Impact of NOMA on Age of Information: A Grant-Free Transmission Perspective

Zhiguo Ding, Fellow, IEEE, Robert Schober, Fellow, IEEE, and H. Vincent Poor, Life Fellow, IEEE

Abstract—The aim of this paper is to characterize the impact of non-orthogonal multiple access (NOMA) on the age of information (AoI) of grant-free transmission. In particular, a low-complexity form of NOMA, termed NOMA-assisted random access, is applied to grant-free transmission in order to illustrate the two benefits of NOMA for AoI reduction, namely increasing channel access and reducing user collisions. Closed-form analytical expressions for the time average AoI achieved by NOMA assisted grant-free transmission are obtained, and asymptotic studies are carried out to demonstrate that the use of the simplest form of NOMA is already sufficient to reduce the AoI of orthogonal multiple access (OMA) by more than 40%. In addition, the developed analytical expressions are also shown to be useful for optimizing the users’ transmission attempt probabilities, which are key parameters for grant-free transmission.

Index Terms—Non-orthogonal multiple access (NOMA), age of information (AoI), grant-free transmission.

I. INTRODUCTION

G rant-free transmission is an important enabling technique to support envisioned sixth-generation (6G) services, including ultra massive machine type communications (umMTC) and enhanced ultra-reliable low latency communications (euRLLC) [1], [2]. Unlike conventional grant-based transmission, grant-free transmission enables users to avoid multi-step signaling and directly transmit data signals together with access control information, which can reduce the system overhead significantly, particularly for scenarios with massively many users requiring short-package transmission. Grant-free transmission can be realized by applying the random access protocols developed for conventional computer networks, such as ALOHA random access [3], [4]. Alternatively, massive multi-input multi-output (MIMO) can also be applied to support grant-free transmission by using the excessive spatial degrees of freedom offered by massive MIMO [5], [6], [7].

Recently, the application of non-orthogonal multiple access (NOMA) to grant-free transmission has received significant attention due to the following two reasons. First, the NOMA principle is highly compatible, and the use of NOMA can significantly improve the reliability and spectral efficiency of random access and massive MIMO based grant-free protocols [8], [9], [10], [11]. Second, more importantly, the use of NOMA alone is sufficient to support grant-free transmission. For example, NOMA-based grant-free transmission has been proposed in [12], where a Bayesian learning based scheme has been designed to ensure successful multi-user detection, even if the number of active grant-free users is unknown. The principle of NOMA can also be used to develop so-called semi-grant-free transmission protocols, where the bandwidth resources which would be solely occupied by grant-based users are released for supporting multiple grant-free users in a distributed manner [13]. In addition, NOMA-based grant-free transmission has also been shown to be robust and efficient in various communication scenarios, such as satellite communication networks, secure Internet of Things (IoT), intelligent reflecting surface (IRS) networks, marine communication systems, etc., see [14], [15], [16], and [17].

The aim of this paper is to characterize the impact of NOMA on the performance of grant-free transmission with respect to a recently developed performance metric, termed the age of information (AoI) [18], [19], [20]. In particular, the AoI describes the freshness of data updates collected in the network, and is an important metric to measure the success of the anticipated 6G services, including umMTC and euRLLC. We note that several existing works have focused on the impact of NOMA on grant-based networks [21], [22], [23], [24], [25]. For example, for two-user grant-based networks, the capability of NOMA to reduce the AoI has been shown to be related to the spectral efficiency gain of NOMA over orthogonal multiple access (OMA) [26]. To the authors’ best knowledge, there is only a single existing work that has applied NOMA to reduce the AoI of grant-free transmission [27], where the strong assumption that the base station estimates all users’ channel state information (CSI) was made.

In this paper, the impact of NOMA on the AoI of grant-free transmission is investigated from the perspective of...
performance analysis, which is different from the existing work focusing on resource allocation [28]. In particular, a low-complexity form of NOMA, which was originally termed NOMA-assisted random access [29] and recently also termed ALOHA with successive interference cancellation (SIC) [30], [31], is adopted in order to illustrate the two benefits of NOMA for AoI reduction, namely increasing channel access and reducing user collisions. The key element of the proposed performance analysis is the modeling of the channel competition among the grant-free users as a Markov chain, which is different from the performance analysis for grant-based NOMA networks [21], [22], [23], [26]. The main contributions of this paper are two-fold:

- Analytical expressions for the time average AoI achieved by NOMA assisted grant-free transmission are obtained, by characterizing the state transition probabilities of the considered Markov chain. We note that by using NOMA-assisted random access, the base station creates multiple preconfigured receive signal-to-noise ratio (SNR) levels, which makes NOMA-assisted grant-free transmission similar to multi-channel ALOHA. As a result, the calculation of the state transition probabilities for the NOMA case is more challenging than that for the OMA case, which can be viewed as single-channel ALOHA. By exploiting the properties of the considered Markov chain and also the characteristics of SIC, closed-form expressions for the state transition probabilities are developed for NOMA assisted grant-free transmission.

- Valuable insights regarding the relative performance of NOMA and OMA assisted grant-free transmission are also obtained. For example, for the case where users always have updates to deliver, asymptotic expressions are developed to demonstrate that the use of NOMA can almost halve the AoI achieved by OMA, even if the simplest form of NOMA is implemented. In addition, the optimal choices of the users’ transmission probabilities for random access with NOMA and OMA are obtained and compared. This study reveals that NOMA-assisted grant-free transmission is fundamentally different from multi-channel ALOHA due to the use of SIC. Furthermore, simulation results are provided to verify the developed analytical expressions and also demonstrate that the use of NOMA can significantly reduce the AoI compared to OMA, particularly for the case of low transmit SNR and a massive number of grant-free users.

II. NOMA ASSISTED GRANT-FREE TRANSMISSION

Consider a grant-free communication network with $M$ users, denoted by $U_m$, $1 \leq m \leq M$, communicating with the same base station. Assume that each time frame comprises $N$ time slots, each of duration of $T$ seconds, where the $n$-th time slot of the $i$-th frame is denoted by $T_S^n_i$, and the starting time of $T_S^n_i$ is denoted by $t_1^n_i$, $1 \leq n \leq N$ and $i \geq 1$, as shown in Fig. 1.

A. Data Generation Models

For the considered grant-free transmission scenario, each user tries to deliver one update to the base station in each time frame. When the users’ updates are generated depends on which of the following two data generation models is used [21].

1) Generate-at-Request (GAR): For GAR, the base station requests all users to simultaneously generate their updates at the beginning of each time frame. GAR is applicable to many important IoT applications, such as structural health monitoring and autonomous driving.

2) Generate-at-Will (GAW): For GAW, a user’s update is generated right before its transmit time slot. GAW has been commonly used in the AoI literature, since it ensures that the delivered updates are freshly generated. We note that the GAW model considered in this paper is a simplified one, and there are more sophisticated GAW models, e.g., a user can generate a new update every few time slots, store the updates in its buffer, and carry out retransmission between the update generation time points. The study of these advanced GAW models is beyond the scope of this paper.

The main focus of this paper is to study the impact of NOMA on grant-free transmission, where both the GAR and the simplified GAW model are used. Here, we focus on GAR for the following two reasons. First, the time-average AoI expression for grant-free transmission for the simplified GAW model can be straightforwardly obtained from that for GAR, as shown in the next section. Second, if there are retransmissions within one time frame, GAW requires a user to repeatedly generate updates, and hence, causes a higher energy consumption than GAR. For grant-based transmission, this increase in energy consumption is not severe since the number of retransmissions in one frame is small [21]. However, in the grant-free case, a user might have to carry out a large number of retransmissions due to potential collisions, which means that the simplified GAW can cause a significantly higher energy consumption than GAR.

B. Channel Access Modes

1) Orthogonal Multiple Access (OMA): Based on OMA, a user’s transmission can be successful only if it solely occupies the bandwidth resource block, i.e., a time slot. A simple example of OMA based grant-free transmission is slotted ALOHA, as described in the following.\(^1\)

Prior to $T_S^n_i$, assume that $j$ users have successfully delivered their updates to the base station. For grant-free transmission, each of the remaining $M-j$ users will independently make a transmission attempt with the same transmit power, denoted by $P_TX$. $P_TX$ can be based on a fixed choice, or be state-dependent, i.e., $P_TX = \frac{1}{M-j}$ [35]. It is assumed that the base station can inform the users about the outcome of their updates via a dedicated control channel.

There are three possible events which cause an update failure for a user: i) the user does not make a transmission 

\(^1\)In this paper, a simple slotted ALOHA scheme is considered, where each user can use any of the time slots in the frame. For more sophisticated random access schemes, e.g., irregular repetition slotted ALOHA (IRSA) [32], [33], [34], which allows a user to choose a subset of time slots for transmission, the principle of NOMA can be also applied as an add-on, e.g., a group of users, instead of a single user, share the same subset of time slots.

\(^2\)Because the noise power is assumed to be normalized to one, $P$ can also be interpreted as the transmit SNR.
attempt; ii) a collision occurs, i.e., there are more than one concurrent transmissions; iii) an outage occurs due to the user’s weak channel condition, i.e., \( \log(1 + P|h_{im}^{1,n}|^2) \leq R \), where \( R \) denotes the user’s target data rate, and \( h_{im}^{1,n} \) denotes \( U_m \)’s channel gain in \( T_S^i \). In this paper, all users are assumed to have the identical target data rates, and their channel gains are assumed to be independent and identically complex Gaussian distributed with zero mean and unit variance. To facilitate the AoI analysis carried out in the paper, only small-scale multi-path fading is considered. This assumption can be justified by smart home and indoor sensing applications, where the users are indoor sensors that try to send their updates to a closely-located data fusion center. We note that the impact of large-scale path-loss on the AoI of grant-free transmission will be illustrated Section IV by using computer simulations, where the case with non-identically distributed channel gains will be also considered.

