mathbf{x}_i^n}_{\text{Near users}} + \underbrace{\sum_{j=1}^{J_2} \sqrt{P_j^f}\mathrm{diag}\left(\mathbf{h}_j^f\right)\mathbf{x}_j^f}_{\text{Far users}} \quad \in \mathbb{C}^K, \tag{1}$$

where P_i^n and P_j^f denote the powers assigned to the ith near and the jth far user respectively. The total numbers of users in the near and the far group are denoted by J_1 and J_2 respectively. $\mathbf{h}_i^n \in \mathbb{C}^K$ and $\mathbf{h}_j^f \in \mathbb{C}^K$ are the channel gain vectors of the ith and the jth user in the near and the far group, respectively. $\mathbf{x}_i^n \in \mathbb{C}^K$ and $\mathbf{x}_j^f \in \mathbb{C}^K$ denote the SCMA codewords for the ith and the jth user in the near and the far group, respectively. Further, assuming $P_i^n = P^n \; \forall i$ and $P_j^f = P^f \; \forall j$, the received signal is expressed as

$$\mathbf{r} = \sqrt{P^n}\sum_{i=1}^{J_1} \mathbf{H}_i^n\mathbf{x}_i^n + \sqrt{P^f}\sum_{j=1}^{J_2} \mathbf{H}_j^f\mathbf{x}_j^f \quad \in \mathbb{C}^K. \tag{2}$$

Due to many IoT devices with HNOMA schemes, an accurate channel estimation at the BS is highly complex. The BS needs to estimate more parameters in the uplink scenario than in the downlink case. The CIR **H** for a single user at the BS is expressed as $\mathbf{H} = \mathbf{H}_{\mathrm{est}} + \mathbf{E}$, where $\mathbf{H}$, $\mathbf{H}_{\mathrm{est}}$, and $\mathbf{E}$ are the actual CIR, the estimated CIR and the error in channel estimation, respectively. **E** can be modeled as complex Gaussian random vector with zero mean vector and the covariance matrix $\sigma_{\mathrm{e}}^2\mathbf{I}_{\mathcal{K}}$. The variance of the error **E** depends on the used SNR level in estimation, and it is expressed as [26] $\sigma_{\mathrm{e}}^2 = \sigma_{\mathrm{h}}^2/(1 + \rho\mathrm{SNR})$, where $\sigma_{\mathrm{h}}^2\mathbf{I}$ is the covariance matrix of **H** and ρ indicates quality of channel estimation. Further, $\rho = \infty$ represents perfect channel estimation in signal detection. Therefore, the impact of channel estimation should be considered for massive IoT applications in practice [27]. An efficient channel estimation algorithm can also be designed in future for NOMA and HNOMA-based systems.

The greater the significant channel gain differences among the groups in a cluster, the better is the channel estimation in HNOMA. Imperfect channel information results in a residual error in the MPA-based SIC detection for HNOMA, while subtracting the stronger users' signal from the received signal. Therefore, design of spectrally-efficient HNOMA-based systems, channel estimation algorithms, and their effects can be explored in the near future.

5 Example of HNOMA-based System in Uplink Scenario

This section presents an example of HNOMA in IoT system to achieve massive connectivity. To reduce the processing complexity and latency, low-power IoT devices are divided into multiple clusters and each cluster has several groups of devices depending on their channel gains. For a single cluster, the received signal $\mathbf{r}$ at BS is the superposition of each group's signal and complex additive white Gaussian noise, and it is expressed as

$$\mathbf{r} = \sum_{k=1}^{K} \mathbf{r}_{\text{group}-k} + \mathbf{n}, \tag{3}$$

where $\mathbf{r}_{\text{group}-k} = \sum_{j=1}^{J_k} \text{diag}(\mathbf{h}_j)\,\mathbf{x}_j$ and $\mathbf{n}$ denote the kth group's received signal and noise at the BS, respectively. The detection performance in HNOMA depends on the channel estimation error and noise as shown in Fig. 10. We consider channel variance σ_{h}^2=1, overloading factor $\lambda = 300\%$ (two groups and each group has six users), $P_1 = 0.2$, $P_2 = 0.8$ and $\mathcal{K} = 4$ and flat Rayleigh fading channel. The imperfect channel estimation degrades the performance of HNOMA-based system, for example $\rho = 100$, the SER performance degrades by 2dB at SER = 10^{-3}, as observed in Fig. 10. However, some powerful coding techniques and channel estimation algorithms can be adopted in the system to achieve higher reliability in IoT networks.

