Title	Distributed Spatial Modulation Aided NOMA
Authors(s)	Shehni, Amir, Flanagan, Mark F.
Publication date	2019-06-21
Publication information	Shehni, Amir, and Mark F. Flanagan. "Distributed Spatial Modulation Aided NOMA." IEEE, 2019.
Conference details	The 2019 European Conference on Networks and Communication (EuCNC), Valenica, Spain, 18-21 June 2019
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/11136
Publisher's statement	© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/eucnc.2019.8801970

Distributed Spatial Modulation Aided NOMA

Amir Shehni and Mark F. Flanagan
University College Dublin, Belfield, Dublin 4, Ireland
Email: amir.shehni@ucd.ie, mark.flanagan@ieee.org

***Abstract*—In this paper, a novel cooperative diversity protocol based on the association of non-orthogonal multiple access (NOMA) and distributed spatial modulation (DSM) is introduced. In the proposed protocol, two source symbols are multiplexed in the power domain, while one source symbol obtains a diversity gain due to its being relayed according to the DSM principle; this doubles the data rate for the source-to-destination link as compared with conventional DSM. We propose two demodulators for use at the destination: an error-aware demodulator which is robust to demodulation errors at the relays, and a suboptimal demodulator which assumes error-free demodulation at the relays. Simulation results demonstrate that while the proposed protocol achieves a source data throughput equal to that of a full-duplex system, its BER performance also significantly outperforms the full-duplex relaying benchmarks of successive relaying and virtual full-duplex DSM.**

I. Introduction

Multiple-input multiple-output (MIMO) systems have received significant research attention over the last two decades due to their impressive benefits. Compared to single-antenna systems, MIMO possesses advantages such as higher data rates as well as increased diversity and spatial multiplexing gains [1]. In spite of the many benefits of MIMO, one of the key challenges is to devise new transmission protocols which deliver the benefits of a multiple-antenna system while activating only one or a small number of antennas at any one time. Apart from removing the need for multiple RF chains at the transmitter, this approach also addresses issues such as inter-antenna synchronization (IAS), which is inevitable for any space-time communication, and inter-channel interference (ICI) which occurs due to the superposition of different signals transmitted by different antennas [2]. Conventional solutions to these issues inevitably increase the complexity and cost of the MIMO system.

The MIMO technique of *spatial modulation* (SM) [2], [3] uses only *one* of the transmit antennas at any one time. This allows the use of a single RF chain while still retaining the many benefits of MIMO. By using SM, the effects of ICI and IAS will be removed, so the system will have a lower complexity and cost [3]. SM operates by activating a particular antenna based on the matching of part of the data with an antenna's unique ID, along with the use of a conventional modulation technique such as PSK/QAM [4]. The simultaneous use of the *spatial* domain and the *signal* domain gives the system extra dimensions for the conveyance of information. Indeed, SM has been recognized as a promising spectrally efficient and energy-efficient transmission method for 5G mobile communication [5].

Non-orthogonal multiple access (NOMA) is one of the most promising multiple access technologies for 5G communication [6]. It multiplexes multiple users in the energy domain and/or code domain in the same time-frequency resource block, which enables the system to achieve higher spectral efficiency as well as higher cell edge throughput. It has also a good backward compatability with other technologies and can be easily used in conjunction with other technologies like OFDMA, massive MIMO or spatial modulation [7]. In particular, in [8] the authors proposed a novel SM-aided NOMA system and used the mutual information metric to characterize its achievable spectral efficiency.

Cooperative communication has established itself as a strong candidate for future wireless networks. It can effectively improve achievable rate and coverage, especially at the cell edge. It has been shown that cooperation is a beneficial way to combat shadowing and multipath fading [9], [10]. Relaying systems usually employ single-antenna devices in order to obtain the benefits of MIMO by forming a distributed antenna array. However, single-antenna devices usually follow the restrictions of half-duplex communication including: i) the necessity of extra bandwidth resources, ii) usage of relays' resources (e.g. signal processing and energy) for cooperation, and iii) the delay experienced by the relays' own data while they are prioritized to retransmit the source data.

Since spatial modulation requires many antennas to achieve high spectral efficiency, for the uplink, the assumption of the availability of many antennas is not very practical, as the transmitting devices usually have only one antenna. However, the necessity for multiple transmit antennas can be bypassed by employing cooperative relays to form a virtual MIMO system. Spatial modulation has been generalized in this direction, leading to the introduction of a distributed version of SM (DSM) in half-duplex relaying. DSM increases the aggregate throughput of the network by allowing the relays to *explicitly* transmit their own data while *implicitly* forwarding the source data. This is possible since the relaying process operates by using the *index* of the activated (transmitting) relay [11], [12]. Targeting an increased source-to-destination data rate, a network coded version of DSM was introduced in [13] which increases the source data rate by 33% compared to DSM. Other research considers a virtual full-duplex DSM (VFD-DSM) system which is able to transmit a new source symbol in every time slot [14], [15]. The application of NOMA in cooperative communication has also been considered in the literature [16]–[18]. Since NOMA and cooperative communication are both capable of improving the spectral efficiency (particularly at the cell edge), it is desirable to jointly leverage the advantages of these technologies.

