

A New QoS-Guarantee Strategy for NOMA Assisted Semi-Grant-Free Transmission

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

Abstract

Semi-grant-free (SGF) transmission has recently received significant attention due to its capability to accommodate massive connectivity and reduce access delay by admitting grant-free users to channels which would otherwise be solely occupied by grant-based users. In this paper, a new SGF transmission scheme that exploits the flexibility in choosing the decoding order in non-orthogonal multiple access (NOMA) is proposed. Compared to existing SGF schemes, this new scheme can ensure that admitting the grant-free users is completely transparent to the grant-based users, i.e., the grant-based users' quality-of-service experience is guaranteed to be the same as for orthogonal multiple access. In addition, compared to existing SGF schemes, the proposed SGF scheme can significantly improve the robustness of the grant-free users' transmissions and effectively avoid outage probability error floors. To facilitate the performance evaluation of the proposed SGF transmission scheme, an exact expression for the outage probability is obtained and an asymptotic analysis is conducted to show that the achievable multi-user diversity gain is proportional to the number of participating grant-free users. Computer simulation results demonstrate the performance of the proposed SGF transmission scheme and verify the accuracy of the developed analytical results.

I. INTRODUCTION

The next generation Internet of Things (NGIoT) is envisioned to be an important use case for beyond 5G mobile networks [1]. The key challenge for supporting NGIoT is, given the scarce radio spectrum, how to support a massive number of devices, each of which might send a small number of packets only. For this emerging application, conventional grant-based transmission is not suitable, since the amount of signalling needed for handshaking could exceed

Z. Ding and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA. Z. Ding is also with the School of Electrical and Electronic Engineering, the University of Manchester, Manchester, UK (email: zhiguo.ding@manchester.ac.uk, poor@princeton.edu). R. Schober is with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nurnberg (FAU), Germany (email: robert.schober@fau.de).

the amount of data sent by the devices. This motivates the development of grant-free transmission schemes, which grant the devices access without lengthy handshaking protocols [2]. Most existing grant-free schemes can be categorized into three groups. The first group applies random access protocols originally developed for computer networks [3], the second group relies on the excess spatial degrees of freedom offered by multiple-input multiple-output (MIMO) techniques [4], [5], and the third group employs non-orthogonal multiple access (NOMA) which encourages spectrum sharing among the devices [6]–[10]. It is noted that there are many works which use a combination of the three types of grant-free schemes and hence can potentially offer a significant performance improvement in terms of connectivity and transmission robustness [11]–[13].

In this paper, we focus on a special case of NOMA based grant-free transmission, termed semi-grant-free (SGF) transmission [14]. Unlike the aforementioned pure grant-free schemes, SGF transmission does not assume that a certain number of resource blocks, such as time slots or subcarriers, are reserved for contention among the grant-free users, since this assumption would put a strict cap on the number of grant-free users which can be served, particularly if the base station has a limited number of antennas and cannot use massive MIMO to improve connectivity. The key idea of SGF transmission is to opportunistically admit grant-free users to those resource blocks which would otherwise be solely occupied by grant-based users. An immediate advantage of SGF over conventional grant-free schemes is that the number of grant-free users is constrained not by the number of resource blocks reserved for grant-free transmission, but by the total number of resource blocks available in the system. Take an orthogonal frequency-division multiple access (OFDMA) system with 128 subcarriers as an example. If only 8 subcarriers are reserved for grant-free transmission, at most 8 grant-free users can be served, but the use of SGF transmission can potentially provide service to 120 additional grant-free users.

In SGF transmission, a crucial task is how to guarantee a grant-based user's quality of service (QoS) experience when admitting grant-free users to the same resource block. In [14], two SGF transmission schemes, termed SGF Scheme I and Scheme II, were developed to realize this goal. In particular, SGF Scheme I requires the base station to decode the grant-based user's signal first by treating the grant-free users' signals as interference, and schedules grant-free users with weak channel conditions in order to limit the interference they cause to the grant-based user. Therefore, Scheme I is ideal for situations, where the grant-free users are cell-edge users, i.e., their connections to the base station are weaker than that of the grant-based user. SGF Scheme II schedules grant-free users with strong channel conditions, and requires the base station to

decode the grant-free users' signals first. Therefore, Scheme II is ideal for situations, where the grant-free users' connections to the base station are strong. Two types of distributed contention control were applied in [14] to reduce the system overhead and to control the number of admitted grant-free users [15]–[17].

In this paper, we consider the same grant-free communication scenario as in [14], i.e., one grant-based user and M grant-free users communicate in one resource block with the same base station. A new SGF transmission scheme is proposed which can be interpreted as an opportunistic combination of the two existing SGF schemes and offers the following three advantages:

- Recall that SGF Scheme I decodes the grant-based user's signal first by treating the grant-free users' signals as interference. Hence, it is inevitable that the grant-based user's QoS experience is negatively affected by the admission of the grant-free users into the channel. The new SGF scheme can strictly guarantee that admitting grant-free users is transparent to the grant-based user, and the grant-based user's QoS experience is the same as when it occupies the channel along.
- Recall that SGF Scheme II directly decodes the grant-free users' signals by treating the grant-based user's signal as interference, which means that interference always exists for the grant-free users. Hence, for SGF Scheme II, the data rates available for the grant-free users can be small. The new SGF scheme can realize interference-free transmission for the grant-free users, and hence can offer significantly improved achievable data rates for grant-free transmission.
- For both existing SGF schemes, their outage probabilities exhibit error floors, when there is no transmit power control, e.g., the grant-free and grant-based users increase their transmit powers without coordination. Take Scheme I as an example, which decodes the grant-based user's signal first and then decodes the grant-free users' signals via successive interference cancellation (SIC). An outage probability error floor exists because increasing the grant-free users' transmit powers might help the second stage of SIC but increases the outage probability in the first stage. A similar error floor exists for Scheme II. The new SGF scheme can effectively avoid these error floors and significantly improve transmission robustness, even without careful power control among the users.

In order to facilitate the performance analysis, an exact expression for the outage probability achieved by the proposed SGF transmission scheme is developed based on order statistics.

Because the outage probability achieved by the proposed SGF scheme can be a function of four random variables, including three dependent order statistics, the developed exact expression has an involved form and hence cannot provide much insight into the properties of SGF transmission. Therefore, two high SNR approximations are developed based on an asymptotic analysis of the derived exact expression. The asymptotic expressions demonstrate that the proposed SGF transmission scheme avoids an outage probability error floor and realizes a multi-user diversity gain of M .

The remainder of the paper is organized as follows. In Section II, the existing SGF schemes are briefly introduced first, and then, the proposed new SGF scheme is described. In Section III, the outage performance achieved by the proposed SGF transmission scheme is analyzed, where an asymptotic analysis is also conducted to illustrate the multi-user diversity gain realized by the proposed scheme. Computer simulations are provided in Section IV, and Section V concludes the paper. We collect the details of all proofs in the appendix.

II. EXISTING AND NEWLY PROPOSED SGF SCHEMES

Consider an SGF communication scenario, where M grant-free users compete with each other for admission to a resource block which would otherwise be solely occupied by a grant-based user for conventional orthogonal multiple access (OMA). Denote the grant-based user's channel gain by g , and the grant-free users' channel gains by h_m , $1 \leq m \leq M$. We assume that the SGF system operates in quasi-static Rayleigh fading environments, i.e., all the channel gains are complex Gaussian distributed with zero mean and unit variance. Without loss of generality, we also assume that the grant-free users' channel gains are ordered as follows:

$$|h_1|^2 \leq \dots \leq |h_M|^2. \quad (1)$$

We note that this ordering assumption is to facilitate the performance analysis, and that this information is not available to any of the nodes in the system, including the base station. Prior to transmission, we assume that the grant-free users can overhear the information exchange between the grant-based user and the base station, and hence know the grant-based user's channel state information (CSI) as well as the grant-based user's transmit power, denoted by P_0 . In addition, each grant-free user has acquired the knowledge of its own CSI, by exploiting the pilot signals broadcasted by the base station.

A. Two Existing SGF Schemes

SGF Scheme I in [14] requires the base station to decode the grant-based user's signal during the first stage of SIC. If grant-free user m is admitted to the channel, the grant-based user's achievable data rate is $\log\left(1 + \frac{P_0|g|^2}{P_s|h_m|^2+1}\right)$, and grant-free user m 's data rate is $\log(1 + P_s|h_m|^2)$ if the first stage of SIC is successful, where P_s denotes the transmit power of the grant-free users, and the noise power is assumed to be normalized to one. In order to guarantee the grant-based user's QoS requirement, the base station broadcasts a predefined threshold, denoted by τ_I , and only the grant-free users whose channel gains fall below the threshold participate in contention. In this way, a user which has a strong channel and hence can cause strong interference to the grant-based user will not be granted access.

The use of distributed contention control ensures that contention can be carried out in a distributed manner. Thereby, each grant-free user's backoff period is proportional to its channel gain and therefore the user with the weakest channel gain will be granted access¹. Therefore, for SGF Scheme I, the admitted grant-free user's data rate is given by

$$R_I = \begin{cases} \log(1 + P_s|h_1|^2), & \text{if } |h_1|^2 \leq \tau_I \quad \& \quad \log\left(1 + \frac{P_0|g|^2}{P_s|h_1|^2+1}\right) > R_0 \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

where R_0 denotes the grant-based user's target data rate.

SGF Scheme II in [14] requires the base station to decode a grant-free user's signal during the first stage of SIC. Similar to Scheme I, the base station broadcasts a threshold, denoted by τ_{II} , and only the grant-free users whose channel gains are stronger than the threshold participate in contention. By using distributed contention control, each participating grant-free user sets its backoff time inversely proportional to its channel gain, which means that the grant-free user with the strongest channel condition is granted access, and its achievable data rate is given by

$$R_{II} = \begin{cases} \log\left(1 + \frac{P_s|h_M|^2}{P_0|g|^2+1}\right), & \text{if } |h_M|^2 \geq \tau_{II} \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

Remark 1: We note that, when $P_s \rightarrow \infty$ and $P_0 \rightarrow \infty$, there is an error floor for the admitted grant-free user's outage probability. Take SGF Scheme II as an example. $\log\left(1 + \frac{P_s|h_M|^2}{P_0|g|^2+1}\right)$ becomes a constant, when $P_s \rightarrow \infty$ and $P_0 \rightarrow \infty$, which means that there will be an error

¹In this paper, we focus on the case, where a single grant-free user is admitted to the channel. However, as discussed in [14] and [17], more than one user can be granted access via distributed contention, which is beyond the scope of this paper.

floor for the outage probability suffered by Scheme II. This error floor can be reduced, if P_s is much larger than P_0 . In other words, the existing SGF schemes require careful power control to guarantee the grant-free user's target outage performance, which might not be possible in practice.

B. Proposed SGF Scheme

In the proposed SGF scheme, prior to transmission, the base station broadcasts a threshold, denoted by $\tau(|g|^2)$, which needs to ensure the following inequality:

$$\log \left(1 + \frac{P_0 |g|^2}{\tau(|g|^2) + 1} \right) \geq R_0. \quad (4)$$

The proposed SGF scheme chooses $\tau(|g|^2)$ such that the above inequality constraint holds with equality:

$$\tau(|g|^2) = \max \left\{ 0, \frac{P_0 |g|^2}{2^{R_0} - 1} - 1 \right\}, \quad (5)$$

where $\max(a, b)$ denotes the maximum of a and b .