2) Non-Orthogonal Multiple Access (NOMA): With NOMA, a user can still succeed in its updating, even if multiple users choose the same time slot. Recall that the principle of NOMA can be implemented in different forms. For the purpose of illustration, a particular form of NOMA, termed NOMA-assisted random access, is adopted to reduce the AoI of grant-free transmission [29]. In particular, prior to transmission, the base station configures \( K \) receive SNR levels, denoted by \( P_1 \geq \cdots \geq P_K \). If \( U_m \) chooses \( P_k \) during \( T_S^i \), it will scale its transmitted signal by \( \sqrt{\frac{P_k}{|h_{im}^{1,n}|^2}} \). The base station carries out SIC by decoding the signal delivered at SNR level, \( P_k \), before decoding the one at \( P_{k+1} \), \( 1 \leq k \leq K - 1 \). The SNR levels are preconfigured to guarantee the success of SIC, i.e., the following conditions need to be satisfied:

\[
\log \left( 1 + \frac{P_k}{1 + (M - 1)P_{k+1}} \right) = R, \quad 1 \leq k \leq K - 1,
\]

and

\[
\log \left( 1 + P_{K_1} \right) = R, \quad \text{which means} \quad P_K = 2^R - 1 \quad \text{and} \quad P_k = \left( 2^R - 1 \right) \left( 1 + (M - 1)P_{k+1} \right), \quad \text{where the noise power is assumed to be normalized to one.}
\]

We note that the condition in (1) is stricter than the condition \( 1 + \frac{P_k}{1 + \sum_{i, j \neq k} P_j} = R \), and ensures the success of SIC, even if one user chooses \( P_k \) and the remaining users choose the SNR level which contributes the most interference, i.e., \( P_{k+1} \). We further note that the case in which all \( M - 1 \) remaining users choose \( P_{k+1} \) is the worst case, since some users may choose SNR levels other than \( P_{k+1} \) or even decide not to transmit at all.

Again assume that there are \( j \) users which have successfully sent their updates to the base station prior to \( T_S^i \). Each of the remaining \( M - j \) users will first randomly choose an SNR level with equal probability, denoted by \( P_k = \frac{R}{K} \), and independently make a transmission attempt with probability \( P_{\text{TX}} \). For illustrative purposes, assume that \( U_m \) is among the \( M - j \) remaining users, and chooses \( P_k \). The possible events which cause \( U_m \)’s update to fail are listed as follows:

- The user does not make an attempt for transmission;
- The receive SNR level chosen by the user is not feasible due to the user’s transmit power budget, i.e., \( P_k \) is not feasible for \( U_m \) in \( T_S^i \) if \( \frac{P_k}{|h_{im}^{1,n}|^2} > P_i \);
- Another user also chooses \( P_k \), which leads to a collision at \( P_k \) and hence a failure at the \( k \)-th stage of SIC;
- Prior to the \( k \)-th stage of SIC, SIC has already been terminated due to one or more failures in the previous SIC stages.

We note that for both the OMA and NOMA cases, a user keeps resending its update to the base station until either the user succeeds or the time frame is finished.

3) Implementation Issues of NOMA: As shown in the next section, the use of NOMA can significantly reduce the AoI of grant-free transmission compared to OMA. However, it is important to note that the implementation of NOMA introduces additional system complexity. For example, NOMA requires SIC; however, for the considered uplink scenario, SIC is to be carried out by the base station, which means that SIC does not cause additional hardware complexity at the user side.

Considering the CSI requirement, one benefit of the adopted form of NOMA is that the base station does not need to estimate all users’ CSI for implementing SIC, as is commonly required by other schemes, e.g., [27] and [28]. In particular, for the adopted form of NOMA, the users require access to their own CSI only. In practice, this CSI assumption can be realized by asking the base station to broadcast pilot signals at the beginning of a time slot, where the users can perform channel estimation individually. For many indoor communication scenarios, such as smart home and intelligent cities, the users are static sensors, and hence the pilot signals can be sent infrequently, which can further reduce the system overhead.

We note that the main reason for using NOMA to reduce the AoI of grant-free networks is that in NOMA, multiple users are allowed to transmit simultaneously, which requires that the users are time synchronized. In this paper, NOMA is implemented in a slotted ALOHA network, where global time synchronization has already been realized in order to reduce collisions between users [36]. We further note that NOMA can also be implemented in an asynchronous manner as shown in [37].

C. AoI Model

AoI is an important performance metric for quantifying the freshness of the updates delivered to the base station. Recall that the transmission attempt probabilities of all users are assumed to be identical. In addition, the users’ channel gains are also assumed to be independent and identically distributed. Therefore, all the users experience the same AoI, and hence, without loss of generality, \( U_j \)’s instantaneous AoI at time \( t \) is focused on and defined as follows [18]:

\[
\Delta(t) = t - \hat{T}(t),
\]

where \( \hat{T}(t) \) denotes the generation time of \( U_j \)’s freshest update successfully delivered to the base station. \( U_j \)’s long-term average AoI of the considered network is given by

\[
\bar{\Delta} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \Delta(t) dt.
\]

The averaged AoI achieved by OMA and NOMA assisted grant-free transmission will be analyzed in the following section.

3 The reason why the adopted NOMA form does not require the base station to have access to the users’ CSI is explained in the following. Without loss of generality, assume that \( U_k \), \( 1 \leq k \leq K \), is active, and chooses \( P_k \). Because each user scales its transmitted signal by its channel gain, the base station receives \( y = \sum_{k=1}^{K} \sqrt{P_k} s_k + w \), where \( s_k \) denotes \( U_k \)’s signal and \( w \) denotes the noise. Therefore, the base station does not need to know the users’ CSI for carrying out SIC. However, it is assumed that the base station knows the users’ choices for \( P_k \) via a reliable feedback channel.
III. AOI OF GRANT-FREE TRANSMISSION

As discussed in the previous section, $U_1$ is treated as the tagged user and its AoI will be focused on in this section, without loss of generality. For the example shown in Fig. 2, $U_1$ successfully sends its updates to the base station in the $n$-th time slot of frame $i$ ($TS_n^i$), and the $l$-th time slot of frame $i+2$ ($TS^{i+2}_l$), but fails in frame $i+1$. For the AoI analysis, the following metrics are required:

- $S_j$: The time duration between the generation time and the receive time of the $j$-th successful update. For the example shown in Fig. 2, $S_{j-1} = nT$ and $S_j = lT$.
- $Y_j$: The time duration between the $(j-1)$-th and the $j$-th successful updates. For the example shown in Fig. 2, $Y_j = (N-n)T + NT + lT$.
- $X_j$: The number of frames between the $(j-1)$-th and the $j$-th successful updates. An example of $X_j$ is shown in Fig. 2.

We note that in the literature of random access, $S_j$ is termed the service delay and $Y_j$ is termed the inter-departure time [35].

By using the aforementioned metrics, for GAR, $U_1$'s averaged AoI is given by

$$
\bar{\Delta} = \lim_{J \to \infty} \sum_{j=1}^{J} \frac{Q_j}{J} = \lim_{J \to \infty} \sum_{j=1}^{J} \left( S_{j-1} Y_j + \frac{1}{2} Y_j^2 \right)
$$

where $J$ denotes the total number of successful updates, $Q_j$ denotes the area of the shaded region shown in Fig. 2, $\mathbb{E}\{Y_j^2\} = \sum_{j=1}^{J} \frac{1}{2} Y_j^2$, $\mathbb{E}\{Y_j\} = \sum_{j=1}^{J} Y_j$, and $\mathbb{E}\{\cdot\}$ denotes the expectation. We note that the expression $Q_j = S_{j-1} Y_j + \frac{1}{2} Y_j^2$ is used since the shaded region in Fig. 2 can be divided into two parts. In particular, the area for the bottom rectangle shape is $S_{j-1} Y_j$, and the area for the upper triangle shape is $\frac{1}{2} Y_j^2$.

Remark 1: We note that the AoI expressions in (4) of this paper and [35, Eq. (3)] are consistent. The reason why there is an extra factor of $\frac{1}{2}$ in [35, Eq. (3)] is that the users' instantaneous AoI was assumed to be discrete-valued in [35], instead of continuous-valued as in this paper.

Remark 2: For GAW, the user’s averaged AoI is given by

$$
\bar{\Delta}^{GAW} = \lim_{J \to \infty} \frac{\sum_{j=1}^{J} \left( T Y_j + \frac{1}{2} Y_j^2 \right)}{\sum_{j=1}^{J} Y_j} = T + \frac{\mathbb{E}\{Y_j^2\}}{2 \mathbb{E}\{Y_j\}},
$$

which is simpler than the AoI expression in (4). Therefore, the analytical results developed for GAR are straightforwardly applicable to the case for GAW.

A. Generic Expressions for $\mathbb{E}\{S_{j-1} Y_j\}$, $\mathbb{E}\{Y_j\}$, and $\mathbb{E}\{Y_j^2\}$

As shown in (4), the AoI is a function of $\mathbb{E}\{S_{j-1} Y_j\}$, $\mathbb{E}\{Y_j\}$, and $\mathbb{E}\{Y_j^2\}$, and generic expressions for these metrics have been derived in [35], and will be briefly introduced in this subsection. In particular, the considered grant-free transmission can be modeled by a Markov chain with $M+1$ states, denoted by $s_k$, $0 \leq k \leq M$. In particular, $s_0$, $0 \leq k \leq M-1$, denotes the transient state, whereas $s_M$ means that $U_1$ has successfully delivered its update to the base station. Define the state transition probability from $s_j$ to $s_{j+k}$ by $P_{j,k}$, $0 \leq i,j \leq M$. Build an $M \times M$ matrix, denoted by $P_M$, whose element in the $(i+1)$-th column and the $(j+1)$-th row is $P_{j+1,i}$, $0 \leq i,j \leq M-1$. Furthermore, build an $M \times 1$ vector, denoted by $p$, whose $(i+1)$-th element is $P_{i,M}$. Once $P_M$ and $p$ are available, $\mathbb{E}\{S_{j-1} Y_j\}$, $\mathbb{E}\{Y_j\}$, and $\mathbb{E}\{Y_j^2\}$ can be obtained as follows.

