

UC Irvine

UC Irvine Previously Published Works

Title

Trellis-Coded Non-Orthogonal Multiple Access.

Permalink

<https://escholarship.org/uc/item/4tr1j5qb>

Journal

IEEE Wirel. Commun. Lett., 9

Authors

Zou, Xun

Ganji, Mehdi

Jafarkhani, Hamid

Publication Date

2020

DOI

10.1109/LWC.2019.2961899

Peer reviewed

Trellis-Coded Non-Orthogonal Multiple Access

Xun Zou, *Student Member, IEEE*, Mehdi Ganji, *Student Member, IEEE*, and Hamid Jafarkhani, *Fellow, IEEE*

Abstract—In this letter, we propose a trellis-coded non-orthogonal multiple access (NOMA) scheme. The signals for different users are produced by trellis coded modulation (TCM) and then superimposed on different power levels. By interpreting the encoding process via the tensor product of trellises, we introduce a joint detection method based on the Viterbi algorithm. Then, we determine the optimal power allocation between the two users by maximizing the free distance of the tensor product trellis. Finally, we manifest that the trellis-coded NOMA outperforms the uncoded NOMA at high signal-to-noise ratio (SNR).

Index Terms—Non-orthogonal multiple access, trellis coded modulation, superposition coding.

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) is envisaged as one of the potential technologies in the next generation wireless communication systems. Users in NOMA systems can share the non-orthogonal resources, e.g., the frequency spectrum and the time slot. From a unified perspective, NOMA consists of code-domain NOMA and power-domain NOMA [1].

Both code-domain and power-domain NOMA have been extensively studied in the existing literature. In the power-domain NOMA systems, the signals of different users are assigned different powers. Then, one major challenge is the optimal power allocation as discussed, for example, in [2, 3]. The optimal power can be determined according to the channel conditions to maximize users' achievable rates. Superposition coding and successive interference cancellation (SIC) techniques are utilized at the transmitter and the receiver, respectively. Again, there are many studies on how to perform these techniques efficiently, for example [4, 5].

The code-domain NOMA has its origin in code division multiple access (CDMA), including sparse code multiple access (SCMA) [6] and trellis coded multiple access (TCMA) [7]. The signals of multiple users are separated by user-specific features, e.g., the uniquely assigned codeword of each user. In the code-domain NOMA, the main efforts are devoted to the multi-user detection, for example, the design of multidimensional constellations [6, 8]. To the best of our knowledge, the joint design of the code-domain and power-domain NOMA has never been studied.

In this work, we apply trellis coded modulation (TCM) to the power-domain NOMA, taking advantages of the coding gain and the power optimization. Utilizing superposition coding, the signals for multiple users are superimposed on different power levels. Compared with [8], the main contribution of this work is introducing the power allocation to code-domain

NOMA. The performance can be improved by allocating proper powers to the signals of different users. Instead of utilizing TCM purely for codeword design in [8], TCM is employed in this work to jointly optimize the error control coding and modulation. Therefore, the Viterbi algorithm can be directly applied to the proposed scheme. By interpreting the modulating process via the tensor product of trellises [9, 10], we implement the maximum likelihood sequence detection (MLSD) based on the Viterbi algorithm [11]. Furthermore, we derive the optimal power allocation between the two users by maximizing the free distance of the tensor product trellis.

The key difference between the trellis-coded NOMA and the traditional TCMA lies in the multiple access scheme. In TCMA, the signals of multiple users are differentiated by their unique features, for example, convolutional encoder, constellation, or interleaver [12]. However, in the trellis-coded NOMA, the signals are differentiated only by the power levels. Furthermore, for the first time, we provide insight into the power optimization for the superimposed TCM signals.