Upon receiving this threshold, each grant-free user compares its channel gain with the threshold individually. Unlike the two existing schemes, the proposed SGF scheme allows all the grant-free users to participate in contention. Each user's backoff time is determined by how its channel gain compares to $\tau(|g|^2)$, as shown in the following:

- Group 1 contains the users whose channel gains are above the threshold, i.e., $P_s |h_m|^2 > \tau(|g|^2)$. If a user in Group 1 is granted access, its signal has to be detected during the first stage of SIC. Otherwise, $P_s |h_m|^2 > \tau(|g|^2)$ leads to $\log \left(1 + \frac{P_0 |g|^2}{P_s |h_m|^2 + 1} \right) < R_0$, which would mean that the grant-based user's signal cannot be decoded correctly². Therefore, if a user in Group 1 is granted access, its achievable data rate is $\log \left(1 + \frac{P_s |h_m|^2}{P_0 |g|^2 + 1} \right)$, and hence its backoff time is set to be inversely proportional to this achievable rate.
- Group 2 contains the users whose channel gains are below the threshold, i.e., $P_s |h_m|^2 < \tau(|g|^2)$. For a user in Group 2, its signal can be decoded in either one of the two SIC stages, without affecting the grant-based user's QoS. In particular, if its signal is decoded in the first stage of SIC, its achievable data rate is $\log \left(1 + \frac{P_s |h_m|^2}{P_0 |g|^2 + 1} \right)$. If its signal is decoded

²In this paper, the grant-free users are assumed to use the same fixed transmit power, P_s . The use of distributed power control can further improve the performance of SGF transmission without increasing system overhead, but is beyond the scope of this paper.

in the second stage of SIC, its achievable data rate is $\log(1 + P_s|h_m|^2)$. We note that if a user from Group 2 is granted access, it is guaranteed that the grant-based user's signal can be successfully decoded in the first stage of SIC, since $P_s|h_m|^2 < \tau(|g|^2)$ leads to $\log\left(1 + \frac{P_0|g|^2}{P_s|h_m|^2+1}\right) > R_0$. In other words, $\log(1 + P_s|h_m|^2)$ is always achievable for a user from Group 2. Therefore, its backoff time is set to be inversely proportional to $\log(1 + P_s|h_m|^2)$, since $\log(1 + P_s|h_m|^2) \geq \log\left(1 + \frac{P_s|h_m|^2}{P_0|g|^2+1}\right)$.

By carrying out distributed contention control [15]–[17], either a user from Group 1 with the largest $\log\left(1 + \frac{P_s|h_m|^2}{P_0|g|^2+1}\right)$ or a user from Group 2 with the largest $\log(1 + P_s|h_m|^2)$ is granted access in a distributed manner.

Remark 2: The proposed SGF scheme can be viewed as a hybrid version of the two existing schemes. In particular, under the condition that admitting a grant-free user needs to be transparent to the grant-based user, the users in Group 1 can support SGF Scheme II, whereas the users in Group 2 can support either of the two schemes. The proposed scheme will select the grant-free user with the largest achievable data rate in an opportunistic manner .

Remark 3: We note that, among the M grant-free users, only two users have the chance of being granted access, if the grant-free users' channel gains are ordered as in (1). One is grant-free user M , if Group 1 is not empty, since $\log\left(1 + \frac{P_s|h_m|^2}{P_0|g|^2+1}\right) \leq \log\left(1 + \frac{P_s|h_j|^2}{P_0|g|^2+1}\right)$ always holds for any $m \leq j$. The other one is the grant-free user which has the strongest channel gain in Group 2, if Group 2 is not empty, since $\log(1 + P_s|h_m|^2) \leq \log(1 + P_s|h_j|^2)$ for $m \leq j$. If a grant-based protocol is used, i.e., global CSI is available at the base station, the base station can decide which user is to be admitted by simply comparing the two users' data rates. The use of the proposed distributed contention control can ensure that the same goal is achieved without acquiring global CSI at the base station.

III. OUTAGE PERFORMANCE ANALYSIS

It is straightforward to show that the use of the proposed SGF scheme can strictly guarantee that admitting grant-free users is completely transparent to the grant-based user, and the grant-based user's experience is the same as with OMA. Therefore, in this paper, we mainly focus on the outage performance of the admitted grant-free user, where we assume that all the grant-free users have the same target data rate, denoted by R_s .

To characterize the outage event, denote the event that there are m users in Group 2 by E_m , where E_m can be explicitly defined as follows:

$$E_m = \left\{ |h_m|^2 < \frac{\tau(|g|^2)}{P_s}, |h_{m+1}|^2 > \frac{\tau(|g|^2)}{P_s} \right\}, \quad (6)$$

for $1 \leq m \leq M-1$. Furthermore, the two extreme cases with no user in Group 1 and Group 2 can be defined as $E_M = \left\{ |h_M|^2 < \frac{\tau(|g|^2)}{P_s} \right\}$ and $E_0 = \left\{ |h_1|^2 > \frac{\tau(|g|^2)}{P_s} \right\}$, respectively.

The overall outage probability experienced by the admitted grant-free users is given by

$$\begin{aligned} P_{out} &= \sum_{m=1}^{M-1} P(E_m, \max\{R_{k,I}, 1 \leq k \leq m\} < R_s, \max\{R_{k,II}, m < k \leq M\} < R_s) \\ &\quad + P(E_M, \max\{R_{k,I}, 1 \leq k \leq M\} < R_s) + P(E_0, \max\{R_{k,II}, 1 \leq k \leq M\} < R_s), \end{aligned} \quad (7)$$

where $R_{k,I} = \log(1 + P_s|h_k|^2)$ and $R_{k,II} = \log\left(1 + \frac{P_s|h_k|^2}{P_0|g|^2+1}\right)$.

Because the grant-free users' channel gains are ordered as in (1), the outage probability can be simplified as follows:

$$P_{out} = \sum_{m=1}^M P(E_m, \max\{R_{m,I}, R_{M,II}\} < R_s) + P(E_0, R_{M,II} < R_s). \quad (8)$$

Define $\epsilon_0 = 2^{R_0} - 1$, $\epsilon_s = 2^{R_s} - 1$, $\alpha_0 = \frac{\epsilon_0}{P_0}$, and $\alpha_s = \frac{\epsilon_s}{P_s}$. We note that if $|g|^2 < \alpha_0$,

$$\tau(|g|^2) = \max\left\{0, \frac{P_0|g|^2}{2^{R_0}-1} - 1\right\} = 0. \quad (9)$$

By using (9), the outage probability can be rewritten as follows:

$$\begin{aligned} P_{out} &= \sum_{m=1}^M \underbrace{P(E_m, |g|^2 > \alpha_0, \max\{R_{m,I}, R_{M,II}\} < R_s)}_{Q_m} \\ &\quad + \underbrace{P(E_0, |g|^2 > \alpha_0, R_{M,II} < R_s)}_{Q_0} + \underbrace{P(R_{M,II} < R_s, |g|^2 < \alpha_0)}_{Q_{M+1}}. \end{aligned} \quad (10)$$

The following theorem provides an exact expression for the outage probability achieved by the proposed SGF scheme.

Theorem 1. Assume that $\epsilon_s \epsilon_0 < 1$ and $M \geq 2$. The outage probability achieved by the proposed SGF transmission scheme can be expressed as follows:

$$\begin{aligned}
P_{out} = & \sum_{m=1}^{M-2} \bar{\eta}_m \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \tilde{\mu}_4 \phi(p, \tilde{\mu}_2) \\
& + \sum_{i=0}^{M-1} \binom{M-1}{i} \frac{(-1)^i \tilde{\eta}_0}{M-1} \left(e^{\frac{1}{P_s}} \phi(i, \mu_7) - e^{-\alpha_s} \phi(i, \mu_8) \right) \\
& + \frac{\tilde{\eta}_0}{M(M-1)} \sum_{l=0}^M (-1)^l \binom{M}{l} e^{-l\alpha_s} e^{\frac{M-l}{P_s}} g_{\tilde{\mu}_{12}}(\alpha_0, \alpha_2) + \sum_{i=0}^M \binom{M}{i} (-1)^i e^{\frac{i}{P_s}} g_{\frac{i}{\alpha_0 P_s}}(\alpha_0, \alpha_1) \\
& + (1 - e^{-\alpha_s})^M e^{-\alpha_1} + \sum_{i=0}^M \binom{M}{i} (-1)^i e^{-i\alpha_s} \frac{1 - e^{-(1+i\alpha_s P_0)\alpha_0}}{1 + i\alpha_s P_0}, \tag{11}
\end{aligned}$$

where $\bar{\eta}_m = \frac{M!}{m!(M-m)!}$, $\tilde{\eta}_0 = \frac{M!}{(M-2)!}$, $\tilde{\mu}_2 = l\alpha_s P_0 + (M-m-l)\frac{\epsilon_0^{-1}P_0}{P_s}$, $\tilde{\mu}_4 = e^{-l\alpha_s + (M-m-l)\frac{1}{P_s}}$, $\mu_7 = \frac{1}{P_s \alpha_0}$, $\mu_8 = \alpha_s P_0$, $\tilde{\mu}_{12} = l\alpha_s P_0 + (M-l)\frac{\alpha_0^{-1}}{P_s}$, $\alpha_1 = (1 + \epsilon_s)\alpha_0$, $\alpha_2 = \frac{\epsilon_0(\epsilon_s + 1)}{(1 - \epsilon_0 \epsilon_s)P_0}$, $g_\mu(x_1, x_2) = \frac{e^{-(1+\mu)x_1} - e^{-(1+\mu)x_2}}{1+\mu}$, and $\phi(p, \mu) = e^{-p\alpha_s} g_\mu(\alpha_1, \alpha_2) + e^{\frac{p}{P_s}} g_{\mu + \frac{p}{P_s \alpha_0}}(\alpha_0, \alpha_1)$.

Proof. See Appendix A. □

Following the steps in the proof for Theorem 1, the outage probability for the case $M = 1$ can be obtained straightforwardly as shown in the following corollary.

Corollary 1. Assume that $\epsilon_s \epsilon_0 < 1$ and $M = 1$. The outage probability achieved by the proposed SGF transmission scheme can be expressed as follows:

$$P_{out} = e^{\frac{1}{P_s}} g_{\frac{\alpha_0^{-1}}{P_s}}(\alpha_0, \alpha_2) - e^{-\alpha_s} g_{\alpha_s P_0}(\alpha_0, \alpha_2) + \sum_{i=0}^1 \binom{1}{i} (-1)^i e^{\frac{i}{P_s}} g_{\frac{i}{\alpha_0 P_s}}(\alpha_0, \alpha_1) + (1 - e^{-\alpha_s}) e^{-\alpha_1}. \tag{12}$$

Remark 4: In this paper, we mainly focus on the case $\epsilon_s \epsilon_0 < 1$ because the error floor of P_{out} can be avoided in this case, i.e., the scenario with $\epsilon_s \epsilon_0 < 1$ is ideal for the application of the proposed SGF scheme. $\epsilon_s \epsilon_0 < 1$ means that R_s needs to be small for a given R_0 , which is a realistic assumption in practice since SGF is invoked to encourage spectrum sharing between a grant-based user and a grant-free user with a small target data rate.

Remark 5: We note that for the case $\epsilon_s \epsilon_0 \geq 1$, the proposed SGF scheme still works and offers significant performance gains compared to the two existing SGF schemes, as shown in the simulation section. However, for $\epsilon_s \epsilon_0 \geq 1$, the outage probability achieved by the proposed

SGF scheme exhibits an error floor, similar to the existing schemes. More detailed discussions will be provided in Section IV.

Remark 6: The outage probability expression shown in Theorem 1 is complicated, mainly due to the fact that Q_m depends on the choice of m . For example, for $1 \leq m \leq M - 2$, Q_m is a function of the four channel gains, h_m, h_{m-1}, h_M , and g , whereas Q_{M-1} is a function of only h_{M-1}, h_M , and g . The fact that the h_m, h_{m-1} , and h_M are dependent order statistics makes the expression even more involved. However, at high SNR, insightful approximations can be obtained as shown in the following theorem.