Denote by $Z$ the number of time slots required by $U_1$ to successfully deliver its update to its base station. For the example shown in Fig. 2, $Z = n$ for the $(j-1)$-th successful update and $Z = l$ for the $j$-th successful update. Then, the probability mass function (pmf) of $Z$ is given by

$$
P(Z = n) = \frac{s_0^T P_M^{n-1} p}{1 - P_{fail}}, \quad n = 1, 2, \ldots, N,
$$

where $s_0 = [1 \ 0 \ \cdots \ (M-1)^T]^T$ denotes the initial probability vector and $0_{m \times n}$ denotes an all-zero $m \times n$ matrix. Therefore, the probability that $U_1$ cannot complete an update within one frame is given by $P_{fail} = P(Z > N) = s_0^T P_M^{N-1} 1$, where $1$ denotes an $M \times 1$ all-one vector. Therefore, the pmf of access delay, $S_j$, can be written as follows:

$$
P(S_j = nT) = \frac{P(Z = n)}{1 - P_{fail}} = \frac{s_0^T P_{n-1}^{M-1} p}{1 - s_0^T P_M^{N-1} 1},
$$

for $1 \leq n \leq N$, which means $\mathbb{E}\{S_j\} = T \sum_{n=1}^{N} n \frac{s_0^T P_{n-1}^{M-1} p}{1 - s_0^T P_M^{N-1} 1}$ and $\mathbb{E}\{S_j^2\} = T^2 \sum_{n=1}^{N} n^2 \frac{s_0^T P_{n-1}^{M-1} p}{1 - s_0^T P_M^{N-1} 1}$.

Similarly, the pmf of $X_j$ is given by

$$
P(X_j = n) = P_{fail}^{n-1} (1 - P_{fail}),
$$

which means that $\mathbb{E}\{X_j\} = \frac{1}{1 - P_{fail}}$ and $\mathbb{E}\{X_j^2\} = \frac{1 + P_{fail}}{(1 - P_{fail})^2}$. The expressions of $\mathbb{E}\{X_j\}$ and $\mathbb{E}\{X_j^2\}$ can be further evaluated to $\mathbb{E}\{Y_j\}$ and $\mathbb{E}\{Y_j^2\}$, since $\mathbb{E}\{Y_j\} = T N \mathbb{E}\{X_j\}$ and $\mathbb{E}\{Y_j^2\} = N^2 T^2 \mathbb{E}\{X_j^2\} + 2 T \mathbb{E}\{S_j\} - 2 \mathbb{E}\{S_j^2\} + 2 \mathbb{E}\{S_j\}^2 - 2 \mathbb{E}\{S_j\}^2$. Furthermore, $\mathbb{E}\{S_j\}$,
\[ \mathcal{E}\{S_j^2\}, \text{ and } \mathcal{E}\{Y_j\} \text{ can be used to evaluate } \mathcal{E}\{S_{j-1}Y_j\} \text{ which can be expressed as follows:} \]
\[ \mathcal{E}\{S_{j-1}Y_j\} = \mathcal{E}\{S_j\} \mathcal{E}\{Y_j\} - \mathcal{E}\{S_j^2\} + \mathcal{E}\{S_j\}, \]
where the last step follows by the fact that \( S_{j-1} \) and \( Y_j - (NT - S_{j-1}) \) are independent.

As discussed above, the crucial step to evaluate the AoI is to find \( \mathbf{P} \) and \( \mathbf{p} \), which depends on the used multiple access schemes.

**B. OMA-Based Grant-Free Transmission**

The state transition probabilities for the OMA case can be straightforwardly obtained, as shown in the following. With OMA, a single user can be served in each time slot, which means that the number of successful users after one time slot can be increased by one at most. Therefore, most of the state transition probabilities in matrix \( \mathbf{P} \) are zero, except for \( P_{j,j}, P_{j,j+1} \), and \( P_{j,M}, 0 \leq j \leq M - 1 \). In particular, \( P_{j,j} \) denotes the probability of the event that no user succeeds, and is given by [35]
\[ P_{j,j} = 1 - (M - j)\mathbb{P}_{TX}e^{-\nu} \left( 1 - \mathbb{P}_{TX} \right)^{M-j-1}, \] (8)
where \( \nu = 2R - 1 \). \( P_{j,j+1} \) denotes the probability of the event that a single user, but not \( U_j \), succeeds and is given by
\[ P_{j,j+1} = (M - j - 1)\mathbb{P}_{TX}e^{-\nu} \left( 1 - \mathbb{P}_{TX} \right)^{M-j-1}. \] (9)
Furthermore, the \( j \)-th element of \( \mathbf{p} \), denoted by \( P_{j,M} \), is given by
\[ P_{j,M} = \mathbb{P}_{TX}e^{-\nu} \left( 1 - \mathbb{P}_{TX} \right)^{M-j-1}. \] (10)

**C. NOMA-Based Grant-Free Transmission**

The benefit of using NOMA is that more users can be admitted simultaneously than for OMA. In particular, with NOMA, the number of successful users after one time slot can be increased by \( K \) at most, whereas the number of successful users was no more than 1 for OMA. This means that the non-zero state transition probabilities in matrix \( \mathbf{P} \) include \( P_{j,j}, P_{j,j+1}, \) and \( P_{j,M}, 0 \leq j \leq M - 1, 1 \leq i \leq K \) and \( j + i \leq M - 1 \).

The analysis of the state transition probabilities for the NOMA case is more challenging than that for the OMA case, mainly due to the application of SIC. For example, a collision at SNR level \( P_k \) can prevent all those users, which choose SNR level \( P_i, i > k \), from being successful. The following lemma provides a high-SNR approximation for the state transition probabilities.

**Lemma 1:** At high SNR, the state transition probability, \( P_{j,j}, 0 \leq j \leq M - 1 \), can be approximated as follows:
\[ P_{j,j} \approx 1 - \sum_{m=1}^{M-j} \binom{M-j}{m} \mathbb{P}_{TX}(1 - \mathbb{P}_{TX})^{M-j-m} \times \sum_{k=1}^{K} m\mathbb{P}_K(1 - k\mathbb{P}_K)^{m-1}, \] (11)
the state transition probability, \( P_{j,j+1}, 0 \leq j \leq M - 2 \), can be approximated as follows:
\[ P_{j,j+1} \approx (M - j)\mathbb{P}_{TX}(1 - \mathbb{P}_{TX})^{M-j-1} \frac{M-j-1}{M-j} K\mathbb{P}_K \]
\[ + \sum_{m=2}^{M-j} \binom{M-j}{m} \mathbb{P}_{TX}(1 - \mathbb{P}_{TX})^{M-j-m} \sum_{k=1}^{K} \frac{M-j-1}{M-j} m\mathbb{P}_K \times \left[ (1 - k\mathbb{P}_K)^{m-1} - \sum_{\kappa=k+1}^{\min\{M-1, j, K\}} (m-1)\mathbb{P}_K(1 - k\mathbb{P}_K)^{m-2} \right], \] (12)
and the state transition probability, \( P_{j,j+i}, 0 \leq j \leq M - 3 \) and \( 2 \leq i \leq \min\{M-1-j, K\} \), can be approximated as follows:
\[ P_{j,j+i} \approx \binom{M-j}{i} \mathbb{P}_{TX}(1 - \mathbb{P}_{TX})^{M-j-i} \times \sum_{k_1=1}^{K-i+1} \sum_{k_2=k_1+i-1}^{M-j} \frac{M-j-i}{M-j} \mathbb{P}_K(1 - k\mathbb{P}_K)^{m-1} \times \left[ (1 - k_2\mathbb{P}_K)^{m-i} - \sum_{\kappa=k_2+1}^{\min\{M-1-j, K\}} (m-i)\mathbb{P}_K(1 - k\mathbb{P}_K)^{m-i-3} \right]. \] (13)

**Proof:** See Appendix A.

Once the transition probability matrix \( \mathbf{P} \) is obtained, the elements of \( \mathbf{p} \) can be obtained straightforwardly by applying \( \mathbf{p}^1 + \mathbf{p} = \mathbf{1} \), where recall that \( \mathbf{1} \) denotes an \( M \times 1 \) all-one vector.

The closed-form analytical expressions shown in Lemma 1 allow the evaluation of the impact of NOMA on the AoI without carrying out intensive Monte Carlo simulations. However, the expressions of the state transition probabilities shown in Lemma 1 are quite involved, which makes it difficult to obtain insights about the performance difference between OMA and NOMA. For this reason, the special case of \( K = 2 \) and \( N = 1 \) is focused on in the remainder of this section. \( K = 2 \) means that there are two SNR levels, i.e., the base station needs to carry out two-stage SIC only, which is an important case in practice due to its low system complexity. \( N = 1 \) implies that there is one time slot in each frame, i.e., in each time slot, all users have updates to deliver and hence always participate in contention.

For this case, the following lemma provides the optimal choice for the transmission probability \( \mathbb{P}_{TX} \).

**Lemma 2:** For the special case of \( K = 2 \) and \( N = 1 \), the optimal choice for the transmission probability \( \mathbb{P}_{TX} \) is given by
\[ \mathbb{P}_{TX}^* = \frac{\eta}{M}, \] (14)
for \( M \to \infty \) and \( P \to \infty \), where \( \eta \) is the root of the following equation:
\[ (1 - \frac{\eta}{2}) e^{-\frac{\eta}{2}} + (1 - \frac{\eta^2}{2}) e^{-\eta} = 0. \]

**Proof:** See Appendix B.

**Remark 3:** Note that \( (1 - \frac{\eta}{2}) e^{-\frac{\eta}{2}} + (1 - \frac{\eta^2}{2}) e^{-\eta} = 0 \) is not related to \( M \), which means that \( \eta \) is not a function of \( M \). By applying off-shelf root solvers, the exact value of \( \eta \) can be straightforwardly obtained as follows: \( \eta \approx 1.6646 \).
By using Lemma 2, the AoI performance difference between NOMA and OMA is analyzed in the following proposition.