II. SYSTEM MODEL

In this letter, we consider a downlink NOMA system consisting of one base station (BS) and two users. Superposition coding is employed at the transmitter. The power allocated to User i 's signal is denoted as P_i , $i = 1, 2$. The channel coefficient between the BS and User i is represented by h_i . We adopt the block fading channel model, i.e., the channel remains static within each block and changes independently from one block to another [2, 3]. We assume that the channel state information is perfectly known by the BS and users. Without loss of generality, we assume that $|h_1|^2 > |h_2|^2$. To stipulate the user fairness, we set $P_2 > P_1$. In what follows, the 8-phase-shift keying (PSK) 4-state TCM serves as an example of TCM [13], which is depicted in Fig. 1. The trellis diagram and the 8-PSK mapping are shown in Figs. 2 (a) and (b), respectively. In Figs. 1 and 2, x_1 and x_2 represent the uncoded bits while z_0 and z_1 denote the coded bits via the convolutional encoder. For the sake of brevity, we employ the signal constellation with unit signal power, i.e., $E_b = 1$. Note that the proposed scheme can be applied to the case where two users employ different modulations/trellises and also the case of more than two users.

In the proposed trellis-coded NOMA, the signals for Users 1 and 2 are first modulated by TCM, as shown in Figs. 1 and 2, and then superimposed on different power levels. Using superposition coding, the n th transmitted symbol at the BS is given by $\sqrt{P_1}a_1(n) + \sqrt{P_2}a_2(n)$ where $a_i(n)$ is the n th symbol for User i after TCM. Then, the n th received sample at User i is given by

$$y_i(n) = h_i \left[\sqrt{P_1}a_1(n) + \sqrt{P_2}a_2(n) \right] + w_i(n), \quad (1)$$

This work was supported in part by the NSF Award CCF-1526780. The authors are with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, 92697 USA (email: {xzou4, mganji, hamidj}@uci.edu).

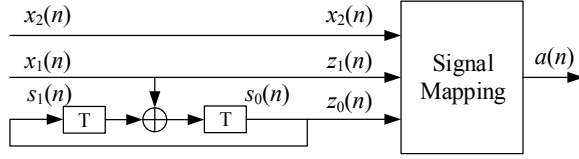


Fig. 1: Illustration of an 8-PSK 4-state TCM encoder.

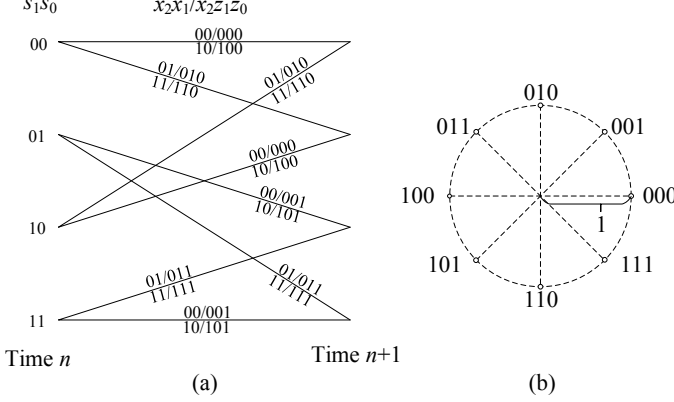


Fig. 2: (a) Trellis representation of 8-PSK 4-state TCM. (b) The mapping of 8-PSK constellation.

where $w_i(n) \sim \mathcal{CN}(0, \sigma_i^2)$ is the additive noise. At users, the modulated symbols are detected and then the binary information bits are recovered from the modulated symbols, which will be explained in the next section.

III. TENSOR PRODUCT OF TRELLISES AND DETECTION DESIGN

In this section, we first present the separate detection method with SIC. Then, we propose the joint detection method based on a novel trellis structure known as “tensor product of trellises”.

A. Separate Detection with SIC

In the separate detection scheme, the signals for Users 1 and 2 are detected separately. User 2 (the weak user) detects its own signal by considering User 1’s signal as noise. User 1 (the strong user) utilizes SIC, i.e., first detects User 2’s signal, removes it from the superimposed signal, and then detects its own signal. The Viterbi algorithm [11] can be employed to determine the sequence with the minimum Euclidean distance from the received sequence using the 4-state trellis in Fig. 2 (a).