Theorem 2. *Assuming that $\epsilon_s \epsilon_0 < 1$, $M \geq 2$ and $P_s = P_0 \rightarrow \infty$, the outage probability P_{out} can be approximated at high SNR as follows:*

$$\begin{aligned}
P_{out} \approx & \sum_{m=1}^{M-2} \frac{\tilde{\eta}_m}{P_s^{M+1}} \epsilon_s^m \sum_{i=0}^{M-m} \binom{M-m}{i} (\epsilon_s + 1)^{M-m-i} (\epsilon_s - \epsilon_0^{-1})^i \epsilon_0^{i+1} \frac{\tilde{\alpha}_2^{i+1} - (1 + \epsilon_s)^{i+1}}{i+1} \quad (13) \\
& + \sum_{m=1}^{M-2} \frac{\tilde{\eta}_0 (1 + \epsilon_s) \epsilon_s^{M-1} \epsilon_0}{P_s^{M+1} (M-1)} \left((\tilde{\alpha}_2 - 1 - \epsilon_s) + \frac{\epsilon_s}{M} \right) \\
& + \frac{\tilde{\eta}_m}{P_s^{M+1}} \sum_{i=0}^{M-m} \binom{M-m}{i} (\epsilon_s + \epsilon_0 \epsilon_s)^{M-m-i} (\epsilon_s - \epsilon_0^{-1})^i \epsilon_0^{i+1} \frac{\epsilon_s^{m+i+1}}{m+i+1} \\
& + \frac{\tilde{\eta}_0}{P_s^{M+1} M (M-1)} e^{\frac{M}{P_s}} \sum_{i=0}^M \binom{M}{i} (\epsilon_s + 1)^{M-i} (\epsilon_s - \epsilon_0^{-1})^i \epsilon_0^{i+1} \frac{\tilde{\alpha}_2^{i+1} - 1}{i+1} \\
& + \frac{1}{P_s^{M+1} (M+1) \epsilon_0^M} \epsilon_0^{M+1} \epsilon_s^{M+1} + \frac{\epsilon_s^M}{P_s^M} + \frac{\epsilon_s^M (1 + \epsilon_0)^{M+1} - 1}{P_s^{M+1} (M+1)},
\end{aligned}$$

where $\tilde{\alpha}_2 = \frac{(\epsilon_s + 1)}{(1 - \epsilon_0 \epsilon_s)}$.

Proof. See Appendix B. □

Remark 7: Following the same steps as in the proof of Theorem 2, the outage probability for the case $M = 1$ can be approximated as follows:

$$P_{out} \approx \frac{1}{2P_s^2} \epsilon_0 \epsilon_s^2 + \frac{\epsilon_s}{P_s} + \frac{1}{P_s^2} (1 + \epsilon_s) \epsilon_0 (\tilde{\alpha}_2 - 1). \quad (14)$$

By comparing the terms in Theorem 2, one can find that there is one term proportional to $\frac{1}{P_s^M}$, and the other ones are proportional to $\frac{1}{P_s^{M+1}}$. Therefore, a further approximation can be straightforwardly obtained as shown in the following corollary.

Corollary 2. Assuming that $\epsilon_s \epsilon_0 < 1$ and $P_s = P_0 \rightarrow \infty$, the outage probability P_{out} can be further approximated as follows:

$$P_{out} \approx \frac{\epsilon_s^M}{P_s^M}. \quad (15)$$

A diversity gain of M is achievable for the proposed SGF transmission scheme.

Remark 8: Recall that for the two existing SGF schemes, their outage probabilities suffer from error floors, when P_s and P_0 go to infinity. Corollary 2 demonstrates that not only can the proposed SGF transmission scheme avoid these error floors, but also can it ensure that the achievable diversity gain is proportional to the number of participating grant-free users, i.e., the more grant-free users there are, the better the outage performance.

Remark 9: The main reason why the proposed SGF scheme avoids an error floor can be explained as follows. By using (8), an upper bound on the outage probability achieved by the proposed SGF scheme can be obtained as follows:

$$\begin{aligned} P_{out} &= \sum_{m=1}^M P(E_m, \max\{R_{m,I}, R_{M,II}\} < R_s) + P(E_0, R_{M,II} < R_s) \\ &\leq \sum_{m=1}^M P(R_{m,I} < R_s) + P(E_0, R_{M,II} < R_s) \\ &= \sum_{m=1}^M P(\log(1 + P_s|h_m|^2) < R_s) + P\left(E_0, \log\left(1 + \frac{P_s|h_M|^2}{P_0|g|^2 + 1}\right) < R_s\right), \end{aligned} \quad (16)$$

where the last step follows from the definitions of $R_{m,I}$ and $R_{M,II}$.

By using the fact that the users are ordered as in (1), P_{out} can be further upper bounded as follows:

$$P_{out} \leq MP(\log(1 + P_s|h_1|^2) < R_s) + \underbrace{P\left(E_0, \log\left(1 + \frac{P_s|h_M|^2}{P_0|g|^2 + 1}\right) < R_s\right)}_{Q_u}. \quad (17)$$

Recall that the outage probability for SGF Scheme II is $P_{out}^{II} \triangleq P\left(\log\left(1 + \frac{P_s|h_M|^2}{P_0|g|^2 + 1}\right) < R_s\right)$, where an error floor exists since its signal-to-interference-plus-noise ratio (SINR) becomes a constant when P_s and P_0 go to infinity. The probability Q_u is quite similar to P_{out}^{II} , but Q_u does not exhibit an error floor, as explained in the following. By using the definition of E_0 , Q_u can be rewritten as follows:

$$Q_u = P\left(|h_1|^2 > \frac{\tau(|g|^2)}{P_s}, |h_M|^2 < \alpha_s(P_0|g|^2 + 1)\right). \quad (18)$$

Since $|h_1|^2 \leq |h_M|^2$, the lower bound on $|h_1|^2$, $\frac{\tau(|g|^2)}{P_s}$, needs to be smaller than the upper bound on $|h_M|^2$, $\alpha_s(P_0|g|^2 + 1)$, which introduces an additional constraint $|g|^2 < \alpha_2$, if $\epsilon_s \epsilon_0 < 1$, as shown in (23) - (25). This additional constraint $|g|^2 < \alpha_2$ effectively removes the error floor since

$$Q_u = \mathbb{P} \left(|h_1|^2 > \frac{\tau(|g|^2)}{P_s}, \log \left(1 + \frac{P_s |h_M|^2}{P_0 |g|^2 + 1} \right) < R_s, |g|^2 < \alpha_2 \right) \quad (19)$$

$$\leq \mathbb{P} (|g|^2 < \alpha_2) = 1 - e^{-\alpha_2} \rightarrow 0,$$

for $P_0 \rightarrow \infty$. On the other hand, it is straightforward to show that the first term in (17), $\mathbb{P}(\log(1 + P_s|h_1|^2) < R_s)$, also goes to zero at high SNR. Therefore, P_{out} does not have an error floor.

IV. SIMULATION RESULTS

In this section, the performance of the proposed SGF transmission scheme is studied via computer simulations, where the accuracy of the developed analytical results is also evaluated. To facilitate performance evaluation, the two existing SGF schemes proposed in [14] are used as benchmark schemes. We note that the proposed SGF scheme allows all the users to participate in contention. Therefore, for a fair comparison, we choose $\tau_I = \infty$ and $\tau_{II} = 0$ for the two benchmarking schemes, which allow all grant-free users to participate in contention and hence yield the best performance for the two schemes.

In Fig. 1, the outage performance achieved by the proposed SGF transmission scheme is compared to those of the two existing schemes for different choices of P_s and P_0 . In particular, in Fig. 1(a), we assume $P_s = \frac{P_0}{10}$, which is equivalent to the case where the grant-free users have weaker channel conditions than the grant-based user, if all the users use the same transmit power. Recall that SGF Scheme I first decodes the grant-based user's signal by treating the grant-free user's signal as interference. Therefore, the situation with $P_s = \frac{P_0}{10}$ is ideal for the application of SGF Scheme I, and Fig. 1(a) confirms this conclusion since SGF Scheme I outperforms SGF Scheme II. We note that for SGF Scheme I, the outage probability for $M = 1$ can be better than that for $M = 5$, since a larger M can reduce $\mathbb{P} \left(\log \left(1 + \frac{P_0 |g|^2}{P_s |h_1|^2 + 1} \right) < R_0 \right)$ but may increase $\mathbb{P}(\log(1 + P_s|h_1|^2) < R_s)$. In Fig. 1(b), we focus on the situation, where $P_s \rightarrow \infty$ and P_0 is kept constant. This is equivalent to the case where the grant-free users have stronger channel conditions than the grant-based user, if all the users use the same transmit power. Therefore, this situation is ideal for the application of SGF Scheme II, and Fig. 1(b) shows that SGF

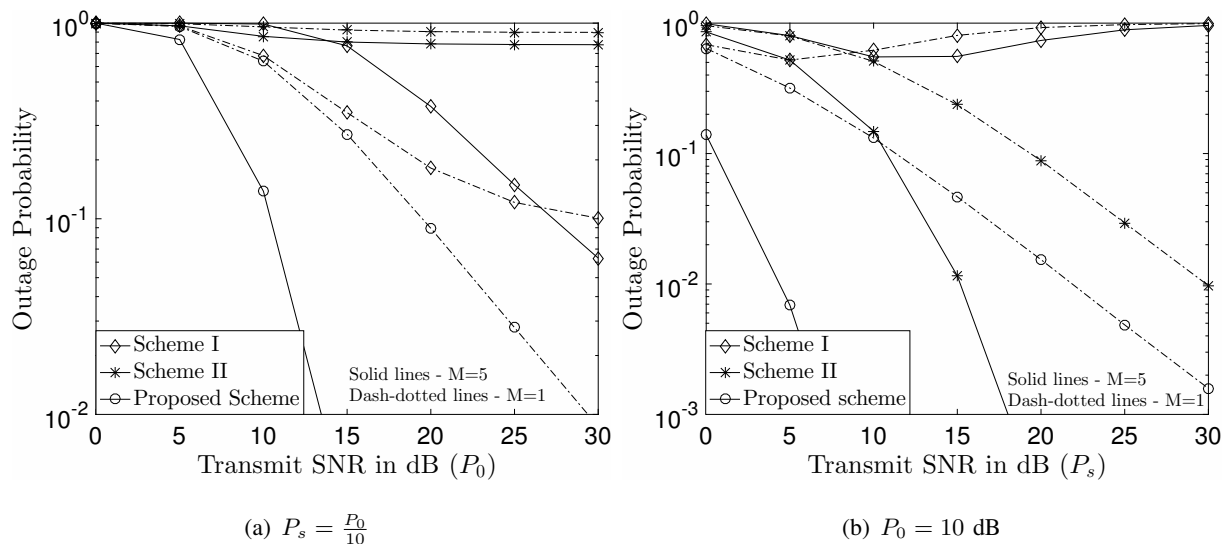


Fig. 1. Comparison of the three SGF transmission schemes. $R_0 = 1$ bit per channel use (BPCU), and $R_s = 0.9$ BPCU.

Scheme II indeed outperforms SGF Scheme I. For both considered scenarios, the proposed SGF scheme outperforms the two existing schemes, and can also effectively avoid error floors in both considered scenarios, as shown in the two figures.

In Fig. 2, we examine the accuracy of the developed analytical results for the outage probability. In Fig. 2(a), the exact expressions for the outage probabilities shown in Theorem 1 and Corollary 1 are used, and the figure shows that the curves for the analytical results perfectly match the curves obtained by simulations, which demonstrates the accuracy of the result provided in Theorem 1. In Fig. 2(b), the accuracy of the approximations developed in Theorem 2 and Corollary 2 is studied. As can be observed from the figure, both approximations are accurate at high SNR. We note that the approximation in Corollary 2 becomes less accurate as M increases. This is due to the fact that the approximation in Corollary 2 disregards the terms, Q_m , $0 \leq m \leq M - 1$, and Q_{M+1} , and considers Q_M only. When M is small, such an approximation is accurate. But the gap between the approximation and the actual probability becomes noticeable when M becomes large. The fact that the curves for the approximation in Corollary 2 are below the other curves is also due to the same reason.

