IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 26, 2017, accepted May 17, 2017, date of publication May 23, 2017, date of current version October 12, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2707182

Two Decades of MIMO Design Tradeoffs and Reduced-Complexity MIMO Detection in Near-Capacity Systems

CHAO XU[1], (Member, IEEE), SHINYA SUGIURA[2], (Senior Member, IEEE), SOON XIN NG[1], (Senior Member, IEEE), PEICHANG ZHANG[3], (Member, IEEE), LI WANG[4], (Member, IEEE), AND LAJOS HANZO[1], (Fellow, IEEE)

[1]School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.
[2]Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan
[3]College of Information Engineering, Shenzhen University, Shenzhen 518060, China
[4]Huawei Technology Sweden Research and Development Competence Center, 169 40 Stockholm, Sweden

Corresponding author: Lajos Hanzo (lh@ecs.soton.ac.uk)

This work was supported by the European Research Council's Advanced Fellow Grant. The work of S. Sugiura was supported by the Japan Society for the Promotion of Science KAKENHI under Grant 26709028 and Grant 16KK0120. The research data of this paper can be found at the DOI 10.5258/SOTON/D0254<https://ww w.outlook.soton.ac.uk/owa/UrlBlockedError.aspx>

ABSTRACT A pair of salient tradeoffs have driven the multiple-input multiple-output (MIMO) systems developments. More explicitly, the early era of MIMO developments was predominantly motivated by the multiplexing-diversity tradeoff between the Bell Laboratories layered space-time and space-time block coding. Later, the linear dispersion code concept was introduced to strike a flexible tradeoff. The more recent MIMO system designs were motivated by the performance-complexity tradeoff, where the spatial modulation and space-time shift keying concepts eliminate the problem of inter-antenna interference and perform well with the aid of low-complexity linear receivers without imposing a substantial performance loss on generic maximum-likelihood/max *a posteriori* -aided MIMO detection. Against the background of the MIMO design tradeoffs in both uncoded and coded MIMO systems, in this treatise, we offer a comprehensive survey of MIMO detectors ranging from hard decision to soft decision. The soft-decision MIMO detectors play a pivotal role in approaching to the full-performance potential promised by the MIMO capacity theorem. In the near-capacity system design, the soft-decision MIMO detection dominates the total complexity, because all the MIMO signal combinations have to be examined, when both the channel's output signal and the *a priori* log-likelihood ratios gleaned from the channel decoder are taken into account. Against this background, we provide reduced-complexity design guidelines, which are conceived for a wide-range of soft-decision MIMO detectors.

INDEX TERMS MIMO design tradeoffs, soft-decision detectors, near-capacity systems, reduced-complexity design.

I. INTRODUCTION

The technical breakthrough of Turbo Codes (TCs) [1], [2] has initiated two decades of exciting developments, leading to a suite of near-capacity tranceiver techniques [3]–[12]. Moreover, the recent developments in the millimeter-wave band [13]–[15] facilitate the employment of a large number of antennas, especially at the Base Station (BS) [15]–[18]. Driven by the growing demand for more advanced wireless communication technologies, in line with Moor's Law, wireless communications systems have gradually become more and more complex. Fig. 1 offers a glimpse of a few key factors that directly affect the design of wireless communications systems. The nomenclature of terminologies used in this paper are summarized in Table 1. The factors in the first category of system modelling seen in Fig. 1 play a fundamental role in efficient system planing and deployment. Once the system model is established, the transceiver design featured in Fig. 1 revolves around achieving the best possible throughput versus BER performance of the second category at the lowest delay and complexity of the third category. Invariably, there is a tradeoff between the performance attained and the complexity imposed, since a complexity reduction is often associated with a performance degradation.

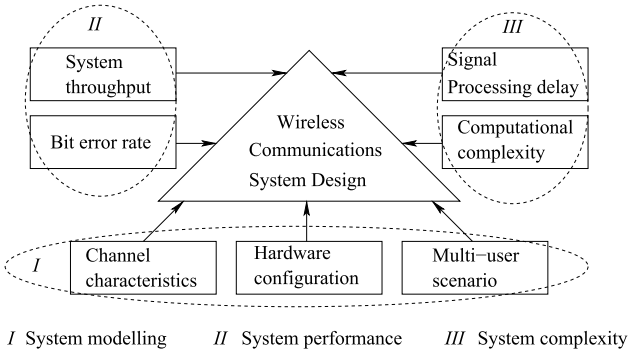


FIGURE 1. Factors affecting the design of wireless communications systems.

As an example, the classic V-BLAST MIMO system is portrayed in Fig. 2, where both the transmitter and the receiver are equipped with multiple antennnas. The N_T Transmit Antenna (TA) elements independently transmit a total number of N_T modulated symbols, which are drawn from the MPSK constellation diagram. The N_T data streams experience fading channels and arrive at the Receive Antenna (RA) elements simultaneously. As a result, the classic Maximum-Likelihood (ML) V-BLAST MIMO receiver [12] of Fig. 2 has to jointly consider all the N_T MPSK constellation diagrams, which imposes a potentially excessive computational complexity that grows exponentially with N_T. In order to mitigate this complexity problem, it is desirable to visit the individual MPSK constellation diagrams separately. However, in practice, Inter-Antenna Interference (IAI) is encountered, because the multiple data streams act as interference imposed on each other. An attractive option is to invoke a Sphere Decoder (SD) [19]–[21] as seen in Fig. 2, which only detects a single symbol at a time, while the previous decisions made by visiting other constellation diagrams are fed back in order to cancel out the known interference. The SD may continue to examine new constellation points of the next constellation diagram, until the search scope exceeds the SNR-dependent sphere radius. Therefore, the performance and complexity of SD is explicitly determined by the sphere radius, where the ML performance may be retained at the cost of a high complexity, whilst visiting less candidates may result in a degraded performance. Another option is to mitigate the IAI by a Linear Filter (LF) [22]–[25], and then the individual constellation digrams may be visited completely separately, which results in a substantially reduced complexity that grows only linearly with N_T. Nonetheless, the residual IAI after LF may still severely degrade the MIMO system's performance.

In this paper, we pay special attention to the important tradeoff between the performance and complexity. We design reduced-complexity algorithms that are tailored for near-capacity communications systems. The basic philosophy of reduced-complexity design is illustrated by the example of SD seen in Fig. 2, where the complex detector may be decomposed into steps so that less decision candidates have to be considered. Moreover, the interaction between the detection steps should be carefully taken into account, so that the optimum full-search-based performance may be retained.

The performance versus complexity tradeoff also plays salient role in MIMO system design. Recently, it has motivated the development of Spatial Modulation (SM) [26], [27], which has been considered as an attractive candidate for large-scale MIMO systems [15], [28]. In more detail, the first era of MIMO development was driven by the classic tradeoff between the attainable multiplexing and diversity gain [29]. The V-BLAST MIMO systems [30]–[32] have a capacity

TABLE 1. Nomenclature.

AO-STBC	Amicable Orthogonal Space-Time Block Coding
BCJR	Bahl-Cocke-Jelinek-Raviv
BICM-ID	Bit-Interleaved Coded Modulation relying on Iterative Decoding
BLAST	Bell Laboratories Layered Space-Time
BS	Base Station
CCMC	Continuous-input Continuous-output Memoryless Channel
CSI	Channel State Information
DCMC	Discrete-input Continuous-output Memoryless Channel
DSTM	Differential Space-Time Modulation
EXIT	EXtrinsic Information Transfer
GSM	Generalized Spatial Modulation
GSSK	Generalized Space Shift Keying
GSTSK	Generalized Space-Time Shift Keying
HR-STBC	Half-Rate Space-Time Block Coding
IAI	Inter-Antenna Interference
IRCC	IRregular Convolutional Code
LDC	Linear Dispersion Code
LDPC	Low-Density Parity Check
LF	Linear Filter
LLR	Log Likelihood Ratio
MAP	Max A Posteriori
MF	Matched Filter
MIMO	Multiple-Input Multiple Output
ML	Maximum-Likelihood
MLC	Multi-Level Coding
MLSE	Maximum Likelihood Sequence Estimation
MMSE	Minimum Mean Squared Error
MRC	Maximum Ratio Combining
MSE	Mean Squared Error
MUD	Multi-User Detection
PCC	Parallel Concatenated Code
PDF	Probability Density Function
PED	Partial Euclidean Distance
PEP	Pairwise Error Probability
PSED	Pairwise Squared Euclidean Distance
QO-STBC	Quasi-Orthogonal Space-Time Block Coding
QS	Quasi-Static
RA	Receive Antenna
RS	Reed-Solomon
RSC	Recursive Convolutional Code
RTS	Repeated Tree Search
SCC	Serial Concatenated Code
SD	Sphere Decoder
SDMA	Space-Division Multiple Access
SIC	Successive Interference Cancelling
SIMO	Single-Input Multiple-Output
SISO	Single-Input Single-Output
SM	Spatial Modulation
SSK	Space-Shift Keying
STBC	Space-Time Block Coding
STM	Space-Time Modulation
STS	Single Tree Search
STSK	Space-Time Shift Keying
TA	Transmit Antenna
TC	Turbo Code
TCM	Trellis Coded Modulation
URC	Unity Rate Code
VLSI	Very-Large-Scale Integration
ZF	Zero-Forcing

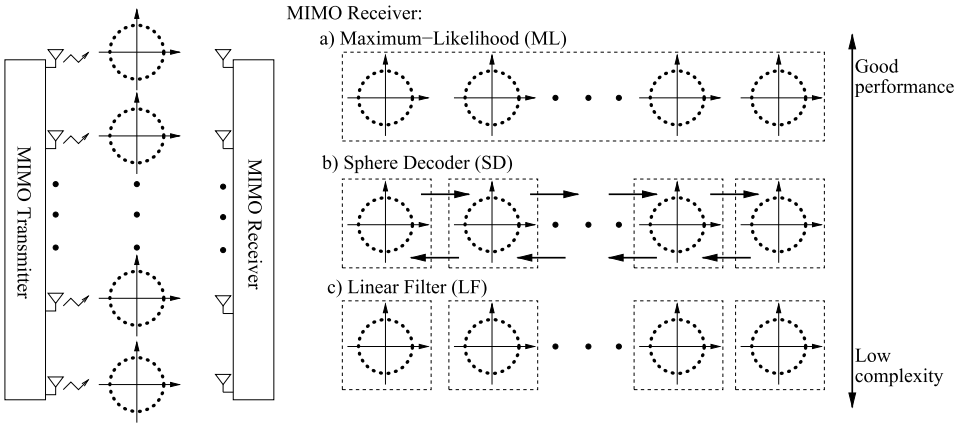


FIGURE 2. An example of striking a tradeoff between the performance attained and the complexity imposed by Bell Laboratories Layered Space-Time (BLAST) systems.

that may be increasing linearly with the number of antennas, but they are not designed for achieving a transmit diversity gain for combating the effects of fading. By contrast, the family of Space-Time Block Codes (STBCs) [33]–[35] offers a benefical transmit diversity gain, but the STBCs cannot achieve the full MIMO capacity. In order to circumvent this problem, the Linear Dispersion Code (LDC) concept [36]–[38] may be introduced to resolve this tradeoff, where a total number of N_Q modulated MPSK/QAM symbols are dispersed across both the N_T-element spatial domain and the N_P-element time domain of the transmission matrix. The LDC of [38] may attain both the full MIMO capacity and the full transmit diversity gain, provided that the parameters satisfy $N_Q \geq N_T N_P$. Nonetheless, since the STBC's orthogonality requirements are dropped by the LDC design, the LDC receiver has to employ the family of V-BLAST detectors in order to tackle the IAI. As a result, the performance versus complexity tradeoff illustrated by Fig. 2 surfaces again. Against this background, the SM scheme [26], [27] activates a single one out of N_T TAs in order to transmit a single modulated MPSK/QAM symbol, which results in a reduced transmitter hardware complexity, since only a single RF chain is employed. Moreover, the receiver's signal processing complexity may also be reduced, where the TA activation index and the modulated symbol index are detected separately. Moreover, the concept of Space-Time Shift Keying (STSK) [39] once again achieves a beneficial diversity gain, where a single one out of N_Q dispersion matrices is activated for dispersing a single modulated MPSK/QAM symbol. The STSK receiver may employ the low-complexity SM detectors in order to recover both the activated dispersion matrix index and the modulated symbol index.

Against this background, in this paper, we consider the pair of key MIMO design tradeoffs, which are portrayed by Fig. 3. The unified mathematical measures of capacity and error probability, which are used for quantifying the multiplexing and diversity tradeoff, are also invoked for characterizing the performance of SM and STSK.

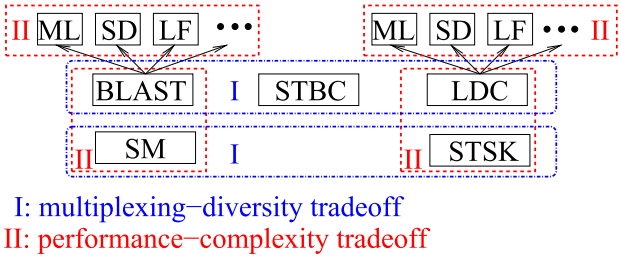


FIGURE 3. Key Multiple-Input Multiple Output (MIMO) schemes and the design tradeoffs that motivated their development.

A. A HISTORICAL PERSPECTIVE ON NEAR-CAPACITY COMMUNICATIONS SYSTEM DESIGN

The communications theoretic capacity limit was established by Shannon [40] in the late 1940s, which quantified a channel's capacity as the maximum mutual information between the input signal and the output signal. Shannon proposed in [40, Th. 11] that the channel capacity, which is the maximum data rate that can be transmitted over the channel at an infinitesimally low error rate, can be achieved with the aid of channel coding at the unconstrained cost of delay and complexity. In the 1950s, the single-error correcting Hamming code was proposed in [41], while the convolutional coding concept was proposed by Elias [42]. Following this, the multiple error correcting Bose-Chaudhuri-Hocquenghem (BCH) code was proposed in [43]–[45]. Furthermore, the Maximum-Likelihood Sequence Estimation (MLSE) of convolutional codes was proposed by Viterbi [46] in 1967. This classic Viterbi algorithm was further interpreted by Forney [47] in 1973, and it was also applied to block codes by Wolf [48] in 1978. As a major milestone, the optimum Log-Max A Posteriori (MAP) decoding algorithm was proposed by Bahl *et al.* [49] in 1974, which is often referred to as the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm. More explicitly, Viterbi's MLSE algorithm aims for maximizing the sequence estimation probability. By contrast, the BCJR Log-MAP aims for maximizing the probability for correctly

decoding each bit. The BCJR Log-MAP algorithm was shown to be capable of achieving a lower Bit Error Rate (BER) in [49] than the Viterbi algorithm [46]–[48]. However, owing to the fact that the BCJR Log-MAP algorithm imposed a substantially higher computational complexity than the Viterbi ML algorithm, it had not attracted much attention until the revolutionary development of near-capacity system design emerging in the 1990s. Most notably, the BCJR Log-MAP algorithm was simplified by the approximation of $\ln\left[\sum_{\forall i}\exp(d^i)\right] \approx \max_{\forall i} d^i$ by Koch and Baier [50] in 1990, which is often referred to as the Max-Lag-MAP algorithm, so that the computationally complex exponential operations may be avoided. Furthermore, Robertson *et al.* [51] proposed the near-optimum Approx-Log-MAP algorithm in 1995, which aimed for compensating the difference between the two terms of $\ln\left[\sum_{\forall i}\exp(d^i)\right]$ and $\max_{\forall i} d^i$ by invoking a lookup table.

On the voyage of persuing the near-capacity performance predicted by Shannon, the construction of powerful channel code became the greatest challenge. It was observed in [52] that the coding gain, which is the E_b/N_0-reduction provided by channel coding, grows linearly with the convolutional code's memory, but the associated decoding complexity grows exponentially. In order to mitigate this problem, the concept of concatenated codes [53] was introduced, where simple component codes were concatenated in order to construct a powerful channel code. The concatenated code concept was first proposed by Elias [54] in 1954, where an idealistic ''error-free'' performance predicted by Shannon's theory was shown to be possible. The concatenated code constituted by a convolutional code and a Reed-Solomon (RS) code stood out among the known candidates [53], [55], [56], which was capable of providing a performance that was only 2.0 $\sim$ 3.0 dB away from Shannon's capacity. In 1979, Battail *et al.* [57] proposed to place a interleaver between the component codes of a concatenated code, which was also referred to as a product code, so that the error bursts may be effectively interleaved. Battail *et al.* [57] also suggested in that the good performance of concatenated codes may be guaranteed if the component decoders can exchange their decisions. Inspired by the development of the Soft-Output Viterbi Algorithm (SOVA) and its application to concatenated codes developed by Hagenauer and Hoeher [58] in 1989, Lodge *et al.* [59] proposed in 1992 that the soft-decision iterative decoding conceived for concatenated block codes inched closer to Shannon's capacity. This scheme was further improved by Lodge *et al.* [60] in their ICC'93 paper, where the performance of half-rate channel coded BPSK transmitted over Additive White Gaussian Noise (AWGN) channels achieved an impressive closest ever 1.3 dB distance from Shannon capacity. It was also predicted by Lodge *et al.* [60] that the concatenated convolutional codes assisted by soft-decision iterative decoding may provide an even better performance. At the same ICC conference in 1993, the groundbreaking Turbo Coding (TC) technique was independently proposed by Berrou *et al.* [1], where a low BER of 10^{-5} was recorded at $E_b/N_0 = 0.7$ dB for half-rate channel coded BPSK transmitted over AWGN channel, which was achieved by the parallel concatenation of a pair of Recursive Convolutional Code (RSC) components exchanging their soft-bit information with the aid of iterative decoding, as previously predicted by Lodge *et al.* [59], [60].

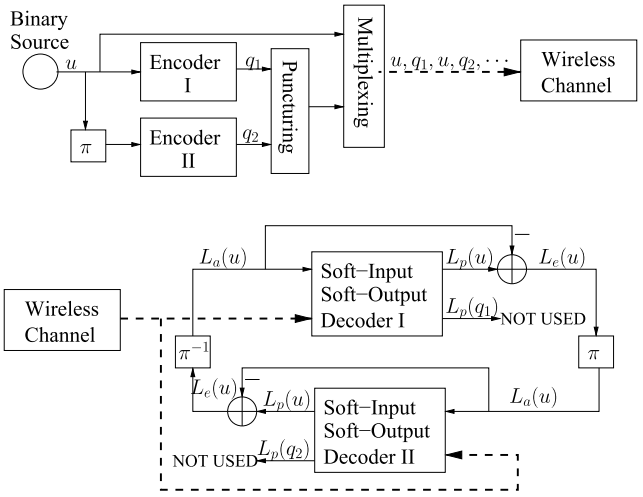


FIGURE 4. The schematic of a Parallel Concatenated Code (PCC) assisted by iterative decoding, which is adopted by Turbo Codes (TCs) [1], [2]. BPSK transmission over AWGN channels is assumed, unless otherwise stated.

Let us now elaborate a little further on TC and its revolutionary effect on channel coding science. The schematic of the Parallel Concatenated Code (PCC) adopted by TC [1], [2] is portrayed in Fig. 4. It can be seen in Fig. 4 that the information bits are encoded twice by a pair of component RSC encoders, where an interleaver is inserted between them in order to ensure that the bit-dependencies imposed by the two RSC codes are eliminated between them. At the receiver, the pair of component RSC decoders exchange their so-called extrinsic information[1] in order to achieve a near-capacity performance. The soft-bit processed by the soft-input soft-output decoders of Fig. 4 is in the form of Log Likelihood Ratio (LLR) [50], [58], where L_a, L_p and L_e represent the *a priori* LLR, *a posteriori* LLR and extrinsic LLR, respectively. BPSK transmission over an AWGN channel was assumed by the TC scheme of [1], [2]. However, it is straightforward to extend this scheme to more complex modulations, where an arbitrary modulator and a demodulator is placed before and after the wireless channel block of Fig. 4, respectively.

Following the groundbreaking invention of TC and considering that the block codes have relatively simple trellis structures [61], Pyndiah *et al.* [62] proposed to replace the convolutional codes of Fig. 4 by block codes, which also achieved a near-capacity performance [62], [63].

[1]The terminology of extrinsic information stems from the fact that as a benefit of the interleaver, they are capable of providing an independent 'extended' source of information for each bit.

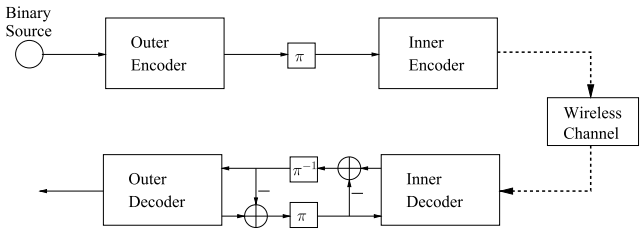


FIGURE 5. The schematic of generalized Serial Concatenated Code (SCC) assisted by iterative decoding.

Hagenauer *et al.* [64] generalized PCC, where any combination of block and convolutional codes was deemed to be possible. Owing to the fact that the TC component decoders in Fig. 4 only updated the LLRs for the information bits, but not for the parity bits, an error floor was experienced for a limited number of decoding iterations, Benedetto and Montorsi [3] and Benedetto *et al.* [4] proposed the concept of Serial Concatenated Code (SCC). The schematic diagram of a SCC is depicted in Fig. 5. Unlike for the PCC of Fig. 4, the SCC component decoders of Fig. 5 exchange their extrinsic information based on the exact same binary bits without any puncturing.

The Low-Density Parity Check (LDPC) coding concept that was originally proposed by Galager [65] in 1962 was popularized by MacKay and Neal [5] in 1996, where a near-capacity performance was achieved by constructing sparse random parity check matrices and by iteratively improving the decoding performance [5]–[7]. Hence the LDPC concept preceded TC by 31 years.

In order to optimize the communications schemes, the modulation scheme, which defines the format of signal transmission and determines the effective throughput should also be taken into account. During their infancy, channel coding and modulation were treated as separate entities [66], [67]. The first attempt of jointly designing channel coding and modulation is due to Mecklenburg *et al.* [68] in 1973, when the conventional Gray-labelling designed for modulation was revised in order to also impose bit-dependency on the channel coded source bits. As the benefit, the demodulator and the channel decoder act in liaison in order to jointly decide upon the modulated symbol. Inspired by this idea, Multi-Level Coding (MLC) was proposed by Imai and Hirakawa [69] in 1977, where the coded bits were mapped to the different - integrity protection - classes of multi-level modulus. The bits mapped to the lower-integrity modem sub-channels were protected by stronger channel codes, which were then detected first by the MLC scheme's multistage decoder followed by the other bits of the MLC scheme. In 1982, Ungerboeck [70] proposed the landmark concept of Trellis Coded Modulation (TCM), where the channel code's parity bits were accommodated by the modem by increasing the number of bits per symbol, because this required no bandwidth expansion for FEC. More explicitly, instead of using Gray-labelling for the modulated symbols, the TCM constellation diagram is divided into subsets by a techniqe referred to as set partitioning, where each bit determines a pair of subsets, and the Euclidean distance between the neighbouring constellation points within a subset is increased at every partitioning step. Similar to the MLC of [69], the TCM of [70] assigned stronger component channel codes associated with longer memories to protect the bits associated with lower Euclidean distances. However, instead of invoking a multistage decoder as the MLC scheme [69], the TCM decoder was originally designed for relying on a single trellis for jointly deciding on all the information bits.

Inspired by the invention of MLC and TCM, a lot of research efforts had been dedicated to developing multi-dimensional constellations for TCM [71]–[73] in the 1980s, where instead of set-partitioning the constellation diagram of a single symbol, a block of data were mapped to higher dimensional constellations, so that a beneficial coding gain was achieved by the joint channel coding and modulation design. However, as described in [74], the number of metrics to be calculated for the TCM decoder's trellis state transitions inevitably increases as the modulation-order increases. In order to mitigate the escalating complexity, the trellis construction of the TCM decoder was decomposed into lower-dimensional problems with the aid of multistage decoding [75]–[77] following the philosophy of the MLC receiver of [69].

A specific TCM scheme conceived for fading channels was conceived by Simon and Divsalar [78], [79] in 1988, which once again separated the channel code and modulation by placing a symbol-based interleaver between the two entities. Moreover, it was observed in [78] and [79] that the TCM scheme's maximized Euclidean distance became less important in fading channels than in case of AWGN channels [80]. Against this background, the classic Bit-Interleaved Coded Modulation (BICM) arrangement was proposed by Zehavi [81] in 1992, which was further developed by Caire *et al.* [82]. It was proven in [81] and [82] that the achievable time-diversity order of the BICM was determined by the minimum Hamming distance of the channel code. As a benefit of bit-based interleaving, every coded bits may be modulated to any modulation constellation point, and hence BICM is not designed for achieving the maximized free Euclidean distance of TCM. As a result, the TCM scheme still performs better than BICM in AWGN channels, but BICM outperforms TCM in fading channels, especially when the SNR is relatively high and hence the fading characteristics dominate the attainable performance. In order to further improve the performance of BICM, the landmark Bit-Interleaved Coded Modulation concept relying on Iterative Decoding (BICM-ID) was proposed by Li and Ritcey [8] in 1997. More explicitly, BICM-ID constitutes an instance of the generalized SCC portrayed in Fig. 5, where the channel code and the modulation scheme constitute the outer code and the inner code, respectively. The BICM-ID scheme was initially proposed for exchanging hard-decisions in [8] and [9] and then it was further developed for exchanging soft-bit decisions in [10] with the aid of a turbo receiver. It was

explicitly demonstrated in [11] that since the BICM-ID receiver's demodulator was capable of mapping any bit back to the constellation subset pairs with the aid of the *a priori* knowledge of all other bits, the free Euclidean distance was once again increased after the demodulator received feedback from the channel decoder, which assisted BICM-ID in outperforming TCM both in AWGN channels and in fading channels.

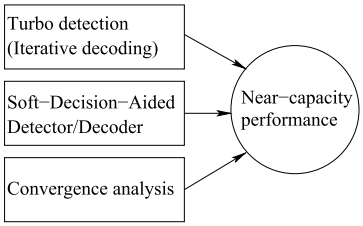


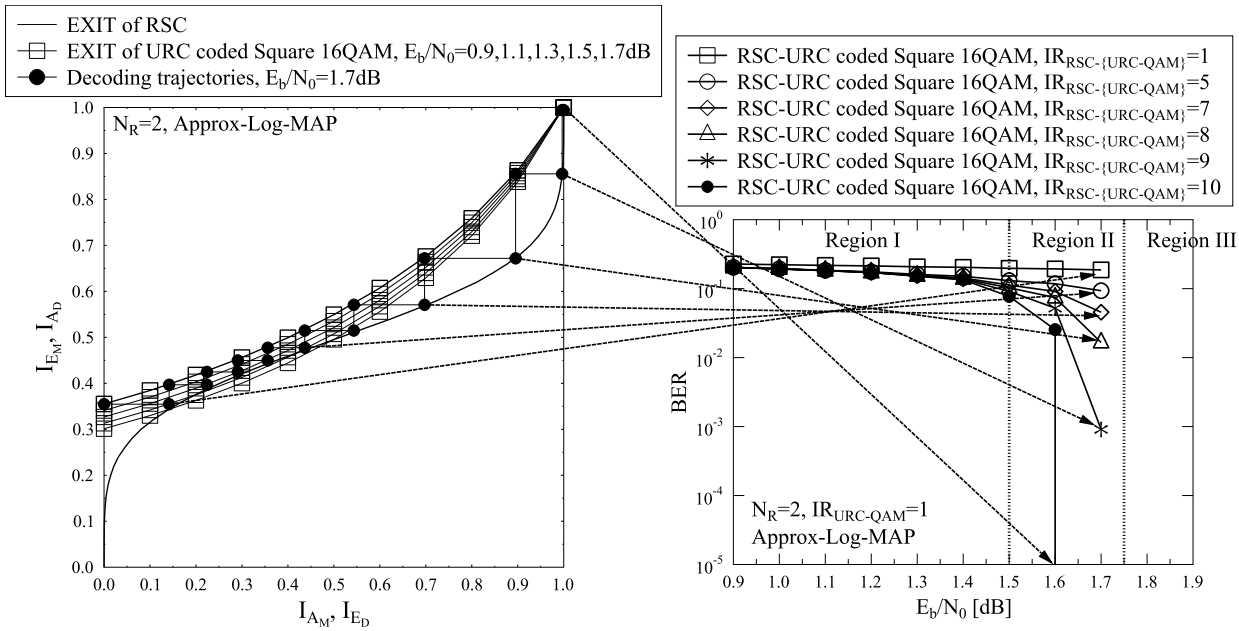
FIGURE 6. The key driving factors behind achieving a near-capacity performance.

It was gradually realized by the community that the ''turbo principle'' [83] may in fact be extended to a variety of areas in order to achieve the full potential of different communications systems. The revolutionary development of near-capacity system design has attracted substantial research interest from the late 1990s onwards, which covers the areas of channel coding [3], [4], source coding [84], equalization [85]–[87], multi-user detection [88]–[91], MIMO systems [22], [92], [93], etc. The three driving factors behind near-capacity system design are summarized in Fig. 6. Clearly, in order to perform iterative decoding/turbo detection, the constituent detectors/decoders have to be revised both to be able to accept and to produce soft-bit LLRs. In this treatise, the terminologies of iterative decoding and turbo detection are used interchangeably in order to address the involvement of potentially any detector/decoder in iterative decoding. The last key factor in Fig. 6 that has not received much attention is the convergence analysis.

The BER versus E_b/N_0 performance curve of a near-capacity system may be generally divided into three regions according to the noise level. In the low SNR region, the component channel codes are unable to correct large bursts of errors. At a specific SNR, which is not much higher than the capacity limit, a ''turbo cliff'' or a ''waterfall'' may be observed as the BER curve drops rapidly, which is the result of decoding convergence. When the SNR is increased beyond this specific region, the BER is expected to become infinitesimally low. An example of such BER performance curve is shown in Fig. 7b. Owing to the fact that the asymptotic union bounds derived based on the distance properties of channel codes are only tight at high SNRs [4], this tool becomes less useful for predicting the performance of turbo detected concatenated codes, which generally operate at a relatively low SNR that is close to the capacity limit. Recall that the error performance of coded modulation at a low SNR associated with a high noise level is more related to the modem's Euclidean distance than to the channel code's Hamming distance. As a result, the modulation scheme's capacity limit itself may be regarded as a loose performance prediction of the decoding convergence. In general, a communications system may be considered to be capable of ''near-capacity'' operation, when a turbo-like performance is achieved, which may be interpreted as attaining decoding convergence at an SNR that is within 1.0 dB distance from the capacity limit, provided that optimum or near-optimum decoding/detecting algorithms are employed.

Naturally, the prediction of the BER curve's ''turbo cliff'' SNR is important for near-capacity system design, but it is also important to optimize the number of iterations between the turbo detected component detectors/decoders so that no futile complexity wastage is imposed. In 1993, Moher [94] proposed to analyse the iterative convergence behavior with the aid of the cross-entropy metric, which was further developed to an iterative detection ''stopping criterion'' in [95]. The concept of cross-entropy allows us to keep track of the Probability Density Function (PDF) of the extrinsic LLRs produced by the component decoders, where decoding convergence is expected to occur, when the extrinsic LLR PDFs of the component decoders converge to the same decisions. Following this idea, Richardson and Urbanke [6] and Richardson *et al.* [7] proposed the density evolution concept for predicting the LDPC decoding convergence, where the belief propagation was also characterized by tracing the PDFs. Inspired by the development of density evolution, ten Brink [96] proposed the powerful tools of EXtrinsic Information Transfer (EXIT) charts in 1999, which visualized the convergence of turbo detection. More explicitly, the PDF of the extrinsic LLRs of a component decoder may be obtained by feeding Gaussian-distributed *a priori* LLRs [97], [98] to the decoder, so that the mutual information between the extrinsic LLRs and the source bits may be evaluated. As a benefit of iterative soft information exchanging between a pair of component decoders, the extrinsic information produced by a component decoder becomes the *a priori* information of another component decoder and vice versa. When the EXIT curves of two component decoders only intersect each other at the (1.0,1.0) point of the EXIT chart as seen in the example protrayed by Fig. 7a, decoding convergence is expected to occur. It can be seen in Fig. 7 that both the SNR and the number of iterations required for decoding convergence are accurately predicted by the EXIT charts. This technique was further extended for SCC in [99] and for PCC in [100] and [101]. Furthermore, it was proposed in [102]–[104] that the mutual information may be calculated without having access to the source bits. As a result, the EXIT charts may be constructed ''on-line'', because as soon as new extrinsic LLRs become available at the receiver, they can be used for updating the current estimate of the mutual information [105].

In summary, the major contributions on near-capacity system design are summarized in Table 2, while Fig. 8 offers a further historic perspective. It is interesting to see in Fig. 8 that the complexity reduction of channel decoding has



a) EXIT charts and decoding trajectories

b) BER performance

FIGURE 7. An example of EXIT charts analysis and BER performance of the RSC and URC coded Square 16QAM scheme. The Discrete-input Continuous-output Memoryless Channel (DCMC) capacity limit of this scheme is given by $E_b/N_0 = 0.1$ dB. a) EXIT charts and decoding trajectories. b) BER performance.

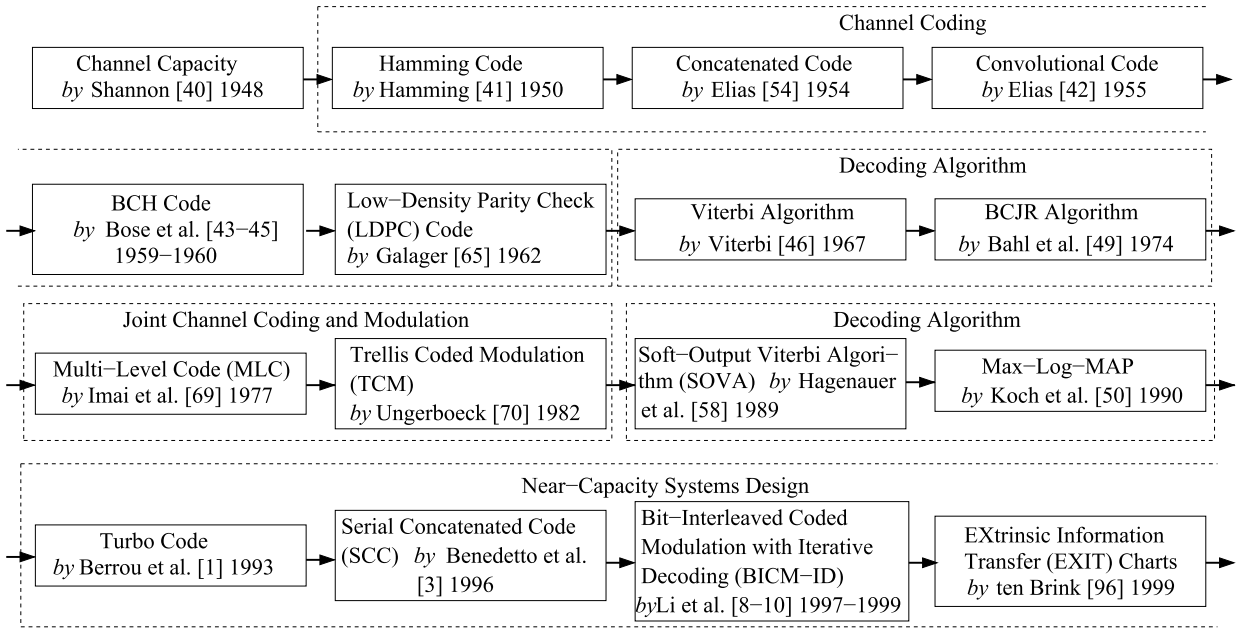


FIGURE 8. Historical chart for major milestones of near-capacity systems design.

motivated major breakthroughs for the entire suite of wireless communication systems twice in history. For the first time, when both the Viterbi and the BCJR algorithms have facilitated joint channel coding and modulation design in the context of MLC and TCM during the era spanning from the late 1970s to the 1980s. For the second time in history, the developments of SOVA and Max-Log-MAP have further inspired near-capacity system design since 1990s. In fact, at the time of writing, soft-decision modulated signal detection typically contributes a substantial fraction of the total complexity, especially when powerful MIMO schemes are employed. Therefore, the reduced-complexity detection algorithms introduced in this treatise may become more beneficial, especially when the soft-decision MIMO signal

TABLE 2. Summary of major contributions on near-capacity system design.

