IEEE Open Journal of the Communications Society

Received 28 August 2023; revised 18 September 2023; accepted 30 September 2023. Date of publication 12 October 2023; date of current version 6 November 2023.

Digital Object Identifier 10.1109/OJCOMS.2023.3323410

Enhanced Hybrid Automatic Repeat Request Scheduling for Non-Terrestrial IoT Networks

GAUTHAM PRASAD[1], VISHNU RAJENDRAN CHANDRIKA[2], LUTZ LAMPE[3] (Senior Member, IEEE), AND GUS VOS[4] (Senior Member, IEEE)

[1]6G Innovation Lab, Ofinno, Reston, VA 20190, USA
[2]Baseband Systems, Ericsson Canada Inc., Ottawa, ON K2K 2V6, Canada
[3]Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[4]Advance Technology Group, Semtech, Richmond, BC V6V 3A4, Canada

CORRESPONDING AUTHOR: G. PRASAD (e-mail: gautham.prasad@alumni.ubc.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and in part by Sierra Wireless. A part of this paper was presented at the 3GPP TSG-RAN meeting #104e, Jan./Feb., 2021 [1].

ABSTRACT Non-terrestrial networks (NTNs) complement their terrestrial counterparts in enabling ubiquitous connectivity globally by serving unserved and/or underserved areas of the world. Supporting enhanced mobile broadband (eMBB) data over NTNs has been extensively studied in the past. However, focus on massive machine type communication (mMTC) over NTNs is currently growing. Evidence for this are the work items included into the 3rd generation partnership project (3GPP) agenda for commissioning standards for Internet-of-Things (IoT) communications over NTNs. Supporting mMTC in non-terrestrial cellular IoT (C-IoT) networks requires jointly addressing the unique challenges introduced in NTNs and C-IoT communications. In this paper, we tackle one such issue caused due to the extended round-trip time and increased path loss in NTNs resulting in a degraded network throughput. We propose smarter transport blocks scheduling methods that can increase the efficiency of resource utilization. We conduct end-to-end link-level simulations of C-IoT traffic over NTNs. Our numerical results of throughput show the improvement in performance achieved using our proposed solutions against legacy scheduling methods.

INDEX TERMS Non-terrestrial networks (NTN), HARQ scheduling, new radio (NR), machine type communication (MTC), narrowband Internet-of-Things (NB-IoT).

I. BACKGROUND

NON-TERRESTRIAL networks (NTNs), including those enabled by satellites in the low earth orbit (LEO), medium earth orbit (MEO), and geostationary earth orbit (GEO), as well as high-altitude platform stations (HAPS) and other unmanned/uncrewed aerial vehicles (UAVs), complement their conventional terrestrial counterparts in enhancing cellular coverage by serving unserved and/or underserved areas [2], [3]. NTNs are critical in remote areas with low/no cellular connectivity. They are applicable across many different industries, such as transportation (maritime, road, rail, and air) and logistics, farming, mining, utilities, and environment monitoring [4], [5], [6]. As a result, fifth generation (5G) and beyond 5G cellular networks are increasingly focusing on supporting NTNs for both enhanced mobile broadband (eMBB) and low power wide area network applications [7], [8], [9].

Support for C-IoT and massive machine type communication (mMTC) systems in NTNs is critical across several domains. NTN serves C-IoT scenarios in both wide- and local-area IoT services. The former includes gathering data on a macro-level from sensors deployed across broad geographical areas. Automotive support (e.g., enabling over-the-air auto upgrades, vehicle platooning, traffic flow optimization), large-scale infrastructure monitoring in energy distribution systems, and managing livestock and farming in the agriculture industry form the core of wide-area IoT application scenarios [10]. On the other hand, sensors and actuators serving a more confined area, such as on board a maritime vessel or a neighborhood advanced metering

infrastructure in a smart grid sub-system, make up local-area IoT networks [10]. As with eMBB applications, NTN allows for expanding IoT networks, both wide- and local-area services, in a ubiquitous manner and guaranteeing continuity of service across geographical areas [11].

In this paper, we propose methods for enhancing resource scheduling tailored for such C-IoT and mMTC systems to operate in NTNs, which are collectively referred to as *IoT-NTN*. In particular, we focus on the impact of the extended round trip time (RTT) of bidirectional signals caused due to the increased propagation distance in NTNs on the network throughput. We draw inspiration from semi-static and codebook-based scheduling methods for bandwidth-limited low-complexity and coverage enhanced user equipmemts (UEs) in the long term evolution (LTE) standard [12, Cl. 7.3]. Such a method is designed for terrestrial networks to accommodate varying low-complexity UE capabilities, control and shared channel processing delays, and network scheduling constraints. We further build on such a method to design a dynamic data and control channel delay-based scheduling in both the uplink and the downlink to additionally account for varying RTTs experienced in NTN scenarios. Our design is therefore compatible with existing scheduling techniques and suitable for NTN environments. We show in this paper that adapting scheduling delays to changing propagation times caused due to satellite movements can enhance the effective uplink and downlink throughput in an IoT-NTN.

Typically, the achievable throughput may not be the primary target performance indicator for C-IoT and mMTC systems. One of the reasons for this is the greater resource consumption (e.g., extended bandwidth requirement) and higher transmit power associated with achieving increased throughput. However, we show in this work that the network throughput can be improved without demanding additional resources or increasing transmit power, but instead by efficiently scheduling the hybrid automatic repeat request (HARQ) processes. As a result, IoT-NTNs can support a larger number of IoT devices, which is critical in NTNs due to the significantly larger cell size compared to conventional terrestrial networks. Toward this end, we exploit the lengthy RTT in NTN links to overlap bidirectional signals in the air. We further utilize the frequency-division full-duplex nature of base stations (BSs) to allow simultaneous transmission and reception at the satellites and BSs.

A. RELATED WORK

IoT-NTN studies have thus far, rightly, focused predominantly on initial access aspects, such as analysis of link budget [13], [14], challenges of Doppler effects and solutions to counter them [14], [15], [16], and issues related to random access [14], [17], [18], [19]. While system-level scheduling enhancements have been considered in the past [20], [21], link-level adaptations, which we focus on in this work, are largely missing in the literature.

In the broader context of NTN, solutions that have previously been proposed to counter the impact of large propagation delay can be roughly classified into three categories:

A) targeting a low initial block error rate (iBLER),
B) disabling the use of HARQ feedback at the medium access control (MAC) layer, and
C) increasing the number of HARQ processes.

1) LOW TARGET IBLER

The issue of stop-and-wait gaps produced due to the lengthy RTT is recognized in [22], [23] for GEO systems. These works acknowledge that packet errors and the resulting re-transmissions can significantly increase transmission gaps especially for high altitude GEO links and thereby reduce the throughput. While [22] only discusses the open issue of stop-and-wait gaps, [23] proposes a solution to counter this by simply improving robustness, i.e., targeting a low iBLER to avoid re-transmissions. Although using low iBLERs can help in reducing re-transmissions, previous simulation studies have shown that a high target iBLER instead is optimal to reduce the overall UE power consumption. This is because a larger target iBLER demands fewer repetitions and can accommodate a higher coding rate [24]. The study in [24] however only considers terrestrial networks operating with negligible propagation delay. When a large iBLER is used in NTNs, it does result in increased stop-and-wait gaps. Therefore, our solution, aimed precisely at reducing these stop-and-wait gaps with smart scheduling, presents a method to use a high iBLER for battery life optimization for IoT-NTN devices, while at the same time also increasing UE battery life by enabling a higher target iBLER. Nevertheless, note that our proposed solution is also compatible with designs that use a sub-optimal low iBLER transmission, but with reduced throughput gains.

2) DISABLED HARQ

Although link-level scheduling research in IoT-NTN is still in its infancy, the issue has been studied extensively for supporting eMBB applications in NR-NTN. An overview of these methods is reported in [7]. We evaluate if the methods suggested for NR-NTN can be adapted also for IoT-NTN. One solution to solve the issue of stop-and-wait gaps in HARQ transmissions is to disable the use of HARQ feedback [7, Sec. 6], [25]. In this case, re-transmissions can be handled at the radio link control (RLC) layer. However, several problems exist with this technique. First, the benefits of HARQ combining (e.g., Chase combining or incremental redundancy) at the physical (PHY) or MAC layer are absent, and thus, more re-transmissions may be needed, which then results in poor overall spectral efficiency. Furthermore, relying on RLC layer for re-transmission introduces additional user-plane latency and increased jitter when errors occur due to the high re-transmission timeout. This is because, without HARQ feedback, the network relies on the RLC feedback. In this case, the status report, which contains the

acknowledgment (ACK) or negative ACK (NACK) feedback, can be severely delayed. For example, when there is no uplink data to transmit, the UE must send a scheduling request for scheduling uplink data just to send the status report [26]. This can take longer than three times the RTT, depending on the resource configuration. Furthermore, when there are burst errors, the status report can only be sent after a valid packet is received, which also causes increased latency. Additionally, if the last physical data unit is in error, the status report is not sent until the transmitter requests it. A possible solution to counter the delayed status reporting is to let the transmitter request an RLC status report often via the *poll bit* [27]. While this reduces signaling latency, it introduces a large signaling overhead, which may be tolerable for eMBB applications but is significant for IoT-NTN scenarios.

