

A DOUBLE K-BEST VITERBI-SPHERE DECODER FOR TRELIS-CODED GENERALIZED SPATIAL MODULATION WITH MULTIPLE CODE RATES

Zhe-Yu Wang¹, Pei-Yun Tsai¹, Yuan-Hao Huang², and I-Wei Lai³

¹Department of Electrical Engineering, National Central University, Taiwan

²Department of Electrical Engineering, National Tsing Hua University, Taiwan

³Department of Electrical Engineering, National Taiwan Normal University, Taipei, Taiwan

ABSTRACT

The trellis-coded generalized spatial modulation (TCGSM) system with multiple code rates is investigated so as to provide configurability and to enhance efficiency in different fading channels. Various high-radix stages are then adopted to support the optimal sequence detection for different configurations. A double K-best technique is then proposed to strike a good balance between performance and complexity of Viterbi-sphere decoding. From the analysis and simulation results, 29%-69% complexity can be saved with about 1dB performance degradation for the code rate of 4/5 in the spatial domain with QPSK and 16-QAM constellations. The first K-best selection for generating branch metrics from sphere decoding can reduce the detection complexity of the constellation while the second K-best selection can be considered if a high-radix trellis is used when the number of transmit antennas is large.

Index Terms— K-best, sphere decoding, trellis-coded generalized spatial modulation (TCGSM), Viterbi decoding.

1. INTRODUCTION

Index modulation, which conveys digital data into on/off keying mechanism associated with the indexes of a certain building block in the communication systems, attracts much attention recently. Due to the new dimension it exploits, the spectral efficiency is increased. In addition, some element is deactivating for transmitting information, energy efficiency is improved without incurring more hardware complexity [1]. Thus, it is regarded as a promising modulation scheme for advanced communication systems.

Spatial modulation (SM), which utilizes antenna indexes to transmit digital data, is the most well-known style of index modulation. Conventional SM chooses single transmit antenna to carry information and thus avoids the inter-antenna interference. The optimal maximum likelihood (ML) symbol detection is performed in [2] and better system performance is achieved as opposed to the maximal-ratio-combining (MRC) detector. To upgrade the spectral efficiency, generalized spatial modulation (GSM) is introduced in [3]. More than one antenna can be activated in one time slot.

As the number of transmit or receive antennas increases that may entail a smaller antenna spacing at the transmitter or the receiver, spatial correlation takes place and it deteriorates the detection performance of the spatial modulation. Hence, trellis-coded spatial modulation (TCSM) [4] and trellis coded generalized spatial modulation [5] have been proposed. The information bits carried in the spatial domain are then encoded by a convolutional encoder. The ML symbol detector can be

used for detecting the spatially-multiplexed data. On the other hand, Viterbi decoder based on Hamming distance is employed for decoding the convolutional coded information bits in the spatial domain.

The sphere decoding algorithm and Viterbi algorithm are both famous for solving the spatial multiplexing problem and convolutional decoding problem, respectively. However, the integration of both algorithms is first proposed in [6] to detect the TCGSM that combines spatial multiplexing and trellis coded spatial modulation. This ML sequence detector is named Viterbi-Sphere decoder (VSD) [6]. The VSD calculating Euclidean distance outperforms the ML symbol detector. The sphere-decoding-embedded sequence detector can also be seen in [7] for detection of spatial permutation modulation.

In this paper, to offer flexibility and configurability, a TCGSM system with multiple convolutional code rates is proposed. Hence, it has the advantages of supporting different numbers of transmit antennas and various spectral efficiencies. In addition, it can combat fading channels with either low or high spatial correlations. To reduce the complexity of the VSD so as to fulfill the requirement of energy efficiency, a double K-best technique is also proposed. The K-best sphere decoding can be elegantly embedded in the K-best Viterbi decoder using high-radix trellis. The tradeoffs between the performance and complexity are discussed. The design of a low-complexity soft-value TCGSM detector is revealed and its feasibility is demonstrated.

