This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3214927

Multi-Type Services Coexistence in Uplink NOMA for Dual-Layer LEO Satellite Constellation

Qifan Hu, Jian Jiao, *Member, IEEE,* Ye Wang, *Member, IEEE,* Shaohua Wu, *Member, IEEE,* Rongxing Lu, *Fellow, IEEE* and Qinyu Zhang, *Senior Member, IEEE*

***Abstract*—The upcoming mega low-earth orbit (LEO) high-throughput satellite constellation is regarded as an emerging paradigm shift in the fifth generation-advance (5GA) networks. In this paper, we propose a multi-type services coexistence handover (MSCH) non-orthogonal multiple access (NOMA) scheme for a dual-layer mega LEO satellite constellation, which can simultaneously and efficiently provide uplink NOMA for three types of 5G user equipments (UEs): mission critical communications (MCC) UEs (CUs), massive machine-type communications (mMTC) UEs (MUs) and enhanced mobile broadband (eMBB) UEs (EUs). The EUs are mainly served in the higher layer satellites for longer service duration and may handover to the lower layer satellites to coexist with CUs or MUs. Moreover, the CUs and MUs perform grant-based (GB) and grant-free (GF) NOMA on resource blocks (RBs) in the lower layer satellites, respectively. Then, we derive the closed-form expressions of three specific key performance indicators (KPIs), i.e., outage probability (OP), system throughput (ST), and ergodic capacity (EC) in the MSCH NOMA scheme, and design five corresponding NOMA algorithms. Simulation results verify the accuracy of our theoretical derivations, and show that the proposed NOMA schemes can achieve better KPI performance than the state-of-art ones.**

***Index Terms*—Mega LEO HTS constellation, uplink NOMA transmission, multi-layer handover, outage probability, ergodic capacity, system throughput**

I. Introduction

Non-terrestrial networks (NTN) are expected to foster the roll out of limited terrestrial fifth generation (5G) network in un-served and underserved areas, and enable 5G-advance (5GA) services at *anywhere and anytime* in cost effective manner [1]. Recently, several giant low earth orbit (LEO) high-throughput satellite (HTS) constellations are planned and begin to launch [2], [3], such as Starlink and OneWeb, both have tens of thousands of LEO HTS distributed in several layers at 300 $\sim$ 1500 km to provide global coverage, and are expected to support these three types of user equipments (UEs) [4], [5]: 1) Mission critical communications (MCC) UE (CU) requires the packet delay budget of 30 ms with packet error rate (PER) less than 10^{-6} [6]. 2) Massive machine-type communications (mMTC) UE (MU) needs massive connectivity for 4 MUs per km^2 [7]. 3) Enhanced mobile broadband (eMBB) UE (EU) requires stationary connectivity with data rate up to 25 Mbit/s, such as video surveillance and vehicle mounted UE [7]. Moreover, considering the limited line of sight (LoS) duration and capacity of a single LEO HTS to the covered ground UEs, the SpaceX Gen2 system proposes an 8-layer LEO HTS constellation at low and very low altitudes [4], which enables the UEs in the same area to view multiple HTSs. Therefore, we assume a dual-layer LEO HTS constellation in this paper, where the EUs are usually served in the higher layer for longer service duration, and the CUs and MUs are served in the lower layer.

On one hand, non-orthogonal multiple access (NOMA) is widely pursued in satellite communications to provide a degree of freedom on power domain by transmitting the superposed signal in the same time-frequency resource block (RB) [8]–[10], which can enhance several key performance indicators (KPI), such as ergodic capacity (EC), outage probability (OP) and system throughput (ST) [11]–[14]: 1) Multiple CUs can simultaneously uplink to HTS via NOMA to reduce the transmission phases [15], since the worst two-way propagation latency is expected to be 26 ms for LEO at 600 km [6]. Note that the PER in the quasi-static fading channel is quickly converged to the OP in finite block-length regime [16], i.e., with 10^3 bits packet length, we utilize OP to evaluate the NOMA performance of CUs. 2) Obviously, NOMA is viewed as a potential enabler of mMTC for delay tolerant MUs [17], and the ST can quantitatively depict the successfully accessed MUs. Thus, we can maximize ST under an appropriate OP threshold for the MUs [7], [18]. 3) Further, EUs usually only have the statistical channel state information (CSI) in satellite communications, and the design objective in NOMA scheme for EUs should be maximum EC rather than the actual data rate [8], [11], [14].

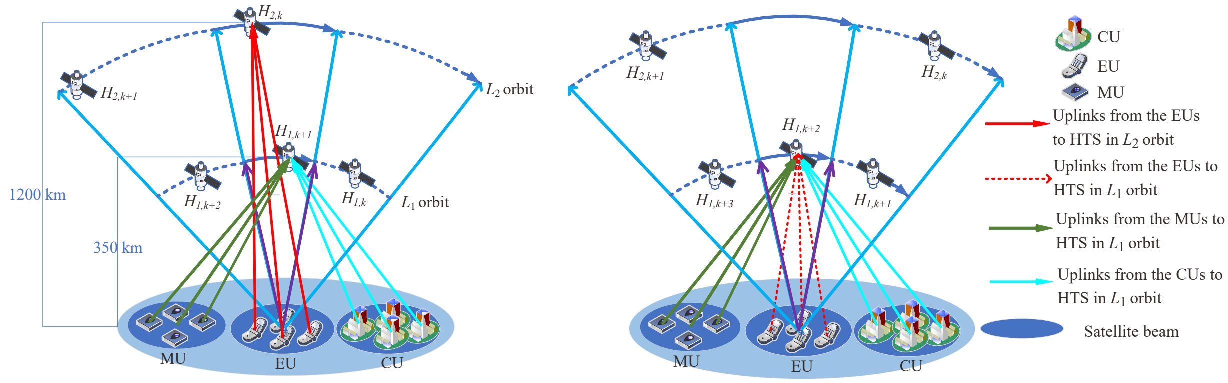
On the other hand, there is still lack of work on the multi-type services coexistence NOMA scheme for satellite communications. Recently, Ding et al. in [19] have proposed the semi-grant free (SGF) NOMA scheme for the grant-based (GB) and grant free (GF) UEs coexistence in terrestrial networks, where several GF UEs can perform power-domain NOMA with one GB UE on the same RB to enhance the system

Manuscript received xxx, 2022. This work was supported in part by the National Natural Sciences Foundation of China (NSFC) under Grant 62071141, Grant 61871147, Grant 61831008, and Grant 62027802. in part by the Shenzhen Basic Research Program under Grant GXWD20201230155427003-20200822165138001, in part by the Natural Science Foundation of Guangdong Province under Grant 2020A1515010505, in part by the Guangdong Science and Technology Planning Project under Grant 2018B030322004, and in part by the Major Key Project of PCL under Grant PCL2021A03-1. (*Corresponding author: Jian Jiao.*)

Q. Hu, J. Jiao, S. Wu, and Q. Zhang are with the Communication Engineering Research Centre, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China, and also with Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: 20S152060@stu.hit.edu.cn; jiaojian@hit.edu.cn; hitwush@hit.edu.cn; zqy@hit.edu.cn).

Y. Wang is with Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: wangy02@pcl.ac.cn).

R. Lu is with the Faculty of Computer Science, University of New Brunswick, Fredericton, NB E3B 5A3, Canada, (e-mail: rlu1@unb.ca).



(a) Scenario of EU access to the HTS in L_2

(b) Scenario of EU access to the HTS in L_1

Fig. 1. System model of a dual-layer LEO HTS constellation.

spectral efficiency without deteriorating the OP performance of GB UEs. Moreover, the different channel gains from different received uplink signals lead to a challenge to analyze the uplink NOMA [13], and the imperfect successive interference cancellation (SIC) decoding uplink NOMA schemes for satellite communications are not analyzed yet.

Furthermore, in order to provide broadband access capabilities, HTS adopts the millimeter wave (mmWave) frequency band, such as Ka, V, Q bands [20], [21]. However, the performance of the satellite-to-ground channel in the mmWave band would be seriously affected by rainfall attenuation and the surrounding environment of UEs [22], [23]. To address the characteristics of mmWave band channels, we utilize the widely used log-normal distribution to model our satellite-to-ground channel [24]–[27], which take two log-normal factors into account for the rainfall attenuation [28]. Moreover, although the EC and OP in log-normal channel are analyzed in OMA systems, there is still lack of derivation for the KPIs in uplink NOMA [23].

Therefore, we propose a multi-type services coexistence handover (MSCH) NOMA scheme in the dual-layer LEO HTS constellation, and the main contributions of this paper are outlined as follows.

- First, we propose the MSCH NOMA scheme to guarantee the KPI requirements of multi-type UEs coexistence in the dual-layer LEO HTS constellation. Inspired by the SGF NOMA, we divide the system bandwidth of lower layer satellites into RBs for CUs and MUs. Specifically, CUs select the RBs as reserved RBs (RRBs) to perform GB NOMA access to satisfy the critical access requirement, and MUs can perform GF NOMA access on the rest RBs as non-reserved RBs (NRBs). Further, the EUs can handover between the two-layer satellites to perform broadband access, thus they can perform NOMA on the higher layer satellites, and joint NOMA access with the CU or MU in the lower layer satellites, respectively.
- Then, considering the different KPI requirements of multi-type UEs and the impact of imperfect SIC, we derive the closed-form expressions of three corresponding KPIs, i.e., OP, ST, and EC in the MSCH NOMA scheme over the log-normal rain attenuation channel. Further, we propose three uplink NOMA algorithms for each type of UEs when EUs perform access to the higher layer satellite, where the "improve ergodic capacity" (IEC) NOMA algorithm can maximum the EC of EUs under an appropriate OP threshold for each EU, and the "improve outage probability" (IOP) NOMA algorithm can minimize OP of each CU, and the "improve system throughput" (IST) NOMA algorithm can maximum the ST of MUs under an appropriate OP threshold for each MU.
- Moreover, when the EUs handover to the lower layer satellites to coexist with CUs or MUs, we propose two uplink NOMA algorithms for the multi-type UEs coexistence, named "EU and CU coexist" (ECco) and "EU and MU coexist" (EMco) NOMA algorithms, which both can maximize the EC of EUs under an appropriate OP threshold for CUs or MUs, respectively. In addition, we propose a "RB selection" (RBS) algorithm for EUs performing the IEC, ECco or EMco NOMA algorithms on empty RB (eRB), RRB and NRB, respectively, which can jointly achieve the maximum EC of EUs in the lower layer satellites. Finally, simulation results validate the accuracy of our theoretical derivations, and show the superior performance compared with the related "discrete uplink power control" (DUPC) NOMA algorithm [29] and "maximize energy efficiency" (MEE) NOMA algorithm [30].

The rest of the paper is outlined as follows. Section II introduces the dual-layer LEO HTS constellation, the handover transmission scheme, channel model, and the MSCH NOMA scheme. In Section III, we derive the closed-form expressions of the OP, EC and ST, respectively, and then design the NOMA algorithms to meet the different KPI requirements of different types of UEs in Section IV. Section V provides the simulation results united with corresponding analysis. Section VI draws a conclusion.

