Bit Error Probability of SM-MIMO Over Generalized Fading Channels

Marco Di Renzo, *Member, IEEE*, and Harald Haas, *Member, IEEE*

Abstract—In this paper, we study the performance of spatial modulation (SM) multiple-input–multiple-output (MIMO) wireless systems over generic fading channels. More precisely, a comprehensive analytical framework to compute the average bit error probability (ABEP) is introduced, which can be used for any MIMO setup, for arbitrary correlated fading channels, and for generic modulation schemes. It is shown that, when compared with state-of-the-art literature, our framework 1) has more general applicability over generalized fading channels, 2) is, in general, more accurate as it exploits an improved union-bound method, and, 3) more importantly, clearly highlights interesting fundamental trends about the performance of SM, which are difficult to capture with available frameworks. For example, by focusing on the canonical reference scenario with independent identically distributed Rayleigh fading, we introduce very simple formulas that yield insightful design information on the optimal modulation scheme to be used for the signal constellation diagram, as well as highlight the different roles played by the bit mapping on the signal and spatial constellation diagrams. Numerical results show that, for many MIMO setups, SM with phase-shift-keying (PSK) modulation outperforms SM with quadrature-amplitude modulation (QAM), which is a result never reported in the literature. In addition, by exploiting asymptotic analysis, closed-form formulas of the performance gain of SM over other single-antenna transmission technologies are provided. Numerical results show that SM can outperform many single-antenna systems and that, for any transmission rate, there is an optimal allocation of the information bits onto spatial and signal constellation diagrams. Furthermore, by focusing on the Nakagami-m fading scenario with generically correlated fading, we show that fading severity plays a very important role in determining the diversity gain of SM. In particular, the performance gain over single-antenna systems increases for fading channels less severe than Rayleigh fading, whereas it gets smaller for more severe fading channels. In addition, it is shown that the impact of fading correlation at the transmitter is reduced for less severe fading. Finally, analytical frameworks and claims are substantiated through extensive Monte Carlo simulations.

M. Di Renzo is with the Laboratoire des Signaux et Systèmes, Unité Mixte de Recherche 8506, Centre National de la Recherche Scientifique-École Supérieure d'Électricité-Université Paris-Sud XI, 91192 Gif-sur-Yvette Cedex, France (e-mail: marco.direnzo@lss.supelec.fr).

H. Haas is with the Institute for Digital Communications (IDCOM), School of Engineering, College of Science and Engineering, The University of Edinburgh, EH9 3JL Edinburgh, UK (e-mail: h.haas@ed.ac.uk).

Index Terms—Large-scale antenna systems, "massive" multiple-input–multiple-output (MIMO) systems, performance analysis, single-RF MIMO design, spatial modulation (SM).

I. INTRODUCTION

SPATIAL MODULATION (SM) is a digital modulation concept for multiple-input–multiple-output (MIMO) wireless systems that has recently been introduced to increase the data rate of single-antenna systems by keeping a low-complexity transceiver design and by requiring no bandwidth expansion [1]–[4]. Unlike conventional spatial multiplexing schemes [2], [5], in SM, the multiplexing gain is realized through mapping a block of information bits into two information-carrying units: 1) the conventional signal constellation diagram [e.g., phase-shift-keying (PSK) or quadrature-amplitude modulation (QAM)] and 2) the so-called spatial constellation diagram, which is the antenna array at the transmitter. Like in conventional modulation schemes, the first subblock of information bits determines the point of the signal constellation diagram that is actually transmitted. Specific to SM is that the second subblock identifies the single active transmit antenna. As a result, the point of the signal constellation diagram is transmitted through a single active antenna belonging to the spatial constellation diagram. This simple modulation concept brings two main advantages: 1) For each channel use, the data rate increases by a factor equal to the logarithm of the number of antennas at the transmitter [2], [5], and 2) the receiver can detect the whole block of bits through single-stream demodulation, as the second subblock of bits is only implicitly transmitted through activation of the transmit antenna [6]. With respect to single-antenna systems, the net gain is a multiplexing gain for the same decoding complexity, whereas the price to pay is the need of more antennas at the transmitter. Recent results have showcased the performance gain of SM with respect to other state-of-the-art transmission technologies for single- and multi-antenna systems [5]–[19]. Finally, it is worth mentioning that SM seems to be an appealing transmission technology for high-rate and low-complexity MIMO implementations that exploit the recently proposed "massive MIMO" or "large-scale antenna systems" paradigm [20], [21]. In fact, in these systems, it is envisaged that improved performance and energy efficiency can be achieved by using tens or hundreds of antenna elements at the base station instead of exploiting base station cooperation. In this perspective, SM can be regarded as a low-complexity modulation scheme that exploits the "massive MIMO" idea but with a single active RF chain. The design of MIMO schemes that retain the benefits of multiple-antenna transmission while having a single active RF element is another recent and major trend in current and future MIMO research [22].

Since its introduction, many researchers have been studying the performance of SM-MIMO over fading channels either through time-consuming Monte Carlo simulations or through analytical modeling. Despite being more challenging, analytical modeling is, in general, preferred because of the following:

1) It allows a deeper understanding of the system performance.
2) It enables a simpler comparison with other competing transmission technologies.
3) It provides opportunities for system optimization.

A careful look at current state-of-the-art reveals the following contributions. The vast majority of analytical frameworks are useful for a special case of SM, which is called space-shift-keying (SSK) modulation [8]. SSK modulation is a low-complexity and low-data-rate version of SM where only the spatial constellation diagram is exploited for data modulation. This transmission technology is extensively studied in [23]–[29] for various MIMO setups and channel models. The analytical study conducted in [23]–[29] has highlighted fundamental properties of the spatial constellation diagram with respect to fading severity, channel correlation, power imbalance, transmit diversity, as well as robustness to multiple-access interference and channel estimation errors. However, these frameworks are of limited use to understand the performance of SM as the signal constellation diagram is neglected. On the other hand, analytical modeling of SM is limited to a very few papers, which have various limitations. In [5] and [30], the authors study a suboptimal receiver design, and the symbol error probability is computed by resorting to numerical integrations, which are not easy to compute and, in some cases, are numerically unstable. In [6], the authors study the average bit error probability (ABEP) of the maximum-likelihood (ML) optimum receiver over independent identically distributed (i.i.d.) Rayleigh fading. The framework is based on the union-bound method. Due to the absence of a scaling factor in the final formula [31], this bound is rather weak. Furthermore, and more importantly, the framework is valid for real-valued signal constellation points, and thus, it cannot be used, e.g., for PSK and QAM. In [9], the authors provide a first closed-form framework to compute the ABEP of SM-MIMO over generically correlated Rician fading and for arbitrary modulation schemes. In addition, the framework highlights some fundamental behaviors of SM, such as its incapability to achieve transmit diversity [17]. However, [9] has the following important limitations.

1) The analysis is applicable to Rician fading only.
2) The framework is based on conventional union-bound methods, whose accuracy degrades for high modulation orders and small numbers of receive antennas.
3) Signal and spatial constellation diagrams are treated as a single entity, which does not highlight the role played by each of them for various fading channels and modulation schemes.

In this depicted context, this paper is aimed at proposing a comprehensive analytical framework to study the ABEP of SM-MIMO over generalized fading channels. More specifically, we are interested in studying the following:

1) the interplay of signal and spatial constellation diagrams and whether an optimal allocation of the information bits between them exists;
2) the effect of adding the spatial constellation diagram on top of the signal constellation diagram, and whether conventional signal modulation schemes (e.g., PSK and QAM) are the best choice for SM, or whether new optimal modulation schemes should be designed to fully exploit the benefits of this hybrid modulation scheme;
3) the advantages and disadvantages of SM with respect to conventional single-antenna PSK/QAM and SSK modulations as a function of the MIMO setup and fading scenario.

To this end, we propose a new analytical framework that foresees to write the ABEP as the summation of the following three contributions:

1) a term that mainly depends on the signal constellation diagram;
2) a term that mainly depends on the spatial constellation diagram;
3) a joint term that depends on both constellation diagrams and highlights their interactions.

This new analytical formulation allows us to introduce an improved union-bound method that is more accurate than conventional union-bound methods and enables a deeper understanding of the role played by both information carrying units for various channel models and MIMO setups. Some of the most important and general results of this paper are as follows.

1) We show that SM outperforms single-antenna PSK/QAM schemes only for data rates greater than 2 bits per channel use (bpcu), and that SM with QAM-modulated points in the signal constellation diagram is never superior to single-antenna QAM for single-antenna receivers. On the other hand, for multi-antenna receivers and higher data rates, SM can significantly outperform single-antenna PSK/QAM. Closed-form expressions of this performance gain over i.i.d. Rayleigh fading are given.
2) Unlike single-antenna systems, where QAM always outperforms PSK, we show that SM with PSK-modulated points in the signal constellation diagram can outperform SM with QAM-modulated points. This is due to the interactions of signal and spatial constellation diagrams, and for i.i.d. Rayleigh fading, we show analytically that the ABEP of SM does not depend only on the Euclidean distance of the points in the signal constellation diagram. This provides important information on how to conceive new modulation schemes that are specifically optimized for SM.
3) By considering, as a case study, Nakagami-m fading, we show that the fading severity m_{Nak} plays an important role on the performance of SM. More specifically, like conventional modulation schemes, the ABEP gets worse for wireless channel with fading more severe ($0.5 \leq m_{\text{Nak}} < 1$) than Rayleigh. However, with respect to

single-antenna PSK/QAM modulation, the performance gain of SM increases thanks to the higher diversity gain experienced by the information bits mapped onto the spatial constellation diagram. On the contrary, the performance gain decreases for less severe ($m_{\text{Nak}} > 1$) fading because the diversity gain of the bits mapped onto the spatial constellation diagram is independent of the fading parameter m_{Nak}. In addition, it is shown that channel correlation at the transmitter has less impact when $0.5 \leq m_{\text{Nak}} < 1$.

The remainder of this paper is organized as follows. In Section II, the system model is described. In Section III, the improved union bound is introduced, and the specific contribution of spatial and signal constellation diagrams is shown. In Section IV, closed-form expressions of the ABEP for various fading channels and modulation schemes are provided. In Section V, the canonical i.i.d. Rayleigh fading channel is studied in detail, and closed-form expressions of the gain with respect to single-antenna PSK/QAM and SSK modulations are given. In Section VI, numerical results are shown to substantiate claims and analytical derivations. Finally, Section VII concludes this paper.

II. SYSTEM MODEL

We consider a generic $N_t \times N_r$ MIMO system, where N_t and N_r denote the antennas at the transmitter and receiver, respectively. We assume that the transmitter can send digital information via M complex symbols. In SM literature [4], the set of N_t antennas ($n_t = 1, 2, \ldots, N_t$) is called spatial constellation diagram, whereas the set of M complex points (χ_l for $l = 1, 2, \ldots, M$) is called signal constellation diagram. The basic idea of SM is to map blocks of information bits into two information carrying units [5]: 1) a symbol, which is chosen from the complex signal constellation diagram, and 2) a single active transmit antenna, which is chosen from the spatial constellation diagram.

More specifically, SM works as follows. At the transmitter, the bitstream is divided into blocks containing $\log_2(N_t) + \log_2(M)$ bits each, with $\log_2(N_t)$ and $\log_2(M)$ being the number of bits needed to identify a transmit antenna in the spatial constellation diagram and a symbol in the signal constellation diagram, respectively. Each block is split into two subblocks of $\log_2(N_t)$ and $\log_2(M)$ bits each. The bits in the first subblock are used to select the transmit antenna that is switched on for transmission, whereas all the other antennas are kept silent. The bits in the second subblock are used to choose a symbol in the signal constellation diagram. At the receiver, the detector can recover the whole block of $\log_2(N_t) + \log_2(M)$ information bits by solving an $N_t \times M$ hypothesis detection problem, which jointly estimates the transmit antenna that is not idle and the signal waveform that is transmitted from it.

In this paper, the generic block of $\log_2(N_t) + \log_2(M)$ bits is called "message," and it is denoted by $\mu(n_t, \chi_l)$, where $n_t = 1, 2, \ldots, N_t$ and $l = 1, 2, \ldots, M$ univocally identify the active transmit antenna n_t and the complex symbol χ_l transmitted from it, respectively. The $N_t \times M$ messages are assumed to be equiprobable.

A. Notation

Throughout this paper, we use the notation as follows.
1) We adopt a complex envelope signal representation.
2) $j = \sqrt{-1}$ is the imaginary unit.
3) $(\cdot)^*$ is the complex-conjugate operator.
4) $(x \otimes y)(t) = \int_{-\infty}^{+\infty} x(\xi) y(t - \xi) d\xi$ is the convolution of $x(\cdot)$ and $y(\cdot)$.
5) $|\cdot|$ is the absolute value.
6) $E\{\cdot\}$ is the expectation operator.
7) $Re\{\cdot\}$ and $Im\{\cdot\}$ are real and imaginary part operators.
8) $\Gamma(x) = \int_0^{+\infty} \xi^{x-1} \exp(-\xi) d\xi$ is the Gamma function.
9) $Q(x) = (1/\sqrt{2\pi}) \int_x^{+\infty} \exp(-t^2/2) dt$ is the Q-function.
10) E_m is the average energy per transmission.
11) T_m is the transmission time slot of each message.
12) $w(\cdot)$ is the unit energy, i.e., $\int_{-\infty}^{+\infty} |w(t)|^2 dt = 1$, elementary transmitted pulse waveform that is nonzero only in $[0, T_m]$.
13) The signal related to $\mu(n_t, \chi_l)$ and transmitted from antenna n_t is denoted by $s(t|\mu(n_t, \chi_l)) = \sqrt{E_m} \chi_l w(t)$.
14) The generic point of the signal constellation diagram χ_l is defined as $\chi_l = \chi_l^R + j\chi_l^I = \kappa_l \exp(j\phi_l)$, where $\chi_l^R = Re\{\chi_l\}$, $\chi_l^I = Im\{\chi_l\}$, $\kappa_l = \sqrt{(\chi_l^R)^2 + (\chi_l^I)^2}$, and $\phi_l = \arctan(\chi_l^I/\chi_l^R)$.
15) $Pr\{\cdot\}$ denotes probability.
16) The noise η_{n_r} at the input of the n_rth ($n_r = 1, 2, \ldots, N_r$) receive antenna is a complex additive white Gaussian noise (AWGN) process with power spectral density N_0 per dimension. Across the receive antennas, the noises η_{n_r} are statistically independent.
17) We introduce $\bar{\gamma} = E_m/(4N_0)$.
18) $\delta(\cdot)$ is the Dirac delta function.
19) $\lfloor x \rceil$ is the function that rounds x to the closest integer.
20) $\lfloor \cdot \rfloor$ is the floor function.
21) $G_{p,q}^{m,n}\left(\cdot \Big| \begin{matrix} (a_p) \\ (b_q) \end{matrix} \right)$ is the Meijer-G function defined in [32, Ch. 8, p. 519].
22) $\mathcal{M}_X(s) = E\{\exp(-sX)\}$ is the moment generating function (MGF) of random variable (RV) X.
23) $X \overset{d}{=} Y$ denotes that the RVs X and Y are equal in distribution or law, i.e., they have the same probability density function (pdf).
24) $(x!)$ is the factorial of x.
25) $(\cdot)^{-1}$ is the inverse of a square matrix.
26) $I_v(\cdot)$ is the modified Bessel function of first kind and order v [33, Ch. 9].
27) $\binom{\cdot}{\cdot}$ is the binomial coefficient.
28) $N_H((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l))$ is the Hamming distance of messages $\mu(\tilde{n}_t, \chi_{\tilde{l}})$ and $\mu(n_t, \chi_l)$, i.e., the number of positions where the information bits are different, with $0 \leq N_H((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l)) \leq \log_2(N_t M)$.

B. Channel Model

We consider the frequency-flat slowly-varying fading channel model as follows:
1) $h_{n_t, n_r}(\xi) = \alpha_{n_t, n_r} \delta(\xi - \tau_{n_t, n_r})$ is the channel impulse response of the wireless link from the n_tth transmit

antenna to the n_rth receive antenna. $\alpha_{n_t,n_r} = \alpha^R_{n_t,n_r} + j\alpha^I_{n_t,n_r} = \beta_{n_t,n_r}\exp(j\varphi_{n_t,n_r})$ is the complex channel gain, and τ_{n_t,n_r} is the propagation time delay. No specific distribution for the channel envelopes $\beta_{n_t,n_r} = \sqrt{(\alpha^R_{n_t,n_r})^2 + (\alpha^I_{n_t,n_r})^2}$, the channel phases $\varphi_{n_t,n_r} = \arctan(\alpha^I_{n_t,n_r}/\alpha^R_{n_t,n_r})$, and $\alpha^R_{n_t,n_r} = \mathrm{Re}\{\alpha_{n_t,n_r}\}$, $\alpha^I_{n_t,n_r} = \mathrm{Im}\{\alpha_{n_t,n_r}\}$ is assumed *a priori*.

2) The delays τ_{n_t,n_r} are assumed to be known at the receiver, i.e., perfect time synchronization is considered. In addition, we assume $\tau_{1,1} \cong \tau_{1,2} \cong \ldots \cong \tau_{N_t,N_r}$, which is a realistic assumption when the distance between transmitter and receiver is much larger than the spacing of the antenna elements [24]. Due to these assumptions, the delays τ_{n_t,n_r} are neglected in the next sections.

C. ML Optimum Detector

Let $\mu(\tilde{n}_t, \chi_{\tilde{l}})$ be the transmitted message.[1] The signal received by the n_rth receive antenna, if $\mu(\tilde{n}_t, \chi_{\tilde{l}})$ is transmitted, is

$$z_{n_r}(t) = s_{\mathrm{ch},n_r}\left(t\,|\,\mu(\tilde{n}_t,\chi_{\tilde{l}})\right) + \eta_{n_r}(t) \tag{1}$$

where $s_{\mathrm{ch},n_r}(t|\mu(\tilde{n}_t,\chi_{\tilde{l}})) = (s(\cdot|\mu(\tilde{n}_t,\chi_{\tilde{l}})) \otimes h_{\tilde{n}_t,n_r})(t) = \sqrt{E_m}\alpha_{\tilde{n}_t,n_r}\chi_{\tilde{l}}w(t)$ for $\tilde{n}_t = 1,2,\ldots,N_t$, $n_r = 1,2,\ldots,N_r$, and $\tilde{l} = 1,2,\ldots,M$.

Equation (1) is a general $N_t \times M$ hypothesis detection problem [34, Sec. 7.1], [35, Sec. 4.2, pp. 257] in AWGN when conditioning upon fading channel statistics. Thus, the ML optimum detector with full channel state information and perfect time synchronization at the receiver is given by [6], [34, Sec. 7.1]

$$
\begin{aligned}
(\hat{n}_t, \chi_{\hat{l}}) &= \operatorname*{arg\,max}_{\text{for } n_t=1,2,\ldots,N_t \text{ and } l=1,2,\ldots,M} \{D(n_t,\chi_l)\} \\
&= \operatorname*{arg\,max}_{\text{for } n_t=1,2,\ldots,N_t \text{ and } l=1,2,\ldots,M}
\end{aligned}
$$

---

[1]We emphasize that symbols with $\tilde{\ }$ identify the *actual* message that is transmitted, while symbols without $\tilde{\ }$ denote the *trial* message that is tested by the detector to solve the $N_t \times M$-hypothesis detection problem. Also, symbols with $\hat{\ }$ denote the message *estimated* by the detector. This notation does not apply to the antenna-index n_r at the receiver since there is no hypothesis-testing in this case.

$$
\left\{ \sum_{n_r=1}^{N_r} \left[\int_{T_m} z_{n_r}(t) s^*_{\mathrm{ch},n_r}\left(t|\mu(n_t,\chi_l)\right) dt \right.\right.
$$
$$
\left.\left. - \frac{1}{2}\int_{T_m} \left|s_{\mathrm{ch},n_r}\left(t|\mu(n_t,\chi_l)\right)\right|^2 dt \right] \right\}. \tag{2}
$$

The outcome of (2) is the estimated message $\mu(\hat{n}_t, \chi_{\hat{l}})$. Thus, the receiver is successful in decoding the whole block of bits if and only if $\mu(\hat{n}_t, \chi_{\hat{l}}) = \mu(\tilde{n}_t, \chi_{\tilde{l}})$, i.e., $\hat{n}_t = \tilde{n}_t$ and $\chi_{\hat{l}} = \chi_{\tilde{l}}$.