In this paper, we propose a new variant of half-duplex DSM which also uses the concept of NOMA. The new protocol, called NOMA-DSM, enables two source symbols to be communicated in two time slots, i.e., the system has a throughput comparable to VFD-DSM. We also propose two demodulators for use at the destination: an error-aware maximum likelihood (ML) demodulator which is robust to demodulation errors at the relays, and a suboptimal demodulator which assumes error-free demodulation at the relays. Simulation results indicate that while the BER of one source symbol in the proposed NOMA-DSM is comparable to that of VFD-DSM at high signal to noise ratio (SNR), the other source symbol can achieve a diversity order of 2.

II. System Model

A brief review of DSM is given in this section, as DSM forms the basis for this work. The proposed protocol, NOMA-DSM, is introduced, and some relevant definitions and notations are also provided.

A. Review of Distributed Spatial Modulation

In DSM, a unique ID is assigned to each relay. The source broadcasts its symbol in the first time slot and it is received by all relays and by the destination. All relays demodulate the source data according to the ML criterion, while the destination stores the received signal for later processing. In the second time slot, each relay whose ID matches to the demodulated data at the relay will be active and transmit its own symbol to the destination. This means that a total of two time slots is required for a complete transmission of each source symbol. Note that although the relays transmit their own data (via straightforward signal modulation), they also *implicitly* carry a copy of the source data through the *index* of the activated relay. The operation of the DSM protocol is illustrated in Fig. **??**(a) for the case of two relays with the assumption of error-free demodulation at the relay. Here the source data in time slot 1 is $p_s = -1$ which leads to the activation of the relay with the same ID ($\mathrm{ID}_{R_2} = -1$) in time slot 2.

B. New protocol: NOMA-DSM

In NOMA-DSM, in the first time slot the source broadcasts a linear combination of two symbols p_1 and p_2 with different allocated energies, where the energies of the symbols are weighted by factors α_1 and α_2 respectively. In this paper we assume $\alpha_1 > \alpha_2$, and accordingly we will refer to symbols p_1 and p_2 as the *stronger symbol* and the *weaker symbol*, respectively. The resulting transmitted signal is $p_s = \sqrt{\alpha_1 E_s}\ p_1 + \sqrt{\alpha_2 E_s}\ p_2$. It is received by all relays and by the destination. All of these nodes perform ML detection to demodulate the stronger symbol (p_1). Then, successive interference cancellation (SIC) is performed at each relay for the detection of the weaker symbol (p_2). In the destination, SIC is used to detect symbol p_1, but detection of p_2 will be performed

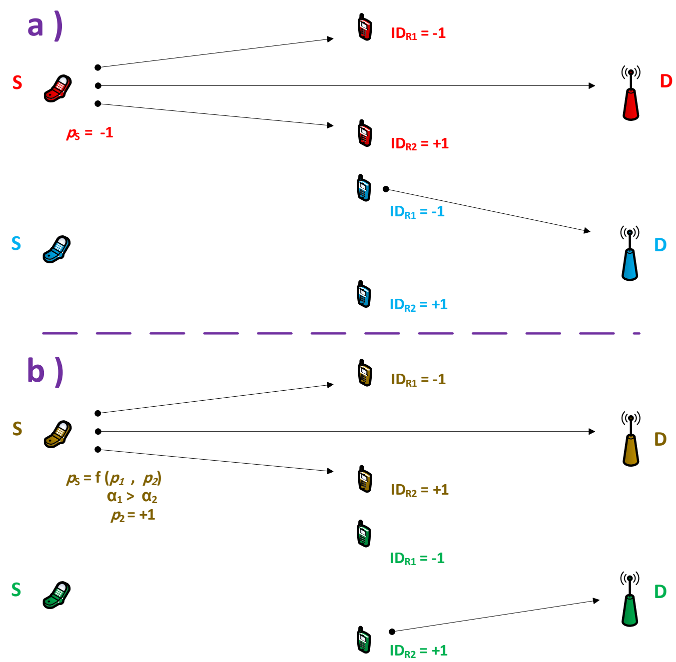

Fig. 1. a) Illustration of distributed spatial modulation for the source symbol $p_s = -1$. b) Illustration of the proposed NOMA-DSM. As can be seen, while the transmitted signal p_s is the combination of two different symbols p_1 and p_2, the activation is based on the symbol p_2. For both protocols, BPSK modulation is assumed, and for simplicity it is assumed that demodulation at the relays is error-free.