2. SYSTEM MODEL OF TCGSM WITH MULTIPLE CODE RATES

Fig. 1 shows the block diagram of the TCGSM system. Input data in the n th time slot are parsed into $b_S(n)$ conveyed in the spatial domain and $b_C(n)$ for constellation mapping. The convolutional encoder generates outputs from $b_S(n)$. Instead of using one convolutional code rate in the conventional TCSM or TCGSM systems [4][5], multiple code rates are employed to provide high flexibility. Thus, the outputs are punctured according to the setting of code rate r . The encoded bits are then used as an index to select activated antennas $a(n) = \{a_{u,1}, \dots, a_{u,N_S}\}$ from the codebook, where u is the index. Given N_T transmit antennas, N_R receive antennas, and N_S M -QAM symbols to be transmitted, usually $N_T \geq N_R \geq N_S > 1$ for GSM with spatial multiplexing. The transmitted symbols are denoted by $q(n) = \{q_1, \dots, q_{N_S}\}$ with $E\{|q_i|^2\} = 1$ and $q_i \in \Omega$, the set of the constellation points. Let the set of the codebook for antenna mapping be Γ_r . Then, the cardinality of the codebook, $|\Gamma_r|$, must

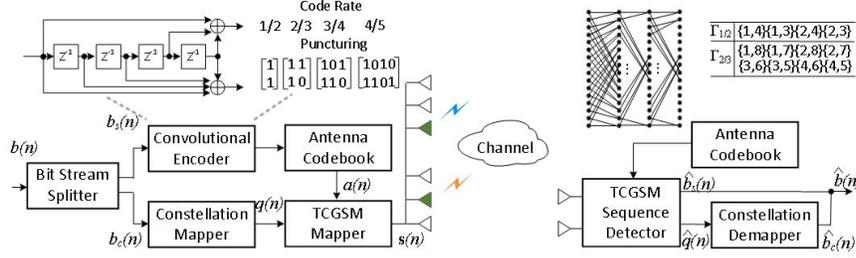


Fig. 1 System block diagram of TCGSM with punctured convolutional encoder for multiple code rates.

satisfy $|\Gamma_r| \leq \binom{N_T}{N_S}$, where $\binom{N_T}{N_S}$ means the number of N_S -combinations from N_T . The spectral efficiency of the TCGSM system can be expressed by

$$\eta = \eta_s + \eta_c = r \log_2 |\Gamma_r| + N_S \log_2 M. \quad (1)$$

To keep sparse nature of the index modulation or spatial modulation, N_S is set to 2 here. With punctured convolutional encoding, code rate r can be 1/2, 2/3, 3/4, or 4/5, and the corresponding codebook cardinality $|\Gamma_r|$ is 4, 8, 16, or 32, respectively. The number of transmit antennas can be varied from 4 to 16. Define

$$s_i(n) = \begin{cases} \frac{q_k}{\sqrt{N_S}} & i = a_{u,k}, \\ 0 & i \neq a_{u,k}, \end{cases} \quad \text{for } 1 \leq k \leq N_S, \quad (2)$$

where $s_i(n)$ is the i th element of vector $\mathbf{s}(n)$ of size $N_T \times 1$. Let $\mathbf{H}(n) = \mathbf{R}_R^{1/2} \mathbf{H}_{\text{IID}}(n) \mathbf{R}_T^{1/2}$. The elements of $\mathbf{H}_{\text{IID}}(n)$ are independent and identically-distributed Gaussian random variables. Matrixes \mathbf{R}_R and \mathbf{R}_T describe spatial correlation at the receiver and transmitter, which take the form of [8]

$$\mathbf{R}_Z = \begin{bmatrix} 1 & c & c^2 & \dots & c^{(N_Z-1)^2} \\ c & 1 & c & \dots & c^{(N_Z-2)^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c^{(N_Z-1)^2} & c^{(N_Z-2)^2} & \dots & \dots & 1 \end{bmatrix} \quad (3)$$