B. Joint Detection with Tensor Product of Trellises

First, we review the concept of the tensor product of trellises [9, 10]. Let us consider trellises T_1 and T_2 with r_1 and r_2 states, respectively, and $S_i^{(l)}$, $i = 1, \dots, r_l$, denotes the i th state of T_l . The tensor product of T_1 and T_2 , denoted as $T_1 \otimes T_2$, can be represented as a trellis with $r_1 \times r_2$ states. Each state in $T_1 \otimes T_2$ is given by $S_i^{(1)} S_j^{(2)}$, $i = 1, \dots, r_1$, $j = 1, \dots, r_2$. The state transition from $S_i^{(1)} S_j^{(2)}$ to $S_k^{(1)} S_l^{(2)}$ exists if and only if there exist transitions from $S_i^{(1)}$ to $S_k^{(1)}$

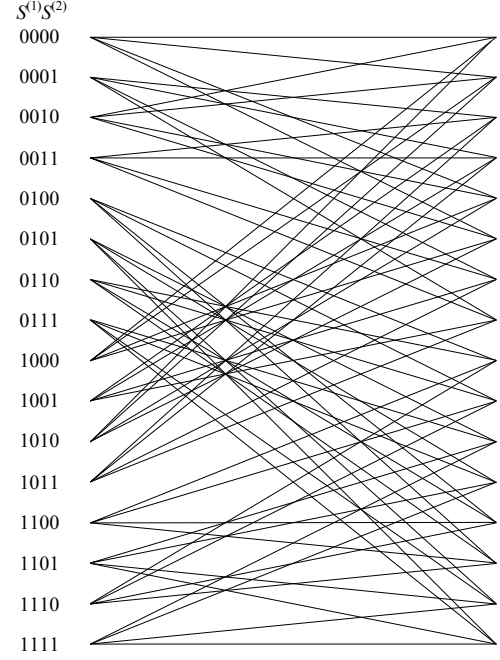


Fig. 3: Underlying tensor product of trellises.

in T_1 and from $S_j^{(2)}$ to $S_l^{(2)}$ in T_2 . One can easily extend the definition of the tensor product trellis to the case of more than two trellises.

Let us revisit the modulating process of two users’ signals in Section II. The symbols for Users 1 and 2 are modulated independently through the 4-state trellis, shown in Fig. 2 (a). Let T_1 and T_2 stand for the trellises employed to modulate the symbols for Users 1 and 2, respectively. The tensor product trellis $T_1 \otimes T_2$ is the 16-state trellis in Fig. 3. Every pair of state transitions in T_1 and T_2 can be represented by a unique transition path in $T_1 \otimes T_2$. For example, let us assume that the state of T_1 transits from $S_i^{(1)}$ to $S_k^{(1)}$ producing the modulated symbol a_1 and the state of T_2 transits from $S_j^{(2)}$ to $S_l^{(2)}$ generating the modulated symbol a_2 . From the perspective of $T_1 \otimes T_2$, the state transits from $S_i^{(1)} S_j^{(2)}$ to $S_k^{(1)} S_l^{(2)}$ and the superimposed symbol $\sqrt{P_1} a_1 + \sqrt{P_2} a_2$ is produced. Since every state transition can be realized by two parallel paths in T_1 and T_2 , as shown in Fig. 2 (a), every state transition in $T_1 \otimes T_2$ includes $2 \times 2 = 4$ parallel paths.

The description of the tensor product trellis demonstrates the equivalence of the trellis-coded NOMA and the TCM using the tensor product trellis. The joint detection is to detect both users’ signals jointly by treating the trellis-coded NOMA as a regular TCM with the tensor product trellis. In the joint detection, the Viterbi algorithm is implemented using the tensor product trellis. It is worth mentioning that there is no necessity to modulate the signals for Users 1 and 2 jointly using the tensor product trellis at the transmitter. The transmitted symbols for each user can be modulated independently according to its own trellis by applying an appropriate power allocation scheme to ensure a good decoding performance (as shown in Section V).