In Fig. 3, the impact of different choices for the target data rate and the transmit power on the outage performance is studied. The figure shows that reducing the grant-free user's target rate can affect the outage probability more significantly than reducing the grant-based user's target rate. In addition, the figure shows that, for a fixed P_0 , increasing P_s can improve the grant-free user's outage performance, i.e., a grant-free user can improve its performance by increasing its own transmit power. This is not true for SGF Scheme I since increasing P_s deteriorates the

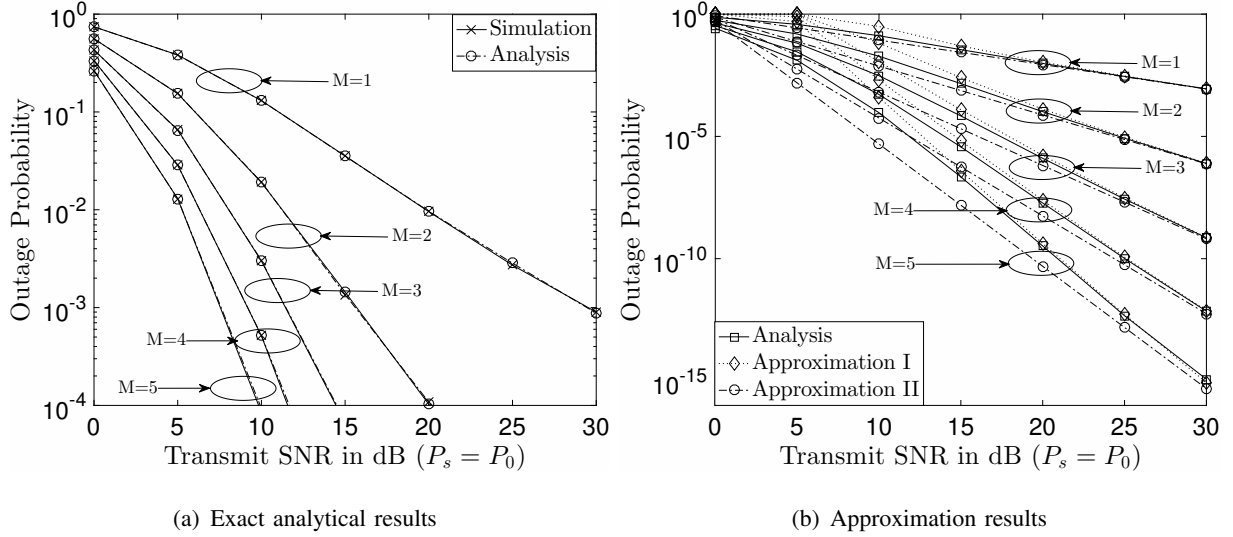


Fig. 2. Accuracy of the developed analytical results. $P_0 = P_s$, $R_0 = 1$ bit per channel use (BPCU), and $R_s = 0.9$ BPCU. The curves for Analysis are based on Theorem 1 and Corollary 1, the curves for Approximation I are based on Theorem 2, and the curves for Approximation II are based on Corollary 2.

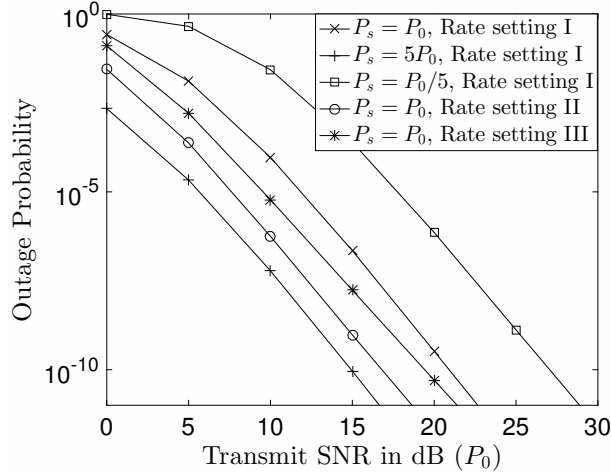


Fig. 3. Performance of the proposed SGF transmission scheme for different choices for the transmit power and the target rates. For Rate Setting I, $R_0 = 1$ BPCU and $R_s = 0.9$ BPCU. For Rate Setting II, $R_0 = 1$ BPCU and $R_s = 0.5$ BPCU. For Rate Setting III, $R_0 = 0.5$ BPCU and $R_s = 0.9$ BPCU. $M = 5$.

probability $P\left(\log\left(1 + \frac{P_0|g|^2}{P_s|h_1|^2+1}\right) < R_0\right)$.

In Fig. 4, the performance of the proposed SGF transmission scheme is evaluated under the condition that $\epsilon_s \epsilon_0 \geq 1$. As discussed in Remarks 3 and 7, the condition $\epsilon_s \epsilon_0 < 1$ is important to avoid error floors. If this condition does not hold, error floors appear, as shown in Fig. 4. However, we note that the outage performance achieved by the proposed SGF transmission scheme is still significantly better than those of the two existing schemes. For example, for the case with $R_0 = 1.5$ bits per channel use (BPCU), and $R_s = 1$ BPCU, the proposed scheme can achieve an outage probability of 1×10^{-4} , whereas the outage probabilities achieved by the two

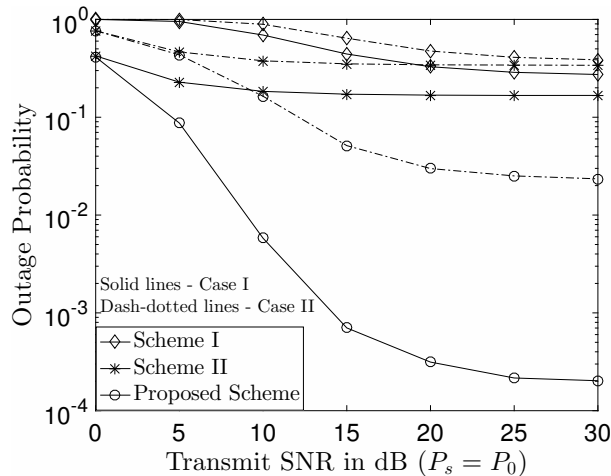


Fig. 4. Performance of the proposed SGF transmission scheme when $\epsilon_s \epsilon_0 \geq 1$. For Case I, $R_0 = 1.5$ BPCU, $R_s = 1$ BPCU. For Case II, $R_0 = 2$ BPCU, $R_s = 1.5$ BPCU. $M = 5$.

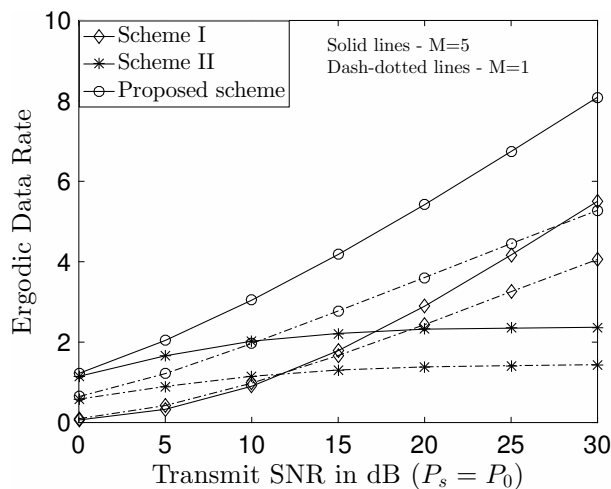


Fig. 5. Ergodic data rate comparison of the three considered SGF transmission schemes. $R_0 = 1$ BPCU.

existing schemes exceed 1×10^{-1} .

In Fig. 5, the ergodic data rate is used to evaluate the performance of the considered SGF schemes. As can be observed from the figure, the proposed SGF scheme outperforms both existing schemes, particularly at high SNR, which is consistent with the figures showing the outage probability. In addition, Fig. 5 shows that the slope of the curves for the proposed SGF scheme is larger than those of the two existing schemes, which demonstrates that the proposed scheme can effectively exploit multi-user diversity. An interesting observation is that, for high transmit powers, the curves for SGF Scheme II become flat, whereas the curves for the other two schemes do not. This is due to the fact that the data rate achieved by Scheme

It is $\log\left(1 + \frac{P_s|h_M|^2}{P_0|g|^2+1}\right)$, which becomes a constant when both P_s and P_0 approach infinity. On the other hand, once the grant-based user's target data rate can be realized, the achievable data rates for SGF Scheme I and the proposed SGF scheme are of the form $\log(1 + P_s|h_m|^2)$, which means that their ergodic data rates are not bounded when P_s goes to infinity, as confirmed by the figure. The performance gain of the proposed scheme over Scheme I is due to the fact that, when the grant-based user's signal cannot be decoded correctly in the first stage of SIC, the data rate of Scheme I becomes zero, but the proposed scheme can still offer a non-zero data rate by changing the SIC order.

Fig. 6 shows the grant-free users' admission probabilities, i.e., which grant-free user is admitted to the resource block by the proposed SGF scheme, for different choices of R_0 and the users' transmit powers. We first note that the admission probability for grant-free user m is given by

$$P_m = P(E_m, R_{m,I} > R_{M,II}), \quad (20)$$

for $1 \leq m \leq M - 1$, and

$$P_M = \sum_{n=1}^{M-1} P(E_m, R_{m,I} < R_{M,II}) + P(E_M) + P(E_0).$$

Fig. 6 shows that at low SNR, grant-free user M , the user with the strongest channel gain, is preferred over the other users, as explained as follows. At low SNR, the threshold $\tau(|g|^2) \triangleq \max\left\{0, \frac{P_0|g|^2}{2^{R_0}-1} - 1\right\}$ is very likely to be zero, which means that Group 2 is empty, i.e., E_0 happens. As a result, grant-free user M is granted access. In addition, Fig. 6 shows that at high SNR, the users' admission probabilities become constant, and increasing R_0 increases the admission probabilities of the grant-free users whose channel gains are weak, which can be explained as follows. By assuming $P_s = P_0 \rightarrow \infty$, P_m , $1 \leq m \leq M - 1$, can be approximated as follows:³

$$\begin{aligned} P_m &= P(E_m, R_{m,I} > R_{M,II}) \rightarrow P(E_m) \\ &= P\left(|h_m|^2 < \frac{\tau(|g|^2)}{P_s}, |h_{m+1}|^2 > \frac{\tau(|g|^2)}{P_s}\right) \rightarrow P\left(|h_m|^2 < \frac{|g|^2}{2^{R_0}-1}, |h_{m+1}|^2 > \frac{|g|^2}{2^{R_0}-1}\right), \end{aligned} \quad (21)$$

which is indeed a constant and not a function of P_s or P_0 . Since P_m , $1 \leq m \leq M - 1$, are constant at high SNR and $\sum_{m=1}^M P_m = 1$, P_M must also be constant at high SNR, as confirmed

³Obtaining an exact expression for P_m is not a trivial task since P_m is a function of four random variables, $|g|^2$, $|h_m|^2$, $|h_{m+1}|^2$ and $|h_M|^2$. The fact that $|h_m|^2$, $|h_{m+1}|^2$ and $|h_M|^2$ are not independent makes it more difficult to analyze P_m , which is an important direction for future research.