for $Z = \{R, T\}$. The received signal is described by

$$\mathbf{y}(n) = \mathbf{H}(n)\mathbf{s}(n) + \mathbf{v}(n). \quad (4)$$

3. OPTIMAL VSD WITH MULTIPLE CODE RATES

Define P as the sequence length and

$$\Phi(\bar{\mathbf{s}}(n)) = \sum_{m=n-P+1}^n \|\mathbf{y}(m) - \mathbf{H}(m)\mathbf{s}(m)\|^2, \quad (5)$$

where $\bar{\mathbf{s}}(n) = [\mathbf{s}(n-P+1) \dots \mathbf{s}(n-1) \mathbf{s}(n)]^T$. By QR decomposition, $\mathbf{H}_{[u]}(m) = \mathbf{Q}_u(m)\mathbf{R}_u(m)$, where $\mathbf{H}_{[u]}(m)$ denotes the column selection by the set specified by codebook index u . Let $\mathbf{d}_u(m) = [s_{a_{u,1}}(m) \dots s_{a_{u,N_S}}(m)]^T$. Given $\mathbf{z}_u(m) = \mathbf{Q}_u^H(m)\mathbf{y}(m)$, (5) can be rewritten as

$$\begin{aligned} \Phi(\bar{\mathbf{s}}(n)) &= \sum_{m=n-P+1}^n \|\mathbf{z}_u(m) - \mathbf{R}_u(m)\mathbf{d}_u(m)\|^2 \\ &= \sum_{m=n-P+1}^n \left(\sum_{i=1}^{N_S} |z_{u,i}(m) - \sum_{k=1}^{N_S} r_{u,i,k}(m)d_{u,k}(m)|^2 \right), \end{aligned} \quad (6)$$

where $z_{u,i}(m)$ and $d_{u,k}(m)$ are the i th entry and k th entry of $\mathbf{z}_u(m)$ and $\mathbf{d}_u(m) \in \Omega^{N_S \times 1}$, respectively, and $r_{u,i,k}(m)$ is the (i,k) th element of $\mathbf{R}_u(m)$. The maximum likelihood (ML) sequence detector finds

$$[\hat{\mathbf{u}}, \hat{\mathbf{d}}] = \arg \min_{\mathbf{u} \in \Gamma^{1 \times P}, \mathbf{d} \in \Omega^{N_S \times P}} \Phi(\bar{\mathbf{s}}(n)), \quad (7)$$

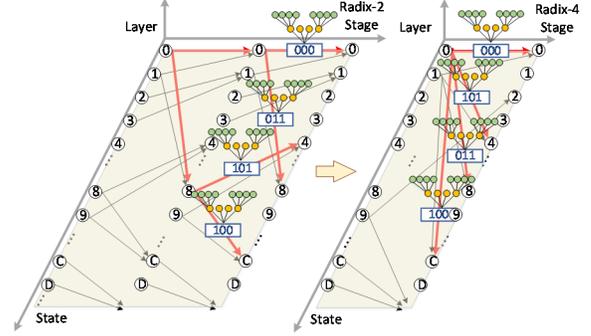


Fig. 2 Tree-embedded trellis with radix-2 and radix-4 stages.

where $\hat{\mathbf{u}}$ and $\hat{\mathbf{d}}$ are the detected results of P time slots. From (6), the sphere decoding related with the inner summation and Viterbi decoding associated with the outer summation can be integrated, named VSD, to detect TCGSM signals [6]. Here, we will extend the VSD in [6] to the double K-best VSD that can support multiple code rates with reduced complexity. To the best of our knowledge, the performance and complexity of double K-best VSD with multiple code rates for TCGSM systems has not been investigated in literature. The advantages of multiple convolutional code rates for TCGSM systems are twofold. When the channel is highly spatially-correlated, lower code rates can protect the spatial-domain data to improve the bit error rate. If the spatial correlation is low, high code rate can increase the spectral efficiency.