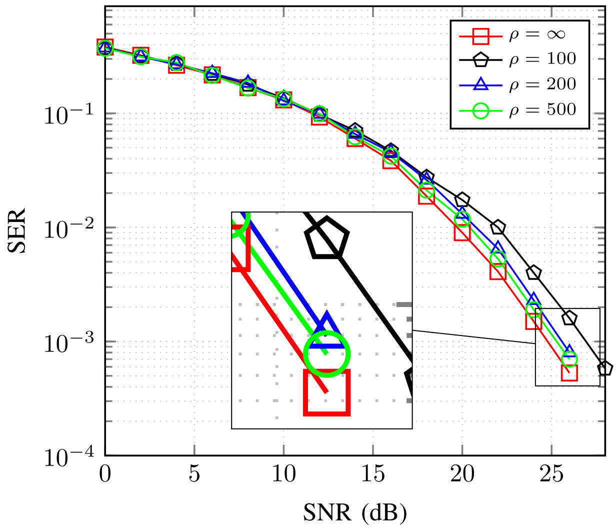


Fig. 10 Impact of imperfect channel estimation on HNOMA's SER performance

6 Research Challenges in Upcoming IoT Networks

6.1 Efficient Energy Enhancement

In next-generation wireless networks, several kinds of BS like micro, femto, and pico will be deployed to support ultra-high density of users. Optimization of energy efficiency will facilitate green future cellular networks infrastructure. In general, using smaller cells, the BS power consumption can be reduced. Further, letting the BS go to sleep mode during light or no traffic load scenarios result in higher energy efficiency in the network. Therefore, dynamic management of users and BS in the wireless network reduces power consumption; however, it is difficult at a large-scale level. Thus, the virtual distribution of resources and their optimization can be a strategic approach to optimize the energy efficiency of 5G and beyond networks. Further, simultaneous wireless information and power transfer (SWIPT) and wireless power transmission can be used in IoT networks.

6.2 Asynchronous Communications

Most of the data in upcoming applications will be short, bursty, and asynchronous. Therefore, we need some signal processing techniques which will shift the paradigm from current systems to asynchronous and non-orthogonal systems. Further, different mobility levels of devices in the massive IoT will pose severe challenges to achieve lower latency. Therefore, detection and estimation must be robust to asynchronous signaling and must possess the capability to deal with uncoordinated interference in next-generation wireless networks. The performance of HNOMA and NOMA-based systems can be analyzed in asynchronous and high mobility scenarios for 5G and beyond networks.

6.3 Optimum Grouping of Devices and Activity Detection

The performance of HNOMA mainly depends on the grouping of devices in a cluster. The grouping of devices can be done using their SNR values at the BS. The devices with similar SNR are included in the same group in a cluster. However, high mobility of devices and fading can make the user paring difficult in the massive networks to achieve higher spectral efficiency. A few sub-optimal iterative methods are proposed for efficient user paring [28–30]. Using a large data set based on users' history, efficient user paring can be achieved by employing a machine learning approach in complex and large wireless networks.

6.4 Robust and Efficient Signal Processing

In some applications like the smart power grid and underwater communications, ambient noise may not be Gaussian in nature i.e., may be impulsive. Therefore, the performance of optimal detection methods (designed for Gaussian noise) can deteriorate significantly. To remove the outliers in large data size using conventional methods like time series before the detection/estimation is a challenging task in massive IoT network due to a higher dimension of the data. Hence, non-linear signal processing techniques can be used to efficiently remove/reduce the impact of the impulse noise before the decision in the IoT network. Deep learning based wireless communication detection and estimation method can also be used to overcome the effect of impulse noise in wireless network. Furthermore, robust detection/estimation methods are required to overcome the impact of channel estimation error and hardware impairments in HNOMA-based IoT applications.