3) INCREASING THE NUMBER OF HARQ PROCESSES

An alternative solution to reduce or eliminate the stop-and-wait gaps is to increase the number of HARQ processes, N_{HARQ}. Increasing N_{HARQ} is associated with two major issues. First, it requires a larger soft buffer size at the receiver. This introduces cost and complexity overheads, and is not suitable for low-cost IoT-NTN devices. Second, implementing a higher number of HARQ processes for half-duplex (HD) frequency division duplexed (FDD) IoT-NTN UEs using state-of-the-art scheduling methods requires the use of multiple transport block grant (MTBG) and ACK bundling. MTBG and ACK bundling increases the system complexity and may not be implemented in all networks, especially those catering to low-cost IoT devices. Furthermore, ACK bundling is only efficient when the iBLER is small. But as discussed earlier in Section I-A1, a lower iBLER results in poorer spectral efficiency, lower speed, and higher UE power consumption. Therefore, increasing the N_{HARQ} with the existing scheduling methods to fill the stop-and-wait gaps is not ideal for IoT-NTN systems. Nevertheless, we show in the following sections that our proposed technique for scheduling IoT NTN TBs not only improves spectral efficiency without necessarily requiring an increase in N_{HARQ}, but also allows the usage of an arbitrary number of parallel HARQ processes. This thereby also allows for increasing N_{HARQ} in the uplink without incurring the two issues described above.

With this backdrop, we propose our flexible HARQ scheduling solutions in the following sections, that do not require disabling HARQ processes or demand any *mandatory* increase in N_{HARQ}.

B. CONTRIBUTIONS

The major contributions and highlights of our paper can be listed as follows:

- We propose HARQ scheduling designs for IoT-NTN systems to increase the uplink and downlink resource utilization efficiency and throughput. To achieve this, we develop flexible downlink data-to-acknowledgment (DD2A) delays and uplink grant-to-data (UG2D) delays for more efficient NTN signaling. We ensure that our solution works for different satellite orbit architectures, and also with the usage of a varied number of HARQ processes.
- We propose methods to signal and calculate DD2A and UG2D delays for dynamic transmission adaptation conditions. We address both scenarios where the number of transport block repetitions are the same and different within a given scheduling interval to render to our solution future-proof.
- We provide a simulation based evaluation for determining the transport block size (TBS) and forward error correction strategy for improving spectral efficiency under varying IoT-NTN link conditions. Using these settings, we also present the evaluation results of our proposed solutions in terms of the achieved throughput gains.

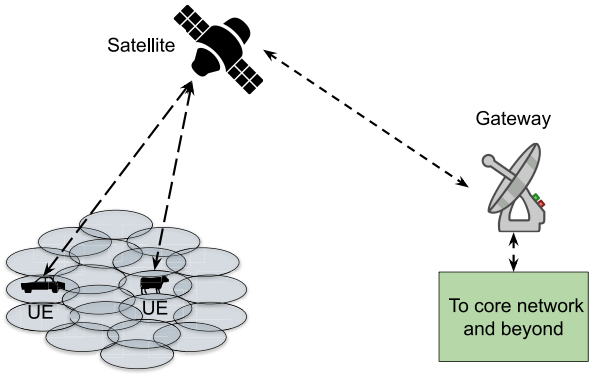

FIGURE 1. An illustration of an IoT-NTN system with terrestrial UEs being served by a non-terrestrial satellite.

C. OUTLINE

The rest of the paper is organized as follows. In Section II, we introduce the system model. We propose our solutions in Section III, which we evaluate and present their performance results in Section IV. We conclude the paper in Section V. We also include a list of important notations and their meanings in Table 1 and a list of acronyms in Table 2.

II. SYSTEM MODEL

We show the overall system model of a typical IoT-NTN in Fig. 1. In this work, we consider the satellite to be in a LEO revolving around the earth in a circular orbit. But our study is directly applicable to higher satellite altitudes of MEO and GEO satellites and also for lower altitudes of HAPS and UAVs. Our study is agnostic to the data processing/forwarding architecture, and thus supports both transparent or bent-pipe payloads and regenerative ones [7], [28]. Our proposed solutions are also applicable to both earth-moving and earth-fixed NTN cell-types [29].

An illustration of our considered system architecture is shown in Fig. 1. The beam footprint of each beam determines the RTT range of the satellite link [7]. We present the

TABLE 1. List of important notations used in the paper.

Notation	Meaning	Notation	Meaning
η_{SUF}	SF utilization factor	n_{DD2A}	Delay between the downlink data and the associated ACK in units of SFs
N_{data}	Number of downlink or uplink data SFs	n_{UG2D}	Delay between the uplink grant and the associated uplink data in units of SFs
N_{rep}	Number of repetitions of TB	N_{DG2D}	Delay between the downlink grant and the associated downlink data in units of SFs
N_{HC}	Lengths of HARQ cycle in units of SFs	$n_{\text{DD2A, min}}$	Minimum mandatory DD2A in units of SFs
R	Throughput	$n_{\text{UG2D, min}}$	Minimum mandatory UG2D in units of SFs
n_{TB}	Transport block size in bits	n_{bundle}	Number of ACKs bundled within one transmission time interval
t_{TB}	Transport block duration in seconds	γ	Signal-to-noise ratio
N_{HARQ}	Number of HARQ processes	β_{PL}	Free space path loss
τ_{RTT}	RTT in seconds	δ	Signal bandwidth
$n_{\text{rep,data}}$	Number of data block repetitions	P_{EIRP}	Effective isotropically radiated power of UE
$n_{\text{rep,PDSCH}}$	Number of PDSCH repetitions	G/T	Satellite antenna-gain-to-noise-temperature
$n_{\text{rep,PDCCH}}$	Number of PDCCH repetitions	k	Boltzmann constant
$n_{\text{rep,PUCCH}}$	Number of PUCCH repetitions	f	Carrier frequency
$n_{\text{rep,PUSCH}}$	Number of PUSCH repetitions	β_{atm}	Atmospheric loss
N_{switch}	Switching delay in units of SFs	β_{shadow}	Shadow fading margin
N_{TBPHC}	Number of TBs scheduled in one HARQ cycle	β_{scint}	Scintillation loss
N_{A2G}	ACK processing delay in units of SFs	β_{polar}	Polarization loss

RTTs of IoT-NTN links for different LEO altitudes for both types of payload architectures in Table 3. The minimum and the maximum RTTs correspond to the maximum (90°) and minimum (10°) beam elevation angles, respectively.

We borrow the uplink and downlink communication mechanism between the UE and BS for IoT-NTN directly from the LTE-MTC (LTE-M) and narrowband IoT (NB-IoT) specifications [12], [26]. Our reasoning behind this is two-fold. First, LTE-M and NB-IoT are the industry chosen standards for enabling low-power wide area networks. Several works have been presented in the past to demonstrate the benefits of using these 3GPP standards, e.g., [30], [31]. The use of LTE-M and NB-IoT provides the advantage of reusing existing infrastructure and operating on licensed bands. These help in providing low-cost, stable, reliable, and predictable performance across application scenarios. Second, similar to using the new radio (NR) standard as the starting point for eMBB over NTN [7, Sec. 6], IoT-NTN standardization activities in 3GPP have agreed to build on the existing LTE-M and NB-IoT standards to expand them into the NTN realm [32]. By showing the effectiveness of our proposed solution for LTE-M and NB-IoT, we aim to demonstrate that integrating our cost-efficient method into legacy systems is practically feasible.

According to both the LTE-M and NB-IoT standards, uplink and downlink data bits are grouped into transport blocks (TBs) of varying sizes for transmission [12]. The TBS is dependent on the adaptive modulation and coding scheme chosen based on the operating conditions and target block error rate (BLER). The transmission time interval transmission time interval, which is the time spanned by one unit of transmission corresponds to one sub-frame (SF) of 1 ms duration, during which one or more TBs are transmitted. A timing diagram of the downlink and uplink transmissions for LTE-M are shown in Figs. 2(a) and (b), respectively. Figs. 2(a) and (b) demonstrate the HARQ-based downlink and uplink communication [33, Ch. 10], which is used in both LTE-M and NB-IoT. The need for using a HARQ-based design in cellular low-power wide area network technologies, such as LTE-M and NB-IoT, has been extensively shown in the past, e.g., [34].

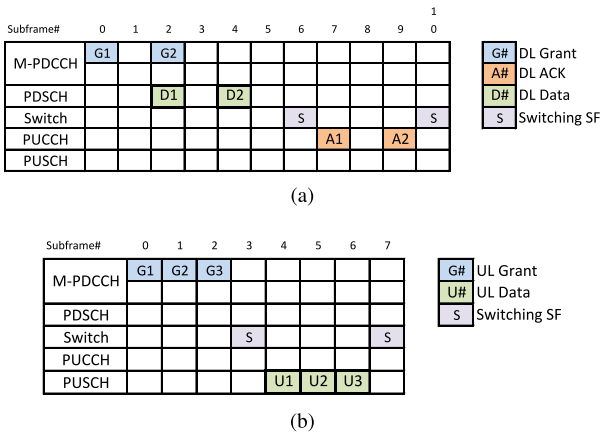

FIGURE 2. Timing diagrams for LTE-M operation in (a) the downlink and (b) the uplink. M-PDCCH: machine-type communication physical downlink control channel; PDSCH: physical downlink shared channel; PUCCH: physical uplink control channel; PUSCH: physical uplink shared channel. In (a), G1 and G2 are the downlink grants scheduling the downlink data D1 and D2, respectively, and A1 and A2 are acknowledgment subframes of D1 and D2, respectively. In (b), G1, G2, G3 are the uplink grants scheduling the uplink data U1, U2, and U3, respectively. S represents a switching subframe in both sub-figures.