III. Average Bit Error Probability Over Generalized Fading Channels: Improved Union Bound

The exact ABEP of the detector in (2) can be computed in closed-form for arbitrary fading channels and modulation schemes as in (3), shown at the bottom of the page [36, eqs. (4) and (5)], where 1) $\boldsymbol{\alpha}$ is a shorthand to denote the set of $N_t \times N_r$ complex channel gains, i.e., α_{n_t,n_r} for $n_t = 1,2,\ldots,N_t$, $n_r = 1,2,\ldots,N_r$, and 2) $\mathrm{E}_{\boldsymbol{\alpha}}\{\cdot\}$ is the expectation computed over all the fading channels.

For arbitrary MIMO systems, the estimation of $\Pr\{(\hat{n}_t,\chi_{\hat{l}}) = (n_t,\chi_l)|(\tilde{n}_t,\chi_{\tilde{l}})\}$ is very complicated as it requires, in general, the computation of multidimensional integrals. Because of that, it is common practice to exploit union-bound methods [34] to compute the ABEP in (3) as in (4), shown at the bottom of the page, where $\mathrm{APEP}((\tilde{n}_t,\chi_{\tilde{l}}) \rightarrow (n_t,\chi_l)) = \mathrm{E}_{\boldsymbol{\alpha}(n_t,\tilde{n}_t)}\{\Pr\{(\tilde{n}_t,\chi_{\tilde{l}}) \rightarrow (n_t,\chi_l)\}\}$ is the average pairwise error probability (APEP), i.e., the probability of detecting $\mu(n_t,\chi_l)$ when, instead, $\mu(\tilde{n}_t,\chi_{\tilde{l}})$ is transmitted, under the assumption that $\mu(n_t,\chi_l)$ and $\mu(\tilde{n}_t,\chi_{\tilde{l}})$ are the only two messages possibly being transmitted, and $\mathrm{E}_{\boldsymbol{\alpha}(n_t,\tilde{n}_t)}\{\cdot\}$ is the expectation computed over the fading channels from the n_tth and $\tilde{n}_t$th transmit antennas and the N_r receive antennas. The APEP is equal to [26, eqs. (10)–(11)]

$$\mathrm{APEP}\left((\tilde{n}_t,\chi_{\tilde{l}}) \rightarrow (n_t,\chi_l)\right)$$
$$= \mathrm{E}_{\boldsymbol{\alpha}(n_t,\tilde{n}_t)}\left\{\Pr\{D(\tilde{n}_t,\chi_{\tilde{l}}) < D(n_t,\chi_l)\}\right\}$$
$$= \mathrm{E}_{\boldsymbol{\alpha}(n_t,\tilde{n}_t)}\left\{Q\left(\sqrt{\bar{\gamma}\sum_{n_r=1}^{N_r}|\alpha_{\tilde{n}_t,n_r}\chi_{\tilde{l}} - \alpha_{n_t,n_r}\chi_l|^2}\right)\right\}. \tag{5}$$

$$\mathrm{ABEP} = \mathrm{E}_{\boldsymbol{\alpha}}\left\{\frac{1}{N_t M}\frac{1}{\log_2(N_t M)}\sum_{n_t=1}^{N_t}\sum_{l=1}^{M}\sum_{\tilde{n}_t=1}^{N_t}\sum_{\tilde{l}=1}^{M}\left[N_H\left((\tilde{n}_t,\chi_{\tilde{l}}) \rightarrow (n_t,\chi_l)\right)\Pr\left\{(\hat{n}_t,\chi_{\hat{l}}) = (n_t,\chi_l)\,\middle|\,(\tilde{n}_t,\chi_{\tilde{l}})\right\}\right]\right\} \tag{3}$$

$$\mathrm{ABEP} \leq \frac{1}{N_t M}\frac{1}{\log_2(N_t M)}\sum_{n_t=1}^{N_t}\sum_{l=1}^{M}\sum_{\tilde{n}_t=1}^{N_t}\sum_{\tilde{l}=1}^{M}\left[N_H\left((\tilde{n}_t,\chi_{\tilde{l}}) \rightarrow (n_t,\chi_l)\right)\mathrm{APEP}\left((\tilde{n}_t,\chi_{\tilde{l}}) \rightarrow (n_t,\chi_l)\right)\right] \tag{4}$$

The union bound in (4) has been used in [6],[2] [9], and [24]–[26]. However, as mentioned in Section I, it has some limitations, as follows.

1) The roles played by spatial and signal constellation diagrams (and the related bit mapping) are hidden in the fourfold summation.
2) It is not accurate enough for large M and small N_r [37], as shown in Section VI.
3) Its computational complexity is the same, irrespective of modulation scheme and fading channel, when, instead, simpler formulas can be obtained in several cases.

A. Improved Upper Bound

To avoid the limitations of the conventional union bound when used for performance analysis of SM, and, more importantly, to get more insights about the expected performance of SM, we propose an improved upper bound. The new bound is summarized in Proposition 1.

Proposition 1: The ABEP in (3) can be tightly upper bounded as follows:

$$\text{ABEP} \leq \text{ABEP}_{\text{signal}} + \text{ABEP}_{\text{spatial}} + \text{ABEP}_{\text{joint}} \quad (6)$$

where $\text{ABEP}_{\text{signal}}$, $\text{ABEP}_{\text{spatial}}$, and $\text{ABEP}_{\text{joint}}$ are defined as

$$
\begin{cases}
\text{ABEP}_{\text{signal}} = \frac{1}{N_t} \frac{\log_2(M)}{\log_2(N_t M)} \sum_{n_t=1}^{N_t} \text{ABEP}_{\text{MOD}}(n_t) \\[2mm]
\text{ABEP}_{\text{spatial}} = \frac{1}{M} \frac{\log_2(N_t)}{\log_2(N_t M)} \sum_{l=1}^{M} \text{ABEP}_{\text{SSK}}(l) \\[2mm]
\text{ABEP}_{\text{joint}} = \frac{1}{N_t M} \frac{1}{\log_2(N_t M)} \sum_{n_t=1}^{N_t} \sum_{l=1}^{M} \sum_{\tilde{n}_t \neq n_t=1}^{N_t} \sum_{\tilde{l} \neq l=1}^{M} \\[2mm]
\quad \times \left\{ \left[N_H(\tilde{n}_t \to n_t) + N_H(\chi_{\tilde{l}} \to \chi_l) \right] \Upsilon(n_t, \chi_l, \tilde{n}_t, \chi_{\tilde{l}}) \right\}
\end{cases}
\tag{7}
$$

$$
\begin{cases}
\text{ABEP}_{\text{MOD}}(n_t) = \frac{1}{M} \frac{1}{\log_2(M)} \sum_{l=1}^{M} \sum_{\tilde{l}=1}^{M} \\[2mm]
\quad \times \left[N_H(\chi_{\tilde{l}} \to \chi_l) \text{E}_{\boldsymbol{\alpha}(n_t)} \left\{ \Pr\left\{ \chi_{\hat{l}} = \chi_l \,|\, \chi_{\tilde{l}} \right\} \right\} \right] \\[2mm]
\text{ABEP}_{\text{SSK}}(l) = \frac{1}{N_t} \frac{1}{\log_2(N_t)} \sum_{n_t=1}^{N_t} \sum_{\tilde{n}_t=1}^{N_t} \\[2mm]
\quad \times \left[N_H(\tilde{n}_t \to n_t) \Psi_l(n_t, \tilde{n}_t) \right]
\end{cases}
\tag{8}
$$

and

1) $N_H(\tilde{n}_t \to n_t)$, $N_H(\chi_{\tilde{l}} \to \chi_l)$ are the Hamming distances of the bits transmitted through spatial and signal constellation diagrams, respectively;
2) $\text{E}_{\boldsymbol{\alpha}(n_t)}\{\cdot\}$ is the expectation computed over the fading channels from the n_tth transmit antenna to the N_r receive antennas;
3) $\gamma_{(n_t, \tilde{n}_t)} = \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t, n_r} - \alpha_{n_t, n_r}|^2$;
4) $\gamma_{(n_t, \chi_l, \tilde{n}_t, \chi_{\tilde{l}})} = \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t, n_r} \chi_{\tilde{l}} - \alpha_{n_t, n_r} \chi_l|^2$;
5) $\Psi_l(n_t, \tilde{n}_t) = (1/\pi) \int_0^{\pi/2} \mathcal{M}_{\gamma_{(n_t, \tilde{n}_t)}} (\bar{\gamma} \kappa_l^2 / 2 \sin^2(\theta)) d\theta$;
6) $\Upsilon(n_t, \chi_l, \tilde{n}_t, \chi_{\tilde{l}}) = (1/\pi) \int_0^{\pi/2} \mathcal{M}_{\gamma_{(n_t, \chi_l, \tilde{n}_t, \chi_{\tilde{l}})}} (\bar{\gamma}/2 \times \sin^2(\theta)) d\theta$.

---

[2]In [6], the scaling factor $1/\log_2(N_t M)$ is not present, which yields a weak upper-bound [31].

Proof: See Appendix A. ∎

Let us analyze each term in (6) as follows:

1) $\text{ABEP}_{\text{signal}}$ is the summation of N_t addends $\text{ABEP}_{\text{MOD}}(\cdot)$. By direct inspection, we notice that each addend is the ABEP of a conventional modulation scheme whose points belong to the signal constellation diagram of SM and are transmitted only through the n_tth transmit antenna. Therefore, $\text{ABEP}_{\text{MOD}}(\cdot)$ depends only on the Euclidean distance of the points in the signal constellation diagram, and thus, $\text{ABEP}_{\text{signal}}$ can be regarded as the term that shows how the signal constellation diagram affects the performance of SM.
2) $\text{ABEP}_{\text{spatial}}$ is the summation of M addends $\text{ABEP}_{\text{SSK}}(\cdot)$. From, e.g., [24, eq. (35)], we observe that $\text{ABEP}_{\text{SSK}}(\cdot)$ is the ABEP of an equivalent SSK-MIMO scheme, where $\bar{\gamma}$ is replaced by $\bar{\gamma} \kappa_l^2$. Except for this scaling factor, $\text{ABEP}_{\text{SSK}}(\cdot)$ only depends on the Euclidean distance of the points in the spatial constellation diagram, and thus, $\text{ABEP}_{\text{spatial}}$ can be regarded as the term that shows how the spatial constellation diagram affects the performance of SM.
3) $\text{ABEP}_{\text{joint}}$ has a more complicated structure, and it depends on the Euclidean distance of points belonging to signal and spatial constellation diagrams. Thus, it is called "joint" because it shows how the interaction of these two non-orthogonal diagrams affects the ABEP of SM.

Finally, let us emphasize that 1) although Proposition 1 might seem like a simple and less compact rearrangement of (3), in Sections IV and V, we show that (6)–(8) allow us to get very simple and often closed-form expressions for specific modulation schemes and fading channels, and 2) unlike $\text{ABEP}_{\text{SSK}}(\cdot)$ and $\text{ABEP}_{\text{joint}}$, which are obtained through conventional union-bound methods, $\text{ABEP}_{\text{MOD}}(\cdot)$ is the exact error probability related to the signal constellation diagram. In other words, no union bound is used to compute this term. The exact computation of $\text{ABEP}_{\text{MOD}}(\cdot)$ avoids the inaccuracies of using the union-bound method for performance analysis of conventional modulation schemes, especially for large M and small N_r [34], [37]. For this reason, we call the framework in (6)–(8) improved union bound. The better accuracy of this new bound is substantiated in Section VI through Monte Carlo simulations. For the convenience of the reader, in Table I, we report the exact expression of $\text{ABEP}_{\text{MOD}}(\cdot)$ in (8) for PSK and QAM modulations. The formulas in Table I are useful for arbitrary fading channels and when Gray coding is used to map the information bits onto the signal constellation diagram.

IV. SIMPLIFIED EXPRESSIONS OF THE AVERAGE BIT ERROR PROBABILITY

Proposition 1 provides a very general framework to compute the ABEP for arbitrary fading channels and modulation schemes. By direct inspection, we notice that (6)–(8) can be computed in closed-form if the MGFs of the signal-to-noise ratios (SNRs) $\gamma(n_t)$, $\gamma_{(n_t, \tilde{n}_t)}$, and $\gamma_{(n_t, \chi_l, \tilde{n}_t, \chi_{\tilde{l}})}$ are available in closed-form. If so, the ABEP can be obtained through the

TABLE I

$\text{ABEP}_{\text{MOD}}(\cdot)$ OF PSK AND QAM MODULATIONS WITH MRC AT THE RECEIVER AND GRAY CODING. FOR QAM MODULATION, WE CONSIDER A GENERIC RECTANGULAR MODULATION SCHEME WITH $M = I_M \times J_M$. SQUARE-QAM MODULATION IS OBTAINED BY SETTING $I_M = J_M = \sqrt{M}$. THE MGF OF $\gamma(n_t) = \sum_{n_r=1}^{N_r} |h_{n_t,n_r}|^2$, $\mathcal{M}_{\gamma(n_t)}(\cdot)$, IS AVAILABLE IN CLOSED-FORM IN [34] FOR MANY CORRELATED FADING CHANNELS. NOTE THAT FADING CORRELATION AT THE TRANSMITTER DOES NOT AFFECT $\text{ABEP}_{\text{MOD}}(\cdot)$, BUT FADING CORRELATION AT THE RECEIVER DOES

Generic Fading Channels	
PSK [34, Eq.(8.29)] [38, Eq.(2), Eq.(7)]	$\begin{cases} \text{ABEP}_{\text{MOD}}(n_t) = \frac{1}{\log_2(M)} \sum_{l=1}^{M-1} \left[\left(2\left\lfloor \frac{l}{M} - \lfloor \frac{l}{M} \rfloor \right\rfloor + 2 \sum_{k=2}^{\log_2(M)} \left\lfloor \frac{l}{2^k} - \lfloor \frac{l}{2^k} \rfloor \right\rfloor \right) P_l(n_t) \right] \\ P_l(n_t) = \frac{1}{2\pi} \int_0^{\pi[1-(2l-1)/M]} T_l^-(\theta;n_t)\, d\theta - \frac{1}{2\pi} \int_0^{\pi[1-(2l+1)/M]} T_l^+(\theta;n_t)\, d\theta \\ T_l^-(\theta;n_t) = \mathcal{M}_{\gamma(n_t)}\left(2\bar{\gamma}\frac{\sin^2[\pi(2l-1)/M]}{\sin^2(\theta)} \right); \; T_l^+(\theta;n_t) = \mathcal{M}_{\gamma(n_t)}\left(2\bar{\gamma}\frac{\sin^2[\pi(2l+1)/M]}{\sin^2(\theta)} \right) \end{cases}$
QAM [34, Eq.(4.2)] [39, Eq.(22)]	$\begin{cases} \text{ABEP}_{\text{MOD}}(n_t) = \frac{1}{\log_2(M)} \left[\sum_{l=1}^{\log_2(I_M)} P_l(I_M;n_t) + \sum_{l=1}^{\log_2(J_M)} P_l(J_M;n_t) \right] \\ P_l(K;n_t) = \frac{2}{K} \sum_{k=0}^{(1-2^{-l})K-1} \left\{ (-1)^{\lfloor \frac{2^{l-1}k}{K} \rfloor} \left(2^{l-1} - \lfloor \frac{2^{l-1}k}{K} + \frac{1}{2} \rfloor \right) T_k(n_t) \right\} \\ T_k(n_t) = \frac{1}{\pi} \int_0^{\pi/2} \mathcal{M}_{\gamma(n_t)}\left(\bar{\gamma}\frac{6(2k+1)^2}{(I_M^2+J_M^2-2)\sin^2(\theta)} \right) d\theta \end{cases}$
i.i.d. Rayleigh Fading ($\text{ABEP}_{\text{MOD}}^{\text{Rayleigh}} = \text{ABEP}_{\text{MOD}}(n_t) = \text{ABEP}_{\text{MOD}}$)	
	$\mathcal{R}(\xi) = \left[\frac{1}{2}\left(1 - \sqrt{\frac{\xi}{2+\xi}} \right) \right]^{N_r} \sum_{n_r=0}^{N_r-1} \left\{ \binom{N_r+1-r}{r} \left[\frac{1}{2}\left(1 + \sqrt{\frac{\xi}{2+\xi}} \right) \right]^{n_r} \right\}$
PSK	$\begin{cases} P_l(n_t) = P_l = (1/2) I_{N_r}(c^-, \vartheta^-) - (1/2) I_{N_r}(c^+, \vartheta^+) \\ I_{N_r}(c^\pm, \vartheta^\pm) \text{ is available in } [34, \text{Eq.}(5A.24)] \\ c^\pm = 4\sigma_0^2\bar{\gamma}\sin^2[\pi(2l\pm1)/M]; \; \vartheta^\pm = \pi - \pi[(2l\pm1)/M] \end{cases}$
PSK approx. [34, Eq.(8.119))]	$\text{ABEP}_{\text{MOD}} \cong \frac{2}{\max\{\log_2(M),2\}} \sum_{k=1}^{\max\{M/4,1\}} \mathcal{R}\left(4\sigma_0^2\bar{\gamma}\sin^2\left[\frac{(2k-1)\pi}{M} \right] \right)$
QAM [34, Eq.(5A.4b)]	$T_k(n_t) = T_k = \mathcal{R}\left(\frac{24\sigma_0^2\bar{\gamma}(2k+1)^2}{I_M^2+J_M^2-2} \right)$
i.i.d. Rayleigh Fading – High–SNR ($\text{ABEP}_{\text{signal}} = [\log_2(M)/\log_2(N_tM)] \text{ABEP}_{\text{MOD}}^{\text{Rayleigh}}$)	
PSK [34, Eq.(8.119)] [40, Eq.(14.4.18)]	$\begin{cases} G_{\text{MOD}}^{\text{PSK}}(M) = \frac{2}{\max\{\log_2(M),2\}} \sum_{k=1}^{\max\{M/4,1\}} \left\{ \sin\left[\frac{(2k-1)\pi}{M} \right] \right\}^{-2N_r} \\ \text{ABEP}_{\text{MOD}}^{\text{Rayleigh}} \stackrel{\bar{\gamma}\gg1}{\approx} 2^{-2N_r} \binom{2N_r-1}{N_r} G_{\text{MOD}}^{\text{PSK}}(M) (4\sigma_0^2\bar{\gamma})^{-N_r} \end{cases}$
QAM [40, Eq.(14.4.18)]	$\begin{cases} G_{\text{MOD}}^{\text{QAM}}(K;k) = \frac{2}{K} \sum_{k=0}^{(1-2^{-l})K-1} \left\{ (-1)^{\lfloor \frac{2^{l-1}k}{K} \rfloor} \left(2^{l-1} - \lfloor \frac{2^{l-1}k}{K} + \frac{1}{2} \rfloor \right) (2k+1)^{-2N_r} \right\} \\ G_{\text{MOD}}^{\text{QAM}}(K) = \left[\frac{1}{\log_2(M)} \left(\frac{6}{I_M^2+J_M^2-2} \right)^{-N_r} \right] \sum_{l=1}^{\log_2(K)} G_{\text{MOD}}^{\text{QAM}}(K;k) \\ \text{ABEP}_{\text{MOD}}^{\text{Rayleigh}} \stackrel{\bar{\gamma}\gg1}{\approx} 2^{-N_r} \binom{2N_r-1}{N_r} \left[G_{\text{MOD}}^{\text{QAM}}(I_M) + G_{\text{MOD}}^{\text{QAM}}(J_M) \right] (4\sigma_0^2\bar{\gamma})^{-N_r} \end{cases}$

computation of simple single integrals and summations. More specifically, $\mathcal{M}_{\gamma(n_t)}(\cdot)$ is available in [34] for many correlated fading channels, which allows us to compute $\text{ABEP}_{\text{MOD}}(\cdot)$ and, eventually, $\text{ABEP}_{\text{signal}}$. On the other hand, the computation of $\mathcal{M}_{\gamma(n_t,\tilde{n}_t)}(\cdot)$ and $\mathcal{M}_{\gamma(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}(\cdot)$ deserves further attention as they are not available in the literature for arbitrary fading channels. Thus, the objective of this section is threefold:

1) to compute closed-form expressions of $\mathcal{M}_{\gamma(n_t,\tilde{n}_t)}(\cdot)$ and $\mathcal{M}_{\gamma(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}(\cdot)$ for the most common fading channel models;

2) to provide simplified formulas of the ABEP in (7) and (8) for specific modulation schemes and fading channels;

3) to analyze the obtained formulas to better understand SM.