Since the Viterbi algorithm can be employed in joint decoding, the computational complexity increases linearly with

the number of decoded symbols, N . More specifically, if the number of states in T_i ($i = 1, 2$) is K_i and the total number of edges in T_i is L_i , the computational complexity of the joint detection method is given by $O(N(K_1K_2 + L_1L_2))$ while that of the separate detection method with SIC is $O(N(K_1 + K_2 + L_1 + L_2))$.

IV. POWER OPTIMIZATION

In this section, we study the power allocation to optimize the performance of the joint detection scheme. The power allocation is optimized under two power constraints. One is the sum power constraint, i.e., $P_1 + P_2 \leq P$ where P is the total transmit power. The other constraint is $P_1 < P_2$ which is added with no loss of generality. We adopt the free distance of the tensor product trellis, d_{free} , to measure the performance, which is widely used in the existing TCM studies, for example [14]. A larger free distance results in a better performance at high signal-to-noise ratio (SNR). As will be illustrated later, the free distance is a function of the power coefficients P_1 and P_2 . We obtain the optimal powers by maximizing the free distance.

The free distance is defined as the minimum Euclidean distance between any pair of valid and distinct sequences produced by a given trellis, i.e., $d_{\text{free}} = \arg \min_{\mathbf{a}_1, \mathbf{a}_2 \in V, \mathbf{a}_1 \neq \mathbf{a}_2} \|\mathbf{a}_1 - \mathbf{a}_2\|$ where V is the set of all valid sequences. The free distance can be determined by choosing the minimum of two candidates: the minimum Euclidean distance between the symbols produced by the parallel paths, i.e., d_{parallel} , and that between the sequences which diverge from the same state and then merge at the same state, i.e., $d_{\text{D\&M}}$. The subscript D\&M is the acronym for “diverging and merging”. In what follows, we analyze these two distances separately. Assume that there are two different paths in $T_1 \otimes T_2$ producing $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$ and $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$, where u_1 and u_2 are the modulated symbols of T_1 and v_1 and v_2 are those of T_2 .

A. Parallel Paths

First, we study the case where $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$ and $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$ are produced by the parallel paths in $T_1 \otimes T_2$. Fig. 4 illustrates the possible positions of $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$ and $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$ in the superimposed constellation when v_1 and v_2 are chosen from $\{1, -1\}$. Because of symmetry, the minimum Euclidean distance for all the other choices will be the same. In Fig. 4, there are four different markers, hollow/solid square/circle. The superimposed symbols depicted by the same marker are the symbols produced by the parallel paths for a specific state transition in $T_1 \otimes T_2$. Every state transition in $T_1 \otimes T_2$ can be realized by four parallel paths. Therefore, there are four positions for every marker. The minimum Euclidean distance between parallel paths can be found by calculating the Euclidean distance between the points sharing the same marker. It is clear from Fig. 4 that the minimum Euclidean distance is either δ_1 or δ_2 . Thus,

$$d_{\text{parallel}} = \min \{\delta_1, \delta_2\} = \min \left\{ 2\sqrt{P_2} - 2\sqrt{P_1}, 2\sqrt{P_1} \right\}. \quad (2)$$

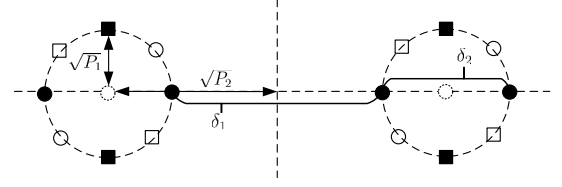


Fig. 4: Illustration of the minimum Euclidean distance in the superimposed constellation.