II. System Model and MSCH NOMA

We assume a dual-layer LEO HTS constellation providing KPI-guaranteed uplink NOMA for the CUs, MUs and EUs as

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3214927

TABLE I
LIST OF ABBREVIATIONS

Abbreviation	Full name
LEO	Low-earth orbit
HTS	High-throughput satellite
5GA	Fifth generation-advance
MSCH	Multi-type service coexistence handover
NOMA	Non-orthogonal multiple access
UE	User equipment
MCC	Mission critical communication
CU	Mission critical communication user equipment
mMTC	Massive machine-type communication
MU	Massive machine-type communication user equipment
eMBB	Enhanced mobile broadband
EU	Enhanced mobile broadband user equipment
GF	Grant-free
SGF	Semi-grant free
GB	Grant-based
GA	Grant-based access
KPI	Key performance indicator
OP	Outage probability
ST	System throughput
EC	Ergodic capacity
NTN	Non-terrestrial network
PER	Packet error rate
LoS	Line-of-sight
RB	Resource block
RRB	Reserved resource block
NRB	Non-reserved resource block
IEC	Improve ergodic capacity
IOP	Improve outage probability
IST	Improve system throughput
ECco	EU and CU coexist
EMco	EU and MU coexist
RBS	RB selection
eRB	Empty resource block
DUPC	Discrete uplink power control
MEE	maximize energy efficiency
PDF	probability density function
SINR	Signal-to-interference and noise ratio
MUMS	Multi-user multi symbol
mmWave	Millimeter wave
SIC	Successive interference cancellation
CSI	Channel state information
AWGN	Additive white Gaussian noise
RAP	Random access preamble
RAM	Resource acquisition message
CID	CU identifier
PI	Power information
RAR	Random access response
RRC	Radio resource control
SSS	Switching service signal
RAS	Resource allocation signal
EID	EU identifier
CDF	Cumulative distribution function
RA	Random access

shown in Fig. 1 [4], [5] [1]. The area sandwiched by the purple arrows in Fig. 1 is the actual communication range of the UEs for high-quality access, while the area sandwiched by the blue arrows on the leftmost and rightmost sides is the theoretical LoS range of the ground UEs. There are multiple HTSs in L_1 and L_2, where the LEO HTSs in L_1 layer are with a lower altitude of 350 km, and the LEO HTSs in L_2 layer are about 1200 km [2], and the number of HTSs in L_1 is large enough to form a seamless coverage for the service area, while the UEs in this coverage area only access one HTS in its LoS duration

[1]Different types of UEs are grouped as shown in Fig. 1, however, these UEs are coexisting in the same area in practical.

TABLE II
PARAMETERS DESCRIPTIONS

Notation	Definition
L_b	The b-th layer of HTS
t_b	The LoS duration of the HTS in b-th layer
Δ_t	The time cost for handover of HTS in one layer to another
x_b	The time cost for handover of two adjacent HTSs in L_b
$H_{a,b}$	The a-th HTS in the b-th layer
S_i	The i-th frame in the L_1-layer HTS
$S_{i,j}$	The j-th subframe of the i-th frame in the HTS
B	The total system bandwidth of HTS
D	The number of RB in each HTS
N_c	The number of CU accessing the RB in L_1 layer HTS
N_m	The number of MU accessing the RB in L_1 layer HTS
N_e	The number of EU accessing the RB in L_1 layer HTS
P	The maximum transmit power of each UE
α_k	The power control coefficient of UE$_k$, where $\alpha_k \in [0,1]$
β_k	The residual coefficient, where $\beta_k \in [0,1]$. Specifically, $\beta_k = 0$ means perfect SIC and $\beta_k = 1$ means SIC is failed
h	The channel gain from UE to HTS
l	The beam gain factor
σ^2	The variance of additive white Gaussian noise
γ	The receiving SINR of UE at HTS

[3]. Moreover, we assume that the Doppler shifts caused by the high dynamic LEO HTSs are identical for different UEs in the same coverage area due to the high altitude orbit of HTSs [31], and set a guard bandwidth as double than the Doppler shifts to relieve the influence of Doppler shifts on the system [32]. In addition, the MUs and CUs only access the L_1 layer HTS due to the low propagation latency and their short packet communications [2]. For convenience, the related notations are summarized in Table II.

A. Handover Transmission Scheme for EUs

Without loss of generality, we assume that EUs perform long-term stationary connectivity (e.g., video surveillance, vehicular mounted and fix installation services [7]). Note that the LoS duration of HTS in L_2 is longer than that of L_1. Therefore, the EUs usually prefer access to the HTS in L_2 for longer LoS duration to reduce the frequent handovers as shown in Fig. 2.

Further, we assume that the LoS of HTSs at L_2 and L_1 are t_2 and t_1, respectively, and the time cost for handover of two adjacent HTSs in L_1 and L_2 are x_1 and x_2, respectively. Moreover, the time cost for handover from an HTS in L_2 to an HTS in L_1 is Δ_t,as well as from an HTS in L_1 to an HTS in L_2.

In addition, let R_1 and R_2 denote the EC of EUs to perform access to HTSs in L_1 and L_2[2], respectively. Therefore, the handover transmission scheme for EUs has three options due to the system parameters as shown in Fig. 2.

[2]R_1 and R_2 are related to the system parameters and the proposed IEC, ECco and EMco NOMA algorithms with RBS algorithm at L_1, and the proposed IEC NOMA algorithm at L_2, respectively.

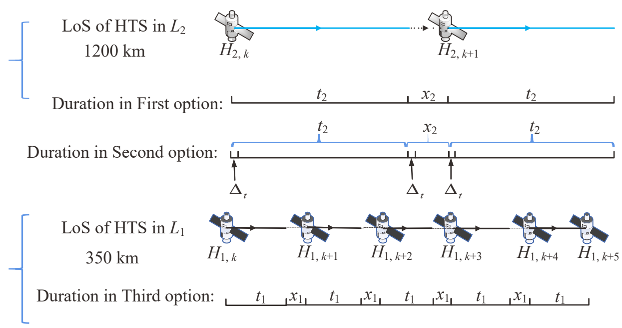


Fig. 2. Schematic diagram of EUs perform handovers between HTSs in L_1 and L_2.

The first option is the EUs only perform handover to the adjacent HTSs in L_2 if $x_2 \leq 2\Delta_t$, and the long-term EC, which is the average EC of the UEs during a period of uplink transmission, which can be expressed as

$$E_1 = \frac{t_2 R_2}{t_2 + x_2}. \tag{1}$$

If $x_2 > 2\Delta_t$, the second option is the EUs can handover to an HTS at L_1 with Δ_t when the LoS of current HTS $H_{2,k}$ at L_2 is finished, and the EUs would handover back to the $H_{2,k+1}$ after x_2 at L_1. Thus, the long-term EC of EUs can be expressed as

$$\begin{aligned} \frac{(t_2 - \Delta_t)R_2 + (x_2 - \Delta_t - nx_1)R_1}{t_2 + x_2} &\leq E_2 \\ \leq \frac{(t_2 - \Delta_t)R_2 + (x_2 - \Delta_t)R_1}{t_2 + x_2}&. \end{aligned} \tag{2}$$

where n is the largest number of HTSs at L_1 that the EUs need to perform handover within x_2.

The third option is the EUs only utilize the HTSs at L_1 for uplink transmission, which may experience frequent handovers during the transmission due to the short LoS duration of HTS in L_1, and the long-term EC can be expressed as

$$E_3 = \frac{t_1 R_1}{t_1 + x_1}. \tag{3}$$

Therefore, the EUs can select an appropriate option to perform handover according to the above long-term EC performance, which is determined by the relationship of x_2 and $2\Delta_t$ in the dual-layer LEO HTS constellation.

B. Channel Model

We utilize the widely used log-normal distribution to model the mmWave channel in our dual-layer LEO HTS constellation [23]. The probability density function (PDF) of the channel gain $|h_{k,a,b}|^2$ from the k-th terrestrial UE U_k to the a-th HTS in the b-th layer $H_{a,b}$ is given by

$$f_{|h_{k,a,b}|^2}(x) = \frac{\varepsilon_{k,a,b}^{\mathcal{M}_{k,a,b}}}{\Gamma\left(\mathcal{M}_{k,a,b}\right)} x^{\mathcal{M}_{k,a,b}-1} \exp\left(-\varepsilon_{k,a,b} \cdot x\right), \tag{4}$$

where $\Gamma(\cdot)$ is the Gamma function, $\varepsilon_{k,a,b} = \mathcal{M}_{k,a,b}/\Omega_{k,a,b}$, and $\mathcal{M}_{k,a,b} = 1/\left(\exp\left(\sigma_{k,a,b}\right) - 1\right)$ is a measure of the fading severity, and $\Omega_{k,a,b} = q_{k,a,b}\sqrt{\frac{\mathcal{M}_{k,a,b}+1}{\mathcal{M}_{k,a,b}}}$ represents the average power of each link, where $q_{k,a,b} = \exp\left(\mu_{k,a,b}\right)$ is the constant area average power, and $\mu_{k,a,b}$ and $\sigma_{k,a,b}$ represent the log-normal location and scale parameters, respectively. Without loss of generality, assume that the HTSs in the same layer have the same SIC performance, then we can ignore the subscript a in the following.

C. Signal Model

Assume that the i-th RB can accommodate at most Z_i UEs by employing NOMA [19], and each UE can only access one RB. Without loss of generality, the signals of Z_i UEs in the i-th RB are arranged in a descending order according to their channel gain $h_{k,b}$ and gain factor $l_{k,b}$ between U_k and H_b as follows

$$|h_{1,b} l_{1,b}| \geq ... \geq |h_{Z_i,b} l_{Z_i,b}|, \tag{5}$$

where $l_{k,b} = \sqrt{G_b \cdot G_k / F_{k,b}}$, and G_b and G_k are the antenna gain at H_b and U_k, respectively, and $F_{k,b} = 92.4 + 20\log f + 20\log d_{k,b}$ is the free space path loss between U_k and H_b, which includes the distance $d_{k,b}$ between U_k and H_b, and the carrier frequency f of U_k.

Then, H_b utilize SIC to recover the received signal at each RB. Considering the impact of imperfect SIC, the signal-to-interference and noise ratio (SINR) of U_k in H_b can be expressed as:

$$\gamma_{k,b} = \frac{\alpha_{k,b} P_{k,b} |h_{k,b} l_{k,b}|^2}{X + Y + \sigma^2}, \tag{6}$$

where

1) $X = \sum_{j=k+1}^{Z} P_{j,b}\alpha_{j,b}|h_{j,b}l_{j,b}|^2$*:* Indicates the interference caused by other UEs in the same NOMA group with worse signals than that of U_k, where $P_{j,b}$ is the transmit power of U_j, and $0 \leq \alpha_{k,b} \leq 1$ is the power coefficient of U_k which can be adjusted by G_b and G_k.

2) $Y = \sum_{i=1}^{k-1} P_{i,b}\beta_{i,b}\alpha_{i,b}|h_{i,b}l_{i,b}|^2$*:* Represents the residual component of U_k's signal after SIC, where $\beta_{i,b}$ represents the residual coefficient after the i-th SIC at H_b.

3) σ^2*:* Additive white Gaussian noise (AWGN) with zero mean and variance σ^2.

D. MSCH NOMA Scheme

In our MSCH NOMA scheme in the dual-layer LEO HTS constellation, each HTS has multiple RBs to provide SGF NOMA transmission. First, considering the HTSs in L_1 as shown in Fig. 3, recall that the RBs selected by CUs are RRBs, the rest RBs selected by MUs called NRBs, and the RBs not selected by any CU and MU are eRBs. Moreover, the LoS duration t_1 in each HTS H_1 is divided into K frames, where the CUs and MUs are randomly activated and perform SGF NOMA at the beginning of each frame. Further, each frame S_i can be further divided into three subframes as $S_{i,1}$, $S_{i,2}$ and $S_{i,3}$, where $S_{i,1}$ and $S_{i,1} + S_{i,2}$ are the packet delay budgets of CUs and MUs, respectively.