Fig. 2(a) illustrates signal flow in the donwlink direction, where a downlink data TB is carried on the physical downlink shared channel (PDSCH). This is preceded by a corresponding single TB grant (STBG) sent on the MTC physical downlink control channel (M-PDCCH). For the case of NB-IoT, the grants are sent on the NB-IoT PDCCH (N-PDCCH). Henceforth, we drop the prefix for brevity and refer to it only as PDCCH. In the example of Fig. 2(a), the BS schedules two downlink TBs via two grants. However, a single grant may also configure multiple TBs, and such grants are referred to as multiple TB grants (MTBGs). For every data TB received, the UE responds with an acknowledgment (ACK) TB, which is typically carried via the physical uplink control channel (PUCCH) as shown in Fig. 2(a). The ACK TB can either acknowledge one or more TBs using unbundled or bundled ACKs, respectively. Uplink transmissions are analogous to the downlink, where uplink data TBs are configured by uplink grants in an MTBG or STBG fashion, where the latter is shown in Fig. 2(b). The data

TABLE 2. List of acronyms used in the paper.

Acronym	Expansion
3GPP	3rd Generation Partnership Project
5G	Fifth Generation
ACK	Acknowledgment
BS	Base Station
C-IoT	Cellular Internet-of-Things
DD2A	Downlink Data-to-Acknowledgment
DG2D	Downlink Grant-to-Data
eMBB	Enhanced Mobile Broadband
FDD	Frequency Division Duplex
GEO	Geostationary Earth Orbit
GNSS	Global Navigation Satellite System
HAPS	High Altitude Platform Station
HARQ	Hybrid Automatic Repeat reQuest
HD	Half Duplex
iBLER	Initial Block Error Rate
IoT	Internet-of-Things
LEO	Low-Earth Orbit
LTE	Long Term Evolution
LTE-M	Long Term Evolution Machine-type Communication
MAC	Medium Access Control
MEO	Medium-Earth Orbit
mMTC	Massive Machine Type Communication
MOPS	Million Operations per Second
M-PDCCH	Machine Type Communication PDCCH
MTBG	Multi Transport Block Grant
MTC	Machine Type Communication
NACK	Negative Acknowledgment
NB-IoT	Narrowband Internet-of-Things
NLOS	Non Line-of-Sight
N-PDCCH	NB-IoT PDCCH
NR	New Radio
NTN	Non-Terrestrial Network
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PHY	Physical Layer
PRB	Physical Resource Block
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
RLC	Radio Link Control
RTT	Round Trip Time
RX	Receive
SF	Sub-frame
SNR	Signal-to-Noise Ratio
STBG	Single Transport Block Grant
SUF	Sub-frame Utilization Factor
TB	Transport Block
TBS	Transport Block Size
TDL	Tapped Delay Line
TX	Transmit
UAV	Uncrewed Aerial Vehicle
UE	User Equipment
UG2D	Uplink Grant-to-Data

TABLE 3. RTTs for LEO NTNs.

Satellite Altitude	Payload Type	Min. RTT (ms)	Max. RTT (ms)
600 km	Regenerative	4	13
600 km	Transparent	8	26
1200 km	Regenerative	8	21
1200 km	Transparent	16	42

TBs are transmitted on the physical uplink shared channel (PUSCH) by the UE. For this work, we consider low-cost IoT UEs that operate in an HD-FDD manner, which is the industry preferred design for cost and complexity reduction [35], [36], [37]. As a result, the UE uses one or more SFs to switch between transmission and reception modes.

The timing diagrams shown in Fig. 2 are nearly identical at both the BS and UE for terrestrial networks due to the negligible signal propagation delay. However, in NTNs, the timing shown in Fig. 2 is valid at the UE, while the downlink transmissions are sent in advance from the BS by half the RTT. Similarly, the uplink packets reach the BS also after half the RTT. We show an example of this in the timing diagram of Fig. 3. A grant transmitted on SF#0 by the BS is received by the UE on SF#8 due to the lengthy propagation delay of 8 ms, which is typical of LEO satellite setups. Similar delays in the reception of both downlink and uplink TBs on all control and shared channels can be seen in Fig. 3. We also notice that the overhead caused by the processing delays and the time of flight between the UE and BS results in several idle SFs at the UE-end, which significantly reduce the achievable throughput.

To quantify the throughput obtainable, we define the concepts of HARQ cycle and SF utilization factor (SUF). We define one HARQ cycle as the total duration of time during which the UE receives on the PDCCH and/or PDSCH and the duration of time it transmits on PUCCH or PUSCH. A HARQ cycle also includes the switching SF(s). For example, Fig. 2(a) and Fig. 2(b) show one HARQ cycle for downlink and uplink, respectively, where the lengths of a HARQ cycle, N_{HC}, in units of SFs are $N_{\text{HC}} = 11$ and $N_{\text{HC}} = 8$, respectively.

Next, we define the SUF, η_{SUF}, as

$$\eta_{\text{SUF}} = \frac{N_{\text{data}}}{n_{\text{rep}} N_{\text{HC}}}, \tag{1}$$

where N_{data} is the total number of SFs occupied by uplink or downlink data TBs and n_{rep} is the number of repetitions used for each TB, i.e., the number of times each TB is repeated including its first transmission. Note that the TB repetitions introduce redundancy to improve the error rates, and are not the same as HARQ re-transmissions. The value of n_{rep} is chosen based on the adaptive modulation and coding scheme used by the UE. HARQ re-transmissions, on the other hand, are repeated transmissions of one or more repetitions of a TB when the TB is not successfully acknowledged. A detailed explanation of the HARQ operation can be found extensively in the literature, e.g., [33, Ch. 10].

Eq. (1) shows that SUF indicates the proportion of time spent on transmitting the payload as opposed to transmission overheads. We therefore compute the useful throughput, R, as

$$R = \eta_{\text{SUF}} \frac{n_{\text{TB}}}{t_{\text{TB}}}, \tag{2}$$

where n_{TB} and t_{TB} are the TBS in bits and time spanned by one TB, respectively. Therefore, for a given TBS corresponding to the adaptive modulation and coding scheme chosen, we can maximize the throughput by maximizing the SUF, i.e., by reducing the transmission overheads. Since higher values of RTT can increase N_{HC}, it can be seen from (1) that it thus deteriorates SUF and the useful throughput.

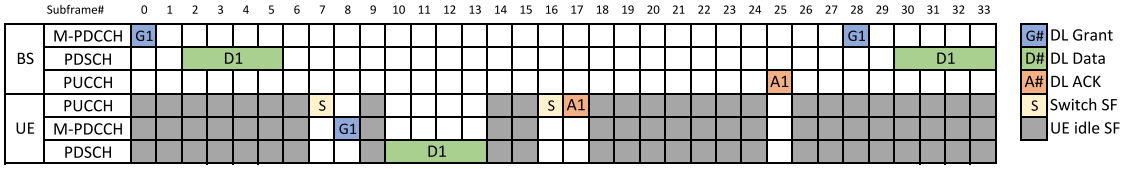

FIGURE 3. Timing diagram for downlink operation in NTN. BS: base station; UE: user equipment; M-PDCCH: machine-type communication physical downlink control channel; PDSCH: physical downlink shared channel; PUCCH: physical uplink control channel. G1 is the downlink grant scheduling the downlink data D1. D1 is repeated four times and therefore spans four subframes. A1 is the acknowledgment subframe of D1. S represents a switching subframe. Subframes shaded grey represent occasions where the UE is neither transmitting nor receiving, and therefore contributing to a reduction of the throughput.

III. PROPOSED SOLUTIONS

As discussed in Section II, the long RTT in NTN causes lengthy stop-and-wait gaps. The value of N_{HARQ} required to fill up the stop-and-wait gaps in its entirety is

$$N_{\text{HARQ}} \geq \frac{\tau_{\text{RTT}}}{t_{\text{TB}}}, \tag{3}$$

where τ_{RTT} is the RTT in seconds. With one TB spanning 1 SF, we can observe from Table 3 that N_{HARQ} must be increased to up to 42 under typical operating conditions to satisfy (3). The current LTE-M and NB-IoT specifications only allow a maximum of 8 and 2 HARQs, respectively [12], [26]. This increase in N_{HARQ} is impractical for low cost mMTC and C-IoT UEs, where a high N_{HARQ} introduces increased complexity due to the required size of the soft-buffer. Furthermore, it also demands corresponding increments in the size of the downlink control information bits.