later. In the second time slot, the symbol p_2 detected at each relay will be checked against the relay's unique ID; if there is a match, the relay will be active and transmit its own data in the second time slot. Finally, at the destination, a joint detection is performed using the signal received in the first time slot with the effect of symbol p_1 removed, together with the signal received in the second time slot (which indicates the relay activation). Note that in conventional power domain NOMA, it is usual that higher energy will be given to the symbol which should be transmitted through the weaker channel. But in the proposed NOMA-DSM, the lower energy will be assigned to the symbol which will be aided by the cooperative link (i.e. p_2), and the higher energy will be given to the symbol which is detected in the conventional manner for point-to-point (i.e., non-cooperative) communication systems. The operation of the NOMA-DSM protocol is illustrated in Fig. **??**(b) for the case of binary source data, BPSK modulation, two relays, and error-free demodulation at the relays. Here the weaker symbol is $p_2 = +1$ and each relay will demodulate this symbol using SIC; since this data matches the ID of relay R_2, this relay will be active (and transmit its own data) in the second time slot.

C. Notation and definitions

A network including one source (S), one destination (D) and $M = 2^q$ relays $(R_1, R_2, \ldots, R_M)$ is considered, with the assumption of half-duplex devices equipped with a single antenna. Each source symbol from the complex constellation $\mathcal{A}_s$ is assumed to consist of phase shift keying (PSK) or quadrature amplitude modulation (QAM) symbols with unit energy. The two source symbols for power-domain multiplexing are denoted by $p_n = \mathcal{M}_s(\mathbf{x}_n)$ for $n \in \{1,2\}$, where each $\mathbf{x}_n$ represents an information bit sequence of length $\log_2(M)$, and $\mathcal{M}_s(.)$ represents the bit-to-symbol mapping at the source (this mapping is assumed to be the same for both symbols). The fading channel between nodes X and Y is a complex Gaussian random variable h_{XY} with zero mean and variance σ^2_{XY}; the channel coefficients h_{XY} are assumed to be independent and identically distributed random variables (i.i.d channels). The complex additive white Gaussian noise (AWGN) with mean zero and variance $\sigma^2 = N_0/2$ per dimension at any node B is given by n_B (here N_0 denotes the noise power spectral density). Each relay F uses a unit-energy PSK/QAM constellation $\mathcal{A}_r$ of size N for transmitting its own data symbol $p_F = \mathcal{M}_F(\mathbf{x}_F)$, where $\mathbf{x}_F$ is the information bit sequence of relay F having length $\log_2(N)$, and $\mathcal{M}_F(.)$ represents the bit-to-symbol mapping at relay F. The received signal at any node B is denoted by y_B. Also, a unique digital identifier ID_{R_r} for $r = 1, 2, \ldots, M$ will be assigned to each relay in which ID_{R_r} is a binary vector of length $\log_2(M)$. For example, if $M = 4$ we could have $\mathrm{ID}_{R_1} = 00$, $\mathrm{ID}_{R_2} = 01$, $\mathrm{ID}_{R_3} = 10$ and $\mathrm{ID}_{R_4} = 11$ (of course, the binary ID can trivially be replaced by the corresponding symbols in the source constellation $\mathcal{A}_s$, as was done in the example in Fig. 1).

III. Transmission Phases in NOMA-DSM

A. Source broadcasting phase

In first time slot, the source broadcasts its symbol $p_s = \sqrt{\alpha_1 E_s}\ p_1 + \sqrt{\alpha_2 E_s}\ p_2$, where E_s is the average transmit energy at the source. It is assumed that $\alpha_1 > \alpha_2$. For every $F \in \Phi = \{R_1, R_2, ..., R_M\}$, the received signal at relay F is given by

$$\begin{aligned} y_F &= h_{SF}\ p_s +\ n_F \\ &= h_{SF}\ (\sqrt{\alpha_1 E_s}\ p_1 + \sqrt{\alpha_2 E_s}\ p_2) +\ n_F. \end{aligned} \tag{1}$$

Applying the ML criterion, the demodulation of symbol p_1, at the relay proceeds via

$$\hat{p}_1^F = \underset{\tilde{p}_s \in \mathcal{A}_s}{\arg\min} \{\ |\ y_F - \sqrt{\alpha_1 E_s}\ h_{SF}\ \tilde{p}_s|^2\}, \tag{2}$$

and the corresponding data vector is estimated via $\hat{\mathbf{x}}_1^F = \mathcal{M}_s^{-1}(\ \hat{p}_1^F\ )$. Then, SIC operates by subtracting the effect of $\hat{p}_1$ and performing ML detection again to detect $\hat{p}_2$, an estimate of the weaker symbol. This ML detection proceeds via

$$\hat{p}_2^F = \underset{\tilde{p}_s \in \mathcal{A}_s}{\arg\min} \{\ |\ (y_F - \sqrt{\alpha_1 E_s}\ h_{SF}\ \hat{p}_1^F) - \sqrt{\alpha_2 E_s}\ h_{SF}\ \tilde{p}_s|^2\}, \tag{3}$$

and the data corresponding to $\hat{p}_2$ at relay F is estimated via $\hat{\mathbf{x}}_2^F = \mathcal{M}_s^{-1}(\ \hat{p}_2^F\ )$.