Therefore, as shown in Fig. 3 (a), there are only CUs and MUs accessing to the HTS in L_1 when the EUs access to the HTS in L_2. The CUs perform GB NOMA in the RRBs,

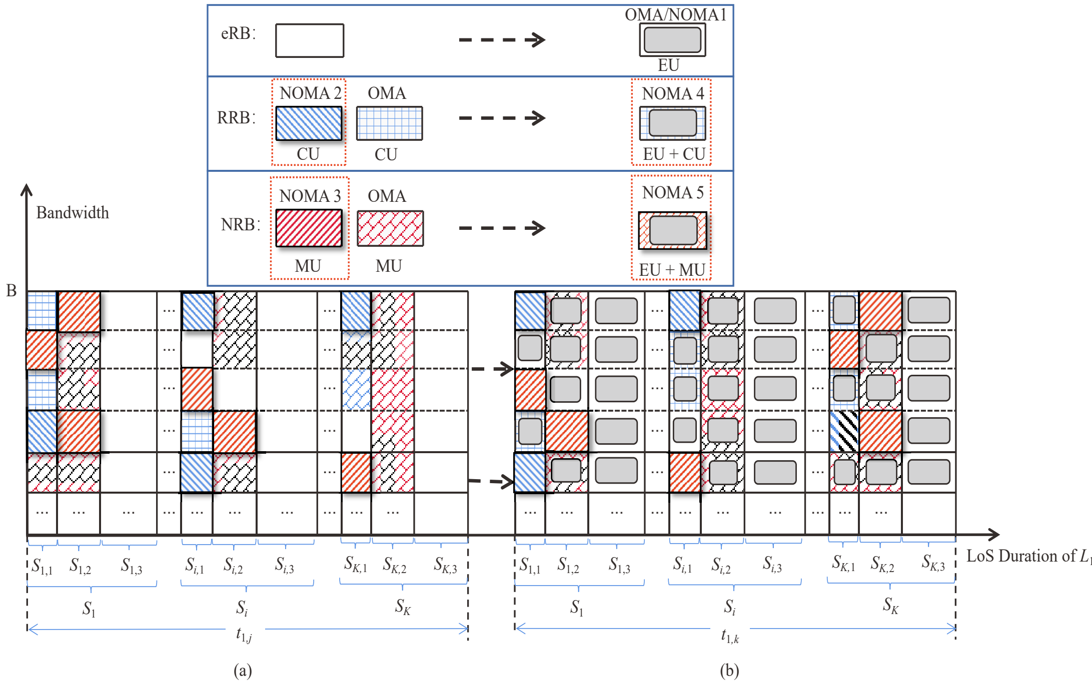


Fig. 3. The illustration of RBs and frames in our MSCH NOMA scheme. (a) CUs and MUs access to the HTS in L_1, (b) EUs access to the HTS in L_1 and perform coexist NOMA with CUs and MUs.

and the MUs perform GF NOMA the NRB. When the EUs perform handover to the HTS in L_1, the CUs, MUs and EUs perform SGF NOMA to the HTS in L_1 as shown in Fig. 3 (b). Specifically, the proposed MSCH NOMA scheme is designed under the following assumptions: 1) The CSI remains constant during S_i and randomly varies between different S_i. 2) Each RB corresponds to a random access preamble (RAP), and HTS can distinguish the selected RBs by the RAPs from received signals [35], [36]. 3) HTS can identify the number of uplink UEs and acquire the CSI of each UE in each RB through the new transceiver structure proposed in [37], which consists of a secondary preamble and a multi-user multi symbol (MUMS) detector. 4) The CUs would finish their service in $S_{i,1}$, and the MUs can perform random access with RRB in $S_{i,2}$, and with NRB in $S_{i,1}$ and $S_{i,2}$ as shown in Fig. 3.

Furthermore, the proposed MSCH NOMA scheme has three parts: 1) The SGF random access protocols for CUs, MUs and EUs as shown in Fig. 4; 2) The resource acquisition messages (RAM) generator as shown in Fig. 5; 3) The NOMA algorithms for different UEs, which are designed in Section IV. In the following, we introduce the first and second parts of the MSCH NOMA scheme.

E. SGF Random Access Protocols in MSCH NOMA Scheme

1) GB Random Access Protocol for CUs:

Step 1 The HTS broadcasts all RAPs at the beginning of $S_{i,1}$ as shown in Fig. 4 (a).

Step 2 The CUs upload the selected RAPs and their CU identities (CIDs) to mark the RRBs, and transmit power information (PI).

Step 3 The HTS broadcasts random access response (RAR), including CSI, maximum data rate and synchronization message, and RAM1 to all UEs.

Step 4 The CUs that get the successful RAR from the HTS make a radio resource control (RRC) connection request by sending their CIDs to the HTS.

Step 5 The HTS sends ACKs to all granted access CUs.

Step 6 The granted access CUs select appropriate NOMA algorithm according to the RAR and RAM1, and begin to upload data.

2) GF Random Access Protocol for MUs:

Step 1 The HTS broadcasts all RAPs at the beginning of $S_{i,1}$ as shown in Fig. 4 (b).

Step 2 The MUs upload the selected RAPs to mark the NRBs and transmit PI.

Step 3 The HTS broadcasts RAR and RAM2 to all terrestrial UEs.

Step 4 The MUs (part 1) select appropriate NOMA algorithm according to the RAR and RAM2, and begin to uplink data as shown in Fig. 3, and the rest of MUs that selecting RRBs (part 2) are waiting in the $S_{i,1}$.

Step 5 At the beginning of $S_{i,2}$, the part2 MUs upload the re-selected RAPs to mark NRBs and transmit PI.

Step 6 The HTS broadcasts RAR and RAM3 to the part 2 MUs.

Step 7 The part 2 MUs select appropriate NOMA algorithm according to the RAR and RAM3, and begin to upload data.

3) GB Access Protocol for EUs:

Step 0 The EUs perform GB NOMA access to the HTS in L_2 for the longer LoS duration [35] as shown in Fig. 4 (c). When the condition of the first option in the handover transmission scheme is satisfied, the EUs wait the time cost x_2 after t_2 to perform handover to the next HTS in L_2. Otherwise,

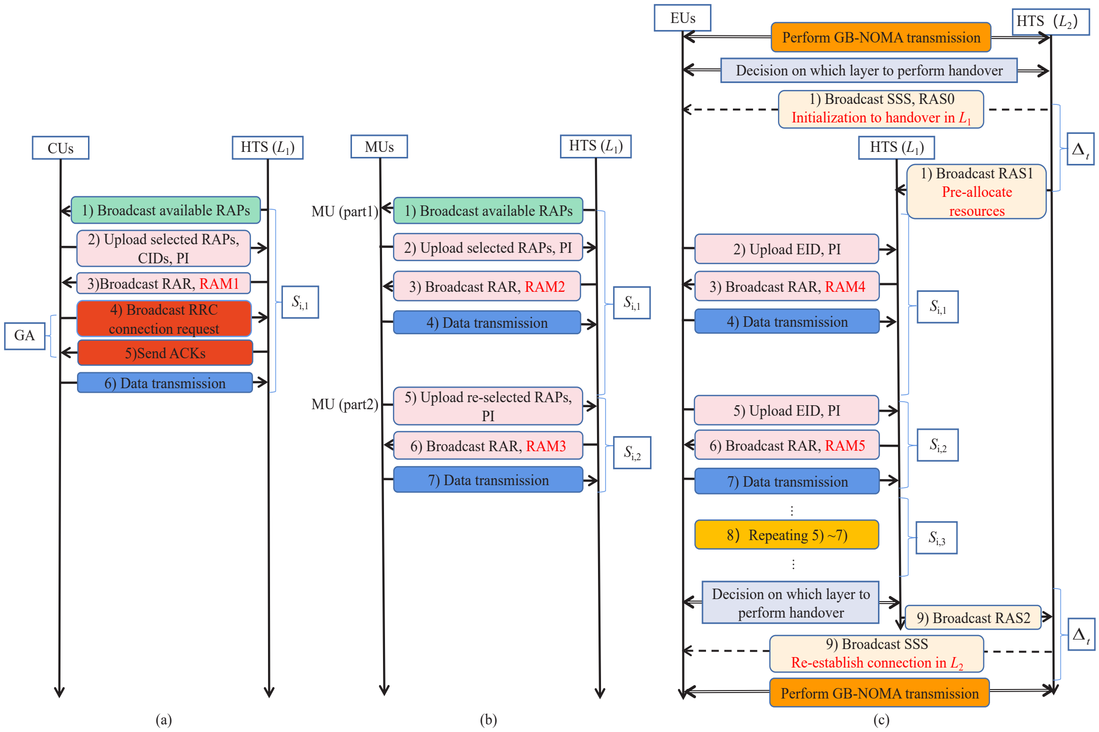


Fig. 4. The SGF random access protocols for CUs, MUs and EUs. (a) GB random access protocol for CUs. (b) GF random access protocol for MUs. (c) GB access protocol for EUs.

the current served HTS in L_2 informs an appropriate HTS in L_1 that the EUs would perform handover with time cost Δ_t [2], and the GB access protocol for EUs in L_1 is as follows.

Step 1 The current served HTS in L_2 broadcasts switching service signals (SSS) and resource allocation signal 0 (RAS0) to EUs, and transmits the EU identities (EIDs) in RAS1 to the HTS in L_1.

Step 2 After receiving SSS, the EUs begin to upload PI to the L_1 layer HTS.

Step 3 The L_1 layer HTS broadcasts the RAR and RAM4 to the EUs.

Step 4 The granted access EUs selects the appropriate RBs and corresponding NOMA algorithm according to RAR and RAM4 for data transmission.

Step 5 At the beginning of the $S_{i,2}$, the EUs upload EIDs and PI to the L_1 layer HTS.

Step 6 The L_1 layer HTSs broadcasts the RAR and RAM5 to the EUs.

Step 7 The granted access EUs selects the appropriate RBs and corresponding NOMA algorithm according to RAR and RAM5 for data transmission.

Step 8 Repeat above **Steps 5-7** in $S_{i,3}$.

Step 9 Repeat above **Steps 2-8** until the EUs can perform handover to the L_2 layer HTS. Then, the L_1 layer HTS transmits RAS2 to the L_1 layer HTS with the time cost Δ_t, and then the L_2 layer HTS broadcasts the SSS to the EUs.

Step 10 Then, EUs begin to perform GB NOMA access to the L_2 layer HTS.

F. RAM Generator in MSCH NOMA Scheme

The generation logic of RAM signals in above SGF random access protocols is shown in Fig. 5. When the L_1 layer HTS receives RAPs, CIDs, EIDs and PI from CUs, MUs and EUs, as well as RAS from L_2 layer HTS, it estimates the CSI of UEs according to the PIs and identifies the number of CUs and MUs in different RBs. In addition, the HTS utilizes the RBS algorithm in Section IV-F to select an appropriate RB for EUs. Therefore, the number of CUs, MUs and EUs in each RB can be obtained as N_c, N_m and N_e, respectively. Then, the HTS generates access denied signal as ⑧ in RAM according to N_c, N_m, N_e and CSI, or further generates control signals in RAM for the UEs to select appropriate NOMA algorithms.

III. Derivation of KPIs

In this section, we first derive specific KPIs (i.e., OP, EC and ST) to measure the performance of different types of UEs.