For the case of IoT-NTN, which uses few physical resource blocks (PRBs) and operates under high path loss environments, multiple repetitions of the TB are often required to achieve reliable communication. Therefore, the required N_{HARQ} is [1]

$$N_{\text{HARQ}} \geq \frac{\tau_{\text{RTT}}}{n_{\text{rep,data}} t_{\text{TB}}}, \tag{4}$$

where $n_{\text{rep,data}}$ is the number of data block repetitions, i.e., $n_{\text{rep,data}} = n_{\text{rep,PDSCH}}$ or $n_{\text{rep,data}} = n_{\text{rep,PUSCH}}$, for indicating repetitions on the PDSCH or the PUSCH, respectively. In Section IV, we present the results from a comprehensive link-level simulation evaluation to show that IoT-NTN does not require any further increase in N_{HARQ} for MTC applications, and up to $N_{\text{HARQ}} = 4$ for NB-IoT applications to achieve reliable communications while also filling the stop-and-wait gaps introduced due to the increased RTT.

A. THE PROBLEM WITH EXISTING TRANSPORT BLOCK SCHEDULING ALGORITHMS

The issue with the current HARQ implementation is that the delay between the downlink data and the corresponding ACK, n_{DD2A}, and similarly in the uplink, the delay between the uplink grant and uplink data, n_{UG2D}, are fixed. Therefore,

$$i_{\text{ACK},j} = i_{\text{DLdata},j} + n_{\text{DD2A}} + 1, \quad \text{for downlink} \tag{5}$$

$$i_{\text{ULdata},j} = i_{\text{UG},j} + n_{\text{UG2D}} + 1, \quad \text{for uplink}, \tag{6}$$

where $i_{\text{ACK},j}$, $i_{\text{DLdata},j}$, $i_{\text{ULdata},j}$, and $i_{\text{UG},j}$ are the positions in SFs of the ACK for the jth TB, the last SF of the jth downlink TB, the first SF of the jth uplink TB, and the last SF of the jth UG, respectively. When $n_{\text{rep}} > n_{\text{DD2A}}$, $i_{\text{ACK},j}$ overlaps with $i_{\text{DLdata},k}$ for $k > j$ or the switching SF. Similarly, when $n_{\text{rep}} > n_{\text{UG2D}}$, $i_{\text{DLdata},k}$ for $k > j$ overlaps with $i_{\text{ULdata},j}$. We show an example of this issue in the downlink in Fig. 4, with $n_{\text{rep}} = 4$ and $n_{\text{DD2A}} = 3$. We notice that $i_{\text{ACK},1}$ on the PUCCH overlaps with $i_{\text{DLdata},2}$ on the PDSCH, which is inoperable for HD-FDD UEs. While Fig. 4 shows the downlink operation, similar overlap also occurs on the uplink between PDCCH and PUSCH due to conflict between the uplink data transmission and downlink grant reception SFs.

Thus, the current HARQ configuration can only support one TB per HARQ cycle when $n_{\text{rep}} > n_{\text{DD2A}}, n_{\text{UG2D}}$. We can re-write (1) under such an operation as

$$\eta_{\text{SUF,DL}} = \big(n_{\text{rep,PDCCH}} + n_{\text{rep,PDSCH}} + n_{\text{rep,PUCCH}} + N_{\text{DG2D}} + n_{\text{DD2A}} + N_{\text{switch}}\big)^{-1} \tag{7}$$

and

$$\eta_{\text{SUF,UL}} = \big(n_{\text{rep,PDCCH}} + n_{\text{rep,PUSCH}} + n_{\text{UG2D}} + N_{\text{switch}}\big)^{-1}, \tag{8}$$

for downlink and uplink, respectively, where $n_{\text{rep,PDCCH}}$ and $n_{\text{rep,PUCCH}}$ represent the number of repetitions of grant on PDCCH and ACK on PUCCH, respectively, N_{DG2D}, n_{DD2A}, and n_{UG2D} and the processing delays between downlink grant and downlink data, downlink data and ACK, and uplink grant and uplink data, respectively, and N_{switch} is the switching delay, all in units of SFs.

We show in Section IV that $n_{\text{rep,PUSCH}} > n_{\text{DD2A}}, n_{\text{UG2D}}$ always holds true, especially with the delay values specified in the legacy LTE-M protocol. Therefore, irrespective of the N_{HARQ} that can be used, only one TB can be scheduled per HARQ cycle. We refer back to the example of NTN downlink HARQ transmissions with fixed DD2A and a time of flight of 8 SF units in Fig. 3. In this example, we use a fixed DD2A of 3 SFs, 4 SFs of data repetitions, and 1 HARQ process. We notice that the processing delays and the time of flight between the UE and BS result in several idle SFs in the UE due to the single TB in one HARQ cycle.

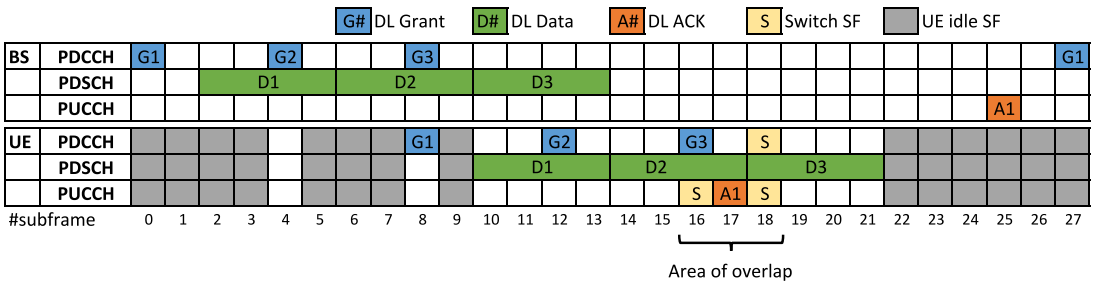

FIGURE 4. An illustration of the scheduling conflict with TB overlap using the current-day method. BS: base station; UE: user equipment; PDCCH: physical downlink control channel; PDSCH: physical downlink shared channel; PUCCH: physical uplink control channel; G1, G2, and G3 are the downlink grants scheduling the downlink data D1, D2, and D3, respectively. D1, D2, and D3 are each repeated four times and therefore span four subframes each. A1 is the acknowledgment subframe of D1. S represents a switching subframe. Subframes shaded grey represent occasions where the UE is neither transmitting nor receiving, and therefore contributing to a reduction of the throughput. The *area of overlap* represents the subframes where the HD-FDD UE is erroneously expected to transmit on PUCCH and simultaneously receive on PDCCH and PDSCH, which renders this scheduling scenario inoperable.

B. THE PROPOSED SOLUTION

When multiple TBs are scheduled on one HARQ cycle, the burden of additional delays can be spread over those multiple TBs. This reduces the overall overheads per TB and improves the SUF. It is even more beneficial in the context of NTN, since the over-the-air travel time can be exploited to overlap the bidirectional uplink and downlink signals in air. To this end, we propose using flexible values of n_{DD2A} and n_{UG2D} to accommodate multiple TBs per HARQ cycle.

1) THROUGHPUT AND TB PER HARQ CYCLE COMPUTATION

Our proposal of using TB-specific DD2A and UG2D delays,[1] i.e., $n_{\mathrm{DD2A},j}$ and $n_{\mathrm{UG2D},j}$, eliminates the overlap between the data TBs and/or the switching SFs. Thereby, we ensure that an arbitrary number of TBs can be scheduled within one HARQ cycle. The SUF in this case can be expressed as in (9) and (10), shown at the bottom of the page, respectively, where N_{TBPHC} is the number of TBs scheduled in one HARQ cycle and $n_{\mathrm{DD2A,\,min}}$ and $n_{\mathrm{UG2D,\,min}}$ are the minimum mandatory DD2A and UG2D delays that must be used to ensure that sufficient processing time is available at the UE to process downlink data and uplink grants, respectively. We then use $\eta_{\mathrm{SUF,DL}}$ and $\eta_{\mathrm{SUF,UL}}$ from (9) and (10), respectively, to derive the downlink and uplink throughput as

$$R_{\mathrm{DL}} = \eta_{\mathrm{SUF,DL}} \frac{n_{\mathrm{TB}}}{t_{\mathrm{TB}}}, \tag{11}$$

$$R_{\mathrm{UL}} = \eta_{\mathrm{SUF,UL}} \frac{n_{\mathrm{TB}}}{t_{\mathrm{TB}}}, \tag{12}$$

respectively.

While any $N_{\mathrm{TBPHC}} > 1$ can be chosen to obtain higher $\eta_{\mathrm{SUF,DL}}$ and $\eta_{\mathrm{SUF,UL}}$ (and consequently, greater R_{DL} and R_{UL}), the limit on the maximum N_{TBPHC} that can be scheduled is set by the maximum N_{HARQ} supported by the standard. The relation between N_{TBPHC} and N_{HARQ} is dependent on the RTT of the network and can be expressed as in (13), shown at the bottom of the page, where N_{A2G} is the ACK processing delay at the BS before scheduling the HARQ process whose TB is acknowledged.

2) MANAGEMENT OF THE VARIABLE DELAYS

Next, we present the methods to compute, choose, and signal the variable delays from the BS to the UE. Toward this end, we consider three different scenarios of HARQ scheduling. The use of MTBG and ACK bundling may not be considered together due to their complexity, and has also already been investigated for delay flexibility in terrestrial networks [38] (and can therefore be extended to NTNs if needed). Therefore, we consider the three other cases of *grouping* grants and ACKs, namely, MTBG without ACK bundling, STBG without ACK bundling, and STBG with ACK bundling.