**Proposition 1:** For the special case of $K = 2$ and $N = 1$, for $M \to \infty$ and $P \to \infty$, the ratio between the AoI achieved by NOMA and OMA is given by

$$\frac{\bar{\Delta}^N}{\bar{\Delta}^O} \approx \frac{2e^{\eta-1}}{\eta(\frac{1}{2} + \frac{9}{4})},$$

where $\bar{\Delta}^N$ and $\bar{\Delta}^O$ denote the AoI achieved by NOMA and OMA, respectively.

**Proof:** See Appendix C.

**Remark 4:** Recall that $\tilde{\eta} \approx 1.6646$, which means that the ratio in Proposition 1 is $\frac{\bar{\Delta}^N}{\bar{\Delta}^O} \approx 0.5653$, i.e., the use of NOMA can almost halve the AoI achieved by OMA. We note that this significant performance gain is achieved by using NOMA with two SNR levels only, i.e., the simplest form of NOMA. By implementing NOMA with more than two SNR levels, the performance gain of NOMA over OMA can be further increased, as shown in the next section.

**Remark 5:** As shown in the proof for Proposition 1, the optimal choice of $P_{TX}$ for OMA is $\frac{1}{M}$. This is expected as explained in the following. With $M$ users competing for access in a single channel, the use of a transmission probability of $\frac{1}{M}$ is reasonable. By using the same rationale, one might expect that $\frac{1}{M}$ should be optimal for the NOMA transmission probability with two SNR levels. However, Lemma 2 shows that the optimal value of $P_{TX}$ is $\frac{1.6646}{M}$, which is a more conservative choice for transmission than $\frac{1}{M}$. The reason for this is potential SIC errors. In particular, although there are two channels (or two SNR levels, $P_1$ and $P_2$), a collision at $P_1$ causes SIC to immediately terminate, which means that $P_2$ can no longer be used to serve any users, i.e., the number of the effective channels is less than 2.

**Remark 6:** Motivated by the results shown in Lemma 2 and Proposition 1, for the general case of $K > 2$, a simple choice of $P_{TX} = \min \{1, \frac{K}{2M} \}$ can be used for NOMA. In fact, the simulation results presented in the next section show that this choice is sufficient to realize a significant performance gain of NOMA over OMA. We note that this choice of $P_{TX}$ depends on $M$ only, unlike the state-dependent choice of $P_{TX}$ used for OMA, which is $P_{TX} = \frac{1}{M - 1}$ [35]. Therefore, an important direction for future research is to find a more sophisticated state-dependent choice of $P_{TX}$ for NOMA-assisted grant-free transmission.

**Remark 7:** Proposition 1 implies that the application of NOMA is particularly beneficial for grant-free transmission, compared to its application to grant-based transmission. Recall that in grant-based networks, one important benefit of using NOMA for AoI reduction is that a user can be scheduled to transmit earlier than with OMA [21]. However, for a user which has already been scheduled to transmit early in OMA, the impact of NOMA on the user’s AoI can be insignificant, particularly at high SNR. Unlike grant-based networks, Proposition 1 shows that in grant-free networks, the use of NOMA can reduce the AoI of OMA by more than 40%, and this significant performance gain applies for all users in the network.

**IV. SIMULATION RESULTS**

In this section, simulation results are presented to demonstrate the AoI achieved by the considered grant-free transmission schemes and to also verify the developed analytical results.

In Fig. 3, the impact of the number of users on the average AoI achieved by the considered grant-free transmission schemes is investigated. As can be seen from the figure, the AoI achieved with NOMA is significantly lower than that of OMA. In addition, Fig. 3 demonstrates that the performance gain of NOMA over OMA increases as the number of users grows. Therefore, the use of NOMA is particularly important for grant-free transmission with a massive number of users, an important use case for 6G. Between the two choices of $P_{TX}$, the adaptive choice yields a better AoI than the fixed choice. For the two subfigures in Fig. 3, different numbers of SNR levels, $K$, are used. By comparing the two subfigures, one can observe that the AoI achieved by the NOMA scheme can be reduced by increasing the number of SNR levels. This is because the use of more SNR levels makes user collisions less likely to happen. Fig. 3 also demonstrates the accuracy of the AoI expressions developed in Lemma 1.

![Fig. 3. Impact of the number of users on the average AoI achieved with OMA and NOMA assisted grant-free transmission for GAR. T = 6, P = 20 dB, R = 0.5 BPCU, and N = 8. For the fixed choice of $P_{TX}$, $P_{TX} = 0.05$, and for the adaptive choice of $P_{TX}$, $P_{TX} = \min \{1, \frac{K}{M} \}$ for NOMA, and $P_{TX} = \frac{1}{M}$ for OMA, i.e., a state-dependent choice is used for OMA as discussed in Section II-B. For the analytical results, Lemma 1 is used.](image-url)
In Fig. 4, the impact of the number of time slots in each frame on the average AoI achieved by the OMA and NOMA assisted grant-free transmission schemes is studied. As can be seen from the figure, the use of NOMA can always realize lower AoI than OMA, regardless of the choices of $N$. An interesting observation from Fig. 4 is that a small increase of $N$, e.g., from 1 to 5, can reduce the AoI. This is because the likelihood for users to deliver their updates to the base station is improved if there are more time slots in each frame. However, after $N \geq 10$, further adding more time slots in each frame increases the AoI, which can be explained with the following example. Assume that $U_1$ can always successfully update its base station in the first time slot of each frame. For this example, $U_1$’s AoI is simply the length of one time frame, and hence its AoI is increased if there are more time slots in one frame.

As discussed in the previous section, the special case with $K = 2$ and $N = 1$ is important in practice, and hence the AoI realized by the OMA and NOMA assisted grant-free transmission schemes is investigated in Fig. 5. In particular, the figure shows that the performance gain of NOMA over OMA is particularly large at low SNR. This is a valuable property in practice since most AoI sensitive applications, such as IoT and sensor networks, are energy constrained and operate in the low SNR regime. Fig. 5(a) also demonstrates that for the case with small $R$, even if the SNR is low, i.e., $-5$ and 0 dB, the difference between the analytical and simulation results is negligible. This property of the developed analytical results is particularly important, given the fact that, for many important applications of grant-free transmission, such as IoT and uMTC, the users’ target data rates are indeed small. The aforementioned conclusions are also confirmed by Fig. 6, where the AoI is shown as a function of the transmit SNR. In particular, the developed analytical results provide accurate estimates in the medium-to-high SNR regions regardless of the choices of $R$.

In Fig. 7, the AoI of grant-free transmission is shown as a function of the transmission probability, $P_{TX}$, and the figure demonstrates that the choices of $P_{TX}$ are crucial to the AoI performance of grant-free transmission. Furthermore, Fig. 7 shows that NOMA assisted grant-free transmission always yields a smaller AoI than the OMA case, if the same values for the transmission probability are used for both of the schemes. As shown in Lemma 2 and the proof for Proposition 1, $P_{TX} = \frac{1}{M}$ is optimal for OMA, and $P_{TX} = \frac{R}{M}$ is optimal for NOMA in the case with $K = 2$ and $N = 1$. Fig. 7 verifies the optimality of these choices of $P_{TX}$, since the minimal Aois achieved by the fixed choices of $P_{TX}$ match perfectly with the Aois realized by the optimal choices of $P_{TX}$.

In Fig. 8, the performance of the considered OMA and NOMA grant-free schemes are compared by using the following Aoi ratio, $\frac{\Delta}{\Delta N}$. For the special case of $N = 1$ and $K = 2$, Proposition 1 predicts that this ratio is $0.5653$ for large $M$, which is confirmed by Fig. 8. If $K$ is fixed, i.e., $K = 2$, an increase of $N$ does not change the ratio significantly, particularly in the case of large $M$. By introducing more SNR levels, i.e., increasing $K$, the AoI ratio can be further reduced,
Fig. 7. Illustration of the impact of $P_{TX}$ on the AoI achieved by the considered grant-free schemes for GAR. $T = 6$, $P = 20$ dB, $R = 0.5$ BPCU, $K = 2$, $N = 1$, and $M = 8$. For the analytical results, Lemma 2 is used.

Fig. 8. Impact of the number of users on the ratio between the AoI achieved with the NOMA and OMA, i.e., $\frac{\Delta \bar{\tau}}{\Delta \bar{\tau}}$, for GAR. $T = 6$, $P = 20$ dB, $R = 0.5$ BPCU, and the adaptive choices of $P_{TX}$ are used. For the analytical results, Proposition 1 is used.

Fig. 9. AoI achieved by the considered grant-free transmission schemes for different data generation models. $N = 8$, $T = 6$, $R = 0.5$ BPCU, and the adaptive choices of $P_{TX}$ are used. For the analytical results, Lemma 1 is used.

which means that the performance gain of NOMA over OMA can be increased by introducing more SNR levels. This is expected since increasing $K$ reduces the likelihood of user collisions and ensures that users can update the base station earlier.

For all the previous simulation results, GAR has been considered, which means that a user’s update is generated at the beginning of a time frame, instead of at the beginning of a time slot as for GAW. In Fig. 9, the AoI achieved by the considered grant-free transmission schemes for the two different data generation models is illustrated. As can be seen from the figure, for both GAR and GAW, NOMA assisted grant-free transmission always outperforms the OMA based scheme. In addition, the figure shows that the AoI realized by the considered schemes for GAW is smaller than that for GAR, because, for GAW, each update is generated right before its delivery time, i.e., there is no service delay $S_j$.

Recall that the AoI analysis shown in this paper is based on the assumption that the users’ channels are independent and identically complex Gaussian distributed, i.e., only small-scale multi-path fading is considered. In Fig. 10(a), the AoI achieved by OMA and NOMA is investigated for the case where all users are uniformly distributed within a disc with radius $D_R$. The adaptive choices of $P_{TX}$ are used. $D_1$ denotes the distance between $U_1$ and the base station, $T = 6$, $R = 0.5$ BPCU, and $N = 8$.

Fig. 10. Impact of the number of users on the average AoI for GAR, when the users are uniformly distributed within a disc with radius $D_R$. The adaptive choices of $P_{TX}$ are used, $D_1$ denotes the distance between $U_1$ and the base station, $T = 6$, $R = 0.5$ BPCU, and $N = 8$.