A. Problem Formulation Under Two UEs Model

Note that the number of UEs in power domain NOMA is usually small in practical [11], thus, we set $Z = 2$ in each RB to derive the closed-form expressions of KPIs in the following. First, recall Eq. (6), the SINR of U_1 at H_b can be expressed as:

$$\gamma_{1,b}=\frac{\alpha_{1,b}P_{1,b}|h_{1,b}|^2|l_{1,b}|^2}{\alpha_{2,b}P_{2,b}|h_{2,b}|^2|l_{2,b}|^2+\sigma^2}, \tag{7}$$

and the SINR of U_2 at H_b is

$$\gamma_{2,b}=\frac{\alpha_{2,b}P_{2,b}|h_{2,b}|^2|l_{2,b}|^2}{\beta_{1,b}\cdot\alpha_{1,b}P_{1,b}|h_{1,b}|^2|l_{1,b}|^2+\sigma^2}. \tag{8}$$

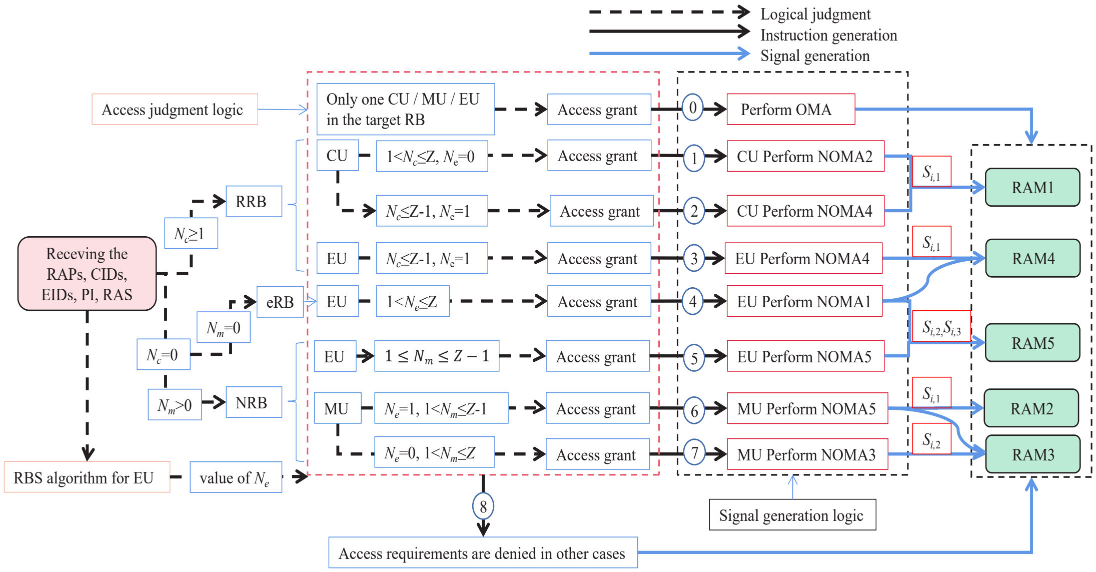


Fig. 5. Logic generation diagram of RAM signal in RRB and NRB for CU, MU and EU.

B. EC Performance Analysis

Recall that EC is defined as the expected value of instantaneous mutual information of SINR $\gamma_{k,b}$ of U_k at H_b, and we have

$$C(\gamma_{k,b}) = E[\log(1+\gamma_{k,b})]. \tag{9}$$

With the help of [39, Eq. (6)], $C(\gamma_{k,b})$ can be well approximated as

$$C(\gamma_{k,b}) = \log_2(e) \cdot \left\{ \ln[1+E(\gamma_{k,b})] - \frac{E\left({\gamma_{k,b}}^2\right) - E(\gamma_{k,b})^2}{2(1+E(\gamma_{k,b}))^2} \right\} \tag{10}$$

Further, with the help of [38, Eq. (6.455.1)], we can get the closed-form expression of $E(\gamma_{1,b})$ and $E(\gamma_{1,b}^2)$ for U_1 at H_b as follows

$$E(\gamma_{1,b}) = \frac{\rho_{1,b}\alpha_{1,b}\Gamma\left(\mathcal{M}_{1,b}+1\right)}{(1+\rho_{2,b}\alpha_{2,b}\frac{\Gamma(\mathcal{M}_{2,b}+1)}{\varepsilon_{2,b}\Gamma(\mathcal{M}_{2,b})})\varepsilon_{1,b}\Gamma\left(\mathcal{M}_{1,b}\right)}, \tag{11}$$

and

$$E(\gamma_{1,b}^2) = \frac{\rho_{1,b}^2\alpha_{1,b}^2\Gamma\left(\mathcal{M}_{1,b}+2\right)}{\left(1+\rho_{2,b}\alpha_{2,b}\frac{\Gamma(\mathcal{M}_{2,b}+1)}{\varepsilon_{2,b}\Gamma(\mathcal{M}_{2,b})}\right)^2\varepsilon_{1,b}^2\cdot\Gamma\left(\mathcal{M}_{1,b}\right)}, \tag{12}$$

where $\rho_{i,b} = \frac{P_{i,b}}{\sigma^2}|l_{i,b}|^2$ denotes the transmission average signal-noise ration (SNR) of U_i at H_b. Then, we can get EC $C(\gamma_{1,b})$ for U_1 at H_b by substituting Eq. (11) and Eq. (12) into Eq. (10).

Similarly, we can obtain the closed-form expression of $E(\gamma_{2,b})$ and $E(\gamma_{2,b}^2)$ for U_2 at H_b as follows

$$E(\gamma_{2,b}) = \frac{\rho_{2,b}\alpha_{2,b}\Gamma\left(\mathcal{M}_{2,b}+1\right)}{(1+\beta_{1,b}\rho_{1,b}\alpha_{1,b}\frac{\Gamma(\mathcal{M}_{1,b}+1)}{\varepsilon_{1,b}\Gamma(\mathcal{M}_{1,b})})\varepsilon_{1,b}\Gamma\left(\mathcal{M}_{1,b}\right)}, \tag{13}$$

and

$$E(\gamma_{2,b}^2) = \frac{\rho_{2,b}^2\alpha_{2,b}^2\Gamma\left(\mathcal{M}_{2,b}+2\right)}{\left(1+\beta_{1,b}\rho_{1,b}\alpha_{1,b}\frac{\Gamma(\mathcal{M}_{1,b}+1)}{\varepsilon_{1,b}\Gamma(\mathcal{M}_{1,b})}\right)^2\varepsilon_{1,b}^2\cdot\Gamma\left(\mathcal{M}_{1,b}\right)}, \tag{14}$$

and we can obtain EC $C(\gamma_{2,b})$ for U_2 at H_b by substituting Eq. (13) and Eq. (14) into Eq. (10). The detailed derivations for Eqs. (11)-(14) are shown in the Appendix A.

C. OP Performance Analysis

The OP of each UE is defined as the probability that the instantaneous SINR $\gamma_{k,b}$ of U_k at H_b lower than a given SINR threshold $\gamma_{k,b}^{\text{th}}$, and we have

$$F_{k,b}(\gamma_{k,b}^{\text{th}}) = \Pr(\gamma_{k,b} < \gamma_{k,b}^{\text{th}}), \tag{15}$$

where $F_{k,b}(\gamma_{k,b}^{\text{th}})$ is the cumulative distribution function (CDF) of $\gamma_{k,b}$.

With the help of [38, Eq. (3.351.2)], we can obtain the closed-form expression of OP for U_1 at H_b as follows

$$F_{1,b}\left(\gamma_{1,b}^{\text{th}}\right) = 1 - \frac{\Gamma\left(\mathcal{M}_{1,b}, \frac{\varepsilon_{1,b}\gamma_{1,b}^{\text{th}}(1+\rho_{2,b}\alpha_{2,b}\frac{\Gamma(\mathcal{M}_{2,b}+1)}{\varepsilon_{2,b}\Gamma(\mathcal{M}_{2,b})})}{\rho_{1,b}\alpha_{1,b}}\right)}{\Gamma\left(\mathcal{M}_{1,b}\right)}. \tag{16}$$

Similarly, we can obtain the closed-form expression of OP for U_2 at H_b via [38, Eq. (3.351.2)] as follows

$$F_{2,b}\left(\gamma_{2,b}^{\text{th}}\right) = 1 - \frac{\Gamma\left(\mathcal{M}_{2,b}, \frac{\varepsilon_{2,b}\gamma_{2,b}^{\text{th}}(1+\beta_{1,b}\rho_{1,b}\alpha_{1,b}\frac{\Gamma(m_{1,b}+1)}{\varepsilon_{1,b}\Gamma(\mathcal{M}_{1,b})})}{\rho_{2,b}\alpha_{2,b}}\right)}{\Gamma\left(\mathcal{M}_{2,b}\right)}. \tag{17}$$

The detailed derivations for Eq. (16) and Eq. (17) are shown in the Appendix B.

Further, note that the channel conditions of U_1 and U_2 are independent, then the joint OP expression of this NOMA group

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2022.3214927

at H_b can be obtained as

$$
\begin{aligned}
F_b\left(\gamma_b^{\mathrm{th}}\right) &= 1-\Pr(\gamma_{1,b}>\gamma_{1,b}^{\mathrm{th}})\Pr(\gamma_{2,b}>\gamma_{2,b}^{\mathrm{th}}) \\
&= 1-\frac{\Gamma\left(\mathcal{M}_{1,b}, \frac{\varepsilon_{1,b}\gamma_{1,b}^{\mathrm{th}}\left(1+\rho_{2,b}\alpha_{2,b}\frac{\Gamma(\mathcal{M}_{2,b}+1)}{\varepsilon_{2,b}\Gamma(\mathcal{M}_{2,b})}\right)}{\rho_{1,b}\alpha_{1,b}}\right)}{\Gamma\left(\mathcal{M}_{1,b}\right)} \\
&\quad \cdot \frac{\Gamma\left(\mathcal{M}_{2,b}, \frac{\varepsilon_{2,b}\gamma_{2,b}^{\mathrm{th}}\left(1+\beta_{1,b}\rho_{1,b}\alpha_{1,b}\frac{\Gamma(\mathcal{M}_{1,b}+1)}{\varepsilon_{2,b}\Gamma(\mathcal{M}_{1,b})}\right)}{\rho_{2,b}\alpha_{2,b}}\right)}{\Gamma\left(\mathcal{M}_{2,b}\right)}.
\end{aligned} \tag{18}
$$

With the given SINR threshold $\gamma_{i,b}^{\mathrm{th}}$, we can establish two inequations according to Eq. (7) and Eq. (8) as follows, which should be satisfied by α_i when decoding the signal of U_1 at H_b,

$$\frac{\alpha_{1,b}\rho_{1,b}|h_{1,b}|^2}{\alpha_{2,b}\rho_{2,b}|h_{2,b}|^2+1} \geq \gamma_{1,b}^{\mathrm{th}}, \tag{19}$$

and

$$\frac{\alpha_{2,b}\rho_{2,b}|h_{2,b}|^2}{\beta_{1,b}\cdot\alpha_{1,b}\rho_{1,b}|h_{1,b}|^2+1} \geq \gamma_{2,b}^{\mathrm{th}}. \tag{20}$$

D. ST Performance Analysis

The ST of U_k at H_b can be expressed as the product of EC and decoding success probability $\Pr_{(k,b)}$ of each UE, which can be expressed as

$$T_{k,b} = C(\gamma_{k,b})\Pr_{(k)}(\gamma_{k,b}^{\mathrm{th}}), \tag{21}$$

where $\Pr\limits_{(k,b)}(\gamma_{k,b}^{\mathrm{th}}) = 1-F_{k,b}(\gamma_{k,b}^{\mathrm{th}})$.

IV. Algorithms Design for MSCH NOMA Scheme

In this section, we focus on the design of NOMA algorithms to guarantee the KPI requirements of different types of UEs in our MSCH NOMA scheme, and we propose the RBS algorithm to enhance the EC of EUs in L_1 layer HTS.

A. IEC Algorithm for EU

As shown in Fig. 4 (c) in the aforementioned Section II-E, when EUs perform NOMA to the HTS in L_2 and the eRBs of the HTS in L_1, considering the high data rate requirements of EUs, the optimization goal is maximizing the sum EC C_{sum} of each EU NOMA group, and the optimization problem is established as follows

$$\max\ C_{sum} = C(\gamma_{e_1,b}) + C(\gamma_{e_2,b}), \tag{22a}$$

$$s.t.: 0 \leq \alpha_{e_k,b} \leq 1, \tag{22b}$$

$$F_{e_k,b}(\gamma_{e_k,b}^{\mathrm{th}}) \leq \xi_{e_i,b}, \tag{22c}$$

where (22b) indicates the power coefficient $\alpha_{e_k,b}$ for EU_k at H_b, (22c) guarantees an OP threshold $\xi_{e_i,b}$ for each EU at H_b.