Year	Author(s)	Topic	Contribution
1948	Shannon [40]	Capacity Theorem	Proposed that the channel capacity, which is the maximum data rate that can be transmitted over the channel at an infinitesimally low error rate, can be achieved with the aid of channel coding at the unconstrained cost of delay and complexity.
1950	Hamming [41]	Channel Code	Proposed the single-error correcting Hamming code.
1954	Elias [54]	Concatenated Code	Proposed the concatenated code concept, where an idealistic "error-free" performance predicted by Shannon's theory was shown to be possible.
1955	Elias [42]	Channel Code	Proposed the classic convolutional coding concept.
1959 $\sim$ 1960	Bose *et al.* [43]–[45]	Channel Code	Proposed the classic multiple-error correcting BCH code, which was named after the authors.
1967	Viterbi [46]	Decoding Algorithm	Proposed the Maximum Likelihood Sequence Estimation (MLSE) decoding algorithm of convolutional code, which was later termed as Viterbi algorithm [47] and was applied to block codes in [48].
1973	Mecklenburg *et al.* [68]	Coded Modulation	Proposed to jointly design channel coding and modulation, where the demodulator and the channel decoder act in liaison in order to jointly decide upon the modulated symbol.
1974	Bahl *et al.* [49]	Decoding Algorithm	Proposed the major milestone of the optimum Log-Max A Posteriori (MAP) decoding algorithm, which is also known as the BCJR algorithm named after the authors.
1977	Imai and Hirakawa [69]	Coded Modulation	Proposed Multi-Level Code (MLC), where the bits mapped to the lower-integrity modem sub-channels were protected by stronger channel codes, which were then detected first by the MLC scheme's multistage decoder followed by the other bits of the MLC scheme.
1979	Battail *et al.* [57]	Concatenated Code	Proposed to place a interleaver between the component codes of a concatenated code and proposed to exchange decisions between the component decoders.
1982	Ungerboeck [70]	Coded Modulation	Proposed the concept of Trellis Coded Modulation (TCM), which increased the constellation Euclidean distance by set-partitioning, while modulation and channel code were jointly designed by a single trellis.
1988	Simon and Divsalar [78], [79]	Coded Modulation	Proposed to place a symbol-based interleaver between the channel code and the modulation for the TCM scheme conceived for fading channels.
1989	Hagenauer *et al.* [58]	Decoding Algorithm	Proposed to modify the Viterbi algorithm to be able to process soft-bit decisions, which is also known as the Soft-Output Viterbi Algorithm (SOVA) algorithm.
1990	Koch and Baier [50]	Decoding Algorithm	Proposed to simplify the BCJR Log-MAP algorithm by the approximation of $\ln\left[\sum_{\forall i}\exp(d^i)\right] \approx \max_{\forall i} d^i$ in order to avoid the computationally complex exponential operations, which is often referred to as the Max-Lag-MAP algorithm.
1992	Zehavi [81]	Coded Modulation	Proposed the classic Bit-Interleaved Coded Modulation (BICM), which replaced the TCM's symbol-based interleaver [78], [79] by a bit-based interleaver in order to improve the achievable time-diversity order of the BICM in fading channels.
1992	Lodge *et al.* [59]	Concatenated Code	Proposed the soft-decision iterative decoding conceived for concatenated block codes that inched closer to Shannon's capacity, which was further improved by the authors in [60].
1993	Berrou *et al.* [1]	Concatenated Code	Proposed the groundbreaking Turbo Code (TC), which achieved a near-capacity performance by the parallel concatenation of a pair of RSCs exchanging their soft-bit decisions with the aid of iterative decoding. It was later summarized in detail by the authors in [2].
1995	Robertson *et al.* [51]	Decoding Algorithm	Proposed the near-optimum Approx-Log-MAP which compensated the difference between the BCJR Log-MAP [49] and the Max-Log-MAP [50] by invoking a lookup table.
1996	Hagenauer *et al.* [64]	Concatenated Code	Proposed to generalize the Parallel Concatenated Code (PCC), which included TC as an special case.
1996	Benedetto *et al.* [3]	Concatenated Code	Proposed to generalize the Searial Concatenated Code (SCC), which was later summarized in detail by the authors [4].
1997 $\sim$ 1999	Li and Ritcey [8]–[10]	Coded Modulation	Proposed the Bit-Interleaved Coded Modulation concept relying on Iterative Decoding (BICM-ID), which improved BICM [81], [82] by introducing iterative decoding between the demodulator and the channel decoder.
1999	ten Brink [96]	Convergence Analysis	Proposed the powerful tools of EXtrinsic Information Transfer (EXIT) charts, which visualized the extrinsic information exchanged in iterative decoding and accurately predicted both the SNR and the number of iterations required for decoding convergence.
2000	Divsalar *et al.* [106]	Concatenated code	Proposed to further place an Unity Rate Code (URC) as an intermediate component in the SCC, so that the error floor of the two-stage turbo detector may be eliminated by the resultant three-stage turbo detector.
2001	ten Brink [101]	Convergence Analysis	Extended EXIT charts to the PCC system design.
2004	Tuchler [99]	Convergence Analysis	Extended EXIT charts to the SCC system design, and proposed the IRregular Convolutional Code (IRCC) concept in order to inch closer to the capacity limit.
2009	Hanzo *et al.* [12]	Coded Modulation	Summarized guidelines for general near-capacity system design and offered design examples for a wide range of communications systems.

detectors are invoked several times in order to approach the performance potential promised by the capacity theorem.

Moreover, it is also worth noting that in line with the reduced-complexity design philosophy, Fig. 8 also shows two examples of major breakthroughs being made by decomposing a very high-complexity detector into lower-complexity parts while taking account the interaction between the constituent parts. The first example is that a high-complexity Convolutional Code (CC) was decomposed into a pair of low-complexity CCs and an interleaver, yielding a concatenated code, which led to the success of TC. The second example is that channel coding and modulation were jointly designed in MLC and TCM in order to achieve a better overall system performance. The BICM-ID scheme once again separated these two entities, where turbo detection exchanging extrinsic information between the channel decoder and signal demodulator was invoked in order to attain the best possible performance.

B. A HISTORICAL PERSPECTIVE ON MULTIPLE-INPUT MULTIPLE OUTPUT SCHEMES

Multiple-Input Multiple Output (MIMO) techniques have been one of the most vibrant areas in communications, where exciting progress has been made over the past two

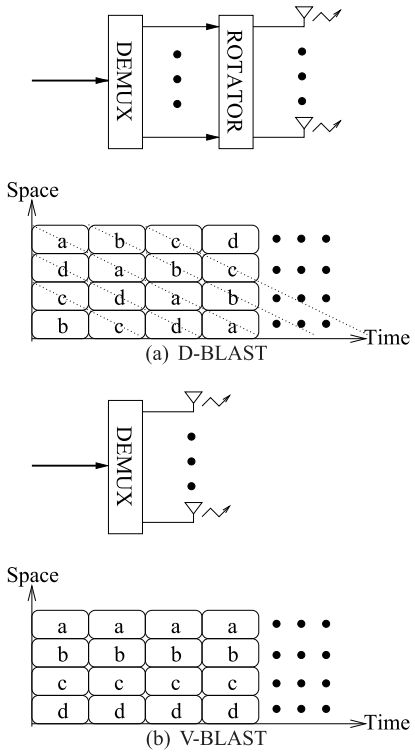


FIGURE 9. Schematics of D-BLAST and V-BLAST. (a) D-BLAST. (b) V-BLAST.

decades. The proposal of employing multiple antennas for a single user was motivated by its substantial capacity gain. In more details, the multiplexing-oriented MIMO concept was proposed by Paulraj and Kailath [107] in 1994, where a high data-rate transmission was carried out by splitting it into low data-rate signals transmitted by spatially separated Space-Division Multiple Access (SDMA) users. In order to pursue the multiplexing gain using co-located antennas, Foschini [30] proposed the ground-breaking layered space-time architecture in 1996, which was later termed as the BLAST. In particular, the original encoding method proposed by Foschini [30] was diagonal-encoding, which may be termed as D-BLAST. As portrayed in Fig. 9(a), the D-BLAST transmitter de-multiplexes a single data stream to N_T separate data streams, where channel coding and modulation may be performed either before or after the de-multiplexing, and then the N_T data streams of the N_T TAs are rotated in a round robin fashion, so that the code words are transmitted in diagonal layers. El-Gamal and Hammons [31] further extended this D-BLAST structure in 2001, where each layer constitutes more than one consecutive diagonal lines. The benefit of D-BLAST's diagonal-encoding is that the signal components of a diagonal layer experience independent fading, which may lead to a potential temporal diversity gain.

In order to simplify the real-time implementation, in 1998, Wolniansky *et al.* [32] proposed V-BLAST that invokes vertical-encoding. As portrayed by Fig. 9(b), the rotator of the D-BLAST was avoided by the V-BLAST transmitter. Owing to the fact that all the signals transmitted from N_T TAs are simultaneously received by N_R RAs, the same detection methods are shared by both D-BLAST and V-BLAST, which was exemplied in Fig. 2. It was demonstrated in [30] and [32] that both D-BLAST and V-BLAST may achieve an improved spectral efficiency that increases linearly with the number of antennas at realistic SNRs and error rates. It was further confirmed by Foschini and Gans [108] in 1998 and then by Telatar [109] in 1999 that compared to the family of Single-Input Multiple-Output (SIMO) systems where multiple antennas may only be used at the receiver, the BLAST MIMO systems have an ergodic capacity that may grow linearly, rather than logarithmically, with the number of antennas, provided that the BLAST MIMO system employs a large number of antennas and that both the input signals and the output signals are independent and identically Gaussian-distributed.

In order to exploit the full potential of BLAST MIMO systems and to approach the impressive capacity results, the BLAST receivers have to employ ML detection in uncoded systems, or the MAP detection in coded systems, which have to evaluate all M^{N_T} combinations of a total of N_T transmitted MPSK/QAM symbols [12]. This implies that the BLAST detection complexity increases exponentially with the number N_T of TAs, which may be particularly unaffordable, when the BLAST detector is invoked several times in turbo coded systems. In order to mitigate this problem, the BLAST schemes [30], [32] were originally proposed to employ the Multi-User Detector (MUD) of the classic Code Division Multiple Access (CDMA) systems [110], [111]. More explicitly, in order to separate the N_T data streams impinging at the BLAST receiver, Linear Filter (LF) based receivers, such as Zero Forcing (ZF) and MMSE receivers may be invoked, where all the other data streams, i.e. the interferers, may be nulled when detecting a particular data stream. However, the LFs suffer from inevitable performance limitations, since ZF enhances the noise, while the MMSE receiver only minimizes, rather than eliminates, the interferers. In order to further improve the attainable performance, the decision-feedback techniques of [112]–[114], which have been widely used for equalization may be employed for cancelling an interfer from the BLAST scheme's received signal immediately after a data stream has been detected, so that the ensuing detection stages suffer less from the interference problem. Nonetheless, the LFs aided BLAST receivers generally suffer from a performance penalty compared to the optimum nonlinear BLAST detection, but the LF aided BLAST detection complexity becomes comparable to that of Single-Input Single-Output (SISO) or SIMO systems, because the constellation diagrams of the N_T data streams are visited completely separately.

In order to achieve a further improved performance in coded systems, the LFs may be revised to be able to both accept and produce soft-bit decisions. The first soft-decision MMSE filter was proposed by Douillard *et al.* [85] for turbo equalization in 1995. However, in the presence of soft-bits, the *a priori* probabilities are no longer equal for all

constellation points, which poses a major design challenge for the MMSE solution of coded systems. In order to solve this problem, the exact MMSE solution incorporating the non-constant *a priori* probabilities was derived for CDMA MUD by Wang and Poor [90] in 1999, and then this solution was invoked for turbo equalization by Tuchler *et al.* [87] in 2002 and finally for turbo BLAST by Sellathurai and Haykin [22] also in 2002.

In order to strike a performance-complexity tradeoff between the BLAST scheme's optimum detector and the LF-aided detectors, Damen *et al.* [115] proposed to apply sphere decoding for BLAST detection in 2000, where the ML performance may be retained at a substantially reduced complexity. As illustrated by Fig. 2, the SD visits the constellation diagrams one-by-one in order to find the best candidates that lie within the decoding radius, and then these constellation diagrams may be visited again by the SD in order to check for other possible candidates. The termination of SD is determined by the SNR-dependent sphere radius. The SD algorithms designed for BLAST detection were extensively documented by Damen *et al.* [19] in 2003. Inspired by the turbo codes, the first soft-decision SD aided BLAST was proposed by Hochwald and ten Brink [116], where a list of BLAST signal candidates was established by the hard-decision SD and then the candidates in this list were processed by the MAP decoding algorithm. However, the *a priori* information gleaned from the channel decoder was not utilized for establishing the candidate-list in [116], which prevented it from achieving BLAST's full potential. In order to mitigate this problem, in 2004, Vikalo *et al.* [117] proposed the soft-decision SD for BLAST, which incorporated the *a priori* information in sphere decoding. Furthermore, in 2008, Studer *et al.* [118] proposed the soft-output SD's Very-Large-Scale Integration (VLSI) implementation, where a single SD tree search was invoked just once for all the soft-bit decisions output for a BLAST detection block. Studer and Bolcskei further developed their work of [118] in [119] in 2010, where the *a priori* LLRs were once again incorporated into the SD's VLSI implementation.

The BLAST systems enjoy a beneficial multiplexing gain, where the system throughput may be N_T times higher than that of their SISO/SIMO counterparts using the same MPSK/QAM constellation. Alternatively, the mutliple TAs may be exploited for achieving a diversity gain, where multiple replicas of the modulated symbols may be transmitted by multiple TAs over multiple symbol periods, so that the receiver becomes capable of recovering the data-carrying symbols from several independently faded observations. This revolutionary invention was originally proposed by Alamouti [34] for the case of using $N_T = 2$ TAs in 1998, where the full transmit diversity was achieved by a SISO receiver at a low detection complexity. More explicitly, the transceiver of Alamouti's transmit diversity technique is portrayed in Fig. 10, where the space-time mapper forms a two-by-two unitary matrix from the $N_Q = 2$ independently modulated MPSK/QAM symbols, which are transmitted by

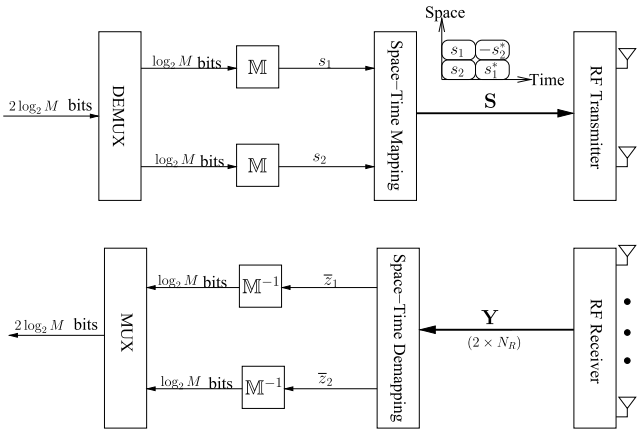


FIGURE 10. Schematic of Alamouti's G2 STBC transceiver.

$N_T = 2$ TAs over $N_P = 2$ symbol periods. Owing to the orthogonality provided by the unitary matrix design, the receiver of Fig. 10 is capable of decoupling the $N_Q = 2$ data streams without encountering BLAST's IAI problem. The class of transmit diversity techniques generated from orthogonal design has been termed as the set of Space-Time Block Code (STBC) arrangements. In particular, as the first member in the STBC family, Alamouti's scheme is often referred to as G2 STBC.

The gravest challenge of STBC design is to construct the unitary matrix from orthogonal design for any arbitrary number of TAs. Alamouti's G2 STBC has a unity normalized throughput of $\overline{R} = \frac{N_Q}{N_P} = 1$, which implies that its throughput is the same as that of its SISO/SIMO counterpart, when using the same MPSK/QAM constellation. Owing to its transmit diversity gain, Alamouti's G2 STBC has a better BER performance than its BLAST MIMO and SIMO counterparts associated with the same system throughput. However, it was proven by Tarokh *et al.* [35] in 1999 that Alamouti's G2 STBC is the only full unity-rate code in the family of STBCs. Nonetheless, Tarokh *et al.* [35] discovered that full unity-rate real-valued STBCs do exist for $N_T = 2$, 4 or 8, which may be generated by the Hurwitz-Radon theory [120], [121]. As a result, the class of Half-Rate (HR) STBCs may be obtained by vertically concatenating the real-valued STBC codeword and its conjugates, which forms the family of HR STBCs that are represented by the terminology of HR-GN_T-STBC for using N_T TAs. For the case of N_T not being a power of 2, the HR-GN_T-STBC transmission matrix may be obtained by taking the first N_T columns of the HR-G$2^{\lceil \log_2 N_T \rceil}$-STBC's codeword. Although the HR-GN_T-STBCs created for $N_T > 8$ were not explicitly constructed, Tarokh *et al.* [35] proved that such a design may impose a substantial transmission delay, which increases exponentially with N_T. For example, we have [35] $N_P = 16 \times 16^{(N_T/8-1)}$ for $N_T > 8$ and being a power of 2.

In order to improve the throughput of STBCs with $N_T > 2$, Ganesan and Stoica [122]–[124] invented the

Amicable Orthogonal (AO) STBCs in 2001 according to the theory of amicable orthogonal design [120]. An AO STBC scheme having N_T TAs may be represented by the terminology of AO-GN_T-STBC. For the case of N_T being a power of 2 as $N_T = 2^\iota$, where ι denotes a positive integer, the AO-G2^ι-STBC schemes have a reduced delay of $N_P = N_T$, and they also have $N_Q = \iota + 1$ transmitted symbols. More explicitly, the AO-G2^ι-STBC's transmission matrix is constructed based both on the lower-level AO-G$2^{\iota-1}$-STBC's transmission matrix having ι symbols as well as on a an extra the $(\iota + 1)$-th modulated symbol. Hence, the construction of AO-G2^ι-STBCs may commence from $\iota = 1$, where the AO-STBC associated with $\iota = 1$ corresponds to Alamouti's G2 STBC. As a result, rate-3/4 STBCs associated with a reduced delay of $N_P = 4$ may be constructed for the AO-STBCs having $N_T = 3$ or $N_T = 4$, while half-rate STBCs associated with a reduced delay of $N_P = 8$ may be constructed for the AO-STBCs having $5 \leq N_T \leq 8$. However, owing to the fact that the AO-STBC's number of transmitted symbols N_Q only increases logarithmically with the number of TAs N_T as $N_Q = \lceil \log_2 N_T \rceil + 1$, the attainable throughput of AO-STBC is expected to be lower than the half-rate of $\overline{R} = \frac{1}{2}$ for $N_T > 8$.

Against this background, it has emerged that there is a tradeoff between the attainable multiplexing and diversity gain in MIMO system design. The development of STBCs was motivated by their improved BER performance, especially in the high SNR region, which is the benefit of their diversity gain. However, it was recognized by Sandhu and Paulraj [125] in 2000 that STBCs cannot achieve the full MIMO capacity except for a special case, which is Alamouti's G2-STBC system associated with a single RA, i.e. with $N_R = 1$. On the other hand, the BLAST systems have the full MIMO capacity, but they are not designed for achieving a transmit diversity gain for combating the effects of fading. This classic MIMO design tradeoff was quantified by Zheng and Tse [29] in 2003, where the relationship between the diversity gain d and the multiplexing gain r is given by $d = (N_T - r)(N_R - r)$, which portrays the diversity and multiplexing gains as rivals in MIMO system design.

If the STBC throughput is to be improved, the first step is to relax the orthogonality requirement. In the light of this principle, the concept of Quasi-Orthogonal (QO) STBC design was proposed by Jafarkhani [126] in 2001, where the QO STBC's transmission matrix is formed by subgroups of orthogonal STBCs. For the QO STBCs, the signals are orthogonal to each other within the subgroups, but they are not orthogonal to the signals from the other subgroups. As a result, the IAI problem resurfaces in the QO STBC design, and hence the signals that cannot be decoupled have to be jointly detected. It was suggested by Papadias and Foschini [127] in 2003 that linear MIMO receivers such as the MMSE detector or the ZF detector may be invoked for QO-STBC systems. However, this may not be an ideal solution, because the sub-optimal linear MIMO receivers may erode the performance advantage of the QO-STBC's diversity gain.

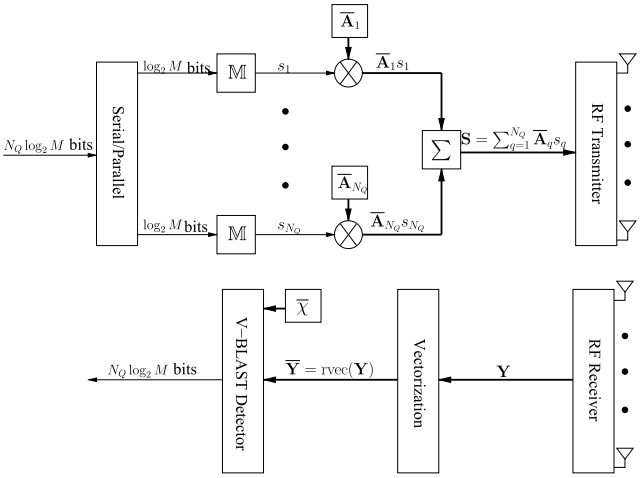


FIGURE 11. Schematic of the capacity-achieving LDC transceiver of [38].

In 2002, Hassibi and Hochwald [36] proposed the new class of Linear Dispersion Code (LDC), which completely droped the STBC's orthogonality requirements in order to further improve the STBC capacity while retaining the full transmit diversity gain. In more details, the LDC's transmission matrix may be represented by $\mathbf{S} = \sum_{q=1}^{N_Q} \left[\overline{\mathbf{A}}_q \Re(s_q) + j\overline{\mathbf{B}}_q \Im(s_q)\right]$, where the real and imaginary parts of a total number of N_Q modulated MPSK/QAM symbols $\{s_q\}_{q=1}^{N_Q}$ are dispersed into both spatial and temporal dimensions by the dispersion matrices $\{\overline{\mathbf{A}}_q\}_{q=1}^{N_Q}$ and $\{\overline{\mathbf{B}}_q\}_{q=1}^{N_Q}$. The dispersion matrices are obtained from random search, where the capacity is maximized while the error probability is aimed to be minimized. Although the LDCs proposed by Hassibi and Hochwald [36] effectively improve the attainable STBC capacity, and the LDC may even outperform the STBC in certain scenarios, the full MIMO capacity still cannot be achieved by the LDC design of [36]. In order to further improve the LDC design, Heath and Paulraj [38] proposed in 2002 that jointly dispersing the real and imaginary parts of the N_Q modulated MPSK/QAM symbols $\{s_q\}_{q=1}^{N_Q}$ may allow the LDC to achieve the full MIMO capacity, which results in a simplified form of the transmission matrix given by $\mathbf{S} = \sum_{q=1}^{N_Q} \left[\overline{\mathbf{A}}_q s_q\right]$. For the sake of clarity, the original LDC design proposed by Hassibi and Hochwald [36] is referred to as the capacity-improving LDC in this treatise, while the further optimized LDC design conceived by Heath and Paulraj [38] is termed as the capacity-achieving LDC, whose transceiver is portrayed in Fig. 11. The vectorization process seen in Fig. 11 may transform the LDC's received signal to a form that is equivalent to the received signal of a V-BLAST system equipped with N_Q TAs and $N_R N_P$ RAs, so that the classic V-BLAST detectors may be invoked for LDC detection. Owing to the fact that the dispersion matrices $\{\overline{\mathbf{A}}_q\}_{q=1}^{N_Q}$ are populated with random elements, they can be designed under

the constraint of having a transmission delay of $N_P = N_T$, which is a more relaxed condition compared to the delay of STBCs [34], [35], [122]–[124]. Furthermore, it was demonstrated by Heath and Paulraj [38] that satisfying the condition of $N_Q \geq N_T N_P$ is required for the LDC to achieve the full MIMO capacity, which implies that the LDC throughput is flexibly adjusted and it may even be higher than that of its BLAST counterpart using the same MPSK/QAM constellations. Upon finding the MIMO matrix capable of achieving the full MIMO capacity, the random search for the capacity-achieving LDC of [38] may aim for minimizing the error probability. It was demonstrated by Heath and Paulraj [38] that powerful LDCs exist that are also capable of outperforming their STBC counterparts. The error probability of LDCs was further improved in [128]–[132], which also tackle the problem of having a diminishing distance between legitimate codewords, when aiming for the high-throughput LDC codeword generation. In general, the random search carried out for populating LDC matrix according to the original guidelines of [38] is capable of producing powerful LDCs that achieve both a full multiplexing gain and a full transmit diversity gain.

The development of LDC successfully resolves the diversity versus multiplexing tradeoff, where both full MIMO capacity and full diversity gain may be attained following the optimized codeword construction guidelines of [38], provided that the parameters satisfy $N_Q \geq N_T N_P$. However, the LDC design becomes a retrograde step for the tradeoff between the performance attained and the complexity imposed. As the STBC's orthogonality requirement is abandoned, the LDC receivers have to invoke conventional V-BLAST detectors in order to deal with the IAI problem. As discussed before, the performance versus complexity tradeoff has an even more significant impact on the family of coded systems. More explicitly, on one hand, optimal MAP aided MIMO receivers exhibit a potentially excessive detection complexity, which may become especially unaffordable when the MIMO detector is invoked several times in the context of turbo detection. On the other hand, suboptimal non-MAP receivers are at risk of producing over-confident output LLRs that deviate from the true probabilities, which cannot be readily corrected by the channel decoder.

Against this background, a newly-developed MIMO technique referred to as Spatial Modulation (SM) was proposed by Song *et al.* [26] in 2004, which is a modulated extension of a scheme proposed in 2001 by Chau and Yu [133]. Then SM was analysed by Mesleh *et al.* [27] in 2008. The SM transmitter is portrayed in Fig. 12, where $\log_2 M$ bits are assigned to modulate a single MPSK/QAM symbol by the MPSK/QAM modulator, while $\log_2 N_T$ bits are assigned to activate a single one out of N_T TA by the TA index activation encoder in order to transmit the single modulated MPSK/QAM symbol. It can be seen in Fig. 12 that only a single RF-chain associated with a TA is activated at a time, which effectively reduce the MIMO's transmission complexity. Moreover, one of the most important motivations behind the SM design is the hope that the TA activation index and the classic

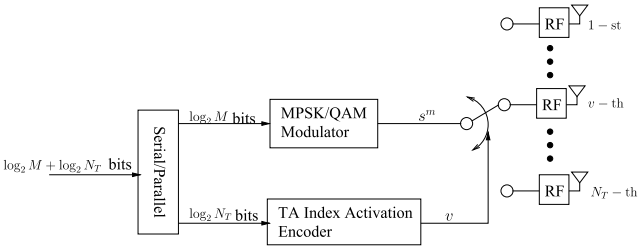


FIGURE 12. Schematic of the SM transmitter.

modulated symbol index may be separately detected, so that the optimal ML MIMO detection performance may be achieved for SM at a substantially reduced complexity. Therefore, Mesleh *et al.* [27] proposed a Maximum Ratio Combining (MRC) based SM detector, which firstly "decouples" the received signal to N_T matched filter output elements. Following this, the TA activation index may be detected by comparing the absolute values of the matched filter output elements, and then the classic MPSK/QAM demodulator may be invoked for demodulating the specific matched filter output element according to the detected TA activation index. As a result, the SM detector does not have to jointly detecting the N_T TA index candidates and the M modulated symbol candidates by evaluating a total of $N_T M$ combinations of SM signals. Instead, the N_T TA index candidates and the M modulated symbol candidates are evaluated separately, which reduces the SM detection complexity order from $O(N_T M)$ to $O(N_T + M)$. However, it was demonstrated by Jeganathan *et al.* [134] in 2008 that completely independently detecting the two indices results in an error floor, unless the fading channels are known and compensated at the transmitter by a precoder. This is because the erroneous TA activation index detection may mislead the MPSK/QAM demodulator into detecting the wrong symbol. As a remedy, Jeganathan *et al.* [134] streamlined the ML MIMO detector's calculations for SM, which takes advantage of the fact that the SM transmit vector contains $(N_T - 1)$ zero elements and a single non-zero element. As a benefit, the computational complexity imposed may be reduced by this simplification, but the detection complexity order remains $O(N_T M)$, where the N_T TA index candidates and the M modulated symbol candidates are still jointly evaluated. As a remedy, Space-Shift Keying (SSK) was proposed by Jeganathan *et al.* [135] and [136] in 2008, where simply the TA activation index conveys the source information. However, the SSK schemes inevitably suffers from a capacity loss compared to the SM schemes.

Inspired by this open problem, SM detector design has been developed in two major directions in the open literature. The first option is to develop the optimal SM ML detectors [137]–[140] endeavouring to reduce the complexity order of the simplified SM detector of [134] without imposing any performance loss. The second approach elaborated on in [141]–[147] aims for improving the attainable performance of the sub-optimal MRC based SM detector of [27], but

achieving the optimal ML SM performance is not guaranteed. In more details, for the optimum ML SM detection, in 2008, Yang and Jiao [137] proposed to invoke classic MPSK/QAM demodulators for all matched filter output elements first, and then the TA activation index detection was performed with the aid of the demodulated MPSK/QAM symbols. This method was also considered by Rajashekar *et al.* [140] in 2014, which was termed as the hard-limiter-based SM detector. Owing to the fact that in the absence of *a priori* information gleaned from a channel decoder, the hard-decision MPSK/QAM demodulators may directly map the matched filter's output signal to the nearest MPSK/QAM constellation point. As a result, the hard-limiter-based SM detection has a low detection complexity order of $O(2N_T)$, which does not increase with the number of modulation levels M. However, this method cannot be directly applied to the soft-decision SM detectors in coded systems, because the channel decoder is unaware of which constellation diagram is employed. As a result, the soft-decision SM detectors have to evaluate and compare all the TA index and classic modulated symbol index combinations, when both the *a priori* information gleaned from the channel decoder and the matched filter output are taken into account, which increases the detection complexity order back to $O(N_T \times M)$.

In order to mitigate this problem, in 2013 Xu *et al.* [139] proposed a SM detector, which aims for reducing the SM detection search scope while maintaining the optimum detection capability. In more detail, by exploring the symmetry provided by the Gray-labelled MPSK/QAM constellation diagrams, the normalized matched filter output elements may be first partially demodulated, so that the correlation between the TA index and the classic modulated symbol index may be taken into account, when the TA index is detected. Following this, only a single MPSK/QAM demodulation action has to be carried out according to the already detected TA activation index. Based on these processing steps, this may be referred to as the reduced-scope SM detector. This method was then also applied to the soft-decision SM detector of [139], which exploited the symmetry of the Gray-labelled constellation diagrams to perform the above mentioned reduced-scope MPSK/QAM demodulation. As a result, the reduced-scope SM detector [139] may achieve a substantial complexity reduction compared to the simplified SM detector of [134] without imposing any performance loss.

Considerable research efforts have also been dedicated to the family of sub-optimal low-complexity SM detectors in recent years. It was discovered and demonstrated by Guo *et al.* [141] in 2010 and by Naidoo *et al.* [142] in 2011 that the error performance of the TA activation index detection of the MRC based SM detector of [27] may be improved by normalizing the matched filter output signals by the fading norm, which leads to the concept of normalized-MRC-based SM detection. The so-called signal-vector-based SM detector proposed by Wang *et al.* [143] in 2012 operates based on the fact that the Square MQAM symbol does not change the direction of the received signal vector, which hence attains the same performance results as the normalized-MRC-based SM detectors. Furthermore, in order to avoid the situation of missing the optimum TA index candidate, Guo [141], Naidoo *et al.* [142], and Zheng [144] proposed to allow the TA activation index detector to produce a list of candidates, and then the MPSK/QAM demodulator may be invoked for all the TA indices in this list. This method may be termed as the list-normalized-MRC-based SM detector. Moreover, Sugiura *et al.* [145] conceived a unity-constellation-power-based SM detector in 2011, where a reduced number of non-negative constellation points associated with a unity constellation power are taken into account for the sake of achieving a more reliable TA index estimation. In 2012, Yang *et al.* [146] further improved the performance of the unity-constellation-power-based SM detector by invoking a list of TA indices as used in [141], [142], and [144], which may be termed as the list-unity-constellation-power-based SM detector. The decision metrics used by the unity-constellation-power-based SM detector were further improved by Tang *et al.* [147] in 2013, which is termed as the distance-ordered-based SM detector. It is also worth mentioning that a sphere decoder was invoked for single-stream SM by Younis *et al.* [148]–[150], which exhibits a reduced complexity compared to the sphere decoder invoked by V-BLAST.

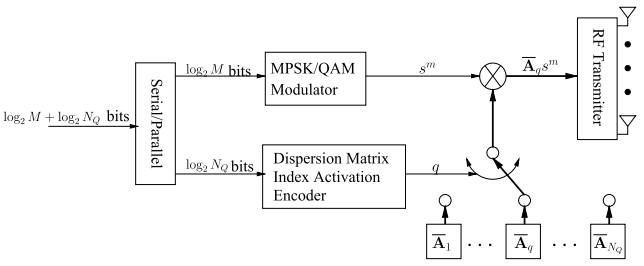


FIGURE 13. Schematic of the STSK transmitter.

In order to be able to benefit from a transmit diversity gain, the concept of Space-Time Shift Keying (STSK) was proposed by Sugiura *et al.* [39] in 2010, which is a combination of SM and LDC. The schematic of the STSK transmitter is portrayed in Fig. 13, which evolved from the LDC transmitter of Fig. 11. In more detail, the STSK transmitter of Fig. 13 assigns $\log_2 M$ bits to modulate a single MPSK/QAM symbol by the MPSK/QAM modulator, while $\log_2 N_Q$ bits are assigned to the dispersion matrix index activation encoder in order to select a single one out of a total number of N_Q dispersion matrices. Then the modulated symbol s^m is dispersed into both spatial and temporal dimensions by the activated dispersion matrix $\overline{\mathbf{A}}_q$, so that the STSK transmission matrix seen in Fig. 13 is given by $\mathbf{S} = s^m \overline{\mathbf{A}}_q$. It was demonstrated by Sugiura *et al.* [39] that after vectorizing the STSK scheme's received signal matrix, the SM detectors may be invoked for detecting the STSK's dispersion matrix index and modulated symbol index. As a result, the SM may rely on a low-complexity single-stream ML detector derived from

TABLE 3. Summary of major contributions on MIMO schemes.

Year	Author(s)	Topic	Contribution
1994	Paulraj and Kailath [107]	MIMO	Proposed the concept of multiplexing-oriented MIMO, where a high data-rate transmission was carried out by splitting it into low data-rate signals transmitted by spatially separated SDMA users.
1996	Foschini [30]	BLAST	Proposed the concept of D-BLAST, where a single data stream is de-multiplexed and then rotated and transmitted by N_T co-located TAs, so that the multiplexing gain may be pursued for a single user.
1998	Wolniansky *et al.* [32]	BLAST	Proposed the concept of V-BLAST, which elimiates the D-BLAST transmitter's diagonal-encoding rotator in order to simplify the real-time implementation.
1998	Alamouti [34]	STBC	Proposed the transmit diversity technique for the case of $N_T = 2$, which is often referred to as Alamouti's G2 STBC.
1999	Telatar [109]	BLAST	Proved that the BLAST MIMO systems have an ergodic capacity that may grow linearly with the number of antennas, provided that the BLAST MIMO system employs a large number of antennas.
1999	Tarokh *et al.* [35]	STBC	Proved that Alamouti's G2 STBC is the only full unity-rate code in the family of STBCs, and proposed Half-Rate (HR) STBCs for any number of TAs according to the Hurwitz-Radon theory [120], [121].
2000	Damen *et al.* [115]	BLAST	Proposed to inoke a sphere decoder for BLAST in order to strike a performance-complexity tradeoff between the ML BLAST detector and the LF-aided detectors.
2000	Sandhu and Paulraj [125]	STBC	Demonstrated that STBCs cannot achieve the full MIMO capacity except for a single case, which is Alamouti's G2-STBC systems associated with $N_R = 1$.
2001	Jafarkhani [126]	STBC	Proposed the concept of Quasi-Orthogonal (QO) STBC, which relaxed the STBC's orthogonality requirement in order to improve the STBC throughput.
2001	Ganesan and Stoica [122]–[124]	STBC	Proposed Amicable Orthogonal (AO) STBCs for any number of TAs according to the theory of amicable orthogonal design [120], which reduced the transmission delay of HR-STBCs and achieved an improved normalized throughput of 3/4 for the AO-STBCs associated with $N_T = 3$ and 4.
2002	Hassibi and Hochwald [36]	LDC	Proposed the capacity-improving LDC, which completely droped the STBC's orthogonality requirements in order to further improve the STBC capacity while retaining the full transmit diversity gain.
2002	Heath and Paulraj [38]	LDC	Proposed the capacity-achieving LDC, which simplified the LDC design of [36] so that both the full MIMO capacity and the full transmit diversity gain may be attained, provided that the parameters satisfy $N_Q \geq N_T N_P$.
2002	Sellathurai and Haykin [22]	BLAST	Proposed the exact MMSE solution incorporating the non-constant *a priori* probabilities for coded BLAST systems.
2003	Damen *et al.* [19]	BLAST	A comprehensive summary paper for the sphere decoding algorithms invoked by the uncoded BLAST systems.
2003	Zheng and Tse [29]	BLAST STBC	Quantified the classic MIMO design tradeoff between the attainable multiplexing and diversity gain.
2003	Hochwald and Brink [116]	BLAST	Proposed the first soft-decision-aided SD for BLAST, where a list of BLAST signal candidates was established by the hard-decision-aided SD and then the candidates in this list were processed by the MAP decoding algorithm.
2004	Vikalo *et al.* [117]	BLAST	Proposed the soft-decision-aided SD for BLAST, which incorporated the *a priori* information in sphere decoding.
2008	Studer *et al.* [118]	BLAST	Proposed the soft-output SD's VLSI implementation, which is further developed by the authors in [119], where the *a priori* LLRs are once again incorporated into the SD's VLSI implementation.
2008	Mesleh *et al.* [27]	SM	Analysed the SM that was firstly appeared in [26], where a single out of N_T TA is activated in order to transmit a single modulated MPSK/QAM symbol, so that a substantial complexity reduction may be achieved for the SM receiver by separately detecting the TA activation index and the classic modulated symbol index.
2008	Jeganathan *et al.* [134]	SM	Demonstrated that completely independently detecting the TA index and the modulated symbol index as seen in [27] resulted in an error floor. The authors further streamlined the ML MIMO detector's calculations for SM.
2008	Jeganathan *et al.* [135], [136]	SM (SSK)	Proposed the concept of SSK, where simply the TA activation index conveys the source information.
2008	Yang and Jiao [137]	SM	Proposed to invoke demodulator before detecting the TA activation index, and demonstrated that the SM capacity is higher than that of the SISO/SIMO systems.
2010	Sugiura *et al.* [39]	STSK	Proposed the concept of STSK in order to be able to benefit from a transmit diversity gain for the SM techniques, where a single out of N_Q LDC's dispersion matrix is activated in order to disperse a single modulated MPSK/QAM symbol, so that the low-complexity SM detectors may be invoked by the STSK receiver.
2011	Basar *et al.* [152]	SM	Proposed to achieve a transmit diversity gain for the original SM by activating more than one TAs in order to convey STBC codewords.
2011	Sugiura *et al.* [151]	STSK	Proposed the concept of Generalized STSK (GSTSK), where virtually all the MIMO schemes including V-BLAST, STBC, LDC, SM and STSK are included in the framework of dispersion matrix-aided space-time modulation.
2013	Xu *et al.* [139]	SM STSK	Proposed the reduced-scope SM detector both for uncoded and coded SM systems, which reduced the detection search scope while maintaining the optimum detection capability. The correlation between the TA index and the modulated symbol index was taken into account when detecting the TA index, and then only a single MPSK/QAM demodulator was invoked according to the already detected TA index.
2014	Rajashekar *et al.* [140]	SM	Summarized the Yang and Jiao's [137] technique as the hard-limiter-based SM detector, which invoked classic MPSK/QAM demodulators for all matched filter output elements first, and then the TA activation index detection was performed with the aid of the already demodulated MPSK/QAM symbols. The hard-limiter-based SM detector cannot be directly applied to coded SM systems.
2014	Renzo *et al.* [15]	SM STSK	A comprehensive summary paper for the recent development of SM.

the optimum V-BLAST MIMO detector at a lower detection complexity. Similarly, STSK is also capable of effectively reducing the LDC's detection complexity. Although a beneficial transmit diversity gain is obtained, the STSK's disadvantage over the SM is that all the STSK transmitter's RF chains have to be activated at the same time, as seen in Fig. 13, which loses the SM's advantage of using a single RF chain at any symbol-instant, as seen in Fig. 12. In 2011, Sugiura *et al.* [151] proceeded to conceive the concept of Generalized Space-Time Shift Keying (GSTSK), where virtually all the MIMO schemes including V-BLAST, STBC, LDC, SM and STSK are included in the framework of dispersion matrix-aided space-time modulation. Furthermore, in 2011, Basar *et al.* [152] arranged for achieving a transmit

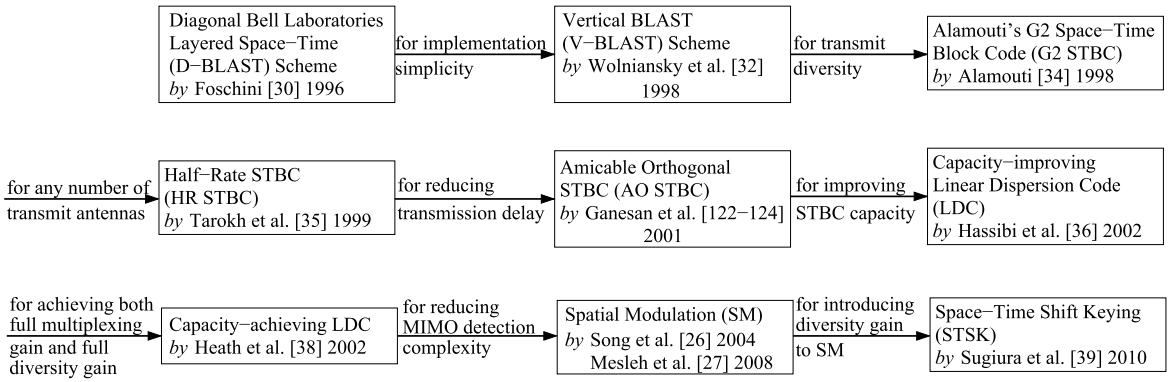


FIGURE 14. Historical chart for major milestones of Multiple-Input Multiple-Output (MIMO) schemes.

diversity gain for the original SM by activating more than one TAs in order to convey STBC codewords. This method has been further developed in [153]–[155] and all these schemes can be categorized under the framework of GSTSK according to the STBC's dispersion matrix expression. By contrast, Renzo and Haas [156] Renzo *et al.* [157], and Renzo and Haas [158] conceived an STBC transmit diversity aided SM scheme by employing idealistic orthogonal shapping filters, while Yang [159] proposed to employ orthogonal frequency-hopping codes. These schemes are however beyond the scope of GSTSK and they impose extra stringent hardware requirements.