7 Deep Learning Based Detection

Humans' daily living activities will fully integrate to various devices to get a comfortable, sustainable and efficient lifestyle [31]. Upcoming 5G and beyond networks will be based on data-driven and machine learning (ML) approaches due to extremely high data rate with heterogenous devices and applications. ML-based smart networks will be adaptive, efficient, and predictive in nature. Therefore, human intervention will be significantly reduced in smart IoT applications. Moreover, large IoT network of numerous connected devices will be realized using ML techniques.

ML or deep learning will also play a vital role in the health-care system to get an easy medical facility in remote areas and early detection of diseases. Further, seamlessly continuous wireless monitoring and assistance will also benefit old-age people in society. Therefore, ML-based solutions give adaptability, scalability, and flexibility in IoT applications.

Higher spectral and power-efficient communication systems can be accomplished using deep learning-based detection and estimation methods. The availability of large data size will benefit the deep neural network (DNN) for efficient training and parameters optimization. DNN can provide simple and power efficient optimization of large and complex 5G and beyond networks. Conventional approaches may be complex and power inefficient in 6G network. For example, recently deep learning based techniques have been used in massive MIMO (multiple-input and multiple-output), channel estimation, D2D (device-to-device) for symbol detection, power optimization, and localization in wireless systems [8]. Deep learning can also be used in cognitive radio network for primary user's detection and adaptive spectrum management in 6G networks [15].

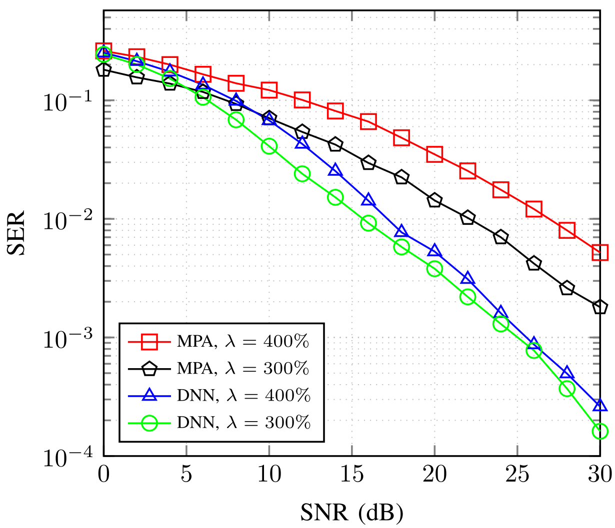


Fig. 11 Comparison between DNN and MPA based detection of HNOMA-based system over Rayleigh fading channel

Personal data security and privacy will be of great concern in next-generation wireless networks. ML-based methods can also aid in achieving security by detecting abnormal behaviors and intrusions in the network before the onset of the attack.

Next, we highlight the importance of DNN for symbol detection in HNOMA. We consider DNN with input and output, and with three hidden layers (with 480, 10, 10 neurons). The training is carried out in off-line mode at SNR = 30 dB. In Fig. 11, the average SER performances of HNOMA using DNN and MPA based detection are shown. DNN outperforms MPA at high SNR due to effective learning of DNN. The performance of MPA is also limited by interference in HNOMA system. A special aspect of DNN-based approach is that it does not require channel information explicitly to decode the users' data symbols, unlike MPA. The training SNR and the number of hidden layers are optimized through multiple trials. Thus, DNN is suitable to decode the information with low real-time complexity in large systems, especially in interference-limited and imperfect channel scenarios.

HNOMA can be considered with massive MIMO technology using conventional and ML-based detection methods to achieve the massive connectivity in low powered IoT networks. However, HNOMA with massive MIMO needs simple and efficient detection methods for deployment in practice. The processing power and the transmitted power in complex wireless systems need to be minimized to fulfill the demands of next-generation green networks. At a time, the number of active devices is very less as compared to the total number of connected devices in the network. Therefore, ML-based active device detection and their data symbol estimation can also be employed in IoT networks for grant-free communication to achieve low latency.