Recall that EU_1 has better channel condition, and $C(\gamma_{e_1,b})$ has higher impact on C_{sum} than $C(\gamma_{e_2,b})$. Then, we can straightforwardly derive that $\alpha_{e_1,b}$ is strictly positive with $C(\gamma_{e_1,b})$ and $\alpha_{e_2,b}$ is negative with $C(\gamma_{e_1,b})$, which is proofed in Appendix C. Therefore, we can maximize C_{sum} with $\alpha_{e_1,b} = 1$. In addition, according to Eq. (16) and Eq. (17), we can derive that $\alpha_{e_2,b}$ is strictly positive with $F_{e_1,b}(\gamma_{e_1,b}^{\mathrm{th}})$, and negative with $F_{e_2,b}(\gamma_{e_2,b}^{\mathrm{th}})$. Thus, in order to satisfy the OP threshold $\xi_{e_i,b}$ in (22c), we should find a minimum $\alpha_{e_2,b}$ to satisfy $F_{e_k,b}(\gamma_{e_k,b}^{\mathrm{th}}) \leq \xi_{e_k,b}$ according to Eq. (19) and Eq. (20). Therefore, we propose the IEC algorithm as NOMA 1 in Fig. 3, which is concluded in Algorithm 1.

Algorithm 1: IEC Algorithm for EUs and IST Algorithm for MUs

Input: $\mathcal{M}_{k,b}$, $\varepsilon_{k,b}$, $\beta_{k,b}$, $\gamma_{k,b}^{\mathrm{th}}$, $\rho_{k,b}$, $h_{k,b}$, $l_{k,b}$, $\xi_{k,b}$, current system transmit power P_{cur}, maximum UE power P_{max}, step length $\Delta_{\alpha_m} > 1$, iterations χ, N_c;
Output: Power coefficients $\alpha_{1,b}$ and $\alpha_{2,b}$;

if $N_c = 1$ *or* $N_e = 1$ **then**
    Return $\alpha_{1,b} = 1$, and $\alpha_{2,b} = 0$;
end
for $P_{cur} \leq P_{max}$ **do**
    Initialize $right = 1$ and $left = 0$;
    Substitute $\alpha_{1,b} = 1$ and $\alpha_{2,b} = left$ into Eq. (17) to get $F_{2,b}(\gamma_{2,b}^{\mathrm{th}})$.
    for $\chi > 0$ **do**
        let $mid = left + \frac{left+right}{\Delta_{\alpha_m}}$,
        Substitute $\alpha_{1,b} = 1$ and $\alpha_{2,b} = mid$ into Eq. (17) to update $F_{2,b}(\gamma_{2,b}^{\mathrm{th}})$.
        if $F_{2,b}(\gamma_{2,b}^{\mathrm{th}}) > \xi_{2,b}$ **then**
            $left = mid$;
        else
            Substitute $\alpha_{1,b} = 1$ and $\alpha_{2,b} = mid$ into Eq. (16) to update $F_{1,b}(\gamma_{1,b}^{\mathrm{th}})$.
            if $F_{1,b}(\gamma_{2,b}^{\mathrm{th}}) > \xi_{1,b}$ **then**
                $right = mid$;
            else
                break;
            end
        end
        $\chi = \chi - 1$;
    end
end
Return $\alpha_{1,b}$ and $\alpha_{2,b}$;

B. IOP Algorithm for CU

First, since CUs and MUs only access the L_1 layer HTS, we can ignore the subscript b in the following.

Then, according to Eq. (44) in Appendix A, the OP of U_1 at H_b can be expressed as $F_1(\gamma_1^{\mathrm{th}}) = 1 - \int_{\frac{\gamma_1^{\mathrm{th}}(\alpha_2\rho_{2,b}|h_{2,b}|^2+1)}{\alpha_{1,b}\rho_1}}^{\infty} f_{|h_1|^2}(y)dy$, and $f_{|h_1|^2}(y)$ is larger than zero. Therefore, $\int_{\frac{\gamma_1^{\mathrm{th}}(\alpha_2\rho_2|h_2|^2+1)}{\alpha_1\rho_1}}^{\infty} f_{|h_1|^2}(y)dy$ is positive with α_1, which means $F_1(\gamma_1^{\mathrm{th}})$ has negative relationship with α_1. Similarly, we can derive that each α_k has a negative relationship with $F_k(\gamma_k^{\mathrm{th}})$.

Therefore, when CUs perform NOMA on RRBs of the HTS in L_1 as shown in Fig. 4 (a) in the aforementioned Section II-E, considering the high reliability requirements of CUs, the optimization problem is to minimize the OP $F_k(\gamma_k^{\text{th}})$ of both CUs, which indicates the power constraints of $\alpha_{c_k}=1$ due to the negative relationship for each α_{c_k} with $F_k(\gamma_k^{\text{th}})$, and we have

$$\min\max\ F_k(\gamma_k^{\text{th}}), \tag{23a}$$

$$s.t.: \log\left(1+\gamma_{c_k}^{\text{th}}\right) \geq R_{c_k}, \tag{23b}$$

where (23b) is the SINR threshold $\gamma_{c_k}^{\text{th}}$ to guarantee the required data rate R_{c_k} of the k-th CU.

Moreover, $\gamma_{c_k}^{\text{th}}$ in (23c) can be expressed as

$$\gamma_{c_k}^{\text{th}} \geq 2^{R_{c_k}}-1. \tag{24}$$

Note that a fixed SINR threshold $\gamma_{c_k}^{\text{th}}$ would lead to error floor [40]. Therefore, we set a varying SINR threshold $\gamma_{c_k}^{\text{th}}$ as follows.

Since (23a) is established for both CUs in the NOMA group, it can be expressed as

$$F_{c_1}(\gamma_{c_1}^{\text{th}}) = F_{c_2}(\gamma_{c_2}^{\text{th}}). \tag{25}$$

By substituting Eq. (16), Eq. (17) into Eq. (25), we can get the expression of $\gamma_{c_k}^{\text{th}}$ to satisfy Eq. (25) as

$$\frac{\gamma_{c_1}^{\text{th}}}{\rho_{c_1}\left(1+\beta_{c_1}\rho_{c_1}\frac{\Gamma\left(\mathcal{M}_{c_1}+1\right)}{\varepsilon_{c_1}\Gamma\left(\mathcal{M}_{c_1}\right)}\right)} = \frac{\gamma_{c_2}^{\text{th}}}{\rho_{c_2}(1+\rho_{c_2}\frac{\Gamma(\mathcal{M}_{c_2}+1)}{\varepsilon_{c_2}\Gamma(\mathcal{M}_{c_2})})}. \tag{26}$$

Therefore, we can set the varying SINR threshold $\gamma_{c_k}^{\text{th}}$ as follows:

$$\gamma_{c_1}^{\text{th}} = \max\left\{2^{R_{c_1}}-1, \frac{1}{\rho_{c_1}\left[1+\beta_{c_1}\rho_{c_1}\frac{\Gamma\left(\mathcal{M}_{c_1}+1\right)}{\varepsilon_{c_1}\Gamma\left(\mathcal{M}_{c_1}\right)}\right]}\right\}, \tag{27}$$

and

$$\gamma_{c_2}^{\text{th}} = \max\left\{2^{R_{c_2}}-1, \frac{1}{\rho_{c_2}(1+\rho_{c_2}\frac{\Gamma(\mathcal{M}_{c_2}+1)}{\varepsilon_{c_2}\Gamma(\mathcal{M}_{c_2})})}\right\}. \tag{28}$$

Therefore, we can solve (23a) by calculating Eq. (27) and Eq. (28), which is the NOMA 2 in Fig. 3.

C. IST Algorithm for MU

Further, as shown in Fig. 4 (b) in the aforementioned Section II-E, when MUs perform NOMA on the NRBs of the HTS in L_1, considering the massive connectivity requirements of MUs, the optimization problem is maximizing the sum ST T_{sum} of each MU NOMA group as follows,

$$\max T_{sum} = T_{m_1}+T_{m_2}, \tag{29a}$$

$$s.t.: 0 \leq \alpha_{m_k} \leq 1, \tag{29b}$$

$$F_{m_k}(\gamma_{m_k}^{\text{th}}) \leq \xi_{m_k}, \tag{29c}$$

where (29b) is the power coefficient α_{m_k}, (29c) guarantees the OP performance of each MU at H.

Note that $\Pr_{(m_k)}(R_{m_k}^{\text{th}})$ is close to 1 under the condition of high SNR, thus, T_{m_k} approaches to $C(\gamma_{m_k})$. Therefore, the optimization problem (29a) is similar to (22a), and the IST algorithm is similar to the IEC Algorithm as shown in Algorithm 1, which is the NOMA 3 in Fig. 3.

D. ECco NOMA Algorithm for EU with CU

Moreover, as shown in Fig. 3 (b) and Fig. 4 (c) in the aforementioned Section II-D and Section II-E, respectively, when EUs perform NOMA with CUs on the RRBs of the HTS in L_1. Considering the stringent requirements of CUs, the signals of CUs have higher priority to perform SIC in the RRBs. Therefore, the SINR of CUs and EUs on RRBs can be expressed as

$$\gamma_c = \frac{\alpha_c\rho_c|h_c|^2}{\alpha_e\rho_e|h_e|^2+1}, \tag{30}$$

and

$$\gamma_e = \frac{\alpha_e\rho_e|h_e|^2}{\alpha_c\beta_c\rho_c|h_c|^2+1}. \tag{31}$$

In addition, the EC of CU $C(\gamma_c)$ and EU $C(\gamma_e)$ should satisfy the inequalities as follows

$$\begin{cases} C(\gamma_e) \geq C_c^{\text{th}}, \\ C\left(\gamma_c\right) \geq C_e^{\text{th}}, \end{cases} \tag{32}$$

where C_c^{th} and C_e^{th} are the EC requirements of CUs and EUs, respectively. Moreover, C_c^{th} is effected by the delay budget of CU [6], and C_e^{th} is determined by the data requirement (25 Mbit/s [7]) of EU.

By substituting (10) into (32), we can derive that

$$\begin{cases} \alpha_e \geq \frac{x\cdot\varepsilon_e\Gamma(\mathcal{M}_e)(1+y\cdot\beta_c)}{\rho_e\Gamma(\mathcal{M}_e+1)(1-\beta_c\cdot x\cdot y)} = \partial_e^1, \\ \alpha_c \geq \frac{y\cdot\varepsilon_c\Gamma(\mathcal{M}_c)(1+x)}{\rho_c\Gamma(\mathcal{M}_c+1)(1-\beta_c x\cdot y)} = \partial_c, \end{cases} \tag{33}$$

where $x=\left(2^{C_e^{\text{th}}}-1\right)$, and $y=\left(2^{C_c^{\text{th}}}-1\right)$.