Considering the fact that there are always $(N_T - 1)$ zero elements and a single non-zero element in a SM's transmission vector, any pair of SM codewords will share a total number of $(N_T - 2)$ zero-element positions. As a result, the average pairwise Euclidean distance between SM codewords is lower than that of its V-BLAST counterpart, which implies that SM may have a higher pairwise symbol error probability than its V-BLAST counterpart. For this reason, it is not likely for SM to outperform V-BLAST at the same system throughput and under the same hardware and software conditions. Indeed, this would only be possible for SM systems, under the employment of extra hardware for creating transmit diversity techniques [15], [158], [160], orthogonal shapping filters [15], [156], [158], or when aiming for a reduced SM throughput [161] or when using more complex ML SM detectors while opting for suboptimal V-BLAST detectors [27], [140], [141], [149], [152]. In summary, the ubiquitous performance versus complexity tradeoff manifests itself in the context of V-BLAST and SM, which is also the case for the LDC and the STSK arrangements. However, although SM may not be capable of outperforming V-BLAST, the performance differences between them are almost negligible compared to the performance loss imposed by employing an MMSE detector for V-BLAST. The same claim is valid, when STSK is compared to LDC.

The capacity of SM was evaluated by Yang and Jiao [137], who confirmed that the SM capacity is higher than that of the SISO/SIMO systems, but the full MIMO capacity cannot be achieved by the family of SM systems. Similarly, STSK also suffers from the same capacity loss against LDC. In order to mitigate this problem, the GSTSK proposed by Sugiura *et al.* [151] advocates transmitting more than one symbols. However, considering SM as an example, if more than one TAs are activated to transmit different symbols, the problem of IAI resurfaces, unless STBC codewords are transmitted. In order to tackle this IAI problem, Wang *et al.* [160] and Sugiura *et al.* [162] proposed sub-optimal interference-suppression receivers for Generalized SM and for the GSTSK, respectively. However, these arrangements are not consistent with the SM/STSK motivation of relying on low-complexity optimum ML receiver design. Against this background, Fu *et al.* [163] and Younis *et al.* [164] proposed the a Generalized SM (GSM) design, where multiple activated TAs may transmit the same symbol. In this way, although the IAI problem is avoided, the capacity improvement provided by the GSM remains limited, because the ergodic capacity is only maximized, when the signals transmitted by multiple TAs are independent and identically distributed (i.i.d.) [108], [109]. Therefore, the concept of a systematically normalized GSM/GSTSK arrangement that achieves an improved capacity without imposing IAI requires further research efforts.

In a netshell, the major contributions on MIMO schemes are summarized in Table 3 at a glance. Moreover, the key MIMO schemes are presented in the historical chart of Fig. 14, where the associated motivations are also indicated. It may be seen in Fig. 14 that the first stage of MIMO developments was motivated by the classic multiplexing and diversity tradeoff, while the SM scheme initiated a new stage of MIMO system design that aims for a reduced hardware and signal processing complexity, which is particularly promising in the context of large-scale MIMO systems [15]–[18] employing tens to hundreds of transmit/receive antennas.

C. NOVEL CONTRIBUTIONS AND STRUCTURE OF THE PAPER

The contributions offered by this paper are summarized as follows:

(1) First of all, we offer a survey on the above-mentioned pair of salient tradeoffs. Specifically, the multiplexing-diversity tradeoff motivated the classic V-BLAST, STBC and LDC designs. By contrast, the performance-complexity tradeoff motivated the conception of SM and STSK. The associated capacities and error probabilities are analysed and compared for different MIMO schemes.

(2) Secondly, we offer a comprehensive survey of MIMO detectors, including both the family of hard-decision to soft-decision schemes. EXIT charts are employed for analysing the performance of MIMO schemes in coded systems. The LLR accuracy test is also introduced in order to guarantee that the soft-decision MIMO detectors are capable of producing reliable LLRs for turbo detection assisted coded systems.

(3) Thirdly, we highlight the performance-complexity tradeoff, where reduced-complexity design guidelines are surveyed in the context of a wide range of MIMO detectors relying on both hard-decision and soft-decision techniques.

The structure of this paper is portrayed by Fig. 15. More explicitly, the classic MIMO schemes that are motivated by the multiplexing-diversity tradeoff are surveyed in Sec. II. The recently-developed MIMO schemes that are motivated by the performance-complexity tradeoff are surveyed in Sec. III. Finally, our conclusions are offered in Sec. IV.

The following notations are used throughout the paper. The notations $\ln(\cdot)$ and $\exp(\cdot)$ refer to natural logarithm and natural exponential functions, respectively. The notations $p(\cdot)$ and $\mathrm{E}(\cdot)$ denote the probability and the expectation, respectively. The notations $\Re(\cdot)$ and $\Im(\cdot)$ take the real part and the imaginary part of a complex number, respectively. The operations $\mathbb{M}(\cdot)$ and $\mathbb{M}^{-1}(\cdot)$ refer to the MPSK/QAM modulation and demodulation, respectively. The operation dec2bin$(\cdot)$ converts a decimal integer to binary bits, while bin2dec$(\cdot)$ converts binary bits to a decimal integer. The operations $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote the conjugate of a complex number, the transpose of a matrix and the Hermitian transpose of a complex matrix, respectively. The notations $\mathbf{A}_{u,-}$ and $\mathbf{A}_{-,v}$ refer to the u-th row and v-th column in matrix $\mathbf{A}$, respectively. The operation $\otimes$ represents the Kronecker product. The notation rvec$(\mathbf{A})$ forms a row-vector by taking the rows of matrix $\mathbf{A}$ one-by-one. Moreover, the operations diag$\{\mathbf{a}\}$ and Toeplitz$(\mathbf{a})$ create a diagonal matrix and a symmetric Toeplitz matrix from vector $\mathbf{a}$, respectively.

The acronyms V-BLAST(N_T,N_R)-MPSK/QAM as well as SM(N_T,N_R)-MPSK/QAM refer to the V-BLAST scheme and to the SM scheme equipped with N_T TAs and N_R RAs. Furthermore, the LDC and STSK schemes are denoted by the acronyms of LDC(N_T,N_R,N_P,N_Q)-MPSK/QAM and STSK(N_T,N_R,N_P,N_Q)-MPSK/QAM, respectively, where N_P and N_Q represent the number of symbol periods per transmission block and the total number of dispersion matrices employed, respectively.

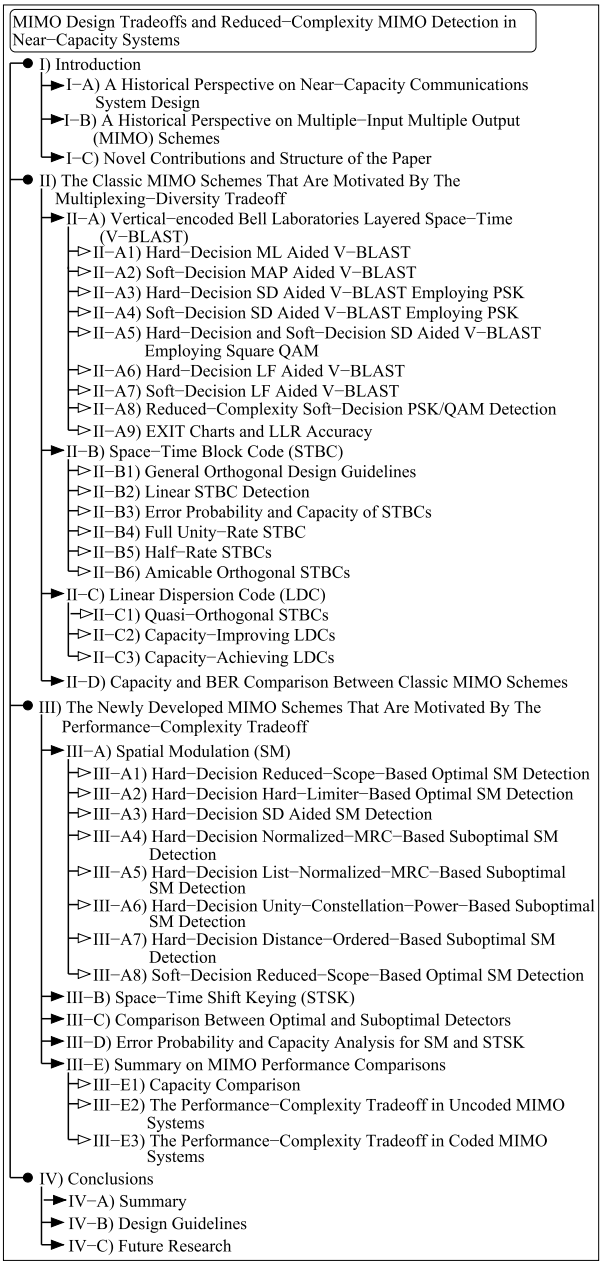


FIGURE 15. Structure of this paper.

II. THE CLASSIC MIMO SCHEMES THAT ARE MOTIVATED BY THE MULTIPLEXING-DIVERSITY TRADEOFF

As portrayed by Fig. 1, a typical MIMO system may employ N_T TAs and N_R RAs. Moreover, a transmission block of MIMO signals may be constituted by a total number of N_Q modulated MPSK/QAM symbols, and this transmission block may be transmitted over N_P symbol periods. Therefore, in the presence of the ubiquitous multipath fading as well as the Gaussian-distributed noise, the signal received by the N_R RAs over N_P Time Slot (TS) at the receiver may be modelled as:

$$\mathbf{Y} = \mathbf{SH} + \mathbf{V}, \tag{1}$$

TABLE 4. A brief summary of the transmitted matrices and parameters of classic MIMO representatives.

MIMO Scheme	Transmitted Matrix	Parameters					
V-BLAST [30]	$\mathbf{S} = [\, s_1, \ \cdots, \ s_{N_T} \,]$	$N_T > 1$	$N_R \geq 1$	$N_P = 1$	$N_Q = N_T$	$I = M^{N_T}$	$R = N_T \log_2 M$
G2-STBC [34]	$\mathbf{S} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}$	$N_T = 2$	$N_R \geq 1$	$N_P = 2$	$N_Q = 2$	$I = M^2$	$R = \log_2 M$
LDC [38]	$\mathbf{S} = \sum_{q=1}^{N_Q} \overline{\mathbf{A}}_q s_q$	$N_T > 1$	$N_R \geq 1$	$N_P > 1$	$N_Q \geq N_T N_P$	$I = M^{N_Q}$	$R = \frac{N_Q \log_2 M}{N_P}$
SM [27]	$\mathbf{S} = [\underbrace{0 \ \cdots \ 0}_{v-1}, s^m, \underbrace{0 \ \cdots \ 0}_{N_T - v}]$	$N_T > 1$	$N_R \geq 1$	$N_P = 1$	$N_Q = 1$	$I = M \cdot N_T$	$R = \log_2 N_T + \log_2 M$
STSK [39]	$\mathbf{S} = \overline{\mathbf{A}}_q s_q$	$N_T > 1$	$N_R \geq 1$	$N_P > 1$	$N_Q \geq 1$	$I = M \cdot N_Q$	$R = \frac{\log_2 N_Q + \log_2 M}{N_P}$

where the $(N_P \times N_T)$-element matrix $\mathbf{S}$ and the $(N_P \times N_R)$-element matrix $\mathbf{Y}$ represent the input and output signals of the MIMO channels. Furthermore, the $(N_T \times N_R)$-element $\mathbf{H}$ in (1) models the MIMO's Rayleigh fading channels, which is assumed to be time-invariant over N_P symbol periods. The $(N_P \times N_R)$-element AWGN matrix $\mathbf{V}$ in (1) models the independent and identically distributed (i.i.d.) zero-mean Gaussian random variables with a common complex variance of N_0, whose PDF is given by:

$$p(\mathbf{V}) = p(\mathbf{Y}|\mathbf{S}^i) = \frac{1}{(\pi N_0)^{N_R N_P}} \exp(-\frac{\|\mathbf{Y} - \mathbf{S}^i\mathbf{H}\|^2}{N_0}), \quad (2)$$

where there are a total of I combinations $\{\mathbf{S}^i\}_{i=0}^{I-1}$ for the MIMO transmission matrix $\mathbf{S}$ in (1). The MIMO transmission matrices and parameters are briefly summarized in Table 4. The details of these classic MIMO schemes will be introduced later.

The Continuous-input Continuous-output Memoryless Channel (CCMC) capacity of the MIMO channels is given by maximizing the mutual information between the input signal and the output signal per channel use as [40]:

$$C^{CCMC}(SNR) = \max_{p(\mathbf{S})} \frac{1}{N_P} H(\mathbf{Y}) - \frac{1}{N_P} H(\mathbf{Y}|\mathbf{S}). \quad (3)$$

Based on $p(\mathbf{V})$ of (2), we have $H(\mathbf{Y}|\mathbf{S}) = H(\mathbf{V}) = H\left[\text{rvec}(\mathbf{V})\right] = \log_2 \det\left[\pi e N_0 \mathbf{I}_{N_P N_R}\right]$. Furthermore, in order to maximize the entropy $H(\mathbf{Y})$ in (3), both the input signal and the output signal have be Gaussian distributed. As a result, the autocorrelation of the i.i.d. Gaussian-distributed input signals is given by $\mathrm{E}\left[\text{rvec}(\mathbf{S})^H \text{rvec}(\mathbf{S})\right] = \frac{1}{N_T}\mathbf{I}_{N_P N_T}$, which complies with the transmit power constraint of $\mathrm{E}\left\{\text{tr}\left[\text{rvec}(\mathbf{S})^H \text{rvec}(\mathbf{S})\right]\right\} = N_P$. Furthermore, the resultant entropy of the vectorized received signal $\text{rvec}(\mathbf{Y}) = \text{rvec}(\mathbf{S})\left(\mathbf{I}_{N_P} \otimes \mathbf{H}\right) + \text{rvec}(\mathbf{V})$ is given by $H(\mathbf{Y}) = H\left[\text{rvec}(\mathbf{Y})\right] = \log_2 \det\{\pi e[\frac{1}{N_T}\left(\mathbf{I}_{N_P} \otimes \mathbf{H}^H\right)\left(\mathbf{I}_{N_P} \otimes \mathbf{H}\right) + N_0 \mathbf{I}_{N_P N_R}]\}$. Therefore, the ergodic CCMC capacity of (3) that is averaged over all channel realizations is given by:

$$\begin{aligned} C^{CCMC}(SNR) &= \frac{1}{N_P}\mathrm{E}\left\{\log_2 \det(\mathbf{I}_{N_P N_R} + \frac{\eta}{N_T}\left[\mathbf{I}_{N_P} \otimes (\mathbf{H}^H\mathbf{H})\right]\right\} \\ &= \mathrm{E}\left[\log_2 \det(\mathbf{I}_{N_R} + \frac{\eta}{N_T}\mathbf{H}^H\mathbf{H})\right], \end{aligned} \quad (4)$$

where $SNR = 10\log_{10}\eta$ is the normalized signal-to-noise ratio $\eta = \frac{1}{N_0}$ represented on the logarithmic decibel scale. When the number of TAs grows towards infinity, the mutual information of (4) may be further extended as [108], [109] $\lim_{N_T\to\infty} C^{CCMC}(SNR) = \log_2 \det(\mathbf{I}_{N_R} + \eta\mathbf{I}_{N_R}) = N^{\min}\log_2(1+\eta)$, where we have $\lim_{N_T\to\infty} \mathrm{E}\left(\frac{1}{N_T}\mathbf{H}^H\mathbf{H}\right) = \mathbf{I}_{N_R}$, while $N^{\min} = \min(N_T, N_R)$ represents the minimum of the number of the TAs and RAs. This implies that as the number of antennas grows, the MIMO capacity may grow linearly with $\min(N_T, N_R)$. Let us recall that the CCMC capacity of SIMO systems is given in [40], [109], and [165] $C^{\text{CCMC}}(SNR) = E\left[\log_2(1+\eta\cdot\|\mathbf{h}\|^2)\right]$, which grows logarithmically with N_R, where $\mathbf{h}$ refer to the N_R-element SIMO fading vector. Therefore, compared to SIMO systems, the MIMO systems are capable of providing a higher data rate without requiring more signal bandwidth. We will demonstrate in Sec. that the STBCs based on orthogonal design [34], [35], [166] and those relying on the Amicable orthogonal design [122]–[124] cannot achieve the full MIMO capacity of (4). This is because the diversity-oriented STBC schemes transmit symbols that are repetitive in both space and time, which implies that the i.i.d. input signal condition of $\mathrm{E}\left[\text{rvec}(\mathbf{S})^H \text{rvec}(\mathbf{S})\right] = \frac{1}{N_T}\mathbf{I}_{N_P N_T}$ is not satisfied.

When the supposedly continuous Gaussian-distributed input signal is discretized for transmitting practical MPSK/QAM symbols, the CCMC capacity of (3) has to be replaced by the more realistic measure of Discrete-input Continuous-output Memoryless Channel (DCMC) capacity of [12], [66], [167]:

$$\begin{aligned} C^{DCMC}(SNR) = \max_{\{p(\mathbf{S}^i)\}_{i=0}^{I-1}} \frac{1}{N_P} \sum_{i=0}^{I-1} \int p(\mathbf{Y}|\mathbf{S}^i)p(\mathbf{S}^i) \\ \cdot \log_2 \frac{p(\mathbf{Y}|\mathbf{S}^i)}{\sum_{\tilde{i}=0}^{I-1} p(\mathbf{Y}|\mathbf{S}^{\tilde{i}})p(\mathbf{S}^{\tilde{i}})} d\mathbf{Y}. \end{aligned} \quad (5)$$

The DCMC capacity of (5) is maximized, when the MIMO transmission matrix candidates are equiprobable, i.e. we have $\{p(\mathbf{S}^i) = \frac{1}{I}\}_{i=0}^{I-1}$. Based on the PDF $p(\mathbf{Y}|\mathbf{S}^i)$ given by (2), the DCMC capacity of (5) may be further simplified as:

$$\begin{aligned} C^{DCMC}(SNR) &= \frac{1}{I \cdot N_P} \sum_{i=0}^{I-1} \mathrm{E}\left\{\log_2\left[\frac{I \cdot p(\mathbf{Y}|\mathbf{S}^i)}{\sum_{\tilde{i}=0}^{I-1} p(\mathbf{Y}|\mathbf{S}^{\tilde{i}})}\right]\right\} \\ &= R - \frac{1}{I \cdot N_P} \sum_{i=0}^{I-1} \mathrm{E}\left\{\log_2\left[\sum_{\tilde{i}=0}^{I-1} \exp(\Psi_{i,\tilde{i}})\right]\right\}, \end{aligned} \quad (6)$$

where we have $\Psi_{i,\bar{i}} = \frac{-\|(\mathbf{S}^i-\mathbf{S}^{\bar{i}})\mathbf{H}+\mathbf{V}\|^2+\|\mathbf{V}\|^2}{N_0}$, while the MIMO throughput is given by $R = \frac{\log_2 I}{N_P}$, which is expected to be achieved by the DCMC capacity of $C^{DCMC}(SNR)$ in the high-SNR region.

In particular, the STBC throughput is given by $R = \frac{N_Q \text{BPS}}{N_P}$. It was demonstrated in [35] that Alamouti's G2-STBC associated with $N_T = N_P = N_Q = 2$ is the only full unity-rate orthogonal STBC employing a complex-valued signal constellation, where the STBC's normalized throughput is defined as $\overline{R} = N_Q/N_P$. When $N_T > 2$ is used for orthogonal STBC schemes, we always have $\overline{R} < 1$ for complex-valued signalling. Hence the orthogonal STBCs do not have the advantage of a higher data rate than SIMO schemes. By contrast, the V-BLAST's maximum achievable rate of (6) is given by $R = N_T$BPS, which is N_T times higher than that of the SIMO. Again, the V-BLAST's feature of maximized MIMO throughput is often interpreted as *multiplexing gain*.

Nonetheless, a higher attainable capacity cannot guarantee a lower error probability. Let us now consider the average BER of a MIMO scheme, which is given in [157], [168], and [169]:

$$\overline{P}_{e,bit} = \mathrm{E}\left\{\sum_{i=0}^{I-1}\sum_{\bar{i}=0,\bar{i}\neq i}^{I-1}\frac{d_H(i,\bar{i})}{I\log_2 I}p(\hat{\mathbf{S}}=\mathbf{S}^{\bar{i}}|\mathbf{S}^i)\right\}, \quad (7)$$

where $d_H(i,\bar{i})$ refers to the Hamming distance between the bit-mappings of $\mathbf{S}^i$ and $\mathbf{S}^{\bar{i}}$, which may be directly obtained by conveying the indices i and $\bar{i}$ back to $\log_2 I$ bits. Furthermore, the average Pairwise Error Probability (PEP) $\mathrm{E}\left\{p(\mathbf{S}^i \to \mathbf{S}^{\bar{i}})\right\}$, which is the average probability $\mathrm{E}\left\{p(\hat{\mathbf{S}}=\mathbf{S}^{\bar{i}}|\mathbf{S}^i)\right\}$ of choosing $\mathbf{S}^{\bar{i}}$ when $\mathbf{S}^i$ was transmitted, may be expressed as [66], [165], [170]–[173]:

$$\begin{aligned}\mathrm{E}\left\{p(\mathbf{S}^i \to \mathbf{S}^{\bar{i}})\right\} &= \mathrm{E}\left\{p\left(\|\mathbf{Y}-\mathbf{S}^{\bar{i}}\mathbf{H}\|^2 < \|\mathbf{V}\|^2\right)\right\}\\ &\leq \mathrm{E}\left\{Q\left[\sqrt{\frac{\left\|(\mathbf{S}^i-\mathbf{S}^{\bar{i}})\mathbf{H}\right\|^2}{2N_0}}\right]\right\}\\ &\leq (0.25\eta)^{-\mathrm{rank}(\mathbf{\Delta})\cdot N_R}\left[\prod_{k=1}^{\mathrm{rank}(\mathbf{\Delta})}\lambda_k(\mathbf{\Delta})\right]^{-N_R},\end{aligned} \quad (8)$$

where $Q(\cdot)$ represents the integral form of the Q-function, while $\{\lambda_k(\mathbf{\Delta})\}$ and $\mathrm{rank}(\mathbf{\Delta})$ refer to the k-th eigenvalue of matrix $\mathbf{\Delta}$ and the rank of $\mathbf{\Delta}$, respectively.

As discussed in [171]–[173], (8) suggests that there are two major factors that may minimize the error probability in the high-SNR region, which are often referred to as the rank criterion and the determinant criterion in the literature of analysing the MIMO systems' performance. In more detail, firstly, it may be observed in (8) that the rate of decline for the term $(0.25\eta)^{-\mathrm{rank}(\mathbf{\Delta})\cdot N_R}$ with respect to the SNR is explicitly determined by $\mathrm{rank}(\mathbf{\Delta})\cdot N_R$. Therefore, this rank criterion indicates that the full MIMO diversity is given by $\min(N_T, N_P)\cdot N_R$, where the full rank of $\mathbf{\Delta}$ is the minimum between N_T and N_P. Furthermore, when $\mathbf{\Delta}$ achieves full rank, the second term $\left[\prod_{k=1}^{\mathrm{rank}(\mathbf{\Delta})}\lambda_k(\mathbf{\Delta})\right]^{N_R}$ in (8) is a function of the determinant of $\mathbf{\Delta}$. As a result, this so-called determinant criterion indicates that a higher gain is achieved by maximizing the minimum determinant $\det(\mathbf{\Delta})$ over all legitimate combinations of $\mathbf{S}^i$ and $\mathbf{S}^{\bar{i}}$, which is achieved when $\mathbf{\Delta}$ is unitary. This condition may be guaranteed by both the classic STBCs based on orthogonal design [34], [35], [166] and the STBCs relying on the Amicable orthogonal design criterion [122]–[124]. By contrast, the classic V-BLAST associated with $N_P = 1$ does not minimize the error probability in rank and determinant criteria, which is due to the fact that V-BLAST has neither transmit diversity - since we have $\min(N_T, N_P) = 1$ - nor has it unitary transmission matrices. The STBC's feature of minimizing the PEP in the high-SNR region according to the rank and determinant criteria is often referred to as *diversity gain*.

The tradeoff between the diversity gain $\overline{D}$ and the multiplexing gain $\overline{R} = \frac{N_Q}{N_P}$ is quantified as $\overline{D} = (N_T - \overline{R})(N_R - \overline{R})$ in [29], which portrays the diversity and multiplexing as rivals in MIMO systems design. As a breakthrough, the development of LDC [36]–[38] suceeded in perfectly accommodating this tradeoff. In more detail, the LDC transmission model of [38], which is summarized in Table 4, is capable of achieving the full MIMO capacity, provided that the parameters satisfy $N_Q \geq N_T N_P$, which results in a maximized multiplexing gain of $\overline{R} \geq N_T$. Furthermore, the LDC of [38] may also retain the full diversity gain of $\overline{D} = \min(N_T, N_P)\cdot N_R$, when the dispersion matrices $\{\overline{\mathbf{A}}_q\}_{q=1}^{N_Q}$ are generated according to the rank and determinant criteria. As a result, the best LDCs generated from a sufficiently exhaustive random search are capable of outperforming both V-BLAST and STBC in MIMO systems.

In this section, we focus our attention on the classic MIMO schemes that are motivated by the multiplexing-diversity tradeoff, where V-BLAST, STBC and LDC are introduced in Sec. II-A, Sec. II-B and Sec. II-C, respectively.

A. VERTICAL-ENCODED BELL LABORATORIES LAYERED SPACE-TIME (V-BLAST)

The classic V-BLAST MIMO, which multiplexes N_T data streams with the aid of N_T TAs, maximizes the MIMO capacity and throughput. The challenge in V-BLAST system design is to deal with the IAI at an affordable signal processing complexity, which clearly strikes a tradeoff between performance and complexity, as portrayed by Fig. 2. Against this background, the ML/MAP detectors, the SD and the LF are introduced for both uncoded and coded V-BLAST systems.

1) HARD-DECISION ML AIDED V-BLAST

The schematic of V-BLAST transmitter is portrayed by Fig. 9(b), where a total of $N_Q = N_T$ modulated symbols are transmitted by the N_T TAs during $N_P = 1$ symbol

periods. Therefore, the N_T-element V-BLAST transmission row-vector

$$\begin{aligned}\mathbf{S} &= \left[s_1, \quad \cdots, \quad s_{N_T} \right] \\ &= \left[\tfrac{1}{\sqrt{N_T}} s^{m_1}, \quad \cdots, \quad \tfrac{1}{\sqrt{N_T}} s^{m_{N_T}} \right], \end{aligned} \tag{9}$$

where the MPSK/QAM symbols are separately modulated as $\{s^{m_v} = \mathbb{M}(m_v)\}_{v=1}^{N_T}$. Upon obtaining the N_R-element received signal row-vector $\mathbf{Y}$ of (1), the following *a posteriori* probability may be maximized over the entire set of $I = M^{N_T}$ candidates $\{\mathbf{S}^i\}_{i=0}^{I-1}$ for the transmit vector $\mathbf{S}$ in (1) as:

$$p(\mathbf{S}^i|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{S}^i)p(\mathbf{S}^i)}{\sum_{\forall \mathbf{S}^i} p(\mathbf{Y}|\mathbf{S}^i)p(\mathbf{S}^i)}. \tag{10}$$

In uncoded V-BLAST systems, the *a priori* probability $\{p(\mathbf{S}^i)\}_{\forall \mathbf{S}^i}$ may be assumed to be a constant of $\frac{1}{M^{N_T}}$ for the equiprobable source. Furthermore, the conditional probability $\{p(\mathbf{Y}|\mathbf{S}^i)\}_{\forall \mathbf{S}^i}$ in (10) is given by (2). Therefore, the hard-decision ML aided uncoded V-BLAST detection may be expressed as:

$$\hat{\mathbf{S}} = \arg \min_{\forall \mathbf{S}^i} \|\mathbf{Y} - \mathbf{S}^i\mathbf{H}\|^2. \tag{11}$$

2) SOFT-DECISION MAP AIDED V-BLAST

In coded V-BLAST systems, the soft-decision MAP detector may produce the *a posteriori* LLRs as [12], [50], [174]–[176]:

$$\begin{aligned} L_p(b_k) &= \ln \frac{\sum_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=1}} p(\mathbf{S}^i|\mathbf{Y})}{\sum_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=0}} p(\mathbf{S}^i|\mathbf{Y})} \\ &= \ln \frac{\sum_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=1}} p(\mathbf{Y}|\mathbf{S}^i)p(\mathbf{S}^i)}{\sum_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=0}} p(\mathbf{Y}|\mathbf{S}^i)p(\mathbf{S}^i)}, \end{aligned} \tag{12}$$

where the subsets $\{\mathbf{S}^i\}_{b_k=1}$ and $\{\mathbf{S}^i\}_{b_k=0}$ represent the V-BLAST combinations set for $\mathbf{S}$ in (1), when the specific bit b_k is fixed to be 1 and 0, respectively. Let us assume that all source information bits are mutually independent. Then the *a priori* probabilities $\{p(\mathbf{S}^i)\}_{i=0}^{I-1}$ may be expressed as $p(\mathbf{S}^i) = \prod_{\bar{k}=1}^{N_T \mathrm{BPS}} \frac{\exp[\widetilde{b}_{\bar{k}} L_a(b_{\bar{k}})]}{1+\exp[L_a(b_{\bar{k}})]}$, where $\{L_a(b_{\bar{k}})\}_{\bar{k}=1}^{N_T \mathrm{BPS}}$ denote the *a priori* LLRs gleaned from a channel decoder, while $[\widetilde{b}_1 \cdots \widetilde{b}_{N_T \mathrm{BPS}}] = \mathrm{dec2bin}(i)$ refers to the bit mapping of V-BLAST to the signal $\mathbf{S}^i$. Therefore, the Log-MAP algorithm of (12), may be rewritten as:

$$L_p(b_k) = \ln \frac{\sum_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=1}} \exp(d^i)}{\sum_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=0}} \exp(d^i)}, \tag{13}$$

where the probability metric d^i in (13) is given by:

$$d^i = -\frac{\|\mathbf{Y} - \mathbf{S}^i\mathbf{H}\|^2}{N_0} + \sum_{\bar{k}=1}^{N_T \mathrm{BPS}} \widetilde{b}_{\bar{k}} L_a(b_{\bar{k}}). \tag{14}$$

We note that the common constant of $\frac{1}{\prod_{\bar{k}=1}^{N_T \mathrm{BPS}} \{1+\exp[L_a(b_{\bar{k}})]\}}$ in all $\{p(\mathbf{S}^i)\}_{i=0}^{I-1}$ is eliminated by the division operation in (12). The resultant extrinsic LLRs produced by the Log-MAP algorithm may be further expressed as $L_e(b_k) = L_p(b_k) - L_a(b_k)$.

In practice, the Log-MAP algorithm of (13) may be simplied by the low-complexity Max-Log-MAP [50] as:

$$L_p(b_k) = \max_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=1}} d^i - \max_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=0}} d^i, \tag{15}$$

which imposes a performance loss owing to the fact that only the pair of maximum *a posteriori* probabilities associated with $b_k = 1$ and $b_k = 0$ are taken into account. In order to mitigate this problem, the so-called Approx-Log-MAP algorithm [51], [177] may be invoked as:

$$L_p(b_k) = \mathrm{jac}_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=1}} d^i - \mathrm{jac}_{\forall \mathbf{S}^i \in \{\mathbf{S}^i\}_{b_k=0}} d^i, \tag{16}$$

where the corrected Jacobian algorithm jac compensates for the inaccuracy imposed by the maximization operation of (15) as $\mathrm{jac}(d^1, d^2) = \max(d^1, d^2) + \delta(|d^1 - d^2|)$. The additional term of $\delta(|d^1 - d^2|)$ takes into account the difference between d^1 and d^2 according to a lookup table [51], [177].

The ML/MAP aided V-BLAST detection introduced in this section requires us to evaluate and compare all $I = 2^{N_T \mathrm{BPS}}$ combinations of MIMO signals, which imposes an unaffordable detection complexity. Therefore, as portrayed by Fig. 2, the SD and the family of linear receivers may be introduced in order to visit all the N_T parallel MPSK/QAM constellation diagrams separately, so that the signal processing complexity may be reduced for V-BLAST detection.

3) HARD-DECISION SD AIDED V-BLAST EMPLOYING PSK

In order to invoke the SD, the V-BLAST receiver may apply the classic QR decomposition to $\mathbf{H}^H$ [116]–[119] as follows:

$$\mathbf{H}^H = \left[\mathbf{Q}, \mathbf{Q}' \right] \begin{bmatrix} \mathbf{U} \\ \mathbf{0} \end{bmatrix}, \tag{17}$$

where $\left[\mathbf{Q}, \mathbf{Q}' \right]$ is a $(N_R \times N_R)$-element unitary matrix, and the $(N_R \times N_T)$-element submatrix $\mathbf{Q}$ has orthogonal columns satisfying $\mathbf{Q}^H\mathbf{Q} = I_{N_T}$. Furthermore, $\mathbf{U}$ in (17) is a $(N_T \times N_T)$-element upper triangular matrix, while $\mathbf{0}$ refers to a $[(N_R - N_T) \times N_T]$-element all-zero matrix. It is a natural requirement that we have $N_R \geq N_T$, so that the QR decomposition of (17) may proceed. The generalized rank-deficient scenario of $N_R < N_T$ is discussed in [178]–[180], where the SD is recommended for detecting N_R symbols, while the ML detector is invoked for the remaining symbols. For the sake of simplicity, we only consider the situation of $N_R \geq N_T$ for the SD aided V-BLAST in this section, which is generally compatible with the industrial MIMO standards [181], [182]. According to (17), the $(N_T \times N_R)$-element fading channel matrix $\mathbf{H}$ may now be represented as:

$$\mathbf{H} = (\mathbf{QU})^H = \mathbf{LQ}^H, \tag{18}$$

where $L = \mathbf{U}^H$ is a $(N_T \times N_T)$-element lower triangular matrix. As a result, the received signal model of (1) may be modified as:

$$\mathbf{YQ} = \mathbf{SL} + \mathbf{VQ}, \tag{19}$$

where **VQ** has exactly the same statistics as the AWGN matrix **V**. Therefore, the ML decision metric of (11), which may also be referred to as Euclidean Distance (ED), may be rewritten as:

$$\left\|\widetilde{\mathbf{Y}} - \mathbf{SL}\right\|^2 = \sum_{v=1}^{N_T} \left| \widetilde{Y}_v - \sum_{t=v}^{N_T} l_{t,v} s_t \right|^2. \tag{20}$$

The N_T-element row-vector $\widetilde{\mathbf{Y}} = \mathbf{YQ}$ in (20) is defined in (19), and $\{\widetilde{Y}_v\}_{v=1}^{N_T}$ are elements taken from $\widetilde{\mathbf{Y}}$. Furthermore, $\{\{l_{t,v}\}_{t=v}^{N_T}\}_{v=1}^{N_T}$ and $\{s_t\}_{t=v}^{N_T}$ in (20) are elements from the lower triangular matrix **L** defined in (18) and elements from the V-BLAST transmit vector **S** in (1), respectively.