Simultaneously, the destination receives the signal

$$\begin{aligned} y_{D_1} &= h_{SD}\ p_s +\ n_{D_1} \\ &= h_{SD}\ (\sqrt{\alpha_1\ E_s}\ p_1 + \sqrt{\alpha_2 E_s}\ p_2) +\ n_{D_1}, \end{aligned} \tag{4}$$

and performs ML detection to estimate the stronger symbol p_1, i.e.,

$$\hat{p}_1^{D_1} = \underset{\tilde{p}_s \in \mathcal{A}_s}{\arg\min} \{\ |\ y_{D_1} - \sqrt{\alpha_1 E_s}\ h_{SD}\ \tilde{p}_s|^2\}, \tag{5}$$

and the corresponding data vector is estimated via $\hat{\mathbf{x}}_1^{D_1} = \mathcal{M}_s^{-1}(\ \hat{p}_1^{D_1}\ )$. Now, p_1 is detected at the destination but for detection of p_2, only the first step of SIC is performed to compute

$$\bar{y}_{D_1} =\ y_{D_1} -\ \sqrt{\alpha_1 E_s}\ h_{SD}\ \hat{p}_1^{D_1}. \tag{6}$$

The destination will wait until the second time slot transmission before completing the demodulation of symbol p_2.

B. Cooperative phase

After the demodulation of p_2 at each relay using (3), each relay F applies the SM principle in order to simultaneously transmit the relays' own data and forward the weaker source symbol p_2. In particular, the modulated symbol of relay F is chosen as follows:

$$p_F = \begin{cases} \mathcal{M}_F(\mathbf{x}_F) & \text{if} \quad \mathrm{ID}_F = \hat{\mathbf{x}}_2^F \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

where $\mathbf{x}_F$ denotes the relays' own data vector to be transmitted to the destination. Finally, the signal received at the destination in the relaying phase can be written as

$$y_{D_2} = \sum_{F \in \Phi} \sqrt{E_r}\ h_{FD}\ p_F + n_{D_2}, \tag{8}$$

where E_r denotes the average transmit energy per symbol at the relay.

Thus, in the absence of demodulation errors at the relays, the index of the active relay conveys extra information regarding the second source symbol p_2 according to the spatial modulation concept; this boosts the achievable diversity gain for the destination's detection of the weaker source symbol, assuming that the stronger symbol has already been detected correctly. Note that there is also the possibility of more than one relay being simultaneously active due to demodulation errors at the relays. A more robust (albeit complex) demodulator can be designed which takes into account such errors. This is discussed in detail in Section IV-A.

IV. Detection Process at the Destination

In this section, two detection schemes are presented for implementation at the destination. First, a robust detector, called the *error-aware* demodulator, which takes into account the possibility of demodulation errors at the relays is derived. Second, a suboptimal detector, called the *low-complexity* demodulator, which assumes error-free demodulation at the relays, is presented. Note that both demodulators are for joint detection of the relay data as well as the weaker source symbol, where the stronger source symbol has already been detected via (5).

A. Error-aware ML demodulator

In cooperative wireless systems, the assumption of error-free demodulation at the relays will not hold in general, especially at low SNR. Therefore, in this section we develop an optimal demodulator which is robust in the presence of demodulation errors at the relay.

The destination's ML detector seeks to determine in a joint manner the most likely source symbol p_2, relay activations, and relays' own data. Therefore, it seeks to maximize

$$P_{D_2} = P\left(\tilde{p}_2, \tilde{\mathbf{p}}_R \mid \bar{y}_{D_1}, y_{D_2}\right), \tag{9}$$

where $\tilde{p}_2$ and $\tilde{\mathbf{p}}_R$ represent a hypothesis for the source symbol and a hypothesis for the *vector* of symbols transmitted from the relays, respectively, where $\tilde{\mathbf{p}}_R = (\tilde{p}_{R_1}\ \tilde{p}_{R_2}\ \ldots\ \tilde{p}_{R_M})^T$. Note that here the silent condition $\tilde{p}_F = 0$ is allowed for each F, which corresponds to the case of *non-activation* of relay F. Applying Bayes' rule, factorizing, and ignoring constant terms, (9) becomes

$$P_{D_2} = p\left(\bar{y}_{D_1} \mid \tilde{p}_2\right) p\left(y_{D_2} \mid \tilde{\mathbf{p}}_R\right) P\big(\tilde{p}_2\big) P\big(\tilde{\mathbf{p}}_R \mid \tilde{p}_2\big). \tag{10}$$

The destination's detector will assume that p_1 has been detected correctly at the destination, so that the first term in (10) is given by

$$p\big(\bar{y}_{D_1}|\tilde{p}_2\big) = \frac{1}{\pi N_0} \exp\left(-\frac{\left|\bar{y}_{D_1} - \sqrt{\alpha_2 E_s}\ h_{SD}\ \tilde{p}_2\right|^2}{N_0}\right). \tag{11}$$