Further, since the SIC first recovers the signal of CU, the OP $F_c(\gamma_c^{\text{th}})$ of CU can be expressed as

$$F_c\left(\gamma_c^{\text{th}}\right) = 1-\frac{\Gamma\left(\mathcal{M}_c, \frac{\varepsilon_c\gamma_c^{\text{th}}(1+\rho_e\alpha_e\frac{\Gamma(\mathcal{M}_e+1)}{\varepsilon_e\Gamma(\mathcal{M}_e)})}{\rho_c\alpha_c}\right)}{\Gamma\left(\mathcal{M}_c\right)}. \tag{34}$$

Therefore, the optimization goal is maximizing the EC $C(\gamma_e)$ of EUs under the an appropriate OP threshold ξ_c for CUs, and the optimization problem is established as follows

$$\max C(\gamma_e), \tag{35a}$$

$$s.t.: \partial_c \leq \alpha_c \leq 1, \tag{35b}$$

$$\partial_e^1 \leq \alpha_e \leq 1, \tag{35c}$$

$$F_c(\gamma_c^{\text{th}}) \leq \xi_c. \tag{35d}$$

where (35b) and (35c) indicate the constraints of α_c and α_e to satisfy the EC requirements of CU and EU, respectively, and (35d) guarantees the OP of CU ξ_c.

Considering the high EC requirements of EUs and the positive relationship for $C(\gamma_e)$ and α_e, we first initialize $\alpha_e=1$. Then, considering the stringent OP requirement of CU, we need to maximize EC of EU under the OP threshold of CU, which may need further adjust α_e and α_c, simultaneously. Thus, we propose the ECco algorithm as NOMA 4 in Fig. 3, which is summarized in Algorithm 2.

Algorithm 2: ECco NOMA Algorithm for CU and EU coexist

```
Input: M_k, ε_k, β_k, γ_k^th, ρ_k, h_k, l_k, ξ_c, C_k^th, P_cur, P_max,
       step length Δ_{α_k};
Output: α_c and α_e;
for P_cur ≤ P_max do
    Calculate ∂_c and ∂_e^1 according to Eq. (33);
    Initialize α_e = 1, and α_c = min(1, ∂_c).
    Substitute α_e and α_c into Eq. (34) to get F_c(γ_c^th).
    while α_c < 1 do
        Substitute α_e and α_c into Eq. (34) to update
        F_c(γ_c^th).
        if F_c(γ_c^th) > ξ_c then
            α_c = min(1, α_c + Δ_{α_c});
        else
            break;
        end
    end
    while α_e > ∂_e^1 do
        Substitute α_e and α_c into Eq. (34) to update
        F_c(γ_c^th).
        if F_c(γ_c^th) > ξ_c then
            α_e = max(∂_e^1, α_e − Δ_{α_e});
        else
            break;
        end
    end
end
Return α_c and α_e;
```

E. EMco NOMA Algorithm for EU with MU

Furthermore, when the EU performs NOMA with MU on the NRBs of the HTS in L_1 as shown in Fig. 3 (b) and Fig. 4 (c) in the aforementioned Section II-D and Section II-E, respectively, the signals of EU could decoded first than that of MU due to the delay tolerant of MUs and high EC requirements of EUs. Therefore, the SINR of EUs and MUs on NRBs can be expressed as

$$\gamma_e = \frac{\alpha_e \rho_e |h_e|^2}{\alpha_m \rho_m |h_m|^2 + 1}, \tag{36}$$

and

$$\gamma_m = \frac{\alpha_m \rho_m |h_m|^2}{\beta_e \alpha_e \rho_e |h_e|^2 + 1}. \tag{37}$$

Considering the high EC requirements of EUs, the optimization goal is maximizing $C(\gamma_e)$ of EU under an appropriate ST threshold for MU. Note that T_{m_k} approaches $C(\gamma_{m_k})$ under high SNR, the optimization problem is maximizing $C(\gamma_e)$ under the an appropriate EC threshold C_m^{th} for MU. Similarly, we have the following inequalities

$$\begin{cases} C(\gamma_m) \geq C_m^{\text{th}}, \\ C(\gamma_e) \geq C_e^{\text{th}}, \end{cases} \tag{38}$$

where C_m^{th} and C_e^{th} are the EC requirements of MU and EU, respectively. By substituting (10) into (38), we have

$$\begin{cases} \alpha_e \geq \frac{x \cdot \varepsilon_e \Gamma(\mathcal{M}_e)(1+z)}{\rho_e \Gamma(\mathcal{M}_e+1)(1-\beta_e x \cdot z)} = \partial_e^2, \\ \alpha_m \geq \frac{z \cdot \varepsilon_m \Gamma(\mathcal{M}_m)(1+x \cdot \beta_e)}{\rho_m \Gamma(\mathcal{M}_m+1)(1-\beta_e x \cdot x)} = \partial_m, \end{cases} \tag{39}$$

where $x = \left(2^{C_e^{\text{th}}} - 1\right)$, and $z = \left(2^{C_m^{\text{th}}} - 1\right)$.

Moreover, the signal of MU is decoded after EU in the NRB, and the OP $F_m(\gamma_m^{\text{th}})$ of MU can be expressed as

$$F_m\left(\gamma_m^{\text{th}}\right) = 1 - \frac{\Gamma\left(\mathcal{M}_m, \frac{\varepsilon_m \gamma_m^{\text{th}} (1+\beta_e \rho_e \alpha_e \frac{\Gamma(m_e+1)}{\varepsilon_e \Gamma(\mathcal{M}_e)})}{\rho_m \alpha_m}\right)}{\Gamma\left(\mathcal{M}_m\right)}. \tag{40}$$

Therefore, the corresponding optimization problem is established as follows

$$\max C(\gamma_e), \tag{41a}$$

$$s.t. : \partial_m \leq \alpha_m \leq 1, \tag{41b}$$

$$\partial_e^2 \leq \alpha_e \leq 1, \tag{41c}$$

$$F_m(\gamma_m^{\text{th}}) \leq \xi_m. \tag{41d}$$

where (41b) and (41c) indicate the constraints of α_m and α_e to satisfy the EC requirements of MU and EU, respectively, and (41d) guarantees the OP of MU ξ_m.

Similar to the ECco algorithm, we set $\alpha_e = 1$ due to the high EC requirements of EUs. Then, we adjust α_e and α_c to maximize EC of EU under the OP threshold of MU, and propose the EMco algorithm as NOMA 5 in Fig. 3, which is summarized in Algorithm 3.

F. RBS Algorithm for EU in L_1 Layer

In addition, as shown in Fig. 5 in the aforementioned Section II-F, when the EUs perform GB access to the HTS in L_1, there are three kinds of RBs such as eRBs, NRBs and RRBs for the EUs to perform IEC, EMco NOMA, ECco NOMA algorithms, respectively. We define the statistical EC as the average EC of the UEs at HTS, which is obtained by the HTS after long-term interaction with the UEs. Let $\bar{C}_1$ denote the statistic EC difference of two EUs and one EU on an eRB with IEC algorithm, and $\bar{C}_2$ and $\bar{C}_3$ denote the statistic EC of EU on the NRBs and RRBs, respectively. Thus, in order to achieve the maximum EC for the EUs, we propose the RBS algorithm as follows.

If the L_1 layer HTS knows the above statistic EC $\bar{C}_1$, $\bar{C}_2$ and $\bar{C}_3$, and let N_1, N_2 and N_3 denote the number of eRBs, NRBs and RRBs that can be allocated to the EUs, respectively. Then, the HTS broadcasts the RAM4 and RAM5 to grant the j-th EU ($1 \leq j \leq J$) to access an RB as $EU_j^{w,d}$ to achieve the maximum EC by allocating adaptive number of RBs for EUs, where w, d denote the w-th RB in the d-th class of RBs, for example, $EU_5^{2,3}$ means the 5-th EU perform ECco NOMA to the second RRB. The RBS algorithm is summarized in Algorithm 4.

Else, the HTS allocates N_1 EUs with better channel conditions to access the eRBs, and the rest EUs perform random access to the available RBs.

Algorithm 3: EMco NOMA Algorithm for MU and EU coexist

Input: $\mathcal{M}_k$, ε_k, β_k, γ_k^{th}, ρ_k, h_k, l_k, ξ_m, C_k^{th}, P_{cur}, P_{max}, step length Δ_{α_m};
Output: α_m and α_e;

```
for P_cur ≤ P_max do
    Calculate ∂_m and ∂_e^2 according to Eq. (39);
    Initialize α_e = 1, and α_m = min(1, ∂_m).
    Substitute α_e and α_m into Eq. (40) to get F_m(γ_m^th).
    while α_m < 1 do
        Substitute α_e and α_m into Eq. (40) to update F_m(γ_m^th).
        if F_m(γ_m^th) > ξ_m then
            α_m = min(1, α_m + Δ_{α_m});
        else
            break;
        end
    end
    while α_e > ∂_e^2 do
        Substitute α_e and α_m into Eq. (40) to update F_m(γ_m^th).
        if F_m(γ_m^th) > ξ_m then
            α_e = max(∂_e^2, α_e − Δ_{α_e});
        else
            break;
        end
    end
end
Return α_m and α_e;
```

Algorithm 4: RBS Algorithm for EU in L_1 Layer

Input: N_1, N_2, N_3, $\bar{C}_1$, $\bar{C}_2$ and $\bar{C}_3$, EU_j, S_i;
Output: $[EU_1, ..., EU_J]$;

```
Initialization: j = 1, w = 1, d = 1, Y_1 = N_1, Y_2 = N_2, Y_3 = N_3;
for j ∈ [1, N_1] and j ≤ J, do
    EU_j = EU_j^{w,d}; j = j + 1, w = w + 1;
end
for j ∈ [N_1 + 1, 2N_1 + N_2 + N_3] and j ≤ J do
    if (C̄_1 == max(C̄_1, C̄_2, C̄_3) or Y_2 == 0 or Y_3 == 0) and Y_1 > 0 then
        w = N_1 − Y_1 + 1; d = 1; EU_j = EU_j^{w,d}; Y_1 = Y_1 − 1;
    end
    if (C̄_2 == max(C̄_1, C̄_2, C̄_3) or Y_1 == 0 or Y_3 == 0) and Y_2 > 0 and (EU access in S_{i,1} or S_{i,2}) then
        w = N_2 − Y_2 + 1; d = 2; EU_j = EU_j^{w,d}; Y_2 = Y_2 − 1;
    end
    if (C̄_3 == max(C̄_1, C̄_2, C̄_3) or Y_1 == 0 or Y_2 == 0) and Y_3 > 0 and (EU access in S_{i,1}) then
        w = N_3 − Y_3 + 1; d = 3; EU_j = EU_j^{w,d}; Y_3 = Y_3 − 1;
    end
    j = j + 1;
end
Return [EU_1, ..., EU_J];
```

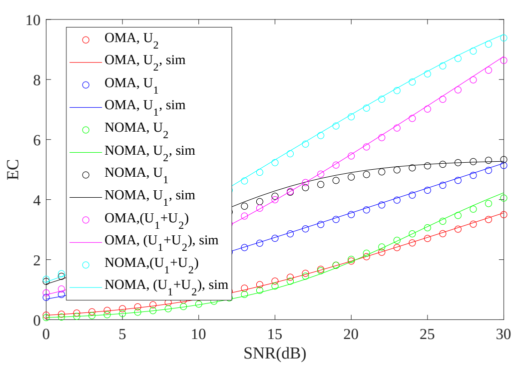


Fig. 6. EC performance of fixed power coefficient NOMA scheme and OMA scheme

TABLE III
SYSTEM PARAMETERS

Parameter	Value
The altitude of HTS in L_1 (km)	350
The altitude of HTS in L_2 (km)	1200
The number of RBs in each HTS D	320
The total number of CUs	200
The total number of MUs	32000
The activation probability of each CU p_c	0.2
The activation probability of each MU p_m [7]	0.01
Additive white Gaussian noise (K)	300
Bandwidth of each RB (MHz)	5
Rain attenuation parameters (μ, σ) (dB) [23]	-2.6, 1.6
Maximum HTS beam gain (dBi) [23]	52.1
UE antenna main lobe gain (dBi)	22.1
Imperfect SIC coefficient	5×10^{-3}

V. Simulation and Analysis

In this section, we present the simulation results to validate the theoretical analysis and show the superiority of our proposed algorithms. The simulation parameters are given in Table III.