V. CONCLUSION

In this paper, the impact of NOMA on the AoI of grant-free transmission has been investigated by applying a particular
form of NOMA, namely NOMA-assisted random access. By modeling grant-free transmission as a Markov chain and accounting for SIC, closed-form analytical expressions for the AoI achieved by NOMA assisted grant-free transmission have been obtained, and asymptotic studies have been carried out to demonstrate that the use of the simplest form of NOMA is already sufficient to reduce the AoI of OMA by more than 40%.

Concise and insightful analytical results have been developed for the special case of $N = 1$ time slot. An important direction for future research is the development of similar insightful results for the general case of $N \geq 1$. As shown in the paper, the choices of the users’ transmission probabilities, $\mathbb{P}_{\text{TX}}$, are crucial for performance maximization of grant-free transmission. A future promising direction is to apply deep reinforcement learning for optimizing $\mathbb{P}_{\text{TX}}$, where the insight regarding $\mathbb{P}_{\text{TX}}$ obtained in this paper can be used for initialization. Furthermore, it is expected that, in future 6G networks, grant-based and grant-free users will coexist, and hence another important future direction is how to apply NOMA and reduce the AoI of both types of users simultaneously.

VI. ACKNOWLEDGMENT

The authors would like to thank Dr. Jinho Choi for his kind suggestions about the implementation of NOMA assisted random access.

APPENDIX A

PROOF FOR LEMMA 1

The proof is divided into three parts to evaluate $P_{j,j}$, $P_{j,j+1}$, and $P_{j,j+1}$, $i \geq 2$, respectively. Throughout the proof, the high SNR assumption is made, which ensures that all the SNR levels, $P_k$, $1 \leq k \leq K$, are feasible for each user, i.e., transmission failures are due to user collisions only. The users’ channel gains in different time slots are assumed to be independent and identically complex Gaussian distributed with zero mean and unit variance.

A. Evaluating $P_{j,j}$

To find the expression for $P_{j,j}$, assume that $j$ users have successfully delivered their updates to the base station. Therefore, each of the remaining $M-j$ users independently makes an attempt to transmit with the probability, $\mathbb{P}_{\text{TX}}$, at a randomly chosen SNR level. $P_{j,j}$ is the probability of the event that none of the $M-j$ users succeeds.

Define $E_{p_{ij}}$ as the event that given $M-j$ remaining users, a user successfully updates the base station by using the $k$-th SNR level, $P_k$, and no user succeeds at $P_i$, $i < k$. The reason to include the constraint that no user succeeds at $P_i$, $i < k$, in the definition of $E_{p_{ij}}$ is to ensure that $E_{p_{ij}}$ and $E_{p_{ij}}$, $k < p$, are uncorrelated. For example, the event that $U_i$ succeeds by using $P_2$ and $U_j$ succeeds by using $P_3$ belongs to $E_{p_{ij}}$, only, and is not included in $E_{p_{ij}}$.

Therefore, the probability $P_{j,j}$ can be expressed as follows:

$$P_{j,j} = 1 - \mathbb{P}(E_{p_{1j}} \cup \cdots \cup E_{p_{Kj}})$$

$$= 1 - \sum_{k=1}^{K} \mathbb{P}(E_{p_{kj}}). \quad (16)$$

Further define $E_{m_{ij}}$ as the event that among the $M-j$ remaining users, there are $m$ active users which make the transmission attempts, and define $E_{p_{ij}|m}$ as the event that among the $m$ active users, a single user successfully updates the base station by using the $k$-th SNR level, $P_k$, and no user chooses $P_i$, $i < k$. Therefore, $\mathbb{P}(E_{p_{ij}|m}) = \sum_{m=1}^{M-j} \binom{M-j}{m} \mathbb{P}(E_{m_{ij}}) \mathbb{P}(E_{p_{ij}|m})$, which means

$$P_{j,j} = 1 - \sum_{k=1}^{K} \sum_{m=1}^{M-j} \binom{M-j}{m} \mathbb{P}(E_{m_{ij}}) \mathbb{P}(E_{p_{ij}|m}). \quad (17)$$

By using the transmission attempt probability of $\mathbb{P}_{\text{TX}}$, the probability, $E_{m_{ij}}$, can be obtained as follows:

$$\mathbb{P}(E_{m_{ij}}) = \mathbb{P}_{\text{TX}}^m(1 - \mathbb{P}_{\text{TX}})^{M-j-m}. \quad (18)$$

Without loss of generality, assume that $U_i$ is one of the $m$ active users which make transmission attempts. The probability for the event that $U_i$ chooses $P_k$, and no user chooses $P_i$, $i < k$, is given by

$$\mathbb{P}_K(1 - k\mathbb{P}_K)^{m-1}, \quad (19)$$

where $(1 - k\mathbb{P}_K)$ is the probability of the event that a user which is not $U_i$ cannot choose $P_p$, $1 \leq p \leq k$. Therefore, $\mathbb{P}(E_{p_{ij}|m})$ can be approximated as follows:

$$\mathbb{P}(E_{p_{ij}|m}) \approx m\mathbb{P}_K(1 - k\mathbb{P}_K)^{m-1}, \quad (20)$$

since each of the $m$ active users can be the successful user with the equal probability, where the high SNR assumption is used, i.e., all SNR levels are feasible to each user and only the errors caused by user collisions are considered. By combining $(17)$, $(18)$, and $(20)$, $P_{j,j}$ can be expressed as follows:

$$P_{j,j} \approx 1 - \sum_{m=1}^{M-j} \binom{M-j}{m} \mathbb{P}_{\text{TX}}^m(1 - \mathbb{P}_{\text{TX}})^{M-j-m}$$

$$\times \sum_{k=1}^{K} m\mathbb{P}_K(1 - k\mathbb{P}_K)^{m-1}. \quad (21)$$

B. Evaluating $P_{j,j+1}$

Recall that $P_{j,j+1}$ is also conditioned on the assumption that $j$ users have successfully updated the base station, and $P_{j,j+1}$ is the probability of the event that there is a single successful update from a user which cannot be $U_i$.

Define $E_{p_{ij}}$ as the event that given $M-j$ remaining users, a single user, other than $U_i$, successfully updates the base station by using the $k$-th SNR level. At high SNR, the following two conclusions can be made regarding $E_{p_{ij}}$.

On the one hand, due to the feature of SIC, $E_{p_{ij}}$ implies that no user chooses $P_i$, $i < k$, which can be shown by contradiction. Assume that $U_p$ chooses $P_i$. If $U_p$ is the only user choosing $P_i$, $U_p$ becomes an additional successful user, which contradicts the assumption that there is a single successful user. If multiple users choose $P_i$, a collision occurs and the assumption that there is a single successful update happens at $P_i$. On the other hand, $E_{p_{ij}}$ does not exclude the event that an SNR level, $P_i$, $i > k$, is chosen by a user; however, $E_{p_{ij}}$ does imply that if $P_i$, $i > k$, is chosen, a collision must happen at this SNR level, otherwise there will be an additional successful user. We also note that $E_{p_{ij}}$ is different from $E_{p_{ij}}$ for the following two reasons. First, $E_{p_{ij}}$ does not exclude the event

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
that the successful user is $U_1$. Second, for $E_{P_{k|j}}$, it is still possible for a user to succeed at SNR levels, $P_i$, $i > k$.

By using $E_{P_{k|j}}$, the probability $P_{j,j+1}$ can be expressed as follows:

$$P_{j,j+1} = \mathbb{P}(E_{L_{j+1}}^1 \cup \cdots \cup E_{L_k}^1) = \sum_{k=1}^{K} \mathbb{P}(E_{L_k}^1),$$

(22)

where the last step follows by the fact that the $E_{L_k}^1$, $1 \leq k \leq K$, are uncorrelated.

Similar to $E_{P_{k|m}}$, define $E_{P_{k|m}}^1$ as the event that among the $m$ active users, a single user, other than $U_1$, successfully updates the base station by using the $k$-th SNR level. By using $E_{m|j}$ and $E_{P_{k|m}}^1$, $P_{j,j+1}$ can be expressed as follows:

$$P_{j,j+1} = \sum_{k=1}^{K} \sum_{m=1}^{M-j} \binom{M-j}{m} \mathbb{P}(E_{m|j}) \mathbb{P}(E_{P_{k|m}}^1).$$

(23)

The analysis of $\mathbb{P}(E_{P_{k|m}}^1)$ is challenging. First, we assume that $m \geq 2$, i.e., there are more than one active users. For illustrative purposes, assume that $U_2$ is an active user. Denote by $E_{P_{k}|m}^+$, the event that $U_2$ succeeds by using $P_k$, and $\mathbb{P}(E_{P_{k}|m}^+)$ is given by

$$\mathbb{P}(E_{P_{k}|m}^+) \approx \mathbb{P}_K(1 - k\mathbb{P}_K)^{m-1},$$

(24)

where the high SNR approximation is used, and the reason to have $(1 - k\mathbb{P}_K)^{m-1}$ is that the $k$ highest SNR levels are no longer available to the other $m - 1$ active users. Because $m \geq 2$, $k \leq K-1$, i.e., $U_2$ cannot succeed by using $P_K$, which can be explained by using a simple example with $U_2$ and $U_3$ being the active users ($m = 2$). As discussed previously, if $U_2$ chooses $P_K$, $U_3$ has to choose $P_p$, $p > K$, which is not possible.

Furthermore, denote by $E_{P_{k}|m}^-$ the event that $U_2$ succeeds by using $P_k$, and there is at least one additional user which succeeds by using $P_i$, $i \geq k + 1$. $\mathbb{P}(E_{P_{k}|m}^-)$ is given by

$$\mathbb{P}(E_{P_{k}|m}^-) \approx \mathbb{P}_K \sum_{k=1}^{K} \binom{M-j}{m} \mathbb{P}_K(1 - k\mathbb{P}_K)^{m-2},$$

(25)

which is obtained in a similar manner as $\mathbb{P}(E_{P_{k}|m}^-)$ in (20).