The second term, corresponding to the relaying phase, is given by

$$p\big(y_{D_2}|\tilde{\mathbf{p}}_R\big) = \frac{1}{\pi N_0} \exp\left(-\frac{\left|y_{D_2} - \sqrt{E_r}\ \mathbf{h}_{RD}^T\ \tilde{\mathbf{p}}_R\right|^2}{N_0}\right), \tag{12}$$

where $\mathbf{h}_{RD} = (h_{R_1D}\ h_{R_2D}\ \ldots\ h_{R_MD})^T$ denotes the vector of channel coefficients of the relay-destination links. The term $P\big(\tilde{p}_2\big)$ in (10) can also be ignored since the source symbols are assumed to be equiprobable *a priori*. The remaining term $P\big(\tilde{\mathbf{p}}_R|\tilde{p}_2\big)$ plays the key role in designing an error-aware demodulator.

An optimal demodulator, which takes into account the possibility of demodulation errors at the relays, operates by using the knowledge of the source-relay channel qualities. The error probability for a single PSK/QAM symbol transmitted over a point-to-point fading channel in additive white Gaussian noise is in general given by

$$P_e = \beta \cdot Q\left(\sqrt{2\gamma \cdot |h|^2 (E/N)}\ \right), \tag{13}$$

where E, N, and h denote the average transmit symbol energy, the variance of the complex AWGN at the receiver, and the channel fading coefficient, respectively. Here γ and β are constants which depend on the particular constellation (PSK/QAM) used at the transmitter.

In the proposed scheme, each relay receives a superposition of the stronger and weaker symbols. Therefore, the probability of error for the symbol p_2 at relay F, which is equal to the probability of incorrect (in)activation, is given by

$$\begin{aligned} P_{e2}^F = &\ P(p_2 \text{ in error} \mid p_1 \text{ correct}) \cdot P(p_1 \text{ correct}) \\ &+ P(p_2 \text{ in error} \mid p_1 \text{ in error}) \cdot P(p_1 \text{ in error}). \end{aligned} \tag{14}$$

At high SNR, the terms $P(p_1 \text{ correct})$ and $P(p_2 \text{ error} \mid p_1 \text{ error})$ in (14) are close to unity, and thus we have

$$P_{e2}^F \simeq P(p_2 \text{ error} \mid p_1 \text{ correct}) + P(p_1 \text{ error}). \tag{15}$$

Using the equation (13) in (15), and considering the assigned energy for each symbol according to the NOMA principle, the probability of error for the weaker symbol is given by

$$\begin{aligned} P_{e2}^F = &\ \beta \cdot Q\left(\sqrt{\frac{2\gamma \cdot |h_{SF}|^2 \cdot \alpha_1 \cdot E_s}{|h_{SF}|^2 \cdot \alpha_2 \cdot E_s + N_0}}\right) \\ &+ \beta \cdot Q\left(\sqrt{2\gamma \cdot |h_{SF}|^2 \cdot \alpha_2 \cdot E_s / N_0}\right). \end{aligned} \tag{16}$$

where γ and β are the constants associated with the constellation $\mathcal{A}_s$.

The second term in (10) can be expressed as

$$P(\tilde{\mathbf{p}}_R \mid \tilde{p}_2) = \prod_{F \in \widetilde{\Phi}^{(\mathrm{ON})}} P(\tilde{p}_F \mid \tilde{p}_2) \cdot \prod_{F \in \widetilde{\Phi}^{(\mathrm{OFF})}} P(\tilde{p}_F \mid \tilde{p}_2), \tag{17}$$

where $\widetilde{\Phi}^{(\mathrm{ON})}$ denotes the set of *active* relays based on the relay symbol vector hypothesis $\tilde{\mathbf{p}}_R$; and $\widetilde{\Phi}^{(\mathrm{OFF})}$ denotes the set of *inactive* relays based on $\tilde{\mathbf{p}}_R$. Then, for $F \in \widetilde{\Phi}^{(\mathrm{ON})}$,

$$P(\tilde{p}_F \mid \tilde{p}_2) = \begin{cases} \frac{1}{N}\ (1 - P_{e2}^F) & \text{if} \quad \mathrm{ID}_F = \tilde{\mathbf{x}}_2 \\ \frac{1}{N}\ P_{e2}^F & \text{otherwise,} \end{cases} \tag{18}$$

while for $F \in \widetilde{\Phi}^{(\mathrm{OFF})}$,

$$P(\tilde{p}_F \mid \tilde{p}_2) = \begin{cases} P_{e2}^F & \text{if} \quad \mathrm{ID}_F = \tilde{\mathbf{x}}_2 \\ 1 - P_{e2}^F & \text{otherwise.} \end{cases} \tag{19}$$

The first case in (18) corresponds to the probability of correct *activation* of relay F and the second case corresponds to the case of incorrect activation. The first case in (19) corresponds to the case where the relay ID matches the true source symbol but the relay is silent (i.e., *incorrect non-activation* of relay F). Finally, the second case in (19) determines the probability of *correct non-activation* of relay F.