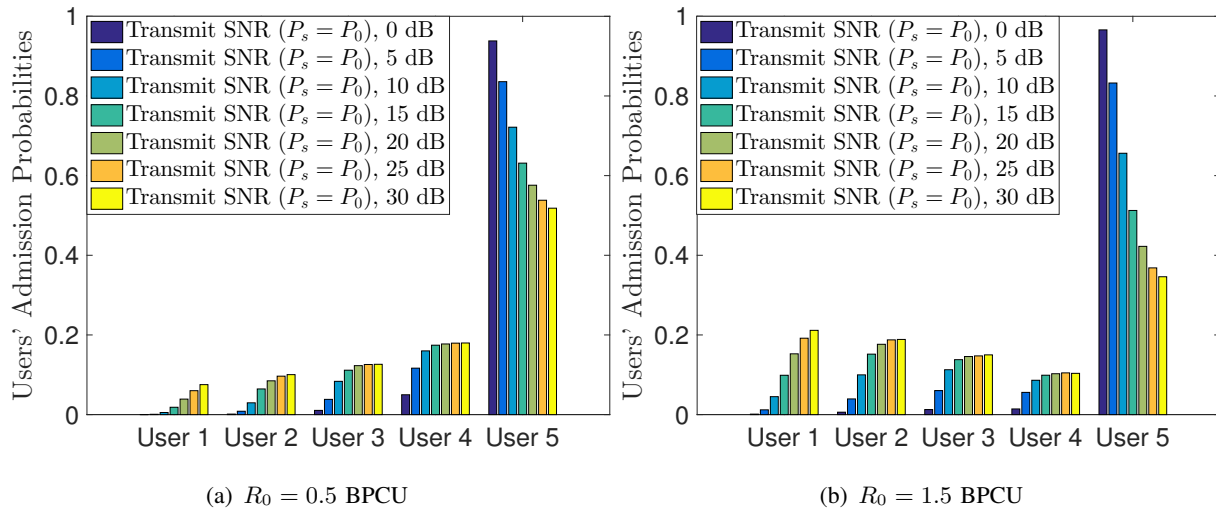


Fig. 6. Users' admission probabilities for the proposed SGF scheme. $M = 5$.

by the figure. By increasing R_0 , $\frac{|g|^2}{2^{R_0}-1}$ is reduced, and (21) indicates that P_m is increased for small m , i.e., the weak users' admission probabilities are increased, as shown in Fig. 6.

V. CONCLUSIONS

In this paper, we have proposed a new NOMA assisted SGF transmission scheme. Compared to the two existing SGF schemes, this new scheme can ensure that admitting a grant-free user is completely transparent to the grant-based user, i.e., the grant-based user communicates with its base station as if it solely occupied the channel. In addition, the proposed SGF scheme significantly improves the reliability of the grant-free users' transmissions compared to the existing SGF schemes. To facilitate the performance evaluation of the proposed SGF scheme, an exact expression for the outage probability was derived, where an asymptotic analysis was also carried out to show that the full multi-user diversity gain of M is achievable. Computer simulation results were provided to demonstrate the performance of the proposed SGF scheme and to verify the accuracy of the developed analytical results.

In this paper, Rayleigh fading is assumed for the users' channel gains. An important direction for future research is to carry out a stochastic geometry analysis by taking the users' path losses into consideration. In addition, we assumed that the admitted grant-free user uses only one time slot for transmission. In practice, the grant-free user may perform retransmission and occupy the channel for a few consecutive time slots. An interesting direction for future research is to develop hybrid automatic repeat request (HARQ) schemes for SGF transmission.

APPENDIX A
PROOF FOR THEOREM 1

The evaluation of probability Q_m in (10) depends on the value of m , as shown in the following subsections.

A. *Evaluation of Q_m , $1 \leq m \leq M - 2$*

In this case, probability Q_m involves three order statistics, h_m , h_{m+1} , and h_M , and can be expressed as follows:

$$\begin{aligned} Q_m &= \mathbb{P} \left(E_m, |g|^2 > \frac{\epsilon_0}{P_0}, \log(1 + P_s |h_m|^2) < R_s, \log \left(1 + \frac{P_s |h_M|^2}{P_0 |g|^2 + 1} \right) < R_s \right) \\ &= \mathcal{E}_{|g|^2 > \frac{\epsilon_0}{P_0}} \left\{ \mathbb{P} \left(|h_m|^2 < \xi, |h_{m+1}|^2 > \frac{P_0 \epsilon_0^{-1} |g|^2 - 1}{P_s}, |h_M|^2 < \frac{\epsilon_s (1 + P_0 |g|^2)}{P_s} \right) \right\}, \end{aligned} \quad (22)$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operation, and $\xi = \min \left\{ \frac{\epsilon_s}{P_s}, \frac{P_0 \epsilon_0^{-1} |g|^2 - 1}{P_s} \right\}$.

For the case $1 \leq m \leq M - 2$, h_{m+1} and h_M are different. As a result, there is a hidden constraint in (22) that the lower bound on h_{m+1} should be smaller than the upper bound on h_M , i.e., $\frac{\epsilon_s (1 + P_0 |g|^2)}{P_s} > \frac{P_0 \epsilon_0^{-1} |g|^2 - 1}{P_s}$. We first note that whether $\frac{\epsilon_s (1 + P_0 |g|^2)}{P_s} > \frac{P_0 \epsilon_0^{-1} |g|^2 - 1}{P_s}$ holds depends on the value of g as shown in the following:

$$\begin{aligned} &\epsilon_s (1 + P_0 |g|^2) - (P_0 \epsilon_0^{-1} |g|^2 - 1) \\ &= (\epsilon_s - \epsilon_0^{-1}) P_0 |g|^2 + \epsilon_s + 1 \begin{cases} < 0, & \text{if } |g|^2 > \frac{\epsilon_0 (\epsilon_s + 1)}{(1 - \epsilon_0 \epsilon_s) P_0} \\ > 0, & \text{otherwise} \end{cases}, \end{aligned} \quad (23)$$

where the assumption that $\epsilon_s \epsilon_0 < 1$ was used. Furthermore, we note that the following inequality always holds

$$\frac{\epsilon_0 (\epsilon_s + 1)}{(1 - \epsilon_0 \epsilon_s) P_0} > \frac{\epsilon_0}{P_0}, \quad (24)$$

since $(\epsilon_s + 1) - (1 - \epsilon_s \epsilon_0) = \epsilon_s + \epsilon_s \epsilon_0 \geq 0$.

Therefore, denoting the probability inside the expectation in (22) by S_m , Q_m can be expressed as follows:

$$\begin{aligned} Q_m &= \mathcal{E}_{\alpha_2 > |g|^2 > \alpha_0} \{S_m\} + \mathcal{E}_{|g|^2 > \alpha_1} \{S_m\} \\ &= \mathcal{E}_{\alpha_2 > |g|^2 > \alpha_0} \{S_m\}, \end{aligned} \quad (25)$$

where the last step follows by using (23).

For the case $1 \leq m \leq M - 2$, S_m is a function of three order statistics, $|h_m|^2$, $|h_{m+1}|^2$, and $|h_M|^2$, whose joint probability density function (pdf) is given by [18]

$$\begin{aligned} & f_{|h_m|^2, |h_{m+1}|^2, |h_M|^2}(x, y, z) \\ &= \eta_m e^{-x} (1 - e^{-x})^{m-1} e^{-y} (e^{-y} - e^{-z})^{M-m-2} e^{-z} \\ &= \eta_m \sum_{i=0}^{M-m-2} \binom{M-m-2}{i} (-1)^i e^{-x} (1 - e^{-x})^{m-1} e^{-y} e^{-(M-m-2-i)y} e^{-iz} e^{-z}, \end{aligned} \quad (26)$$

where $x \leq y \leq z$ and $\eta_m = \frac{M!}{(m-1)!(M-m-2)!}$.

For a fixed g and by using the joint pdf shown in (26), S_m can be expressed as follows:

$$\begin{aligned} S_m &= \eta_m \sum_{i=0}^{M-m-2} \binom{M-m-2}{i} (-1)^i \int_0^\xi e^{-x} (1 - e^{-x})^{m-1} \\ &\quad \times \int_{\frac{\alpha_0^{-1}|g|^2-1}{P_s}}^{\alpha_s(1+P_0|g|^2)} e^{-(M-m-1-i)y} \int_y^{\alpha_s(1+P_0|g|^2)} e^{-(i+1)z} dz dy dx. \end{aligned}$$

With some algebraic manipulations, S_m can be calculated as follows:

$$\begin{aligned} S_m &= \eta_m \sum_{i=0}^{M-m-2} \binom{M-m-2}{i} \sum_{p=0}^m \binom{m}{p} \frac{(-1)^{i+p} e^{-p\xi}}{m(i+1)} \\ &\quad \times \left(\frac{\mu_3 e^{-\mu_1|g|^2} - \mu_5 e^{-\mu_6|g|^2}}{M-m} - \frac{\mu_4 e^{-\mu_2|g|^2} - \mu_5 e^{-\mu_6|g|^2}}{M-m-1-i} \right), \end{aligned}$$

where $\mu_1 = \frac{(M-m)\alpha_0^{-1}}{P_s}$, $\mu_2 = \left((i+1)\alpha_s + (M-m-1-i)\frac{\epsilon_0^{-1}}{P_s} \right) P_0$, $\mu_3 = e^{\frac{M-m}{P_s}}$, $\mu_4 = e^{-(i+1)\alpha_s + (M-m-1-i)\frac{1}{P_s}}$, $\mu_5 = e^{-(M-m)\alpha_s}$, and $\mu_6 = (M-m)\alpha_s P_0$.

Recall that Q_m can be obtained by finding the expectation of S_m for $\alpha_2 > |g|^2 > \alpha_0$, i.e., $Q_m = \mathcal{E}_{\alpha_2 > |g|^2 > \alpha_0} \{S_m\}$. We note that S_m is a function $|g|^2$ via ξ . The complication is that ξ can have two possible forms depending on the value of $|g|^2$ as shown in the following:

$$\begin{aligned} \xi &= \begin{cases} \alpha_s, & \text{if } \epsilon_s < \alpha_0^{-1}|g|^2 - 1 \\ \frac{\alpha_0^{-1}|g|^2-1}{P_s}, & \text{otherwise} \end{cases} \\ &= \begin{cases} \alpha_s, & \text{if } |g|^2 > \alpha_1 \\ \frac{\alpha_0^{-1}|g|^2-1}{P_s}, & \text{otherwise} \end{cases}. \end{aligned} \quad (27)$$

It is important to note that $\alpha_0 \leq \alpha_1 \leq \alpha_2$ always holds since

$$\frac{\epsilon_0}{P_0} \leq \frac{\epsilon_0(1 + \epsilon_s)}{P_0} \leq \frac{\epsilon_0(1 + \epsilon_s)}{P_0(1 - \epsilon_0\epsilon_s)}. \quad (28)$$

Therefore, a key step for evaluating $\mathcal{E}_{\alpha_2 > |g|^2 > \alpha_0} \{S_m\}$ is to calculate the following general expectation:

$$\begin{aligned}
& \mathcal{E}_{\alpha_2 > |g|^2 > \alpha_0} \left\{ e^{-p\xi} e^{-\mu|g|^2} \right\} \\
&= e^{-p\alpha_s} \int_{\alpha_1}^{\alpha_2} e^{-(1+\mu)x} dx + \int_{\alpha_0}^{\alpha_1} e^{-p\frac{\alpha_0^{-1}x-1}{P_s}} e^{-(1+\mu)x} dx \\
&= e^{-p\alpha_s} g_\mu(\alpha_1, \alpha_2) + e^{\frac{p}{P_s}} g_{\mu+\frac{p}{P_s}}(\alpha_0, \alpha_1). \tag{29}
\end{aligned}$$

By using the result shown in (29), Q_m can be calculated as follows:

$$\begin{aligned}
Q_m &= \eta_m \sum_{i=0}^{M-m-2} \binom{M-m-2}{i} \frac{(-1)^i}{m(i+1)} \sum_{p=0}^m \binom{m}{p} (-1)^p \\
&\times \left(\frac{\mu_3\phi(p, \mu_1) - \mu_5\phi(p, \mu_6)}{M-m} - \frac{\mu_4\phi(p, \mu_2) - \mu_5\phi(p, \mu_6)}{M-m-1-i} \right). \tag{30}
\end{aligned}$$