The SD aims for finding the specific detection candidates that lie within the decoding sphere radius R, which is formulated as:

$$\left\|\widetilde{\mathbf{Y}} - \mathbf{SL}\right\|^2 < R^2. \tag{21}$$

This detection problem may be solved step-by-step. According to the ED of (20), the Partial Euclidean Distance (PED) evaluated by the SD may be defined as:

$$d_v = \sum_{\tilde{v}=v}^{N_T} \left| \widetilde{Y}_{\tilde{v}} - \sum_{t=\tilde{v}}^{N_T} l_{t,\tilde{v}} s_t \right|^2 = d_{v+1} + \Delta_v, \tag{22}$$

where the PED increment Δ_v is given by:

$$\Delta_v = \left| \left(\widetilde{Y}_v - \sum_{t=v+1}^{N_T} l_{t,v} s_t \right) - l_{v,v} s_v \right|^2. \tag{23}$$

The only variable in the PED increment of (23) is s_v, as elements $\{s_t\}_{t=v+1}^{N_T}$ are known from previous SD decisions.

The SD constellation search strategies may include both the Fincke-Pohst enumeration strategy of [20], [183], and [184], and the Schnorr-Euchner search strategy of [21] and [185], which define how the SD visits different constellation points at a specific SD index. For a SD based on the PED increment of (23), the Pohst searching enumeration strategy of [20], [183], and [184] requires the SD to enumerate all candidates for s_v within the SNR-dependent decoding sphere, as defined by the condition of $d_v < R^2$, which is examplified by Fig. 16(a). The Schnorr-Euchner search strategy of [21] and [185] efficiently refines the Pohst strategy, where the priorities of all the legitimate candidates for s_v are ranked according to the increasing order of their corresponding PED increment values Δ_v. Therefore, when the SD reaches a specific index v for the first time, the candidate associated with the highest priority is visited. Then, when the SD reaches v again for the m-th time, the candidate associated with the m-th highest priority should be visited. In this way, the SD always knows, which specific candidate should be examined without repeating the enumeration. Considering the hard-decision SD based on the PED increment of (23) as an example, the MPSK candidate associated with the highest priority may be directly obtained by rounding the phase of the decision variable $\widetilde{z}_v^{SD} = (\widetilde{Y}_v - \sum_{t=v+1}^{N_T} l_{t,v} s_t)(l_{v,v})^*$

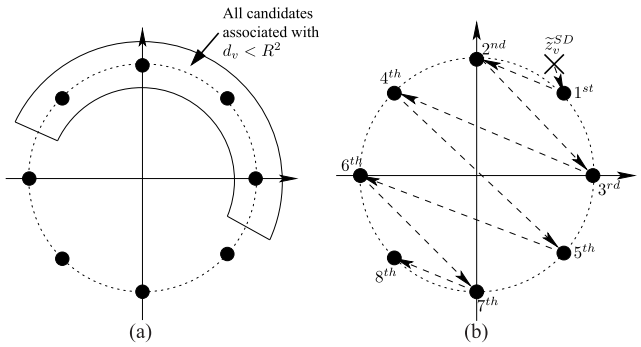


FIGURE 16. Examples of the SD constellation-search strategies of the Fincke-Pohst enumeration strategy of [20], [183], [184] and the Schnorr-Euchner search strategy of [21], [185], where the 8PSK constellation is employed. (a) Fincke-Pohst. (b) Schnorr-Euchner.

to the nearest MPSK index as $\check{m}_v = \lfloor p_v \rceil$, where $p_v = \frac{M}{2\pi} \angle \widetilde{z}_v^{SD}$. Then the remaining MPSK constellation points may be visited in a zigzag fashion by the SD. In more details, if the phasor index $\check{m}_v$ is rounded down from p_v, i.e. we have $\check{m}_v \leq p_v$, then the SD may visit the remaining constellation points according to the steps of $\check{m}_v = \check{m}_v + 1$, $\check{m}_v = \check{m}_v - 2$, $\check{m}_v = \check{m}_v + 3$, etc. By contrast, for the case of $\check{m}_v > p_v$, the SD based steps of visiting constellation points are $\check{m}_v = \check{m}_v - 1$, $\check{m}_v = \check{m}_v + 2$, $\check{m}_v = \check{m}_v - 3$, etc. The Schnorr-Euchner search strategy is examplified by Fig. 16(b).

The SD tree-search strategies may include both the breadth-first (K-Best) [186]–[188] and depth-first solutions [19], [21], [189], which define how the SD traverses across different SD indices $v \in \{1, \cdots, N_T\}$. The breadth-first (K-Best) tree search strategy, reduces the SD index from $v = N_T$ down to $v = 1$, where only K candidates associated with the higher priorities are retained at each level. The major advantage of the breadth-first (K-Best) tree search strategy is that the total number of nodes visited by the SD is constant, but K-best algorithm is unable to guarantee to spot the ML solution. The depth-first tree search strategy, which is also popularly adopted by the Multiple-Symbol Differential Sphere Detection (MSDSD) aided noncoherent schemes of [190]–[194], commences its search by decreasing the SD index from $v = N_T$ down to $v = 1$ as well, but only the best candidate is visited on each level. When the SD index of $v = 1$ is reached, the SD radius is shrunk to be consistent with the newly found contender candidate **S**. Then the SD index is increased again in order to check if there is any other nodes that may lie inside the updated decoding sphere. If a new valid candidate is found within the sphere at any value of the SD index v, the SD index may decrement down towards $v = 1$ again. Otherwise, the search may terminate, once the SD index of $v = N_T$ is reached. Therefore, the depth-first tree search has a nonconstant complexity, but spotting the optimum ML solution may be only guaranteed, if the initial SD radius is set to be sufficiently large.

The breadth-first and depth-first tree-search strategies are exemplified in Fig. 17. First of all, Fig. 17 shows that both

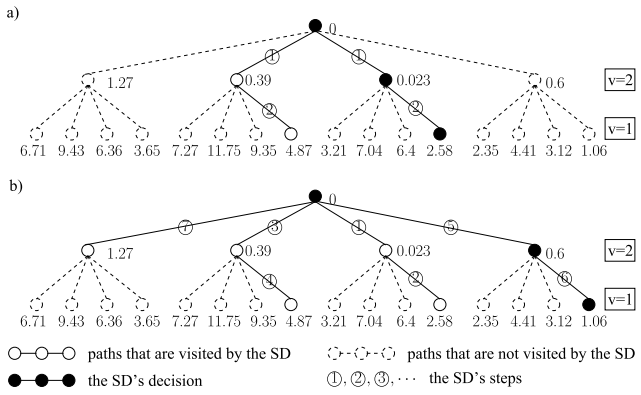


FIGURE 17. Examples of the SD tree-search strategies of the breadth-first (K-Best) [186]–[188] and depth-first solutions [19], [21], [189], where the V-BLAST(2,2)-QPSK is employed at $E_b/N_0 = 0$ dB. The PEDs d_v of (22) are labelled for each node. a) Breadth-First Tree-Search Strategy ($K = 2$. b) Depth-First Tree-Search Strategy.)

strategies effectively avoid visiting all the valid nodes, which results in a reduced complexity compared to the ML aided V-BLAST. For the breadth-first associated with $K = 2$ in Fig. 17, only two nodes that have lower PEDs d_v of (22) are visited for each SD index of $v = 2$ and $v = 1$ in Steps ① and ②, respectively. However, the breadth-first decision associated with the ED of 2.58 is not the ML solution in Fig. 17, where the lowest ED is given by 1.06. For the depth-first strategy in Fig. 17, when the SD visits $v = 2$ and $v = 1$ for the first time in Steps ① and ②, only the nodes associated with the lowest PEDs of $d_2 = 0.023$ and $d_1 = 2.58$ are visited for SD indices $v = 2$ and $v = 1$, respectively. The SD radius R is updated according to the ED of $d_1 = 2.58$ as $R^2 = 2.58$ in Step ②, and then the SD index is increased to $v = 2$ in order to check the next node associated with the second lowest PED of $d_2 = 0.39$ in Step ③, which is lower than $R^2 = 2.58$. Hence the SD index is decreased to $v = 1$, where the best node has a PED $d_1 = 4.87$ that is higher than $R^2 = 2.58$ in Step ④. The SD index is increased and then decreased again in Steps ⑤ and ⑥, respectively, where the SD radius is updated according to $d_1 = 1.06$ in Step ⑥. The SD index is increased again in Step ⑦, where the visited node has a PED of $d_2 = 1.27$ that is higher than $R^2 = 1.06$. Hence the SD terminates the search. It can be seen in Fig. 17 that the depth-first strategy may visit more nodes than the breadth-first strategy, but the ML solution associated with the lowest ED of 1.06 may be obtained by the depth-first decision, provided that the initial SD radius is set to be sufficiently large.

In practice, a possible choice of the initial SD radius R may be found from the statistical properties of the ED of (21) as $R^2 = JN_RN_0 - \mathbf{Y}\left[\mathbf{I}_{N_R} - \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}\right]\mathbf{Y}^H$ [116], where an integer $J \geq 1$ may be selected in order to strike a trade-off between the performance and complexity. Furthermore, it was demonstrated in [195] that both the selection of an SNR-dependent R and the potential SD search failure may be avoided by defining the intial SD radius R as the distance between the received signal and the MMSE solution formulated as $R^2 = \|\mathbf{Y} - \mathbf{Y}^{MMSE}\|^2$, where the MMSE solution is given by $\mathbf{Y}^{MMSE} = \mathbf{Y}(\mathbf{H}^H\mathbf{H} + N_0N_T\mathbf{I}_{N_R})^{-1}\mathbf{H}^H$. The details of this MMSE solution will be elaborated on in Sec. II-A6.

4) SOFT-DECISION SD AIDED V-BLAST EMPLOYING PSK

In order to invoke soft-decision SD for the Max-Log-MAP optimum V-BLAST detection, it may be observed in (15) that the Max-Log-MAP algorithm aims to find the maximum probability metric, which is similar to the action of the hard-decision ML V-BLAST detection of (11). Therefore, the problem of finding the maximum probability metric d^i of (14) may be transformed to the problem of searching for the minimum ED formulated as:

$$d = \frac{\sum_{v=1}^{N_T}\left|\widetilde{Y}_v - \sum_{t=v}^{N_T} l_{t,v}s_t\right|^2}{N_0} - \sum_{v=1}^{N_T}\left\{\sum_{\bar{k}_v=1}^{\text{BPS}}\left[\widetilde{b}_{\bar{k}_v}L_a(b_{\bar{k}_v}) - \overline{C}_{a,\bar{k}_v}^{SD}\right]\right\}, \quad (24)$$

which is obtained by toggling the polarity of the probability metric d^i of (14). The first term in (24) is revised from the hard-decision SD's ED of (20). The second term in (24) is revised from the *a priori* probability term of (14), where an extra constant $\overline{C}_{a,\bar{k}_v}^{SD} = \frac{1}{2}\left[|L_a(b_{\bar{k}_v})| + L_a(b_{\bar{k}_v})\right]$ is introduced in order to guarantee that the ED of (24) remains non-negative all the time [119], [196].

As a result, the maximization operation of the Max-Log-MAP of (15) is transformed into finding the optimal candidate that lies within the decoding sphere radius R, where the SD may evaluate the PED according to the ED of (24) as:

$$d_v = \frac{\sum_{\bar{v}=v}^{N_T}\left|\widetilde{Y}_{\bar{v}} - \sum_{t=\bar{v}}^{N_T} l_{t,\bar{v}}s_t\right|^2}{N_0} - \sum_{\bar{v}=v}^{N_T}\left\{\sum_{\bar{k}_{\bar{v}}=1}^{\text{BPS}}\left[\widetilde{b}_{\bar{k}_{\bar{v}}}L_a(b_{\bar{k}_{\bar{v}}}) - \overline{C}_{a,\bar{k}_{\bar{v}}}^{SD}\right]\right\} = d_{v+1} + \Delta_v, \quad (25)$$

and the PED increment Δ_v is given by:

$$\Delta_v = \frac{\left|\left(\widetilde{Y}_v - \sum_{t=v+1}^{N_T} l_{t,v}s_t\right) - l_{v,v}s_v\right|^2}{N_0} - \sum_{\bar{k}_v=1}^{\text{BPS}}\left[\widetilde{b}_{\bar{k}_v}L_a(b_{\bar{k}_v}) - \overline{C}_{a,\bar{k}_v}^{SD}\right]. \quad (26)$$

It can be seen in (26) that the soft-decision SD's PED increment Δ_v includes two terms, where the first term is revised from the hard-decision SD's PED increment of (23), while the second term is the *a priori* information obtained from the channel decoder. As a result, the soft-decision SD cannot directly utilize the decision variable $\widetilde{z}_v^{SD} = (\widetilde{Y}_v - \sum_{t=v+1}^{N_T} l_{t,v}s_t)(l_{v,v})^*$ in order to find the closest MPSK phase, which is used by the hard-decision SD, as exemplified

in Fig. 16. This is because the second term of *a priori* information in (26) is not included in $\widetilde{z}_v^{SD}$, and in fact, the channel decoder is unaware of which MPSK constellation diagram is considered. As a result, the soft-decision SD invoking the Fincke-Pohst strategy of [117] enumerates all MPSK candidates that lie inside the search bound, while the soft-decision SD invoking the Schnorr-Euchner strategy in [119] has to evaluate and compare all MPSK candidates according to (26) in order to establish their specific priorities for the SD's search order. Against this background, a reduced-complexity soft-decision SD is introduced in [197], where the *a priori* LLRs in the second term of (26) are assigned to the appropriate parts of $\widetilde{z}_v^{SD}$ obtained from the first term in (26), so that the best MPSK candidate associated with the lowest Δ_v may be obtained by visiting a reduced subset of constellation points, and then the remaining MPSK constellation points may be visited in a zigzag fashion that is similar to the hard-decision SD exemplified by Fig. 16.

In summary, with the aid of soft-decision SD, the minimum ED d_{MAP} as well as the optimum V-BLAST candidate $\hat{\mathbf{S}}$ may be obtained. The optimum candidate $\hat{\mathbf{S}}$ may further be translated into hard-bit decisions $\{\hat{b}_k^{MAP}\}_{k=1}^{N_T\text{BPS}}$. In order to produce the soft-bit decisions according to the Max-Log-MAP algorithm of (15), the SD is invoked again for producing the second ED $\bar{d}_{MAP}$, where the search space is halved by fixing the k-th bit b_k to the flipped MAP decision as $b_k = \bar{b}_k^{MAP}$. In summary, the Max-Log-MAP algorithm of (15) may be completed as:

$$L_p(b_k) = \begin{cases} -d_{MAP} + \bar{d}_{MAP}, & \text{if } b_k^{MAP} = 1 \\ -\bar{d}_{MAP} + d_{MAP}, & \text{if } b_k^{MAP} = 0. \end{cases} \tag{27}$$

In this way, the SD has to be invoked $(N_T\text{BPS}+1)/(N_T\text{BPS})$ times for producing a single soft-bit output, which is often referred to as the Repeated Tree Search (RTS) [191], [197], [198]. Alternatively, it's recently proposed in [118] and [119] that the Single Tree Search (STS) [198] may opt to invoke the SD only once for obtaining all the EDs of d_{MAP} and $\bar{d}_{MAP}$, which may induce a potential performance loss. More explicitly, if the hypothesis bit-mapping arrangement for d_{MAP} is updated and changed, all the counter-hypothesis bit-mapping arrangements for $\bar{d}_{MAP}$ have to be changed accordingly. As a result, the previously dismissed candidates that obey the new bit-mapping cannot be taken into account again. As a remedy, the sub-optimal detector has to invoke the LLR correction method [119] for correcting the LLR results. In fact, the STS's motivation of visiting a node at most once can still be accomplished by the RTS, where the previously visited nodes may be labelled so that the repeated calculations may be avoided by reading the previously evaluated PED metrics.

5) HARD-DECISION AND SOFT-DECISION SD AIDED V-BLAST EMPLOYING SQUARE QAM

The PED increments of (23) and (26) may be utilized by the hard-decision and soft-decision SD aided V-BLAST, respectively, when an arbitary PSK/QAM constellation is employed. Nonetheless, it was suggested in [19], [115], and [199] that the real part and the imaginary part of the Square QAM constellation should be separately visited by the SD. To this end, the received signal model of (1) has to be decoupled as:

$$\overline{\mathbf{Y}} = \overline{\mathbf{S}}\overline{\mathbf{H}} + \overline{\mathbf{V}}, \tag{28}$$

where the $2N_R$-element received signal row-vector $\overline{\mathbf{Y}} = [\Re(\mathbf{Y}), \Im(\mathbf{Y})]$, the $2N_T$-element transmit signal row-vector $\overline{\mathbf{S}} = [\Re(\mathbf{S}), \Im(\mathbf{S})]$, the $(2N_T \times 2N_R)$-element fading matrix $\overline{\mathbf{H}} = \begin{bmatrix} \Re(\mathbf{H}) & \Im(\mathbf{H}) \\ -\Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix}$ and the $2N_R$-element AWGN row-vector $\overline{\mathbf{V}} = [\Re(\mathbf{V}), \Im(\mathbf{V})]$ are all real-valued. The V-BLAST receiver may now apply QR decomposition to $\overline{\mathbf{H}}^T$ as expressed in (17), so that the received signal matrix may be decomposed as:

$$\overline{\mathbf{H}} = \mathbf{L}\mathbf{Q}^T, \tag{29}$$

where $\mathbf{L}$ is a $(2N_T \times 2N_T)$-element real-valued lower triangular matrix, while the $(2N_R \times 2N_T)$-element real-valued matrix $\mathbf{Q}$ has orthogonal columns as $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_{2N_T}$. Similar to (17), $N_R \geq N_T$ is also assumed for V-BLAST employing Square QAM. As a result, the received signal model of (28) may be rewritten as:

$$\overline{\mathbf{Y}}\mathbf{Q} = \overline{\mathbf{S}}\mathbf{L} + \overline{\mathbf{V}}\mathbf{Q}, \tag{30}$$

where $\mathbf{Q}$ obtained from (29) does not change the statistics of the AWGN matrix $\overline{\mathbf{V}}$. Therefore, the ED of the ML detection of E(11) may now be expressed as:

$$\left\|\widetilde{\overline{\mathbf{Y}}} - \overline{\mathbf{S}}\mathbf{L}\right\|^2 = \sum_{v=1}^{2N_T}\left(\widetilde{\overline{Y}}_v - \sum_{t=v}^{2N_T} l_{t,v}\bar{s}_t\right)^2, \tag{31}$$

where $\widetilde{\overline{\mathbf{Y}}} = \overline{\mathbf{Y}}\mathbf{Q}$ is defined in (30). By exploiting the structure of the lower triangular matrix $\mathbf{L}$, the PED utilized by the SD may be defined according to the ED of (31) as:

$$d_v = \sum_{\bar{v}=v}^{2N_T}\left(\widetilde{\overline{Y}}_{\bar{v}} - \sum_{t=\bar{v}}^{2N_T} l_{t,\bar{v}}\bar{s}_t\right)^2 = d_{v+1} + \Delta_v, \tag{32}$$

where the PED increment Δ_v is given by:

$$\Delta_v = \left(\widetilde{\overline{Y}}_v - \sum_{t=v}^{2N_T} l_{t,v}\bar{s}_t\right)^2. \tag{33}$$

Similar to (26), the PED increment Δ_v for the soft-decision SD aided V-BLAST employing Square MQAM is given by:

$$\Delta_v = \frac{\left(\widetilde{\overline{Y}}_v - \sum_{t=v}^{2N_T} l_{t,v}\bar{s}_t\right)^2}{N_0} - \sum_{\bar{k}_v=1}^{\text{BPS}/2}\left[\widetilde{b}_{\bar{k}_v} L_a(b_{\bar{k}_v}) - \overline{C}_{a,\bar{k}_v}^{SD}\right]. \tag{34}$$

In summary, with the aid of SD, the complexity of the ML/MAP aided V-BLAST detection may be substantially reduced. More explicitly, when the breadth-first (K-Best)

strategy is invoked, the SD aided V-BLAST detection complexity may be lower-bounded by $O(N_T)$ and by $O(2N_T)$ for the case of MPSK and for the case of Square MQAM, respectively. Moreover, when the depth-first strategy is invoked, the SD aided V-BLAST detection complexity lower bounds are given by $O(2N_T - 1)$ and $O(4N_T - 1)$ for MPSK and Square MQAM, respectively, where only a single constellation point is visited, when the SD index is reduced from $v = N_T$ down to $v = 1$ and then increased from $v = 2$ up to $v = N_T$. However, the SD complexity lower bounds can only be approached in the high-SNR region or when provided with full *a priori* information in coded systems, where the ED differences between the candidates are large so that the optimum solution may be found without any ambiguity. It is also demonstrated in [200] that the average SD complexity is a polynomial function, which is often approximately cubic, while [201] demonstrates that the SD complexity is still exponential at low SNR region. Therefore, in the coming section, we further introduce LF aided V-BLAST receivers, which exhibit a detection complexity that may as low as single-antenna-based detection, but the sub-optimal performance is inevitable.

6) HARD-DECISION LF AIDED V-BLAST

For low-complexity V-BLAST detection, LFs may be conceived for detecting the paralleled data-streams separately, while suppressing the interference as best as possible. More explicitly, under the idealized assumption of having perfect knowledge of the CSI, the basic Matched Filter (MF) output becomes [22]:

$$\mathbf{Z}^{MF} = \mathbf{Y}\mathbf{G}^{MF} = \mathbf{S}\mathbf{H}\mathbf{H}^H + \mathbf{V}\mathbf{H}^H, \tag{35}$$

where the $(N_R \times N_T)$-element MF weight matrix in (35) is given by $\mathbf{G}^{MF} = \mathbf{H}^H$. Furthermore, the v-th element in the N_T-element decision variable row-vector $\mathbf{Z}^{MF}$ of (35) is given by $z_v^{MF} = s_v\|\mathbf{H}_{v,-}\|^2 + \sum_{\forall \bar{v} \neq v} s_{\bar{v}}\mathbf{H}_{\bar{v},-}(\mathbf{H}_{v,-})^H + \mathbf{V}(\mathbf{H}_{v,-})^H$, where the second term of $\sum_{\forall \bar{v} \neq v} s_{\bar{v}}\mathbf{H}_{\bar{v},-}(\mathbf{H}_{v,-})^H$ introduces severe interference. Without dealing with this interference term, directly demodulating the single symbol s_v by carrying out the operation $z_v^{MF}/\|\mathbf{H}_{v,-}\|^2$ results in an irreducible error floor.

In order to mitigate this problem, the Zero-Forcing (ZF) detector aims for cancelling the interference term of the $(N_R \times N_T)$-element ZF weight matrix $\mathbf{G}^{ZF} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}$, so that the ZF filter output is given by [23]–[25]:

$$\mathbf{Z}^{ZF} = \mathbf{Y}\mathbf{G}^{ZF} = \mathbf{S} + \widetilde{\mathbf{V}}, \tag{36}$$

where the N_T-element noise row-vector $\widetilde{\mathbf{V}} = \mathbf{V}\mathbf{G}^{ZF}$ has an increased noise power of $\|\mathbf{G}^{ZF}\|^2 \cdot N_0$. The ZF's problem of noise enhancement may result in a severe performance contamination, especially in the low-SNR region.

The Minimum Mean Squared Error (MMSE) filter may further reduce the noise power by minimizing the Mean Squared Error (MSE), which is defined as the Euclidean distance between the MMSE filter output and the transmitted V-BLAST vector as $\sigma_{MSE}^2 = \mathrm{E}\left(\|\mathbf{Z}^{MMSE} - \mathbf{S}\|^2\right)$ [22], [23], [25]. More explicitly, the MMSE filter output may be expressed as:

$$\mathbf{Z}^{MMSE} = \mathbf{Y}\mathbf{G}^{MMSE} = \mathbf{S}\mathbf{H}\mathbf{G}^{MMSE} + \mathbf{V}\mathbf{G}^{MMSE}, \tag{37}$$

where the $(N_R \times N_T)$-element MMSE weight matrix $\mathbf{G}^{MMSE}$ is conceived for minimizing the MSE σ_{MSE}^2, which may be extended as:

$$\sigma_{MSE}^2 = \mathrm{tr}\left[(\mathbf{G}^{MMSE})^H \mathrm{E}\left(\mathbf{Y}^H\mathbf{Y}\right)\mathbf{G}^{MMSE}\right] - 2\Re\left\{\mathrm{tr}\left[(\mathbf{G}^{MMSE})^H \mathrm{E}\left(\mathbf{Y}^H\mathbf{S}\right)\right]\right\} + 1, \tag{38}$$

where the auto-correlation matrix is given by $\mathrm{E}\left\{\mathbf{Y}^H\mathbf{Y}\right\} = \frac{1}{N_T}\mathbf{H}^H\mathbf{H} + N_0\mathbf{I}_{N_R}$, while the cross-correlation matrix is given by $\mathrm{E}\left\{\mathbf{Y}^H\mathbf{S}\right\} = \frac{1}{N_T}\mathbf{H}^H$. Therefore, the MMSE solution of $\frac{\partial \sigma_{\mathrm{MSE}}^2}{\partial \mathbf{G}^{MMSE}} = 0$ leads us to the MMSE weight matrix of:

$$\mathbf{G}^{MMSE} = \left(\mathbf{H}^H\mathbf{H} + N_0 \cdot N_T \cdot \mathbf{I}_{N_R}\right)^{-1}\mathbf{H}^H, \tag{39}$$

which do not have to be updated, when the fading channel's envelope remains near-constant. As a result, the v-th element in the MMSE filter output vector $\mathbf{Z}^{MMSE}$ of (37) may be rewritten as:

$$z_v^{MMSE} = s_v\mathbf{H}_{v,-}\mathbf{G}_{-,v}^{MMSE} + \sum_{\forall \bar{v} \neq v} s_{\bar{v}}\mathbf{H}_{\bar{v},-}\mathbf{G}_{-,v}^{MMSE} + \mathbf{V}\mathbf{G}_{-,v}^{MMSE}, \tag{40}$$

Finally, the linear MPSK/QAM demodulator may be invoked for recovering the data-carrying modulation indices as $\hat{m}_v = \mathbb{M}^{-1}(\widetilde{z}_v)$ for $v = \{1, \cdots, N_T\}$, where we have $\widetilde{z}_v = \sqrt{N_T} \cdot z_v^{MMSE} \cdot (\mathbf{H}_{v,-}\mathbf{G}_{-,v}^{MMSE})^*/|\mathbf{H}_{v,-}\mathbf{G}_{-,v}^{MMSE}|^2$ according to (40).

It was proposed in [32] and [202]–[206]that the interference cancellation techniques based on either ZF receivers or MMSE receivers may further improve the LF aided V-BLAST detection performance. Moreover, the interference nulling and cancelling proposed for Multi-User Detection (MUD) in CDMA systems [207]–[210] may be adopted by V-BLAST, since the V-BLAST scheme's multiple TAs may be considered to be equivalent to CDMA's multiple users. For example, the Successive Interference Cancelling (SIC) may opt for detecting the data streams one by one from the strongest to the weakest. When the LF makes a decision concerning a single data stream, it may be remodulated and then subtracted from the received signal so that the remaining data streams may be detected successively, while having to cope with a reduced amount of interference.

7) SOFT-DECISION LF AIDED V-BLAST

If we directly revise the hard-decision MMSE aided V-BLAST for employment in coded systems, the Log-MAP of (13), the Max-Log-MAP of (15) and the Approx-Log-MAP of (16) may be invoked, where the signal sets may be replaced by the classic MPSK/QAM of $\forall s^m \in \{s^m\}_{b_k=1}$ and $\forall s^m \in \{s^m\}_{b_k=0}$, while the symbol-by-symbol based

a posteriori probability metrics may be revised for MMSE detection as:

$$d^m = -\frac{\left|z_v^{MMSE} - s^m \mathbf{H}_{v,-}\mathbf{G}_{-,v}^{MMSE}/\sqrt{N_T}\right|^2}{N_0\|\mathbf{G}_{-,v}^{MMSE}\|^2} + \sum_{\bar{k}_v=1}^{\mathrm{BPS}} \widetilde{b}_{\bar{k}_v} L_a(b_{\bar{k}_v}). \quad (41)$$

The decision variable $\{z_v^{MMSE}\}_{v=1}^{N_T}$ is given by (40), while the MMSE filters taps $\{\mathbf{G}_{-,v}^{MMSE}\}_{v=1}^{N_T}$ are formulated in (39). As a result, for producing $\mathrm{BPS} = \log_2 M$ number of *a posteriori* LLRs either by the Log-MAP of (13), or by the Max-Log-MAP of (15) or alternatively by the Approx-Log-MAP of (16), a total of M *a posteriori* probability metrics of (41) have to be evaluated and compared according to the MPSK/QAM constellation points.

However, this simple mechanism does not deliver the exact MMSE solution [22], [92], [93] for coded V-BLAST systems, because the *a priori* knowledge of the V-BLAST symbols is not taken into account by the MSE objective function of (38). In order to improve the MMSE solution, first of all, the output signal produced by the MMSE filter may be extended as:

$$\bar{z}_v^{MMSE} = \mathbf{Y}\overline{\mathbf{G}}_v^{MMSE} = s_v\mathbf{H}_{v,-}\overline{\mathbf{G}}_v^{MMSE} + \mathbf{S}^{\bar{v}}\mathbf{H}^{\bar{v}}\overline{\mathbf{G}}_v^{MMSE} + \mathbf{V}\overline{\mathbf{G}}_v^{MMSE}, \quad (42)$$

where the N_R-element MMSE filter taps column-vector $\overline{\mathbf{G}}_v^{MMSE}$ aims for minimizing the interference term of $u_v = \mathbf{S}^{\bar{v}}\mathbf{H}^{\bar{v}}\overline{\mathbf{G}}_v^{MMSE}$ without increasing the noise power. In the presence of *a priori* LLRs, the residual interference term after MMSE filtering may be further mitigated by the following operations [22], [92], [93], [211]:

$$\check{z}_v^{MMSE} = \bar{z}_v^{MMSE} - \hat{u}_v = \bar{z}_v^{MMSE} - \hat{\mathbf{S}}^{\bar{v}}\mathbf{H}^{\bar{v}}\overline{\mathbf{G}}_v^{MMSE}, \quad (43)$$

where $\hat{\mathbf{S}}^{\bar{v}} = \mathrm{E}(\mathbf{S}^{\bar{v}})$ referred to as the estimate of the interference vector. The t-th $(1 \le t \le N_T - 1)$ element $\hat{s}_t = \mathrm{E}(s_t)$ in $\hat{\mathbf{S}}^{\bar{v}}$ may be obtained from the *a priori* probabilities as [22], [87], [90]:

$$\hat{s}_t = \frac{1}{\sqrt{N_T}}\sum_{m=0}^{M-1} s^m p(s_t = s^m) = \frac{1}{\sqrt{N_T}}\sum_{m=0}^{M-1} s^m \frac{\exp\left[\sum_{\bar{k}=1}^{\mathrm{BPS}} \widetilde{b}_{\bar{k}} L_a(b_{\bar{k}})\right]}{\prod_{\bar{k}=1}^{\mathrm{BPS}}\left\{1+\exp\left[L_a(b_{\bar{k}})\right]\right\}}. \quad (44)$$

It was demonstrated in [22] and [90] that $\frac{\partial \mathrm{E}(|\check{z}_v^{MMSE} - s_v|^2)}{\partial \overline{\mathbf{G}}_v^{MMSE}} = 0$ results in the MMSE weight matrix shown in (45), as shown at the bottom of this page, where the $[(N_T-1)\times(N_T-1)]$-element matrix $\mathbf{R}_{|s|}^{\bar{v}}$ refers to the estimate of the interference powers as $\mathbf{R}_{|s|}^{\bar{v}} = \mathrm{diag}[\mathrm{E}(|s_1|^2), \cdots, \mathrm{E}(|s_{v-1}|^2), \mathrm{E}(|s_{v+1}|^2), \cdots, \mathrm{E}(|s_{N_T}|^2)]$. The estimate of a specific symbol's power $\mathrm{E}(|s_t|^2)$ is given by replacing the constellation point s^m in (44)by its power $|s^m|^2$. As a special case, we have $\mathrm{E}(|s_t|^2) = \frac{1}{N_T}$ for MPSK constellations. It is worth noting that when there is no *a priori* information as represented by $I_A = 0$, we have $\mathbf{R}_{|s|}^{\bar{v}} = \frac{1}{N_T}\mathbf{I}_{N_T-1}$ and $\hat{\mathbf{S}}^{\bar{v}} = \mathbf{0}_{1\times(N_T-1)}$, and the MMSE filter taps of (45) become $\overline{\mathbf{G}}_v^{MMSE} = \left(\mathbf{H}^H\mathbf{H} + N_T N_0 \mathbf{I}_{N_R}\right)^{-1}\mathbf{H}_{v,-}^H$, which is exactly the same as the hard-decision MMSE solution of (39). By contrast, when perfect *a priori* information of $I_A = 1$ is available, we have $\mathbf{R}_{|s|}^{\bar{v}} = \mathrm{diag}\left((\hat{\mathbf{S}}^{\bar{v}})^H\hat{\mathbf{S}}^{\bar{v}}\right)$, which results in the MMSE filter taps of $\overline{\mathbf{G}}_v^{MMSE} = \left(\mathbf{H}_{v,-}^H\mathbf{H}_{v,-} + N_T N_0\mathbf{I}_{N_R}\right)^{-1}\mathbf{H}_{v,-}^H$. We note that the case of $I_A = 1$ leads to the perfect estimation of the interference term $\hat{u}_v = \mathbf{S}^{\bar{v}}\mathbf{H}^{\bar{v}}\overline{\mathbf{G}}_v^{MMSE}$ of (43), which implies that the optimum MIMO detection capability may be achieved by the MMSE detector at $I_A = 1$. As a result, the *a posteriori* probability metric of (41) may be revised for the exact MMSE solution as:

$$d^m = -\frac{\left|\check{z}_v^{MMSE} - \frac{1}{\sqrt{N_T}}s^m\mathbf{H}_{v,-}\overline{\mathbf{G}}_v^{MMSE}\right|^2}{N_0^{MMSE}} + \sum_{\bar{k}=1}^{\mathrm{BPS}}\widetilde{b}_{\bar{k}}L_a(b_{\bar{k}}), \quad (46)$$

where the interference-decontaminated MMSE filter output of (43) may be rewritten as $\check{z}_v^{MMSE} = s_v\mathbf{H}_{v,-}\overline{\mathbf{G}}_v^{MMSE} + (\mathbf{S}^{\bar{v}} - \hat{\mathbf{S}}^{\bar{v}})\mathbf{H}^{\bar{v}}\overline{\mathbf{G}}_v^{MMSE} + \mathbf{V}\overline{\mathbf{G}}_v^{MMSE}$, while the residual interference term $(\mathbf{S}^{\bar{v}} - \hat{\mathbf{S}}^{\bar{v}})\mathbf{H}^{\bar{v}}\overline{\mathbf{G}}_v^{MMSE}$ and the AWGN term $\mathbf{V}\overline{\mathbf{G}}_v^{MMSE}$ have a joint variance of $N_0^{MMSE} = \mathrm{E}[|(\mathbf{S}^{\bar{v}} - \hat{\mathbf{S}}^{\bar{v}})\mathbf{H}^{\bar{v}}\overline{\mathbf{G}}_v^{MMSE} + \mathbf{V}\overline{\mathbf{G}}_v^{MMSE}|^2] = \frac{1}{N_T}(\overline{\mathbf{G}}_v^{MMSE})^H(\mathbf{H}_{v,-})^H - \frac{1}{N_T}\left|(\overline{\mathbf{G}}_v^{MMSE})^H(\mathbf{H}_{v,-})^H\right|^2$.