Since the sphere decoding is combined with the Viterbi decoding, one time slot is regarded as a basic unit for detection. Hence, high-radix trellis is used for solving the detection problem with the punctured convolutional codes or large antenna-codebook sizes adopted in the TCGSM system. Fig. 2 shows an example of tree-embedded trellis with code rate 2/3. The left part is the original radix-2 trellis diagram with two possible state transitions from each state. However, the 4-bit outputs of two stages are punctured and only 3 bits are reserved for antenna-codebook. The trees appear only at the last stage. Note that the index of the tree along the branch represents the associated set $a(n)$ from the antenna codebook for computing Euclidean distance by the sphere decoding. The right part is the radix-4 trellis diagram. Although there are 4×16 branches at one radix-4 stage, only 8 different branch metrics generated from the sphere decoders are required owing to that the cardinality of Γ_r for $r = 2/3$ is 8. Thus, unlike the conventional convolutional decoder, the number of required branch metrics (BMs) per trellis stage corresponding to one time slot of the VSD is equal to the cardinality of Γ_r . Then the entire distance of the route is

accumulated as the path metric (PM) and the Compare-and-Select unit (CSU) decides the survival path for each node. Table I summarizes the required BMs, PMs, and the CSU configurations of the VSD given different code rates. The spectral efficiency in each case is also shown. The minimum requirements of the number of transmit antennas in power of 2 are also provided. Hence, the flexibility of the wide range of the spectral efficiency from 5 to 12 bps/Hz of the TCGSM system with multiple code rates is demonstrated

Table I List of the configurations and spectral efficiencies of VSD for TCGSM with different code rates.

Code Rate (r)	1/2	2/3	3/4	4/5
No. of Ant.	4	8	8	16
$ \Gamma_r $	4	8	16	32
Radix	Radix-2	Radix-4	Radix-8	Radix-16
No. of BMs	4	8	16	32
No. of PMs	32	64	128	256
CSU	2-to-1	4-to-1	8-to-1	16-to-1
η (QPSK) (bps/Hz)	5	6	7	8
η (16-QAM) (bps/Hz)	9	10	11	12

Fig. 3 shows the performance of the TCGSM system with multiple code rates. The numbers of transmit antennas are set as the ones in Table I for the respective code rates. Two received antennas are used. The channel model is spatially-correlated with $c = 0.8$. The channel remains unchanged for the period of the whole sequence P , which is at least five times the constraint length of the convolutional code, in slow-fading cases, but changes independently for each time-slot in fast fading cases to test two extreme conditions. If 16-QAM is used, the average bit-error-rate (BER) performance is dominated by the constellation domain instead of the spatial domain. In addition, more bits are conveyed in the spatial domain for $r = 4/5$. Thus, the average BER curve is influenced more by the spatial modulation. However, if QPSK is used with $r = 4/5$ in spatially-correlated slow-fading channels, the BER in the spatial domain is worse than the BER in the constellation domain.

4. DOUBLE K-BEST VSD WITH MULTIPLE CODE RATES

If the number of transmit antennas increases, the size of the codebook for antenna mapping will grow. From Table I, we can see that the complexity gets higher given a high code rate or a large number of antennas. Consequently, the complexity reduction is necessary to achieve the energy-efficient goal of spatial modulation. The K-best sphere decoding is widely adopted due to its advantages of reduced search space and constrained complexity for MIMO detection problem [9]. On the other hand, the K-best Viterbi decoder has also been proposed in [10] to reduce power consumption and silicon area. Therefore, we propose a double K-best VSD, which combines K-best sphere decoding and K-best Viterbi decoding. In addition, ℓ^1 -norm is used instead of ℓ^2 -norm for Euclidean distance approximation to eliminate the multiplications. Real-value decomposition (RVD) and Schnorr-Euchner (SE) enumeration [11] are also adopted to facilitate the implementation regarding the search complexity. In the following, the tradeoffs between

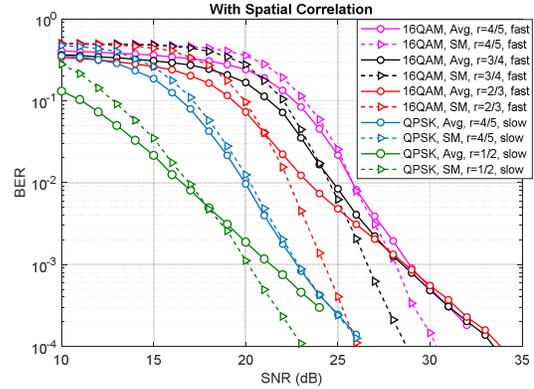


Fig. 3 Performance of TCGSM with multiple code rates in spatially-correlated fast-fading channels given $N_S = 2$, $N_R = 2$, and $c = 0.8$.