A. Validation of Theoretical Analysis

1) EC: The EC performance of a fixed power coefficient ($\alpha_1 = 0.2, \alpha_2 = 0.8$) NOMA scheme and OMA scheme is shown in Fig. 6. We can observe that the Monte Carlo simulations agree well with our analysis results in Fig. 6, which indicates that our derived expressions Eq. (10), Eq. (11), Eq. (12), Eq. (13) and Eq. (14) can accurately evaluate the EC of our system.

2) OP: The OP performance of a NOMA scheme with a fixed power coefficient ($\alpha = 0.2, \alpha_2 = 0.8$) and a fixed SINR threshold ($\gamma^{\text{th}} = 1$ dB) is shown in Fig. 7. The Monte Carlo simulations validate the accuracy of Eq. (16), Eq. (17) and Eq. (18). Note that the ST expression in Eq. (21) is a multiplicative combination of OP and EC, which can be validated with the above results.

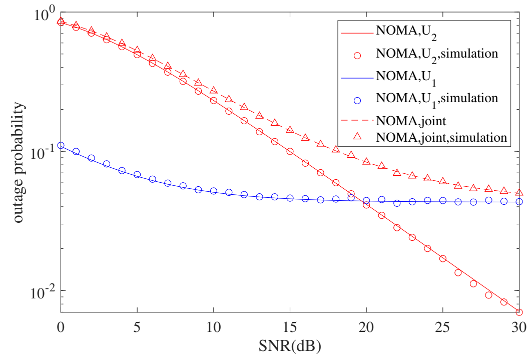


Fig. 7. OP performance of using a fixed power coefficient NOMA scheme

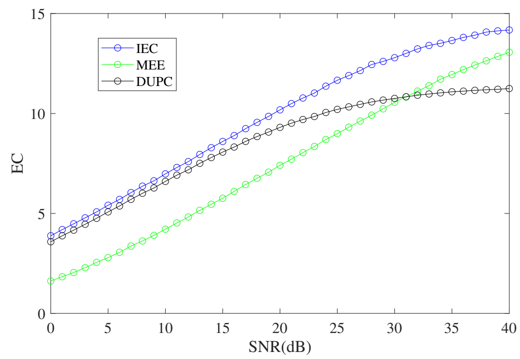


Fig. 8. Comparison of EC performance for EUs.

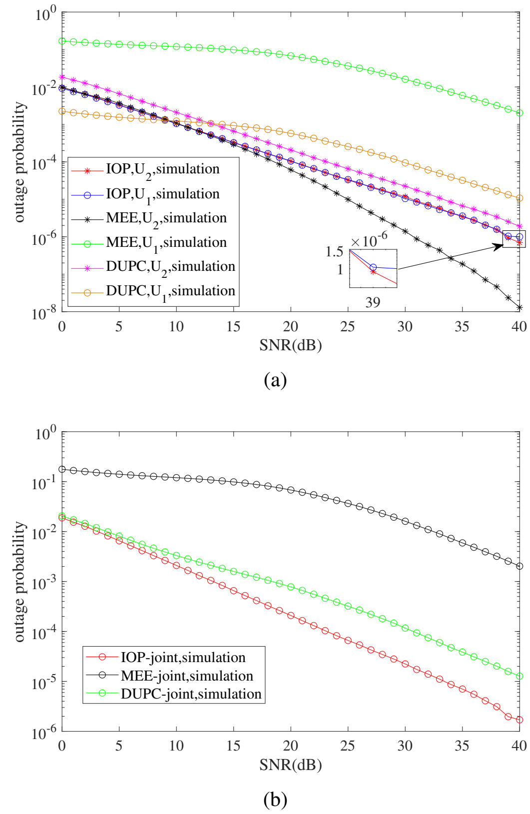


Fig. 9. Comparison of OP performance for (a) U_1 and U_2 separately and (b) CU NOMA group.

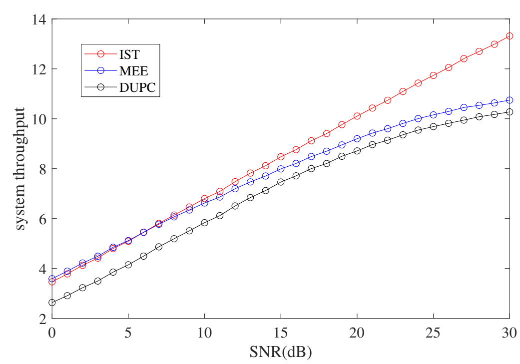


Fig. 10. Comparison of ST performance for MU group.

B. Comparison of NOMA Algorithms

1) Comparison of EC Performance for EUs: As shown in Fig. 8, the IEC algorithm can achieve better EC performance than other NOMA schemes. For example, the IEC algorithm can save about 4 dB and 9 dB compared with the DUPC and MEE algorithm when reaching the same EC (e.g., EC = 10 bit/s/Hz), respectively.

2) Comparison of OP Performance for CUs: As shown in Fig. 9 (a), the IOP algorithm can achieve the goal of minimizing the OP of each CU, and it can achieve an OP of 10^{-6} per CU at SNR about 39 dB. Meanwhile, as shown in Fig. 9 (b), the IOP algorithm can achieve much better joint OP for each CU NOMA group than the MEE NOMA algorithm, and can save 7 dB than the DUPC algorithm when reaching the same OP performance at higher SNR ($>$ 25 dB).

3) Comparison of ST Performance for MUs: As shown in Fig. 10, at higher SNR ($>$ 6dB), the IST algorithm can achieve better ST performance than other NOMA schemes, and save 4 dB and 8 dB compared with MEE and DUPC algorithms when reaching the same ST performance (e.g., ST = 10 bit/s/Hz), respectively.

4) Comparison of EC performance for EU and CU coexistence: As shown in Fig. 11 that although the ECco NOMA algorithm has lower EC of EU due to the OP requirements of CU, the proposed ECco NOMA algorithm can approach higher sum EC performance than that of MEE algorithm before the high SNR region ($\leq$ 30 dB), and can obtain better EC performance than the DUPC algorithm.

5) Comparison of EC performance for EU and MU coexistence: As shown in Fig. 12, the proposed EMco NOMA algorithm can save about 10 dB and 18 dB compared with MEE and DUPC algorithms when reaching the same EC at higher SNR ($>$ 22dB), respectively.

C. Comparison of EC performance of RBS Algorithm

As shown in Fig. 13, when fixing the SNR of each EU as 30dB, the EC gain of RBS algorithm compared with the random access (RA) algorithm in $S_{i,1}$ and $S_{i,2}$ subframes, and the average EC for each EU with RBS algorithm is better than the RA algorithm with different numbers of access EUs. Specifically, when $J = 60$, the RBS algorithm can obtain

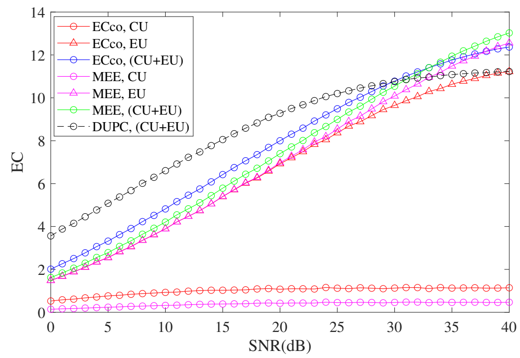


Fig. 11. Comparison of EC performance for CU with EU performing joint NOMA.

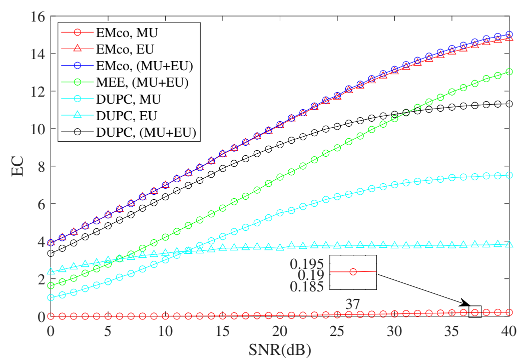


Fig. 12. Comparison of EC performance for MU with EU performing joint NOMA.

about 4.2% EC gain in $S_{i,1}$ compared with the RA algorithm. In addition, when $J = 200$, the EC gain of RBS algorithm compared with RA algorithm in $S_{i,1}$ and $S_{i,2}$ is about 16.4% and 59.9%, respectively. Further, when $J = 270$, the EC gain of RBS algorithm compared with RA algorithm in $S_{i,1}$ and $S_{i,2}$ is about 1% and 11.5%, respectively.

VI. Conclusion

In this paper, we have proposed the MSCH NOMA scheme to guarantee the KPI requirements of multi-type UEs coexistence in the dual-layer LEO HTS constellation. Specifically, we first proposed a handover transmission scheme for EUs in a dual-layer LEO HTS constellation, then we proposed the SGF random access protocols for CUs, MUs and EUs performing coexist uplink transmission, and further proposed the RAM generator to control transmissions of the UEs. Moreover, we have derived the closed-form expressions of KPIs (i.e., OP, EC and ST) to measure the performance of different types of UEs, and utilized them to design the NOMA algorithms, where the IEC and IST NOMA algorithms can maximum the EC of EUs and ST of MUs under an appropriate OP thresholds, respectively, and the IOP NOMA algorithm can minimize OP of each CUs, and the ECco and EMco NOMA algorithms both can maximize the EC of EUs under an appropriate OP threshold for CUs or MUs, respectively. In addition, we have proposed an RBS algorithm to further improve the EC of EUs. Finally, the accuracy of our derived expressions has been verified by Monte Carlo simulations, and extensive simulation

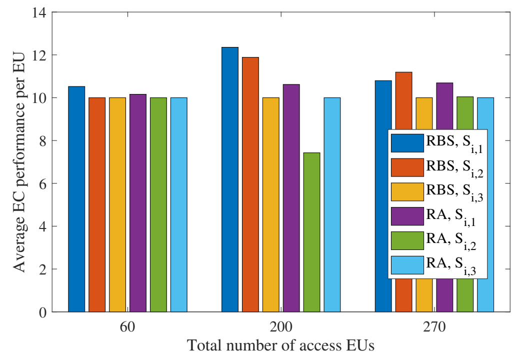


Fig. 13. Comparison of EC performance for EU scheduling strategy.

results demonstrate that the proposed NOMA algorithms have better OP, EC and ST performance than the MEE-NOMA and DUPC-NOMA algorithms.

Appendix A
Derivations of Eq. (11) to Eq. (14)

Since the first order moment expression of $\gamma_{k,b}$ can be calculated by:

$$E\left[\gamma_{k,b}\right] = \int_0^{\infty} [1 - F_{k,b}(x)]dx. \tag{42}$$

By substituting Eq. (16) and Eq. (17) into Eq. (42), and with the help of [37, Eq.(6.455.1)], we can get the first-moment expression of $\gamma_{1,b}$ and $\gamma_{2,b}$ as Eq. (11) and Eq. (13), respectively.

In addition, the second order moment expression of $\gamma_{k,b}$ can be calculated by:

$$E\left[\left(\gamma_{2,b}\right)^2\right] = 2\int_0^{\infty} x\left[1 - F_{2,b}(x)\right]dx. \tag{43}$$

Then, we can get the second-moment expression of $\gamma_{1,b}$ and $\gamma_{2,b}$ as Eq. (12) and Eq. (14), respectively, by substituting Eq. (16) and Eq. (17) into Eq. (47) and via [37, Eq.(6.455.1)].