The calculation of the MMSE filter taps $\{\overline{\mathbf{G}}_v^{MMSE}\}_{v=1}^{N_T}$ of (45) specifically calculated for detecting all the V-BLAST symbols $\{s_v\}_{v=1}^{N_T}$ requires us to perform an matrix-element inversion for a total number of N_T times, which may be excessive for practical implementations. Simplified matrix inversion techniques were proposed in [87] and [212], but the matrix inversion still had to be carried out N_T times. Moreover, it was proposed in [211] that both $\mathbf{R}_{|s|}^{\bar{v}}$ and $\hat{\mathbf{S}}^{\bar{v}}$ may be estimated by averaging over all samples of a detection frame, so that the matrix inversion did not have to be updated for detecting each transmitted V-BLAST symbol. However, this method imposes a substantial performance loss. Against this background, a better choice is proposed in [93], where the matrix inversion only has to be performed once for detecting all the N_T transmitted V-BLAST symbols. More explicitly,

$$\overline{\mathbf{G}}_v^{MMSE} = \left\{\mathbf{H}_{v,-}^H\mathbf{H}_{v,-} + N_T(\mathbf{H}^{\bar{v}})^H\left[\mathbf{R}_{|s|}^{\bar{v}} - \mathrm{diag}\left((\hat{\mathbf{S}}^{\bar{v}})^H\hat{\mathbf{S}}^{\bar{v}}\right)\right]\mathbf{H}^{\bar{v}} + N_T N_0\mathbf{I}_{N_R}\right\}^{-1}\mathbf{H}_{v,-}^H. \quad (45)$$

the *a posteriori* probability metric of (46) may simplified as:

$$d^m = -\frac{\left|\widetilde{z}_v^{MMSE} - s^m\right|^2}{\widetilde{N}_0^{MMSE}} + \sum_{\bar{k}=1}^{\text{BPS}} \widetilde{b}_{\bar{k}} L_a(b_{\bar{k}}), \tag{47}$$

where the decision variable is given by $\widetilde{z}_v^{MMSE} = \left(\mathbf{Y}\widetilde{\mathbf{G}}_v^{MMSE} - \hat{\mathbf{S}}^{\bar{v}}\mathbf{H}^{\bar{v}}\widetilde{\mathbf{G}}_v^{MMSE}\right)/\widetilde{h}_v^{MMSE}$, while the equivalent fading factor $\widetilde{h}_v^{MMSE} = \frac{1}{\sqrt{N_T}}\mathbf{H}_{v,-}\widetilde{\mathbf{G}}_v^{MMSE}$ is supposed to be a real number. Furthermore, the equivalent MMSE filter taps vector $\widetilde{\mathbf{G}}_v^{MMSE}$ is given by (48), as shown at the bottom of this page, where both $\hat{\mathbf{S}} = \mathrm{E}(\mathbf{S})$ and $\mathbf{R}_{|s|} = \mathrm{diag}\left[\mathrm{E}(|s_1|^2), \cdots, \mathrm{E}(|s_{N_T}|^2)\right]$ only have to be evaluated once for detecting a V-BLAST received signal row-vector $\mathbf{Y}$ of (1). Moreover, the equivalent noise power $\widetilde{N}_0^{MMSE}$ in (47) is given by $\widetilde{N}_0^{MMSE} = \frac{1}{\sqrt{N_T}\cdot\widetilde{h}_v^{MMSE}} + N_T\left[|\hat{s}_v|^2 - \mathrm{E}(|s_v|^2)\right]$.

8) REDUCED-COMPLEXITY SOFT-DECISION PSK/QAM DETECTION

The linear soft-decision LF aided V-BLAST effectively separates the superimposed parallel data streams, so that the classic soft-decision PSK/QAM detectors may be invoked, where the detection complexity is on the order of $O(M)$ instead of the MAP aided V-BLAST's order of $O(M^{N_T})$. More specifically, for the soft-decision MMSE, the Log-MAP of (13), the Max-Log-MAP of (15) and the Approx-Log-MAP of (16) may be employed, where the signal sets may be replaced by the classic MPSK/QAM of $\forall s^m \in \{s^m\}_{b_k=1}$ and $\forall s^m \in \{s^m\}_{b_k=0}$, while the *a posteriori* probability metric $\{d^m\}$ of (47) has to be evaluated M times according to the MPSK/QAM constellation points. It is worth noting that the detection complexity order of soft-decision Sqaure MQAM is given by $O(\sqrt{M})$, where the real and imaginary parts of the Square QAM constellation points are visited separately.

Moreover, the bit-metric generation methods introduced in [213]–[215] may further reduce the complexity order to $O(\log_2 M)$, where the approximated LLR values are efficiently evaluated on a bit-by-bit basis. However, these early contributions on bit-metric generation did not consider the *a priori* LLRs. This is because the detection of the Gray-labelled low-order PSK/QAM schemes (e.g BPSK/QPSK and Square 16QAM) generally produces near-horizontal curves in the EXIT chart [99], which means that exchanging information between the soft PSK/QAM detector and the channel decoder may have a negligible benefit.

However, high-order MQAM schemes are routinely utilized in recent commercialized systems. For example, Square 64QAM and Square 256QAM have been included in the ITU-R IMT Advanced 4G standards [217] and in IEEE 802.11ac [218], respectively. As the number of modulation levels M increases, the soft MQAM detectors become capable of producing an improved iteration gain. Against this background, a reduced-complexity design for soft-decision MPSK/QAM detection was proposed in [216], which may be briefly summarized as:

Algorithm 1: Design guidelines for reduced-complexity soft-decision MPSK/QAM detection

1) First of all, each *a priori* LLRs are related to a reduced-size fraction of the channel's output signal constellations.
2) As a result, by further exploring the symmetry provided by Gray-labelled constellations, a reduced subset of positive PAM magnitudes and a reduced subset of constellation points found in the first quadrant are visited by the soft-decision Square QAM detector and by the soft-decision general PSK/QAM detector, respectively, in order to obtain the maximum *a posteriori* probability metric that is required by the Max-Log-MAP of (15).
3) Finally, both the Max-Log-MAP of (15) and the Approx-Log-MAP of (16) may be completed by comparing the terms evaluated by the previous steps.

Let us consider the 8PSK scheme of Fig. 18(b) as an example. First of all, the soft-decision MMSE aided V-BLAST's *a posteriori* probability metric may be extended as:

$$d^m = \frac{\Re(\bar{z}_v)\Re(\bar{s}^m) + \Im(\bar{z}_v)\Im(\bar{s}^m)}{\overline{N}_0} - \frac{|\bar{z}_v|^2}{\widetilde{N}_0} - \frac{|\bar{s}^m|^2}{\widetilde{N}_0} + \sum_{\bar{k}=1}^{\text{BPS}} \widetilde{b}_{\bar{k}}^m L_a(b_{\bar{k}}), \tag{49}$$

where $\widetilde{z}_v^{MMSE}$ and s^m in (47) are replaced by $\bar{z}_v = \widetilde{z}_v^{MMSE}\exp(j\pi/M_P)$ and $\bar{s}^m = s^m\exp(j\pi/M_P)$, respectively, according to the rotated 8PSK constellation of Fig. 18(b), while the superscript of $\widetilde{N}_0^{MMSE}$ is deleted and we have $\overline{N}_0 = \widetilde{N}_0/2$. Moreover, the constant of $\left(-\frac{|\widetilde{z}_n|^2}{\widetilde{N}_0}\right)$ seen in (49) may be ignored. As a result, the four probability metrics $\{d^m\}_{m\in\{0,2,4,6\}}$ for the four constellation points $\{s^m = \pm\cos(\frac{\pi}{8}) \pm \sin(\frac{\pi}{8})\}$ are extended in (56), where the real and imaginary parts of $\bar{z}_v$ are respectively related to the corresponding *a priori* LLRs $L_a(b_2)$ and $L_a(b_1)$ by:

$$t_{\text{Re}}^{G0} = \frac{\cos(\frac{\pi}{8})\Re(\bar{z}_n)}{\overline{N}_0} - \frac{L_a(b_2)}{2},$$
$$t_{\text{Im}}^{G0} = \frac{\sin(\frac{\pi}{8})\Im(\bar{z}_n)}{\overline{N}_0} - \frac{L_a(b_1)}{2}, \tag{50}$$

$$\widetilde{\mathbf{G}}_v^{MMSE} = \left\{N_T\mathbf{H}^H\left[\mathbf{R}_{|s|} - \mathrm{diag}\left(\hat{\mathbf{S}}^H\hat{\mathbf{S}}\right)\right]\mathbf{H} + N_T N_0 \mathbf{I}_{N_R}\right\}^{-1}\mathbf{H}_{v,-}^H. \tag{48}$$

while the constant $C_{8\text{PSK}}$ is given by $[C_{8\text{PSK}} = -\frac{1}{\widetilde{N}_0} + \frac{L_a(b_1)+L_a(b_2)}{2}]$. It can be seen that the four metrics formulated in (56), as shown at the bottom of this page all contain three parts, i.e. they are $\pm t_{\text{Re}}^{G0}$, $\pm t_{\text{Im}}^{G0}$ and $C_{8\text{PSK}}$. As a result, the maximum metric over the four candidates in (56) is given by a simple estimation:

$$d^{G0} = \max_{m=\{0,2,4,6\}} d^m = |t_{\text{Re}}^{G0}| + |t_{\text{Im}}^{G0}| + C_{8\text{PSK}}. \quad (51)$$

Similarly, the maximum of the four probability metrics $\{d^m\}_{m\in\{1,3,5,7\}}$ related to the four constellation points $\{s^m = \pm\sin(\frac{\pi}{8}) \pm \cos(\frac{\pi}{8})\}$ may also be directly expressed as:

$$d^{G1} = \max_{m=\{1,3,5,7\}} d^m = |t_{\text{Re}}^{G1}| + |t_{\text{Im}}^{G1}| + L_a(b_3) + C_{8\text{PSK}}, \quad (52)$$

where the real and imaginary test-variables are given by:

$$t_{\text{Re}}^{G1} = \frac{\sin(\frac{\pi}{8})\Re(\bar{z}_n)}{\overline{N}_0} - \frac{L_a(b_2)}{2},$$
$$t_{\text{Im}}^{G1} = \frac{\cos(\frac{\pi}{8})\Im(\bar{z}_n)}{\overline{N}_0} - \frac{L_a(b_1)}{2}. \quad (53)$$

Therefore, the maximum *a posteriori* probability metric generated by the Max-Log-MAP algorithm is given by:

$$d^{\max} = \max_{g=\{0,1\}} \left(d^{Gg}\right) = \max \left\{ \begin{array}{l} |t_{\text{Re}}^{G0}| + |t_{\text{Im}}^{G0}| + C_{8\text{PSK}} \\ |t_{\text{Re}}^{G1}| + |t_{\text{Im}}^{G1}| + L_a(b_3) + C_{8\text{PSK}} \end{array} \right\}. \quad (54)$$

Therefore, instead of evaluating and comparing (49) for a total of $M = 8$ times, (54) only has to evaluate and compare a reduced number of $(M/4 = 2)$ candidates in order to obtain $d^{\max}$. In other words, $d^{\max}$ of (54) is obtained without visiting all the eight 8PSK constellation points. In fact, only the two constellation points in the first quadrant are of interest, as demonstrated by Fig. 18(b). In summary, the Max-Log-MAP invoked by the reduced-complexity soft-decision 8PSK detection may be completed as:

$$L_p(b_1) = d_{\max}^{b_1=1} - d_{\max}^{b_1=0},$$
$$L_p(b_2) = d_{\max}^{b_2=1} - d_{\max}^{b_2=0},$$
$$L_p(b_3) = |t_{\text{Re}2}| + |t_{\text{Im}2}| + L_a(b_3) - |t_{\text{Re}1}| - |t_{\text{Im}1}|, \quad (55)$$

where $d_{\max}^{b_1=1}$ and $d_{\max}^{b_1=0}$ may be obtained by replacing $\{|t_{\text{Im}}^{Gg}|\}_{g=0}^{1}$ in (54) by $\{-t_{\text{Im}}^{Gg}\}_{g=0}^{1}$ and $\{t_{\text{Im}}^{Gg}\}_{i=1}^{2}$, respectively, while $d_{\max}^{b_2=1}$ and $d_{\max}^{b_2=0}$ are obtained by replacing $\{|t_{\text{Re}}^{Gg}|\}_{i=1}^{2}$ in (54) by $\{-t_{\text{Re}}^{Gg}\}_{g=0}^{1}$ and $\{t_{\text{Re}}^{Gg}\}_{g=0}^{1}$, respectively. The constant $C_{8\text{PSK}}$ in (54) may be omitted.

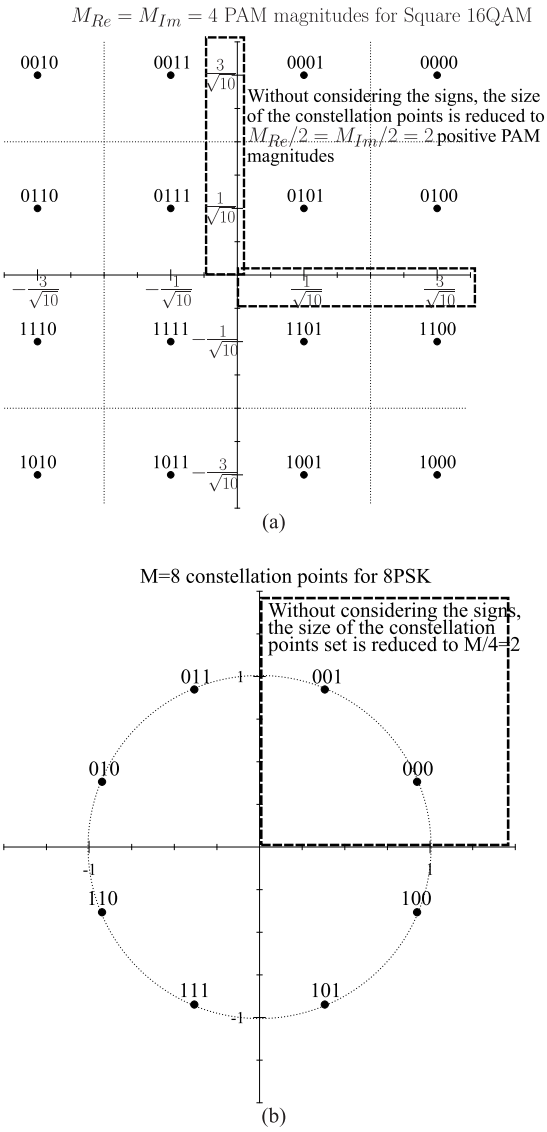


FIGURE 18. Examples of the reduced-complexity soft-decision *M*PSK/QAM detector's constellation diagrams in [216], where all the detected *M*PSK ($M \geq 4$) constellation diagrams and the detected Star *M*QAM constellation diagrams are rotated anti-clockwise by a phase of π/M and π/M_P, respectively, so that there are exactly $M/4$ constellation points in each quadrant. (a) Square 16QAM. (b) Rotated 8PSK.

$$d^0 = \frac{\cos(\frac{\pi}{8})\Re(\bar{z}_v)}{\overline{N}_0} + \frac{\sin(\frac{\pi}{8})\Im(\bar{z}_v)}{\overline{N}_0} - \frac{1}{\widetilde{N}_0} = t_{\text{Re}}^{G0} + t_{\text{Im}}^{G0} + C_{8\text{PSK}},$$
$$d^2 = -\frac{\cos(\frac{\pi}{8})\Re(\bar{z}_v)}{\overline{N}_0} + \frac{\sin(\frac{\pi}{8})\Im(\bar{z}_v)}{\overline{N}_0} - \frac{1}{\widetilde{N}_0} + L_a(b_2) = -t_{\text{Re}}^{G0} + t_{\text{Im}}^{G0} + C_{8\text{PSK}},$$
$$d^4 = \frac{\cos(\frac{\pi}{8})\Re(\bar{z}_v)}{\overline{N}_0} - \frac{\sin(\frac{\pi}{8})\Im(\bar{z}_v)}{\overline{N}_0} - \frac{1}{\widetilde{N}_0} + L_a(b_1) = t_{\text{Re}}^{G0} - t_{\text{Im}}^{G0} + C_{8\text{PSK}},$$
$$d^6 = -\frac{\cos(\frac{\pi}{8})\Re(\bar{z}_v)}{\overline{N}_0} - \frac{\sin(\frac{\pi}{8})\Im(\bar{z}_v)}{\overline{N}_0} - \frac{1}{\widetilde{N}_0} + L_a(b_1) + L_a(b_2) = -t_{\text{Re}}^{G0} - t_{\text{Im}}^{G0} + C_{8\text{PSK}}, \quad (56)$$

The complexities of the conventional soft-decision MPSK/QAM detection algorithms and those of the reduced-complexity detection algorithms of [216] are quantified in terms of the total number of real-valued multiplications required for producing a single soft-bit output in Fig 19, where the Complexity-Reduction Ratio (CRR) is defined as the complexity difference divided by the complexity of the conventional detector. It can be seen in Figs. 19(a) and 19(b) that the CRRs achieved by the reduced-complexity detection algorithms of [216] approach their upper bound of 50% and 75% for Square QAM and for general PSK/QAM respectively as M increases, because 50% of the PAM magnitudes and 75% of the PSK/QAM constellation points have been avoided by the reduced-complexity design. The complexity reduction seen in both Fig. 19(a) and Fig. 19(b) is substantial, especially, when the soft MPSK/QAM detector is invoked several times in the aforementioned turbo detection applications.

9) EXIT CHARTS AND LLR ACCURACY

As introduced in Sec. I-A, one of the major design challenges is to predict and compare their E_b/N_0 convergence thresholds in order to choose the most appropriate channel coding and modulation parameters. Motivated by this challenge, researchers have focussed their attentions on characterizing the convergence behavior of turbo detection [6], [7], [98], [100], [101], [219]. More explicitly, let us consider the classic SCC of Fig. 5 as an example. The inner decoder in Fig. 5 produces *a posteriori* LLRs L_p based on both the channel's output signal and the *a priori* LLRs L_a obtained from the outer decoder. The resultant extrinsic LLRs $L_e = L_p - L_a$ gleaned from the inner decoder are then de-interleaved and fed to the outer decoder as *a priori* LLRs L_a, so that the outer decoder may further produce L_p based on the L_a gleaned from the inner decoder. As the turbo detection continues, the extrinsic LLRs $L_e = L_p - L_a$ of the outer decoder are further interleaved and then fed to the inner decoder as L_a. Therefore, it can be readily seen that the prediction of the SNR and the number of iterations required for decoding convergence is important for turbo detection, so that no futile complexity wastage is imposed.

Against this background, as exemplified by Fig. 7, the EXIT charts [100], [101], [219] may effectively visualize the flow of extrinsic information between the turbo detector components. More explicitly, the transfer characteristics of a decoder/demapper may be formulated as [101], [105]:

$$I_E = T(I_A), \tag{57}$$

where the *a priori* information $I_A = I(b; L_a)$ and the extrinsic information $I_E = I(b; L_e)$ are the input and output of the transfer function T, respectively. In order to virtualize the transfer function T as seen in Fig. 7, the first step is to generate a group of *a priori* LLRs L_a according to I_A. Then $I_E = I(b; L_e)$ may be evaluated based on the extrinsic LLRs L_e, which are obtained by feeding L_a to the decoder/demapper. Recent tutorials on EXIT charts may be found in [220] and [221]. In this treatise, we further offer a brief summary of the EXIT chart technique and provide insights into its practical aspects, such as its area property and LLR accuracy examination.

First of all, the procedures of evaluating the transfer function T of (57) are summarized as follows:

Evaluation of the transfer function T of (57)

1) For a specific *a priori* mutual information I_A, a group of *a priori* LLRs L_a may be generated as Gaussian-distributed random variables as [101]:

$$L_a = \mu_A \cdot x + v \tag{58}$$

where v is a Gaussian random variable having a zero mean and a variance of σ_A^2, while we have $\mu_A = \frac{\sigma_A^2}{2}$. Moreover, $x \in \{+1, -1\}$ in (58) is equivalent to source data bit $b \in \{1, 0\}$. Furthermore, the PDF of the *a priori* LLRs L_a generated by (58) is given by $p(L_a|x) = \frac{1}{\sqrt{2\pi}\sigma_A} \exp\left[-\frac{(L_a - \frac{\sigma_A^2}{2}x)^2}{2\sigma_A^2}\right]$, which satisfies both the symmetry condition of $p(L_a|x) = p(-L_a| - x)$ and the consistency condition of $p(L_a|x = +1) = p(L_a|x = -1)e^{L_a}$. As a result, the relationship between I_A and σ_A for generating L_a of (58) may be formulated as [101]:

$$\begin{aligned} I_A &= J(\sigma_A) \\ &= 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_A} \exp\left[-\frac{(L_a - \frac{\sigma_A^2}{2})^2}{2\sigma_A^2}\right] \\ &\quad \cdot \log_2(1 + e^{-L_a}) dL_a, \end{aligned} \tag{59}$$

where we have the input range of $\sigma_A \geq 0$ and output range of $0 \leq I_A \leq 1$. Observe from Fig. 20 that the output I_A of function $J(\cdot)$ increases monotonically with respect to its input σ_A, hence with given I_A, the corresponding σ_A may be unambiguously obtained by the inverse function of (59) as:

$$\sigma_A = J^{-1}(I_A). \tag{60}$$

It is worth noting that althougth the function $J(\cdot)$ and its inverse function $J^{-1}(\cdot)$ cannot be expressed in closed form [101], it is demonstrated in [222] that they can be approxmiated with negligible error as $J(\sigma_A) \approx \left(1 - 2^{-H_1 \sigma_A^{2H_2}}\right)^{H_3}$ and $J^{-1}(I_A) \approx \left[-\frac{1}{H_1}\log_2(1 - I_A^{1/H_3})\right]^{1/(2H_2)}$, respectively, where the parameters of $H_1 = 0.3073$, $H_2 = 0.8935$ and $H_3 = 1.1064$ were obtained by minimizing the MSE between the functions $J(\cdot)$ of (59) and $J^{-1}(\cdot)$ of (60) and their approximations.

2) Upon feeding the generated soft-valued inputs L_a to the tested component decoder/demapper in the concatenated code, a group of extrinsic LLR outputs L_e may be obtained.
3) Finally, the extrinsic mutual information output of (57) may be computed as:

$$I_E = \frac{1}{2}\sum_{b=1,0}\int_{-\infty}^{\infty} p(L_e|b) \cdot \log_2 \frac{2p(L_e|b)}{p(L_e|b=1)+p(L_e|b=0)} dL_e, \tag{61}$$

where the PDFs $p(L_e|b=1)$ and $p(L_e|b=0)$ may be obtained by evaluating the histograms of L_e [223] with respect to the source data bit being $b=1$ and $b=0$. In order to avoid the histogram evaluation, the following alternative averaging method was proposed in [99]:

$$I_E \approx 1 - \frac{1}{N_c}\sum_{n=1}^{N_c}\left[\log_2(1+e^{-x[n]\cdot L_e[n]})\right], \tag{62}$$

where the PDF $p(L_e|b)$ of the extrinsic LLRs is assumed to be symmetric $p(L_e|x) = p(-L_e|-x)$ and consistent $p(L_e|x=+1) = p(L_e|x=-1)e^{L_e}$. It was discussed in [7], [99], and [224] that the assumption of PDF symmetry may be granted, as long as the input PDFs including the PDFs of both *a priori* LLRs and of the channel's output signal are symmetric. Moreover, in order to further avoid getting access to the source bits, an efficient computation of (61) and (62) was further proposed in [102]–[104] as:

$$I_E \approx \frac{1}{N_c}\sum_{n=1}^{N_c}\left[\frac{e^{L_e[n]}}{1+e^{L_e[n]}}\log_2\left(\frac{2e^{L_e[n]}}{1+e^{L_e[n]}}\right) + \frac{1}{1+e^{L_e[n]}}\log_2\left(\frac{2}{1+e^{L_e[n]}}\right)\right], \tag{63}$$

where the EXIT charts may be constructed "online", because as soon as new extrinsic LLRs become available at the receiver, they can be used for updating the current estimate of the mutual information [105]. Nonetheless, it is worth noting that (61) based on histograms is the most accurate method of evaluating I_E based on L_e.

For the classic two-stage SCC of Fig. 5, the transfer functions of both the inner and of the outer code may be expressed as:

$$I_{E_M} = T_M(I_{A_M}, SNR), \quad I_{E_D} = T_D(I_{A_D}), \tag{64}$$

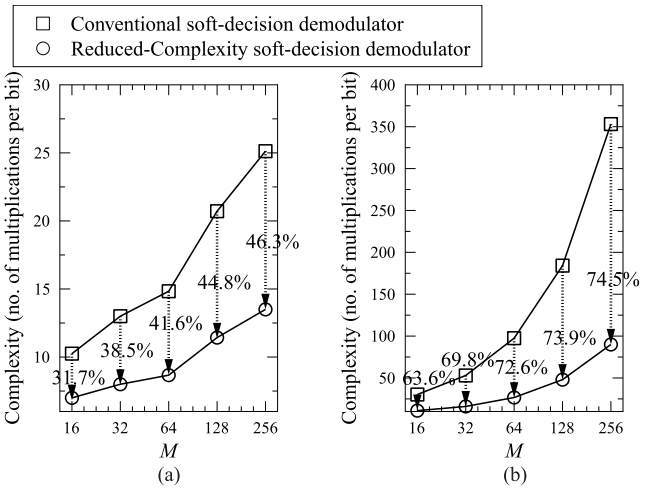


FIGURE 19. Complexity (number of multiplications per bit) comparison between the conventional soft-decision *M*PSK/QAM detection algorithms and the reduced-complexity detection algorithms of [216]. The Complexity Reduction Ratios (CRRs) achieved by the proposed detection algorithms are indicated on the figures. (a) Square *M*QAM. (b) *M*PSK or Star/Cross *M*QAM.

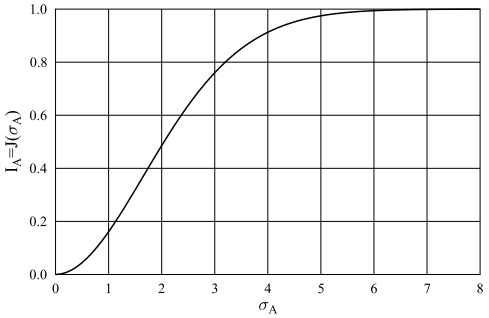


FIGURE 20. The function $J(\cdot)$ of (58).

where the subscripts M and D refer to the demapper and decoder respectively, while naturally T_M of the inner code is a function of both the *a priori* information I_{A_M} and of the channel SNR. Given I_A assuming equi-spaced values from the range of [0, 1], a pair of EXIT curves may be obtained for the transfer functions of (64) with the aid of the algorithm above. Moreover, due to the specific nature of turbo detection, the extrinsic information of the inner code becomes the *a priori* information of the outer code, i.e. we have $I_{E_M} = I_{A_D}$, followed by the extrinsic information of the outer code becoming the *a priori* information of the inner code, i.e. $I_{E_D} = I_{A_M}$. This feature allows us to portray two EXIT curves in a single chart as exemplified by Fig. 7(a). Generally, the intersections between the inner code's EXIT curve and the outer code's EXIT curve get closer to $I_{E_M} = I_{A_D} = 1.0$ as E_b/N_0 increases, which implies that the turbo detector's capability of enhancing our confidence in its input information gradually improves [12]. As a result, in the classic two-stage SCC of Fig. 5, where the inner code is a demodulator, perfect extrinsic information of $I_{E_D} = 1.0$ can only be achieved by the channel decoder, when perfect *a priori* information of

$I_{A_D} = I_{E_M} = 1.0$ is provided by the demodulator, which requires an infinite SNR. This implies that a non-negligible BER exists, unless the inner code's EXIT curve and the outer code's EXIT curve only intersect at the (1.0,1.0) point. In order to achieve this goal, a URC may was introduced as an intermediate code in [106], which equipped the resultant concatenated scheme with a free distance of two that was shown to be the sufficient and neccessary condition for achieving an infinitesimally low BER [106], [225]. For the resultant three-stage turbo receiver, the specific activation order of the component decoders is sometimes also referred to as scheduling in the related literature [222]. Moreover, in this treatise the terminology of 'activation order' is preferred to avoid confusion with 'scheduling' routinely used in resource-allocation. In order to simplify the receiver's analysis, the amalgamated URC and QAM decoder may be viewed as the amalgamated inner code in this treatise.

It can be seen in Fig. 7a that an open tunnel emerges between the inner and outer codes' EXIT curves at $E_b/N_0 = 1.3$ dB, where the only intersection of the two curves is at the (1.0,1.0) point. More explicitly, the requirement for an open EXIT tunnel may be expressed as:

$$\begin{aligned} &T_M(I_{M_A}, SNR) > T_D^{-1}(I_{D_E}), \quad \text{when } I_{M_A} = I_{D_E} \in [0, 1), \\ &T_M(I_{M_A}, SNR) = 1, \quad \text{when } I_{M_A} = I_{D_E} = 1. \end{aligned} \tag{65}$$

Since EXIT curves are obtained by averaging over numerous transmitted frames, the Monte-Carlo simulation based decoding trajectories are subject to small but potentially non-negligible deviations from the EXIT curves' prediction. As a result, it cannot be guaranteed that all Monte-Carlo simulation based decoding trajectories can get through the extremely narrow EXIT tunnel at $E_b/N_0 = 1.3$ dB seen in Fig. 7(a). Nonetheless, Fig. 7(b) shows that an infinitesimally low BER is recorded at a slightly increased E_b/N_0 of 1.7 dB, which implies that all Monte-Carlo simulation based decoding trajectories recorded at $E_b/N_0 = 1.7$ dB can actually get through their open EXIT tunnels. An example of Monte-Carlo simulation based decoding trajectory recorded at $E_b/N_0 = 1.7$ is portrayed in Fig. 7(a).

The concept of mutual information is popularly used for quantifying capacity. More explicitly, the relationship between the DCMC capacity $C^{DCMC} = \max_{p(\mathbf{S})} I(\mathbf{S}; \mathbf{Y})/N_P$ of (5) and the extrinsic information $I_E = I(b; L_e)$ of (61) leads us to the so-called area property of EXIT chart [104], [226] as:

$$\overline{A}_M(SNR) = \int_0^1 T_M(I_A, SNR) dI_A \approx \frac{C^{DCMC}(SNR)}{R}, \tag{66}$$

where the area under the EXIT curve of the inner demapper $\overline{A}_M(SNR)$ is directly linked to the maximum achievable rate. Similarly, the area $\overline{A}_D$ under the EXIT curve of the channel decoder is related to the coding rate as:

$$\overline{A}_D = 1 - \int_0^1 T_D^{-1}(I_E) dI_E \approx 1 - R_c. \tag{67}$$

Against this background, a variety of near-capacity systems have been designed in [12], [99], [101], [227], and [228] by matching the EXIT curve shapes of the inner and outer codes, so that an open tunnel may be encountered at the lowest possible E_b/N_0. In order to approach to this goal, on the one hand, numerous researchers have focused their attention on how to design optimized modulation schemes so that their EXIT curves may match the shape of the outer channel code's EXIT curve. This topic is widely known as bit-to-symbol mapping optimization for BICM-ID [227], [229]–[233]. On the other hand, as the family of modulation schemes keeps evolving, especially in MIMO applications, it becomes more feasible to adjust the channel decoder's transfer characteristics, as seen in [99], [228], and [234].

For example, the 17-point IRCC proposed in [99] is a popular implementation of the aforementioned near-capacity design. More explicitly, the 17-point IRCC is constituted by 17 subcodes associated with code rates of $\{r_k = 0.1 + (k-1) \cdot 0.05\}_{k=1}^{17}$. These subcodes are constructed from a systematic half-rate memory-four mother code, which is defined by the octally represented generator polynomial of $(G_r, G) = (31, 27)_8$. Subcodes with higher rates are obtained by puncturing, while subcodes with lower rates are created by adding more generators and by puncturing. Given the appropriate weighting coefficients of $\{0 \leq \alpha_k \leq 1\}_{k=1}^{17}$, each subcode may encode $\alpha_k \cdot r_k \cdot N_c$ information bits to $\alpha_k \cdot N_c$ coded bits, where N_c refers to the frame length. The IRCC's coefficients have to satisfy the following two conditions:

$$\sum_{k=1}^{17} \alpha_k = 1, \quad \sum_{k=1}^{17} \alpha_k r_k = R_c. \tag{68}$$

As a result, the transfer function of the outer IRCC may be characterized by the weighted superposition of the subcodes' mutual information transfer functions $\{T_{D,k}(I_A)\}_{k=1}^{17}$ as:

$$T_D(I_A) = \sum_{k=1}^{17} \alpha_k T_{D,k}(I_A), \tag{69}$$

where all subcodes are assumed to produce LLRs associated with symmetric and consistent PDFs. In summary, the IRCC's weighting coefficients may be obtained by minimizing the MSE between the mutual information transfer functions of the inner and the outer codes according to:

$$\{\alpha_k\}_{k=1}^{17} = \arg \min_{\{\alpha_k\}_{k=1}^{17}} \int_0^1 |T_D(I) - T_M^{-1}(I, SNR)|^2 dI. \tag{70}$$

We note that the search formulated in (70) may start with the maximum achievable rate's $\overline{SNR}$. If the resultant weighting coefficients cannot produce an EXIT curve for the outer code that matches the inner code's EXIT curve sufficiently well, then SNR shall be increased and (70) is repeatedly evaluated, until a valid group of weighting coefficients of $\{\alpha_k\}_{k=1}^{17}$ is obtained. In this treatise, the MIMO schemes will be tested in the context of RSC, TC as well as IRCC coded systems.

Moreover, as mentioned before, the efficient computation of (62) and (63) is based on the important assumption of

satisfying the symmetric condition of $p(L_e|x) = p(-L_e|-x)$ and the consistency condition of $p(L_e|x=+1) = p(L_e|x=-1)e^{L_e}$. If the symmetricity condition cannot be satisfied, the histogram-based (61) has to be invoked for evaluating I_E. However, if the consistency condition cannot be guaranteed, the LLR definition of $L(b) = \ln \frac{p(b=1)}{p(b=0)}$ will be violated. Let us elaborate a little further here. The consistency condition of $p(L_e|b=1) = p(L_e|b=0)e^{L_e}$ leads to the following relationship:

$$L_e = \ln \frac{p(L_e|b=1)}{p(L_e|b=0)} = \ln \frac{p(b=1|L_e)}{p(b=0|L_e)}, \tag{71}$$

because we have $p(b|L_e) = \frac{p(L_e|b)p(b)}{\sum_{b=\{1,0\}} p(L_e|b)p(b)}$ according to Bayes' law [235], and we have $\{p(b) = 0.5\}_{b=\{1,0\}}$ for equiprobable source bits, the extrinsic LLRs satisfy the relationship defined by (71), their LLR values may be deemed as accurate [7], [99], [224] according to the LLR definition. However, if the extrinsic LLRs produced by a demapper deviates from (71), the excessive LLR values may degrade the turbo detection performance, since they cannot be corrected by the channel decoder after a few iterations.

For example, Figs. 21 and 22 portrays the EXIT charts prediction and the BER performance of coded V-BLAST schemes, where the optimum MAP detectors of Sec. II-A2, the SD of Sec. II-A4 that retains the optimum MAP detection capability and the MMSE detectors of Sec. II-A7 are employed. We note that the "Hard MMSE" seen in Figs. 21 and 22 refers to the soft-decision MPSK/QAM detectors invoking the probability metric of (41), which is directly derived from the hard-decision V-BLAST MMSE of Sec. II-A6. Moreover, the "Exact MMSE" solution refers to the soft-decision MPSK/QAM detectors invoking the probability metric of (47), which is obtained by taking into account the *a priori* LLRs for updating the MMSE filter taps, as derived in Sec. II-A7.

It can be seen in Fig. 21(a) that the "Hard MMSE" used for detecting V-BLAST signals employing BPSK/QPSK exhibits horizontal EXIT curves, while the optimum/SD aided V-BLAST detection benefits from a significant iteration gain. Furthermore, it was discussed in Sec. II-A7 that the exact MMSE solution associated with $I_A = 0$ is equivalent to the hard-decision MMSE detector, while the exact MMSE solution associated with $I_A = 1$ is equivalent to the optimum MAP V-BLAST detector, which is verified by Fig. 21(a). Despite the associated performance loss, Fig. 21(b) shows that the "Hard MMSE" may produce unreliable LLRs, which deviate from the true probabilities. These unreliable LLRs cannot be readily corrected by the channel decoder, hence "Hard MMSE" is not recommended for turbo detection.

Fig. 22 further characterizes the performance of these soft-decision V-BLAST detectors in the context of TC coded systems. In order to achieve an iteration gain, $IR_{TC} = 4$ inner iterations are carried out within the TC and $IR_{TC-MIMO} = 4$ outer iterations are employed between the TC and MIMO receiver for MAP-optimum/SD V-BLAST detection and

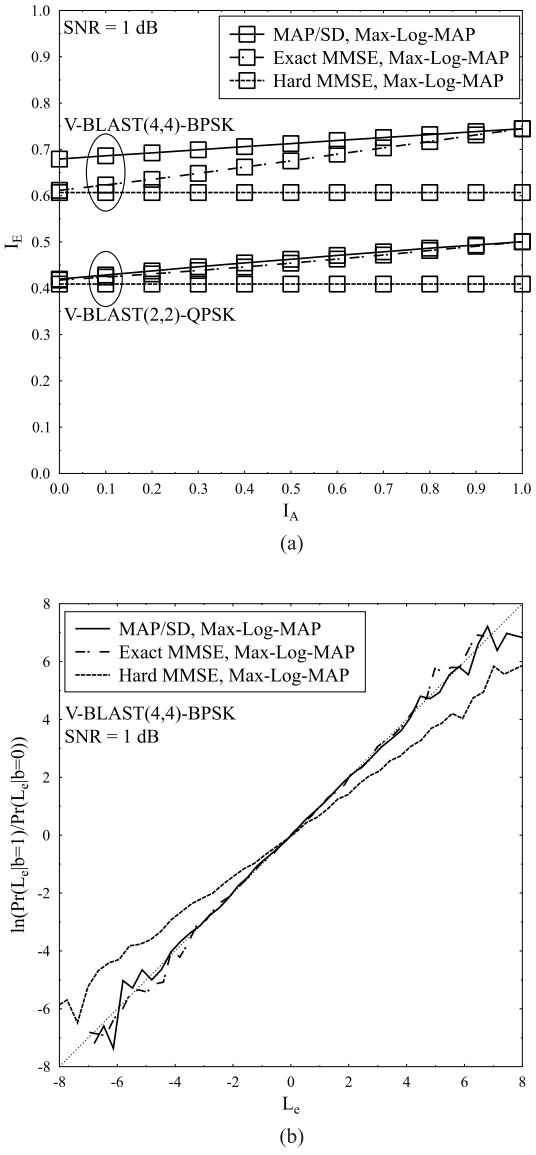


FIGURE 21. EXIT charts and LLR accuracy test for soft-decision MAP/SD and MMSE V-BLAST detectors, where the throughput is given by $R = 4$ bits/block/channel use. (a) EXIT Charts. (b) LLR accuracy test.

exact MMSE solution. Meanwhile, we have $IR_{TC} = 16$ and $IR_{TC-MIMO} = 1$, when "Hard MMSE" is employed, so that the turbo detection is configured to maintain the same total number of iterations. It can be seen in Fig. 22 that the MAP-optimum/SD V-BLAST detectors may achieve an excellent performance that is within 1.0 dB from the maximum achievable rate, which is the E_b/N_0 bound that has to be satisfied for achieving half of the full DCMC capacity of (5). By contrast, the low-complexity "Hard MMSE", which is associated with mutual information loss in the EXIT charts of Fig. 21(a) and with the unreliable LLRs seen in Fig. 21(b), imposes a substantial overall performance degradation, as evidenced by Fig. 22. It is further demonstrated by Fig. 22 that the exact MMSE detector is capable of performing close to the optimum V-BLAST detector.