1. DD2A delay can also be considered as a PDSCH to PUCCH delay, and UG2D delay can also be considered as a PDCCH to PUSCH delay.

$$\eta_{\mathrm{SUF,DL}} = \frac{N_{\mathrm{TBPHC}}}{n_{\mathrm{rep,PDCCH}} + N_{\mathrm{DG2D}} + N_{\mathrm{TBPHC}} n_{\mathrm{rep,PDSCH}} + n_{\mathrm{rep,PUCCH}} + \max\left(n_{\mathrm{DD2A,\,min}}, (N_{\mathrm{TBPHC}} - 1) n_{\mathrm{rep,PUCCH}}\right) + 2N_{\mathrm{switch}}} \tag{9}$$

$$\eta_{\mathrm{SUF,UL}} = \frac{N_{\mathrm{TBPHC}}}{n_{\mathrm{rep,PDCCH}} + \max\left(n_{\mathrm{UG2D,\,min}}, (N_{\mathrm{TBPHC}} - 1) n_{\mathrm{rep,PDCCH}}\right) + N_{\mathrm{TBPHC}} n_{\mathrm{rep,PDSCH}} + 2N_{\mathrm{switch}}} \tag{10}$$

$$N_{\mathrm{HARQ}} = \left\lceil N_{\mathrm{TBPHC}} \left(1 + \frac{\tau_{\mathrm{RTT}} + N_{\mathrm{A2G}}}{t_{\mathrm{TB}} \left(n_{\mathrm{rep,PDCCH}} + N_{\mathrm{DG2D}} + N_{\mathrm{TBPHC}} \left(n_{\mathrm{rep,PDSCH}} + n_{\mathrm{rep,PUCCH}}\right) + 2N_{\mathrm{switch}}\right)}\right)\right\rceil \tag{13}$$

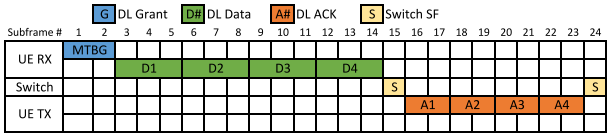

FIGURE 5. Timing diagram for downlink operation with MTBG at the UE-end. UE: user equipment; RX: reception; TX: transmission; MTBG is the multiple transport block grant scheduling the downlink data D1, D2, D3, and D4, respectively. D1, D2, D3, and D4 are each repeated three times and therefore span three subframes each. A1, A2, A3, and A4 are the acknowledgment subframes of D1, D2, D3, and D4, respectively. A1, A2, A3, and A4 are each repeated two times and hence span two subframes each. S represents a switching subframe.

No ACK Bundling: An example of MTBG without ACK bundling for downlink transmission is shown in Fig. 5. The MTBG on SFs#1 − 2 schedules all the following four downlink TBs in SFs#3 − 14. When STBG is used instead, the grants are split to schedule each TB individually. The variable DD2A of the jth TB is the sum of the time left for the remaining TBs scheduled in the MTBG, the ACKs corresponding to all TBs from $1, 2, \ldots, j-1$, and the switching SFs. Therefore,

$$n_{\mathrm{DD2A},j} = (N_{\mathrm{TBPHC}} - j)n_{\mathrm{rep,PDSCH}} + (j-1)n_{\mathrm{rep,PUCCH}} + N_{\mathrm{switch}}. \tag{14}$$

All the parameters used in (14) are already available to the UE to compute $n_{\mathrm{DD2A},j}$ since N_{TBPHC}, $n_{\mathrm{rep,PDSCH}}$, and j are extracted from the MTBG and $n_{\mathrm{rep,PUCCH}}$ is a radio resource control (RRC) configured parameter. Therefore, the UE requires no additional signaling from the BS to compute $n_{\mathrm{DD2A},j}$. However, with the use of STBG, $n_{\mathrm{rep,PDSCH}}$ is conveyed by legacy grants, while the identifier j is to be signaled by the BS when scheduling the jth TB. Similarly, N_{TBPHC} can also be explicitly signaled to the UE if the BS chooses to configure an N_{TBPHC} corresponding to an N_{HARQ} that is lower than the supported maximum value. The same principle can also be applied in the uplink to obtain the variable UG2D delay as

$$n_{\mathrm{UG2D},j} = (N_{\mathrm{TBPHC}} - j)n_{\mathrm{rep,PDCCH}} + (j-1)n_{\mathrm{rep,PUSCH}} + N_{\mathrm{switch}}. \tag{15}$$

Similar to the case of downlink, all parameters required to compute (15) is acquired by the UE as in the case of downlink transmission.

With ACK Bundling: The condition of ACK bundling is only applicable in the downlink since there is no notion of acknowledgment in the uplink. We show an example timing diagram in Fig. 6, where the acknowledgment for four TBs are bundled together into two transmission time intervals. The bundled ACKs1-4 together carry the HARQ-ACK information, either via PUCCH or PUSCH, for all the preceding four TBs D1-D4. The number of transmission time intervals to bundle the plurality of ACKs may be up to the network implementation. When the ACKs are bundled into a single SF, (14) can be modified as

$$n_{\mathrm{DD2A},j} = (N_{\mathrm{TBPHC}} - j)n_{\mathrm{rep,PDSCH}} + \left\lfloor \frac{j-1}{n_{\mathrm{bundle}}} \right\rfloor n_{\mathrm{rep,PUCCH}} + N_{\mathrm{switch}}, \tag{16}$$

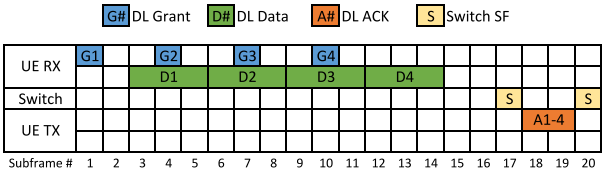

FIGURE 6. Timing diagram for downlink operation with ACK bundling at the UE-end. UE: user equipment; RX: reception; TX: transmission; MTBG is the multiple transport block grant scheduling the downlink data D1, D2, D3, and D4, respectively. D1, D2, D3, and D4 are each repeated three times and therefore span three subframes each. A1-4 represents the bundled acknowledgment carrying the acknowledgment for D1, D2, D3, and D4. A1-4 is repeated two times and hence spans two subframes. S represents a switching subframe.

where n_{bundle} is the number of ACKs that are bundled within one transmission time interval. n_{bundle} can be learned by the UE during RRC configuration and hence requires no signaling overhead.

The above analyses assume that the number of repetitions used by all TBs within a HARQ cycle is the same, as is the case in current LTE-M and NB-IoT operations. However, the BS may choose to use different modulation and coding schemes for each TB based on the type of data being transmitted. While this method is currently not supported in LTE-M or NB-IoT specifications, we address this condition to make our solution future-proof when such an adaptive transmission technique may be implemented in the future. In this case, $n_{\mathrm{rep,PDSCH}}$ and $n_{\mathrm{rep,PUSCH}}$ of each of the remaining or previous TBs are required to compute the DD2A and UG2D delays, respectively. Therefore, (14), (15), and (16) can be modified as

$$n_{\mathrm{DD2A},j} = \sum_{k=j}^{N_{\mathrm{TBPHC}}} n_{\mathrm{rep,PDSCH},k} + (j-1)n_{\mathrm{rep,PUCCH}} + N_{\mathrm{switch}}, \tag{17}$$

$$n_{\mathrm{UG2D},j} = (N_{\mathrm{TBPHC}} - j)n_{\mathrm{rep,PDCCH}} + \sum_{k=1}^{j-1} n_{\mathrm{rep,PUSCH},k} + N_{\mathrm{switch}}, \tag{18}$$

$$n_{\mathrm{DD2A},j} = \sum_{k=j}^{N_{\mathrm{TBPHC}}} n_{\mathrm{rep,PDSCH},k} + \left\lfloor \frac{j-1}{n_{\mathrm{bundle}}} \right\rfloor n_{\mathrm{rep,PUCCH}} + N_{\mathrm{switch}}, \tag{19}$$

respectively, where $n_{\mathrm{rep,PDSCH},k}$ and $n_{\mathrm{rep,PUSCH},k}$ are the number of repeats on the PDSCH and PUSCH for the kth TB, respectively.

IV. RESULTS

In this section, we present numerical results of the increase in throughput achieved with the use of our proposed solutions. The primary reason for the throughput gains obtained by our method can be attributed to the higher number of TBs that can be supported in one HARQ cycle by our variable delay design. However, we recognize that increasing the number of TBs in one HARQ cycle and the consequent possible use of an increased N_{HARQ} (for the case of NB-IoT, but not for LTE-M) specifically to fill the entire RTT results

in a higher complexity at the receiver due to a larger size of the soft-buffer. This condition is typically not preferable when the receiver is the UE (i.e., for downlink communications). On the other hand, a higher receiver complexity introduced by the increased buffer size is negligible in the uplink where the BS receives the TBs. Hence, considering practical implementation scenarios, we focus on the uplink to evaluate the throughput gains achievable using our proposed methods. Note that an increase in N_{HARQ} is not mandatory for the operation of our proposed methods, and performance improvements using our solution can be achieved regardless of the increase.