The optimal error-aware ML demodulator at the destination seeks to maximize the metric P_{D_2} given by (10). After some algebraic simplification, this optimization problem can be written as

$$\begin{aligned} \{\hat{p}_2, \hat{\mathbf{p}}_R\} = &\ \underset{\tilde{p}_2 \in \mathcal{A}_s,\ \tilde{\mathbf{p}}_R \in (\mathcal{A}_r \cup \{0\})}{\arg\min} \Big\{ \\ &\left|\bar{y}_{D_1} - \sqrt{\alpha_2 E_s}\ h_{SD}\ \tilde{p}_2\right|^2 + \left|y_{D_2} - \sqrt{E_r}\ \mathbf{h}_{RD}^T\ \tilde{\mathbf{p}}_R\right|^2 \\ &- N_0 \log P\left(\tilde{\mathbf{p}}_R \mid \tilde{p}_2\right) \Big\}, \end{aligned} \tag{20}$$

and the detected data is then obtained via $\hat{\mathbf{x}}_2 = \mathcal{M}_s^{-1}(\hat{p}_2)$ and $\hat{x}_F = \mathcal{M}_F^{-1}(\hat{p}_F)$ for each $F \in \Phi$.

This demodulator is robust in that the search is conducted over all possible active relay sets, regardless of the hypothesis of the weaker source symbol $\tilde{p}_2$. Thus the complexity of this demodulator is $O\left(M(N+1)^M\right)$, which is practical only for small values of the source constellation size M. At reasonable values of the source-relay SNR, the probability of more than two simultaneous relay activations is negligible. Therefore, for larger values of M, the complexity can be dramatically reduced while sacrificing little optimality, by narrowing the set of hypothesized vectors $\tilde{\mathbf{p}}_R$ in (20) to those for which $\left|\widetilde{\Phi}^{(\mathrm{ON})}\right| \leq 2$. Given this reduction in the size of the search space, the complexity for the demodulator becomes $O\left(M\left[1+MN+\binom{M}{2}N^2\right]\right)$.

B. Low-complexity demodulator

In this subsection, a suboptimal low-complexity demodulator is presented which assumes error-free demodulation at the relays. Under this assumption, the index of the active relay *must* match the weaker source symbol. Therefore, this low-complexity ML demodulator can be formulated as

$$\{\hat{p}_2,\ \hat{p}_F\} = \underset{\tilde{p}_2 \in \mathcal{A}_s,\ \tilde{p}_F \in \mathcal{A}_r}{\arg\min} \Bigg\{ \left|\bar{y}_{D_1} - \sqrt{\alpha_2 E_s}\ h_{SD}\ \tilde{p}_2\right|^2 + \left|y_{D_2} - \sqrt{E_r}\ h_{FD}\ \tilde{p}_F\right|^2 \Bigg\}, \tag{21}$$

where F is the relay matching the hypothesis $\tilde{p}_2$ via $\mathrm{ID}_F = \mathcal{M}_s^{-1}(\tilde{p}_2)$. The estimated data is then obtained via $\hat{\mathbf{x}}_2 = \mathcal{M}_s^{-1}(\hat{p}_2)$ and $\hat{\mathbf{x}}_F = \mathcal{M}_F^{-1}(\hat{p}_F)$. Also, it can be seen that this demodulator has complexity $O(M \cdot N)$.

V. Simulation Results and Discussion

Although the proposed NOMA-DSM protocol is a half-duplex protocol, it communicates two source symbols in two time slots. Hence, concerning source symbols, the system performance can be compared to full-duplex protocols where source sends a new symbol in every time slot. Therefore, for source data, two baseline full-duplex protocols, *successive relaying* [19], [20] and the more recently proposed *virtual full-duplex DSM (VFD-DSM)* [14], [15] are considered as benchmarks. For the relay data, NOMA-DSM follows the restrictions of a half-duplex system, communicating one relay symbol per two time slots. Therefore, the conventional DSM (working in half duplex mode) is chosen as a baseline for NOMA-DSM for evaluation of the BER of the relay data.

The benchmark scheme of successive relaying operates as follows: in every odd time slot $2k-1$, relay R_1 forwards the symbol received from the source in the previous time slot while the other relay R_2 receives the current transmission of the source. In the next (even) time slot $2k$, the relays interchange their roles; R_2 forwards the previous source symbol while R_1 receives. A direct link also exists from source to destination. It should be noted that in successive relaying, the relays do not have their own data and only retransmit the source data to the destination. The destination performs MAP detection over two consecutive time slots.

The benchmark scheme of VFD-DSM is similar to DSM, except that it operates in full-duplex mode where the source is able to transmit new data in every time slot, while the relays can also transmit their own data. Here, the performance of NOMA-DSM is compared with that of the two demodulators for VFD-DSM, called *Local MAP* and *Global MAP*, which were presented in [15].