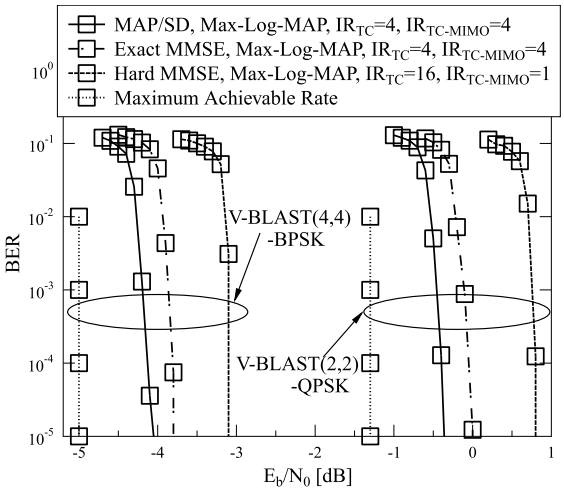


FIGURE 22. BER performance of half-rate TC coded V-BLAST associated with the same system throughput of $R_c R = 2$.

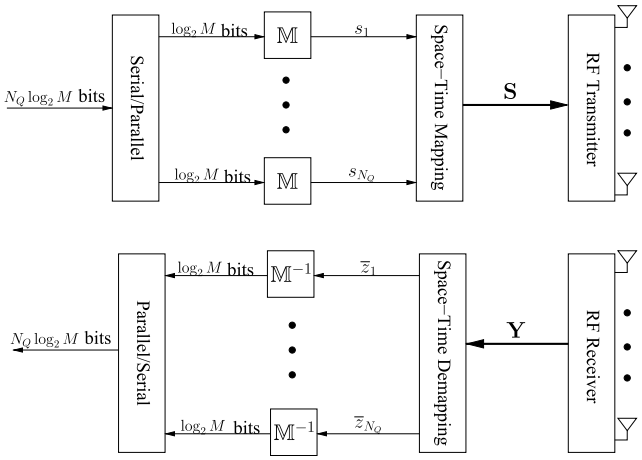


FIGURE 23. Schematic of an orthogonal STBC transceiver.

B. SPACE-TIME BLOCK CODE (STBC)

The schematic of orthogonal STBC transceivers is depicted in Fig. 23. An STBC transmitter firstly encodes the N_QBPS source bits into N_Q modulated MPSK/QAM symbols $\{s_q\}_{q=1}^{N_Q}$. During N_P symbol periods, the $(N_P \times N_T)$-element symbol-matrix transmitted from the N_T TAs may be formulated by:

$$\mathbf{S} = \sqrt{P_t} G_{N_T}(\{s_q\}_{q=1}^{N_Q}) \tag{72a}$$

$$= \sqrt{P_t} \sum_{q=1}^{N_Q} \left[\mathbf{A}_q \Re(s_q) + j\mathbf{B}_q \Im(s_q)\right], \tag{72b}$$

where $G_{N_T}(\cdot)$ represents the real and imaginary parts of the transmission matrix by dispersing the real and imaginary parts of the modulated MPSK/QAM symbols into the $(N_P \times N_T)$-element real-valued matrices $\{\mathbf{A}_q\}_{q=1}^{N_Q}$ and $\{\mathbf{B}_q\}_{q=1}^{N_Q}$, respectively, while the normalization factor P_t is introduced in order to guarantee satisfying the power constraint of $\mathrm{E}\left[\mathrm{tr}(\mathbf{S}^H\mathbf{S})\right] = N_P$.

We note that the V-BLAST transmission matrix shown in Table 4 may also be framed according to (72b). The corresponding dispersion matrices used for V-BLAST are given by:

$$\mathbf{A}_q = \mathbf{B}_q = [\underbrace{0 \cdots 0}_{q-1}, 1, \underbrace{0 \cdots 0}_{N_T-q}], \quad 1 \le q \le N_T, \tag{73}$$

where we have $N_T = N_Q$ and $N_P = 1$. Moreover, the power normalization factor is given by $P_t = \frac{1}{N_T}$. It can be seen in (73) that the V-BLAST transmission matrix is constructed in spatial domain only.

1) GENERAL ORTHOGONAL DESIGN GUIDELINES

The objectives of the STBC design are two-fold: to minimize the error probability of (7) and to employ the low-complexity linear receiver portrayed in Fig. 23 without encountering the V-BLAST's inter-antenna interference problem. In order to achieve the former goal, the PEP of (8) should be minimized by achieving full diversity and maximizing the coding gain. In order to achieve the second objective, the MIMO's inter-antenna interference should be able to be cancelled out before invoking a linear MPSK/QAM demodulator at the receiver. Let us firstly consider the codeword difference formulated in the PEP upper bound of (8) according to the STBC transmission matrix of (72) as:

$$\begin{aligned}\mathbf{S}^i - \mathbf{S}^{\bar{i}} &= \sqrt{P_t} \sum_{q=1}^{N_Q} \left[\mathbf{A}_q \Re(s_q^i - s_q^{\bar{i}}) + j\mathbf{B}_q \Im(s_q^i - s_q^{\bar{i}})\right] \\ &= \sqrt{P_t} G_{N_T}(\{s_q^i - s_q^{\bar{i}}\}_{q=1}^{N_Q}).\end{aligned} \tag{74}$$

Therefore, when Hadamard's inequality [236] is applied to the determinant criterion of (8), it can be seen that the optimality condition is that $\mathbf{\Delta} = (\mathbf{S}^i - \mathbf{S}^{\bar{i}})^H(\mathbf{S}^i - \mathbf{S}^{\bar{i}})$ is unitary, which requires that $\mathbf{S}^i - \mathbf{S}^{\bar{i}} = \sqrt{P_t} G_{N_T}(\{s_q^i - s_q^{\bar{i}}\}_{q=1}^{N_Q})$ have orthogonal columns. This reveals that in general, the STBC transmission matrix $\mathbf{S} = \sqrt{P_t} G_{N_T}(\{s_q\}_{q=1}^{N_Q})$ should always have orthogonal columns, which requires $N_P \ge N_T$. Furthermore, when the signal vectors transmitted by N_T TAs are orthogonal to each other, they are expected to be decoupled at the receiver without encountering the V-BLAST's IAI problem.

If we also take into account all the considerations including performance, cost and delay, the STBC from orthogonal design may be translated into the following stringent design requirements [12], [34], [35], [122]–[124], [166], [237]:

The first requirement (R1) results in the maximum attainable normalized throughput of $\overline{R} = \frac{N_Q}{N_P} = 1$, so that the employment of multiple TAs for STBC systems would not end up with a lower throughput than that of the SISO and SIMO systems. The second requirement (R2) minimizes the transmission delay while maintaining the transmit diversity order, which is given by $\min\{N_T, N_P\}$ according to (8). The third requirement (R3) simplifies the hardware design of the RF amplifiers by minimizing the peak-to-average ratio. Lastly, the orthogonality requirement (R4) is the key both to

the minimized error probability and to the low-complexity interference-free linear STBC receiver, where the multiple streams may be individually detected.

Orthogonal STBC Design Requirements

(R1) Full Unity-Rate Requirement: $N_P = N_Q$.
(R2) Delay Optimality Requirement: $N_P = N_T$.
(R3) Hardware Simplicity Requirement: all the elements in $G_{N_T}(\{s_q\}_{q=1}^{N_Q})$ of (72) should be taken from $\{0, \pm s_q, \pm s_q^*\}_{\forall q=1}^{N_Q}$.
(R4) Orthogonality Requirement: the transmission matrix of (72) should have orthogonal columns so that we have:

$$\mathbf{S}^H\mathbf{S} = \frac{N_P}{N_T}\frac{\sum_{q=1}^{N_Q}|s_q|^2}{N_Q}\mathbf{I}_{N_T}, \tag{75}$$

which complies with the power constraint of $\mathrm{E}\left[\mathrm{tr}(\mathbf{S}^H\mathbf{S})\right] = N_P$.

2) LINEAR STBC DETECTION

Let us now proceed to characterize the interference-free linear STBC receiver by further exploring the orthogonality requirement (R4). First of all, the STBC transmission matrix of (72) may be expressed in the following alternative form:

$$\mathbf{S} = \sqrt{P_t}\sum_{q=1}^{N_Q}\left(\mathbf{D}_q^+ s_q + \mathbf{D}_q^- s_q^*\right), \tag{76}$$

where the alternative dispersion matrices in (76) are given by $\left\{\mathbf{D}_q^+ = \frac{1}{2}(\mathbf{A}_q + \mathbf{B}_q)\right\}_{q=1}^{N_Q}$ and $\left\{\mathbf{D}_q^- = \frac{1}{2}(\mathbf{A}_q - \mathbf{B}_q)\right\}_{q=1}^{N_Q}$. Following this, the matrix norm term in the probability $p(\mathbf{Y}|\mathbf{S})$ expression of (2) may be extended as $\|\mathbf{Y}-\mathbf{S}\mathbf{H}\|^2 = \|\mathbf{Y}\|^2 - \mathrm{tr}(\mathbf{Y}^H\mathbf{S}\mathbf{H}) - \mathrm{tr}(\mathbf{H}^H\mathbf{S}^H\mathbf{Y}) + \mathrm{tr}(\mathbf{H}^H\mathbf{S}^H\mathbf{S}\mathbf{H})$, where both $\|\mathbf{Y}\|^2$ and $\mathrm{tr}(\mathbf{H}^H\mathbf{S}^H\mathbf{S}\mathbf{H}) = \frac{N_P\|\mathbf{H}\|^2}{N_T N_Q}|s_q|^2$ are constants thanks to the associated orthogonality requirement (R4), while $\mathbf{S}$ may be expressed by (76) in order to decouple $\{s_q\}_{q=1}^{N_Q}$. In summary, we have [237]:

$$\|\mathbf{Y}-\mathbf{S}\mathbf{H}\|^2 = \sum_{q=1}^{N_Q}\left(\frac{N_P\|\mathbf{H}\|^2}{N_T N_Q}|\bar{z}_q - s_q|^2\right) + \varrho, \tag{77}$$

where the decision variable is given by:

$$\bar{z}_q = \frac{N_T N_Q\sqrt{P_t}}{N_P\|\mathbf{H}\|^2}\mathrm{tr}\left[\mathbf{Y}^H\mathbf{D}_q^-\mathbf{H} + \mathbf{H}^H(\mathbf{D}_q^+)^H\mathbf{Y}\right], \tag{78}$$

and the constant is given by $\varrho = \|\mathbf{Y}\|^2 - \sum_{q=1}^{N_Q}\frac{N_P\|\mathbf{H}\|^2}{N_T N_Q}|\bar{z}_q|^2$. As a result, the conditional probability of receiving $\mathbf{Y}$, when $\mathbf{S}$ is transmitted in (2) may be decoupled as:

$$p(\mathbf{Y}|\mathbf{S}) = \vartheta\prod_{q=1}^{N_Q} p(\bar{z}_q|s_q), \tag{79}$$

where the constant is given by $\vartheta = \frac{(\pi\overline{N}_0)^{N_Q}}{(\pi N_0)^{N_R N_P}}\exp(-\frac{\varrho}{N_0})$, so that the equivalent conditional probability of receiving $\bar{z}_q$, when s_q is transmitted may be expressed as:

$$p(\bar{z}_q|s_q) = \frac{1}{\pi\overline{N}_0}\exp\left(-\frac{|\bar{z}_q - s_q|^2}{\overline{N}_0}\right), \tag{80}$$

where the equivalent noise power is given by $\overline{N}_0 = \frac{N_T N_Q}{N_P\|\mathbf{H}\|^2}N_0$. The ML/MAP detector aims for maximizing the *a posteriori* probability $p(\mathbf{S}|\mathbf{Y})$ of (10), where the constant ϑ in (79) may be cancelled out by the division operation in Bayes' law seen in (10). Therefore, we may now conclude that the STBC may invoke a linear MPSK/QAM demodulator for recovering s_q from $\bar{z}_q$ without encountering the BLAST MIMO's IAI problem. More explicitly, the hard-decision aided linear STBC detection may be carried out as $\hat{s}_q = \mathbb{M}^{-1}(\bar{z}_q)$ for $q \in \{1, \cdots, N_Q\}$, which is similar to the hard-decision LF-aided V-BLAST introduced in Sec. II-A6. Similarly, the soft-decision linear STBC detection may be carried out in the same way as the soft-decision LF aided V-BLAST introduced in Sec. II-A7, where $\widetilde{z}_v^{MMSE}$ and $\widetilde{N}_0^{MMSE}$ in the probability metric of (47) may be replaced by $\bar{z}_q$ and $\overline{N}_0$ of (80). Naturally, the reduced-complexity soft-decision PSK/QAM demodulators of Sec. II-A8 may be invoked by the linear soft-decision STBC detection.

3) ERROR PROBABILITY AND CAPACITY OF STBCs

It is shown by (79) that the STBC detection in fading channels may be transformed into decoupled MPSK/QAM detection in AWGN channels without any performance loss. Therefore, considering that the average BER of (7) is approximated based on the evaluation of the PEP, which is only accurate in the high-SNR region, the error probability of the STBC in fading channels may be more closely evaluated by the performance of MPSK/QAM schemes in AWGN channels [165], [238], where the noise power is given by $\overline{N}_0 = \frac{N_T N_Q}{N_P\|\mathbf{H}\|^2}N_0$. More specifically, if the full unity-rate requirement (R1) is guaranteed in the STBC design, the equivalent noise power becomes $\overline{N}_0 = \frac{N_T}{\|\mathbf{H}\|^2}N_0 = \frac{1}{\left(\sum_{v=1}^{N_T}\|\mathbf{H}_{v,-}\|^2\right)/N_T}N_0$, which explicitly reveals the benefit of having diversity gain. More explicitly, the divisor of $\left(\sum_{v=1}^{N_T}\|\mathbf{H}_{v,-}\|^2\right)/N_T$ is averaged over the fading samples gleaned from the N_T TAs, which implies that the equivalent noise power would not be readily amplified by a single deep fading channel.

It was recognized in [36], [125], and [239] that STBCs cannot achieve the full MIMO capacity except for a single special case, which is Alamouti's G2-STBC system associated with a single RA $N_R = 1$. Let us now elaborate a little further here on this issue, so that the multiplexing versus diversity tradeoff of MIMO system design may be better augmented.

According to the equivalent input/output relationship of (80), the maximized mutual information of STBC is given

by:

$$C_{STBC}^{CCMC}(SNR) = \max_{\{p(s_q)\}_{q=1}^{N_Q}} \frac{1}{N_P}\sum_{q=1}^{N_Q}\left[H(\bar{z}_q) - H(\bar{z}_q|s_q)\right] = \frac{N_Q}{N_P}\mathrm{E}\left[\log_2\left(1+\frac{N_P\|\mathbf{H}\|^2}{N_T N_Q}\eta\right)\right], \quad (81)$$

where we have $H(\bar{z}_q|s_q) = \log_2\left[\pi e\left(\frac{N_T N_Q}{N_P\|\mathbf{H}\|^2}N_0\right)\right]$ and $H(\bar{z}_q) = \log_2\left[\pi e\left(1+\frac{N_T N_Q}{N_P\|\mathbf{H}\|^2}N_0\right)\right]$ according to $p(\bar{z}_q|s_q)$ of (80) and the assumption of Gaussian input PDFs $\{p(s_q)\}_{q=1}^{N_Q}$.

Considering a V-BLAST MIMO system equipped with N_T' and N_R' antennas operating at an SNR of η', the term of $\frac{\eta'}{N_T'}\mathbf{H}'^H\mathbf{H}'$ in the MIMO capacity of (4) can only be equal to the term of $\frac{N_P\|\mathbf{H}\|^2}{N_T N_Q}\eta$ in the STBC capacity of (81), when we have $N_T' = N_T N_R$, $N_R' = 1$ and $\eta' = \frac{N_R N_P}{N_Q}\eta$. In other words, the relationship between the STBC capacity and the V-BLAST MIMO capacity may be expressed as [36], [125], [239]:

$$C_{STBC}^{CCMC}(N_T, N_R, \eta) = \frac{N_Q}{N_P}C_{MIMO}^{CCMC}(N_T N_R, 1, \frac{N_R N_P}{N_Q}\eta) \le C_{MIMO}^{CCMC}(N_T, N_R, \eta), \quad (82)$$

where the equality only holds, when we have $N_T = N_P = N_Q$ and $N_R = 1$, which may only be satisfied by Alamouti G2-STBC scheme equipped with a single RA of $N_R = 1$.

It becomes clear now that there is a tradeoff amongst the conflicting capacity, performance and complexity in MIMO systems design. More explicitly, the V-BLAST MIMO introduced in Sec. II-A achieves the maximum attainable MIMO throughput that is N_T times higher than a SISO/SIMO system throughput. By contrast, the STBC MIMO introduced in Sec. II-B minimizes the MIMO's PEP bound and benefits from a low signal processing complexity at the receiver, but it cannot achieve the maximum achievable MIMO capacity.

4) FULL UNITY-RATE STBC

When complex-valued high-throughput MPSK/QAM constellations are employed, it was proven in [35] that the only STBC satisfying all the requirements listed in Sec. II-B1 is Alamouti's G2-STBC [34], whose codeword is constructed by:

$$G_2(s_1, s_2) = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}. \quad (83)$$

It can be seen in 83 that Alamouti's G2-STBC transmits ($N_Q = 2$) modulated MPSK/QAM symbols by ($N_T = 2$) TAs over ($N_P = 2$) 'channel uses'. Therefore, the G2-STBC satisfies the full unity-rate requirement (R1), the delay optimal requirement (R2) and the transmitter's hardware requirement (R3) discussed in Sec. II-B1. Furthermore, we also have $[G_2(s_1, s_2)]^H G_2(s_1, s_2) = (|s_1|^2 + |s_2|^2)\mathbf{I}_2$ according to (83). Therefore, according to (72a), the G2-STBC's transmission matrix is given by $\mathbf{S} = \frac{1}{\sqrt{2}}G_2(s_1, s_2)$, since the power normalization factor in (72) is given by $P_t = \frac{1}{2}$, so that the orthogonality requirement (R4) in Sec. II-B1 may also be fully met. As a result, the linear STBC receiver developed in Sec. II-B1 may be applied to Alamouti's G2-STBC as seen in [35].

5) HALF-RATE STBCs

When the family of real-valued constellations is considered, the orthogonal design satisfying the four requirements listed in Sec. II-B1 does exist for $N_T = 2$, 4 or 8 [35], which may be solved by the Hurwitz-Radon theory of [120] and [121]. We note that the conjugation operation $\{s_q^*\}_{q=1}^{N_Q}$ may be eliminated from the requirement (R3) of Sec. II-B1 for real-valued signalling. More specifically, for $N_T = 2$ TAs, the real-valued orthogonal design $G_2^{\Re}(s_1, s_2)$ is the same as Alamouti's G2-STBC design of $G_2(s_1, s_2)$ seen in (83) without the conjugation operations. Moreover, for the cases of $N_T = 4$ and $N_T = 8$, the STBCs constructed from real-valued orthogonal design are given by $G_4^{\Re}(s_1, s_2, s_3, s_4)$ and $G_8^{\Re}(s_1, s_2, s_3, s_4, , s_5, s_6, s_7, s_8)$ according to (4) and (5) in [35], respectively.

In order to accommodate complex-valued MPSK/QAM symbols, the Half-Rate (HR)-G4-STBC may be obtained by vertically concatenating the STBC from real-valued orthogonal design and its conjugates as:

$$G_4(s_1, s_2, s_3, s_4) = \begin{bmatrix} G_4^{\Re}(s_1, s_2, s_3, s_4) \\ G_4^{\Re}(s_1, s_2, s_3, s_4)^* \end{bmatrix}. \quad (84)$$

Furthermore, the HR-G3-STBC design of $G_3(s_1, s_2, s_3, s_4)$ may be constructed by taking the first three columns in $G_4(s_1, s_2, s_3, s_4)$. Similarly, the HR-G8-STBC may also be obtained by vertically concatenating $G_8^{\Re}(s_1, s_2, s_3, s_4, , s_5, s_6, s_7, s_8)$ and its conjugates. Accordingly, the HR-GN_T-STBC design of $G_{N_T}(s_1, s_2, s_3, s_4, , s_5, s_6, s_7, s_8)$ associated with $5 \le N_T \le 7$ may be constructed by taking the first N_T columns in $G_8(s_1, s_2, s_3, s_4, , s_5, s_6, s_7, s_8)$.

It may be observed that all the HR-GN_T-STBCs associated with $3 \le N_T \le 8$ fail to meet the full unity-rate requirement (R1) of Sec. II-B1, resulting in a normalized throughput of $\overline{R} = \frac{N_Q}{N_P} = \frac{1}{2}$. Similarly, the delay optimal requirement (R2) becomes $N_P = 2N_Q$. However, the transmitter's hardware requirement (R3) is still satisfied by the half-rate STBCs. Furthermore, it may be observed that we always have $G_{N_q}(s_1, \cdots, s_{N_Q})^H G_{N_q}(s_1, \cdots, s_{N_Q}) = \sum_{q=1}^{N_Q} 2|s_q|^2\mathbf{I}_{N_T}$ for $3 \le N_T \le 8$ according to the half-rate STBC design, hence the HR-STBC's transmission matrix of (72a) may be expressed as $\mathbf{S} = \sqrt{\frac{N_P}{2N_T N_Q}}G_{N_T}(s_1, \cdots, s_{N_Q})$, since the power normalization factor of (72) is given by $P_t = \frac{N_P}{2N_T N_Q}$. As a result, the orthogonality requirement (R4) facilitating single-stream detection is fully satisfied by the half-rate STBCs. Therefore, the linear STBC receiver developed in Sec. II-B1 may also be applied to them. We note that no STBCs having $N_T > 8$ were explicitly constructed, but it was

proven in [35] that such a design may impose a substantial delay growing exponentially with N_T, which is given by $N_P = 16 \times 16^{(N_T/8-1)}$ for $N_T > 8$ with N_T being a power of 2.

6) AMICABLE ORTHOGONAL STBCs

In order to improve the throughput of STBCs associated with $N_T > 2$, it was demonstrated in [35] and [166] that rate 3/4 STBC exists for $N_T = 3$ and $N_T = 4$. However, these alternative STBCs do not obey the transmitter's hardware requirement of (R3) in Sec. II-B1, which implies that the linear region of the MIMO's amplifier has to be extended. As a remedy, the Amicable Orthogonal (AO) STBCs obtained according to the theory of amicable orthogonal design [120] were presented for $N_T = 4$ and $N_T = 8$ in [122]–[124] and then generalized for any values of N_T in [240]–[242]. In more details, if the number of TAs is a power of 2 as $N_T = 2^\iota$ for a positive integer of $\iota \geq 1$, the general AO-GN_T-STBC design may be formulated as:

$$G_{2^\iota}^{\mathrm{AO}}(s_1, \cdots, s_{\iota+1}) = \begin{bmatrix} G_{2^{\iota-1}}^{\mathrm{AO}}(s_1, \cdots, s_\iota) & s_{\iota+1}\mathbf{I}_{2^{\iota-1}} \\ -s_{\iota+1}^*\mathbf{I}_{2^{\iota-1}} & G_{2^{\iota-1}}^{\mathrm{AO}}(s_1, \cdots, s_\iota)^H \end{bmatrix}. \quad (85)$$

It can be seen in (85) that if the AO-STBC design starts from $\iota = 1$ and $G_1^{\mathrm{AO}}(s_1) = s_1$, then Alamouti's G2-STBC of (83) may be obtained from (85) as $G_2^{\mathrm{AO}}(s_1, s_2) = G_2(s_1, s_2)$. Similarly, for all the cases of AO-GN_T-STBC associated with $N_T = 2^\iota$, the STBC design requirements of (R2), (R3) and (R4) in Sec. II-B1 are satisfied.

For the scenarios of N_T not being a power of 2, the AO-GN_T-STBC design may be obtained by taking the first N_T columns of $G_{2^{\bar{\iota}}}^{\mathrm{AO}}(s_1, \cdots, s_{\bar{\iota}+1})$, where we have $\bar{\iota} = \lceil \log_2 N_T \rceil$. These AO-STBCs do not obey the delay optimal requirement of (R2) in Sec. II-B1. Nonetheless, the transmission delays of AO-STBCs are generally substantially lower than their HR-STBC counterparts discussed in Sec. II-B5. For example, the AO-G3-STBC and AO-G4-STBC have $N_P = 4$, while the AO-GN_T-STBC for $5 \leq N_T \leq 8$ have $N_P = 8$, which are halves of the parameters of the HR-STBCs in Sec. II-B5.

In summary, owing to the fact that we always have $G_{N_T}^{\mathrm{AO}}(s_1, \cdots, s_{N_Q})^H G_{N_T}^{\mathrm{AO}}(s_1, \cdots, s_{N_Q}) = \sum_{q=1}^{N_Q} |s_q|^2 \mathbf{I}_{N_T}$ according to (85), the AO-STBC transmission matrix may be expressed according to (72a) as $\mathbf{S} = \sqrt{\frac{N_P}{N_T N_Q}} G_{N_T}^{\mathrm{AO}}(s_1, \cdots, s_{N_Q})$, where the power normalization factor seen in (72) is given by $P_t = \frac{N_P}{N_T N_Q}$. Since the AO-STBC transmission matrix satisfies the orthogonality requirement (R4) of Sec. II-B1, the linear STBC receiver developed in Sec. II-B1 may be directly invoked for the AO-STBCs.

Furthermore, the number of time slots N_P will not increase exponentially with N_T for the AO-STBC design according to (85), as opposed to the HR-STBCs in Sec. II-B5. However, it can be observed that the AO-STBCs associated with $5 \leq N_T \leq 7$ also have a normalized throughput of $\overline{R} = \frac{1}{2}$, which is exactly the same as that of their HR-STBCs counterparts of Sec. II-B5. Moreover, since the AO-STBC's number of transmitted symbols N_Q only increases logarithmically with N_T according to $N_Q = \lceil \log_2 N_T \rceil + 1$, the normalized throughput of AO-STBC is expected to be lower than $\overline{R} = \frac{1}{2}$ for $N_T > 8$.

C. LINEAR DISPERSION CODE (LDC)

In this section, we firstly introduce the family of Quasi-Orthogonal (QO)-STBCs [126], [127], [242]–[246] as the intermediate step for improving the STBC capacity, which can only be achieved by relaxing the orthogonality requirements detailed in Sec. II-B1. In Sec. II-C2, the STBC capacity is further improved by the high-rate LDC design philosophy of [36] proposing to randomly populate the dispersion matrices of (72) in order to find the specific set, which maximizes the CCMC capacity. However, we will also demonstrate in Sec. II-C2 that the LDCs of [36], which separately disperse the real and imaginary parts of the modulated symbols fail to achieve the maximum attainable MIMO capacity. In order to mitigate this problem, the set of so-called capacity-achieving LDCs proposed in [37] and [38] are summarized in Sec. II-C3, where the MIMO capacity may be approached, while attaining a beneficial diversity gain.

1) QUASI-ORTHOGONAL STBCs

In order to improve the attainable STBC throughput, the first step is to relax the orthogonality requirement of Sec. II-B1 at the cost of encountering IAI and hence requiring multi-stream detection. In the light of this principle, the concept of QO-STBC design was proposed in [126] and [243]. In more details, provided that the number of TAs is a power of 2 according to $N_T = 2^\iota$ and ($\iota > 1$), the QO-STBC transmission codeword is constructed from the AO-STBC of (85) as [126], [242]:

$$G_{2^\iota}^{QO}(s_1, \cdots, s_{2\iota}) = \begin{bmatrix} G_{2^{\iota-1}}^{AO}(s_1, \cdots, s_\iota) & G_{2^{\iota-1}}^{AO}(s_{\iota+1}, \cdots, s_{2\iota}) \\ -G_{2^{\iota-1}}^{AO}(s_{\iota+1}, \cdots, s_{2\iota})^* & G_{2^{\iota-1}}^{AO}(s_1, \cdots, s_\iota)^* \end{bmatrix}. \quad (86)$$

It can be seen that the term $s_{\iota+1}\mathbf{I}_{2^{\iota-1}}$ that can only transmit a single modulated symbol in the context of the AO-STBC design of (85) is replaced by the term $G_{2^{\iota-1}}^{AO}(s_{\iota+1}, \cdots, s_{2\iota})$ that may transmit ι symbols in conjunction with the QO-STBC design of (86). As a result, for any number of TAs, the normalized throughput of QO-STBC is increased to $\overline{R} = \frac{2\bar{\iota}}{2^{\bar{\iota}}}$, where we have $\bar{\iota} = \lceil \log_2 N_T \rceil$.

It may be observed in (86) that we always have $\mathrm{tr}\left[G_{2^\iota}^{QO}(s_1, \cdots, s_{2\iota})^H G_{2^\iota}^{QO}(s_1, \cdots, s_{2\iota})\right] = N_T(\sum_{q=1}^{N_Q} |s_q|^2)$. Therefore, the power normalization factor of (72a) is given by $P_t = \frac{N_P}{N_T N_Q}$, and the QO-STBC transmission matrix may be formulated as $\mathbf{S} = \sqrt{\frac{N_P}{N_T N_Q}} G_{2^\iota}^{QO}(s_1, \cdots, s_{2\iota})$, so that the power constaint of $\mathrm{E}\left[\mathrm{tr}(\mathbf{S}^H\mathbf{S})\right] = N_P$ may be satisfied. However, the orthogonality requirement of (R4) in Sec. II-B1 cannot be satisfied, because the columns in $G_{2^{\iota-1}}^{AO}(s_1, \cdots, s_\iota)$ and the columns in $G_{2^{\iota-1}}^{AO}(s_{\iota+1}, \cdots, s_{2\iota})$ are not orthognal to

each other, despite the fact that the columns are orthogonal within each transmission sub-group.

It was suggested in [127] and [243] that linear MIMO receivers such as the MMSE detector or the ZF detector may be invoked for QO-STBC systems. However, this may not be an ideal solution because the sub-optimal linear MIMO receivers fail to fully exploit QO-STBC's diversity gain. Moreover, a lot of research efforts [127], [242], [244]–[246] have been dedicated to improving both the capacity and the performance of QO-STBC designs by modifying the signal constellations. Nonetheless, the QO-STBC serves as an intermediate solution between the STBC and V-BLAST MIMO design, while the STBC's limitations imposed on the capacity and throughput have not been completely solved. In the following section, we continue by introducing the concept of LDC, which aims for systematically bridging the gap between the STBC and V-BLAST.

2) CAPACITY-IMPROVING LDCs

Motivated by the limitations of STBCs, the LDC concept was proposed in [36] in order to improve the STBC's capacity, while attaining the maximum achievable diversity order. First of all, the STBC's transmission matrix model of (72) may be rewritten for LDCs as:

$$\mathbf{S} = \sum_{q=1}^{N_Q} \left[\overline{\mathbf{A}}_q \Re(s_q) + j\overline{\mathbf{B}}_q \Im(s_q) \right], \tag{87}$$

where the dispersion matrices $\{\overline{\mathbf{A}}_q\}_{q=1}^{N_Q}$ and $\{\overline{\mathbf{B}}_q\}_{q=1}^{N_Q}$ are constructed according to our capacity-improving and diversity-maintaining requirements, which will be detailed later. Moreover, $\{s_q\}_{q=1}^{N_Q}$ represent modulated MPSK/QAM symbols, which are dispersed in both the spatial domain and time domain by the dispersion matrices of (87). We note that the transmission model of (87) may include both the STBC and V-BLAST schemes, where the dispersion matrices of (87) are normalized version of those introduced in Sec. II-B1, so that the power constraint of $\mathrm{E}\left[\mathrm{tr}(\mathbf{S}^H\mathbf{S})\right] = N_P$ may be satisfied.

In order to overcome the throughput disadvantage of STBCs, the number of transmitted symbols N_Q may be increased even beyond N_P, so that the V-BLAST throughput may be approached. Furthermore, the LDCs are still suggested to maintain $N_T = N_P$ in order to retain the maximum attainable transmit diversity order at the lowest transmission delay. According to the MIMO received signal model of (1), the LDC's signal received during the t-th time slot ($1 \leq t \leq N_P$) may be expressed as $\mathbf{Y}_{t,-} = \sum_{q=1}^{N_Q} [\overline{\mathbf{A}}_q^{t,-}\mathbf{H}\Re(s_q) + j\overline{\mathbf{B}}_q^{t,-}\mathbf{H}\Im(s_q)] + \mathbf{V}_{t,-}$, where the N_R-element row-vectors $\{\mathbf{Y}_{t,-}\}_{t=1}^{N_P}$ and $\{\mathbf{V}_{t,-}\}_{t=1}^{N_P}$ are taken from the received signal matrix $\mathbf{Y}$ and the AWGN matrix $\mathbf{V}$ in (1), respectively. Moreover, the N_T-element row-vectors $\{\overline{\mathbf{A}}_q^{t,-}\}_{t=1}^{N_P}$ and $\{\overline{\mathbf{B}}_q^{t,-}\}_{t=1}^{N_P}$ are taken from the dispersion matrices of (87). Let us now decouple the real and imaginary parts of $\mathbf{Y}_{t,-}$ as $\Re(\mathbf{Y}_{t,-}) = \sum_{q=1}^{N_Q}\{[\Re(\overline{\mathbf{A}}_q^{t,-})\Re(\mathbf{H}) - \Im(\overline{\mathbf{A}}_q^{t,-})\Im(\mathbf{H})]\Re(s_q) - [\Re(\overline{\mathbf{B}}_q^{t,-})\Im(\mathbf{H}) + \Im(\overline{\mathbf{B}}_q^{t,-})\Re(\mathbf{H})]\Im(s_q)\} + \Re(\mathbf{V}_{t,-})$ and $\Im(\mathbf{Y}_{t,-}) = \sum_{q=1}^{N_Q}\{[\Re(\overline{\mathbf{A}}_q^{t,-})\Im(\mathbf{H}) + \Im(\overline{\mathbf{A}}_q^{t,-})\Re(\mathbf{H})]\Re(s_q) + [\Re(\overline{\mathbf{B}}_q^{t,-})\Re(\mathbf{H}) - \Im(\overline{\mathbf{B}}_q^{t,-})\Im(\mathbf{H})]\Im(s_q)\} + \Im(\mathbf{V}_{t,-})$, which leads to the following equivalent received signal model for the LDC of (87) as:

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{S}} \cdot \widetilde{\chi} \cdot \widetilde{\mathbf{H}} + \widetilde{\mathbf{V}}, \tag{88}$$

where the matrices are given by:

$$\begin{aligned}
\widetilde{\mathbf{Y}} &= [\Re\{\mathrm{rvec}(\mathbf{Y})\}, \Im\{\mathrm{rvec}(\mathbf{Y})\}], \\
\widetilde{\mathbf{S}} &= \left[\Re(s_1), \cdots, \Re(s_{N_Q}), \Im(s_1), \cdots, \Im(s_{N_Q})\right], \\
\widetilde{\chi} &= \begin{bmatrix} \Re\{\mathrm{rvec}(\overline{\mathbf{A}}_1)\}, & \Im\{\mathrm{rvec}(\overline{\mathbf{A}}_1)\} \\ \vdots & \vdots \\ \Re\{\mathrm{rvec}(\overline{\mathbf{A}}_{N_Q})\}, & \Im\{\mathrm{rvec}(\overline{\mathbf{A}}_{N_Q})\} \\ -\Im\{\mathrm{rvec}(\overline{\mathbf{B}}_1)\}, & \Re\{\mathrm{rvec}(\overline{\mathbf{B}}_1)\} \\ \vdots & \vdots \\ -\Im\{\mathrm{rvec}(\overline{\mathbf{B}}_{N_Q})\}, & \Re\{\mathrm{rvec}(\overline{\mathbf{B}}_{N_Q})\} \end{bmatrix}, \\
\widetilde{\mathbf{H}} &= \begin{bmatrix} \mathbf{I}_{N_P} \otimes \Re(\mathbf{H}) & \mathbf{I}_{N_P} \otimes \Im(\mathbf{H}) \\ -\mathbf{I}_{N_P} \otimes \Im(\mathbf{H}) & \mathbf{I}_{N_P} \otimes \Re(\mathbf{H}) \end{bmatrix}, \\
\widetilde{\mathbf{V}} &= [\Re\{\mathrm{rvec}(\mathbf{V})\}, \Im\{\mathrm{rvec}(\mathbf{V})\}].
\end{aligned} \tag{89}$$

The equivalent dispersion matrix $\widetilde{\chi}$ is known to both the transmitter and receiver. According to the transmit power constraint, we always have $\mathrm{tr}(\widetilde{\chi}^T\widetilde{\chi}) = 2N_P$.