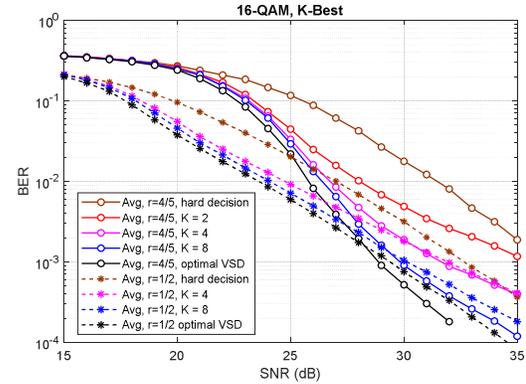


Fig. 4 Performance comparison of VSD with K-best sphere decoding in spatially-correlated fast-fading channels given $N_S = 2$, $N_R = 2$, and $c = 0.8$.

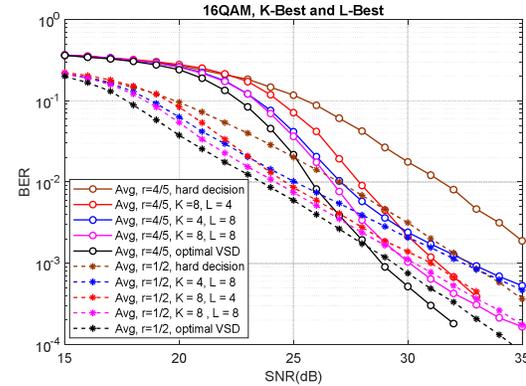


Fig. 5 Performance comparison of VSD with K-best sphere decoding and L-best Viterbi decoding in spatially-correlated fast-fading channels given $N_S = 2$, $N_R = 2$, and $c = 0.8$.

the performance and complexity resulted from these techniques are evaluated and discussed for the TCGSM systems.

To make it clear, we use parameter K for K-best sphere decoding and parameter L for L-best Viterbi decoding. Fig. 4 shows the comparison results for different K values but a complete trellis. ℓ^1 -norm is used together with the K-best

technique. The performances of the ML symbol detector with ℓ^1 -norm approximation and hard-decision Viterbi decoding [5] is also given, which is an upper bound for the selection of K values. On the other hand, the optimal VSD indicates the lower bound of the BER performance. The numbers of transmit antennas for $r = 4/5$ and $r = 1/2$ are 16 and 4, respectively. Thus, the slopes of the BER curves are a little different. The performance of hard-decision Viterbi decoding is worse if the code rate is higher. For 16-QAM, $K = 8$ and ℓ^1 -norm result in the degradation less than 1 dB at the BER of 10^{-3} compared to the optimal VSD. For QPSK, we can use VSD with 4-best sphere decoding which causes small degradation. In Fig. 5, we show the combination of the K-best sphere decoding and L-best Viterbi decoding for detection of TCGSM signals. For 16-QAM, $K = 8$ and $L = 8$ can be considered, additional degradation of about 0.2dB is generated if we compare the curves with those of $K = 8$ in Fig. 4. Furthermore, the degradation caused by a small K value is more severe than the one caused by a small L value in the high SNR region where the BER is usually dominated by the constellation domain rather than the spatial domain.