From (9)–(19), we notice that the throughput achievable with the use of our method relies on the number of data TB repetitions. To this end, we first begin our evaluation campaign with a link-level simulation of a point-to-point IoT-NTN uplink path to determine a suitable value of $n_{\text{rep,PUSCH}}$ required to achieve a target BLER.

A. SIMULATION SETTINGS

We conduct our simulations in a MATLAB environment using the standard-compliant 5G toolbox to support our link-level simulation. We derive a majority of our simulation settings from the relevant 3GPP technical reports and technical documents related to NTN [7], [10], [39]. As suggested by 3GPP, we consider the NTN UEs to be enabled with a global navigation satellite system (GNSS) ability such that it can perform pre- and post-compensation of the frequency offset [40]. Therefore, we apply a maximum residual frequency offset of 34 Hz after post-compensation at the satellite and pre-compensation along with continuous frequency tracking at the UE [41]. We use the 3GPP recommended NTN tapped delay line (TDL) non-line-of-sight (NLOS) channel model to investigate the performance of our solution. We scale the power delay parameters of the reference TDL-A model [10] according to the desired value of delay spread specified for suburban environment [7], [42]. We choose two different values of $n_{\text{TB}} = \{144, 504\}$ to investigate the impact of TBS on BLER and spectral efficiency. We list all simulation parameters in Table 4.

TABLE 4. Evaluation settings.

Parameter	Value
n_{TB}	144, 504 bits
Modulation	QPSK
No. of PRBs	1
SFs for channel estimation	5
No. of HARQ processes	1
Channel model	NTN TDL-A
UE environment	Suburban NLOS
No. of transmit antenna	1
No. of receive antenna	2
UE Speed	15 km/h
UE-satellite elevation angle	30 degrees
Feeder-satellite elevation angle	10 degrees
Residual frequency offset	Uniformly distributed in $[-34$ Hz, 34 Hz$]$
f	2 GHz
Target BLER	10%
P_{EIRP}	23 dBm
G/T	-4.9 dB/T
δ	180 kHz
β_{atm}	0.07 dB
β_{shadow}	3 dB
β_{scint}	2.2 dB
β_{polar}	0 dB
No. of Monte Carlo runs	$10,000$
n_{UG2D} for LTE-M fixed delay	3
n_{UG2D} for NB-IoT fixed delay	8

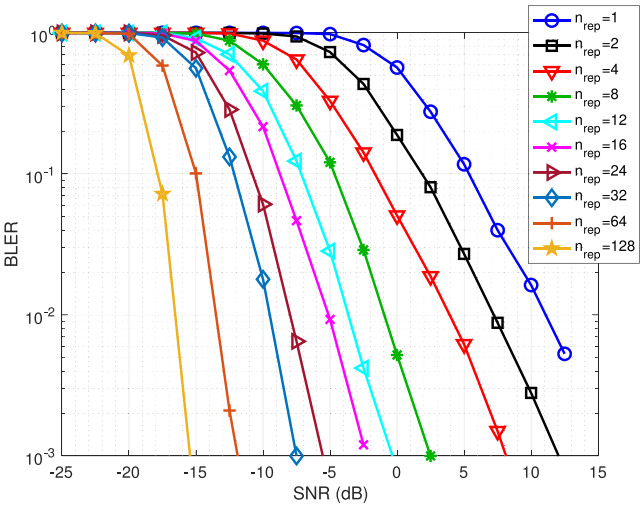

FIGURE 7. BLER vs SNR for different repetitions (n_{rep}) of PUSCH with TBS of 144 bits in NTN TDL-A channel.

B. NUMERICAL RESULTS

We begin by presenting our link-level simulation results.

1) NUMBER OF REPETITIONS

In Figs. 7 and 8, we present the results of the variation of BLER for different possible signal-to-noise ratio (SNR) values for TDL-A channel. Based on our desired operating conditions (i.e., SNRs), we use the results in Figs. 7 and 8 to choose the required number of repetitions to achieve a target BLER. We then use these numbers to present the throughput gains of our proposed solutions in the following.

We find from Figs. 7 and 8 that for a fixed target BLER, choosing a higher value of n_{TB} provides better spectral efficiency. For example, for a target BLER of 10% at the operating value of $\gamma = -5.6$ dB for LEO1200, $n_{\text{rep,PUSCH}} = 12$ and $n_{\text{rep,PUSCH}} = 24$ for $n_{\text{TB}} = 144$ and $n_{\text{TB}} = 504$, respectively. This results in a spectral efficiency of 12 and 21 bits/PRB for $n_{\text{TB}} = 144$ and $n_{\text{TB}} = 504$, respectively. This phenomenon of higher spectral efficiency for larger TBS is also true across satellite access types (i.e., LEO, MEO, GEO) and target BLERs. Therefore, we perform our throughput gain evaluation for our proposed method in Section IV-B2 with $n_{\text{TB}} = 504$. We extract the corresponding numbers for n_{rep} for both satellite access types at their operating SNR values from Fig. 8 as $n_{\text{rep}} = 12$ and $n_{\text{rep}} = 24$ for LEO600 and LEO1200 scenarios, respectively. Furthermore, the results also clearly show that $n_{\text{rep,PUSCH}}$ for

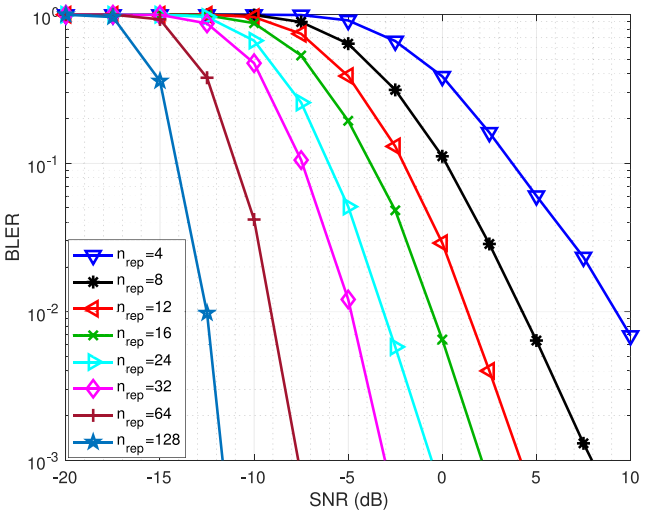

FIGURE 8. BLER vs SNR for different repetitions (n_{rep}) of PUSCH with TBS of 504 bits in NTN TDL-A channel.

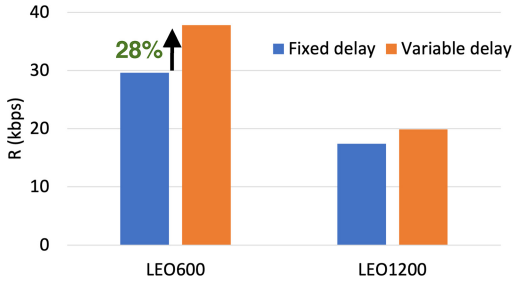

FIGURE 9. Throughput for LEO600 and LEO1200 for the conventional fixed delay method and the proposed variable delay design in an LTE-M UE.

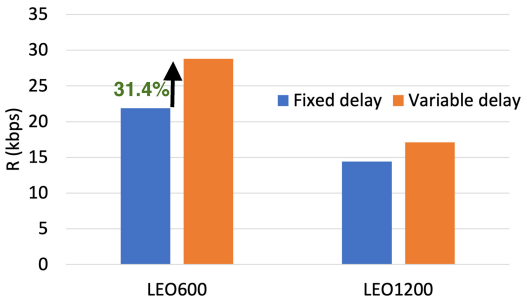

FIGURE 10. Throughput for LEO600 and LEO1200 for the conventional fixed delay method and the proposed variable delay design in an NB-IoT UE.

both LEO600 and LEO1200 cases is greater than n_{DD2A} and n_{UG2D}. As demonstrated in Section III, this condition results in the PUCCH SFs overlapping the PDSCH time slots in the downlink and PDCCH SFs overlapping with PUSCH in the uplink, respectively, for the case of fixed DD2A and UG2D methods of the state-of-the-art.

2) SPECTRAL EFFICIENCY AND THROUGHPUT GAIN

We focus on the uplink throughput gain provided by the use of our variable delay methods using both LTE-M and NB-IoT based UEs. We consider the two LEO satellite access types of LEO600 and LEO1200 that consist of satellites at altitudes of 600 km and 1200 km, respectively. We use an elevation angle of 30 degrees, which provides $\tau_{\text{RTT}} = 20$ ms and $\tau_{\text{RTT}} = 34$ ms for LEO600 and LEO1200 scenarios, respectively. Next, we compute the operating SNR condition to determine a suitable value of n_{rep} to use for evaluating our proposed solutions. We compute the SNR, γ, as

$$\gamma = \frac{P_{\text{EIRP}} G}{kT\beta_{\text{PL}}\beta_{\text{atm}}\beta_{\text{shadow}}\beta_{\text{scint}}\beta_{\text{polar}}\delta}, \tag{20}$$

where β_{PL} is the free space path loss given by [10]

$$\beta_{\text{PL}} = 10^{3.245+\log_{10}(f^2)+\log_{10}(d^2)}, \tag{21}$$

δ is the signal bandwidth, P_{EIRP} is the effective isotropically radiated power from the NTN UE, G/T is the antenna-gain-to-noise-temperature value of the satellite antenna, k is the Boltzmann constant, f is the carrier frequency in GHz, d is the distance between the UE and the satellite, and β_{atm}, β_{shadow}, β_{scint}, β_{polar} are the atmospheric loss, shadow fading margin, scintillation loss, and polarization loss, respectively. We compute d based on the satellite altitude and the elevation angle listed in Table 4. Using the values from Table 4, we obtain $\gamma = -0.2$ dB and $\gamma = -5.6$ dB for LEO600 and LEO1200 scenarios, respectively. These conditions are also consistent with the suggested link budget evaluations presented for NR-NTN in [7].