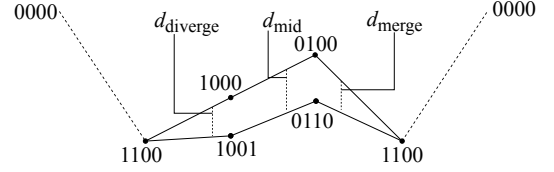


Fig. 5: Illustration of the diverging-and-merging paths with the minimum Euclidean distance.

B. Diverging-and-Merging Paths

Second, we study the Euclidean distance between the sequences which diverge from the same state and then merge at the same state. It can be shown that if two sequences diverge from any state, it takes at least three transitions to merge at the same state. We utilize the exhaustive search to find a pair of sequences with the minimum Euclidean distance among all pairs of distinct sequences, which is shown in Fig. 5. Note that all valid codewords start and end at state zero. However, any common sub-sequence will not contribute to d_{free} . Therefore, to calculate d_{free} in Fig. 5, we need to consider the state transitions $1100 \rightarrow 1000 \rightarrow 0100 \rightarrow 1100$ and $1100 \rightarrow 1001 \rightarrow 0110 \rightarrow 1100$. As shown in Fig. 5, the squared Euclidean distance between the diverging-and-merging paths is given by

$$d_{\text{D\&M}}^2 = d_{\text{diverge}}^2 + d_{\text{mid}}^2 + d_{\text{merge}}^2. \quad (3)$$

First, let us focus on the diverging paths in Fig. 5. According to Fig. 2, the superimposed symbol produced by the path $1100 \rightarrow 1000$ is given by $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$, where $u_1 \in \{e^{j3\pi/4}, e^{j7\pi/4}\}$ and $v_1 \in \{1, -1\}$. Similarly, the superimposed symbol produced by $1100 \rightarrow 1001$ is given by $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$, where $u_2 \in \{e^{j3\pi/4}, e^{j7\pi/4}\}$ and $v_2 \in \{e^{j\pi/2}, e^{j3\pi/2}\}$. The positions of the superimposed symbols can be shown in Fig. 6. The minimum Euclidean distance between the diverging paths is given by

$$d_{\text{diverge}} = \delta_3 = |\sqrt{2P_2} - 2\sqrt{P_1}|.$$

One can employ the same approach to derive the minimum Euclidean distance between the merging paths and find that $d_{\text{merge}} = d_{\text{diverge}}$.

Second, we investigate the Euclidean distance d_{mid} in Fig. 5. The superimposed symbol produced by the path $1000 \rightarrow 0100$ is given by $\sqrt{P_1}u_1 + \sqrt{P_2}v_1$, where $u_1 \in \{1, -1\}$ and $v_1 \in \{1, -1\}$. Similarly, the superimposed symbol produced by the path $1001 \rightarrow 0110$ is given by $\sqrt{P_1}u_2 + \sqrt{P_2}v_2$, where $u_2 \in \{1, -1\}$ and $v_2 \in \{e^{j\pi/4}, e^{j5\pi/4}\}$. The positions of the superimposed symbols can be shown in Fig. 7.

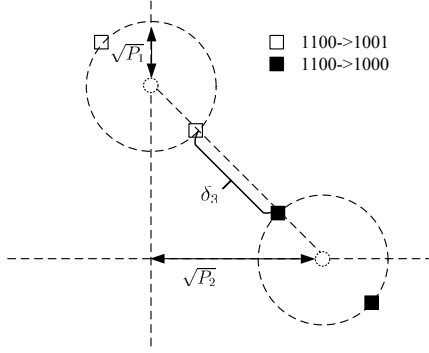


Fig. 6: Illustration of the minimum Euclidean distance between the symbols produced by diverging paths.

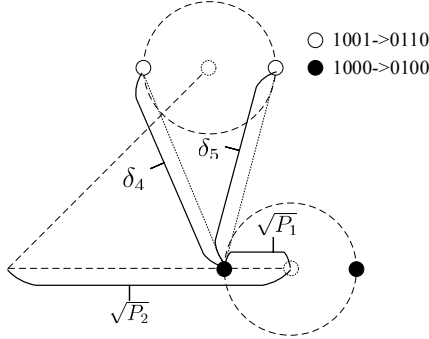


Fig. 7: Illustration of the minimum Euclidean distance between the symbols in the intermediate stage of the diverging-and-merging paths.