To our best knowledge, and according to Section I, such a comprehensive study is not available in the literature.

A. Identically Distributed Fading at the Transmitter

Let us consider the scenario with identically distributed fading at the transmitter. We study uncorrelated and correlated fading, where in the latter case, the term "identically distributed"

means that all pairs of wireless links are equicorrelated. In formulas, this implies $\mathcal{M}_{\gamma(n_t)}(s) = \mathcal{M}_\gamma^{\mathrm{MOD}}(s)$, $\mathcal{M}_{\gamma(n_t,\tilde{n}_t)}(s) = \mathcal{M}_\gamma^{\mathrm{SSK}}(s)$, and $\mathcal{M}_{\gamma(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}(s) = \mathcal{M}_{\gamma(\chi_l,\chi_{\tilde{l}})}(s)$ for $n_t = 1, 2, \ldots, N_t$ and $\tilde{n}_t = 1, 2, \ldots, N_t$, which means that the MGFs are the same for each n_t or for each pair $(n_t, \tilde{n}_t)$. Accordingly, the ABEP in Proposition 1 can be simplified, as shown in Corollary 1.

Corollary 1: For identically distributed fading, (7) in Proposition 1 simplifies to (9), shown at the bottom of the page, where $\mathrm{ABEP}_{\mathrm{MOD}}$ is the error probability in Table I with $\mathcal{M}_{\gamma(n_t)}(s) = \mathcal{M}_\gamma^{\mathrm{MOD}}(s)$. If a constant-modulus modulation is considered, i.e., $\kappa_l = \kappa_0$ for $l = 1, 2, \ldots, M$, then $\mathrm{ABEP}_{\mathrm{spatial}}$ in (9) reduces to

$$\mathrm{ABEP}_{\mathrm{spatial}} = \frac{N_t}{2} \frac{\log_2(N_t)}{\log_2(N_t M)}$$
$$\times \left[\frac{1}{\pi} \int_0^{\pi/2} \mathcal{M}_\gamma^{\mathrm{SSK}} \left(\frac{\bar{\gamma}\kappa_0^2}{2\sin^2(\theta)} \right) d\theta \right]. \quad (10)$$

Likewise, if a constant-modulus modulation, i.e., $\kappa_l = \kappa_0$ for $l = 1, 2, \ldots, M$, and independent and uniformly distributed channel phases are considered, then $\mathrm{ABEP}_{\mathrm{joint}}$ in (9) simplifies to

$$\mathrm{ABEP}_{\mathrm{joint}} = \left[\frac{M(N_t-1)}{2} \frac{\log_2(M)}{\log_2(N_t M)} \right.$$
$$\left. + \frac{N_t(M-1)}{2} \frac{\log_2(N_t)}{\log_2(N_t M)} \right]$$
$$\times \left[\frac{1}{\pi} \int_0^{\pi/2} \mathcal{M}_\gamma^{\mathrm{SSK}} \left(\frac{\bar{\gamma}\kappa_0^2}{2\sin^2(\theta)} \right) d\theta \right]. \quad (11)$$

Proof: $\mathrm{ABEP}_{\mathrm{signal}}$ in (9) follows immediately from (7) by taking into account that for identically distributed fading, the M addends of the summation are all the same. $\mathrm{ABEP}_{\mathrm{spatial}}$ in (9) can be obtained by noticing the following: 1) for identically distributed fading, $\Psi_l(n_t, \tilde{n}_t)$ in (8) is the same for $n_t = 1, 2, \ldots, N_t$ and $\tilde{n}_t = 1, 2, \ldots, N_t$, and thus, it can be moved out of the twofold summation; and 2) $\sum_{n_t=1}^{N_t} \sum_{\tilde{n}_t=1}^{N_t} N_H(\tilde{n}_t \to n_t) = (N_t^2/2) \log_2(N_t)$ for any bit mapping. Finally, some algebraic manipulations lead to (9). Equation (10) follows from (9) with $\kappa_l = \kappa_0$ for $l = 1, 2, \ldots, M$. $\mathrm{ABEP}_{\mathrm{joint}}$ in (9) can be obtained as follows: 1) for identically distributed fading, $\Upsilon(n_t, \chi_l, \tilde{n}_t, \chi_{\tilde{l}})$ in (7) can be moved out of the twofold summa-

tion with indexes n_t and $\tilde{n}_t$ because it is the same for each pair $(n_t, \tilde{n}_t)$, and 2) $\sum_{n_t=1}^{N_t} \sum_{\tilde{n}_t \neq n_t=1}^{N_t} [N_H(\tilde{n}_t \to n_t) + N_H(\chi_{\tilde{l}} \to \chi_l)] = (N_t^2/2) \log_2(N_t) + N_t(N_t-1) N_H(\chi_{\tilde{l}} \to \chi_l)$ for any bit mapping. Finally, some simplifications lead to (9). Equation (11) can be obtained from (9) and the following considerations:

1) If the channel phases are uniformly distributed, then $\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})} = \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r}\chi_{\tilde{l}} - \alpha_{n_t,n_r}\chi_l|^2 \overset{d}{=} \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r}\kappa_{\tilde{l}} - \alpha_{n_t,n_r}\kappa_l|^2$. In fact, since adding a constant phase term to a uniformly distributed phase still yields a uniformly distributed phase, i.e., $(\varphi_{n_t,n_r} + \phi_l) \overset{d}{=} \varphi_{n_t,n_r}$, then $\alpha_{n_t,n_r}\chi_l = [\beta_{n_t,n_r} \exp(j\varphi_{n_t,n_r})] \times [\kappa_l \exp(j\phi_l)] = \beta_{n_t,n_r}\kappa_l \exp(j(\varphi_{n_t,n_r} + \phi_l)) \overset{d}{=} \beta_{n_t,n_r} \times \kappa_l \exp(j\varphi_{n_t,n_r}) = \alpha_{n_t,n_r}\kappa_l$.

2) If $\kappa_l = \kappa_0$ for $l = 1, 2, \ldots, M$, then $\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})} = \kappa_0^2 \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r} - \alpha_{n_t,n_r}|^2 = \kappa_0^2 \gamma_{(n_t,\tilde{n}_t)}$, which for identically distributed fading implies $\mathcal{M}_{\gamma(\chi_l,\chi_{\tilde{l}})}(s) = \mathcal{M}_\gamma^{\mathrm{SSK}}(\kappa_0^2 s)$. Thus, the integral in (9) can be replaced by the integral in (11).

3) For a constant-modulus modulation, the integral in (9) can be moved out of the twofold summation, which can be simplified using the identity $\sum_{l=1}^M \sum_{\tilde{l}=1}^M N_H(\chi_{\tilde{l}} \to \chi_l) = (M^2/2) \log_2(M)$ for any bit mapping. Finally, some algebraic manipulations lead to (11). This concludes the proof. ∎

Corollary 1 leads to two important considerations about the performance of SM: 1) $\mathrm{ABEP}_{\mathrm{signal}}$ and (9) show that, for identically distributed fading, the ABEP of SM is independent of the bit mapping of the spatial constellation diagram. This result is reasonable and agrees with intuition: If the channels are statistically identical, on average, the Euclidean distance of pairs of channel impulse responses is the same. In this case, the bit mapping has no role in determining the ABEP. On the other hand, the complex-valued points of the signal constellation diagram have different Euclidean distances, and this bit mapping plays an important role. 2) Under some realistic assumptions (i.e., constant-modulus modulation and uniform channel phases), $\mathrm{ABEP}_{\mathrm{joint}}$ in (9), which in the most general case depends on both spatial and signal constellation diagrams, depends only on the signal constellation diagram. Thus, since there are no terms in Table I, (10), and (11) that depend on both constellation diagrams, we conclude that they can be optimized individually. In particular, the best bit mapping for the signal constellation diagram turns out to be the conventional one based only on the Euclidean distance.

$$\begin{cases} \mathrm{ABEP}_{\mathrm{signal}} = \frac{\log_2(M)}{\log_2(N_t M)} \mathrm{ABEP}_{\mathrm{MOD}} \\ \mathrm{ABEP}_{\mathrm{spatial}} = \frac{1}{M} \frac{\log_2(N_t)}{\log_2(N_t M)} \frac{N_t}{2} \sum_{l=1}^M \left[\frac{1}{\pi} \int_0^{\pi/2} \mathcal{M}_\gamma^{\mathrm{SSK}} \left(\frac{\bar{\gamma}\kappa_l^2}{2\sin^2(\theta)} \right) d\theta \right] \\ \mathrm{ABEP}_{\mathrm{joint}} = \frac{1}{M} \frac{1}{\log_2(N_t M)} \sum_{l=1}^M \sum_{\tilde{l} \neq l=1}^M \left\{ \left[\frac{N_t \log_2(N_t)}{2} + N_H(\chi_{\tilde{l}} \to \chi_l)(N_t-1) \right] \left[\frac{1}{\pi} \int_0^{\pi/2} \mathcal{M}_{\gamma(\chi_l,\chi_{\tilde{l}})} \left(\frac{\bar{\gamma}}{2\sin^2(\theta)} \right) d\theta \right] \right\} \end{cases} \quad (9)$$

Finally, we notice that, e.g., (10) and (11) are very simple to compute and avoid the computation of fold summations on N_t and M. This is an important difference with respect to other frameworks available in the literature, where fourfold summations are always required, regardless of modulation scheme and channel model [6], [9]. In addition, Corollary 1 simplifies the frameworks in [24] and [26] for SSK modulation, as the twofold summation can be avoided for some fading channels and modulation schemes. Furthermore, we mention that Corollary 1 provides closed-form results if the MGFs, which depend on the specific fading channel model, are available in closed-form as well as if the related finite integral can be computed explicitly. In Section IV-B and C, we show some fading scenarios (Nakagami-m and Rice fading with arbitrary fading parameters and correlation) where the MGFs can be obtained in close-form. In addition, in Section V, we study the i.i.d. Rayleigh fading scenario where fold summations can be avoided and integrals can be computed in closed-form, thus leading to a very simple analytical framework for system analysis and optimization.

B. Nakagami-m Envelopes With Uniform Phases

In this section, the fading envelopes β_{n_t,n_r} are Nakagami-m RVs with fading severity $m_{\text{Nak}}^{(n_t,n_r)} = [\text{E}\{\beta_{n_t,n_r}^2\}]^2 / \text{E}\{[\beta_{n_t,n_r}^2 - \text{E}\{\beta_{n_t,n_r}^2\}]^2\}$ and mean square value $\Omega_{\text{Nak}}^{(n_t,n_r)} = \text{E}\{\beta_{n_t,n_r}^2\}$. We adopt the notation $\beta_{n_t,n_r} \sim \mathcal{N}(m_{\text{Nak}}^{(n_t,n_r)}, \Omega_{\text{Nak}}^{(n_t,n_r)})$. The amplitude correlation coefficient $\rho_{\text{Nak}}^{(n_t,n_r,\tilde{n}_t,\tilde{n}_r)}$ is defined as follows:

$$\rho_{\text{Nak}}^{(n_t,n_r,\tilde{n}_t,\tilde{n}_r)}$$
$$= \frac{\text{E}\{[\beta_{n_t,n_r} - \text{E}\{\beta_{n_t,n_r}\}][\beta_{\tilde{n}_t,\tilde{n}_r} - \text{E}\{\beta_{\tilde{n}_t,\tilde{n}_r}\}]\}}{\sqrt{\text{E}\{[\beta_{n_t,n_r} - \text{E}\{\beta_{n_t,n_r}\}]^2\}}\sqrt{\text{E}\{[\beta_{\tilde{n}_t,\tilde{n}_r} - \text{E}\{\beta_{\tilde{n}_t,\tilde{n}_r}\}]^2\}}}.$$
$$(12)$$

In addition, the channel phases φ_{n_t,n_r} are independent and uniformly distributed RVs in $[0, 2\pi)$. We adopt the notation $\varphi_{n_t,n_r} \sim \mathcal{U}(0, 2\pi)$. Finally, channel phases and fading envelopes are assumed to be independent.

Given this fading model, let us analyze and explicitly compute each term in (6).

1) ABEP$_{\text{signal}}$: $\mathcal{M}_{\gamma_{(n_t)}}(\cdot)$ has been widely studied in the literature, and closed-form expressions for non-identically distributed and arbitrary correlated Nakagami-m fading can be found in [34, Sec. 9.6.4].

2) ABEP$_{\text{spatial}}$: For Nakagami-m fading, $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot)$ is not well known in the literature and only recently it has been analyzed in [24] for single-antenna receivers, i.e., $N_r = 1$. Thus, we need to generalize [24] for our system model. Two case studies are considered: 1) correlated fading at the transmitter and independent fading at the receiver and 2) correlated fading at both ends of the MIMO channel. In the first case study, by exploiting the independence of the fading at the receiver, we have $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(s) = \prod_{n_r=1}^{N_r} \mathcal{M}_{\gamma_{(n_t,\tilde{n}_t;n_r)}}(s)$, where $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t;n_r)}}(\cdot)$ is the MGF of $\gamma_{(n_t,\tilde{n}_t;n_r)} = |\alpha_{\tilde{n}_t,n_r} - \alpha_{n_t,n_r}|^2$. This latter MGF is available in closed-form in [24, Sec. III] for generic correlated fading at the transmitter. The second case study is analytically more complicated as $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot)$ requires the computation of the expectation of $2N_r$ correlated RVs. Proposition 2 provides the final expression of $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot)$ for $N_r = 2$.

Proposition 2: Given $2N_r$ arbitrary distributed and correlated Nakagami-m RVs with fading envelopes (β_{n_t,n_r} and $\beta_{\tilde{n}_t,n_r}$) distributed according to the multivariate Nakagami-m pdf in [41, eq. (2)] and channel phases uniformly and i.i.d. in $[0, 2\pi)$, then $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot)$ for $N_r = 2$ is given in (13), shown at the bottom of the page, where

1) $\boldsymbol{\Sigma}$ is the $2N_r \times 2N_r$ correlation matrix of the Gaussian RVs associated with β_{n_t,n_r} and $\beta_{\tilde{n}_t,n_r}$, which can be computed from the amplitude correlation coefficients $\rho_{\text{Nak}}^{(n_t,n_r,\tilde{n}_t,\tilde{n}_r)}$ by using the method in [42];
2) $\boldsymbol{\Sigma}_{\text{trid}}$ is the tridiagonal approximation of $\boldsymbol{\Sigma}$, which can be obtained as described in [41, Sec. IV] and Appendix B;
3) $p_{ab} = \boldsymbol{\Sigma}_{\text{trid}}^{-1}(a, b)$ are the entries of $\boldsymbol{\Sigma}_{\text{trid}}^{-1}$;
4) $m_{\text{Nak}} = m_{\text{Nak}}^{(n_t,n_r)} = m_{\text{Nak}}^{(\tilde{n}_t,n_r)}$ is the fading parameter common to all links;
5) $\mathcal{F}_{\mathbf{k}}^{(p_{11},p_{33})}(\cdot)$ and $\mathcal{F}_{\mathbf{k}}^{(p_{22},p_{44})}(\cdot)$ are defined as

$$\begin{cases} \mathcal{F}_{\mathbf{k}}^{(p_{11},p_{33})}(s) = (1/4)s_{p_{33}}^{-(m_{\text{Nak}}+k_2+k_3)} \\ \quad \times G_{2,2}^{1,2}\left(-\frac{s^2}{s_{p_{11}}s_{p_{33}}}\middle| \begin{matrix} 1-m_{\text{Nak}}-k_2-k_3 & 1-m_{\text{Nak}}-k_1 \\ 0 & 0 \end{matrix}\right) \\ \mathcal{F}_{\mathbf{k}}^{(p_{22},p_{44})}(s) = (1/4)s_{p_{22}}^{-(m_{\text{Nak}}+k_1+k_2)} s_{p_{44}}^{-(m_{\text{Nak}}+k_3)} \\ \quad \times G_{2,2}^{1,2}\left(-\frac{s^2}{s_{p_{22}}s_{p_{44}}}\middle| \begin{matrix} 1-m_{\text{Nak}}-k_3 & 1-m_{\text{Nak}}-k_1-k_2 \\ 0 & 0 \end{matrix}\right) \end{cases}$$
$$(14)$$

where $s_p = s + (p/2)$.

Proof: See Appendix B. Formulas for $N_r > 2$ can be obtained as described in Appendix B. For arbitrary N_r, the final formula is given by the $(2N_r - 1)$-fold series of the product of N_r terms $\mathcal{F}_{\mathbf{k}}^{(\cdot,\cdot)}(\cdot)$. ∎

$$\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(s) = \frac{\left|\boldsymbol{\Sigma}_{\text{trid}}^{-1}\right|^{m_{\text{Nak}}}}{2^{(4m_{\text{Nak}}-4)}\Gamma(m_{\text{Nak}})}$$

$$\times \sum_{k_1=0}^{+\infty}\sum_{k_2=0}^{+\infty}\sum_{k_3=0}^{+\infty}\left[\left(\frac{1}{4}\right)^{k_1+k_2+k_3}\frac{\left(|p_{12}|^{2k_1}|p_{23}|^{2k_2}|p_{34}|^{2k_3}\right)\mathcal{F}_{\mathbf{k}}^{(p_{11},p_{33})}(s)\mathcal{F}_{\mathbf{k}}^{(p_{22},p_{44})}(s)}{(k_1!)(k_2!)(k_3!)\Gamma(k_1+m_{\text{Nak}})\Gamma(k_2+m_{\text{Nak}})\Gamma(k_3+m_{\text{Nak}})}\right] \quad (13)$$

It is worth mentioning that (14) gives an exact result when Σ is tridiagonal, i.e., $\Sigma = \Sigma_{\text{trid}}$. On the contrary, for arbitrary correlation and by using the Green method [41, Sec. IV], it provides a very tight approximation (see Section VI). Finally, we mention that the series in (14) converge very quickly thanks to the factorial and the Gamma functions in the denominator.

ABEP$_{\text{joint}}$: To compute $\mathcal{M}_{\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}}(\cdot)$, we need Proposition 3.

Proposition 3: For Nakagami-m fading envelopes and uniform phases, $\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}$ reduces to $\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})} = \gamma_{(n_t,\kappa_l,\tilde{n}_t,\kappa_{\tilde{l}})} = \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r}^{(\tilde{l})} - \alpha_{n_t,n_r}^{(l)}|^2$, where $\alpha_{n_t,n_r}^{(l)} \overset{d}{=} \beta_{n_t,n_r}^{(l)} \exp(j\varphi_{n_t,n_r})$, $\beta_{n_t,n_r}^{(l)} = \kappa_l \beta_{n_t,n_r}$, and $\beta_{n_t,n_r}^{(l)} \sim \mathcal{N}(m_{\text{Nak}}^{(n_t,n_r)}, \Omega_{\text{Nak}}^{(n_t,n_r;l)})$ with $\Omega_{\text{Nak}}^{(n_t,n_r;l)} = \kappa_l^2 \Omega_{\text{Nak}}^{(n_t,n_r)}$.