The form in (30) is quite involved and cannot be directly used to obtain a high SNR approximation later. In the following, we will show that the expression of Q_m can be simplified. In particular, Q_m can be first rewritten as follows:

$$\begin{aligned}
Q_m &= \frac{\eta_m}{m(M-m-1)} \sum_{i=0}^{M-m-2} \binom{M-m-1}{i+1} (-1)^i \sum_{p=0}^m \binom{m}{p} (-1)^p \\
&\times \left(\frac{\mu_3\phi(p, \mu_1) - \mu_5\phi(p, \mu_6)}{M-m} - \frac{\mu_4\phi(p, \mu_2) - \mu_5\phi(p, \mu_6)}{M-m-1-i} \right), \tag{31}
\end{aligned}$$

which is obtained by absorbing $i+1$ into the binomial coefficients $\binom{M-m-1}{i+1}$. By letting $l = i+1$, Q_m can be further rewritten as follows:

$$\begin{aligned}
Q_m &= \frac{-\eta_m}{m(M-m-1)} \sum_{l=0}^{M-m-1} \binom{M-m-1}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \\
&\times \left(\frac{\mu_3\phi(p, \mu_1) - \mu_5\phi(p, \mu_6)}{M-m} - \frac{\tilde{\mu}_4\phi(p, \tilde{\mu}_2) - \mu_5\phi(p, \mu_6)}{M-m-l} \right). \tag{32}
\end{aligned}$$

We note that the term $l = 0$ can be added since $\frac{\mu_3\phi(p, \mu_1) - \mu_5\phi(p, \mu_6)}{M-m} - \frac{\tilde{\mu}_4\phi(p, \tilde{\mu}_2) - \mu_5\phi(p, \mu_6)}{M-m-l} = 0$ for $l = 0$.

We further note the fact that μ_1, μ_3, μ_5 and μ_6 are not functions of l . Therefore, by using the fact that $\sum_{l=0}^n (-1)^l \binom{n}{l} = 0$, some terms in Q_m can be eliminated. In particular, Q_m can be simplified as follows:

$$\begin{aligned}
Q_m &= \frac{\eta_m}{m(M-m-1)} \sum_{l=0}^{M-m-1} \binom{M-m-1}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \frac{\tilde{\mu}_4 \phi(p, \tilde{\mu}_2) - \mu_5 \phi(p, \mu_6)}{M-m-l} \\
&\stackrel{(a)}{=} \tilde{\eta}_m \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p (\tilde{\mu}_4 \phi(p, \tilde{\mu}_2) - \mu_5 \phi(p, \mu_6)) \\
&\stackrel{(b)}{=} \tilde{\eta}_m \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \tilde{\mu}_4 \phi(p, \tilde{\mu}_2), \tag{33}
\end{aligned}$$

where step (a) follows by absorbing $M-m-l$ into the binomial coefficients, step (b) follows by using the fact that $\sum_{l=0}^n (-1)^l \binom{n}{l} = 0$. Again, we note that in step (a), the term $l = M-m$ can be added without changing the value of the summation since $\tilde{\mu}_4 \phi(p, \tilde{\mu}_2) - \mu_5 \phi(p, \mu_6) = 0$ for $l = M-m$. Comparing (30) to (33), we note that the expression for Q_m has been simplified.

B. Evaluation of Q_{M-1}

Recall that Q_{M-1} can be expressed as follows:

$$\begin{aligned}
Q_{M-1} &= \mathcal{E}_{|g|^2 > \alpha_0} \left\{ \text{P} \left(|h_{M-1}|^2 < \xi, \right. \right. \\
&\quad \left. \left. |h_M|^2 > \frac{\alpha_0^{-1} |g|^2 - 1}{P_s}, |h_M|^2 < \alpha_s (1 + P_0 |g|^2) \right) \right\}. \tag{34}
\end{aligned}$$

Denote the probability inside of the expectation in (34) by S_{M-1} . Again, by applying (23), Q_{M-1} can be expressed as follows:

$$Q_{M-1} = \mathcal{E}_{\alpha_2 > |g|^2 > \alpha_0} \{ S_{M-1} \}. \tag{35}$$

Unlike S_m , $1 \leq m \leq M-2$, S_{M-1} becomes a function of two order statistics, $|h_{M-1}|^2$ and $|h_M|^2$, whose joint pdf is given by

$$f_{|h_{M-1}|^2, |h_M|^2}(x, y) = \tilde{\eta}_0 e^{-x} (1 - e^{-x})^{M-2} e^{-y}, \tag{36}$$

where $x \leq y$. By using this joint pdf, S_{M-1} can be calculated as follows:

$$\begin{aligned} S_{M-1} &= \frac{\tilde{\eta}_0 (1 - e^{-\xi})^{M-1} \left(e^{-\frac{\alpha_0^{-1}|g|^2-1}{P_s}} - e^{-\alpha_s(1+P_0|g|^2)} \right)}{M-1} \\ &= \sum_{i=0}^{M-1} \binom{M-1}{i} (-1)^i \frac{\tilde{\eta}_0}{M-1} e^{-i\xi} \\ &\quad \times \left(e^{\frac{1}{P_s}} e^{-\mu_7|g|^2} - e^{-\alpha_s} e^{-\mu_8|g|^2} \right). \end{aligned} \quad (37)$$

By applying (29), Q_{M-1} can be obtained as follows:

$$Q_{M-1} = \sum_{i=0}^{M-1} \binom{M-1}{i} (-1)^i \frac{\tilde{\eta}_0}{M-1} \quad (38)$$

$$\times \left(e^{\frac{1}{P_s}} \phi(i, \mu_7) - e^{-\alpha_s} \phi(i, \mu_8) \right). \quad (39)$$

C. Evaluation of Q_M

Unlike Q_m , $1 \leq m \leq M-1$, Q_M is a function of h_M and g . In particular, recall that Q_M can be expressed as follows:

$$\begin{aligned} Q_M &= \text{P} \left(\log(1 + P_s|h_M|^2) < R_s, \right. \\ &\quad \left. \log \left(1 + \frac{P_s|h_M|^2}{P_0|g|^2+1} \right) < R_s, E_M, |g|^2 > \alpha_0 \right) \\ &= \text{P} \left(\log(1 + P_s|h_M|^2) < R_s, |h_M|^2 < \tau(|g|^2)P_s^{-1}, |g|^2 > \alpha_0 \right), \end{aligned} \quad (40)$$

where the last step follows from the fact that $P_s|h_M|^2 \geq \frac{P_s|h_M|^2}{P_0|g|^2+1}$. Therefore, we can rewrite Q_M as follows:

$$\begin{aligned} Q_M &= \text{P} \left(|h_M|^2 < \alpha_s, |h_M|^2 < \frac{\alpha_0^{-1}|g|^2-1}{P_s}, |g|^2 > \alpha_0 \right) \\ &= \text{P} \left(|h_M|^2 < \frac{\alpha_0^{-1}|g|^2-1}{P_s}, \alpha_0 < |g|^2 < \alpha_1 \right) + \text{P} \left(|h_M|^2 < \alpha_s, |g|^2 > \alpha_1 \right), \end{aligned} \quad (41)$$

where we use the fact that $\alpha_s < \frac{\alpha_0^{-1}|g|^2-1}{P_s}$ is guaranteed if $|g|^2 > \alpha_1$. By applying the fact that h_M and g are independent, Q_M can be calculated as follows:

$$\begin{aligned} Q_M &= \int_{\alpha_0}^{\alpha_1} \left(1 - e^{-\frac{\alpha_0^{-1}x-1}{P_s}} \right)^M e^{-x} dx \\ &\quad + (1 - e^{-\alpha_s})^M e^{-\alpha_1}. \end{aligned} \quad (42)$$

With some algebraic manipulations, Q_M can be finally expressed as follows:

$$Q_M = \sum_{i=0}^M \binom{M}{i} (-1)^i e^{\frac{i}{P_s}} g_{\frac{i}{\alpha_0 P_s}} (\alpha_0, \alpha_1) + (1 - e^{-\alpha_s})^M e^{-\alpha_1}. \quad (43)$$

D. Evaluation of Q_0

Q_0 is surprisingly more complicated to analyze, compared to Q_M . Recall that Q_0 can be expressed as follows:

$$Q_0 = \text{P} \left(\log \left(1 + \frac{P_s |h_M|^2}{P_0 |g|^2 + 1} \right) < R_s, |h_1|^2 > \frac{P_0 \epsilon_0^{-1} |g|^2 - 1}{P_s}, |g|^2 > \frac{\epsilon_0}{P_0} \right) \\ = \text{P} \left(|h_M|^2 < \alpha_s (P_0 |g|^2 + 1), |h_1|^2 > \frac{\alpha_0^{-1} |g|^2 - 1}{P_s}, |g|^2 > \alpha_0 \right). \quad (44)$$

Again, by applying the fact that the lower bound on $|h_M|^2$ needs to be larger than the upper bound on $|h_1|^2$ as discussed in (23), the probability Q_0 can be expressed as follows:

$$Q_0 = \mathcal{E}_{\alpha_0 < |g|^2 < \alpha_2} \left\{ \text{P} \left(|h_M|^2 < \alpha_s (P_0 |g|^2 + 1), |h_1|^2 > \frac{\alpha_0^{-1} |g|^2 - 1}{P_s} \right) \right\}.$$

Denote the probability inside the expectation in (45) by S_0 . S_0 is a function of two order statistics, $|h_1|^2$ and $|h_M|^2$, whose joint pdf is given by

$$f_{|h_1|^2, |h_M|^2}(x, y) = \tilde{\eta}_0 e^{-x} (e^{-x} - e^{-y})^{M-2} e^{-y} \\ = \tilde{\eta}_0 \sum_{i=0}^{M-2} (-1)^i \binom{M-2}{i} e^{-(M-1-i)x} e^{-(i+1)y}, \quad (45)$$

for $x \leq y$. For a fixed $|g|^2$ and by applying the joint pdf, S_0 can be calculated as follows:

$$S_0 = \tilde{\eta}_0 \sum_{i=0}^{M-2} (-1)^i \binom{M-2}{i} \\ \times \int_{\frac{\alpha_0^{-1} |g|^2 - 1}{P_s}}^{\alpha_s (P_0 |g|^2 + 1)} e^{-(M-1-i)x} \int_x^{\alpha_s (P_0 |g|^2 + 1)} e^{-(i+1)y} dy dx. \quad (46)$$

With some algebraic manipulations, S_0 can be expressed as follows:

$$S_0 = \tilde{\eta}_0 \sum_{i=0}^{M-2} (-1)^i \binom{M-2}{i} \\ \times \left(\frac{e^{\frac{M}{P_s}} e^{-\mu_{10} |g|^2} - e^{-M \alpha_s} e^{-\mu_{11} |g|^2}}{M(i+1)} \right. \\ \left. - \frac{\left(e^{\frac{M-1-i}{P_s} - (i+1)\alpha_s} e^{-\mu_{12} |g|^2} - e^{-M \alpha_s} e^{-\mu_{11} |g|^2} \right)}{(i+1)(M-1-i)} \right), \quad (47)$$

where $\mu_{10} = \frac{M}{\alpha_0 P_s}$ and $\mu_{11} = M\alpha_s P_0$, and $\mu_{12} = (i+1)\alpha_s P_0 + (M-1-i)\frac{\alpha_0^{-1}}{P_s}$.