The computation complexity of the double K-best VSD is then investigated so as to tradeoff the performance and complexity for implementation. Given K-best sphere decoding with ℓ^1 -norm and RVD, the tree has $2N_s$ layers and each parent node has \sqrt{M} child nodes. The second summation in (6) is changed to

$$\sum_{i=1}^{2N_s} |\tilde{z}_{u,i}(m) - \sum_{k=i}^{2N_s} \tilde{r}_{u,i,k}(m) \tilde{d}_{u,k}(m)| \quad (8)$$

where $\tilde{z}_{u,i}(m)$, $\tilde{r}_{u,i,k}(m)$, and $\tilde{d}_{u,k}(m)$ are the elements from the corresponding matrix after RVD of $\mathbf{z}_u(m) - \mathbf{R}_u(m)\mathbf{d}_u(m)$. The tree search starts from $i = 2N_s$. Define the number of reserved nodes in layer i of the tree as

$$D_i = \min(D_{i+1}\sqrt{M}, K) \quad (9)$$

with $D_{2N_s+1} = 1$ corresponding to the root node. If $D_{i+1}\sqrt{M} > K$, then D_{i+1} enumeration units are needed, each demanding two binary slicers for 16-QAM ($\beta = 2$) and using simply sign bit to enumerate for QPSK ($\beta = 0$). Furthermore, with on-demand expansion, only partial Euclidean distances (PEDs) of $(D_{i+1} + D_i - 1)$ nodes are calculated and are expanded from the previous D_{i+1} survival nodes of the tree [9]. In addition, we only need to pre-calculate the residual component, $\tilde{z}_{u,i-1}(m) - \sum_{k=i}^{2N_s} \tilde{r}_{u,i,k}(m) \tilde{d}_{u,k}(m)$, of the D_i survival nodes for PED calculation in the subsequent layer $i - 1$. Then, the K-best selection needs $[(D_{i+1} - 1) + (D_i - 1)\log_2(D_{i+1})]$ comparisons. On the other hand, if $D_{i+1}\sqrt{M} \leq K$, then PEDs of all $D_{i+1}\sqrt{M}$ nodes are required and enumeration is not necessary as well as the K-best selection. Since $\tilde{d}_{u,k} \in \{-\sqrt{M} + 1, \dots, 1, 3, \dots, \sqrt{M} - 1\}$, $\tilde{r}_{u,i,k}(m) \tilde{d}_{u,k}(m)$ can be computed by α shift-and-add operations and can be shared for all PED calculations of the same layer. Thus, $\alpha = 0$ for QPSK and $\alpha = 1$ for 16-QAM, respectively. The computation complexity in terms of additions/subtractions associated with K-best sphere decoding for one BM is summarized in Table II.

The introduction of the L-best technique in the trellis reduces the number of reserved PMs and changes the number of inputs for comparison in the CSU. If l stages are merged for detection, namely radix- 2^l trellis, there will be 16×2^l PMs and $16 \cdot 2^l$ -to-1 selections by CSU at each radix- 2^l stage. When only L best nodes in the trellis are reserved, $L \times 2^l$ PMs and $16 \cdot L$ -to-1

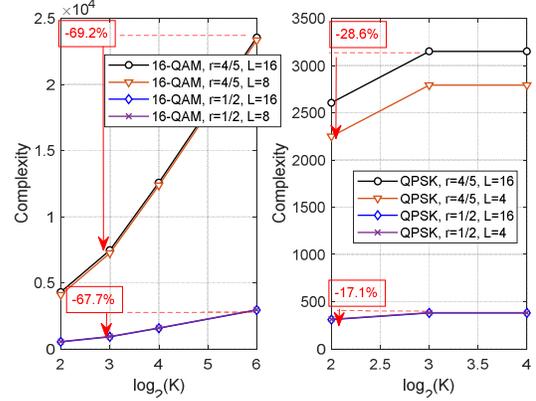


Fig. 6 Complexity reduction of the double K-best techniques

Table II Complexity evaluation with K-best and L-best selection.