Proof: The equality in law $\alpha_{n_t,n_r}^{(l)} \overset{d}{=} \beta_{n_t,n_r}^{(l)} \exp(j\varphi_{n_t,n_r})$ can be obtained by using the same analytical development used for (11) in Corollary 1, and more specifically, the identity in law $(\varphi_{n_t,n_r} + \phi_l) \overset{d}{=} \varphi_{n_t,n_r}$. This concludes the proof. ∎

Proposition 3 points out that $\gamma_{(n_t,\tilde{n}_t)}$ and $\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}$ are related by a scaling factor in the mean power of each channel envelope, i.e., $\Omega_{\text{Nak}}^{(n_t,n_r;l)} / \Omega_{\text{Nak}}^{(n_t,n_r)} = \kappa_l^2$. Accordingly, $\mathcal{M}_{\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}}(\cdot)$ can be computed by using the same frameworks used to compute $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot)$. In other words, for arbitrary correlation, $\mathcal{M}_{\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}}(\cdot)$ is still given by (13) but with a different correlation matrix Σ.

Diversity Analysis: The accurate analysis of ABEP$_{\text{signal}}$, ABEP$_{\text{spatial}}$, and ABEP$_{\text{joint}}$ through closed-form expressions of the MGFs allows us to provide important considerations about the diversity gain [43] of SM in Nakagami-m fading as well as to understand the constellation diagram that dominates the performance of SM for high SNR. The main result is given in Proposition 4.

Proposition 4: Let us assume, for the sake of simplicity, that $m_{\text{Nak}} = m_{\text{Nak}}^{(n_t,n_r)} = m_{\text{Nak}}^{(\tilde{n}_t,n_r)}$ for each wireless link. The diversity gain Div$_{\text{SM}}$ of SM is equal to Div$_{\text{SM}} = \min\{N_r, m_{\text{Nak}} N_r\}$.

Proof: From (6), we have Div$_{\text{SM}} = \min\{\text{Div}_{\text{signal}}, \text{Div}_{\text{spatial}}, \text{Div}_{\text{joint}}\}$, where Div$_{\text{signal}}$, Div$_{\text{spatial}}$, and Div$_{\text{joint}}$ are the diversity gains of ABEP$_{\text{signal}}$, ABEP$_{\text{spatial}}$, and ABEP$_{\text{joint}}$, respectively. In fact, for high SNR, the worst term dominates the slope of the ABEP and thus the diversity gain [43]. From Section IV-B1 and [34, Sec. 9.6.4], it follows that Div$_{\text{signal}} = m_{\text{Nak}} N_r$. From Section IV-B2 and B3, [24], and [44], it follows that Div$_{\text{spatial}} = \text{Div}_{\text{joint}} = N_r$. In fact, as analytically shown in [44], $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t;n_r)}}(\cdot)$ and $\mathcal{F}_{\mathbf{k}}^{(\cdot,\cdot)}(\cdot)$ have unit diversity gain regardless of the fading severity m_{Nak}, and thus, the diversity is determined only by the number N_r of antennas at the receiver. This concludes the proof. ∎

Proposition 4 unveils important properties of SM and provides information about the best scenarios where SM should be used. More specifically, we have the following:

1) In scenarios with less severe fading than Rayleigh, i.e., $m_{\text{Nak}} > 1$, we have Div$_{\text{SM}} = \text{Div}_{\text{spatial}} = \text{Div}_{\text{joint}} = N_r$. We conclude that the ABEP is mainly determined by the spatial constellation diagram (i.e., ABEP$_{\text{spatial}} \gg$

ABEP$_{\text{signal}}$ and ABEP$_{\text{joint}} \gg$ ABEP$_{\text{signal}}$), and that the diversity gain is independent of fading severity.

2) In scenarios with more severe fading than Rayleigh, i.e., $0.5 \leq m_{\text{Nak}} < 1$, we have Div$_{\text{SM}} = \text{Div}_{\text{signal}} = m_{\text{Nak}} N_r$. We conclude that the ABEP is mainly determined by the signal constellation diagram (i.e., ABEP$_{\text{signal}} \gg$ ABEP$_{\text{spatial}}$ and ABEP$_{\text{signal}} \gg$ ABEP$_{\text{joint}}$), and that the diversity gain strongly depends on fading severity.

3) Due to the increasing diversity gain of ABEP$_{\text{signal}}$ with m_{Nak} [34], it is expected that ABEP$_{\text{signal}}$ provides a negligible contribution for increasing m_{Nak}, and that the ABEP gets better with m_{Nak}. This behavior is similar to conventional modulations [34] but different from SSK [24], [26].

4) From [34], it is known that conventional single-antenna systems have the same diversity gain as ABEP$_{\text{signal}}$, i.e., Div$_{\text{conventional}} = \text{Div}_{\text{signal}} = m_{\text{Nak}} N_r$. Thus, with respect to conventional modulations, the performance gain of SM is expected to increase for $0.5 \leq m_{\text{Nak}} < 1$, whereas it is expected to decrease for $m_{\text{Nak}} > 1$. This conclusion agrees with intuition since SM encodes part of the information bits onto the spatial constellation diagram whose points are more closely spaced if $m_{\text{Nak}} > 1$.

5) From [24] and [44], it is known that the diversity gain of SSK modulation is the same as ABEP$_{\text{spatial}}$, i.e., Div$_{\text{SSK}} = \text{Div}_{\text{spatial}} = N_r$, which is independent of m_{Nak}. Thus, unlike conventional modulation schemes and SM, SSK modulation does not experience any diversity reduction when $0.5 \leq m_{\text{Nak}} < 1$, and it can be concluded that, thanks to the higher diversity gain, it turns out to be, among SM and conventional modulations, the best transmission scheme in scenarios with fading less severe than Rayleigh. The price to be paid is the need of many antennas at the transmitter to achieve the same rate. On the contrary, in more severe fading channels, SSK modulation turns out to be worse than conventional modulation.

In conclusion, the performance of SM in Nakagami-m fading strongly depends on m_{Nak}, and there is no clear transmission technology better than others for any m_{Nak}. This important result suggests the adoption of a multimode adaptive transmission scheme, which can switch among the best modulations according to the fading severity and the desired rate.

Finally, we close this section with the following corollary.

Corollary 2: For Nakagami-m fading envelopes, uniform channel phases, and a constant-modulus modulation, i.e., $\kappa_l = \kappa_0$ for $l = 1, 2, \ldots, M$, ABEP$_{\text{spatial}}$ and ABEP$_{\text{joint}}$ in (7) can be simplified as in (15), shown at the bottom of the next page.

Proof: Equation (15) can be obtained through analytical steps similar to (10) and (11) but without the assumption of identically distributed fading. ∎

Corollary 2 shows that, for a constant-modulus modulation, ABEP$_{\text{spatial}}$ and ABEP$_{\text{joint}}$ can be computed only through $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot) = \mathcal{M}_{\gamma}^{\text{SSK}}(\cdot)$. This makes even more evident the connection established between $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot)$ and $\mathcal{M}_{\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}}(\cdot)$ in Proposition 3. We note that in (15), neither ABEP$_{\text{spatial}}$ nor ABEP$_{\text{joint}}$ depend on the bit mapping used for

the signal constellation diagram. Thus, the optimality of usual bit mappings adopted for $\text{ABEP}_{\text{signal}}$ seems to be preserved.

C. Rician Fading

Let us consider a generic Rician fading [26], [45]. In this case, α_{n_t,n_r} are generically correlated complex Gaussian RVs, and α_{n_t,n_r}^R and α_{n_t,n_r}^I are independent by definition. We adopt the notation $\mu_{n_t,n_r}^R = \text{E}\{\alpha_{n_t,n_r}^R\}$, $\mu_{n_t,n_r}^I = \text{E}\{\alpha_{n_t,n_r}^I\}$, and $\sigma_{n_t,n_r}^2 = \text{E}\{(\alpha_{n_t,n_r}^R - \mu_{n_t,n_r}^R)^2\} = \text{E}\{(\alpha_{n_t,n_r}^I - \mu_{n_t,n_r}^I)^2\}$. In addition, we use the shorthands $\alpha_{n_t,n_r}^R \sim \mathcal{G}(\mu_{n_t,n_r}^R, \sigma_{n_t,n_r}^2)$ and $\alpha_{n_t,n_r}^I \sim \mathcal{G}(\mu_{n_t,n_r}^I, \sigma_{n_t,n_r}^2)$.

The analysis of Rician fading is simpler than Nakagami-m fading, and a unified framework can be used to compute (6). The main enabling result is summarized in Proposition 5.

Proposition 5: Given a complex Gaussian RV α_{n_t,n_r}, then $\alpha_{n_t,n_r}^{(l)} = \chi_l \alpha_{n_t,n_r}$ is still a complex Gaussian RV such that $\text{Re}\{\alpha_{n_t,n_r}^{(l)}\} \sim \mathcal{G}(\mu_{n_t,n_r,l}^R, \kappa_l^2 \sigma_{n_t,n_r}^2)$, $\text{Im}\{\alpha_{n_t,n_r}^{(l)}\} \sim \mathcal{G}(\mu_{n_t,n_r,l}^I, \kappa_l^2 \sigma_{n_t,n_r}^2)$, with $\mu_{n_t,n_r,l}^R = \mu_{n_t,n_r}^R \kappa_l \cos(\phi_l) - \mu_{n_t,n_r}^I \kappa_l \sin(\phi_l)$ and $\mu_{n_t,n_r,l}^I = \mu_{n_t,n_r}^I \kappa_l \cos(\phi_l) + \mu_{n_t,n_r}^R \times \kappa_l \sin(\phi_l)$. In addition, $\text{Re}\{\alpha_{n_t,n_r}^{(l)}\}$ and $\text{Im}\{\alpha_{n_t,n_r}^{(l)}\}$ are independent RVs.

Proof: By definition, $\alpha_{n_t,n_r}^{(l)} = \kappa_l \exp(j\phi_l)(\alpha_{n_t,n_r}^R + j\alpha_{n_t,n_r}^I) = \text{Re}\{\alpha_{n_t,n_r}^{(l)}\} + j\text{Im}\{\alpha_{n_t,n_r}^{(l)}\}$ with $\text{Re}\{\alpha_{n_t,n_r}^{(l)}\} = \kappa_l \alpha_{n_t,n_r}^R \cos(\phi_l) - \kappa_l \alpha_{n_t,n_r}^I \sin(\phi_l)$ and $\text{Im}\{\alpha_{n_t,n_r}^{(l)}\} = \kappa_l \times \alpha_{n_t,n_r}^I \cos(\phi_l) + \kappa_l \alpha_{n_t,n_r}^R \sin(\phi_l)$. Then, by taking into account that α_{n_t,n_r}^R and α_{n_t,n_r}^I are Gaussian distributed, independent, and have the same variance, simple algebraic manipulations conclude the proof. ∎

From Proposition 5, we conclude that $\gamma_{(n_t)}$, $\gamma_{(n_t,\tilde{n}_t)}$, and $\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}$ in Table I and (6)–(8) are all given by the summation of N_r square envelopes of arbitrary distributed and correlated Gaussian RVs. Thus, the related MGF can be computed by using the so-called Moschopoulos method [45]. More specifically, we have the following.

1) $\mathcal{M}_{\gamma_{(n_t)}}(\cdot)$ can be obtained from [45, eq. (25)].
2) $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot)$ can be found in [26, eq. (15)].
3) $\mathcal{M}_{\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}}(\cdot)$ can still be obtained from [26, eq. (15)] thanks to Proposition 5.

The only differences between $\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(\cdot)$ and $\mathcal{M}_{\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})}}(\cdot)$ are the parameters of each Gaussian RV, which, however, can be related to one another, as shown in Proposition 5. The same applies to the covariance matrices.

1) Diversity Analysis: Diversity can be studied by using the Moschopoulos method. In fact, [45, Sec. IV-B] and [26, Sec. III-C] show that each ABEP term in (6) has diversity gain equal to N_r, i.e., $\text{Div}_{\text{SM}} = N_r$. Thus, unlike Nakagami-$m$ fading, in Rician fading, the ABEPs in (6) have the same slope. However, [26, Sec. V] has pointed out that $\text{ABEP}_{\text{spatial}}$ and $\text{ABEP}_{\text{signal}}$ have opposite behavior with the Rician factor: $\text{ABEP}_{\text{spatial}}$ increases and $\text{ABEP}_{\text{signal}}$ decreases when the Rician factor increases, respectively. Therefore, the ABEPs in (6) have different coding gains depending on the fading severity.

The Moschopoulos framework can be simplified for some fading channels, as shown in Corollary 3.

Corollary 3: For Rician fading, a constant-modulus modulation, and zero-mean fading, i.e., $\mu_{n_t,n_r}^R = \mu_{n_t,n_r}^I = 0$, then $\text{ABEP}_{\text{spatial}}$ and $\text{ABEP}_{\text{joint}}$ in (7) can be simplified as shown in (15).

Proof: For zero-mean fading and a constant-modulus modulation, $\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})} \overset{d}{=} \kappa_0^2 \gamma_{(n_t,\tilde{n}_t)}$ because $\mu_{n_t,n_r,l}^R$ and $\mu_{n_t,n_r,l}^I$ are independent of ϕ_l. Then, considerations similar to Corollary 2 lead to (15). ∎

We emphasize that, although Corollaries 2 and 3 provide the same result, the assumptions are different. In Nakagami-m fading, the channel phases need to be uniformly distributed. On the other hand, in Rician fading, the complex channel gains need to have zero mean (i.e., zero Rician factor). Furthermore, it should be noted that although Rayleigh fading is a special case of either Rician or Nakagami-m fading, for correlated channels, exploiting the framework for Rician fading leads to a more straightforward analytical derivation.

V. AVERAGE BIT ERROR PROBABILITY OVER INDEPENDENT AND IDENTICALLY DISTRIBUTED RAYLEIGH FADING

In this section, we study the canonical i.i.d. Rayleigh fading scenario. Our contribution is threefold:

1) i.i.d. Rayleigh fading has already been studied in [6]. However, [6] is useful only for real-valued signal constellation points, whereas our framework is simple and applicable to generic signal constellation diagrams. In addition, we provide asymptotically tight bounds that highlight fundamental properties of SM.
2) i.i.d. Rayleigh fading is a special case of either Rician or Nakagami-m fading [34]. We show how the integrals in (6) can be computed in closed-form.
3) Closed-form expressions of the SNR difference between SM and other similar transmission technologies are provided. This allows us to understand the best transmission technology to use for every MIMO setup and data rate. To our best knowledge, these contributions make this section novel and important to understand the achievable performance of SM.

$$\begin{cases} \text{ABEP}_{\text{spatial}} = \frac{1}{N_t \log_2(N_t M)} \sum_{n_t=1}^{N_t} \sum_{\tilde{n}_t=1}^{N_t} \left[\frac{1}{\pi} \int_0^{\pi/2} \mathcal{M}_\gamma^{\text{SSK}}\left(\frac{\bar{\gamma}\kappa_0^2}{2\sin^2(\theta)}\right) d\theta \right] \\ \text{ABEP}_{\text{joint}} = \frac{1}{N_t \log_2(N_t M)} \sum_{n_t=1}^{N_t} \sum_{\tilde{n}_t \neq n_t=1}^{N_t} \left\{ \left[\frac{M \log_2(M)}{2} + N_H(\tilde{n}_t \to n_t)(M-1) \right] \left[\frac{1}{\pi} \int_0^{\pi/2} \mathcal{M}_\gamma^{\text{SSK}}\left(\frac{\bar{\gamma}\kappa_0^2}{2\sin^2(\theta)}\right) d\theta \right] \right\} \end{cases} \quad (15)$$

Let us consider the channel model in Section IV-C, which for i.i.d. Rayleigh fading reduces to $\mu_{n_t,n_r}^R = \mu_{n_t,n_r}^I = 0$ and $\sigma_{n_t,n_r}^2 = \sigma_0^2$. Corollary 4 summarizes the ABEP of SM over i.i.d. Rayleigh fading.

Corollary 4: For i.i.d. Rayleigh fading, the ABEP in (6) reduces to (16), shown at the bottom of the page, where $\text{ABEP}_{\text{MOD}}^{\text{Rayleigh}}$ and $\mathcal{R}(\cdot)$ are defined in Table I. Furthermore, for a constant-modulus modulation, $\text{ABEP}_{\text{spatial}}$ and $\text{ABEP}_{\text{joint}}$ simplify as

$$\begin{cases} \text{ABEP}_{\text{spatial}} = \left[\frac{N_t}{2} \frac{\log_2(N_t)}{\log_2(N_t M)} \right] \mathcal{R}\left(4\sigma_0^2 \bar{\gamma} \kappa_0^2\right) \\ \text{ABEP}_{\text{joint}} = \left[\frac{N_t(M-1)}{2} \frac{\log_2(N_t)}{\log_2(N_t M)} + \frac{M(N_t-1)}{2} \frac{\log_2(M)}{\log_2(N_t M)} \right] \\ \qquad\qquad \times \mathcal{R}\left(4\sigma_0^2 \bar{\gamma} \kappa_0^2\right). \end{cases}$$

(17)

Proof: $\text{ABEP}_{\text{signal}}$ follows from Corollary 1 by using the known results summarized in Table I. $\text{ABEP}_{\text{spatial}}$ in (16) can be obtained from Corollary 1 with $\mathcal{M}_\gamma^{\text{SSK}}(s) = (1 + 4\sigma_0^2 s)^{-N_r}$ [34, eq. (2.8)] and by computing the related integral with [34, eq. (5A.4b)]. $\text{ABEP}_{\text{joint}}$ in (16) can be computed from Corollary 1, i.e., $\gamma_{(n_t,\chi_l,\tilde{n}_t,\chi_{\tilde{l}})} = \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r} \chi_{\tilde{l}} - \alpha_{n_t,n_r} \chi_l|^2 \overset{d}{=} \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r} \kappa_{\tilde{l}} - \alpha_{n_t,n_r} \kappa_l|^2$, which for i.i.d. Rayleigh fading leads to $\mathcal{M}_{\gamma(\chi_l,\chi_{\tilde{l}})}(s) = (1 + 2\sigma_0^2(\kappa_l^2 + \kappa_{\tilde{l}}^2)s)^{-N_r}$. The final integral can be computed using [34, Eq. (5A.4b)]. Finally, (17) follows from (16) with $\kappa_l = \kappa_0$ for $l = 1, 2, \ldots, M$, $\sum_{l=1}^{M} \sum_{\tilde{l}=1}^{M} N_H(\chi_{\tilde{l}} \to \chi_l) = (M^2/2)\log_2(M)$, and simple algebraic manipulations. This concludes the proof. ∎

The formulas in (16) and (17) provide important considerations about the performance of SM. For example, (16) shows that, regardless of the signal constellation diagram, the bit mapping on the spatial constellation diagram has no influence on the performance of SM. On the other hand, the bit mapping on the signal constellation diagram plays an important role in $\text{ABEP}_{\text{joint}}$. In particular, while conventional bit mappings (e.g., Gray coding) based on the Euclidean distance of the signal constellation points turn out to be optimal to minimize $\text{ABEP}_{\text{signal}}$, additional constraints might be introduced on the optimal choice of the signal constellation diagram and on the related bit mapping to minimize $\text{ABEP}_{\text{joint}}$ (see Corollary 5 as well). On the other hand, for a constant-modulus modulation,

we notice that $\text{ABEP}_{\text{joint}}$ is independent of the properties of the signal constellation diagram and only depends on its cardinality M. Thus, the optimization criterion based on the Euclidean distance, which is optimal for $\text{ABEP}_{\text{signal}}$, turns out to be optimal for the overall ABEP. Finally, we remark that $\text{ABEP}_{\text{spatial}}$ and $\text{ABEP}_{\text{joint}}$ are independent of the phases of the complex points of the signal constellation diagram. Only the moduli of these points play a role. This result suggests that, to minimize $\text{ABEP}_{\text{spatial}}$ and $\text{ABEP}_{\text{joint}}$, we can focus our attention only on the moduli and neglect the phases.