8 Conclusion

We have highlighted the importance of HNOMA to connect low-power devices in upcoming next-generation IoT applications. HNOMA can achieve higher spectral efficiency by serving multiple users at a time in the network. Further, the impact of channel estimation error on the detection performance is highlighted. The inclusion of polar codes can benefit the massive IoT networks to achieve reliability in adverse and channel mismatch scenarios, as highlighted. A few research challenges and their solutions are suggested to fulfill the demand of next-generation IoT networks in this paper. At last, we have advocated for a deep learning-based smart wireless system design for HNOMA. The interference and the channel impact can be minimized in HNOMA/NOMA-based systems using deep learning. Therefore, deep learning and data-driven solutions will play an important role in next- generation wireless networks.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Stoica, R.-A., & de Abreu, G.T.F. (2019). 6G: the wireless communications network for collaborative and AI applications, *arXiv preprint*arXiv:1904.03413.
2. Sharma, S., Deka, K., Bhatia, V., & Gupta, A. (2019). Joint power-domain and SCMA-based NOMA system for downlink in 5G and beyond. *IEEE Communications Letters, 23*(6), 971–974.
3. Nam, Y. Y. (2019). Multiuser activity and data detection via sparsity-blind greedy recovery for uplink grant-free NOMA, *IEEE Communications Letters*, vol. PP, no. PP, pp. 1–4.
4. Liu, Y., Zhang, H., Long, K., Nallanathan, A., & Leung, V. C. M. (2019). Energy-efficient subchannel matching and power allocation in NOMA autonomous driving vehicular networks. *IEEE Wireless Communications, 26*(4), 88–93.
5. Ding, Z., Lei, X., Karagiannidis, G. K., Schober, R., Yuan, J., & Bhargava, V. K. (2017). A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends. *IEEE Journal on Selected Areas in Communications, 35*(10), 2181–2195.
6. Mathur, H., Deepa, T., & A. (2021). *Survey on Advanced Multiple Access Techniques for 5G and Beyond Wireless Communications, Wireless Personal Communications* (pp. 1–18)
7. Reddy, B. S. K., Mannem, K., & Jamal, K. (2021). Software Defined Radio Based Non-orthogonal Multiple Access (NOMA) Systems. *Wireless Personal Communications,* 1–23.
8. Yang, P., Xiao, Y., Xiao, M., & Li, S. (2019). 6G wireless communications: Vision and potential techniques. *IEEE Network, 33*(4), 70–75.
9. Wang, Q., Zhang, R., Yang, L.-L., & Hanzo, L. (2018). Non-orthogonal multiple access: A unified perspective. *IEEE Wireless Communications, 25*(2), 10–16.
10. Wan, D., Wen, M., Ji, F., Yu, H., & Chen, F. (2018). Non-orthogonal multiple access for cooperative communications: Challenges, opportunities, and trends. *IEEE Wireless Communications, 25*(2), 109–117.
11. Yang, K., Yang, N., Ye, N., Jia, M., Gao, Z., & Fan, R. (2018). Non-orthogonal multiple access: achieving sustainable future radio access. *IEEE Communications Magazine, 57*(2), 116–121.
12. Dang, J., Zhang, Z., & Wu, L. (2016). A New Framework for Non-orthogonal Multiple Access Based on Generalized Energy Spreading Transform. *Wireless Personal Communications, 91*(2), 847–860.
13. Liang, L., Ye, H., & Li, G. Y. (2018). Toward intelligent vehicular networks: A machine learning framework. *IEEE Internet of Things Journal, 6*(1), 124–135.
14. Sun, Y., Ding, Z., Dai, X., & Karagiannidis, G. K. (2018). A feasibility study on network NOMA. *IEEE Transactions on Communications, 66*(9), 4303–4317.
15. Jiang, C., Zhang, H., Ren, Y., Han, Z., Chen, K.-C., & Hanzo, L. (2016). Machine learning paradigms for next-generation wireless networks. *IEEE Wireless Communications, 24*(2), 98–105.
16. Tse, D., & Viswanath, P. (2005). *Fundamentals of wireless communication*. Cambridge: Cambridge University Press.
17. Arikan, E. (2009). Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory, 55*(7), 3051–3073.
18. Tal, I., & Vardy, A. (2015). List decoding of polar codes. *IEEE Transactions on Information Theory, 61*(5), 2213–2226.
19. Balatsoukas-Stimming, A., Parizi, M. B., & Burg, A. (2015). LLR-Based Successive Cancellation List Decoding of Polar Codes. *IEEE Transactions on Signal Processing, 63*(19), 5165–5179.
20. Tal, I., & Vardy, A. (2013). How to construct polar codes. *IEEE Transactions on Information Theory, 59*(10), 6562–6582.
21. Mori, R., & Tanaka, T. (2009). Performance of Polar Codes with the Construction using Density Evolution. *IEEE Communications Letters, 13*(7), 519–521.
22. Trifonov, P. (2012). Efficient design and decoding of polar codes. *IEEE Transactions on Communications, 60*(11), 3221–3227.
23. Zhu, Peiying. Polar code for 5G NR, ITW 2018 Keynote, available online: http://itw2018.org/static/resource/Keynote-Peiying.pdf.
24. Deka, K., Priyadarsini, M., Sharma, S., & Beferull-Lozano, B. (2020). Design of SCMA Codebooks using Differential Evolution. In *IEEE International Conference on Communications Workshops (ICC Workshops),2020*, 1–7.
25. Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2).