By using $E_{P_{k}|m}^+$ and $E_{P_{k}|m}^-$, the probability for the event that among the $m$ active users, only $U_2$ succeeds by using $P_k$, denoted by $E_{2k}$, is given by

$$\mathbb{P}(E_{2k}) = \mathbb{P}(E_{P_{k}|m}^+) - \mathbb{P}(E_{P_{k}|m}^-) \approx \mathbb{P}_K \left[ (1 - k\mathbb{P}_K)^{m-1} - \sum_{k=1}^{K} \binom{M-j}{m} \mathbb{P}_K(1 - k\mathbb{P}_K)^{m-2} \right].$$

(26)

Intuitively, $\mathbb{P}(E_{P_{k}|m}^+)$ should be simply $\mathbb{P}(E_{P_{k}|m}^+)/(m-1)$, i.e., including $(m-1)$ cases corresponding to $U_i$, $2 \leq i \leq m$. If $U_1$ is one of the active users, indeed, $\mathbb{P}(E_{P_{k}|m}^-) = (m-1)\mathbb{P}(E_{2k})$. However, if $U_1$ is not an active user, $\mathbb{P}(E_{P_{k}|m}^-) = m\mathbb{P}(E_{2k})$. Therefore, by taking into account the fact that $U_1$ might not be an active user, $\mathbb{P}(E_{P_{k}|m}^-)$ can be expressed as follows:

$$\mathbb{P}(E_{P_{k}|m}^-) \approx \left[ \frac{m}{M-j} \binom{M-j}{m} - \frac{M-j-m}{M-j} \right] \mathbb{P}(E_{2k})$$

$$= \frac{M-j-1}{M-j} \mathbb{P}_K [ (1 - k\mathbb{P}_K)^{m-1} - \sum_{k=1}^{K} (m-1)\mathbb{P}_K(1 - k\mathbb{P}_K)^{m-2} ],$$

(27)

for the case $m \geq 2$.

For the case $m = 1$, i.e., there is a single active user, $\mathbb{P}(E_{P_{k}|m}^-)$ is simply zero if this user is $U_1$. Otherwise, $\mathbb{P}(E_{P_{k}|m}^-) = \mathbb{P}_K$, i.e., $P_k$ is selected by the active user.

Therefore, for $m = 1$, $\mathbb{P}(E_{P_{k}|m}^-)$ is given by

$$\mathbb{P}(E_{P_{k}|m}^-) \approx \frac{M-j-1}{M-j} \mathbb{P}_K.$$

(28)

By combining (23), (27) and (28), the probability, $P_{j,j+1}$, can be obtained as shown in the lemma.

C. Evaluating $P_{j,j+i}$, $i \geq 2$

We note that for $P_{j,j+1}$, there is a single successful update, and $P_{j,j+1}$ has been analyzed by first specifying which SNR level, i.e., $P_k$, is used for this successful update. The same method could be applied to analyze $P_{j,j+i}$, $i \geq 2$; however, the resulting expression can be extremely complicated if the number of the used SNR levels is large.

Instead, the feature of SIC can be used to simplify the analysis of $P_{j,j+i}$. Consider the following example with $i = 3$ and $K = 8$. Assume that the following three SNR levels, $P_2$, $P_4$, and $P_5$, are used by the successful users. The key observation for simplifying the performance analysis is that those SNR levels prior to $P_2$ and between the chosen ones are not selected by any users, e.g., no user chooses $P_1$ and $P_3$. This observation can be explained by taking $P_3$ as an example. If this SNR level has been selected, a collision must happen at this level, otherwise there should be an additional successful user. However, if a collision does happen at $P_3$, SIC needs to be terminated in the third SIC stage, and hence no successful update can happen at $P_3$, which contradicts to the assumption that $P_3$ is used by a successful user. As a result, no one chooses $P_3$.

By using this observation, we note that only two SNR levels are significant to the analysis of $P_{j,j+i}$, namely the highest and the lowest SNR levels, which are denoted by $P_{k_1}$ and $P_{k_2}$, respectively. For the aforementioned example, $P_{k_1} = P_3$ and $P_{k_2} = P_5$. Define $E_{P_{k|j+1}}$ as the event that given $M-j$ remaining users, i users which are not $U_1$ successfully update the base station, where $P_{k_1}$ and $P_{k_2}$ are the highest and lowest used SNR levels. By using $E_{P_{k|j+1}}$, the probability $P_{j,j+i}$ can be expressed as follows:

$$P_{j,j+i}$$

$$= \sum_{k_1=1}^{K} \sum_{k_2=k_1+i-1}^{K} \mathbb{P}(E_{L_{k+1}}^1).$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where $E_{P_i|m}$ is defined similar to $E_{P_k|j}$ with the assumption that there are $m$ active users.

In order to find $\mathbb{P}(E_{P_i|m})$, again, we first focus on the case $m \geq i + 1$, i.e., there are more $i$ active users. Define $E_{ik}$ as the probability for the particular event that among the $m$ active users, $U_2$ succeeds by using $P_{k_i}$, $U_j$ succeeds by using $P_{k_{j-1}}$, $2 \leq j \leq i$, and $U_{i+1}$ succeeds by using $P_{k_i}$. Similar to $\mathbb{P}(E_{2k})$, $\mathbb{P}(E_{ik})$ can be approximated at high SNR as follows:

$$\mathbb{P}(E_{ik}) \approx \mathbb{P}_K \left[(1 - k_2 P_K)^{m-i} - \sum_{k=k_2+1}^{K} (m-i)\mathbb{P}_K(1 - k P_K)^{m-i-1} \right],$$  

(30)

where the term, $(1 - k_2 P_K)^{m-i}$, is due to the fact that the remaining $m-i$ active users can choose $P_i$, $i > k_2$, only, and the second term in the bracket is obtained similar to $\mathbb{P}(E_{P_k|m})$ in (20).

Following steps similar to those to obtain $\mathbb{P}(E_{P_{i-1}|m})$ in (27), $\mathbb{P}(E_{P_{i-1}|m})$ can be obtained from $\mathbb{P}(E_{ik})$ as follows:

$$\mathbb{P}(E_{P_{i-1}|m}) \approx \mathbb{P}(E_{ik}) \binom{k_2 - k_1 - 1}{i-2} \left[\frac{m}{M-j}(m-1)(m-2)\cdots(m-i)\right] \frac{M-j-m}{M-j}m \cdots (m-i+1),$$  

(31)

for the case $m \geq i + 1$, where $(k_2-k_1-1)$ is the number of the possible choices for the $(i-2)$ SNR levels which are between $P_{k_1}$ and $P_{k_2}$.

The special case $m = i$ means that there are $m$ active users and each of the active users is a successful user. Therefore, the probability in (30), $\mathbb{P}(E_{ik})$, is simply $P_i^i$, and hence for the case $m = i$, $\mathbb{P}(E_{P_{i-1}|m})$ can be expressed as follows:

$$\mathbb{P}(E_{P_{i-1}|m}) \approx \mathbb{P}_K \binom{k_2 - k_1 - 1}{i-2} \frac{M-j-i}{M-j}m \cdots (m-i+1).$$  

(32)

By combining (31) and (32), the expression of $P_{j,i+1}$ can be obtained as shown in the lemma. Therefore, the proof for the lemma is complete.

APPENDIX B

PROOF FOR LEMMA 2

The proof comprises first simplifying the state transition probabilities, then developing an asymptotic expression of the AoI, and finally finding the optimal choice of $\mathbb{P}_{TX}$.

A. Simplifying the State Transition Probabilities

For the case of $N = 1$ and $K = 2$, only the following transition probabilities, $P_{0,0}$, $P_{0,1}$, $P_{0,2}$, need to be focused on.

In particular, the expression of $P_{0,0}$ can be simplified as follows:

$$P_{0,0} \approx 1 - \sum_{m=1}^{M} \binom{M}{m} \mathbb{P}_{TX}(1 - \mathbb{P}_{TX})^{M-m} \times \sum_{k=1}^{K} m \mathbb{P}_K(1 - k \mathbb{P}_K)^{m-1} \left[\binom{m}{1} \mathbb{P}_{TX}^{m-2} + (1 - \mathbb{P}_{TX})^{M-1} \mathbb{P}_K \right].$$  

$$P_{0,1} \approx M \mathbb{P}_{TX}(1 - \mathbb{P}_{TX})^{M-1} \frac{M-1}{M} \mathbb{P}_K \times \sum_{k=2}^{K} \binom{M}{m} \mathbb{P}_{TX}^{m-2} \times \left[\binom{m}{1} \mathbb{P}_K(1 - k \mathbb{P}_K)^{m-2} \right].$$  

(33)

Define $\tau(m) = \sum_{k=1}^{K} \frac{M-1}{M} m \mathbb{P}_K[(1 - k \mathbb{P}_K)^{m-1} - \sum_{k=k+1}^{K} \binom{m-1}{1} \mathbb{P}_K(1 - k \mathbb{P}_K)^{m-1}]$. Note that in the expression of $\tau(m)$, $k \geq 1$, and hence $\kappa \geq 2$. For the case of $K = 2$, $1 - 2 \mathbb{P}_K = 0$. By using this observation, $\tau(m)$ can be simplified as follows:

$$\tau(m) = \frac{m(M-1)}{M} \mathbb{P}_K(1 - \mathbb{P}_K)^{m-1},$$  

for $m > 2$, and

$$\tau(2) = \left(\frac{m(M-1)}{M} + \frac{M-m}{M} \right) \mathbb{P}_K[(1 - \mathbb{P}_K) - \mathbb{P}_K] = 0.$$  

(34)
By using the simplified expression of \( \tau(m) \), \( P_{0,1} \) can be simplified as follows:

\[
P_{0,1} \approx MP_{TX}(1 - P_{TX})^{M-1} \frac{M-1}{M} \sum_{m=0}^{\frac{P_{TX}}{M}} \binom{M}{m} P_{TX}^m
\]

where the last step follows by the fact that \( 1 - P_K = P_K \).