A. Simulation setup

For all protocols, it is assumed that there are two relays, and that BPSK modulation is employed at the source and relays ($M = N = 2$). A node geometry is assumed where the distance between all pairs of communicating nodes is equal, i.e., $\sigma_{SD}^2 = \sigma_{SF}^2 = \sigma_{FD}^2 = 1$ for $F \in \{R_1, R_2\}$. The average transmit symbol energy is the same for both source and relays, i.e., $E_s = E_r$. The values for the energy weighting factors α_1 and α_2 are 0.9 and 0.1, respectively. For all protocols, the effect of inter-relay interference (IRI) is neglected; note however that the benchmark full-duplex protocols will suffer more severely from IRI than the proposed protocol.

B. Simulation Results

In Fig. 1, the BER of the source data at the destination for NOMA-DSM is compared with that of other full-duplex protocols. For the stronger source symbol $\hat{p}_1$, the results for the optimal and sub-optimal demodulators are similar, as they follow the same detection procedure as in point-to-point communication, and there is no cooperative link. The successive relaying protocol and the VFD-DSM protocol with Local MAP both perform 1.3 dB worse in BER compared to the detection quality of the stronger symbol $\hat{p}_1$, due to the existence of relay links as interference in full-duplex communications. However, the BER for the source symbol in the VFD-DSM protocol with Global MAP is slightly better than that of the stronger source symbol $\hat{p}_1$ in NOMA-DSM, partly because in the former case, the symbol detection procedure exploits the benefit of the cooperative link. However, for demodulation of the weaker source symbol p_2, unlike in VFD-DSM, it can be seen that the proposed scheme can achieve a diversity order of 2 (symbol p_2 is transmitted via the *index* of the activated relay as well as via the direct link). At a BER of 10^{-3}, this results in a gain of 3.7 dB over VFD-DSM with Global MAP and 5.5 dB over VFD-DSM with Local MAP.

The comparison of the BER of the relay data at the destination for both DSM and NOMA-DSM is shown in Fig. 2. Since in NOMA-DSM the activation of each relay is based on the detection of the weaker source symbol, the probability of erroneous relay activation for NOMA-DSM is higher than its counterpart in conventional DSM. This means that when the suboptimal low-complexity detector is employed, the significant spectral efficiency advantage of NOMA-DSM comes at the cost of a significantly reduced error rate performance in the relay data (a loss of approximately 5dB, as can be seen in Fig. 2), as this detector does not take into account the

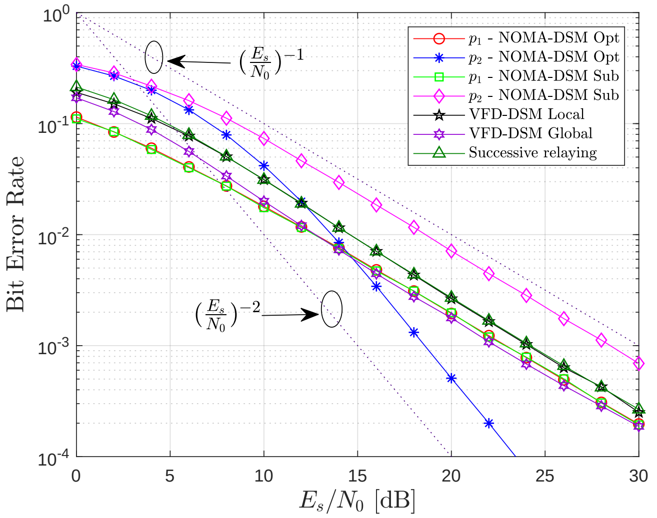

Fig. 2. BER of the source data at the destination, for the proposed NOMA-DSM and the benchmark full-duplex protocols. For NOMA-DSM, both the error-aware ML demodulator and the low-complexity demodulator are considered.

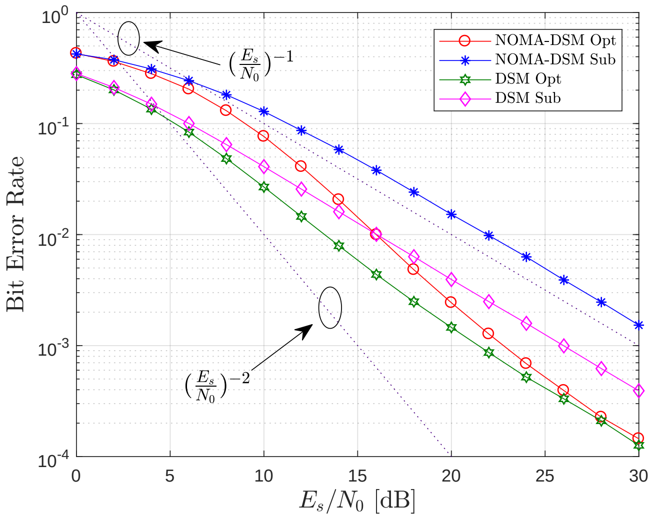

Fig. 3. BER of the relay data at the destination, for the proposed NOMA-DSM and DSM. For NOMA-DSM, both the error-aware ML demodulator and the low-complexity demodulator are considered.

possibility of demodulation errors at the relay. However, when the optimal error-aware demodulator is employed, the BER of the relay data in NOMA-DSM is close to that of conventional DSM at high SNR, while the throughput attained for the source data is double that of conventional DSM.