To enable a deeper understanding of the achievable performance and a simpler comparison with other transmission technologies, in Corollary 5, we provide a tight high-SNR approximation of (16) and (17).

2) Corollary 5: For high-SNR, $\text{ABEP}_{\text{spatial}}$ and $\text{ABEP}_{\text{joint}}$ in (16) can be simplified as in (18), shown at the bottom of the page, where $\Theta_{\text{joint}}^{(M,N_r,H)} = \sum_{l=1}^{M} \sum_{\tilde{l}=1}^{M} [N_H(\chi_{\tilde{l}} \to \chi_l) \times (\kappa_l^2 + \kappa_{\tilde{l}}^2)^{-N_r}]$, $\Theta_{\text{joint}}^{(M,N_r)} = \sum_{l=1}^{M} \sum_{\tilde{l}\neq l=1}^{M} (\kappa_l^2 + \kappa_{\tilde{l}}^2)^{-N_r}$, and $\Theta_{\text{spatial}}^{(M,N_r)} = \sum_{l=1}^{M} \kappa_l^{-2N_r}$. For a constant-modulus modulation, they simplify as follows: $\Theta_{\text{joint}}^{(M,N_r,H)} = (2\kappa_0^2)^{-N_r}(M^2/2)\log_2(M)$, $\Theta_{\text{joint}}^{(M,N_r)} = M(M-1)(2\kappa_0^2)^{-N_r}$, and $\Theta_{\text{spatial}}^{(M,N_r)} = M\kappa_0^{-2N_r}$. Formulas for $\text{ABEP}_{\text{signal}}$ can be found in Table I.

Proof: Equation (18) follows from $\mathcal{R}(\xi) \overset{\bar{\gamma}\gg1}{=} 2^{-N_r} \times \binom{2N_r-1}{N_r} \xi^{-N_r}$ [40, eq. (14.4.18)] and some algebra. ∎

The high-SNR framework in (18) is simple and accurate and sheds light on the performance of SM. By using [43], it enables us to compute coding and diversity gains. In particular, the diversity gain is N_r, whereas the coding gain depends on the MIMO setup, i.e., N_t, M, and the spatial constellation diagram. The impact of the signal constellation diagram comes into play only through $\Theta_{\text{spatial}}^{(M,N_r)}$, $\Theta_{\text{joint}}^{(M,N_r)}$, and $\Theta_{\text{joint}}^{(M,N_r,H)}$. More specifically, (18) provides the criterion to choose the points of the signal constellation diagram, i.e., the moduli κ_l^2 that minimize the ABEP: $\Theta_{\text{spatial}}^{(M,N_r)}$, $\Theta_{\text{joint}}^{(M,N_r)}$, and $\Theta_{\text{joint}}^{(M,N_r,H)}$ should be kept as small as possible for a given average energy constraint. Thus, the Euclidean distance criterion used for $\text{ABEP}_{\text{signal}}$ along with the minimization of the coefficients above give the cost functions that need to be jointly considered to optimize the performance of SM. In Section VI, we show the very

$$\begin{cases} \text{ABEP}_{\text{signal}} = \frac{\log_2(M)}{\log_2(N_t M)} \text{ABEP}_{\text{MOD}}^{\text{Rayleigh}} \\ \text{ABEP}_{\text{spatial}} = \frac{1}{M} \frac{\log_2(N_t)}{\log_2(N_t M)} \frac{N_t}{2} \sum_{l=1}^{M} \mathcal{R}\left(4\sigma_0^2 \bar{\gamma} \kappa_l^2\right) \\ \text{ABEP}_{\text{joint}} = \frac{1}{M} \frac{1}{\log_2(N_t M)} \sum_{l=1}^{M} \sum_{\tilde{l}\neq l=1}^{M} \left[\left(\frac{N_t \log_2(N_t)}{2} + (N_t - 1)N_H(\chi_{\tilde{l}} \to \chi_l) \right) \mathcal{R}\left(2\sigma_0^2 \bar{\gamma}\left(\kappa_l^2 + \kappa_{\tilde{l}}^2\right)\right) \right] \end{cases}$$

(16)

$$\begin{cases} \text{ABEP}_{\text{spatial}} \overset{\bar{\gamma}\gg1}{=} \left[\frac{N_t}{2} \frac{1}{M} \frac{\log_2(N_t)}{\log_2(N_t M)} 2^{-N_r} \binom{2N_r-1}{N_r} \Theta_{\text{spatial}}^{(M,N_r)} \right] \left(4\sigma_0^2 \bar{\gamma}\right)^{-N_r} \\ \text{ABEP}_{\text{joint}} \overset{\bar{\gamma}\gg1}{=} \left[\frac{N_t}{2} \frac{1}{M} \frac{\log_2(N_t)}{\log_2(N_t M)} \binom{2N_r-1}{N_r} \Theta_{\text{joint}}^{(M,N_r)} + \frac{N_t-1}{M} \frac{1}{\log_2(N_t M)} \binom{2N_r-1}{N_r} \Theta_{\text{joint}}^{(M,N_r,H)} \right] \left(4\sigma_0^2 \bar{\gamma}\right)^{-N_r} \end{cases}$$

(18)

TABLE II
SNR Difference (in Decibels) Between Transmission Technology X and Y. PSK and QAM Denote Single-Antenna Schemes With Gray Coding. SM-PSK and SM-QAM Denote SM With PSK and QAM Modulation With Gray Coding in the Signal Constellation Diagram. (M, N_t) Is Referred to SM, With $M = I_M \times J_M$ for SM-QAM. M^{PSK} and $M^{\mathrm{QAM}} = I_M^{\mathrm{QAM}} \times J_M^{\mathrm{QAM}}$ Are Referred to PSK and QAM, Respectively. N_t^{SSK} Is Referred to SSK. The Comparison Is Made by Considering the Same Data Rate R for Each Transmission Technology, Which Implies $\log_2(N_t M) = \log_2(M^{\mathrm{QAM}}) = \log_2(M^{\mathrm{PSK}}) = \log_2(N_t^{\mathrm{SSK}})$. Other Symbols Are Defined in Table I

$$\Delta_{\mathrm{SNR}}^{(X/Y)} = 10 \log_{10}\left(\mathrm{SNR}_X / \mathrm{SNR}_Y\right) = -(10/N_r) \log_{10}\left(\Pi_{\mathrm{SNR}}^{(X/Y)}\right)$$

$$\Pi_{\mathrm{SNR}}^{(\mathrm{PSK/SM-PSK})} = \frac{\left[\frac{N_t M}{2} \frac{\log_2(N_t)}{\log_2(N_t M)} + 2^{-N_r} \frac{\log_2(M)}{\log_2(N_t M)} G_{\mathrm{MOD}}^{\mathrm{PSK}}\left(M^{\mathrm{PSK}}\right) + \frac{(N_t-1)M}{2} \frac{\log_2(M)}{\log_2(N_t M)}\right]}{2^{-N_r} G_{\mathrm{MOD}}^{\mathrm{PSK}}(M^{\mathrm{PSK}})}$$

$$\Pi_{\mathrm{SNR}}^{(\mathrm{QAM/SM-QAM})} = \frac{\frac{1}{M}\left[\frac{N_t}{2}\log_2(N_t)\Theta_{\mathrm{spatial}}^{(M,N_r)} + \frac{N_t}{2} 2^{N_r} \log_2(N_t) \Theta_{\mathrm{joint}}^{(M,N_r)} + (N_t-1)2^{N_r}\Theta_{\mathrm{joint}}^{(M,N_r,H)}\right] + \log_2(M)\left[G_{\mathrm{MOD}}^{\mathrm{QAM}}(I_M) + G_{\mathrm{MOD}}^{\mathrm{QAM}}(J_M)\right]}{\log_2(M^{\mathrm{QAM}})\left[G_{\mathrm{MOD}}^{\mathrm{QAM}}(I_M^{\mathrm{QAM}}) + G_{\mathrm{MOD}}^{\mathrm{QAM}}(J_M^{\mathrm{QAM}})\right]}$$

$$\Pi_{\mathrm{SNR}}^{(\mathrm{SSK/SM-QAM})} = \frac{\frac{1}{M}\left[\frac{N_t}{2}\log_2(N_t)\Theta_{\mathrm{spatial}}^{(M,N_r)} + \frac{N_t}{2} 2^{N_r} \log_2(N_t) \Theta_{\mathrm{joint}}^{(M,N_r)} + (N_t-1)2^{N_r}\Theta_{\mathrm{joint}}^{(M,N_r,H)}\right] + \log_2(M)\left[G_{\mathrm{MOD}}^{\mathrm{QAM}}(I_M) + G_{\mathrm{MOD}}^{\mathrm{QAM}}(J_M)\right]}{\frac{N_t^{\mathrm{SSK}}}{2}\log_2\left(N_t^{\mathrm{SSK}}\right)}$$

$$\Pi_{\mathrm{SNR}}^{(\mathrm{SSK/QAM})} = \frac{G_{\mathrm{MOD}}^{\mathrm{QAM}}(I_M^{\mathrm{QAM}}) + G_{\mathrm{MOD}}^{\mathrm{QAM}}(J_M^{\mathrm{QAM}})}{\frac{N_t^{\mathrm{SSK}}}{2}\log_2\left(N_t^{\mathrm{SSK}}\right)}$$

$$\Pi_{\mathrm{SNR}}^{(\mathrm{SSK/PSK})} = \frac{2^{-N_r} G_{\mathrm{MOD}}^{\mathrm{PSK}}\left(M^{\mathrm{PSK}}\right)}{\frac{N_t^{\mathrm{SSK}}}{2}}$$

interesting, and apparently unexpected, result that, for some MIMO setups and data rates, SM with PSK-modulated points (SM-PSK) outperforms SM with QAM-modulated points (SM-QAM) for the same average energy constraint. On the other hand, it is well known that $\mathrm{ABEP}_{\mathrm{signal}}$ with QAM modulation is never worse than $\mathrm{ABEP}_{\mathrm{signal}}$ with PSK modulation. This result can be well understood with the help of (18): unlike PSK, QAM has points with moduli that can be either smaller or larger than 1, which has an impact on $\Theta_{\mathrm{spatial}}^{(M,N_r)}$, $\Theta_{\mathrm{joint}}^{(M,N_r)}$, and $\Theta_{\mathrm{joint}}^{(M,N_r,H)}$. Since the ABEP of SM is a weighted summation of all these terms, it turns out that SM-PSK may outperform SM-QAM. This leads to two important conclusions: 1) The best modulation scheme (between PSK and QAM) to use depends on M and N_t for a given data rate, and 2) neither PSK nor QAM seem to be optimal signal modulation schemes for SM. However, (18) provides the criterion to compute the optimal modulation scheme that minimizes the ABEP.

A. Comparison With Single-Antenna PSK/QAM and SSK Modulations

By exploiting Corollary 5, in this section, we aim at computing in closed-form the SNR difference between SM and other transmission technologies with similar complexity, such as single-antenna PSK/QAM and SSK modulations. The high-SNR framework for PSK/QAM can be found in Table I, whereas for SSK, we get $\mathrm{ABEP}_{\mathrm{SSK}} = (N_t/2)\mathcal{R}(4\sigma_0^2\bar\gamma) \overset{\bar\gamma \gg 1}{=} 2^{-N_r}(N_t/2)\binom{2N_r-1}{N_r}(4\sigma_0^2\bar\gamma)^{-N_r}$ from $\mathrm{ABEP}_{\mathrm{spatial}}$ in (18).

Due to space constraints, we cannot report all the details of the analytical derivation, but we can only summarize the main procedure used to compute the formulas in Table II. From (18) and Table I, for any transmission technology X, the ABEP is $\mathrm{ABEP}_X = K_X(\sigma_0^2\bar\gamma_X)^{-N_r} = K_X(\mathrm{SNR}_X)^{-N_r}$. Then, for any pair ABEP_X and ABEP_Y, we have $\mathrm{ABEP}_X = \mathrm{ABEP}_Y \Rightarrow$ $K_X(\mathrm{SNR}_X)^{-N_r} = K_Y(\mathrm{SNR}_Y)^{-N_r}$. If we define the SNR difference (in decibels) as $\Delta_{\mathrm{SNR}}^{(X/Y)} = 10 \log_{10}(\mathrm{SNR}_X/\mathrm{SNR}_Y)$, then we get $\Delta_{\mathrm{SNR}}^{(X/Y)} = -(10/N_r)\log_{10}(\mathrm{K}_Y/\mathrm{K}_X) = -(10/N_r)\log_{10}(\Pi_{\mathrm{SNR}}^{(X/Y)})$. If $\Delta_{\mathrm{SNR}}^{(X/Y)} > 0$, then for the same ABEP, Y needs $\Delta_{\mathrm{SNR}}^{(X/Y)}$ decibels less transmit energy than X, i.e., $\Delta_{\mathrm{SNR}}^{(X/Y)}$ is the energy gain of Y with respect to X.

Using Table II, in Table III, we show some examples of the SNR advantage/disadvantage of SM with respect to SSK and single-antenna PSK/QAM. Further comments are postponed to Section VI.

VI. Numerical and Simulation Results

The aim of this section is to substantiate frameworks and claims through Monte Carlo simulations. Two case studies are considered: 1) i.i.d. Rayleigh fading (see Section V), and 2) identically distributed Nakagami-m fading (see Section IV-B). In the first case study, we focus our attention on the better accuracy provided by our upper bound, on the comparison of SM with other modulations, and on understanding the role played by the signal and spatial constellation diagrams. In the second case study, we turn our attention to analyze the effect of fading correlation and fading severity on the achievable diversity. Without loss of generality, we consider the identically distributed setup to keep the chosen parameters and variables reasonably low to maintain a sensible set of simulation results. This allows us to focus our attention on fundamental behaviors and to show the main trends. In particular, in the presence of channel correlation, we consider the constant correlation model [41]. The reason of this choice is twofold: 1) to reduce the number of parameters needed to identify the correlation profile and 2) to study a worst-case scenario, which arises when assuming that the constant correlation coefficient corresponds to the pair of antennas that are most closely spaced.

TABLE III

SNR ($\bar{\gamma}$ in Decibels) Difference (See Table II for Definition) Between SM-PSK/QAM and Single-Antenna PSK/QAM Modulation, as Well as SM-QAM and SSK Modulation. SM Outperforms (i.e., It Requires Less Transmit Energy per Single Transmission) the Competing Transmission Technology if $\Delta_{\text{SNR}}^{(X/Y)} > 0$. For a Given Rate R in bpcu, the Constellation Size Is 1) $M^{(\text{PSK},\text{QAM})} = 2^R$ for Single-Antenna PSK/QAM Modulation; 2) $N_t^{\text{SSK}} = 2^R$ for SSK Modulation; and 3) $MN_t = 2^R$ for SM, Where $N_t = 2, 4, 8$ in the First/Second/Third Line of Each Row, Respectively. N.A. Means "Not Available"

$N_r = 1$					
Rate (R) / $\Delta_{\text{SNR}}^{(X/Y)}$	2 bpcu	3 bpcu	4 bpcu	5 bpcu	6 bpcu
(PSK, SM–PSK)	−2.4304	−0.9691	0.5799	2.0412	3.2906
	N.A.	−1.5761	0.2803	2.2640	4.2597
	N.A.	N.A.	0.0684	2.1512	4.3511
(QAM, SM–QAM)	−2.4304	−1.0939	−3.3199	−2.2055	−4.1422
	N.A.	−1.7009	−2.7542	−3.3406	−4.3064
	N.A.	N.A.	−2.9661	−2.3416	−4.9156
(SSK, SM–QAM)	0.5799	0.7918	−0.2854	0.2460	−0.3481
	N.A.	0.1848	0.2803	−0.8890	−0.5123
	N.A.	N.A.	0.0684	0.1100	−1.1215
$N_r = 2$					
Rate / $\Delta_{\text{SNR}}^{(X/Y)}$	2 bpcu	3 bpcu	4 bpcu	5 bpcu	6 bpcu
(PSK, SM–PSK)	−1.0543	1.9011	4.5154	5.6931	5.9642
	N.A.	1.6453	5.3471	8.8845	11.1650
	N.A.	N.A.	5.2585	9.2429	13.1632
(QAM, SM–QAM)	−1.0543	1.7709	0.1040	2.3751	0.9242
	N.A.	1.5152	2.0064	2.2836	2.6976
	N.A.	N.A.	1.9177	4.2581	2.5484
(SSK, SM–QAM)	0.4509	0.3959	−1.7622	−1.8280	−3.6242
	N.A.	0.1401	0.1401	−1.9196	−1.8508
	N.A.	N.A.	0.0515	0.0550	−2.0000
$N_r = 3$					
Rate / $\Delta_{\text{SNR}}^{(X/Y)}$	2 bpcu	3 bpcu	4 bpcu	5 bpcu	6 bpcu
(PSK, SM–PSK)	−0.6461	3.0103	5.5248	5.9627	6.0094
	N.A.	2.8560	7.2677	10.9352	11.9378
	N.A.	N.A.	7.2144	11.8624	16.1295
(QAM, SM–QAM)	−0.6461	2.7651	0.9978	3.3520	1.6807
	N.A.	2.6108	3.6577	3.8842	4.4339
	N.A.	N.A.	3.6044	6.5402	4.7666
(SSK, SM–QAM)	0.3574	0.2639	−2.5664	−3.1516	−5.7457
	N.A.	0.1096	0.0934	−2.6194	−2.9926
	N.A.	N.A.	0.0401	0.0367	−2.6598

A. Better Accuracy of the Improved Upper Bound

In Figs. 1 and 2, we study the accuracy of the improved upper bound in Section III-A against Monte Carlo simulations and the conventional union bound. The frameworks for single-antenna PSK/QAM are obtained from Table I. It can be noticed that our framework is, in general, more accurate than the conventional union bound, and that it well overlaps with the Monte Carlo simulations. In particular, our bound is more accurate than the conventional union bound for large M and small N_r. In addition, the figures compare the ABEP of SM and single-antenna PSK/QAM. In particular, the worst-case scenario with only $N_t = 2$ is considered. We observe two different trends: 1) in Fig. 1, SM-PSK always outperforms PSK, regardless of N_r, and the gain increases with N_r; on the other hand, 2) in Fig. 2, SM-QAM is worse than QAM if $N_r = 1$, and it outperforms QAM if $N_r = 3$. This result is substantiated by the high-SNR framework in Table II. The general outcome of our study for i.i.d. Rayleigh fading is the following: 1) SM-QAM never outperforms QAM for $N_r = 1$, and 2) SM-QAM never outperforms QAM for data rates R less than $R = 2$ bpcu. Further comments about this outcome are given in Section VI-B.

B. Comparison With PSK, QAM, and SSK Modulations

Motivated by Fig. 2, we exploit the framework in Table II to understand deeper the possible performance advantage of SM with respect to SSK and single-antenna PSK/QAM. The accuracy of the frameworks in Table II has been validated through Monte Carlo simulations, and a perfect match has been found. In particular, the interested reader might verify the accuracy of Table II by looking at the SNR difference estimated through Monte Carlo simulations in Figs. 3–8. Table II provides the following outcomes.