Finally, \( P_{0,2} \) can be rewritten as follows:

\[
P_{0,2} \approx \left( \frac{M}{2} \right) P_{TX}(1 - P_{TX})^{M-2} \sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^{K} \frac{M-i}{M}
\]

\[
\times \binom{k_2-k_1-1}{0} \prod_{p=0}^{1} (2-p)
\]

\[
+ \sum_{m=0}^{M} \binom{P_{TX}}{m} (1 - P_{TX})^{M-m} \sum_{k_1=1}^{K-2} \sum_{k_2=k_1+1}^{K-1} \frac{M-i}{M}
\]

\[
\times \binom{k_2-k_1-1}{0} \prod_{p=0}^{1} (m-p)
\]

\[
\times \left[ (1-k_2P_K)^{m-2} - \sum_{k=1}^{K-3} (m-2)P_K (1-k^2P_K)^{m-3} \right]
\]

\[
= \left( \frac{M}{2} \right) P_{TX}(1 - P_{TX})^{M-2} \frac{M-2}{M} P_{TX}^2
\]

\[
= \left( \frac{M-1}{M} \right) P_{TX}(1 - P_{TX})^{M-2} P_{TX}^2,
\]

where step (a) follows by employing the properties of the binomial coefficients.

**B. Asymptotic Studies of AoI**

By using the above transition probabilities, the probability for the event that \( U_1 \) cannot complete an update within one frame, \( P_{\text{fail}} \), can be simplified as follows:

\[
P_{\text{fail}} \approx P(Z > N) = s_0T_P N M 1 = \sum_{j=0}^{2} P_{0,j}
\]

\[
= 1 - M(x + 1 - P_{TX})^{M-1}x - M(1 - P_{TX})^{M-1}x
\]

\[
+ 2(M-1)(1 - P_{TX})^{M-1}x
\]

\[
+ (M-1) \sum_{i=2}^{M} \binom{M-1}{i} (1 - P_{TX})^{M-i-1}x ^{i+1}
\]

\[
+ \frac{(M-1)!}{(M-3)!} (1 - P_{TX})^{M-2}x^2
\]

\[
\tau_0 = x(M-1) \sum_{i=2}^{M} \binom{M-1}{i} (1 - P_{TX})^{M-i-1}x ^{i}
\]

\[
r = (1 - P_{TX} + x)(M-1)x - (1 - P_{TX})^{M-1}(M-1)x
\]

\[
- (M-1)(1 - P_{TX})^{M-2}(M-1)x^2.
\]

By using the simplified expression of \( \tau_0 \), \( P_{\text{fail}} \) can be rewritten as follows:

\[
P_{\text{fail}} \approx 1 - M(x + 1 - P_{TX})^{M-1}x - M(1 - P_{TX})^{M-1}x
\]

\[
+ 2(M-1)(1 - P_{TX})^{M-1}x + (M-1)(M-2)(1 - P_{TX})^{M-2}x^2
\]

\[
= 1 - P_{TX}P_K + 1 - P_{TX}P_K
\]

\[
- (1 - P_{TX} + (M-1)P_{TX}P_K)(1 - P_{TX})^{M-2}P_{TX}P_K.
\]

The simplified expression of \( P_{\text{fail}} \) can be used to facilitate the asymptotic studies of the AoI. For the case of \( N = 1 \), the pmf of the access delay can be obtained from \( P_{\text{fail}} \) as follows:

\[
P(S_j = T) = \frac{s_0T_P P_{0,M}^N 1}{1 - s_0T_P P_{0,M} 1} = \frac{P_{0,M} 1}{1 - P_{0,M}},
\]

where \( p_{0,M} \) is the first element of \( p \). As a result, \( E\{S_j\} \) and \( E\{S_j^2\} \) can be obtained as follows:

\[
E\{S_j\} = T \sum_{n=1}^{s_0T_P P_{0,M} 1} n \frac{s_0T_P P_{0,M} 1}{1 - s_0T_P P_{0,M} 1} = T,
\]

\[
E\{S_j^2\} = T^2 \sum_{n=1}^{s_0T_P P_{0,M} 1} n^2 \frac{s_0T_P P_{0,M} 1}{1 - s_0T_P P_{0,M} 1} = T^2,
\]

which are expected since \( S_j \) becomes a deterministic parameter for this special case.

Furthermore, recall that the inter-departure time \( Y_j \) can be rewritten as follows:

\[
Y_j = (NT - S_j) + (X_j - 1)NT + S_j.
\]

Therefore, for the case of \( N = 1 \), the expectation of \( Y_j \) can be simplified as follows:

\[
E\{Y_j\} = E\{(NT - S_j) + (X_j - 1)NT + S_j\}
\]

\[
= TNE\{X_j\} = TN \frac{1}{1 - P_{\text{fail}}},
\]

which is obtained by using the fact that \( X_j \) follows the geometric distribution. Similarly, the expectation of \( Y_j^2 \) can be simplified as follows:

\[
E\{Y_j^2\} = T^2E\{X_j^2\} + 2E\{S_j^2\} - 2E\{S_j\}^2
\]

\[
= T^2 \frac{1 + P_{\text{fail}}}{1 - P_{\text{fail}}} + 2T^2 - 2T^2.
\]

Finally, for the case of \( N = 1 \), the averaged AoI can be expressed as follows:

\[
\Delta N = \frac{E\{S_j\}E\{Y_j\} - E\{S_j^2\} + E\{S_j\}E\{Y_j\}}{E\{Y_j\} + E\{S_j\}} + \frac{E\{Y_j^2\}}{2E\{Y_j\}}.
\]
\[
\frac{\bar{E} \{ Y_j \} - T^2 + T^2}{2 \bar{E} \{ Y_j \}} = T + \frac{NT \{ 2 - f(P_{TX}) \}}{2 f(P_{TX})}, \quad (47)
\]
where \( f(P_{TX}) = \left( P_{TX} P_K + 1 - P_{TX} \right)^{M-1} P_{TX} P_K + (1 - P_{TX} + (M - 1) P_{TX} P_K) (1 - P_{TX})^{M-2} P_{TX} P_K. \)

### C. Finding the Optimal Choice of \( P_{TX} \)

The considered AoI minimization problem can be expressed as follows: \( \min_{0 \leq P_{TX} \leq 1} \Delta^N \). It is challenging to show whether \( \Delta^N \) is a convex function of \( P_{TX} \), given the complex expression of \( \Delta^N \). However, it can be shown that \( \Delta^N \) first decreases and then increases as \( P_{TX} \) grows, as shown in the following.

The first order derivative of \( \Delta^N \) with respect to \( P_{TX} \) is given by
\[
\Delta^N \approx \frac{NT}{2} \left( \frac{1}{f(P_{TX})} - \frac{2 f'(P_{TX}) f(P_{TX}) - f'(P_{TX}) \bar{f}(P_{TX})}{f^2(P_{TX})} \right) = -\frac{NT}{2} \frac{f'(P_{TX}) f(P_{TX})}{f^2(P_{TX})},
\]
which shows that the monotonicity of \( \Delta^N \) is decided by the sign of \( f'(P_{TX}) \) only. In the following, we will first show that \( f'(P_{TX}) = 0 \) has a single root for \( 0 \leq P_{TX} \leq 1 \).

With some straightforward algebraic manipulations, \( f'(P_{TX}) \) can be expressed as follows:
\[
f'(P_{TX}) = \left[ 1 - M \frac{P_{TX}}{2} \right] \left( 1 - \frac{P_{TX}}{2} \right)^{M-2} + \left[ 1 - 2P_{TX} - M(M-3) \frac{P_{TX}}{2} \right] (1 - P_{TX})^{M-3}.
\]

Because \( 0 \leq P_{TX} \leq 1 \), both \( \left( 1 - \frac{P_{TX}}{2} \right) ^{-2} \) and \( (1 - P_{TX}) \) are positive.

On the one hand, if \( P_{TX} \geq \frac{2}{M} \), then \( 1 - M \frac{P_{TX}}{2} \geq 0 \). On the other hand, there are two roots for \( -\frac{\sqrt{4 + 2M(M-1) - 4}}{M(M-3)} \) and \( \frac{\sqrt{4 + 2M(M-1) - 4}}{M(M-3)} \).

When \( M \rightarrow \infty \), the two roots can be approximated as follows:
\[
\pm \sqrt{4 + 2M(M-1) - 4} \approx \pm \sqrt{\frac{2}{M}}.
\]
Therefore, \( 1 - 2P_{TX} - M(M-3) \frac{P_{TX}}{2} \leq 0 \) if \( P_{TX} \geq \frac{\sqrt{2}}{M} \), otherwise \( 1 - 2P_{TX} - M(M-3) \frac{P_{TX}}{2} > 0 \). Denote the root(s) of \( f'(P_{TX}) = 0 \) by \( P_{TX}^* \), which is bounded as follows:
\[
\frac{\sqrt{2}}{M} \leq P_{TX}^* \leq \frac{2}{M}.
\]

A key observation from (50) is that the upper and lower bounds on \( P_{TX}^* \) are of the same order of \( \frac{1}{M} \). Therefore, \( P_{TX}^* \) can be expressed as \( P_{TX}^* = \frac{\eta}{M} \), where \( \sqrt{2} \leq \eta \leq 2 \). By using this expression for \( P_{TX}^* \), \( f'(P_{TX}) = 0 \) can be expressed as follows:
\[
0 = \left( 1 - \frac{\eta}{2} \right) \left( 1 - \frac{\eta}{2M} \right)^{M-2} + \left( 1 - \frac{\eta}{M} \right)^{M-3}.
\]

In order to find an explicit expression of \( P_{TX} \), we note that \( f_x(x) = (1 - \frac{a}{x})^2 \) can be approximated at \( x \rightarrow \infty \) as follows:
\[
\ln f_x(x) = \lim_{x \rightarrow \infty} x \ln \left( 1 - \frac{a}{x} \right) = \lim_{z \rightarrow 0} \ln \left( 1 - \frac{1}{z} \right) = -\ln \frac{1}{z} - 1 = -a.
\]

which implies \( (1 - \frac{a}{M})^M = e^{-a} \) for large \( M \).