According to Fig. 7, the minimum Euclidean distance d_{mid} is given by

$$\begin{aligned} d_{\text{mid}}^2 &= \min\{\delta_4^2, \delta_5^2\} \\ &= (2 - \sqrt{2})P_2 + \min\left\{0, 4P_1 + 2\sqrt{P_1P_2}(\sqrt{2} - 2)\right\}. \end{aligned}$$

To summarize, the minimum Euclidean distance between the diverging-and-merging paths is given by

$$\begin{aligned} d_{\text{D\&M}}^2 &= d_{\text{diverge}}^2 + d_{\text{mid}}^2 + d_{\text{merge}}^2 \\ &= (6 - \sqrt{2})P_2 + 8P_1 - 8\sqrt{2P_1P_2} \\ &\quad + \min\left\{0, 4P_1 + 2\sqrt{P_1P_2}(\sqrt{2} - 2)\right\}. \end{aligned} \quad (4)$$

C. Free Distance

The free distance of $T_1 \otimes T_2$ is determined by finding the minimum of d_{parallel} and $d_{\text{D\&M}}$, i.e.,

$$\begin{aligned} d_{\text{free}}^2 &= \min\{d_{\text{parallel}}^2, d_{\text{D\&M}}^2\} \\ &= \min\left\{4P_1, 4\left(\sqrt{P_2} - \sqrt{P_1}\right)^2, (6 - \sqrt{2})P_2 + 8P_1 \right. \\ &\quad \left. - 8\sqrt{2P_1P_2} + \min\left\{0, 4P_1 + 2\sqrt{P_1P_2}(\sqrt{2} - 2)\right\}\right\}. \end{aligned} \quad (5)$$

The optimal powers can be derived by maximizing the free distance, i.e.,

$$[P_1^*, P_2^*] = \arg \max_{P_1, P_2} d_{\text{free}}^2, \text{ s.t. } P_1 + P_2 \leq P, \quad (6)$$

where P is the total transmit power. According to (5), one can derive that d_{free} is maximized when $4P_1 =$

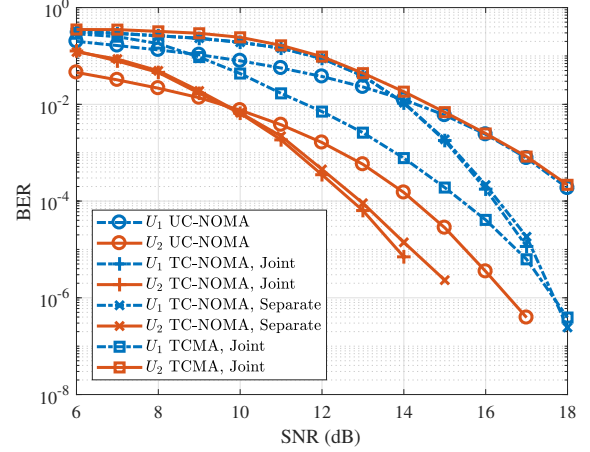


Fig. 8: BER vs. SNR for TCMA, uncoded and trellis-coded NOMA when $P_1 = 0.1$, $P_2 = 1$, $|h_1|^2 = 2$, $|h_2|^2 = 1$.

$(6 - \sqrt{2})P_2 + 8P_1 - 8\sqrt{2P_1P_2}$, which then results in $\frac{P_1^*}{P_2^*} = \left(\frac{2\sqrt{2} - \sqrt{2 + \sqrt{2}}}{2}\right)^2 \approx 0.2404$. Besides, to combat the channel noise, $P_1 + P_2$ should be maximized. As a result, $P_1^* = \frac{0.2404}{1 + 0.2404}P \approx 0.1938P$ and $P_2^* \approx 0.8062P$.