K-best ($D_{i+1}\sqrt{M} > K$) for 1 BM (No. of Add./Sub.)	
PED Calculation	$2(D_{i+1} + D_i - 1) + \alpha$ $+ (2N_s - i + 1)D_i + \alpha(2N_s - i + 1)$
Enumeration	βD_i
K-best Selection	$(D_{i+1} - 1) + (D_i - 1)\log_2(D_{i+1})$
L-best for radix- 2^l trellis (No. of Add./Sub.)	
PM Accumulation	$L \times 2^l$
CSU	$\min(16(2^l - 1), (16(L - 1)))$
L-best Selection	$15 + 4(L - 1)$

comparisons are required in the worst case. However, additional 16-to- L sorting must be taken into consideration for determining the survival nodes in the trellis. The complexity is also listed in Table II.

Fig. 6 shows the complexities reduction achieved by K-best and L-best techniques for different code rates and transmit antenna numbers of the TCGSM system. Note that Viterbi decoding has linear complexity of the finite states. The net reduction will not always be positive if the overhead of the L-best selection is considered. Thus, when radix-8 or radix-16 trellis is adopted, the L-best technique for complexity reduction can be taken into consideration. However, the K-best technique brings substantial advantages. For TCGSM of $r = 4/5$ and constellation of 16-QAM, the complexity reduction is 69.2% with $(K, L) = (8, 8)$ compared to the one with $(K, L) = (64, 16)$, which is almost the full complexity of the VSD.

5. CONCLUSION

In this paper, the TCGSM system with multiple code rates is designed to offer flexibility and configurability in different fading channels. The high-radix stages are used in the trellis to support the decoding variety and the growth of the transmit antenna numbers for the optimal sequence detector that combines Viterbi decoding and sphere decoding. A double K-best VSD is then proposed to reduce the complexity so as to achieve the goal of energy efficiency of the spatial modulation or index modulation systems. From simulation results, 8-best is suitable for 16QAM and 4-best is sufficient for QPSK detection with small performance degradation but significant complexity reduction. The L-best is helpful when a high-radix stage is adopted in the trellis. A complexity saving of 69% can be attained by the double K-best technique for TCGSM with $r = 4/5$ and 16-QAM.

REFERENCES

- [1] E. Basar1, et al. "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693-16746. 2017.
- [2] 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.
- [3] J. Wang, S. Jia, J. Song, "Generalized spatial modulation system with multiple active transmit antennas and low complexity detection scheme," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1605-1615, Apr. 2012.
- [4] R. Mesleh, M. D. Renzo, H. Hass, and P. M. Grant, "Trellis coded spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2349-2361, Jul. 2010.
- [5] Y. Zhou, D. Yuan, X. Zhou, and H. Zhang, "Trellis coded generalized spatial modulation," in *2014 IEEE 79th Veh. Technol. Conf.*, Seoul, Korea, 2014.
- [6] S. S. Long, P. Y. Tsai, Y. H. Huang, and I. W. Lai, "Trellis coded generalized spatial modulation with spatial multiplexing," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2017, pp. 832-837.
- [7] J. C. Chi, Y. C. Yeh, I. W. Lai and Y. H. Huang, "Sphere decoding for spatial permutation modulation MIMO Systems," *IEEE International Conference on Communications (ICC)*, 2017, pp. 1-8.
- [8] A. V. Zelst and J. S. Hammerschmidt, "A single coefficient spatial correlation model for multiple-input multiple-output (MIMO) radio channels," in *Proc. of 27th General Assembly of the URSI*, pp. 1-4, 2002.
- [9] M. Y. Huang, and P. Y. Tsai, "Toward multi-gigabit wireless: design of high-throughput MIMO detectors with hardware efficient architecture," *IEEE Trans. on Circ. and Sys. I: Regular Paper*, vol. 61, Iss: 2, pp. 613-624, 2014.
- [10] H. Kato, T. H. Tran, and Y. Nakashima "ASIC design of a low-complexity K-best viterbi decoder for IoT applications," *Asian Pacific Conference on Circuits and Systems, (APCCAS)* 2016, pp. 396-399.
- [11] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Mathematical Programming*, vol. 66, pp. 181-199, 1994.