Next, we compute the throughput using (10) and (12). We begin with LTE-M systems, where the state-of-the-art method uses a fixed $n_{\text{UG2D}} = 3$ and one switching SF. We compare the achievable throughput for the state-of-the-art technique using a fixed delay scheduling method and our proposed solution with variable n_{UG2D} in Fig. 9. The results demonstrate that we achieve a 28% increase in throughput using our proposed method for LEO600 satellite constellation altitude. For the case of NB-IoT UEs, whose results are shown in Fig. 10, we observe an even higher increase in throughput of over 31% for LEO600 with the use of $n_{\text{UG2D}} = 8$ for the fixed delay design and two SFs allotted for switching the UE between transmission and reception. Note that the results for NB-IoT were with an increased $N_{\text{HARQ}} = 4$, whereas the LTE-M system evaluations were with the current allowed maximum of $N_{\text{HARQ}} = 8$.

We further observe in both Fig. 9 and Fig. 10 that the throughput gains obtained with the use of our method is higher for lower satellite altitudes. This is because the value of n_{rep} required to achieve a target BLER increases with higher satellite altitudes due to the increase in pathloss that demands lower code rates. As a result, the increase in throughput obtained with greater number of TBs per HARQ cycle crosses a point of diminishing returns. This result is also intuitive, since a higher number of repetitions can cover a larger portion of the propagation time delay and therefore presents smaller stop-and-wait gaps to be filled with the use of a larger number of TBs per HARQ cycle. Consequently, this reduces the amount of throughput gain that is achievable

from our proposed solution. Nevertheless, for each of these cases, our method provides a superior value of R by reducing the transmission overheads associated with HARQ-based communications.

3) RUN-TIME COMPLEXITY AND POWER CONSUMPTION

In our final evaluation portion, we present computational complexity results associated with the use of our solution. In particular, we show the run-time complexity and the power consumption resulting from our solution. To this end, we only focus on the computations performed at the battery-powered UE side.

The additional computations at the UE-end associated with the deployment of our solution include calculating the variable delays shown in Section III. We first determine the number of operations involved with computing DD2A and UG2D. We then use a worst-case assumption that the computations are performed at every SF level, i.e., once every 1 ms. Note that this is an exaggerated case. In practice, computations at the UE-side need to be performed only when there is a noticeable change in the RTT, for example, when the BS uses a different N_{TBPHC} in response to the RTT variation. Previous investigations have shown that the RTT varies at a rate of less than 100 μs/s for a LEO satellite at an altitude of 600 km and a UE that is moving at a speed of 1200 km/h in the opposite direction of the satellite movement [43]. This results in an RTT variation of less than a nanosecond per SF.

We compute the additional power consumptions for each of our adaptive delay computation methods (i.e., DD2A, UG2D, and DD2A with bundled ACK) as

$$P_\phi = R_{\text{ops}} \frac{\sigma_\phi}{\eta_{\text{proc}}}, \tag{22}$$

where P_ϕ is the power consumption associated with computing the variable delays for a scheme represented by $\phi \in \{\text{DD2A, UG2D, DD2A-bundled ACK}\}$, σ_ϕ is the number of operations (i.e., additions, subtractions, multiplications, divisions, and floors) in millions determined from (14), (15), and (16) for ϕ = DD2A, ϕ = UG2D, and ϕ = DD2A-bundled ACK, respectively, R_{ops} is the rate of computations, which we set as once every 1 ms as described previously, and η_{proc} is the processor efficiency in million operations per second (MOPS) per mW.

We first present the absolute rate of computations (i.e., $R_{\text{ops}}\sigma_\phi$) in MOPS in Table 5. Note that for the case of bundled ACK (denoted as DD2A-BA in Table 5), we incur an extra division and floor operation every SF to determine the variable delay using (19).

We present the results of P_ϕ in Fig. 11 for different values of η_{proc} [44, Ch. 5]. The results demonstrate that our method introduces less than 60 nW of additional power, even with a relatively less efficient processor that provides an efficiency of 144 MOPS per mW. With a superior processor efficiency, e.g., 970 MOPS/mW, power consumption of our proposed solution drops to less than 7 nW to determine the variable delays.

TABLE 5. MOPS computation.

Operation	No. of Operations		
	DD2A	UG2D	DD2A-BA
Addition	2	2	2
Subtraction	2	2	2
Multiplication	2	2	2
Division	0	0	1
Floor	0	0	1
Total	6	6	8
Duration (s)	0.001	0.001	0.001
MOPS	0.006	0.006	0.008

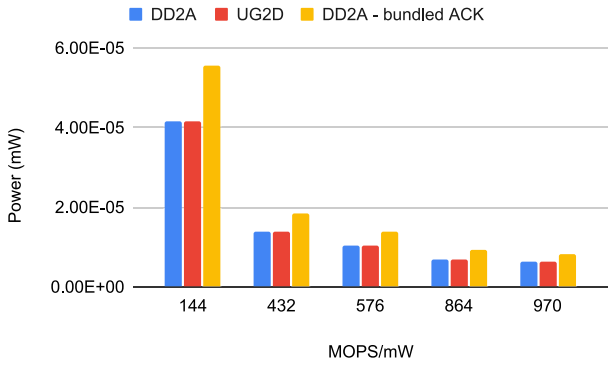

FIGURE 11. Power consumption for implementing our solution in a UE with different processor architectures.

We put these numbers in perspective by comparing them against typical operating power consumed by commercial low-power IoT UEs. To this end, we use the measurement results obtained for power consumption in NB-IoT devices [45]. NB-IoT UEs are shown to consume at least 200 mW and 700 mW of power in the uplink and downlink, respectively. The maximum of the power consumption numbers from Fig. 11 shows that adaptive DD2A with bundled ACK and UG2D computations using our proposed method introduces an additional 60 nW and 40 nW, respectively. It can be seen that these numbers are several orders of magnitude lower than the overall power consumption of the device. Such a low power consumption introduces a truly negligible impact on the battery life of an IoT device.

V. CONCLUSION

We proposed a dynamic HARQ scheduling design targeted at IoT-NTN UEs to exploit the extended signal propagation time encountered in satellite communication links. We presented an analysis to determine a suitable number of HARQs to be supported to extract superior throughput under any propagation condition. Our detailed simulation evaluation demonstrated noticeable gains in the achieved throughput with the use of our proposed methods, by considering suitable coding rates for different transmission link conditions and satellite altitudes. Our solution enables NTNs to serve an increased number of UEs, which is critical, given the extended cell-size in NTNs and the increasing number of interconnected devices.

REFERENCES

[1] "NTN IoT HARQ considerations," Sierra Wireless, Richmond, BC, Canada, document 3GPP TSG RAN WG1 Meeting #104-eR1-2101323, Feb. 2021.

[2] A. Guidotti, A. Vanelli-Coralli, A. Mengali, and S. Cioni, "Non-terrestrial networks: Link budget analysis," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–7.

[3] F. Rinaldi et al., "Broadcasting services over 5G NR enabled multi-beam non-terrestrial networks," *IEEE Trans. Broadcast.*, vol. 67, no. 1, pp. 33–45, Mar. 2021.

[4] J.-H. Lee, J. Park, M. Bennis, and Y.-C. Ko, "Integrating LEO satellite and UAV relaying via reinforcement learning for non-terrestrial networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.

[5] M. Bacco et al., "Networking challenges for non-terrestrial networks exploitation in 5G," in *Proc. IEEE 2nd 5G World Forum (5GWF)*, 2019, pp. 623–628.

[6] X. Lin et al., *5G New Radio Evolution Meets Satellite Communications: Opportunities, Challenges, and Solutions*. Cham, Switzerland: Springer, 2021, pp. 517–531. [Online]. Available: https://doi.org/10.1007/978-3-030-58197-8_18

[7] "Solutions for NR to support non-terrestrial networks (NTN)," 3GPP, Sophia Antipolis, France, Rep. TR 38.821, Dec. 2019.

[8] "Summary #4 of 8.15.1 IoT NTN scenarios," Eutelsat, Paris, France, document 3GPP TSG RAN WG1 Meeting #103e, R1-2008868, Nov. 2020.

[9] N. H. Mahmood et al., "White paper on critical and massive machine type communication towards 6G," 2020, *arXiv:2004.14146*.

[10] "study on new radio (NR) to support non-terrestrial networks," 3GPP, Sophia Antipolis, France, Rep. TR 38.811, Sep. 2020.

[11] O. Liberg et al., "Narrowband Internet of Things for non-terrestrial networks," *IEEE Commun. Stand. Mag.*, vol. 4, no. 4, pp. 49–55, Dec. 2020.

[12] "Evolved universal terrestrial radio access (E-UTRA); physical layer procedures," 3GPP, Sophia Antipolis, France, Rep. TS 36.213, Apr. 2022.