While we presented the results for a two-user scenario with 8-PSK 4-state TCM, our approach can be generalized to any TCM.

V. SIMULATION RESULTS

In this section, we present the simulation results of the 8-PSK 4-state trellis-coded NOMA (TC-NOMA), TCMA, and the uncoded NOMA (UC-NOMA) with 4-PSK. We ensure a fair comparison among these schemes since the TCM is implemented without consuming extra bandwidth compared with the uncoded modulation [13]. In our simulation, we employ bit error ratio (BER) as the measure of performance. In the uncoded NOMA, the maximum likelihood detection is employed. We also present the results for the TCMA where the signals for Users 1 and 2 are modulated by the identical trellis shown in Fig. 2 but differentiated by constellation [7]. In TCMA, the constellation used by one user is the other user's constellation rotated by $\pi/8$. In contrast to the trellis-coded NOMA, the transmitted signal in TCMA is given by $\sqrt{(P_1 + P_2)/2}[a_1(n) + a_2(n)]$, which ensures a fair comparison by using the same sum transmit power.

First, we show the BER as a function of SNR for NOMA and TCMA schemes in Figs. 8 and 9 when $P_2 = 1$ and $P_1 = 0.1$ or 0.3 , respectively. SNR is given by $\frac{1}{\sigma^2}$ where σ^2 is the variance of noise. For $P_1 = 0.1$ or 0.3 , it is manifested that at high-SNR, similar to conventional TCM [13], the trellis-coded NOMA using the joint detection outperforms the uncoded NOMA. Besides, the trellis-coded NOMA using the separate detection achieves a similar performance to that using the joint detection when $P_1 = 0.1$. In contrast, there is a huge gap between the BER curves of the separate detection and those of the joint detection when $P_1 = 0.3$. This is because of the severe inter-user interference when detecting two

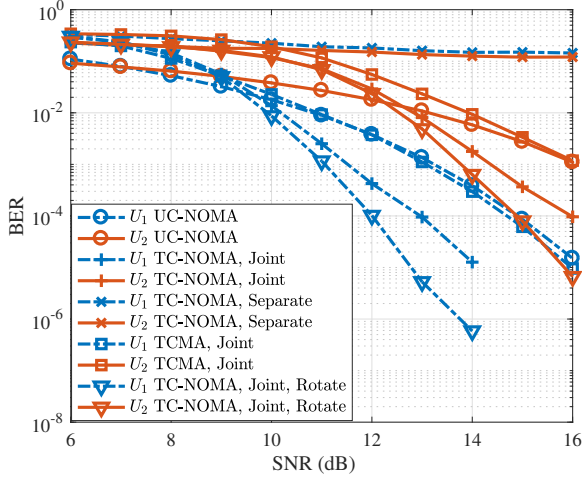


Fig. 9: BER vs. SNR for TCMA, uncoded and trellis-coded NOMA when $P_1 = 0.3$, $P_2 = 1$, $|h_1|^2 = 2$, $|h_2|^2 = 1$.

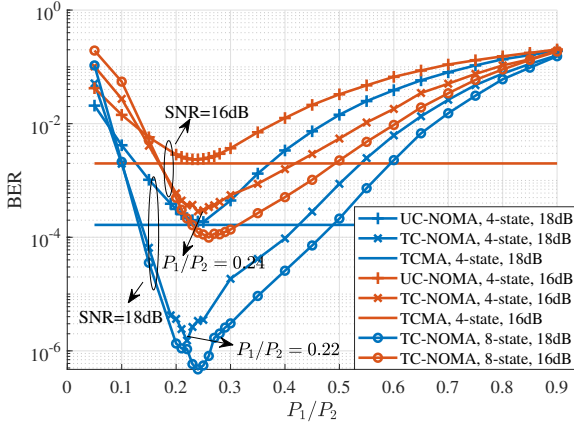


Fig. 10: BER vs. P_1/P_2 for TCMA, uncoded and trellis-coded NOMA schemes at users employing the joint detection when $P_1 + P_2 = 1$.