1) If $N_r = 1$, SM-QAM never outperforms QAM, and the gap increases with the data rate.
2) Whatever N_r is and if $R < 3$ bpcu, SM-PSK and SM-QAM never outperform PSK and QAM, respectively.
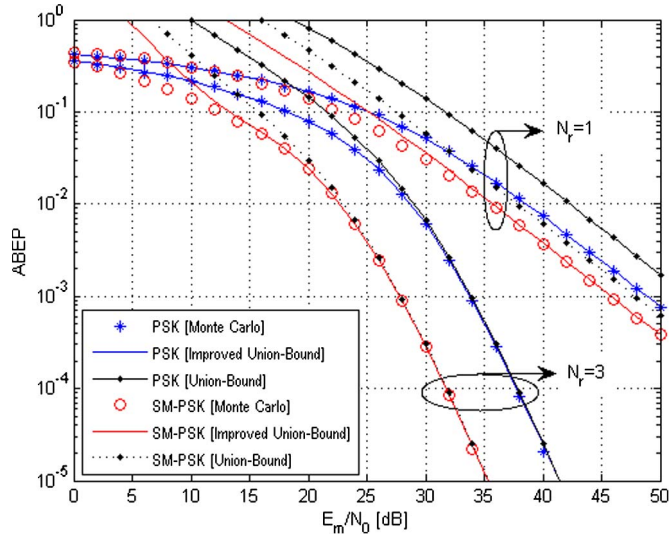
Fig. 1. ABEP of PSK ($M^{\mathrm{PSK}} = 64$) and SM-PSK ($M = 32, N_t = 2$) against E_m/N_0. Accuracy of proposed analytical framework (denoted by "improved union-bound" in the legend) and conventional union bound (denoted by "union-bound" in the legend) for unit-power ($\sigma_0^2 = 1$) i.i.d. Rayleigh fading (the rate is $R = 6$ bpcu).
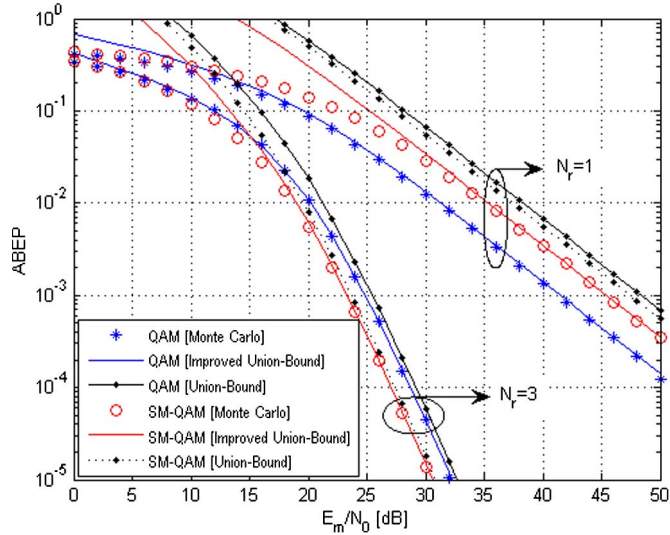

Fig. 3. ABEP of SM-PSK against E_m/N_0. Performance comparison for various sizes of signal and spatial constellation diagrams. Accuracy of proposed analytical frameworks for unit-power ($\sigma_0^2 = 1$) i.i.d. Rayleigh fading (the rate is $R = 4$ bpcu). The setup ($M = 2, N_t = 8$) is not shown as it overlaps with the setup ($M = 4, N_t = 4$).

Fig. 2. ABEP of QAM ($M^{\mathrm{QAM}} = 64$) and SM-QAM ($M = 32, N_t = 2$) against E_m/N_0. Accuracy of proposed analytical framework (denoted by "improved union-bound" in the legend) and conventional union bound (denoted by "union-bound" in the legend) for unit-power ($\sigma_0^2 = 1$) i.i.d. Rayleigh fading (the rate is $R = 6$ bpcu).

Fig. 4. ABEP of SM-QAM against E_m/N_0. Performance comparison for various sizes of signal and spatial constellation diagrams. Accuracy of proposed analytical frameworks for unit-power ($\sigma_0^2 = 1$) i.i.d. Rayleigh fading (the rate is $R = 4$ bpcu). The setup ($M = 2, N_t = 8$) is not shown as it overlaps with the setup ($M = 4, N_t = 4$).

3) Except the former setups, SM always outperforms PSK and QAM, and the gain increases with R, and if more antennas are available at the transmitter, i.e., more information bits can be sent through the spatial constellation diagram.

4) The SNR gain increases with N_r, which means that SM is inherently able to exploit receiver diversity much better than PSK/QAM.

It is important to emphasize here that in Section V-A, we have pointed out that QAM might not be the best modulation scheme for SM. This means that the optimal signal constellation diagram for SM is still unknown, and thus, the ABEP of SM might be reduced further by looking for the signal constellation
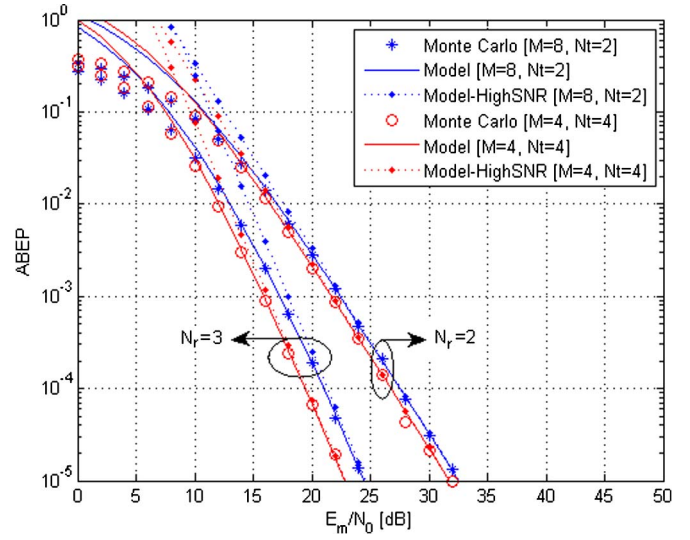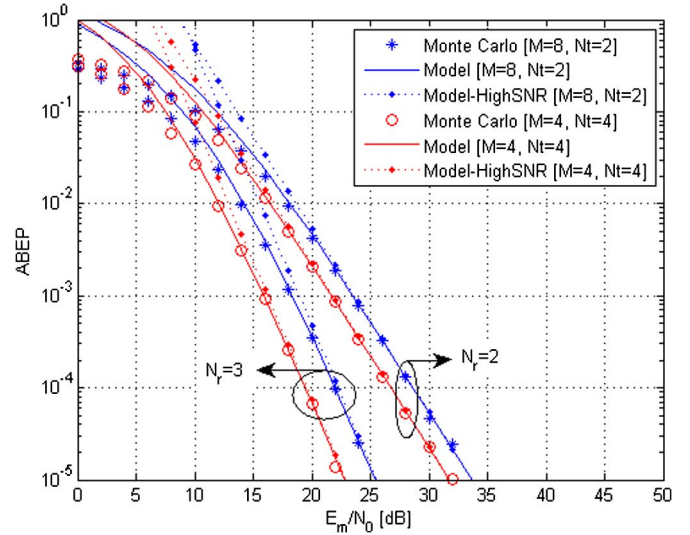
diagram that optimizes the coefficients $\Theta_{\mathrm{spatial}}^{(M,N_r)}$, $\Theta_{\mathrm{joint}}^{(M,N_r)}$, and $\Theta_{\mathrm{joint}}^{(M,N_r,H)}$. In other words, the noticeable gain offered by SM might be increased further, and possibilities of improvement for those setups where SM is worse than state-of-the-art might be found as well. Further comments about the impact of the signal modulation scheme on the performance of SM are available in Section VI-C. This study corroborates our analytical findings and confirms that an adaptive multimode modulation scheme might be a very good choice. Finally, in Table III, we compare SM-QAM with SSK as well. It can be noticed that, particularly for high data rates, SSK outperforms SM-QAM. This result shows that, when R increases, it is convenient to transmit the information bits only through the spatial constellation diagram
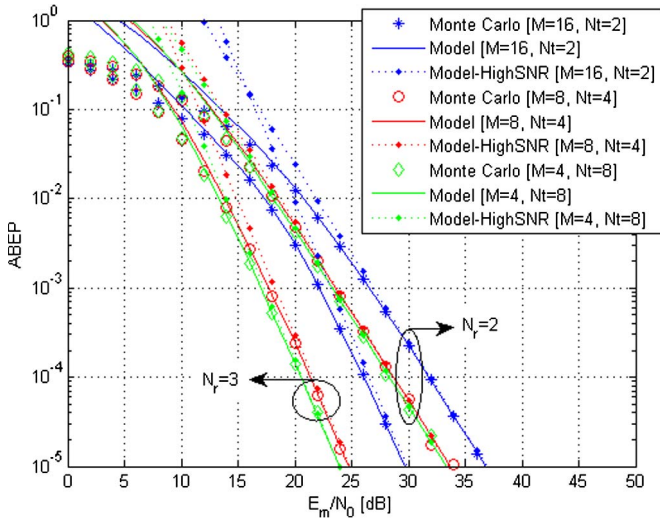
Fig. 5. ABEP of SM-PSK against E_m/N_0. Performance comparison for various sizes of signal and spatial constellation diagrams. Accuracy of proposed analytical frameworks for unit-power ($\sigma_0^2 = 1$) i.i.d. Rayleigh fading (the rate is $R = 5$ bpcu). The setup ($M = 2, N_t = 16$) is not shown as it overlaps with the setup ($M = 4, N_t = 8$).
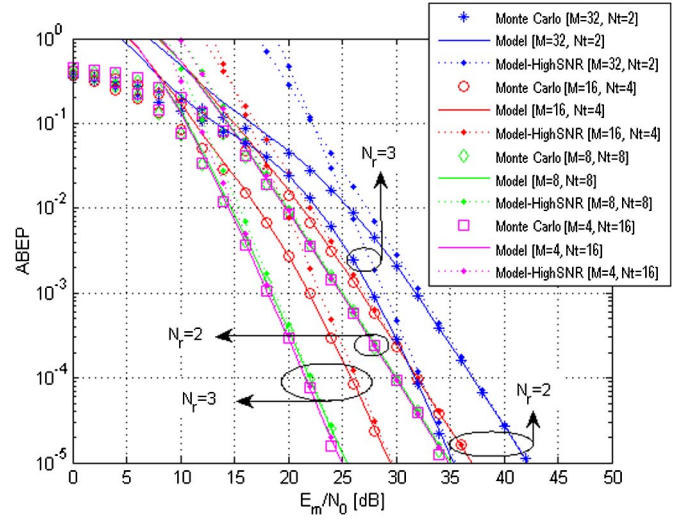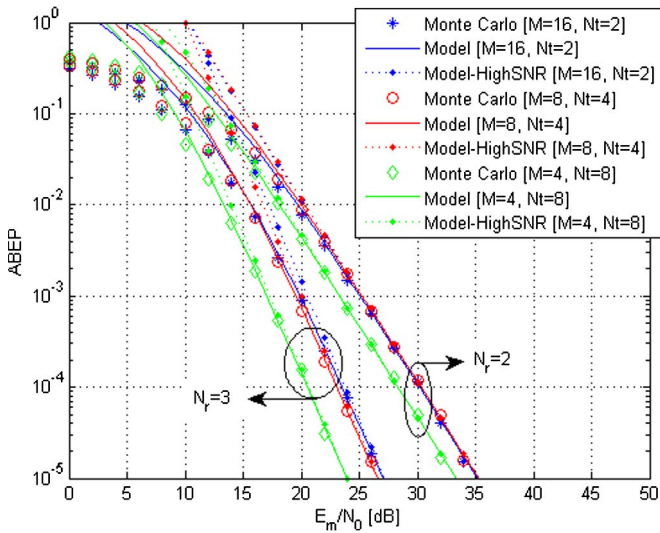

Fig. 7. ABEP of SM-PSK against E_m/N_0. Performance comparison for various sizes of signal and spatial constellation diagrams. Accuracy of proposed analytical frameworks for unit-power ($\sigma_0^2 = 1$) i.i.d. Rayleigh fading (the rate is $R = 6$ bpcu). The setup ($M = 2, N_t = 32$) is not shown as it overlaps with the setup ($M = 4, N_t = 16$).

Fig. 6. ABEP of SM-QAM against E_m/N_0. Performance comparison for various sizes of signal and spatial constellation diagrams. Accuracy of proposed analytical frameworks for unit-power ($\sigma_0^2 = 1$) i.i.d. Rayleigh fading (the rate is $R = 5$ bpcu). The setup ($M = 2, N_t = 16$) is not shown as it overlaps with the setup ($M = 4, N_t = 8$).
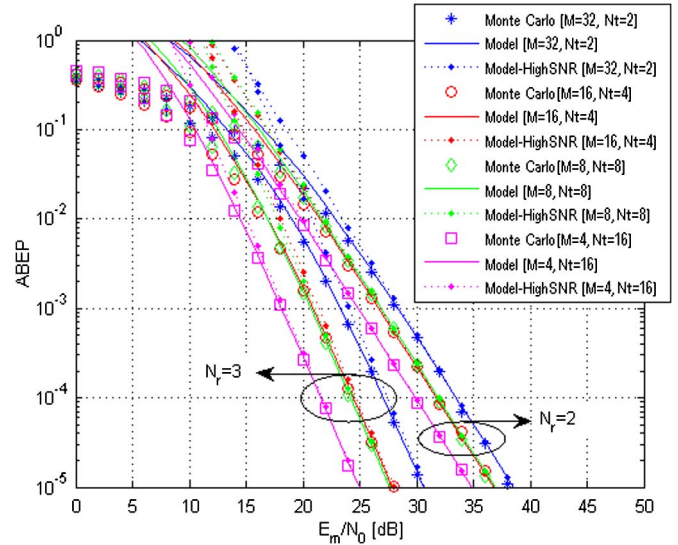

Fig. 8. ABEP of SM-QAM against E_m/N_0. Performance comparison for various sizes of signal and spatial constellation diagrams. Accuracy of proposed analytical frameworks for unit-power ($\sigma_0^2 = 1$) i.i.d. Rayleigh fading (the rate is $R = 6$ bpcu). The setup ($M = 2, N_t = 32$) is not shown as it overlaps with the setup ($M = 4, N_t = 16$).

as this minimizes the ABEP over i.i.d. fading channels. However, the price to pay for this additional improvement is the need of larger antenna arrays at the transmitter. Therefore, there is a clear tradeoff between the achievable performance and the number of antennas that can be put on a transmitter and still being able to keep the i.i.d. assumption. In any case, these numerical examples corroborate the potential performance and energy gain benefits of exploiting SSK for low-complexity "massive" MIMO implementations [20].

C. Interplay of Signal- and Spatial-Constellation Diagrams

In this section, we wish to give a deeper look at the performance of SM for various configurations of signal and spatial

constellation diagrams, as well as at the effect of the adopted modulation scheme. More specifically, we seek to answer two fundamental questions. 1) Is there, for a given data rate R, an optimal pair (N_t, M) that minimizes the ABEP, and 2) is the optimal modulation scheme for single-antenna systems still optimal for SM? The results shown in Figs. 3–8 provide a sound answer to both questions. In particular, if $R = 4$ bpcu, we have the following.

1) The ABEP decreases by increasing N_t, but the improvement is negligible for $N_t > 4$ (thus, $N_t = 4$ can be seen as the optimal choice in this scenario).
2) The SNR gain with N_t is higher in SM-QAM than in SM-PSK.

3) For $N_t = 2$, SM-PSK outperforms SM-QAM, which substantiates the claims in Section V, whereas there is no difference between them for $N_t \geq 4$. In fact, in this latter case, PSK and QAM lead to the same signal constellation diagram. Thus, since PSK modulation is, in general, simpler to be implemented as the power amplifiers at the transmitter have less stringent linearity requirements [46], then SM-PSK seems to be preferred to SM-QAM in all cases.

If $R = 5$ bpcu, we have the following.

1) $N_t = 8$ is the best choice to minimize both the ABEP and the size of the antenna array at the transmitter.
2) For SM-PSK, the setup $N_t = 4$ is a very appealing configuration as the ABEP is close to the optimal value, but the complexity of the transmitter is very low.
3) For $N_t = 2$, SM-QAM is definitely superior to SM-PSK as the spatial constellation diagram has a low impact on the overall performance.
4) For $N_t = 4$, SM-PSK is much better than SM-QAM, and in particular, for SM-QAM, the net improvement when moving from $N_t = 2$ to $N_t = 4$ is negligible.
5) For $N_t \geq 8$, there is no difference between SM-PSK and SM-QAM since they have the same signal constellation diagram, and thus, SM-PSK is the best choice because it is simpler to implement.

In addition, if $R = 6$ bpcu, we have a behavior similar to $R = 4$ bpcu and $R = 5$ bpcu. Thus, we focus only on two main aspects: 1) The best ABEP is obtained when $N_t = 16$. By comparing the best MIMO setup for different rates, we conclude that the best N_t increases with the rate, and the rule of thumb seems to be "double the number of transmit antennas for each 1 bpcu increase of the data rate." Although this increase of the rate might appear to be small for every doubling of the number of antennas at the transmitter, this multiplexing gain is obtained with a single active RF chain and with low (single-stream) decoding complexity. These two features agree with current trends in MIMO research [20], [22], as mentioned in Section I. 2) If $N_t = 8$, SM-PSK is a very appealing choice to achieve very good performance with low complexity. In addition, we emphasize the good accuracy of our framework in all the analyzed scenarios.

Finally, we close this section by mentioning that the good performance offered by SM-PSK against SM-QAM for some MIMO setups and rates brings to our attention that SM-PSK might be a good candidate for energy efficient applications. As a matter of fact, in [46], it is mentioned that a non-negligible percentage of the energy consumption at the base stations of current cellular networks is due to the linearity requirements of the power amplifiers, which are needed to use high-order modulation schemes (such as QAM), and which result in the low power efficiency of the amplifiers. Furthermore, in [47, p. 12], it is clearly stated that this power inefficiency significantly contributes to the so-called quiescent energy, which is independent of the amount of transmitted data and thus should be reduced as much as possible.
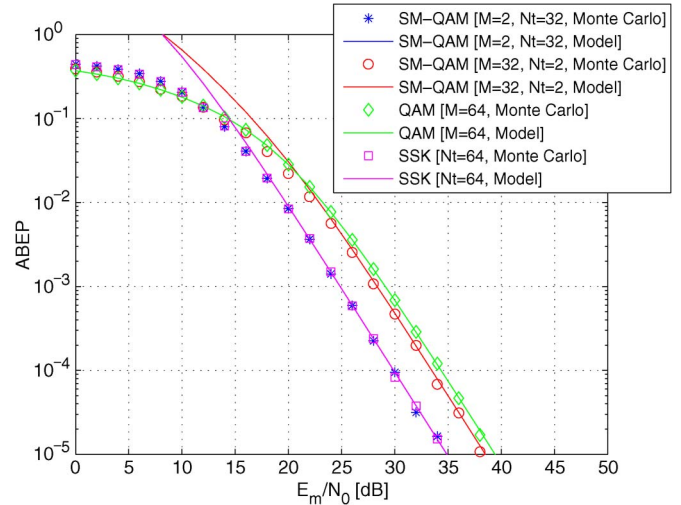

Fig. 9. ABEP against E_m/N_0 over i.i.d. Nakagami-m fading ($m_{\mathrm{Nak}} = 1.0$, i.e., Rayleigh, $N_r = 2$, and rate $R = 6$ bpcu). Performance comparison and accuracy of the analytical framework for SM-QAM, QAM, and SSK.
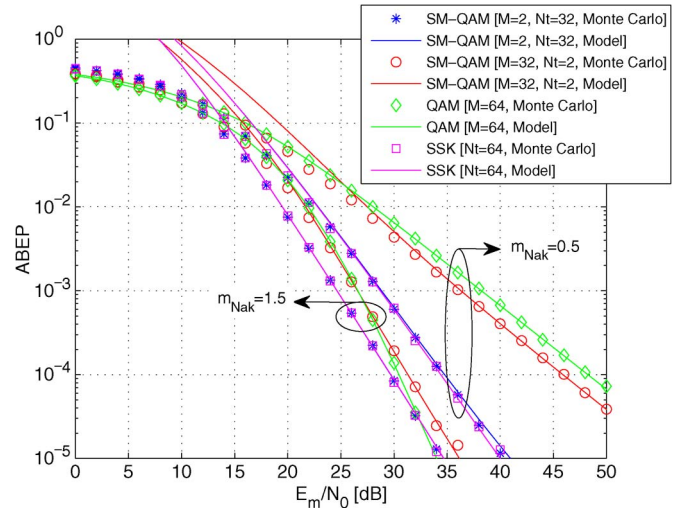

Fig. 10. ABEP against E_m/N_0 over i.i.d. Nakagami-m fading ($m_{\mathrm{Nak}} = 0.5$ and $m_{\mathrm{Nak}} = 1.5$, $N_r = 2$, and rate $R = 6$ bpcu). Performance comparison and accuracy of the analytical framework for SM-QAM, QAM, and SSK.