It can be seen in (88) that the equivalent LDC received signal model is the same as the V-BLAST received signal model of (1), where the LDC's equivalent fading channels matrix is given by $\widetilde{\chi} \cdot \widetilde{\mathbf{H}}$. Therefore, the LDC receiver may invoke the hard/soft-decision V-BLAST detectors introduced in Sec. II-A. It is worth noting that there is a total of $I = M^{N_Q}$ combinations for the LDC codeword of $\{\widetilde{\mathbf{S}}^i\}_{i=0}^{I-1}$, when the V-BLAST detectors are invoked.

According to the LDC's input-output relationship of (88), the CCMC capacity of the LDC is given by:

$$\begin{aligned}
C_{LDC}^{CCMC}(SNR) &= \max_{p(\widetilde{\mathbf{S}})} \frac{1}{2N_P} H(\widetilde{\mathbf{Y}}) - \frac{1}{2N_P} H(\widetilde{\mathbf{Y}}|\widetilde{\mathbf{S}}) \\
&= \frac{1}{2N_P} \mathrm{E}\left[\log_2 \det\left(\mathbf{I}_{2N_PN_R} + \eta\widetilde{\mathbf{H}}^T\widetilde{\chi}^T\widetilde{\chi}\widetilde{\mathbf{H}}\right)\right],
\end{aligned} \tag{90}$$

where the entropies are given by $H(\widetilde{\mathbf{Y}}) = \log_2 \det\left(\frac{\pi e}{2}\widetilde{\mathbf{H}}^T\widetilde{\chi}^T\widetilde{\chi}\widetilde{\mathbf{H}} + \frac{\pi eN_0}{2}\mathbf{I}_{2N_PN_R}\right)$ and $H(\widetilde{\mathbf{Y}}|\widetilde{\mathbf{S}}) = H(\widetilde{\mathbf{V}}) = \log_2 \det\left(\frac{\pi eN_0}{2}\mathbf{I}_{2N_PN_R}\right)$. We note that the CCMC capacity of virtually all MIMO schemes, whose transmission matrix may be expressed in the form of (87), may be evaluated by (90). Obviously, when $\widetilde{\chi}$ is a scaled unitary matrix formulated as:

$$\widetilde{\chi}^T\widetilde{\chi} = \frac{1}{N_T}\mathbf{I}_{2N_TN_P}, \tag{91}$$

the CCMC capacity of the LDC in (90) may achieve its highest possible value of:

$$C_{LDC}^{CCMC}(SNR) = \frac{1}{2N_P} \mathrm{E}\left[\log_2 \det\left(\mathbf{I}_{2N_PN_R} + \frac{\eta}{N_T}\widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}}\right)\right]. \tag{92}$$

$$\widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}} = \begin{bmatrix} \mathbf{I}_{N_P} \otimes \left[\Re(\mathbf{H}^T)\Re(\mathbf{H}) + \Im(\mathbf{H}^T)\Im(\mathbf{H})\right] & \mathbf{I}_{N_P} \otimes \left[\Re(\mathbf{H}^T)\Im(\mathbf{H}) - \Im(\mathbf{H}^T)\Re(\mathbf{H})\right] \\ \mathbf{I}_{N_P} \otimes \left[\Im(\mathbf{H}^T)\Re(\mathbf{H}) - \Re(\mathbf{H}^T)\Im(\mathbf{H})\right] & \mathbf{I}_{N_P} \otimes \left[\Re(\mathbf{H}^T)\Re(\mathbf{H}) + \Im(\mathbf{H}^T)\Im(\mathbf{H})\right] \end{bmatrix}. \tag{93}$$

Furthermore, it may be readily seen that the LDC's capacity of (92) may achieve the maximum MIMO capacity of (4), if and only if we have $\widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}} = \mathbf{I}_{2N_P} \otimes (\mathbf{H}^H\mathbf{H})$. Unfortunately, this is only true when a single RA $N_R = 1$ is used. This is because the term of $\widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}}$ in (92) may be extended as (93), as shown at the top of this page, which only becomes equal to $\mathbf{I}_{2N_P} \otimes (\mathbf{H}^H\mathbf{H})$, when we have $\Re(\mathbf{H}^T)\Im(\mathbf{H}) = \Im(\mathbf{H}^T)\Re(\mathbf{H})$ for $N_R = 1$. In summary, the relationship between the LDC capacity of (92) and the MIMO capacity of (4) may be expressed as $C_{LDC}^{CCMC}(SNR) \leq C_{MIMO}^{CCMC}(SNR)$, where the equality only holds for $N_R = 1$.

Nonetheless, the LDC capacity is expected to be higher than STBC capacity summarized in Sec. II-B3. Considering Alamouti's classic G2-STBC as an example, according to (83), the equivalent dispersion matrix $\widetilde{\chi}$ is given by:

$$\widetilde{\chi} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}, \tag{94}$$

and it may be readily seen that it does not have orthogonal columns, since we have $\widetilde{\chi}^T\widetilde{\chi} \neq \frac{1}{2}\mathbf{I}_8$. We also note that it is straightforward to prove that the capacity of Alamouti's G2-STBC evaluated by (90) based on the equivalent LDC dispersion matrix of (94) is exactly the same as that calculated by (81).

In fact, in order to guarantee that the LDC's equivalent dispersion matrix $\widetilde{\chi}$ has orthogonal columns as specified by (91), we may have $N_Q \geq N_T N_P$. Considering that further increasing N_Q will inevitably reduce the codewords' difference $\|\mathbf{S}^i - \mathbf{S}^{\bar{i}}\|$, which degrades the PEP of (8), the LDC design is suggested to satisfy $N_Q = N_T N_P$. Owing to the earlier suggestion of $N_T = N_P$ recommended owing to its diversity and delay benefits, the LDC may achieve the throughput of $R = \frac{N_Q}{N_P}\text{BPS} = N_T\text{BPS}$, which is exactly the same as the V-BLAST throughput.

In summary, according to (91), the LDC dispersion matrix $\widetilde{\chi}$ may be randomly generated as a $(2N_Q \times 2N_Q)$-element unitary matrix scaled by $\frac{1}{\sqrt{N_T}}$, so that the CCMC capacity is maximized. Moreover, in order to also retain the maximum attainable diversity order, the randomly generated dispersion matrix should have a full rank for all $\mathbf{\Delta} = (\mathbf{S}^i - \mathbf{S}^{\bar{i}})^H(\mathbf{S}^i - \mathbf{S}^{\bar{i}})$ in (8). Since it is also important to minimize the PEP according to the determinant criterion of (8), the optimum LDC dispersion matrix chosen from random search should satisfy:

$$\max \{\det(\mathbf{\Delta})\}_{\min}, \tag{95}$$

where $\{\det(\mathbf{\Delta})\}_{\min}$ is the minimum determinant $\det(\mathbf{\Delta})$ among all legitimate $\mathbf{\Delta}$ values for a randomly generated $\widetilde{\chi}$. Further developments on LDC codeword generation may be found in [128]–[132], which also tackle the problem of having a diminishing distance between legitimate codewords, when aiming for the high-throughput LDC codeword generation. Nonetheless, the random generation according to the above design guidelines is sufficiently effective for producing good LDCs that achieve both a multiplexing gain and a diversity gain.

3) CAPACITY-ACHIEVING LDCs

In order to achieve the maximum attainable MIMO capacity, it was proposed in [38] that the LDC's dispersion matrices in (87) should satisfy $\{\overline{\mathbf{A}}_q = \overline{\mathbf{B}}_q\}_{q=1}^{N_Q}$, just like the V-BLAST characterized in (73), so that the real and imaginary parts of the modulated MPSK/QAM symbols may be dispersed together as:

$$\mathbf{S} = \sum_{q=1}^{N_Q} \left[\overline{\mathbf{A}}_q s_q\right]. \tag{96}$$

In this way, the real and the imaginary parts of the received signal model do not have to be decoupled, as seen in (88). Instead, vectorizing the received MIMO signal matrix $\mathbf{Y}$ of (1) leads to the new received LDC signal model of:

$$\overline{\mathbf{Y}} = \overline{\mathbf{S}} \cdot \overline{\chi} \cdot \overline{\mathbf{H}} + \overline{\mathbf{V}}, \tag{97}$$

where the matrices are given by:

$$\overline{\chi} = \begin{bmatrix} \text{rvec}(\overline{\mathbf{A}}_1) \\ \vdots \\ \text{rvec}(\overline{\mathbf{A}}_{N_Q}) \end{bmatrix}, \quad \begin{aligned} &\overline{\mathbf{Y}} = \text{rvec}(\mathbf{Y}), \ \ \overline{\mathbf{S}} = [s_1, \cdots, s_{N_Q}], \\ &\overline{\mathbf{H}} = \mathbf{I}_{N_P} \otimes \mathbf{H}, \ \ \overline{\mathbf{V}} = \text{rvec}(\mathbf{V}). \end{aligned} \tag{98}$$

It can be seen that the new LDC's received signal model of (97) is equivalent to that of an V-BLAST system equipped with N_Q TAs and $N_R N_P$ RAs. Therefore, all the hard/soft-decision V-BLAST detectors introduced in Sec. II-A may be invoked for LDC detection, where the equivalent multiplexed transmitted symbol vector and the fading channel matrix of the V-BLAST system are given by $\frac{1}{\sqrt{N_Q}}\overline{\mathbf{S}}$ and $\sqrt{N_Q}\overline{\chi}\overline{\mathbf{H}}$, respectively. There is a total of $I = M^{N_Q}$ combinations for the LDC codeword of $\{\widetilde{\mathbf{S}}^i\}_{i=0}^{I-1}$, when the V-BLAST detectors are invoked. In summary, the LDC transceiver is summarized in the schematic diagram of Fig. 11.

According to the new input-output relationship of (97), the CCMC capacity of the LDC model of (96) is given by:

$$C_{LDC}^{CCMC}(SNR) = \max_{p(\overline{\mathbf{S}})} \frac{1}{N_P} H(\overline{\mathbf{Y}}) - \frac{1}{N_P} H(\overline{\mathbf{Y}}|\overline{\mathbf{S}}) = \frac{1}{N_P} \mathrm{E}\left[\log_2 \det\left(\mathbf{I}_{N_P N_R} + \eta \overline{\mathbf{H}}^H \overline{\chi}^H \overline{\chi} \overline{\mathbf{H}}\right)\right], \tag{99}$$

where the related entropies are given by $H(\overline{\mathbf{Y}}) = \log_2 \det\left(\pi e \overline{\mathbf{H}}^H \overline{\chi}^H \overline{\chi} \overline{\mathbf{H}} + \pi e N_0 \mathbf{I}_{N_P N_R}\right)$ and $H(\overline{\mathbf{Y}}|\overline{\mathbf{S}}) = H(\overline{\mathbf{V}}) = \log_2 \det\left(\pi e N_0 \mathbf{I}_{N_P N_R}\right)$. It can be seen in (99) that the CCMC capacity is maximized when the equivalent dispersion matrix $\overline{\chi}$ has orthogonal columns as represented by:

$$\overline{\chi}^H \overline{\chi} = \frac{1}{N_T} \mathbf{I}_{N_T N_P}, \tag{100}$$

which is scaled according to the power constraint of $\mathrm{E}\left[\mathrm{tr}(\mathbf{S}^H \mathbf{S})\right] = N_P$. As a result, the CCMC capacity of (99) becomes:

$$C_{LDC}^{CCMC}(SNR) = \frac{1}{N_P} \mathrm{E}\{\log_2 \det[\mathbf{I}_{N_P N_R} + \frac{\eta}{N_T} (\mathbf{I}_{N_P} \otimes \mathbf{H})^H (\mathbf{I}_{N_P} \otimes \mathbf{H})]\} = \mathrm{E}\left[\log_2 \det(\mathbf{I}_{N_R} + \frac{\eta}{N_T} \mathbf{H}^H \mathbf{H})\right], \tag{101}$$

which is exactly the same as the full MIMO capacity of (4). Therefore, in order to avoid any ambiguity, the terminology of LDCs may generally refer to the capacity-achieving model of (96), rather than to the conventional model of (87).

We note that (100) requires $N_Q \geq N_T N_P$. Hence, for the case of full transmit diversity associated with $N_T = N_P$, $N_Q = N_T N_P$ leads to the LDC throughput being the same as the V-BLAST throughput of $R = N_T$BPS. Similar to the discussions in Sec. II-C2, the generation of LDCs may follow the guidelines of maximizing the CCMC capacity of (99) and of minimizing the PEP of (8), which may be summarized as:

It is worth emphasizing once again that the LDC's CCMC capacity is only maximized when $N_Q \geq N_T N_P$. Nonetheless, $N_Q < N_T N_P$ is acceptable in Step (1) for the sake of meeting specific system requirements, because a lower number of transmitted symbols N_Q normally leads to a higher Euclidean distance among the LDC codewords $\|\mathbf{S}^i - \mathbf{S}^{\bar{i}}\|^2$, which may minimize the PEP union bound of (8).

Furthermore, according to Hadamard's inequality, the determinant $\det(\mathbf{\Delta})$ is maximized when $\mathbf{\Delta}$ is unitary, which is the foundation of the orthogonal STBC design. It was proposed in [37] that the determinant criterion in the LDC design may be translated into making $\mathbf{\Delta}$ as close to unitary as possible, which may be quantified as minimizing the following two metrics:

$$d_1 = \sum_{q=1}^{N_Q} \kappa(\overline{\mathbf{A}}_q) = \sum_{q=1}^{N_Q} \|\overline{\mathbf{A}}_q^{-1}\| \cdot \|\overline{\mathbf{A}}_q\|, \tag{102a}$$

$$d_2 = \sum_{\forall q \neq \bar{q}} \|\overline{\mathbf{A}}_q^H \overline{\mathbf{A}}_{\bar{q}} + \overline{\mathbf{A}}_{\bar{q}}^H \overline{\mathbf{A}}_q\|, \tag{102b}$$

where the operation $\kappa(\cdot)$ refers to the condition number of the matrix [236], where we have $\kappa(\mathbf{A}) \geq 1$ and the equality only holds for unitary matrices. It can be readily seen that orthogonal codes may have $d_1 = N_Q$ and $d_2 = 0$. Moreover, it was also proposed in [12] and [247] that the determinant criterion of $\max\ \{\det(\mathbf{\Delta})\}_{\min}$ in the LDC design may be revised for the sake of maximizing the LDC's DCMC capacity of (6) in order to pursue an improved near-capacity performance. In fact, minimizing the PEP $\mathrm{E}\left\{p\left(\|\mathbf{Y} - \mathbf{S}^{\bar{i}}\mathbf{H}\|^2 < \|\mathbf{V}\|^2\right)\right\} = \mathrm{E}\left\{p\left[\|(\mathbf{S}^i - \mathbf{S}^{\bar{i}})\mathbf{H} + \mathbf{V}\|^2 < \|\mathbf{V}\|^2\right]\right\}$ of (8) would automatically result in minimizing the term $\exp(\frac{-\|(\mathbf{S}^i - \mathbf{S}^{\bar{i}})\mathbf{H} + \mathbf{V}\|^2 + \|\mathbf{V}\|^2}{N_0})$ in the DCMC capacity of (6). Consequently, the LDCs conceived for minimizing the PEP generally also have a maximized DCMC capacity.

LDC generation guidelines

(1) Randomly generate a unitary matrix χ of size $(N \times N)$, where we have $N = \max(N_Q, N_T N_P)$.
   a) If $N_Q > N_T N_P$ is required, the LDC dispersion matrix is given by taking the first $N_T N_P$ columns of the scaled unitary matrix as $\overline{\chi} = \frac{1}{\sqrt{N_T}} \chi \begin{bmatrix} \mathbf{I}_{N_T N_P} \\ \mathbf{0} \end{bmatrix}$, where $\mathbf{0}$ is a $(N_Q - N_T N_P) \times N_T N_P$-element all-zero matrix.
   b) If $N_Q = N_T N_P$ is required, the LDC's dispersion matrix is directly given by $\overline{\chi} = \frac{1}{\sqrt{N_T}} \chi$.
   c) If $N_Q < N_T N_P$ is required, the LDC's dispersion matrix is given by taking the first N_Q rows of the scaled unitary matrix as $\overline{\chi} = \sqrt{\frac{N_P}{N_Q}} \left[\mathbf{I}_{N_Q}, \mathbf{0}\right] \chi$, where $\mathbf{0}$ is a $N_Q \times (N_T N_P - N_Q)$-element all-zero matrix.

(2) Rank criterion: for the resultant $I = M^{N_Q}$ LDC codewords $\{\mathbf{S}^i\}_{i=0}^{I-1}$ of (96), having a full rank should be guaranteed for all combinations of $\mathbf{\Delta} = (\mathbf{S}^i - \mathbf{S}^{\bar{i}})^H (\mathbf{S}^i - \mathbf{S}^{\bar{i}})$ as $\mathrm{rank}(\mathbf{\Delta}) = \min(N_T, N_P)$.

(3) Determinant criterion: The minimum determinant among all combinations of $\mathbf{\Delta}$ is given by $\{\det(\mathbf{\Delta})\}_{\min}$. The related random search may be conducted by repeating Steps (1) as well as (2), and the chosen one should maximize $\{\det(\mathbf{\Delta})\}_{\min}$.

D. CAPACITY AND BER COMPARISON BETWEEN CLASSIC MIMO SCHEMES

Fig. 24 presents the capacity comparison between V-BLAST, STBC and LDC. It can be seen in Fig. 24(a) that both V-BLAST and LDC achieve the highest MIMO capacity, as analysed in Sec. II-A and Sec. II-C3, respectively, but Alamouti's G2-STBC associated with $N_R = 2$ can only achieve the capacity of another V-BLAST system that is associated with $N_T = 4$ and $N_R = 1$ having a doubled SNR of 2η, which was explicitly discussed in Sec. II-B3. The

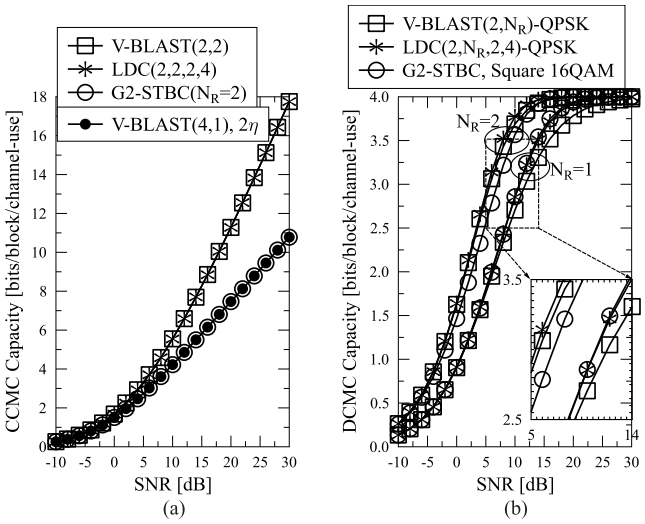


FIGURE 24. Capacity comparison between V-BLAST, Alamouti's G2-STBC and LDC, where $N_T = 2$ TAs are employed and the throughput is given by $R = 4$ bits/block/channel-use. (a) CCMC Capacity. (b) DCMC Capacity.

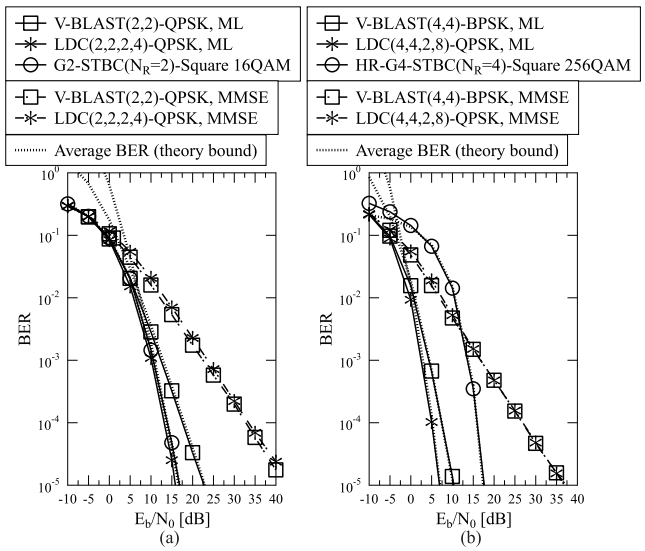


FIGURE 25. Performance comparison between V-BLAST, STBC and LDC associated with the same throughput of $R = 4$ bits/block/channel-use. (a) $N_T = 2$, $N_R = 2$. (b) $N_T = 4$, $N_R = 4$.

MIMO's DCMC capacity of (6) often predicts the achievable performance. For this spirit, it can be seen in Fig. 24(b) that Alamout's G2-STBC and LDC achieve their full DCMC capacity quantified in terms of bits/block/channel-use at a lower SNR than V-BLAST for the case of $N_R = 1$, which confirms the beneficial transmit diversity gain of both STBC and LDC. However, when $N_R = 2$ RAs are used, Alamouti's G2 STBC exhibits a lower DCMC capacity in the low SNR region, as evidenced in Fig. 24(b). We will augment the reasons for this feature later.

Fig. 25 portrays the performance comparison between V-BLAST, STBC and LDC associated with the same throughput of $R = 4$. It is evidenced by Fig. 25(a) that both LDC(2,2,2,4)-QPSK and Alamouti's G2-STBC ($N_R = 2$) employing Square 16QAM significantly outperform their multiplexing-oriented counterpart of V-BLAST(2,2)-QPSK, especially in the high SNR region. Furthermore, Fig. 25(a) demonstrates that LDC(2,2,2,4)-QPSK performs even slightly better than its STBC counterpart. Fig. 25(b) also shows that LDC(4,4,2,8)-QPSK is capable of outperforming both its multiplexing-oriented counterpart of V-BLAST(4,4)-BPSK and its STBC counterpart of HR-G4-STBC ($N_R = 4$) employing Square 256QAM for the case of $N_T = 4$.

However, it is also demonstrated by Fig. 25(b) that although HR-G4-STBC retains its full diversity order, its performance remains modest, unless the SNR is expected to be extremely high. This is because G4-STBC has a low normalized throughput of $\overline{R} = 0.5$, which requires us to employ a high-order 256QAM scheme in order to achieve the required system throughput. The orthogonal STBC design aims for achieving the lowest error probability at high SNRs, when the determinant term $\det(0.25\eta\Delta)$ dominants the divisor of the PEP in (8). Since Δ is unitary as guaranteed by the STBC, recall from Sec. II-B3 that the number of modulation levels M is the only factor that affects the error probability in the low-SNR region, owing to the fact that the error probability of the STBC in fading channels is given by the performance of MPSK/QAM schemes in AWGN channels associated with the equivalent noise power of $\overline{N}_0 = \frac{N_T N_Q}{N_P \|\mathbf{H}\|^2} N_0$. This is also the reason why Alamouti's G2-STBC employing a higher-order 16QAM scheme associated with $N_R = 2$ cannot achieve the best DCMC capacity in the low-SNR region of Fig. (24(b)).

In summary, the LDC was shown in Fig. 24 to be able to achieve the V-BLAST's full MIMO capacity, and it is also capable of retaining the STBC's full diversity gain, hence offering the best performance, as shown in Fig. 25. Therefore, the LDC may resolves the tradeoff between the multiplexing and diversity gain in MIMO systems design, provided that the conditions of (100) as well as the rank and determinant criteria presented in Sec. II-C3 are satisfied.

Nonetheless, as demonstrated in Sec. II-C3, the LDC receivers have to employ the V-BLAST detectors, where the performance versus complexity tradeoff portrayed in Fig. 2 once again emerges. It can be seen in Fig. 25 that without the interference cancellation techniques of [32], [202]–[206], the low-complexity hard-decision MMSE detector imposes a substantial performance loss on both the uncoded V-BLAST and LDC schemes. Furthermore, it was demonstrated by Figs. 21 and 22 that without taking into account the *a priori* information in the MSE, the soft-decision MMSE detector also imposes a significant performance loss on coded V-BLAST systems. Moreover, the hard/soft-decision SD and MMSE aided V-BLAST detectors introduced in Sec. II-A are explicitly designed for uplink MIMO systems associated with $N_T \leq N_R$. For the rank-deficient MIMO systems associated with $N_T > N_R$, which are often encountered in realistic wireless communication systems, the SD aided V-BLAST is recommended for detecting N_R symbols, while the ML detector is invoked for the remaining symbols [178]–[180]. This complication may be avoided by using LDC instead of V-BLAST

as discussed in Sec. II-C3, where we have the design-freedom to adjust the LDC arguments of N_P and N_Q in order to create an equivalent uplink MIMO system. However, considering that the LDC's dispersion matrices are randomly populated, the LDC transmitter may be required to transmit symbols that are not drawn from the classic MPSK/QAM constellations, which further complicates the hardware design of the related MIMO systems. In order to overcome these limitations of the conventional MIMO systems design, the recently developed MIMO schemes of SM and STSK, which are inspired by stricking an attractive performance-complexity tradeoff are introduced in the following section.

III. THE NEWLY-DEVELOPED MIMO SCHEMES THAT ARE MOTIVATED BY THE PERFORMANCE-COMPLEXITY TRADEOFF

The development of LDCs has resolved the tradeoff between the diversity and multiplexing gain, but it is a retrograde step for the tradeoff between performance and complexity. Given that the STBC's orthogonality requirement is abandoned, the LDC receivers have to invoke V-BLAST-style multi-stream detectors, which may exhibit an excessive complexity, when aiming for attaining an optimal performance. Considering that the family of suboptimal V-BLAST detectors would not be deemed desirable, especially not in coded systems, because they tend to produce unreliable soft output LLRs that do not represent the true probabilities, as demonstrated in Sec. II-A9. In this section, we focus our attention on the SM and STSK families, which open a new chapter in the design of MIMO systems that is explicitly motivated by striking a compelling performance versus complexity tradeoff.

A. SPATIAL MODULATION (SM)

The schematic of the SM transmitter is portrayed in Fig. 12. In more details, the first $\text{BPS} = \log_2 M$ bits are assigned to a single MPSK/QAM symbol $s^m = \mathbb{M}(m)$, while the following $\text{BPS}_T = \log_2 N_T$ source information bits are assigned to activate a single TA out of a total of N_T TAs. As a result, the N_T-element SM transmission row-vector is expressed as [26], [27], [133]:

$$\mathbf{S} = [\underbrace{0 \cdots 0}_{v-1}, s^m, \underbrace{0 \cdots 0}_{N_T - v}]. \tag{103}$$

Based on the received MIMO signal model of (1), the full-search hard-decision ML MIMO detection of (11) and the soft-decision MAP MIMO detection using (14) may also be invoked for SM. However, as it was demonstrated in Secs. II-A1 and II-A2, the ML/MAP aided MIMO detection complexity may increase exponentially with the throughput R. More explicitly, the complexity order of the hard-decision ML aided MIMO detection of (11) and the soft-decision MAP MIMO detection using (14) is given by $O(I)$, where the total number of combinations is given by $I = 2^R$ for both V-BLAST and SM.

Owing to the fact that only a single TA is activated, opposed to V-BLAST, SM does not introduce any IAI. Therefore, in order to conceive a single-antenna-based low complexity SM detector, the TA activation index and the classic modulated symbol index are suggested to be detected separately in [27], so that the complexity order of this so-called hard-decision Maximum Ratio Combining (MRC) based SM detection may be reduced to $O(N_T + M)$. In more details, under the assumption of having perfect CSI knowledge at the receiver, the matched filter output may be recorded as:

$$\mathbf{Z} = \mathbf{Y}\mathbf{H}^H, \tag{104}$$

where the v-th element in the N_T-element row-vector $\mathbf{Z}$ is given by $\{z_v = \mathbf{Y}\mathbf{H}^H_{v,-}\}_{v=1}^{N_T}$, and the N_R-element row-vector $\{\mathbf{H}_{v,-}\}_{v=1}^{N_T}$ refers to the v-th row in $\mathbf{H}$. The hard-decision MRC based SM detector may determine the TA activation index by comparing the absolute values of the elements in the matched filter's output vector $\mathbf{Z}$ as [27]:

$$\hat{v} = \arg \max_{\forall v \in \{1,\cdots,N_T\}} |z_v|. \tag{105}$$

Upon obtaining the TA activation index $\hat{v}$, the $\hat{v}$-th element in $\mathbf{Z}$ may be demodulated as:

$$\hat{m} = \mathbb{M}^{-1}(z_{\hat{v}}). \tag{106}$$

Therefore, the complexity order of MRC based SM detection is in fact given by $O(N_T + 1)$, where (106) directly maps $z_{\hat{v}}$ to the closest constellation point.

Unfortunately, as demonstrated in [134], the hard-decision MRC based SM detection suffers from an irreducible error floor. It can be seen in (106) that the demodulator may be misled into detecting the wrong classic modulated symbol, if the TA activation index obtained in (105) is erroneous. In order to restore the ML detection capability, the hard-decision simplified ML aided SM detector of [134] streamlines the hard-decision ML MIMO detector of (11) as:

$$\begin{aligned}\hat{\mathbf{S}} &= \arg \min_{\forall m \in \{0,\cdots,M-1\}, \forall v \in \{1,\cdots,N_T\}} \|\mathbf{Y}\|^2 + \kappa_v^2 |s^m|^2 \\ &\quad - 2\Re\left[(s^m)^* \mathbf{Y}\mathbf{H}^H_{v,-}\right] \\ &= \arg \min_{\forall m \in \{0,\cdots,M-1\}, \forall v \in \{1,\cdots,N_T\}} \kappa_v^2 |s^m|^2 - 2\Re\left[(s^m)^* z_v\right],\end{aligned} \tag{107}$$

where we have $\{\kappa_v = \|\mathbf{H}_{v,-}\|\}_{v=1}^{N_T}$, and the constant of $\|\mathbf{Y}\|^2$ is omitted from the MIMO decision metric of (11). As a benefit of having $(N_T - 1)$ zeros in the SM transmission vector seen in Table 4, the computational complexity of the SM detection of (107) is considerably lower than that of the conventional MIMO detection of (11). Nonetheless, the complexity order of the hard-decision simplified ML aided SM detection of (107) is still given by $O(I)$.

When SM was first proposed as an alternative to V-BLAST MIMO, the two most appealing features of SM were its low hardware transmitter complexity as well as its design objective of imposing a low receiver signal processing complexity. Therefore, in this section, we focus our attention on the strategically important subject of reduced-complexity SM detectors. The hard-decision SM detector design has

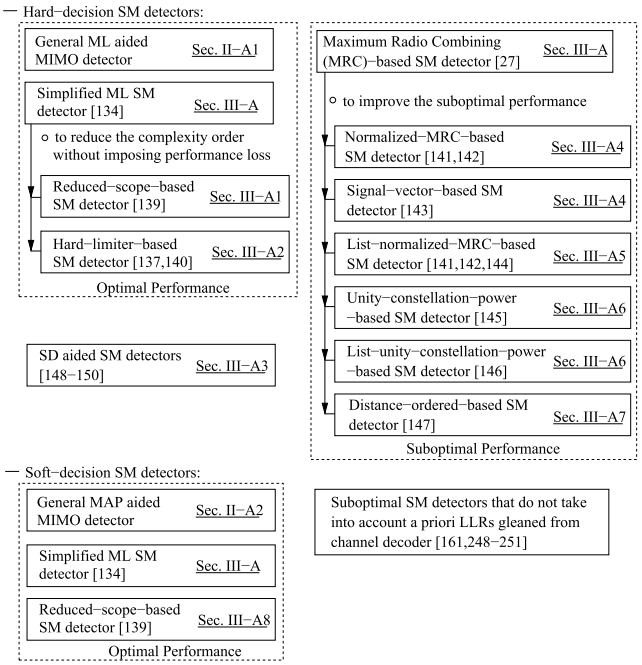


FIGURE 26. Summary of hard/soft-decision optimal/suboptimal SM detectors with their references and their section numbers in this paper.

been developed in two main directions in the open literature, as portrayed in Fig. 26. The first option is to develop the optimal SM detection [137]–[140] that endeavours to reduce the complexity order of the simplified ML aided SM detection of (107) without imposing any performance loss. The second approach elaborated on in [141]–[147] aims for improving the performance of the sub-optimal MRC-based SM detection of (105) and (106), but attaining the optimal SM performance is not guaranteed. Moreover, the SD was also developed for SM in [148]–[150], which exhibits a reduced complexity compared to the SD aided by V-BLAST.

For coded SM schemes, instead of using the general MAP aided MIMO detector introduced in Sec. II-A2, the simplified hard-decision ML aided SM detector of [134] may be readily revised to the simplified soft-decision MAP aided SM detector. More explicitly, based on (107), the probability metric of (14) invoked by the general MAP aided MIMO detectors may be simplified for SM as:

$$d^i = -\frac{\kappa_v^2|s^m|^2 - 2\Re\left[(s^m)^* z_v\right]}{N_0} + \sum_{\bar{k}=1}^{\log_2 I} \widetilde{b}_{\bar{k}} L_a(b_{\bar{k}}). \quad (108)$$

The relationship between the SM index i, the TA activation index v and the classic modulated symbol index m is given by $i = v - 1 + mN_T$ according to the SM transmitter design. The only difference between (14) and (108) is a constant of $-\frac{\|\mathbf{Y}\|^2}{N_0}$, which may be eliminated by the division operation of the Log-MAP of (12). Therefore, all general MIMO's detection algorithms including Log-MAP of (12), Max-Log-MAP of (15) and Approx-Log-MAP of (16) may invoke the probability metric of (108) instead of (14) for SM detection without imposing any performance loss.

Similarly, the hard-decision reduced-scope-based SM detector may also be revised for coded SM systems as suggested in [139], where the SM TA activation index v and the modulated symbol index m are detected separately, while the correlation between the two terms is also taken into account in order to retain the optimal performance. In other words, only a reduced subset of the SM combinations $\{\mathbf{S}^i\}_{i=0}^{N_T M-1}$ has to be examined by the reduced-scope-based SM detector. However, the soft-decision version of the hard-limiter-based SM detector of [137], [140] is the same as the simplified soft-decision MAP-aided SM detector using (108). More explicitly, the substantial complexity reduction provided by the hard-limiter-based SM detection of [137], [140] relies on the low-complexity implementation of hard-decision MPSK/QAM demodulators, where a certain decision variable obtained from the matched filter output may be directly demapped to the nearest constellation point, which is similar to the feature portrayed by Fig. 16(b). However, when the *a priori* LLRs gleaned from the channel decoder are also taken into account in coded SM systems, both the channel's output signal as well as the *a priori* LLRs have to be transformed back into modulated symbols according to the constellation diagram. As a result, the SM TA activation index v and the modulated symbol index m once again have to be jointly detected according to all SM combinations $\{\mathbf{S}^i\}_{i=0}^{N_T M-1}$, which results in the simplified soft-decision MAP-aided SM detector using (108).

Furthermore, the sub-optimal hard-decision SM detectors [141]–[147] are not recommended for employment in coded SM systems. This is because these sub-optimal SM detectors may falsify the reliability of the output LLRs, which may fail to reflect the true *a posteriori* probabilities by producing LLRs having excessively high values. This flawed situation cannot be readily rectified by the channel decoder, as discussed in Sec. II-A9. The sub-optimal soft-decision SM detectors may also be found in [161] and [248]–[251], where the beneficial *a priori* information is not exploited by the SM detectors.