By applying the integration result in (29), Q_0 can be expressed as follows:

$$Q_0 = \tilde{\eta}_0 \sum_{i=0}^{M-2} (-1)^i \binom{M-2}{i} \times \left(\frac{e^{\frac{M}{P_s}} g_{\mu_{10}}(\alpha_0, \alpha_2) - e^{-M\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2)}{M(i+1)} - \frac{\left(e^{\frac{M-1-i}{P_s} - (i+1)\alpha_s} g_{\mu_{12}}(\alpha_0, \alpha_2) - e^{-M\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2) \right)}{(i+1)(M-1-i)} \right). \quad (48)$$

The expression in (48) is quite involved, and cannot be used directly to obtain a high SNR approximation. In the following, we will show that (48) can be simplified. First, the expression for Q_0 is modified as follows:

$$Q_0 = \frac{\tilde{\eta}_0}{M-1} \sum_{i=0}^{M-2} (-1)^i \binom{M-1}{i+1} \times \left(\frac{e^{\frac{M}{P_s}} g_{\mu_{10}}(\alpha_0, \alpha_2) - e^{-M\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2)}{M} - \frac{\left(e^{\frac{M-1-i}{P_s} - (i+1)\alpha_s} g_{\mu_{12}}(\alpha_0, \alpha_2) - e^{-M\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2) \right)}{(M-1-i)} \right), \quad (49)$$

which is obtained by absorbing $i+1$ into the binomial coefficients. The expression for Q_0 can be further modified as follows:

$$Q_0 = -\frac{\tilde{\eta}_0}{M-1} \sum_{l=0}^{M-1} (-1)^l \binom{M-1}{l} \times \left(\frac{e^{\frac{M}{P_s}} g_{\mu_{10}}(\alpha_0, \alpha_2) - e^{-M\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2)}{M} - e^{-l\alpha_s} \frac{\left(e^{\frac{M-l}{P_s}} g_{\tilde{\mu}_{12}}(\alpha_0, \alpha_2) - e^{-(M-l)\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2) \right)}{(M-l)} \right) \quad (50)$$

which is obtained by substituting $l = i+1$. We note that the term $l = 0$ can be added without changing the value of the summation, since the terms in the second and third lines in (50) cancel each other when $l = 0$.

By using the fact that $\sum_{l=0}^n (-1)^l \binom{n}{l} = 0$, Q_0 can be further simplified as follows:

$$\begin{aligned}
Q_0 &= \frac{\tilde{\eta}_0}{M-1} \sum_{l=0}^{M-1} (-1)^l \binom{M-1}{l} e^{-l\alpha_s} \\
&\quad \times \left(\frac{e^{\frac{M-l}{P_s}} g_{\tilde{\mu}_{12}}(\alpha_0, \alpha_2) - e^{-(M-l)\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2)}{(M-l)} \right) \\
&= \frac{\tilde{\eta}_0}{M(M-1)} \sum_{l=0}^M (-1)^l \binom{M}{l} e^{-l\alpha_s} \\
&\quad \times \left(e^{\frac{M-l}{P_s}} g_{\tilde{\mu}_{12}}(\alpha_0, \alpha_2) - e^{-(M-l)\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2) \right),
\end{aligned} \tag{51}$$

where the last step is obtained by absorbing $(M-l)$ into the binomial coefficients. In addition, we also note that the term $l = M$ can be added since $e^{\frac{M-l}{P_s}} g_{\tilde{\mu}_{12}}(\alpha_0, \alpha_2) - e^{-(M-l)\alpha_s} g_{\mu_{11}}(\alpha_0, \alpha_2) = 0$ when $l = M$.

Again, by using the fact that $\sum_{l=0}^n (-1)^l \binom{n}{l} = 0$, Q_0 can be further simplified as follows:

$$Q_0 = \frac{\tilde{\eta}_0}{M(M-1)} \sum_{l=0}^M (-1)^l \binom{M}{l} e^{-l\alpha_s} e^{\frac{M-l}{P_s}} g_{\tilde{\mu}_{12}}(\alpha_0, \alpha_2). \tag{52}$$

Q_{M+1} can be evaluated similarly to Q_M since both are functions of h_M and g , and it can be expressed as follows:

$$\begin{aligned}
Q_{M+1} &= \text{P} \left(\log \left(1 + \frac{P_s |h_M|^2}{P_0 |g|^2 + 1} \right) < R_s, |g|^2 < \alpha_0 \right) \\
&= \sum_{i=0}^M \binom{M}{i} (-1)^i e^{-i\alpha_s} \frac{1 - e^{-(1+i\alpha_s P_0)\alpha_0}}{1 + i\alpha_s P_0}.
\end{aligned} \tag{53}$$

Therefore, by combining (33), (38), (44), (52) and (53), the overall outage probability is obtained as shown in the theorem and the proof is complete.

APPENDIX B

PROOF FOR THEOREM 2

As discussed in the proof for Theorem 1, Q_m depends on the value of m . Therefore, the high SNR approximations for different Q_m will be discussed separately in the following subsections.

A. High SNR Approximation for Q_m , $1 \leq m \leq M - 2$

Among all the terms in (10), the expression for Q_m , $1 \leq m \leq M - 2$, is the most complicated one, as is evident from the proof of Theorem 1. First, recall Q_m can be expressed as follows:

$$\begin{aligned} Q_m &= \bar{\eta}_m \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \tilde{\mu}_4 \phi(p, \tilde{\mu}_2) \\ &= \bar{\eta}_m \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \\ &\quad \times \left(\tilde{\mu}_4 e^{-p\alpha_s} g_{\tilde{\mu}_2}(\alpha_1, \alpha_2) + \tilde{\mu}_4 e^{\frac{p}{P_s}} g_{\tilde{\mu}_2 + \frac{p}{P_s\alpha_0}}(\alpha_0, \alpha_1) \right). \end{aligned} \quad (54)$$

In order to facilitate a high SNR approximation, Q_m is rewritten as follows:

$$\begin{aligned} Q_m &= \bar{\eta}_m \int_{\alpha_1}^{\alpha_2} \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \tilde{\mu}_4 e^{-p\alpha_s} e^{-(1+\tilde{\mu}_2)x} dx \\ &\quad + \bar{\eta}_m \int_{\alpha_0}^{\alpha_1} \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \tilde{\mu}_4 e^{\frac{p}{P_s}} e^{-(1+\tilde{\mu}_2 + \frac{p}{P_s\alpha_0})x} dx, \end{aligned} \quad (55)$$

By applying the approximation, $e^{-x} \approx 1 - x$ for $x \rightarrow 0$ and also using the definitions of $\tilde{\mu}_2$ and $\tilde{\mu}_4$, Q_m can be approximated as follows:

$$\begin{aligned} Q_m &\approx \bar{\eta}_m \int_{\alpha_1}^{\alpha_2} \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \\ &\quad \times e^{-l\alpha_s - \frac{l}{P_s}} e^{-p\alpha_s} e^{-(l\epsilon_s - l\epsilon_0^{-1})x} dx \\ &\quad + \bar{\eta}_m \int_{\alpha_0}^{\alpha_1} \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l \sum_{p=0}^m \binom{m}{p} (-1)^p \\ &\quad \times e^{-l\alpha_s - l\frac{1}{P_s}} e^{\frac{p}{P_s}} e^{-(l\epsilon_s - l\epsilon_0^{-1} + \frac{p}{P_s\alpha_0})x} dx. \end{aligned} \quad (56)$$

By using the fact that $\sum_{l=0}^n (-1)^l \binom{n}{l} a^l = (1-a)^n$, the approximation of Q_m can be further simplified as follows:

$$\begin{aligned} Q_m &\approx \bar{\eta}_m (1 - e^{-\alpha_s})^m \int_{\alpha_1}^{\alpha_2} \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l e^{-l\alpha_s - \frac{l}{P_s}} e^{-(l\epsilon_s - l\epsilon_0^{-1})x} dx \\ &\quad + \bar{\eta}_m \int_{\alpha_0}^{\alpha_1} \left(1 - e^{\frac{1}{P_s} - \frac{x}{P_s\alpha_0}}\right)^m \sum_{l=0}^{M-m} \binom{M-m}{l} (-1)^l e^{-l\alpha_s - l\frac{1}{P_s}} e^{-(l\epsilon_s - l\epsilon_0^{-1})x} dx \\ &= \bar{\eta}_m (1 - e^{-\alpha_s})^m \int_{\alpha_1}^{\alpha_2} \left(1 - e^{-(\alpha_s + \frac{1}{P_s} + (\epsilon_s - \epsilon_0^{-1})x)}\right)^{M-m} dx \\ &\quad + \bar{\eta}_m \int_{\alpha_0}^{\alpha_1} \left(1 - e^{\frac{1}{P_s} - \frac{x}{P_s\alpha_0}}\right)^m \left(1 - e^{-(\alpha_s + \frac{1}{P_s} + (\epsilon_s - \epsilon_0^{-1})x)}\right)^{M-m} dx. \end{aligned} \quad (57)$$

A more simplified form of Q_m can be obtained by carrying out the following high SNR approximations:

$$Q_m \approx \bar{\eta}_m (1 - e^{-\alpha_s})^m \int_{\alpha_1}^{\alpha_2} \left(\alpha_s + \frac{1}{P_s} + (\epsilon_s - \epsilon_0^{-1})x \right)^{M-m} dx \quad (58)$$

$$+ \frac{\bar{\eta}_m}{P_s^m \alpha_0^m} \int_{\alpha_0}^{\alpha_1} (x - \alpha_0)^m \left(\alpha_s + \frac{1}{P_s} + (\epsilon_s - \epsilon_0^{-1})x \right)^{M-m} dx,$$

With some algebraic manipulations, the high SNR approximation for Q_m can be obtained as follows:

$$Q_m \approx \frac{\bar{\eta}_m}{P_s^{M+1}} \epsilon_s^m \sum_{i=0}^{M-m} \binom{M-m}{i} (\epsilon_s + 1)^{M-m-i} (\epsilon_s - \epsilon_0^{-1})^i$$

$$\times \epsilon_0^{i+1} \frac{\tilde{\alpha}_2^{i+1} - (1 + \epsilon_s)^{i+1}}{i+1} + \frac{\bar{\eta}_m}{P_s^{M+1} \epsilon_0^m} \sum_{i=0}^{M-m} \binom{M-m}{i}$$

$$\times (\epsilon_s + \epsilon_0 \epsilon_s)^{M-m-i} (\epsilon_s - \epsilon_0^{-1})^i \epsilon_0^{m+i+1} \frac{\epsilon_s^{m+i+1}}{m+i+1}. \quad (59)$$

B. High SNR Approximation for Q_0

To facilitate the asymptotic analysis, Q_0 can be rewritten as follows:

$$Q_0 = \frac{\tilde{\eta}_0}{M(M-1)} \sum_{l=0}^M (-1)^l \binom{M}{l} e^{-l\alpha_s} e^{\frac{M-l}{P_s}} g_{\tilde{\mu}_{12}}(\alpha_0, \alpha_2)$$

$$= \frac{\tilde{\eta}_0}{M(M-1)} \sum_{l=0}^M (-1)^l \binom{M}{l} e^{-l\alpha_s} e^{\frac{M-l}{P_s}} \frac{e^{-(1+\tilde{\mu}_{12})\alpha_0} - e^{-(1+\tilde{\mu}_{12})\alpha_2}}{1 + \tilde{\mu}_{12}}.$$

To carry out high SNR approximations, Q_0 can be first rewritten as follows:

$$Q_0 = \frac{\tilde{\eta}_0}{M(M-1)} \sum_{l=0}^M (-1)^l \binom{M}{l} e^{-l\alpha_s} e^{\frac{M-l}{P_s}} \int_{\alpha_0}^{\alpha_2} e^{-(1+\tilde{\mu}_{12})x} dx \quad (60)$$

$$= \frac{\tilde{\eta}_0}{M(M-1)} e^{\frac{M}{P_s}} \int_{\alpha_0}^{\alpha_2} e^{-x} e^{-M\epsilon_0^{-1}x} \sum_{l=0}^M (-1)^l \binom{M}{l}$$

$$\times e^{-l(\alpha_s + \frac{1}{P_s} + \epsilon_s x - \epsilon_0^{-1}x)} dx.$$

By using the fact that $\sum_{l=0}^n (-1)^l \binom{n}{l} a^l = (1-a)^n$, Q_0 can be expressed as follows:

$$Q_0 = \frac{\tilde{\eta}_0}{M(M-1)} e^{\frac{M}{P_s}} \int_{\alpha_0}^{\alpha_2} \left(1 - e^{-l(\alpha_s + \frac{1}{P_s} + \epsilon_s x - \epsilon_0^{-1}x)} \right)^M dx \quad (61)$$

$$\approx \frac{\tilde{\eta}_0}{M(M-1)} e^{\frac{M}{P_s}} \int_{\alpha_0}^{\alpha_2} \left(\alpha_s + \frac{1}{P_s} + \epsilon_s x - \epsilon_0^{-1}x \right)^M dx,$$

where the last step is obtained by applying the following approximation, $e^{-x} \approx 1 - x$ for $x \rightarrow 0$.