D. Impact of Fading Severity

In Figs. 9 and 10, we study the impact of fading severity on the performance of QAM, SM, and SSK modulations. Fig. 9 shows the basic scenario with i.i.d. Rayleigh fading ($m_{\mathrm{Nak}} = 1.0$), where from Section IV-B4, we know that all modulations have the same diversity. Fig. 10 highlights the effect of more ($m_{\mathrm{Nak}} = 0.5$) and less ($m_{\mathrm{Nak}} = 1.5$) severe fading. The figures provide the following three important outcomes, which are well captured by the framework in Section IV-B4.

1) Overall, the ABEP gets better for increasing values of m_{Nak}.
2) The SNR gain of SM with respect to QAM increases if $m_{\mathrm{Nak}} = 0.5$, as a consequence of the steeper slope of some components of the ABEP of SM. Furthermore, we notice that SSK is the only modulation scheme with no reduction of the diversity gain. If $N_t = 32$, SM has
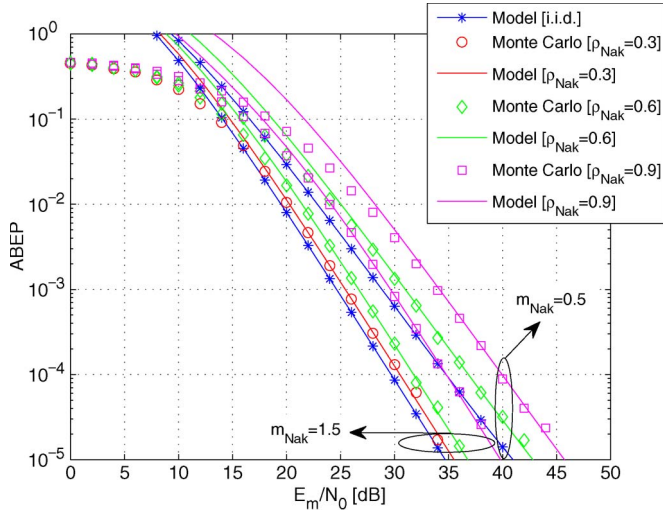
Fig. 11. ABEP of SM-QAM against E_m/N_0 over correlated (at the transmitter) and identically distributed Nakagami-m fading ($m_{\mathrm{Nak}} = 0.5$ and $m_{\mathrm{Nak}} = 1.5$, $N_r = 2$, and rate $R = 6$ bpcu). Performance comparison and accuracy of the analytical framework for $M = 2$ and $N_t = 32$.

Fig. 12. ABEP of SM-QAM against E_m/N_0 over correlated (at the transmitter) and identically distributed Nakagami-m fading ($m_{\mathrm{Nak}} = 0.5$ and $m_{\mathrm{Nak}} = 1.5$, $N_r = 2$, and rate $R = 6$ bpcu). Performance comparison and accuracy of the analytical framework for $M = 32$ and $N_t = 2$.

performance very close to SSK, but the different slope is noticeable even for moderate SNRs.

3) If $m_{\mathrm{Nak}} = 1.5$, QAM provides the best diversity gain, but at low SNR, the high coding gain introduced by SM and SSK is still advantageous. However, a crossing point can be observed for high SNR, which shows that QAM should be preferred in this case.

In conclusion, these results substantiate the diversity analysis conducted in Section IV-B4 and show, once again, that the characteristics of fading are of paramount importance to assess the superiority of a modulation scheme with respect to another one. An adaptive multimode modulation scheme might be an appealing choice to use always the best modulation scheme for any fading scenario.

E. Impact of Fading Correlation

Finally, in Figs. 11–14, we study the impact of fading correlation at the transmitter and receiver over Nakagami-m fading. The analytical framework is available in Section IV-B4, and in particular, in the analyzed scenario, $\mathcal{M}_{\gamma(n_t)}(s) = \mathcal{M}_\gamma(s)$ can be found in [34, eq. (9.173)]. We use a constant correlation model, and ρ_{Nak} denotes the correlation coefficient of pairs of Nakagami-m envelopes. We consider two case studies: 1) channel correlation only at the transmitter (see Figs. 11 and 12), and 2) channel correlation only at the receiver (see Figs. 13 and 14). The rationale of this choice is to investigate the different effects that correlation might have at either ends of the communication link. In fact, according to (5), correlation might have a different impact at the transmitter and at the receiver: correlation at the transmitter affects the distance of points in the spatial constellation diagram, whereas correlation at the receiver reduces the diversity gain of maximal ratio combining (MRC) at the destination.

In Figs. 11 and 12, we study the impact of correlation at the transmitter. It can be noticed, as expected, that performance
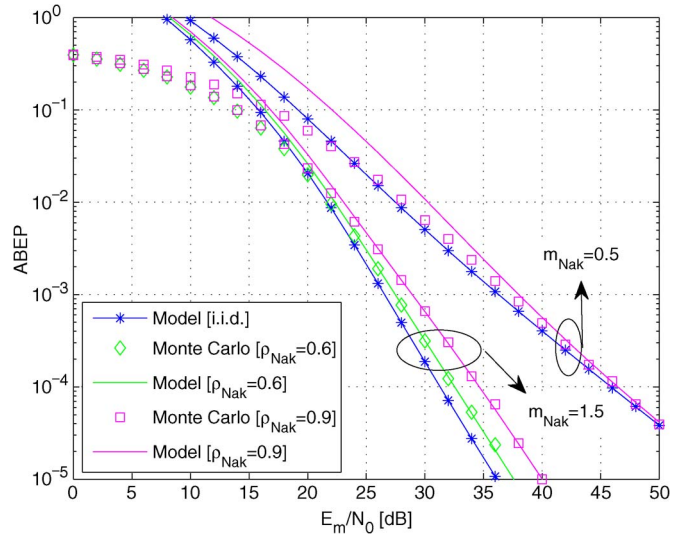

Fig. 13. ABEP of SM-QAM against E_m/N_0 over correlated (at the receiver) and identically distributed Nakagami-m fading ($m_{\mathrm{Nak}} = 0.5$ and $m_{\mathrm{Nak}} = 1.5$, $N_r = 2$, and rate $R = 6$ bpcu). Performance comparison and accuracy of the analytical framework for $M = 2$ and $N_t = 32$.
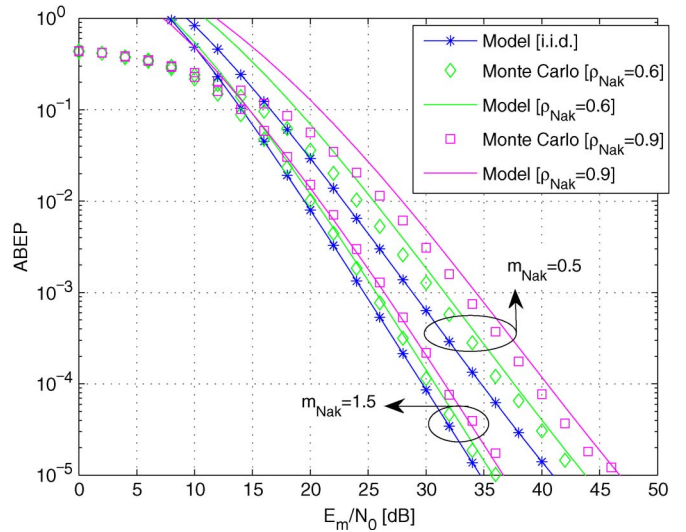
degrades with channel correlation. In addition, the impact of correlation increases with N_t, which is a reasonable outcome in our scenario. However, the SNR degradation with increasing values of ρ_{Nak} is tolerable if $\rho_{\mathrm{Nak}} < 0.6$, whereas for higher values, a few decibel loss can be observed. Very interestingly, Fig. 12 shows that channel correlation has a negligible effect if $m_{\mathrm{Nak}} = 0.5$. This result is very interesting, particularly if compared to the same curves in Fig. 11 and with the ABEP of QAM in Fig. 10 (QAM uses just one transmit antenna, and thus, it is not affected by fading correlation at the transmitter). In particular, we note the following: 1) If $N_t = 2$, SM is always superior to QAM, regardless of fading correlation; and 2) if $N_t = 32$, SM is much better than QAM, even for a high fading correlation ($\rho_{\mathrm{Nak}} = 0.9$). The net outcome is the following: For
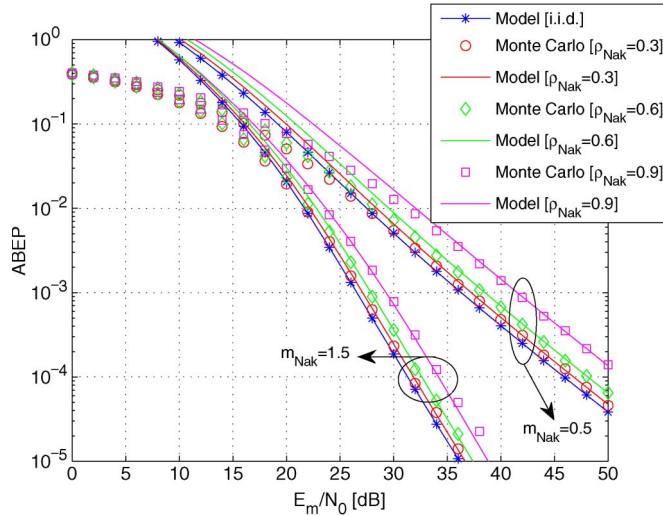
Fig. 14. ABEP of SM-QAM against E_m/N_0 over correlated (at the receiver) and identically distributed Nakagami-m fading ($m_{\text{Nak}} = 0.5$ and $m_{\text{Nak}} = 1.5$, $N_r = 2$, and rate $R = 6$ bpcu). Performance comparison and accuracy of the analytical framework for $M = 32$ and $N_t = 2$.

severe fading channels, correlation degrades the ABEP, but it does not offset the SNR gain that, for independent fading, SM has with respect to QAM. On the other hand, if $m_{\text{Nak}} = 1.5$, the superiority of QAM becomes even more pronounced if compared to the independent fading scenario. In conclusion, fading correlation at the transmitter poses no problems to SM in severe fading channels, whereas it should be carefully managed in other fading scenarios, particularly if we want to keep the performance advantage over single-antenna QAM (whose ABEP is not affected by this correlation). For SM, solutions to counteract fading correlation have recently been proposed in [9] and [14]. Once again, we emphasize that, because of the constant correlation model, Figs. 11 and 12 show the worst case effect of fading correlation, particularly for large N_t.

In Figs. 13 and 14, we study the impact of correlation at the receiver. Overall, the ABEP degrades for increasing ρ_{Nak}. A higher robustness to fading correlation can be noticed for $m_{\text{Nak}} = 1.5$. If $m_{\text{Nak}} = 0.5$, the diversity advantage of SM with respect to QAM if kept in the presence of channel correlation too. For large antenna arrays at the transmitter (e.g., $N_t = 32$), the diversity loss in $\text{ABEP}_{\text{signal}}$ has a negligible impact even for high correlated channels. If $m_{\text{Nak}} = 1.5$, we observe that the SNR degradation gets smaller for larger antenna arrays at the transmitter. In other words, transmitting more information bits through the spatial constellation diagram (e.g., increasing N_t) can mitigate the effect of channel correlation at the receiver. However, Figs. 11 and 12 point out a clear tradeoff: increasing N_t degrades the ABEP if we have channel correlation at the transmitter. We believe that the exploitation of the proposed frameworks for an end-to-end system optimization by taking into account all these tradeoffs might be a very important research issue: How do we find the optimal SM setup providing the best performance/complexity tradeoff as a function of fading correlation, fading severity, etc.?

Finally, we wish to emphasize the good accuracy of our framework for the very complicated fading scenario under analysis. Our framework agrees with Monte Carlo simulations

in all scenarios. Only in some figures there are negligible errors, which are mainly due to the Green approximation described in Section IV-B. Thus, our frameworks can be exploited for accurate system optimization.

VII. CONCLUSION

In this paper, we have proposed a comprehensive framework for the analysis of SM-MIMO over generalized fading channels. The framework is applicable to a large variety of correlated fading models and MIMO setups. Furthermore, and more importantly, by carefully analyzing the obtained formulas, we have derived important information about the performance of SM over fading channels, including the effect of fading severity, the achievable diversity gain, along with the impact of the signal constellation diagram. It has been shown that the modulation scheme used in the signal constellation diagram significantly affects the performance, and for i.i.d. Rayleigh fading, closed-form expressions for its optimization have been proposed. Finally, we have conducted an extensive simulation campaign to validate the analytical derivation and have showcased important trends about the performance of SM for a large variety of fading scenarios and MIMO setups. We believe that our frameworks can be very useful to understand fundamental behaviors and tradeoffs of SM as well as can be efficiently used for system optimization.

APPENDIX A
PROOF OF PROPOSITION 1

Before going into the details of the proof, let us analyze the Hamming distance $N_H((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l))$ of messages $\mu(\tilde{n}_t, \chi_{\tilde{l}})$ and $\mu(n_t, \chi_l)$. In particular, $N_H((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l))$ is equal to the number of different bits between the messages. Since a bit error might occur when 1) only the antenna index is wrongly detected; 2) only the signal-modulated point is wrongly detected; or 3) both antenna index and signal-modulated point are wrongly detected, then we conclude that the total number of bits in error is given by $N_H((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l)) = N_H(\tilde{n}_t \to n_t) + N_H(\chi_{\tilde{l}} \to \chi_l)$, where $N_H(\tilde{n}_t \to n_t)$ and $N_H(\chi_{\tilde{l}} \to \chi_l)$ are defined in Proposition 1. This remark is used to compute (6)–(8), and it is important to highlight the role played by the bit mapping in each constellation diagram. Proposition 1 can be obtained as follows:

1) $\text{ABEP}_{\text{signal}}$ is obtained from (4) by grouping together all the terms for which $\tilde{n}_t = n_t$ and $\tilde{l} \neq 1$, and by noticing the following: a) $N_H(\tilde{n}_t \to n_t) = 0$ if $\tilde{n}_t = n_t$; and b) (5) reduces to $\text{APEP}((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l)) = \text{E}_{\boldsymbol{\alpha}(n_t)} \{Q(\sqrt{\bar{\gamma}|\chi_{\tilde{l}} - \chi_l|^2 \sum_{n_r=1}^{N_r} |\alpha_{n_t,n_r}|^2})\}$. Then, $\text{ABEP}_{\text{signal}} = \text{ABEP}_{\text{signal}}^{\text{bound}}$ in (4) reduces to (19), shown at the bottom of the next page.

2) It can readily be noticed that $\text{ABEP}_{\text{MOD}}^{\text{bound}}(n_t)$ is the union bound of a conventional modulation scheme [34], where a) only the n_tth transmit antenna is active, and b) we have the same constellation diagram as the signal constellation diagram of SM. More specifically, $\text{ABEP}_{\text{MOD}}^{\text{bound}}(n_t)$ is the ABEP of a single-input–multiple-output system

with MRC. This ABEP is known in closed-form for many modulation schemes and bit mappings without the need of using union-bound methods. Thus, to get more accurate estimates of the ABEP, $\text{ABEP}_{\text{MOD}}^{\text{bound}}(\cdot)$ can be replaced by $\text{ABEP}_{\text{MOD}}(\cdot)$, as shown in (8), which is the exact ABEP of a single-input–multiple-output system with MRC.

3) Likewise, $\text{ABEP}_{\text{spatial}}$ is obtained from (4) by grouping together all the terms for which $\tilde{n}_t \neq n_t$ and $\tilde{l} = l$, and by noticing that a) $N_H(\chi_{\tilde{l}} \to \chi_l) = 0$ if $\tilde{l} = l$, and b) (5) reduces to $\text{APEP}((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l)) = \text{E}_{\boldsymbol{\alpha}(n_t, \tilde{n}_t)}\{Q(\sqrt{\bar{\gamma}\kappa_l^2 \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r} - \alpha_{n_t,n_r}|^2})\}$. Finally, from [34, eq. (4.2)], we have $\Psi_l(n_t, \tilde{n}_t) = \text{APEP}((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l))$, where $\Psi_l(\cdot, \cdot)$ is defined in Section III-A.

4) $\text{ABEP}_{\text{joint}}$ in (7) collects all the terms that are neither in $\text{ABEP}_{\text{signal}}$ nor in $\text{ABEP}_{\text{spatial}}$. More specifically, (7) can be obtained from [34, eq. (4.2)]: $\Upsilon(n_t, l, \tilde{n}_t, \tilde{l}) = \text{APEP}((\tilde{n}_t, \chi_{\tilde{l}}) \to (n_t, \chi_l)) = \text{E}_{\boldsymbol{\alpha}(n_t, \tilde{n}_t)}\{Q(\sqrt{\bar{\gamma} \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r}\chi_{\tilde{l}} - \alpha_{n_t,n_r}\chi_l|^2})\}$, where $\Upsilon(\cdot, \cdot, \cdot, \cdot)$ is defined in Section III-A.