user's signals separately and the error propagation problem in SIC. Furthermore, using the joint detection, the trellis-coded NOMA outperforms TCMA at high-SNR in Figs. 8 and 9. Moreover, in the trellis-coded NOMA, the signals of different users can also employ different constellations. The curves with "TC-NOMA, Joint, Rotate" in Fig. 9 are for the case where the constellation used by one user is the other user's constellation rotated by $\pi/8$. The trellis-coded NOMA with constellation rotation achieves a better performance compared with the trellis-coded or uncoded NOMA without constellation rotation and TCMA. It can be explained intuitively by considering how the constellation rotation affects the Euclidean distance between superimposed symbols. According to Figs. 4 and 6, the minimum Euclidean distance may increase if the constellation of User 1's signal rotates by $\pi/8$, which then improves the performance.

Fig. 10 shows how the average BER changes with the power ratio P_1/P_2 using the joint detection at SNRs 16dB and 18dB for the 4-state trellis in Fig. 2 and the 8-state trellis in Fig. 12.8 of [14]. It is shown that the minimum BER is achieved when

$P_1/P_2 \approx 0.25$ for the uncoded NOMA and the 8-state trellis-coded NOMA. The optimal power ratio for the 4-state trellis-coded NOMA is 0.24 for SNR=16dB and 0.22 for SNR=18dB, which are close to the optimal power ratio of 0.2404 derived in Section IV. Moreover, the trellis-coded NOMA in its best case scenario outperforms the uncoded NOMA in its best case scenario. Besides, the performance of TCMA does not change with P_1/P_2 . By choosing the proper powers, the trellis-coded NOMA outperforms TCMA.

VI. CONCLUSIONS

In this letter, we study the trellis-coded NOMA and propose a joint detection method based on the tensor product of trellises. Besides, we derive the optimal power allocation between the two users by maximizing the free distance of the tensor product trellis. Simulation results demonstrate that the trellis-coded NOMA outperforms the uncoded NOMA and TCMA using an appropriate power allocation. The study of the trellis-coded NOMA systems with more than two users is our future work.

REFERENCES

- [1] Q. Wang, R. Zhang, L.-L. Yang, and L. Hanzo, "Non-orthogonal multiple access: A unified perspective," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 10–16, Apr. 2018.
- [2] X. Liu and H. Jafarkhani, "Downlink non-orthogonal multiple access with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 6151–6164, Sep. 2017.
- [3] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, Oct. 2016.
- [4] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2628–2639, Jul. 2012.
- [5] R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 503–520, Third Quarter 2011.
- [6] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE PIMRC*, London, UK, Sep. 2013, pp. 332–336.
- [7] T. Aulin and R. Espineira, "Trellis coded multiple access (TCMA)," in *Proc. IEEE ICC*, Vancouver, BC, Canada, Jun. 1999, pp. 1177–1181.
- [8] B. Di, L. Song, Y. Li, and G. Y. Li, "TCM-NOMA: Joint multi-user codeword design and detection in trellis coded modulation based NOMA for beyond 5G," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 766–780, Jun. 2019.
- [9] H. Jafarkhani and V. Tarokh, "Design of successively refinable trellis-coded quantizers," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1490–1497, Jul. 1999.
- [10] —, "Multiple description trellis-coded quantization," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 799 – 803, Jun. 1999.
- [11] A. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Trans. Comm. Tech.*, vol. 19, no. 5, pp. 751–772, Oct. 1971.
- [12] F. Brannstrom, T. M. Aulin, and L. K. Rasmussen, "Iterative detectors for trellis-code multiple-access," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1478–1485, Sep. 2002.
- [13] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets part I: Introduction," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 5–11, Feb. 1987.
- [14] S. Benedetto and E. Biglieri, *Principles of digital transmission: with wireless applications*. Springer Science & Business Media, 1999.