Therefore, for \( M \rightarrow \infty \), (51) can be approximated as follows:
\[
0 = (1 - \frac{\eta}{2}) e^{-\eta} + \left( 1 - \frac{\eta^2}{2} \right) e^{-\eta}, \quad (53)
\]

where the value of \( \eta \) can be straightforwardly obtained by applying off-the-shelf root solvers. It is important to point out that (53) has a single root, which means that \( P_{TX} = \frac{\eta}{M} \) is the single root of \( f'(P_{TX}) = 0 \). Therefore, \( P_{TX} = \frac{\eta}{M} \) is the optimal choice of \( P_{TX} \) to minimize \( \Delta^N \), and the proof is complete.

### Appendix C

**Proof for Proposition 1**

Based on Lemma 2, the optimal choice for the transmission probability is given by \( P_{TX} = \frac{\eta}{M} \). By using this optimal choice of \( P_{TX} \), \( P_{fail} \) can be expressed as follows:
\[
P_{fail} \approx 1 - \left( 1 - \frac{\eta}{2M} \right)^{M-1} \frac{\eta}{2M} - \left( 1 - \frac{\eta}{M} + (M - 1) \frac{\eta}{2M} \right) \left( 1 - \frac{\eta}{M} \right)^{M-2} \frac{\eta}{2M}.
\]

Again applying the following limit: \( \lim_{M \rightarrow \infty} (1 - \frac{a}{M})^M = e^{-a} \), \( P_{fail} \) can be approximated for large \( M \) as follows:
\[
P_{fail} \approx 1 - e^{-\frac{\eta}{2M}} - \left( 1 - \frac{\eta}{M} + (M - 1) \frac{\eta}{2M} \right) e^{-\frac{\eta}{2M}} \frac{\eta}{2M} \approx 1 - e^{-\frac{\eta}{2M}} - \left( 1 + \frac{\eta}{2} \right) e^{-\frac{\eta}{2M}} \frac{\eta}{2M}.
\]

By using this approximation of \( P_{fail} \), the AoI achieved by NOMA can be approximated as follows:
\[
\Delta^N \approx \frac{NT}{2} - \frac{e^{-\frac{\eta}{2M}} - \left( 1 + \frac{\eta}{2} \right) e^{-\frac{\eta}{2M}}}{e^{-\frac{\eta}{2M}} + \left( 1 + \frac{\eta}{2} \right) e^{-\frac{\eta}{2M}}} - 1
\]
\[
\approx \frac{NT}{2} \frac{e^{-\frac{\eta}{2M}} - \left( 1 + \frac{\eta}{2} \right) e^{-\frac{\eta}{2M}}}{e^{-\frac{\eta}{2M}} + \left( 1 + \frac{\eta}{2} \right) e^{-\frac{\eta}{2M}}}.
\]

In order to find an explicit expression for the AoI achieved by OMA for the case of \( N = 1 \), the transition probabilities can be expressed as follows:
\[
P_{0,0} = 1 - MP_{TX} (1 - P_{TX})^{M-1},
\]
and
\[
P_{0,1} = (M - 1) P_{TX} (1 - P_{TX})^{M-1}.
\]
Therefore, for OMA, the failure probability, $P_{\text{fail}}$, is given by

$$P_{\text{fail}} = P_{0,0} + P_{0,1} = 1 - P_{TX} (1 - P_{TX})^{-1}. \quad (59)$$

By using the expression for $P_{\text{fail}}$, the AoI achieved by OMA can be expressed as follows:

$$\Delta^O = 1 + \frac{NT}{2} \left( \frac{1}{1 + P_{\text{fail}}} \right) \left( 1 - P_{TX} (1 - P_{TX})^{-1} \right)^{-1}.
= 1 + \frac{NT}{2 \left( 1 - P_{TX} (1 - P_{TX})^{-1} \right)}.
\quad (60)$$

Define $f_O(P_{TX}) = P_{TX} (1 - P_{TX})^{-1}$. Similar to the NOMA case, it is straightforward to show that the optimal choice of $P_{TX}$ for OMA is the root of $f_O(P_{TX}) = 0$. Note that $f_O'(P_{TX})$ can be expressed as follows:

$$f_O'(P_{TX}) = \left( 1 - M P_{TX} (1 - P_{TX})^{-2} \right),
\quad (61)$$

which is the reason why $P_{TX} = \frac{1}{M}$ is optimal for the OMA case. By using $P_{TX} = \frac{1}{M}$, $P_{\text{fail}}$ can be expressed as follows:

$$P_{\text{fail}} = 1 - \frac{1}{M} \left( \frac{1}{1 - \frac{1}{M}} \right)^{M-1},$$

and hence the AoI achieved by OMA is given by

$$\Delta^O = 1 + \frac{MT}{2} \left( \frac{1}{1 - P_{TX} (1 - P_{TX})^{-1}} \right) = 1 + \frac{NT}{2 \left[ \frac{M}{1 - \frac{1}{M}} \right]^{M-1} - 1}. \quad (62)$$

For $M \to \infty$, the AoI achieved by OMA can be approximated as follows:

$$\Delta^O \approx \frac{NT}{2} \left( \frac{2M}{1 - \frac{1}{M}} \right)^{M-1} \approx NT e^M. \quad (63)$$

Therefore, for $M \to \infty$, the ratio between the AoI realized by NOMA and OMA is given by

$$\frac{\Delta^M}{\Delta^O} \approx \frac{NT}{\eta (e^2 + 1 + 2)} \quad (64)$$

which completes the proof.

REFERENCES

[1] A. C. Cirik, N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, “Toward the standardization of grant-free operation and the associated NOMA strategies in 3GP,” IEEE Commun. Standards Mag., vol. 3, no. 4, pp. 60–66, Dec. 2019.
[2] X. You et al., “Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts,” Sci. China Inf. Sci., vol. 64, no. 1, pp. 1–74, Nov. 2021.
[3] R. Abbas, M. Shrivaminoh paddham, Y. Li, and B. Vucetic, “Random access for M2M communications with QoS guarantees,” IEEE Trans. Commun., vol. 65, no. 7, pp. 2889–2903, Jul. 2017.
[4] J. Choi, J. Ding, N.-P. Le, and Z. Ding, “Grant-free random access in machine-type communication: Approaches and challenges,” IEEE Wireless Commun., vol. 29, no. 1, pp. 151–158, Feb. 2022.
[5] L. Liu and W. Yu, “Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation,” IEEE Trans. Signal Process., vol. 66, no. 11, pp. 2933–2946, Jun. 2018.
[6] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, “Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the Internet of Things,” IEEE Signal Process. Mag., vol. 35, no. 5, pp. 88–99, Sep. 2018.
Zhiguo Ding (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications in 2000 and the Ph.D. degree from Imperial College London in 2005.

He is currently a Professor of communications with Khalifa University and has also been affiliated with The University of Manchester and Princeton University. His current research interests include 5G networks, game theory, cooperative and energy harvesting networks, and statistical signal processing. He is a Distinguished Lecturer of IEEE ComSoc and a Web of Science Highly Cited Researcher in two categories 2022. He received the EU Marie Curie Fellowship from 2012 to 2014, the Top IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY Editor in 2017, the IEEE Heinrich Hertz Award in 2018, the IEEE Jack Neubauer Memorial Award in 2018, the IEEE Best Signal Processing Letter Award in 2018, the Friedrich Wilhelm Bessel Research Award in 2020, and the IEEE SPCC Technical Recognition Award in 2021. He was an Editor of IEEE WIRELESS COMMUNICATION LETTERS, IEEE TRANSACTIONS ON COMMUNICATIONS, and IEEE COMMUNICATION LETTERS from 2013 to 2016. He is serving as an Area Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY and an Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS.

Robert Schober (Fellow, IEEE) received the Diplom (University) and Ph.D. degrees in electrical engineering from the Friedrich-Alexander University of Erlangen-Nuremberg (FAU), Germany, in 1997 and 2000, respectively.

From 2002 to 2011, he was a Professor and the Canada Research Chair with The University of British Columbia (UBC), Vancouver, Canada. Since January 2012, he has been an Alexander von Humboldt Professor and the Chair for Digital Communication with FAU. His current research interests include communication theory, wireless and molecular communications, and statistical signal processing.

Dr. Schober is a Fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada and a member of the German National Academy of Science and Engineering. He received several awards for his work, including the 2002 Heinz Maier-Leibnitz Award of the German Science Foundation (DFG), the 2004 Innovations Award of the Vodafone Foundation for Research in Mobile Communications, the 2006 UBC Killam Research Prize, the 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, the 2008 Charles McDowell Award for Excellence in Research from UBC, the 2011 Alexander von Humboldt Professorship, the 2012 NSERC E. W. R. Stacie Fellowship, the 2017 Wireless Communications Recognition Award by the IEEE Wireless Communications Technical Committee, and the 2022 IEEE Vehicular Technology Society Stuart F. Meyer Memorial Award. Furthermore, he received numerous best paper awards for his work, including the 2022 ComSoc Stephen O. Rice Prize and the 2023 ComSoc Leonard G. Abraham Prize. Since 2017, he has been listed as a Highly Cited Researcher by the Web of Science. He served as the Editor-in-Chief for the IEEE TRANSACTIONS ON COMMUNICATIONS, the VP Publications of the IEEE Communication Society (ComSoc), the ComSoc Member at Large, and the ComSoc Treasurer. Currently, he serves as a Senior Editor for the PROCEEDINGS OF THE IEEE and as the ComSoc President-Elect.

H. Vincent Poor (Life Fellow, IEEE) received the Ph.D. degree in EEC's from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton University, where he is currently the Michael Henry Strater University Professor. From 2006 to 2016, he served as the Dean of Princeton’s School of Engineering and Applied Science. He has also held visiting appointments at several other universities, including most recently at Berkeley and Cambridge.

His current research interests include information theory, machine learning and network science, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the recent book Machine Learning and Wireless Communications (Cambridge University Press, 2022). He is a member of the National Academy of Engineering and the National Academy of Sciences and a foreign member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. He received the IEEE Alexander Graham Bell Medal in 2017.