[13] M. Conti, A. Guidotti, C. Amatetti, and A. Vanelli-Coralli, "NB-IoT over non-terrestrial networks: Link budget analysis," in *Proc. IEEE Global Commun. Conf.*, 2020, pp. 1–6.

[14] M. Gineste et al., "Narrowband IoT service provision to 5G user equipment via a satellite component," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2017, pp. 1–4.

[15] G. Charbit, D. Lin, K. Medles, L. Li, and I. -K. Fu, "Space-terrestrial radio network integration for IoT," in *Proc. 6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[16] M. Conti, S. Andrenacci, N. Maturo, S. Chatzinotas, and A. Vanelli-Coralli, "Doppler impact analysis for NB-IoT and satellite systems integration," in *Proc. Int. Conf. Commun. (ICC)*, 2020, pp. 1–7.

[17] O. Kodheli, N. Maturo, S. Chatzinotas, S. Andrenacci, and F. Zimmer, "On the random access procedure of NB-IoT non-terrestrial networks," in *Proc. Adv. Satell. Multimedia Syst. Conf. Sig. Process. Space Commun. Workshop (ASMS/SPSC)*, 2020, pp. 1–8.

[18] H. Chougrani, S. Kisseleff, W. A. Martins, and S. Chatzinotas, "NB-IoT random access for non-terrestrial networks," 2021, *arXiv:2101.08079*.

[19] O. Kodheli et al., "Random access procedure over non-terrestrial networks: From theory to practice," *IEEE Access*, vol. 9, pp. 109130–109143, Jun. 2021.

[20] O. Kodheli, S. Andrenacci, N. Maturo, S. Chatzinotas, and F. Zimmer, "Resource allocation approach for differential Doppler reduction in NB-IoT over LEO satellite," in *Proc. 9th Adv. Satell. Multimedia Syst. Conf. 15th Signal Process. Space Commun. Workshop (ASMS/SPSC)*, 2018, pp. 1–8.

[21] O. Kodheli, S. Andrenacci, N. Maturo, S. Chatzinotas, and F. Zimmer, "An uplink UE group-based scheduling technique for 5G mMTC systems over LEO satellite," *IEEE Access*, vol. 7, pp. 67413–67427, 2019.

[22] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Netw.*, vol. 35, no. 2, pp. 244–251, Mar./Apr. 2021.

[23] R. Barbau, V. Deslandes, G. Jakllari, J. Tronc, J.-F. Chouteau, and A.-L. Beylot, "NB-IoT over GEO satellite: Performance analysis," in *Proc. Adv. Satell. Multimedia Syst. Conf. 16th Signal Process. Space Commun. Workshop (ASMS/SPSC)*, 2020, pp. 1–8.

[24] "Uplink HARQ-ACK feedback," Qualcomm, Inc., San Diego, CA, USA, document 3GPP TSG RAN WG1 Meeting #88bis, R1-1705013, Apr. 2017.

[25] X. Lin, S. Rommer, S. Euler, E. A. Yavuz, and R. S. Karlsson, "5G from space: An overview of 3GPP non-terrestrial networks," 2021, *arXiv:2103.09156*.

[26] "Evolved universal terrestrial radio access (E-UTRA); medium access control (MAC) protocol specification," 3GPP, Sophia Antipolis, France, Rep. TS 36.321, Apr. 2022.

[27] "Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," 3GPP, Sophia Antipolis, France, Rep. TS 36.331, Apr. 2022.

[28] F. Rinaldi et al., "Non-terrestrial networks in 5G & beyond: A survey," *IEEE Access*, vol. 8, pp. 165178–165200, 2020.

[29] G. Geraci, D. Lopez-Perez, M. Benzaghta, and S. Chatzinotas, "Integrating terrestrial and non-terrestrial networks: 3D opportunities and challenges," *IEEE Commun. Mag.*, vol. 61, no. 4, pp. 42–48, Apr. 2023

[30] A. Sørensen et al., "Modeling and experimental validation for battery lifetime estimation in NB-IoT and LTE-M," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9804–9819, Jun. 2022.

[31] A. Khalifeh, K. A. Aldahdouh, K. A. Darabkh, W. Al-Sit et al., "A survey of 5G emerging wireless technologies featuring LoRaWAN, Sigfox, NB-IoT and LTE-M," in *Proc. Int. Conf. Wireless Commun. Signal Process. Netw. (WiSPNET)*, 2019, pp. 561–566.

[32] "Study on Narrow-Band Internet of Things (NB-IoT)/enhanced machine type communication (eMTC) support for non-terrestrial networks (NTN)," 3GPP, Sophia Antipolis, France, Rep. TR 36.763, Jan. 2021.

[33] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution: From Theory to Practice*. Hoboken, NJ, USA: Wiley, 2011.

[34] Y.-P. E. Wang et al., "A primer on 3GPP Narrowband Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 117–123, Mar. 2017.

[35] A. Hoglund et al., "Overview of 3GPP release 14 further enhanced MTC," *IEEE Commun. Stand. Mag.*, vol. 2, no. 2, pp. 84–89, Jun. 2018.

[36] S. R. Borkar, "Long-term evolution for machines (LTE-M)," in *LPWAN Technologies for IoT and M2M Applications*, B. S. Chaudhari and M. Zennaro, Eds. Cambridge, MA, USA: Academic, 2020, pp. 145–166.

[37] V. Rajendran, G. Prasad, L. Lampe, and G. Vos, "SCUBA: An in-device multiplexed protocol for sidelink communication on unlicensed bands," *IEEE Internet Things J.*, vol. 8, no. 22, pp. 16637–16652, Nov. 2021.

[38] G. Vos and S. Dost, "Method and apparatus for resource allocation for half duplex frequency division duplexing in a wireless communication system," U.S. Patent Appl. 66 002, 2021.

[39] "Discussion on the delay-tolerant HARQ operation for NTN," ZTE, Shenzhen, China, document 3GPP TSG RAN1 #99, R1-1913238, Nov. 2019.

[40] "Summary#1 of AI: 8.1.2.4 enhancements on HST-SFN deployment," MCC Support, 3GPP, 3GPP, Sophia Antipolis, France, document 3GPP TSG RAN WG1 103-e, v1.0.0, Nov. 2020.

[41] "On frequency compensation, uplink timing and random access in NTN," Ericsson, Stockholm, Sweden, document 3GPP TSG RAN1 #98, R1-1909107, Aug. 2019.

[42] "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)," 3GPP, Sophia Antipolis, France, Rep. TR 38.901, V16.1.0, Dec. 2019.

[43] "Correction of round trip delay drift rate for NTN scenarios," Nokia, Espoo, Finland, document 3GPP TSG RAN1 #104bis-e, R1-2103719, Apr. 2021.

[44] R. Fasthuber, F. Catthoor, P. Raghavan, and F. Naessens, *Energy-Efficient Communication Processors: Design and Implementation for Emerging Wireless Systems*, vol. 10. New York, NY, USA: Springer, 2013.

[45] M. Lauridsen, R. Krigslund, M. Rohr, and G. Madueno, "An empirical NB-IoT power consumption model for battery lifetime estimation," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, 2018, pp. 1–5.

GAUTHAM PRASAD received the M.S. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2014, and the Ph.D. degree in electrical and computer engineering from The University of British Columbia, Vancouver, BC, Canada, in 2019. He has coauthored and is a co-inventor in numerous conference/journal articles and patent applications, respectively. His research interests broadly lie in all areas covering the applications of communications theory, signal processing, and machine learning. He is a recipient of multiple best paper awards, academic scholarships, travel awards, start-up cohort funds, and research grants.

VISHNU RAJENDRAN CHANDRIKA received the B.Tech. degree in electronics and communication engineering from the University of Kerala, India, in 2011, the M.Tech. degree in electronics and electrical engineering from the Indian Institute of Technology Guwahati, Guwahati, India, in 2014, and the Ph.D. degree from the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada, in 2023. His research interests are device-to-device communications, low-cost cellular machine-type communications, Internet of Things, and cellular non-terrestrial network communications.

LUTZ LAMPE (Senior Member, IEEE) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the University of Erlangen, Erlangen, Germany, in 1998 and 2002, respectively. Since 2003, he has been with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada, where he is a Full Professor. He is a Co-Editor of the book "*Power Line Communications: Principles, Standards and Applications from Multimedia to Smart Grid*" (2nd edition, John Wiley & Sons). His research interests are broadly in theory and application of wireless, optical wireless, optical fiber, power line, and underwater acoustic communications. He has been a (co-)recipient of a number of best paper awards. He has served as an associate editor and a guest editor for several IEEE journals, and as the general and technical program committee co-chair for IEEE conferences. He has been a Distinguished Lecturer of the IEEE Communications Society.

GUS VOS (Senior Member, IEEE) received the B.Eng. degree from the University of Victoria and the M.Eng. degree from Simon Fraser University. As a VP Chief Scientist for technology standards with Semtech, he serves as the company's 3GPP representative, leading the company's efforts to optimize the LTE and NR radio access technology standards for the Internet of Things. He brings to this role more than 30 years of experience in wide-area wireless communications, having started his career working with proprietary protocols before moving to his current role advocating for standardized cellular protocols.