Appendix B
Derivation of Eq. (16) and Eq. (17)

By substituting Eq. (7) into Eq. (15), and with the help of [37, Eq.(3.351.2)], we can derive the OP of U_1 at H_b as follows:

$$\begin{aligned}
F_{\gamma_{1,b}}(\gamma_{1,b}^{\rm th}) &= \Pr(\gamma_{1,b} < \gamma_{1,b}^{\rm th}) \\
&= \Pr\left(\frac{\alpha_{1,b}\rho_{1,b}|h_{1,b}|^2}{\alpha_{2,b}\rho_{2,b}|h_{2,b}|^2+1} < \gamma_{1,b}^{\rm th}\right), \\
&= \Pr\left(|h_{1,b}|^2 < \frac{\gamma_{1,b}^{\rm th}(\alpha_{2,b}\rho_{2,b}|h_{2,b}|^2+1)}{\alpha_{1,b}\rho_{1,b}}\right), \\
&= \int_0^{\frac{\gamma_{1,b}^{\rm th}(\alpha_{2,b}\rho_{2,b}|h_{2,b}|^2+1)}{\alpha_{1,b}\rho_{1,b}}} f_{|h_{1,b}|^2}(y)dy \\
&= 1 - \int_{\frac{\gamma_{1,b}^{\rm th}(\alpha_{2,b}\rho_{2,b}|h_{2,b}|^2+1)}{\alpha_{1,b}\rho_{1,b}}}^{\infty} f_{|h_{1,b}|^2}(y)dy, \\
&\overset{A}{=} 1 - \frac{\Gamma\left(M_{1,b}, \frac{\varepsilon_{1,b}\gamma_{1,b}^{th}(1+\rho_{2,b}\alpha_{2,b}|h_{2,b}|^2)}{\rho_{1,b}\alpha_{1,b}}\right)}{\Gamma\left(M_{1,b}\right)}
\end{aligned} \tag{44}$$

where Eq. A is derived with the help of [37, Eq.(3.351.2)].

Similarly, by substituting Eq. (4) into Eq. (5), and with the help of [37, Eq.(3.351.2)], we can derive the OP of U_2 at H_b as

$$F_{2,b}(\gamma_{2,b}^{\mathrm{th}}) = 1 - \frac{\Gamma\left(M_{2,b}, \frac{\varepsilon_{2,b}\gamma_{2,b}^{th}(1+\beta_{1,b}\rho_{1,b}\alpha_{1,b}|h_{1,b}|^2)}{\rho_{2,b}\alpha_{2,b}}\right)}{\Gamma\left(M_{2,b}\right)}. \tag{45}$$

Since $|h_{k,b}|^2$ of each U_k at H_b obeys the log-normal distribution, we utilized the expected value of $|h_{k,b}|^2$, $|h_{k,b}|^2_{mean}$ to simplify the calculation.

With the help of [37, Eq.(3.381.11), we can obtain the $|h_{k,b}|^2_{mean}$ as

$$\begin{aligned}
|h_{k,b}|^2_{mean} &= \int_{-\infty}^{\infty} x \cdot \frac{\varepsilon_{k,b}^{\mathcal{M}_{k,b}}}{\Gamma\left(\mathcal{M}_{k,b}\right)} x^{\mathcal{M}_{k,b}-1} \exp\left(-\varepsilon_{k,b}\cdot x\right) dx \\
&= \frac{\varepsilon_{k,b}^{\mathcal{M}_{k,b}}}{\Gamma\left(\mathcal{M}_{k,b}\right)} \cdot \int_{-\infty}^{\infty} x^{\mathcal{M}_{k,b}} \exp\left(-\varepsilon_{k,b}x\right) dx \\
&= \frac{\Gamma(\mathcal{M}_{k,b}+1)}{\varepsilon_{k,b}\cdot\Gamma\left(\mathcal{M}_{k,b}\right)}.
\end{aligned} \tag{46}$$

By substituting Eq. (46) into Eq. (44) and Eq. (45), we can get the closed-form expression of OP for U_1 and U_2 at H_b as Eq. (16) and Eq. (17), respectively.

Appendix C
The relationship of $\alpha_{e_1,b}$ and $\alpha_{e_2,b}$ with $C(\gamma_{e_1,b})$

Since $\frac{\rho_{1,b}\alpha_1\Gamma(\mathcal{M}_{1,b}+1)}{(1+\rho_{2,b}\cdot\alpha_2\frac{\Gamma(\mathcal{M}_{2,b}+1)}{\varepsilon_{2,b}\Gamma(\mathcal{M}_{2,b})})\varepsilon_{1,b}\cdot\Gamma(\mathcal{M}_{1,b})}$ is much larger than 1 at higher SNR $\rho_{1,b}$ of U_1 at H_b, the EC $C(\gamma_{1,b})$ of U_k at H_b can be approximated as

$$\begin{aligned}
C(\gamma_{1,b}) =& \log_2(e)\cdot[1+\frac{\rho_{1,b}\alpha_1\Gamma\left(\mathcal{M}_{1,b}+1\right)}{(1+\rho_{2,b}\alpha_2\frac{\Gamma(\mathcal{M}_{2,b}+1)}{\varepsilon_{2,b}\cdot\Gamma(\mathcal{M}_{2,b})})\varepsilon_{1,b}\Gamma\left(\mathcal{M}_{1,b}\right)}] \\
&-\log_2(e)\cdot\frac{\frac{\Gamma(\mathcal{M}_{1,b})}{\Gamma(\mathcal{M}_{1,b}+1)}-1}{2}.
\end{aligned} \tag{47}$$

Note that $-\frac{\frac{\Gamma\left(\mathcal{M}_{1,b}\right)}{\Gamma\left(\mathcal{M}_{1,b}+1\right)}-1}{2}$ is a constant, we can straightforwardly derive that $\alpha_{1,b}$ is strictly positive with $C(\gamma_{1,b})$ and $\alpha_{2,b}$ is negative with $C(\gamma_{1,b})$.

References

[1] J. Jiao, et al., "Massive access in space-based Internet of Things: Challenges, opportunities, and future directions," *IEEE Wireless Communications*, vol. 28, no. 5, pp. 118–125, 2021.
[2] 3GPP, "Solutions for NR to support non-terrestrial networks (NTN)," TR 38.821 V16.1.0, May. 2021.
[3] FCC, "SpaceX non-geostationary satellite system (attachment a)" [Online]. Available: https://fcc.report/IBFS/SAT-MOD-2018110800083/1569860.pdf, 2018.
[4] FCC, "Application for approval for orbital deployment and operating authority for the spaceX GEN2 NGSO satellite system" [Online]. Available: https://fcc.report/IBFS/SAT-LOA-20200526-00055/2378669, 2020.
[5] FCC, "Technical information to supplement schedule S" [Online]. Available: https://fcc.report/IBFS/SAT-LOA-20200526-00055/2378671, 2020.
[6] 3GPP, "System architecture for the 5G system," TS 23.501 V17.2.0, 2021.
[7] 3GPP, "Service requirements for next generation new services and markets," TS 22.261 V18.4.0, Sept. 2021.
[8] X. Zhu, et al, "Non-orthogonal multiple access based integrated terrestrial-satellite networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2253-2267, 2017.
[9] X. Pang, et al., "Energy-efficient design for mmWave-enabled NOMA-UAV networks." *Sci. China Inf. Sci.*, vol. 64, p. 140303, 2021.
[10] N. Zhao, et al., "Security enhancement for NOMA-UAV networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 3994-4005, 2020.
[11] O. Maraqa, et al, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2192-2235, Fourthquarter 2020.
[12] Y. Wu, et al., "Non-orthogonal multiple access assisted federated learning via wireless power transfer: A cost-efficient approach," *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2853–2869, 2022.
[13] N. Huang, et al., "Integrated sensing and communication assisted mobile edge computing: An energy-efficient design via intelligent reflecting surface,*IEEE Wireless Communications Letters*, early access, 2022.
[14] Y. Wu, et al., "NOMA assisted multi-access mobile edge computing: A joint optimization of computation offloading and time allocation," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12244–12258, Dec. 2018.
[15] J. Jiao, et al., "Intelligent hybrid nonorthogonal multiple access relaying for vehicular networks in 6G," *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14773–14786, 2021.
[16] J. Jiao, et al., "Unequal access latency random access protocol for massive machine-type communications," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 5924–5937, 2020.
[17] X. Liu, et al., "QoS-guarantee resource allocation for multibeam satellite industrial internet of things with NOMA," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2052-2061, 2021.
[18] J. Jiao, et al., "Superimposed pilot code-domain NOMA scheme for satellite-based internet of things," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2732-2743, 2021.
[19] Z. Ding, et al., "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4464-4478, 2019.
[20] A. D. Panagopoulos, et al., "Satellite communications at Ku, Ka, and V bands: propagation impairments and mitigation techniques," *IEEE Communications Surveys and Tutorials*, vol. 6, no. 3, pp. 2-14, 2004.
[21] J. Jiao, et al., "Design and analysis of novel Ka band NOMA uplink relay system for Lunar farside exploration," *China Communications*, vol. 17, no. 7, pp. 1–14, July 2020.
[22] X. Liang, et al., "Outage analysis of multirelay multiuser hybrid satellite-terrestrial millimeter-wave networks," *IEEE Wireless Communications Letters*, vol. 7, no. 6, pp. 1046-1049, 2018.
[23] K. An, et al., "Performance limits of cognitive-uplink FSS and terrestrial FS for Ka-band," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 5, pp. 2604-2611, 2019.
[24] X. Fang, et al., "5G embraces satellites for 6G ubiquitous IoT: basic models for integrated satellite terrestrial networks," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14399-14417, 2021.
[25] I. Kostic, et al., "Analytical approach to performance analysis for channel subject to shadowing and fading," *IEE Proceedings Communications*, vol. 152, no. 6, pp. 821–827, Dec. 2005.
[26] K. An, et al., "On the ergodic capacity of multiple antenna cognitive satellite terrestrial networks," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1-5, 2016.
[27] K. An, et al., "Outage performance for the cognitive broadband satellite system and terrestrial cellular network in millimeter wave scenario," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1-6, 2017.
[28] C. Kourogiorgas, et al., "Cognitive uplink FSS and FS links coexistence in Ka-band: propagation based interference analysis," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 1675-1680, 2015.
[29] J. Zhang, et al., "Deep reinforcement learning for throughput improvement of the uplink grant-free NOMA system," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6369-6379, 2020.
[30] M. Zeng, et al., "Energy-efficient joint user-RB association and power allocation for uplink hybrid NOMA-OMA," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5119-5131, 2019.
[31] L. You, et al., "Massive MIMO transmission for LEO satellite communications," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1851-1865, 2020.

[32] D. Goto, et al., "LEO-MIMO satellite systems for high capacity transmission," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, 2018.
[33] P. Popovski, et al., "5G wireless network slicing for eMBB, URLLC, and mMTC: a communication-theoretic view," *IEEE Access*, vol. 6, pp. 55765-55779, 2018.
[34] Z. Ding, et al., "Unveiling the importance of SIC in NOMA systems-part 1: state of the art and recent findings," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2373-2377, 2020.
[35] M. Shahab, et al., "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Communications Surveys & Tutorials*,vol. 22, no. 3, pp. 1805–1838, 2020.
[36] T. Yang, et al., "Grant free age-optimal random access protocol for satellite-based Internet of Things," *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 3947–3961, 2022.
[37] S. Kim, et al., "A new non-orthogonal transceiver for asynchronous grant-free transmission systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1889-1902, 2021.
[38] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products,* 7th ed. New York, NY, USA: Academic, 2007.
[39] D. B. da Costa, et al., "Capacity analysis of cooperative systems with relay selection in Nakagami-m fading," *IEEE Commun. Lett.*, vol. 13, no. 9, pp. 637-640, 2009.
[40] Z. Ding, et al., "A new QoS-guarantee strategy for NOMA assisted semi-grant-free transmission," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7489-7503, 2021.