By applying the binomial expansion, Q_0 can be expressed as follows:

$$\begin{aligned} Q_0 &\approx \frac{\tilde{\eta}_0}{M(M-1)} e^{\frac{M}{P_s}} \sum_{i=0}^M \binom{M}{i} \left(\alpha_s + \frac{1}{P_s}\right)^{M-i} (\epsilon_s - \epsilon_0^{-1})^i \int_{\alpha_0}^{\alpha_2} x^i dx \\ &= \frac{\tilde{\eta}_0}{M(M-1)} e^{\frac{M}{P_s}} \sum_{i=0}^M \binom{M}{i} \left(\alpha_s + \frac{1}{P_s}\right)^{M-i} \\ &\quad \times (\epsilon_s - \epsilon_0^{-1})^i \frac{\alpha_2^{i+1} - \alpha_0^{i+1}}{i+1}. \end{aligned} \quad (62)$$

C. High SNR Approximation for Q_{M-1}

First, we recall that Q_{M-1} can be expressed as follows:

$$\begin{aligned} Q_{M-1} &= \sum_{i=0}^{M-1} \binom{M-1}{i} (-1)^i \frac{\tilde{\eta}_0}{M-1} \\ &\quad \times \left(e^{\frac{1}{P_s}} e^{-i\alpha_s} g_{\mu_7}(\alpha_1, \alpha_2) + e^{\frac{1}{P_s}} e^{\frac{i}{P_s}} g_{\mu_7 + \frac{i}{P_s \alpha_0}}(\alpha_0, \alpha_1) \right. \\ &\quad \left. - e^{-\alpha_s} e^{-i\alpha_s} g_{\mu_8}(\alpha_1, \alpha_2) - e^{-\alpha_s} e^{\frac{i}{P_s}} g_{\mu_8 + \frac{i}{P_s \alpha_0}}(\alpha_0, \alpha_1) \right). \end{aligned} \quad (63)$$

In order to obtain the high SNR approximation, we first express Q_{M-1} as follows:

$$\begin{aligned} Q_{M-1} &= \sum_{i=0}^{M-1} \binom{M-1}{i} (-1)^i \frac{\tilde{\eta}_0}{M-1} \\ &\quad \times \left(\int_{\alpha_1}^{\alpha_2} e^{\frac{1}{P_s}} e^{-i\alpha_s} e^{-(1+\mu_7)x} dx + \int_{\alpha_0}^{\alpha_1} e^{\frac{1}{P_s}} e^{\frac{i}{P_s}} e^{-(1+\mu_7 + \frac{i}{P_s \alpha_0})x} dx \right. \\ &\quad \left. - \int_{\alpha_1}^{\alpha_2} e^{-\alpha_s} e^{-i\alpha_s} e^{-(1+\mu_8)x} dx - \int_{\alpha_0}^{\alpha_1} e^{-\alpha_s} e^{\frac{i}{P_s}} e^{-(1+\mu_8 + \frac{i}{P_s \alpha_0})x} dx \right). \end{aligned} \quad (65)$$

By using the fact that $\sum_{l=0}^n (-1)^l \binom{n}{l} = 0$, Q_{M-1} can be further expressed as follows:

$$\begin{aligned} Q_{M-1} &= \frac{\tilde{\eta}_0}{M-1} \left(e^{\frac{1}{P_s}} (1 - e^{-\alpha_s})^{M-1} \int_{\alpha_1}^{\alpha_2} e^{-(1 + \frac{1}{P_s \alpha_0})x} dx \right. \\ &\quad + \int_{\alpha_0}^{\alpha_1} e^{\frac{1}{P_s}} \left(1 - e^{\frac{1}{P_s} - \frac{1}{P_s \alpha_0} x} \right)^{M-1} e^{-(1 + \frac{1}{P_s \alpha_0})x} dx \\ &\quad - e^{-\alpha_s} (1 - e^{-\alpha_s})^{M-1} \int_{\alpha_1}^{\alpha_2} e^{-(1 + \alpha_s P_0)x} dx \\ &\quad \left. - \int_{\alpha_0}^{\alpha_1} e^{-\alpha_s} \left(1 - e^{\frac{1}{P_s} - \frac{1}{P_s \alpha_0} x} \right)^{M-1} e^{-(1 + \alpha_s P_0)x} dx \right). \end{aligned} \quad (66)$$

Directly applying the approximation, $e^{-x} \approx 1 - x$ for $x \rightarrow 0$, to the above equation results in a very complicated form. In order to facilitate the high SNR approximation, we rearrange the four terms in the above equation as follows:

$$Q_{M-1} = \frac{\tilde{\eta}_0}{M-1} \left((1 - e^{-\alpha_s})^{M-1} \times \left(\int_{\alpha_1}^{\alpha_2} e^{\frac{1}{P_s} - (1 + \frac{1}{P_s \alpha_0})x} - e^{-\alpha_s - (1 + \alpha_s P_0)x} dx \right) + \int_{\alpha_0}^{\alpha_1} \left(1 - e^{\frac{1}{P_s} - \frac{1}{P_s \alpha_0}x} \right)^{M-1} \times \left(e^{\frac{1}{P_s} - (1 + \frac{1}{P_s \alpha_0})x} - e^{-\alpha_s - (1 + \alpha_s P_0)x} \right) dx \right).$$

By applying the approximation, $e^{-x_1} - e^{-x_2} \approx x_2 - x_1$ for $x_1 \rightarrow 0$ and $x_2 \rightarrow 0$, Q_{M-1} can be approximated as follows:

$$Q_{M-1} \approx \frac{\tilde{\eta}_0}{M-1} \left((1 - e^{-\alpha_s})^{M-1} \times \left(\int_{\alpha_1}^{\alpha_2} \left(\frac{1}{P_s} - \frac{1}{P_s \alpha_0}x + \alpha_s + \alpha_s P_0 x \right) dx \right) + \int_{\alpha_0}^{\alpha_1} \left(1 - e^{\frac{1}{P_s} - \frac{1}{P_s \alpha_0}x} \right)^{M-1} \times \left(\frac{1}{P_s} - \frac{1}{P_s \alpha_0}x + \alpha_s + \alpha_s P_0 x \right) dx \right). \quad (67)$$

The approximation shown in (67) can be further approximated as follows:

$$Q_{M-1} \approx \frac{\tilde{\eta}_0}{M-1} \left((1 - e^{-\alpha_s})^{M-1} \left(\int_{\alpha_1}^{\alpha_2} \left(\frac{1}{P_s} + \alpha_s \right) dx \right) + \int_{\alpha_0}^{\alpha_1} \left(1 - e^{\frac{1}{P_s} - \frac{1}{P_s \alpha_0}x} \right)^{M-1} \times \left(\frac{1}{P_s} + \alpha_s \right) dx \right) = \frac{\tilde{\eta}_0 (1 + \epsilon_s) \epsilon_s^{M-1} \epsilon_0}{P_s^{M+1} (M-1)} \left((\tilde{\alpha}_2 - 1 - \epsilon_s) + \frac{\epsilon_s}{M} \right). \quad (68)$$

Following similar steps as for the approximation of Q_m , $1 \leq m \leq M-1$, Q_M can be approximated as follows:

$$Q_M \approx \frac{1}{(M+1)\epsilon_0^M} \alpha_0^{M+1} \epsilon_s^{M+1} + \alpha_s^M, \quad (69)$$

and Q_{M+1} can be approximated as follows:

$$Q_{M+1} \alpha_s^M \frac{(1 + \epsilon_0)^{M+1} - 1}{P_0 (M+1)}. \quad (70)$$

By combining (59), (62), (68), (69) and (70), the high SNR approximation for P_{out} can be obtained and the proof for Theorem 2 is complete.

REFERENCES

- [1] “Roadmap for IoT research, innovation and development in europe,” EU NGIoT, Jan. 2020.
- [2] A. Bayesteh, E. Yi, H. Nikopour, and H. Baligh, “Blind detection of SCMA for uplink grant-free multiple-access,” in *Proc. Int. Symp. Wireless Commun. Systems (ISWCS)*, Barcelona, Spain, Aug. 2014.
- [3] F. Clazzer, A. Munari, G. Liva, F. Lazaro, C. Stefanovic, and P. Popovski, “From 5G to 6G: Has the time for modern random access come?” (submitted) Available on-line at arXiv:1903.03063v1.
- [4] 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, Sept. 2018.
- [5] Z. Chen, F. Sohrabi, and W. Yu, “Multi-cell sparse activity detection for massive random access: Massive MIMO versus cooperative MIMO,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4060–4074, Aug. 2019.
- [6] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Springer Press, 2019.
- [7] Y. Zhang, Q. Guo, Z. Wang, J. Xi, and N. Wu, “Block sparse Bayesian learning based joint user activity detection and channel estimation for grant-free NOMA systems,” *IEEE Trans. Veh. Tech.*, vol. 67, no. 10, pp. 9631–9640, Oct. 2018.
- [8] S. Doğan, A. Tusha, and H. Arslan, “NOMA with index modulation for uplink URLLC through grant-free access,” *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1249–1257, Oct. 2019.
- [9] Y. Du, B. Dong, Z. Chen, X. Wang, Z. Liu, P. Gao, and S. Li, “Efficient multi-user detection for uplink grant-free NOMA: Prior-information aided adaptive compressive sensing perspective,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812–2828, Dec. 2017.
- [10] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, “A novel analytical framework for massive grant-free NOMA,” *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2436–2449, Mar. 2019.
- [11] J. Choi, “NOMA-based compressive random access using gaussian spreading,” *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5167–5177, Jul. 2019.
- [12] J. Seo and B. C. Jung, “Random access games with cost of waiting for uplink NOMA systems,” *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1361–1364, Oct. 2019.
- [13] J. Choi, “NOMA based random access with multichannel ALOHA,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2736–2743, 2017.
- [14] Z. Ding, R. Schober, P. Fan, and H. V. Poor, “Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access,” *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, Jun. 2019.
- [15] Q. Zhao and L. Tong, “Opportunistic carrier sensing for energy-efficient information retrieval in sensor networks,” *EURASIP J. Wireless Commun. Networking*, vol. 2, no. 2, pp. 231–241, Apr. 2005.
- [16] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, “A simple cooperative diversity method based on network path selection,” *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 659–672, Mar. 2006.
- [17] R. Talak and N. B. Mehta, “Feedback overhead-aware, distributed, fast, and reliable selection,” *IEEE Trans. Commu.*, vol. 60, no. 11, pp. 3417–3428, Nov. 2012.
- [18] H. A. David and H. N. Nagaraja, *Order Statistics*. John Wiley, New York, 3rd ed., 2003.