VI. Conclusion

In this paper, a new protocol was proposed for cooperative communication based on the association of distributed spatial modulation and non-orthogonal multiple access. The proposed scheme can increase the data rate of the source-to-destination link while also leveraging the DSM concept for relaying. An optimal error-aware demodulator as well as a low-complexity suboptimal demodulator were proposed for this context. Simulation results confirm that under optimal detection at the destination, the proposed NOMA-DSM protocol can significantly outperform the full-duplex relaying benchmarks of successive relaying and VFD-DSM, as well as conventional half-duplex DSM.

Acknowledgment

This work was supported by Science Foundation Ireland (grant no. 13/CDA/2199).

References

[1] Ezio Biglieri *et al*, *MIMO Wireless Communications*. Cambridge University Press, New York, NY, USA, 2007.
[2] 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.
[3] 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, July 2008.
[4] P. Yang, M. Di Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design guidelines for spatial modulation," *IEEE Commun. Surveys & Tuts.*, vol. 17, no. 1, pp. 6–26, First Quarter, 2015.
[5] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura and L. Hanzo, "Spatial modulation for generalized MIMO: challenges, opportunities, and implementation," *IEEE Proceeding.*, vol. 102, no. 1, pp. 56–103, Jan. 2014.
[6] Q. Wang, R. Zhang, L. Yang, and L. Hanzo. "Non-orthogonal multiple access: A unified perspective." *IEEE Wire. Comm.* vol 25, no. 2, pp. 10-16, April 2018.
[7] D. Wan, M. Wen, F. Ji, H. Yu and F. Chen, "Non-Orthogonal Multiple Access for Cooperative Communications: Challenges, Opportunities, and Trends," *IEEE Wire. Comm.*, vol. 25, no. 2, pp. 109-117, April 2018.
[8] X. Wang, J. Wang, L. He, Z. Tang and J. Song, "On the Achievable Spectral Efficiency of Spatial Modulation Aided Downlink Non-Orthogonal Multiple Access," *IEEE Comm. Letters,* vol. 21, no. 9, pp. 1937-1940, Sept. 2017.
[9] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communications in wireless networks", *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.
[10] J. N. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior", *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
[11] S. Narayanan, M. Di Renzo, F. Graziosi, and H. Haas, "Distributed spatial modulation for relay networks," *IEEE Veh. Technol. Conf.*, pp. 1–6, Sep. 2013.
[12] S. Narayanan, M. Di Renzo, F. Graziosi, and H. Haas, "Distributed spatial modulation: A cooperative diversity protocol for half-duplex relay-aided wireless networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2947-2964, May 2016.
[13] A. Shehni, M. F. Flanagan, "Network Coded Distributed Spatial Modulation for Relay Networks," *25th International Conference on Telecommunications (ICT)*, Saint-Malo, France, June 2018.
[14] A. Shehni, S. Narayanan and M. F. Flanagan. "A Virtual Full-Duplex Distributed Spatial Modulation Technique for Relay Networks," *Proc. 27th IEEE Annual Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Valencia, Spain, 4-7 September 2016.
[15] A. Shehni and M. F. Flanagan, "Virtual full-duplex distributed spatial modulation with SER-optimal and suboptimal detection," *IEEE WCNC conf.*, pp. 1-6, Barcelona, Spain, April 2018.
[16] Z. Ding, M. Peng and H. V. Poor, "Cooperative Non-Orthogonal Multiple Access in 5G Systems," *IEEE Comm. Letters*, vol. 19, no. 8, pp. 1462-1465, Aug. 2015.
[17] M. F. Kader, M. B. Shahab and S. Y. Shin, "Exploiting Non-Orthogonal Multiple Access in Cooperative Relay Sharing," *IEEE Comm. Letters*, vol. 21, no. 5, pp. 1159-1162, May 2017.
[18] Z. Ding, H. Dai and H. V. Poor, "Relay Selection for Cooperative NOMA," *IEEE Wire. Comm. Letters,* vol. 5, no. 4, pp. 416-419, Aug. 2016.
[19] T. Oechtering and A. Sezgin, "A new cooperative transmission scheme using the space-time delay code," in Proc. *ITG Workshop Smart Antenna*, pp. 41-48, Mar. 2004.
[20] S. Yang, J.-C. Belfiore, "On slotted amplify-and-forward cooperative diversity schemes,' in Proc. *2006 IEEE Intl. Sym. Inf. Theory*, July 2006.