APPENDIX B
PROOF OF PROPOSITION 2

By definition, $\mathcal{M}_{\gamma_{(n_t, \tilde{n}_t)}}(\cdot)$ is given by

$$
\mathcal{M}_{\gamma_{(n_t, \tilde{n}_t)}}(s)
$$
$$
= \text{E}\left\{ \exp\left[-s \sum_{n_r=1}^{N_r} |\alpha_{\tilde{n}_t,n_r} - \alpha_{n_t,n_r}|^2 \right] \right\}
$$
$$
= \text{E}\left\{ \prod_{n_r=1}^{N_r} \exp\left(-s|\alpha_{\tilde{n}_t,n_r} - \alpha_{n_t,n_r}|^2 \right) \right\}
$$
$$
= \text{E}_{\boldsymbol{\beta}}\left\{ \prod_{n_r=1}^{N_r} \exp\left(-s\beta_{\tilde{n}_t,n_r}^2 \right) \times \prod_{n_r=1}^{N_r} \exp\left(-s\beta_{n_t,n_r}^2 \right) \right.
$$
$$
\times \text{E}_{\boldsymbol{\varphi}}\left\{ \prod_{n_r=1}^{N_r} \exp\left[2s\beta_{\tilde{n}_t,n_r}\beta_{n_t,n_r} \right. \right.
$$
$$
\left. \left. \left. \times \cos(\varphi_{\tilde{n}_t,n_r} - \varphi_{n_t,n_r}) \right] \right\} \right\} \tag{20}
$$

where the last equality explicitly shows the conditioning over fading envelopes and channel phases, and $\boldsymbol{\beta}$, $\boldsymbol{\varphi}$ are shorthands to denote the set of all fading envelopes and channel phases, respectively. Let us compute $J(s; \beta_{\tilde{n}_t,n_r}, \beta_{n_t,n_r}) =$

$\text{E}_{\boldsymbol{\varphi}}\{\prod_{n_r=1}^{N_r} \exp[2s\beta_{\tilde{n}_t,n_r}\beta_{n_t,n_r}\cos(\varphi_{\tilde{n}_t,n_r} - \varphi_{n_t,n_r})]\}$ in (20). It can be obtained as

$$
J(s; \beta_{\tilde{n}_t,n_r}, \beta_{n_t,n_r})
$$
$$
= \prod_{n_r=1}^{N_r} \text{E}_{\boldsymbol{\varphi}}\left\{ \exp\left[2s\beta_{\tilde{n}_t,n_r}\beta_{n_t,n_r}\cos(\varphi_{\tilde{n}_t,n_r} - \varphi_{n_t,n_r}) \right] \right\}
$$
$$
= \prod_{n_r=1}^{N_r} I_0(2s\beta_{\tilde{n}_t,n_r}\beta_{n_t,n_r}) \tag{21}
$$

where the first equality is due to the independence of the channel phases, and the second equality is obtained from [35, p. 339, eqs. (366) and (367)] and [24, eq. (14)]. Accordingly, $\mathcal{M}_{\gamma_{(n_t, \tilde{n}_t)}}(\cdot)$ simplifies as

$$
\mathcal{M}_{\gamma_{(n_t, \tilde{n}_t)}}(s)
$$
$$
= \text{E}_{\boldsymbol{\beta}}\left\{ \prod_{n_r=1}^{N_r} \left[\exp\left(-s\beta_{\tilde{n}_t,n_r}^2 \right) \exp\left(-s\beta_{n_t,n_r}^2 \right) \right. \right.
$$
$$
\left. \left. \times I_0\left(2s\beta_{\tilde{n}_t,n_r}\beta_{n_t,n_r} \right) \right] \right\}
$$
$$
= \int_{\boldsymbol{\beta}}\left\{ \prod_{n_r=1}^{N_r} \left[\exp\left(-s\beta_{\tilde{n}_t,n_r}^2 \right) \exp\left(-s\beta_{n_t,n_r}^2 \right) \right. \right.
$$
$$
\left. \left. \times I_0\left(2s\beta_{\tilde{n}_t,n_r}\beta_{n_t,n_r} \right) \right] \right\} f_{\boldsymbol{\beta}}(\boldsymbol{\beta})\,d\boldsymbol{\beta} \tag{22}
$$

where $f_{\boldsymbol{\beta}}(\cdot)$ is the multivariate Nakagami-m pdf in [41, eq. (2)].

As an example, and without loss of generality, let us consider $N_r = 2$. For ease of notation, we set $n_t = 1$ and $\tilde{n}_t = 2$. Accordingly, (22) reduces to (23), shown on the top of the next page. Finally, by using the infinite series representation of $I_v(\cdot)$ in [33, eq. (9.6.10)], and after lengthy algebraic manipulations, $\mathcal{M}_{\gamma_{(n_t, \tilde{n}_t)}}(\cdot)$ can be rewritten as shown in (13), where the integrals in (24), shown at the top of the next page, have been introduced. These latter integrals can be computed in closed-form from [24, Sec. III-B], thus obtaining the final result in (14). More specifically, the analytical procedure we have used to compute (14) is as follows: 1) first, the integral on variable $\beta_{2,1}$ is solved in closed-form by using the identities in [32, eq. (8.4.3.1)] and [32, eq. (8.4.22)], as well as by applying the Mellin–Barnes theorem in [32, eq. (2.24.1.1)] on the obtained integral; and 2) second, the obtained single integral on variable $\beta_{1,1}$ is solved in closed-form by using again the identity in [32, eq. (8.4.3.1)] and by applying the Mellin–Barnes theorem in [32, eq. (2.24.1.1)].

$$
\begin{cases}
\text{ABEP}_{\text{signal}}^{\text{bound}} = \frac{1}{N_t}\frac{\log_2(M)}{\log_2(N_t M)} \sum_{n_t=1}^{N_t} \text{ABEP}_{\text{MOD}}^{\text{bound}}(n_t) \\[2ex]
\text{ABEP}_{\text{MOD}}^{\text{bound}}(n_t) = \frac{1}{M}\frac{1}{\log_2(M)} \sum_{l=1}^{M} \sum_{\tilde{l}=1}^{M} \left[N_H(\chi_{\tilde{l}} \to \chi_l)\text{E}_{\boldsymbol{\alpha}(n_t)}\left\{ Q\left(\sqrt{\bar{\gamma}|\chi_{\tilde{l}} - \chi_l|^2 \sum_{n_r=1}^{N_r} |\alpha_{n_t,n_r}|^2} \right) \right\} \right]
\end{cases} \tag{19}
$$

$$
\begin{cases}
\mathcal{M}_{\gamma_{(n_t,\tilde{n}_t)}}(s) = \int\limits_{\boldsymbol{\beta}} \left\{ \left[\exp\left(-s\beta_{1,1}^2\right) \exp\left(-s\beta_{2,1}^2\right) I_0(2s\beta_{1,1}\beta_{2,1}) \right] \left[\exp\left(-s\beta_{1,2}^2\right) \exp\left(-s\beta_{2,2}^2\right) I_0(2s\beta_{1,2}\beta_{2,2}) \right] \right\} f_{\boldsymbol{\beta}}(\boldsymbol{\beta})d\boldsymbol{\beta} \\
f_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \left[\dfrac{\left|\boldsymbol{\Sigma}_{\text{trid}}^{-1}\right|^{m_{\text{Nak}}}}{2^{(m_{\text{Nak}}-1)}\Gamma(m_{\text{Nak}})} \beta_{1,1}^2 \beta_{2,2}^2 \exp\left(-\dfrac{p_{44}}{2}\beta_{2,2}^2\right) \right] \times \left[|p_{12}|^{-(m_{\text{Nak}}-1)}\beta_{1,1} \exp\left(-\dfrac{p_{11}}{2}\beta_{1,1}^2\right) I_{m_{\text{Nak}}-1}\left(|p_{12}|\beta_{1,1}\beta_{1,2}\right) \right] \\
\qquad\qquad \times \left[|p_{23}|^{-(m_{\text{Nak}}-1)}\beta_{1,2} \exp\left(-\dfrac{p_{22}}{2}\beta_{1,2}^2\right) I_{m_{\text{Nak}}-1}\left(|p_{23}|\beta_{1,2}\beta_{2,1}\right) \right] \\
\qquad\qquad \times \left[|p_{34}|^{-(m_{\text{Nak}}-1)}\beta_{2,1} \exp\left(-\dfrac{p_{33}}{2}\beta_{2,1}^2\right) I_{m_{\text{Nak}}-1}\left(|p_{34}|\beta_{2,1}\beta_{2,2}\right) \right]
\end{cases}
\tag{23}
$$

$$
\begin{cases}
\mathcal{F}_{\mathbf{k}}^{(p_{11},p_{33})}(s) = \int\limits_0^{+\infty}\int\limits_0^{+\infty} \beta_{1,1}^{2m_{\text{Nak}}+2k_1-1} \beta_{2,1}^{2m_{\text{Nak}}+2k_2+2k_3-1} \exp\left[-\left(s+\dfrac{p_{11}}{2}\right)\beta_{1,1}^2\right] \\
\qquad\qquad \times \exp\left[-\left(s+\dfrac{p_{33}}{2}\right)\beta_{2,1}^2\right] I_0(2s\beta_{1,1}\beta_{2,1})d\beta_{1,1}d\beta_{2,1} \\
\mathcal{F}_{\mathbf{k}}^{(p_{22},p_{44})}(s) = \int\limits_0^{+\infty}\int\limits_0^{+\infty} \beta_{1,2}^{2m_{\text{Nak}}+2k_1+2k_2-1} \beta_{2,2}^{2m_{\text{Nak}}+2k_3-1} \exp\left[-\left(s+\dfrac{p_{22}}{2}\right)\beta_{1,2}^2\right] \\
\qquad\qquad \times \exp\left[-\left(s+\dfrac{p_{44}}{2}\right)\beta_{2,2}^2\right] I_0(2s\beta_{1,2}\beta_{2,2})d\beta_{1,2}d\beta_{2,2}
\end{cases}
\tag{24}
$$

The analytical development can be generalized to arbitrary N_r by simply inserting in (22) the general pdf in [41, eq. (2)] and solving the integrals as in (21)–(24).

Finally, a few comments about the Green approximation $\boldsymbol{\Sigma} \cong \boldsymbol{\Sigma}_{\text{trid}}$ in (23). 1) The pdf in (23) requires the correlation matrix $\boldsymbol{\Sigma}$ of the Gaussian RVs associated to the fading envelopes. This matrix can be computed from the amplitude correlation coefficient $\rho_{\text{Nak}}^{(n_t,n_r,\tilde{n}_t,\tilde{n}_r)}$ by using the procedure in [42, Sec. III]. 2) For arbitrary and unequal values of $\Omega_{\text{Nak}}^{(n_t,n_r)}$, the Green method in [41], which is given under the assumption that $\Omega_{\text{Nak}}^{(n_t,n_r)} = 1$ for $n_t = 1, 2, \ldots, N_t$ and $n_r = 1, 2, \ldots, N_r$, must be generalized. More specifically, the coefficients u_i in [41, eq. (9)], which are needed to compute $\boldsymbol{\Sigma}_{\text{trid}}$, take the form $u_i = \boldsymbol{\Sigma}(i,i)/v_i$, where $\boldsymbol{\Sigma}(i,i)$ is the entry of $\boldsymbol{\Sigma}$ located in the ith row and in the ith column, and v_i are the coefficients to be computed by solving the nonlinear system of equations in [41, eq. (10)].

REFERENCES

[1] Y. Chau and S.-H. Yu, "Space modulation on wireless fading channels," in *Proc. IEEE Veh. Technol. Conf.*, Oct. 2001, vol. 3, pp. 1668–1671.

[2] S. Song, Y. Yang, Q. Xiong, K. Xie, B.-J. Jeong, and B. Jiao, "A channel hopping technique I: Theoretical studies on band efficiency and capacity," in *Proc. IEEE Int. Conf. Commun., Circuits Syst.*, Jun. 2004, vol. 1, pp. 229–233.

[3] R. Y. Mesleh, H. Haas, C. W. Ahn, and S. Yun, "Spatial modulation—A new low complexity spectral efficiency enhancing technique," in *Proc. IEEE Int. Conf. Commun. Netw. China*, Oct. 2006, pp. 1–5.

[4] M. Di Renzo, H. Haas, and P. M. Grant, "Spatial modulation for multiple-antenna wireless systems—A survey," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 182–191, Dec. 2011.

[5] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, Jul. 2008.

[6] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Spatial modulation: Optimal detection and performance analysis," *IEEE Commun. Lett.*, vol. 12, no. 8, pp. 545–547, Aug. 2008.

[7] Y. Yang and B. Jiao, "Information-guided channel-hopping for high data rate wireless communication," *IEEE Commun. Lett.*, vol. 12, no. 4, pp. 225–227, Apr. 2008.

[8] J. Jeganathan, A. Ghrayeb, L. Szczecinski, and A. Ceron, "Space shift keying modulation for MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3692–3703, Jul. 2009.

[9] T. Handte, A. Muller, and J. Speidel, "BER analysis and optimization of generalized spatial modulation in correlated fading channels," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2009, pp. 1–5.

[10] S. U. Hwang, S. Jeon, S. Lee, and J. Seo, "Soft-output ML detector for spatial modulation OFDM systems," *IEICE Electron. Exp.*, vol. 6, no. 19, pp. 1426–1431, Oct. 2009.

[11] M. M. Ulla Faiz, S. Al-Ghadhban, and A. Zerguine, "Recursive least-squares adaptive channel estimation for spatial modulation systems," in *Proc. IEEE Malaysia Int. Conf. Commun.*, Dec. 2009, pp. 785–788.

[12] M. Di Renzo and H. Haas, "Performance comparison of different spatial modulation schemes in correlated fading channels," in *Proc. IEEE Int. Conf. Commun.*, May 2010, pp. 1–6.

[13] N. Serafimovski, M. Di Renzo, S. Sinanovic, R. Y. Mesleh, and H. Haas, "Fractional bit encoded spatial modulation (FBE-SM)," *IEEE Commun. Lett.*, vol. 14, no. 5, pp. 429–431, May 2010.

[14] R. Y. Mesleh, M. Di Renzo, H. Haas, and P. M. Grant, "Trellis coded spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2349–2361, Jul. 2010.

[15] M. Di Renzo and H. Haas, "Performance analysis of spatial modulation," in *Proc. IEEE Int. Conf. Commun. Netw. China*, Aug. 2010, pp. 1–7.

[16] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3219–3230, Nov. 2010.

[17] M. Di Renzo and H. Haas, "Transmit-diversity for spatial modulation (SM): Towards the design of high-rate spatially-modulated space-time block codes," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2011, pp. 1–6.

[18] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Space-time block coded spatial modulation," *IEEE Trans. Commun.*, vol. 59, no. 3, pp. 823–832, Mar. 2011.

[19] S. Sugiura, S. Chen, and L. Hanzo, "Generalized space-time shift keying designed for flexible diversity-, multiplexing- and complexity-tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1144–1153, Apr. 2011.

[20] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[21] G. Wright, "GreenTouch initiative: Large scale antenna systems demonstration," in *Proc. Spring Meeting*, Seoul, Korea, 2011.

[22] A. Mohammadi and F. M. Ghannouchi, "Single RF front-end MIMO transceivers," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 104–109, Dec. 2011.

[23] M. Di Renzo and H. Haas, "Improving the performance of space shift keying (SSK) modulation via opportunistic power allocation," *IEEE Commun. Lett.*, vol. 14, no. 6, pp. 500–502, Jun. 2010.

[24] M. Di Renzo and H. Haas, "A general framework for performance analysis of space shift keying (SSK) modulation for MISO correlated Nakagami-m fading channels," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2590–2603, Sep. 2010.

[25] M. Di Renzo and H. Haas, "Space shift keying (SSK) modulation with partial channel state information: Optimal detector and performance analysis over fading channels," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3196–3210, Nov. 2010.

[26] M. Di Renzo and H. Haas, "Space shift keying (SSK-) MIMO over correlated Rician fading channels: Performance analysis and a new method for transmit-diversity," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 116–129, Jan. 2011.

[27] M. Di Renzo and H. Haas, "Space shift keying (SSK) modulation: On the transmit-diversity/multiplexing trade-off," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2011, pp. 1–6.

[28] M. Di Renzo and H. Haas, "Bit error probability of space shift keying MIMO over multiple-access independent fading channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3694–3711, Oct. 2011.

[29] M. Di Renzo, D. De Leonardis, F. Graziosi, and H. Haas, "Space shift keying (SSK-) MIMO with practical channel estimates," *IEEE Trans. Commun.*, 2012, to be published. [Online]. Available: http://arxiv.org/pdf/1201.4793v1.pdf

[30] A. Alshamali and B. Quza, "Performance of spatial modulation in correlated and uncorrelated Nakagami fading channel," *J. Commun.*, vol. 4, no. 3, pp. 170–174, Apr. 2009.

[31] M. Di Renzo, R. Y. Mesleh, H. Haas, and P. M. Grant, "Upper bounds for the analysis of trellis coded spatial modulation over correlated fading channels," in *Proc. IEEE Veh. Technol. Conf.*, May 2010, pp. 1–5.

[32] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series: More Special Functions.* Moscow, Russia: Fizmatlit, 2003.

[33] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York: Dover, 1972.

[34] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 1st ed. Hoboken, NJ: Wiley, 2000.

[35] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* Hoboken, NJ: Wiley, 2001.

[36] M. Irshid and I. Salous, "Bit error probability for coherent M-ary PSK systems," *IEEE Trans. Commun.*, vol. 39, no. 3, pp. 349–352, Mar. 1991.

[37] A. Hedayat, H. Shah, and A. Nosratinia, "Analysis of space-time coding in correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2882–2891, Nov. 2005.

[38] J. Lassing, E. G. Strom, E. Agrell, and T. Ottosson, "Computation of the exact bit-error rate of coherent M-ary PSK with Gray code bit mapping," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1758–1760, Nov. 2003.

[39] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.

[40] J. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.

[41] G. K. Karagiannidis, D. A. Zogas, and S. A. Kotsopoulos, "An efficient approach to multivariate Nakagami-m distribution using Green's matrix approximation," *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 883–889, Sep. 2003.

[42] Q. T. Zhang, "A decomposition technique for efficient generation of correlated Nakagami fading channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2385–2392, Nov. 2000.

[43] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1389–1398, Aug. 2003.

[44] M. Di Renzo and H. Haas, "Bit error probability of space modulation over Nakagami-m fading: Asymptotic analysis," *IEEE Commun. Lett.*, vol. 15, no. 10, pp. 1026–1028, Oct. 2011.

[45] J. Cheng and T. Berger, "Capacity of a class of fading channels with channel state information (CSI) feedback," in *Proc. Allerton Conf. Commun. Control, Comput.*, Oct. 2001, pp. 1152–1160.

[46] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, Fourth Quarter, 2011.

[47] S. D. Gray, "Theoretical and practical considerations for the design of green radio networks," in *Proc. IEEE Veh. Technol. Conf.*, Budapest, Hungary, May 2011. [Online]. Available: http://www.ieeevtc.org/conf-admin/vtc2011spring/5.pdf

Marco Di Renzo (SM'05–A'07–M'09) was born in L'Aquila, Italy, in 1978. He received the Laurea (*cum laude*) and Ph.D. degrees in electrical and information engineering from the University of L'Aquila, in April 2003 and January 2007, respectively.

From August 2002 to January 2008, he was with the Center of Excellence for Research Design methodologies of Embedded controllers, Wireless interconnet and Systems-on-chip, University of L'Aquila. From February 2008 to April 2009, he was a Research Associate with the Telecommunications Technological Center of Catalonia, Barcelona, Spain. From May 2009 to December 2009, he was an Engineering and Physical Sciences Research Council Research Fellow with the Institute for Digital Communications, The University of Edinburgh, Edinburgh, U.K. Since January 2010, he has been a Tenured Researcher ("Chargé de Recherche Titulaire") with the French National Center for Scientific Research (CNRS), as well as a research staff member of the Laboratory of Signals and Systems (L2S): a joint research laboratory of the CNRS, the École Supérieure d'Électricité (SUPÉLEC), and the University of Paris-Sud XI, Paris, France. His main research interests are in the areas of wireless communications theory, signal processing, and information theory.

Dr. Di Renzo was the recipient of a special mention for the outstanding five-year (1997–2003) academic career from the University of L'Aquila; the THALES Communications fellowship for doctoral studies (2003–2006), University of L'Aquila; and the Torres Quevedo Award for his research on ultra-wideband systems and cooperative localization for wireless networks (2008–2009), Ministry of Science and Innovation, Spain.

Harald Haas (S'98–A'00–M'03) received the Ph.D. degree from the University of Edinburgh in 2001. He currently holds the Chair of mobile communications with the Institute for Digital Communications, University of Edinburgh, Edinburgh, U.K. His main research interests are in the areas of wireless system design and analysis, as well as digital signal processing, with a particular focus on interference coordination in wireless networks, spatial modulation, and optical wireless communication. He holds more than 15 patents. He has published more than 50 journal papers, including a science article and more than 140 peer-reviewed conference papers. Nine of his papers are invited papers. He has co-authored a book entitled *Next Generation Mobile Access Technologies: Implementing TDD* (Cambridge University Press). Since 2007, he has been a Regular High Level Visiting Scientist supported by the Chinese "111 program" with the Beijing University of Posts and Telecommunications.

Prof. Haas was an invited speaker at the 2011 Technology Entertainment Design (TED) Global Conference. He was shortlisted for the World Technology Award for communications technology (individual) in 2011.