26. Sharma, S., Deka, K., & Beferull-Lozano, B. (2021). Low-complexity detection for uplink massive MIMO SCMA systems. *IET Communications, 15*(1), 51–59.
27. Mahmoodi, S., Omidi, M. J., Mehbodniya, A., & Adachi, F. (2017). Sparsity Enhancement for Sparse Channel Estimation Using Non-orthogonal Basis. *Wireless Personal Communications, 95*(2), 1759–1779.
28. Nasser, A., Muta, O., Gacanin, H., & Elsabrouty, M. (2021). Joint User Pairing and Power Allocation With Compressive Sensing in NOMA Systems. *IEEE Wireless Communications Letters, 10*(1), 151–155.
29. Wang, J., Li, Y., Ji, C., Sun, Q., Jin, S., Quek, T. Q. S., & Location-Based, M.I.M.O.-N.O.M.A. (2020). Multiple Access Regions and Low-Complexity User Pairing. *IEEE Transactions on Communications, 68*(4), 2293–2307.
30. Zhu, L., Zhang, J., Xiao, Z., Cao, X., & Wu, D. O. (2019). Optimal User Pairing for Downlink Non-Orthogonal Multiple Access (NOMA). *IEEE Wireless Communications Letters, 8*(2), 328–331.
31. Qolomany, B., Al-Fuqaha, A., Gupta, A., Benhaddou, D., Alwajidi, S., Qadir, J., & Fong, A. C. (2019). Leveraging machine learning and big data for smart buildings: A comprehensive survey. *IEEE Access, 7,* 90316–90356.

Kuntal Deka received the B. E., and Ph.D degrees from Assam Engineering College Guwahati and Indian Institute of Technology (IIT) Guwahati in 2008 and 2016, respectively, all in Electrical Engineering. He had worked in the Department of Electronics and Communications Engineering at Indian Institute of Information Technology (IIIT) Guwahati from 2015 to 2018 as an Assistant Professor. Currently, he is serving as an Assistant Professor at School of Electrical Sciences, IIT Goa. His area of expertise are error correcting codes, information theory, multiple access and modulation techniques for modern wireless communication standards. He has published around 20 journals and conferences in reputed publication houses.

Sanjeev Sharma received the B. Tech., M.Tech. and Ph.D degrees from Shri Govindram Seksaria Institute of Technology and Science Indore (SGSITS Indore), Indian Institute of Technology Guwahati, and Indian Institute of Technology Indore in 2008, 2010 and 2018, respectively, all in Electrical Engineering. He had worked in the Department of Electrical and Computer Systems Engineering at Monash University, Melbourne, Australia, as a Postdoctoral Fellow. Currently, he is working as an Assistant Professor at Electronics Engineering, IIT (BHU) Varanasi. His research interests lie in signal processing for wireless communications and networking, mathematical modelling, simulation, design, and analysis of wireless systems. He has published around 40 journals and conferences in wireless communication.