1) HARD-DECISION REDUCED-SCOPE-BASED OPTIMAL SM DETECTION

The reduced-scope-based SM detection [139] aims for restoring the ML detection capability of the MRC-based SM detection by separating the TA index and the classic modulated symbol index from the optimal SM detection of (107) without imposing any performance loss. First of all, the optimal SM detection of (107) is extended as:

$$\hat{\mathbf{S}} = \max_{\forall v\in\{1,\cdots,N_T\},\forall m\in\{0,\cdots,M-1\}} \Re(\widetilde{z}_v)\Re(s^m) + \Im(\widetilde{z}_v)\Im(s^m) - \kappa_v^2|s^m|^2, \quad (109)$$

where we have $\{\widetilde{z}_v = 2z_v\}_{v=1}^{N_T}$. Let us now consider QPSK aided SM detection as an example. The QPSK's detected constellation diagram is deliberately rotated anti-clockwisely by $\pi/4$, so that there is only a single constellation point in each quadrant. As a result, the decision variable should be

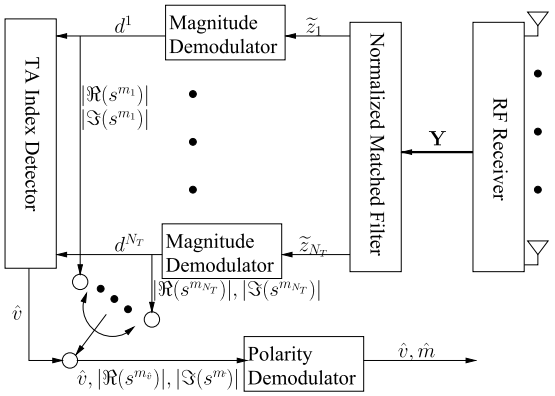


FIGURE 27. Schematic of the hard-decision reduced-scope-based SM receiver.

rotated as $z'_v = \widetilde{z}_v \exp(j\frac{\pi}{4})$, and the detected constellation points are given by $\{s'^m = s^m \exp(j\frac{\pi}{4})\}_{m=0}^{M-1} = \{\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}\}$. For a specific TA index v, the maximum metric over all rotated QPSK constellations is given by:

$$d^v = \max \left\{ \begin{matrix} \frac{\Re(z'_v)}{\sqrt{2}} + \frac{\Im(z'_v)}{\sqrt{2}} - \kappa_v^2, & \frac{\Re(z'_v)}{\sqrt{2}} - \frac{\Im(z'_v)}{\sqrt{2}} - \kappa_v^2, \\ -\frac{\Re(z'_v)}{\sqrt{2}} + \frac{\Im(z'_v)}{\sqrt{2}} - \kappa_v^2, & -\frac{\Re(z'_v)}{\sqrt{2}} - \frac{\Im(z'_v)}{\sqrt{2}} - \kappa_v^2 \end{matrix} \right\} \tag{110a}$$

$$= \left|\frac{\Re(z'_v)}{\sqrt{2}}\right| + \left|\frac{\Im(z'_v)}{\sqrt{2}}\right| - \kappa_v^2. \tag{110b}$$

It can be seen in (110) that

the four comparisons involving four metric evaluations of (110a) may be carried out by a single metric evaluation according to (110b). As a result, the optimum TA activation index $\hat{v}$ may be found by searching for the maximum metric over all the N_T candidates $\{d^v\}_{v=1}^{N_T}$, regardless of which particular QPSK symbol was transmitted. This may be expressed as:

$$\hat{v} = \arg \max_{\forall v \in \{1,\cdots,N_T\}} d^v. \tag{111}$$

Unlike the MRC-based detection of (105), the reduced-scope-based TA index detection of (111) is directly derived from the ML detection of (109), which does not impose any performance loss. Furthermore, upon finding the optimum TA index $\hat{v}$, QPSK demodulation may be concluded by directly testing the $\hat{v}$-th decision variable $z'_{\hat{v}}$ as:

$$\hat{b}_1 = \begin{cases} 1, & \text{if } \Im(z'_{\hat{v}}) < 0 \\ 0, & \text{otherwise,} \end{cases} \quad \hat{b}_2 = \begin{cases} 1, & \text{if } \Re(z'_{\hat{v}}) < 0 \\ 0, & \text{otherwise.} \end{cases} \tag{112}$$

The schematic of the general hard-decision reduced-scope-based SM receiver is portrayed by Fig. 27. More explicitly, its design guideline [139] is breifly summarized as follows:

Algorithm 2: Design guidelines for reduced-scope-based hard-decision SM detection

1) First of all, the N_T normalized matched filter outputs are given by $\{\widetilde{z}_v = 2z_v\}_{v=1}^{N_T}$, where we have $\{z_v = \mathbf{Y}\mathbf{H}_{v,-}^H\}_{v=1}^{N_T}$ according to (104).
2) Secondly, similar to the reduced-complexity design introduced in Sec. II-A8 and exemplified by Fig. 18, only the real PAM magnitudes or the constellation points located in the first quadrant are visited by the "Magnitude Demodulator" of Fig. 27, when either Square QAM or general PSK/QAM is employed, respectively. As a result, the local maximum metrics $\{d^v\}_{v=1}^{N_T}$ associated with N_T TA activation index candidates may be obtained by a reduced SM detection search scope.
3) Thirdly, the decision metrics $\{d^v\}_{v=1}^{N_T}$ are compared by the "TA Index Detector" of Fig. 27, where the detected index $\hat{v}$ is given by (111).
4) In order to detect the modulated symbol index $\hat{m}$ based on $\hat{v}$, the results of the $\hat{v}$-th "Magnitude Demodulator" are passed to the "Polarity Demodulator" of Fig. 27, which completes the PSK/QAM demodulation by determining the specific quadrant of the demodulated symbol.

For example, when Square 16QAM is employed, the local maximum metrics $\{d^v\}_{v=1}^{N_T}$ seen in (109) may be obtained by testing both the real and the imaginary parts of the QAM constellation separately, which is expressed as (113), as shown at the bottom of this page, where each one of them only has to be evaluated once. Furthermore, for a specific TA index v,

$$\begin{aligned}
d_{\text{Re}}^{v,0} &= \max \left\{ \tfrac{1}{\sqrt{10}}\Re(\widetilde{z}_v) - \tfrac{1}{10}\kappa_v^2, \quad -\tfrac{1}{\sqrt{10}}\Re(\widetilde{z}_v) - \tfrac{1}{10}\kappa_v^2 \right\} = \left|\frac{1}{\sqrt{10}}\Re(\widetilde{z}_v)\right| - \frac{1}{10}\kappa_v^2, \\
d_{\text{Re}}^{v,1} &= \max \left\{ \tfrac{3}{\sqrt{10}}\Re(\widetilde{z}_v) - \tfrac{9}{10}\kappa_v^2, \quad -\tfrac{3}{\sqrt{10}}\Re(\widetilde{z}_v) - \tfrac{9}{10}\kappa_v^2 \right\} = \left|\frac{3}{\sqrt{10}}\Re(\widetilde{z}_v)\right| - \frac{9}{10}\kappa_v^2, \\
d_{\text{Im}}^{v,0} &= \max \left\{ \tfrac{1}{\sqrt{10}}\Im(\widetilde{z}_v) - \tfrac{1}{10}\kappa_v^2, \quad -\tfrac{1}{\sqrt{10}}\Im(\widetilde{z}_v) - \tfrac{1}{10}\kappa_v^2 \right\} = \left|\frac{1}{\sqrt{10}}\Im(\widetilde{z}_v)\right| - \frac{1}{10}\kappa_v^2, \\
d_{\text{Im}}^{v,1} &= \max \left\{ \tfrac{3}{\sqrt{10}}\Im(\widetilde{z}_v) - \tfrac{9}{10}\kappa_v^2, \quad -\tfrac{3}{\sqrt{10}}\Im(\widetilde{z}_v) - \tfrac{9}{10}\kappa_v^2 \right\} = \left|\frac{3}{\sqrt{10}}\Im(\widetilde{z}_v)\right| - \frac{9}{10}\kappa_v^2,
\end{aligned} \tag{113}$$

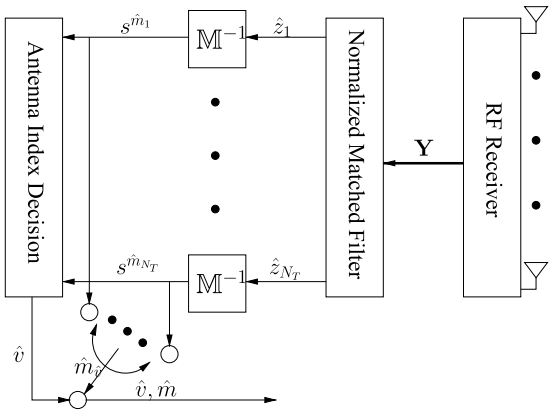


FIGURE 28. Schematic of the hard-decision hard-limiter-based SM receiver. Its difference to Fig. 27 is that full classic demodulation is performed before the TA index detection, because the hard-decision linear *M*PSK/QAM demodulation complexity is quite low in uncoded systems.

the maximum metric is given by:

$$d^v = \max_{g\in\{0,1\}} d_{\text{Re}}^{v,g} + \max_{f\in\{0,1\}} d_{\text{Im}}^{v,f}, \tag{114}$$

where the optimum PAM magnitude index pairs $\hat{g}$ and $\hat{f}$ obtained for each $\{d^v\}_{v=1}^{N_T}$ may be recorded. There are a total of N_T pairs, hence they may be represented by $\{\hat{g}_v\}_{v=1}^{N_T}$ and $\{\hat{f}_v\}_{v=1}^{N_T}$. Based on (114), the TA activation index detection of (111) may be invoked, and then the second part of the Square 16QAM demodulation may be concluded as follows:

$$\hat{b}_1 = \begin{cases} 1, & \text{if } \Im(\widetilde{z}_{\hat{v}}) < 0 \\ 0, & \text{otherwise,} \end{cases} \quad \hat{b}_2 = \begin{cases} 1, & \text{if } \hat{f}_{\hat{v}} = 1 \text{ for } d^{\hat{v}} \\ 0, & \text{otherwise,} \end{cases}$$
$$\hat{b}_3 = \begin{cases} 1, & \text{if } \Re(\widetilde{z}_{\hat{v}}) < 0 \\ 0, & \text{otherwise,} \end{cases} \quad \hat{b}_4 = \begin{cases} 1, & \text{if } \hat{g}_{\hat{v}} = 1 \text{ for } d^{\hat{v}} \\ 0, & \text{otherwise.} \end{cases} \tag{115}$$

The specific index pair $\hat{f}_{\hat{v}}$ and $\hat{g}_{\hat{v}}$ are recovered from (114). It can be readily seen that a reduced number of decision metrics are evluated in (113) according to the reduced-scope search space, and then the only comparisons that are required are those for the following steps.

In summary, the hard-decision reduced-scope optimal SM detection complexity orders of [139] are given by $O(\sqrt{M}N_T)$ and $O(MN_T/4)$, respectively, when Square QAM and general PSK/QAM are employed, respectively.

2) HARD-DECISION HARD-LIMITER-BASED OPTIMAL SM DETECTION

Due to the fact that detecting the TA index is generally much more computationally complex than the hard-decision PSK/QAM demodulation, the hard-limiter-based optimal SM detection portrayed by Fig. 28 invokes the full *M*PSK/QAM demodulators first in order to obtain the optimum modulation indices for all candidate TA indices and then the TA index detection is performed with the aid of the demodulated *M*PSK/QAM symbols. This method was first advocated in [137] and further interpreted in [140].

Let us assume that a tentative TA activation index v is fixed, and then the SM detection of (107) may be rewritten as:

$$\begin{aligned} \hat{m}_v &= \arg \min_{\forall m\in\{0,\cdots,M-1\}} (|\hat{z}_v - s^m|^2 - |\hat{z}_v|^2)\kappa_v^2 \\ &= \arg \min_{\forall m\in\{0,\cdots,M-1\}} |\hat{z}_v - s^m|^2 \\ &= \mathbb{M}^{-1}(\hat{z}_v), \end{aligned} \tag{116}$$

where the demodulator's decision variable is given by $\{\hat{z}_v = z_v/\kappa_v^2\}_{v=1}^{N_T}$. In this way, the optimum modulated symbol index $\hat{m}_v$ associated with all TA activation indices may be obtained by directly demapping $\hat{z}_v$ to the closest legitimate constellation point, which is similar to the feature portrayed by Fig. 16(b).

Upon obtaining the optimum constellation points for all candidate TA activation indices $\{s^{\hat{m}_v}\}_{v=1}^{N_T}$, the optimum TA index may be obtained based on (107) as:

$$\hat{v} = \arg \min_{\forall v\in\{1,\cdots,N_T\}} (\left|\hat{z}_v - s^{\hat{m}_v}\right|^2 - |\hat{z}_v|^2)\kappa_v^2 \tag{117}$$

and then the corresponding ($\text{BPS}_T = \log_2 N_T$) hard-bit decisions may be obtained by translating $\hat{v}$ back to binary bits. Furthermore, the remaining ($\text{BPS} = \log_2 M$) hard-bit decisions may be obtained by directly translating the specific modulated symbol index $\hat{m}_{\hat{v}}$ back to binary bits.

The hard-limiter-based optimal SM detection's complexity order is given by $O(N_T + N_T)$, where the demodulator has to be invoked N_T times before TA index detection. This detection complexity order does not grow with the number of modulation levels M, which is one of the most appealing advantages of hard-limiter-based optimal SM detection, espcially for the case of employing high-order *M*PSK/QAM schemes.

3) HARD-DECISION SD AIDED SM DETECTION

It was suggested in [148] and [149] that the conventional MIMO detector's transmit search space in (11) may be reduced by the so-called Transmitter-centric SD (Tx-SD) as:

$$\{\hat{m}, \hat{v}\} = \arg \min_{\forall\{m,v\}\in\mathbb{S}} \left\|\mathbf{Y} - s^m\mathbf{H}_{v,-}\right\|^2, \tag{118}$$

where $\mathbb{S}$ denotes the Tx-SD search space. In more details, when *M*PSK is employed, the V-BLAST SD's PED increment of (23) may be simplified for SM Tx-SD as:

$$\left|\widetilde{Y}_v - l_{v,v}s_v\right|^2 < R^2, \tag{119}$$

because only one transmit TA is activated. Therefore, the PED increment of (119) defines a new search space for $\{m, v\}$, since only the candidates that lie inside the sphere have to be taken into account by the SM detection of (118). Similarly,

when Square MQAM is employed, the SD's PED increment of (33) may be simplified for SM as:

$$\left(\widetilde{\overline{Y}}_v - l_{v,v}\bar{s}_v\right)^2 < R^2, \quad \forall v \in \{N_T+1, \cdots, 2N_T\}, \tag{120a}$$

$$\left(\widetilde{\overline{Y}}_v - l_{v+N_T,v}\bar{s}_{v+N_T} - l_{v,v}\bar{s}_v\right)^2 < R^2, \quad \forall v \in \{1, \cdots, N_T\}. \tag{120b}$$

The new search space defined in (120) may be further reformulated as $\frac{-R+\widetilde{\overline{Y}}_v}{l_{v,v}} < \bar{s}_v < \frac{R+\widetilde{\overline{Y}}_v}{l_{v,v}}$ for $v \in \{N_T+1, \cdots, 2N_T\}$ and $\frac{-R+(\widetilde{\overline{Y}}_v - l_{v+N_T,v}\bar{s}_{v+N_T})}{l_{v,v}} < \bar{s}_v < \frac{R+(\widetilde{\overline{Y}}_v - l_{v+N_T,v}\bar{s}_{v+N_T})}{l_{v,v}}$ for $v \in \{1, \cdots, N_T\}$ [148], [149]. This Tx-SD-defined search space may effectively reduce the SM detector's search space formulated in (118).

Furthermore, it was also proposed in [150] that the receive search space of the conventional MIMO detection of (11) may be reduced by the so-called Receiver-centric SD (Rx-SD) as:

$$\{\hat{m}, \hat{v}\} = \arg \max_{\forall\{m,v\}} \left\{ n(m,v) | \sum_{r=1}^{n(m,v)} \left|Y_r - s^m H_{v,r}\right|^2 < R^2 \right\}, \tag{121}$$

where Y_r refers to the signal received at the r-th RA, while $H_{v,r}$ models the fading channel spanning from the v-th TA to the r-th RA. The Rx-SD of (121) aims for finding the optimum pair $\{\hat{m}, \hat{v}\}$, which may maximize the counter $n(m, v)$. More explicitly, for a specific data-carrying index pair $\{m, v\}$, the following PED is examined:

$$d_r^{\{m,v\}} = d_{r-1}^{\{m,v\}} + \left|Y_r - s^m H_{v,r}\right|^2 < R^2. \tag{122}$$

If the PED $d_r^{\{m,v\}}$ lies inside the Rx-SD sphere specified by (122), the counter may be incremented according to $n(m, v) = n(m, v)+1$, and the Rx-SD index r may continue to be increased. Otherwise, the PED evaluation of (122) may be terminated, and the next index pair $\{m, v\}$ shall be examined. The sphere radius may be updated as $R^2 = d_{N_R}^{\{m,v\}}$, when the Rx-SD index reaches $r = N_R$. The Rx-SD tree search is supposed to be experienced by all the $I = N_T M$ candidates of the index pair $\{m, v\}$, but a reduced-complexity termination may be expected, when the sphere radius R is swiftly reduced in the high-SNR region. The optimum index pair $\{\hat{m}, \hat{v}\}$ seen in (121) is the one, which maximizes the counter $n_{\max} = \max_{\forall\{m,v\}} n(m, v)$ with the aid of the minimum final PED value as $\{\hat{m}, \hat{v}\} = \arg \min d_{n_{\max}}^{\{m,v\}}$.

As demonstrated in [148] and [149], the Tx-SD of (118) and the Rx-SD of (122) may be combined as:

$$\{\hat{m}, \hat{v}\} = \arg \max_{\forall\{m,v\}\in\mathbb{S}} \left\{ n(m,v) | \sum_{r=1}^{n(m,v)} \left|Y_r - s^m H_{v,r}\right|^2 < R^2 \right\}, \tag{123}$$

where the transmit search space is limited within $\mathbb{S}$, while the receive search space is confined by the Rx-SD counter $n(m, v)$.

For the sake of discussion, it was thoroughly reviewed in [148]–[150] that as a benefit of the SM's specific feature of single TA activation, the SD conceived for SM exhibits a substantially reduced computational complexity compared to the SD conceived for conventional V-BLAST. Furthermore, the optimum SM performance may be attained by the SD, provided that the sphere radius is initialized to be sufficiently large. However, the SD complexity still remains SNR-dependent, since its complexity the lower bound can only be reached in the high-SNR region. Moreover, it is important to note that the hard-limiter-based SM detection presented in Sec. III-A2 may be seen as a special case of the SD aided SM detection. This is because that the transmit search space associated with the classic modulated symbol index m is limited by minimizing the hard-limiter metric in (116), while the receive search space is reduced to a single-antenna-based scenario, since the hard-limiter-based SM detector examines the matched filter outputs $\{z_v\}_{v=1}^{N_T}$ instead of the received signals $\{Y_r\}_{r=1}^{N_R}$.

4) HARD-DECISION NORMALIZED-MRC-BASED SUBOPTIMAL SM DETECTION

First of all, let us introduce the normalized matched filter output as:

$$\overline{\mathbf{Z}} = \mathbf{Y}\left(\overline{\mathbf{H}}\right)^H, \tag{124}$$

where each row in the normalized $(N_T \times N_R)$-element fading channels matrix $\overline{\mathbf{H}}$ is given by $\left\{\overline{\mathbf{H}}_{v,-} = \mathbf{H}_{v,-}/\kappa_v\right\}_{v=1}^{N_T}$, and the v-th element in the N_T-element normalized matched filter output row-vector $\overline{\mathbf{Z}}$ is given by $\{\bar{z}_v = \mathbf{Y}\overline{\mathbf{H}}_{v,-}^H = z_v/\kappa_v\}_{v=1}^{N_T}$.

It was demonstrated in [141]–[143] that a more accurate estimate of the TA activation index may be delivered by testing the normalized matched filter output of (124) instead of the direct matched filter output in the MRC based SM detection of (105). Therefore, the normalized-MRC-based SM detection may determine the TA index by:

$$\hat{v} = \arg \max_{\forall v\in\{1,\cdots,N_T\}} |\bar{z}_v|. \tag{125}$$

Upon obtaining the TA activation index $\hat{v}$, the linear MPSK/QAM demodulator of (116) may be invoked for detecting the classic modulated symbol index as:

$$\hat{m} = \mathbb{M}^{-1}(\bar{z}_{\hat{v}}/\kappa_{\hat{v}}). \tag{126}$$

Therefore, the complexity order of normalized-MRC-based SM detection is also given by $O(N_T + 1)$.

The so-called signal-vector-based detection proposed in [143] operates based on the fact that the Square MQAM symbol does not change the direction of the received signal vector $\mathbf{Y} = s^m \mathbf{H}_{v,-}$. The signal-vector-based detection's estimate of the TA activation index is

given by:

$$\hat{v} = \arg \min_{\forall v \in \{1,\cdots,N_T\}} \arccos\left(\frac{|\mathbf{Y}\mathbf{H}_{v,-}^H|}{\|\mathbf{Y}\|\|\mathbf{H}_{v,-}\|}\right) = \arg \min_{\forall v \in \{1,\cdots,N_T\}} \arccos\left(\frac{|\bar{z}_v|}{\|\mathbf{Y}\|}\right), \quad (127)$$

which is in fact equivalent to the normalized-MRC-based estimation of (125), because $\arccos(\cdot)$ is a function that monotonically decreases with respect to its argument, and $\|\mathbf{Y}\|$ in (127) is a constant.

5) HARD-DECISION LIST-NORMALIZED-MRC-BASED SUBOPTIMAL SM DETECTION

The normalized-MRC-based SM detection still suffers from the problem of imperfect TA index estimation. Therefore, to circumvent this, the list-normalized-MRC-based SM detection is introduced in [141], [142], and [144], where a total of N_{List} TA indices are taken into account in order to avoid the situation of missing the optimum TA index candidate.

More explicitly, instead of selecting a single TA index in (125), a list of N_{List} possible TA candidates is compiled as:

$$[v_1, \cdots, v_{N_{\text{List}}}] = \arg \text{ sortD}_{\forall v \in \{1,\cdots,N_T\}} |\bar{z}_v|. \quad (128)$$

where the operation "sortD" sorts all the elements $\{|\bar{z}_v|\}_{v=1}^{N_T}$ in decreasing order. In the TA index list, v_1 represents the TA index associated with the highest metric $|\bar{z}_{v_1}|$, and $v_{N_{\text{List}}}$ associated with the lowest metric $|\bar{z}_{v_{N_{\text{List}}}}|$. We have $1 \leq N_{\text{List}} \leq N_T$, where the special cases of $N_{\text{List}} = 1$ and $N_{\text{List}} = N_T$ correspond to the normalized-MRC-based SM detection and to the optimum SM detection, respectively. Following this, the demodulator may be invoked N_{List} times for all the candidates on the list as:

$$\hat{m}_{v_t} = \mathbb{M}^{-1}(\bar{z}_{v_t}/\kappa_{v_t}), \ 1 \leq t \leq N_{\text{List}}. \quad (129)$$

The TA activation index may now be confirmed by comparing the N_{List} candidates associated with their respective optimum classic modulated symbol indices according to (107) as:

$$\hat{v} = v_{\hat{t}} = \arg \min_{\forall t \in \{1,\cdots,N_{\text{List}}\}} \kappa_{v_t}^2 |s^{\hat{m}_{v_t}}|^2 - 2\kappa_{v_t} \Re\{(s^{\hat{m}_{v_t}})^* \bar{z}_{v_t}\}. \quad (130)$$

Then the classic modulated symbol index may be given by $\hat{m}_{v_{\hat{t}}}$, which is obtained from (129). The detected classic modulated symbol index as well as the detected TA activation index may now be translated back to bits. The complexity order of the list-normalized-MRC-based SM detector is given by $O(N_T + 2N_{\text{List}})$, where the demodulator has to be invoked N_{List} times in (129) before comparing the N_{List} candidates in (130).

As a further advance, it was proposed in [252] and [253] that a classic modulated symbol index list may be introduced in order to strike a tradeoff between the performance and complexity of the demodulator. More explicitly, a list of constellation points is established for replacing the complete search space for $\{s^m\}_{m=0}^{M-1}$ of (107). In [252], all Square QAM constellations are partitioned into level-1 subsets as well as level-2 subsets, and only $N_{\text{List}-m1}$ constellation points in the level-1 subset and $N_{\text{List}-m2}$ constellation points in the level-2 subset are considered for the demodulation. In [253], the $3 \sim 5$ constellation points that surround the decision variable $\bar{z}_v/\kappa_v$ are considerred for demodulation. Owing to the fact that the hard-decision MPSK/QAM demodulation may be implemented at a very low detection complexity, the further discussion of sub-optimal modulation list establishment in [252] and [253] may be avoided in uncoded systems. We note that the TA index list based SM detection of [141] and [142] may be considered to represent the upper bound for [252], [253] in terms of both performance and complexity.

6) HARD-DECISION UNITY-CONSTELLATION-POWER-BASED SUBOPTIMAL SM DETECTION

The unity-constellation-power-based suboptimal SM detection is proposed in [145], where a total of $\widetilde{M}$ candidates of non-negative constellation points associated with unity constellation power $\{\widetilde{s}^{\widetilde{m}} = \frac{|\Re(s^m)|}{|s^m|} + j\frac{|\Im(s^m)|}{|s^m|}\}_{\widetilde{m}=1}^{\widetilde{M}}$ are taken into account for the sake of more reliable TA index estimation. In more details, the unity-constellation-power-based TA index detection is given by:

$$\hat{v} = \arg \max_{\forall v \in \{1,\cdots,N_T\}, \forall \widetilde{m} \in \{1,\cdots,\widetilde{M}\}} |\Re(\bar{z}_v)|\Re(\widetilde{s}^{\widetilde{m}}) + |\Im(\bar{z}_v)|\Im(\widetilde{s}^{\widetilde{m}}), \quad (131)$$

and then the demodulation regime of (126) may be invoked in order to detect the classic modulated symbol index. The complexity order of unity-constellation-power-based SM detection is given by $O(N_T\widetilde{M} + 1)$. This method was shown to be especially beneficial [145] for Star MQAM detection, because when the constellation power is normalized, only a total of $(M_P/4 + 1)$ M_PPSK phase candidates focused in the first quadrant has to be considered.

In order to improve the TA index detection of (131), list-unity-constellation-power based SM detection was proposed in [146], which may follow the same procedures as the list-normalized-MRC-based SM detection of Sec. III-A5, except that the list establishment of (128) should use the metric of $|\Re(\bar{z}_v)|\Re(\widetilde{s}^{\widetilde{m}}) + |\Im(\bar{z}_v)|\Im(\widetilde{s}^{\widetilde{m}})$ in (131) instead of $|\bar{z}_v|$. As a result, the complexity order of list-unity-constellation-power-based SM detection is given by $O(N_T\widetilde{M} + 2N_{\text{List}})$.

7) HARD-DECISION DISTANCE-ORDERED-BASED SUBOPTIMAL SM DETECTION

The distance-ordered-based suboptimal SM detection of [147] performs classic symbol demodulation first, and then a list of candidate TA indices is established based on the Euclidean distances between the demodulated symbols and the decision variables.

More explicitly, the hard-limiter-based demodulator of (116) is invoked in order to identify the optimum classic modulated symbol indices $\{\hat{m}_v\}_{v=1}^{N_T}$ for all TA index

candidates. Following this, the distance-based TA index list is established by:

$$[v_1, \cdots, v_{N_{\text{List}}}] = \arg \ \text{sortI}_{\forall v \in \{1,\cdots,N_T\}} \ |s^{\hat{m}_v} - \hat{z}_v|\kappa_v, \quad (132)$$

where the sorting operation "sortI" orders all the elements $\{|s^{\hat{m}_v} - \hat{z}_v|\kappa_v\}_{v=1}^{N_T}$ according to their increasing values. Therefore, the TA activation index detected from the list may be decided based on (107) as:

$$\hat{v} = v_{\hat{t}} = \arg \min_{\forall t \in \{1,\cdots,N_{\text{List}}\}} (\left|\hat{z}_{v_t} - s^{\hat{m}_{v_t}}\right|^2 - |\hat{z}_{v_t}|^2)\kappa_{v_t}^2. \quad (133)$$

Naturally, the classic modulated symbol index may be directly obtained by $\hat{m}_{v_{\hat{t}}}$. The complexity order of distance-ordered-based SM detection is given by $O(2N_T + N_{\text{List}})$.

8) SOFT-DECISION REDUCED-SCOPE-BASED OPTIMAL SM DETECTION

In order to further reduce the complexity order $O(N_T M)$ of the soft-decision simplified MAP aided SM detectors using (108), the reduced-scope hard-decision SM detectors introduced in Sec. III-A1 were also revised as soft-decision detectors in [139], which retain the optimal MAP SM detection capability by visiting only a reduced subset of the SM search space. More explicitly, similar to (109), the *a posteriori* probability metric of (108) may be extended as:

$$d^{v,m} = \frac{\Re(\widetilde{z}_v)\Re(s^m)}{N_0} + \frac{\Im(\widetilde{z}_v)\Im(s^m)}{N_0} - \frac{\kappa_v^2|s^m|^2}{N_0} + \sum_{\bar{k}=1}^{\log_2 I} \widetilde{b}_{\bar{k}} L_a(b_{\bar{k}}). \quad (134)$$

For the $\text{BPS}_T = \log_2 N_T$ bits that are assigned to modulate the TA activation index v, the soft-bit decisions produced by the Max-Log-MAP of (15) may be expressed as:

$$L_p(b_k) = \max_{\forall v \in \{1,\cdots,N_T\}_{b_k=1}} d^v - \max_{\forall v \in \{1,\cdots,N_T\}_{b_k=0}} d^v, \quad (135)$$

where $\{1, \cdots, N_T\}_{b_k=1}$ and $\{1, \cdots, N_T\}_{b_k=0}$ refer to the index set for v, when the specific bit $\{b_k\}_{k=\text{BPS}+1}^{\log_2 I}$ is fixed to 1 and 0, respectively. In order to produce the *a posteriori* LLR in (135), we have to obtain the maximum probability metric for each TA activation index v as:

$$d^v = \max_{\forall m \in \{0,\cdots,M-1\}} d^{v,m}. \quad (136)$$

It can be readily seen in (136) that the reduction of the SM detection search space may be achieved by invoking the reduced-complexity soft-decision MPSK/QAM detection of Algorithm 1, where by exploring the symmetry provided by the Gray-labelled MPSK/QAM constellation diagrams, only a reduced subset of positive PAM magnitudes and a reduced subset of constellation points of the first quadrant have to be visited, when Square QAM and general PSK/QAM are employed, respectively.

Let us consider QPSK as an example. Once again, the QPSK's detected constellation diagram is rotated anti-clockwisely by $\pi/4$, so that there is only a single constellation point in each quadrant. As a result, the decision variable should be rotated as $z'_v = \widetilde{z}_v \exp(j\frac{\pi}{4})$, and the detected constellation points are given by $\{s'^m = s^m \exp(j\frac{\pi}{4})\}_{m=0}^{M-1} = \{\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}\}$. Therefore, the maximum probability metric of (136) over four QPSK constellation points is given by (137), as shown at the bottom of this page, where the *a priori* probability metric for the TA activation index v is given by $p_v^a = \sum_{\bar{k}=\text{BPS}+1}^{\log_2 I} \widetilde{b}_{\bar{k}} L_a(b_{\bar{k}})$, while the two new variables associated with testing the real and imaginary parts separately are defined as $t_{\text{Re}}^v = \frac{\Re(z'_v)}{\sqrt{2}N_0} - \frac{L_a(b_2)}{2}$ and $t_{\text{Im}}^v = \frac{\Im(z'_v)}{\sqrt{2}N_0} - \frac{L_a(b_1)}{2}$. As a result, the maximum probability of (137), may be simply given by a one-step evaluation as:

$$d^v = |t_{\text{Re}}^v| + |t_{\text{Im}}^v| - \frac{\kappa_v^2}{N_0} + p_v^a, \quad (138)$$

where a constant of $\frac{L_a(b_1)+L_a(b_2)}{2}$ is discarded from (137), because this term may be eliminated by the subtraction in the Max-Log-MAP of (135). Therefore, instead of evaluating and comparing a total number of $M = 4$ probability metrics corresponding to all QPSK constellation points in (137), the calculation of (138) in fact only visits a single constellation point, which is located in the first quadrant. As a result, the Max-Log-MAP of (135) only has to evaluate and compare the N_T *a posteriori* probability metrics $\{d^v\}_{v=1}^{N_T}$ of (138). Therefore, the SM-QPSK detection complexity order has been reduced from $O(N_T M)$ to $O(N_T)$ for detecting the $\text{BPS}_T = \log_2 N_T$ bits that are assigned to the TA activation index v. Moreover, for the $\text{BPS} = \log_2 M = 2$ bits that are assigned to encode the QPSK's classic modulated symbol index m, when a specific bit $\{b_k\}_{k=1}^2$ is set to 1 or 0 as required by the Max-Log-MAP of (15), the QPSK constellation set has to be updated. More specifically, when the first bit is set to be $b_1 = 1$ or $b_1 = 0$, the QPSK constellation set has to be updated as $\{\frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}, , -\frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}\}$ or $\{\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}\}$, respectively. As a result, the

$$d^v = \max \left\{ \begin{array}{l} \frac{\Re(z'_v)}{\sqrt{2}N_0} + \frac{\Im(z'_v)}{\sqrt{2}N_0} - \frac{\kappa_v^2}{N_0} + p_v^a \\ -\frac{\Re(z'_v)}{\sqrt{2}N_0} + \frac{\Im(z'_v)}{\sqrt{2}N_0} + L_a(b_2) - \frac{\kappa_v^2}{N_0} + p_v^a \\ \frac{\Re(z'_v)}{\sqrt{2}N_0} - \frac{\Im(z'_v)}{\sqrt{2}N_0} + L_a(b_1) - \frac{\kappa_v^2}{N_0} + p_v^a \\ -\frac{\Re(z'_v)}{\sqrt{2}N_0} - \frac{\Im(z'_v)}{\sqrt{2}N_0} + L_a(b_1) + L_a(b_2) - \frac{\kappa_v^2}{N_0} + p_v^a \end{array} \right\} = \max \left\{ \begin{array}{l} t_{\text{Re}}^v + t_{\text{Im}}^v - \frac{\kappa_v^2}{N_0} + p_v^a \\ -t_{\text{Re}}^v + t_{\text{Im}}^v - \frac{\kappa_v^2}{N_0} + p_v^a \\ t_{\text{Re}}^v - t_{\text{Im}}^v - \frac{\kappa_v^2}{N_0} + p_v^a \\ -t_{\text{Re}}^v - t_{\text{Im}}^v - \frac{\kappa_v^2}{N_0} + p_v^a \end{array} \right\} + \frac{L_a(b_1) + L_a(b_2)}{2}. \quad (137)$$

Max-Log-MAP algorithm of (15) may be simplified for producing the first soft-bit decision as:

$$L_p(b_1) = \max_{v\in\{1,\cdots,N_T\}}\left(|t^v_{\mathrm{Re}}| - t^v_{\mathrm{Im}} - \frac{\kappa_v^2}{N_0} + p^a_v\right) - \max_{v\in\{1,\cdots,N_T\}}\left(|t^v_{\mathrm{Re}}| + t^v_{\mathrm{Im}} - \frac{\kappa_v^2}{N_0} + p^a_v\right), \quad (139)$$

where the imaginary term of $|t^v_{\mathrm{Im}}|$ in (138) is replaced by $(-t^v_{\mathrm{Im}})$ and (t^v_{Im}), when b_1 is fixed to 1 and 0, respectively. Similarly, the second soft-bit decision is given by:

$$L_p(b_2) = \max_{v\in\{1,\cdots,N_T\}}\left(-t^v_{\mathrm{Re}} + |t^v_{\mathrm{Im}}| - \frac{\kappa_v^2}{N_0} + p^a_v\right) - \max_{v\in\{1,\cdots,N_T\}}\left(t^v_{\mathrm{Re}} + |t^v_{\mathrm{Im}}| - \frac{\kappa_v^2}{N_0} + p^a_v\right). \quad (140)$$

The complexity order of (139) and (140) is given by $O(2N_T)$. It is worth noting that (139) and (140) of MPSK/QAM demodulation impose the minimum computational complexity, which only involves combinations and comparisons of the quantities that have already been evaluated by the antenna index detection of (135) using (138).

Therefore, the design guidelines for soft-decision reduced-scope-based SM detection of [139] may be summarized as follows:

Algorithm 3: Design guidelines for reduced-scope-based soft-decision SM detection

1) First of all, the maximum probability metric d^v over $\{d^{v,m}\}_{\forall m}$ of (134) for each TA activation index v is given by $d^v = \max_{\forall m\in\{0,\cdots,M-1\}} d^{v,m}$ as seen in (136), whose search space may be reduced by invoking the reduced-complexity soft-decision MPSK/QAM detectors of [216], which is summarized as Algorithm 1 in this paper.
2) Secondly, when the Max-Log-MAP of (15) is invoked, the $\mathrm{BPS}_T = \log_2 N_T$ soft-bit decisions associated with the TA activation index v may be produced as expressed by (135), where the local maximum probability metrics $\{d^v\}_{\forall v}$ have been obtained in Step (1)
3) Finally, the $\mathrm{BPS} = \log_2 M$ number of soft-bit decisions associated with the modulated symbol index m may be produced by updating the MPSK/QAM constellation set for a specific bit $\{b_k\}_{k=1}^{\log_2 M}$ being set to 1 or 0 according to the Max-Log-MAP of (15).

We note that the production of *a posteriori* LLRs in Steps (2) and (3) only involves combinations and comparisons of the quantities that have already been evaluated in Step (1), which requires a low computational complexity. Moreover, the Approx-Log-MAP of (16) may also be invoked by the soft-decision reduced-scope-based SM detection as seen in [139].

In summary, it was demonstrated in [139] that for the $\mathrm{BPS}_T = \log_2 N_T$ bits assigned to the TA index, the 1PSK/BPSK/QPSK aided SM detection operates at the complexity order lower bound of $O(N_T)$, while the Square MQAM aided SM detection and the MPSK/QAM aided SM detection have the complexity order of $O(\sqrt{M}\cdot N_T)$ and $O(\frac{M}{4}\cdot N_T)$, respectively. For the pair of specific bits, which determine the sign of the transmitted MPSK/QAM symbol, the BPSK/QPSK aided SM detection complexity is given by the order of $O(2N_T)$, while the Square MQAM aided SM detection complexity order and the general MPSK/QAM aided SM detection complexity order are given by $O(\sqrt{M}\cdot N_T)$ and $O(\frac{M}{2}\cdot N_T)$, respectively. For the remaining $(\mathrm{BPS}-2)$ bits, which determine the specific magnitudes of the MPSK/QAM symbols, the complexity order of the Square MQAM aided SM detection and that of the general MPSK/QAM aided SM detection are given by $O(\frac{\sqrt{M}}{2}\cdot N_T)$ and $O(\frac{M}{4}\cdot N_T)$, respectively. In summary, the Square MQAM aided SM detection has a lower complexity compared to the general MPSK/QAM aided SM detection, owing to the fact that the real and imaginary parts of Square MQAM constellation may be visited separately.

B. SPACE-TIME SHIFT KEYING (STSK)

The concept of STSK was proposed in [39] as a combination of SM and LDC, so that a transmit diversity gain may be obtained by the family of SM-style low-complexity MIMO systems. The schematic of the STSK transmitter is portrayed in Fig. 13, which is modified from the LDC transmitter of Fig. 11, where only a single one out of a total of N_Q dispersion matrices is selected for dispersing a single MPSK/QAM symbol. As a result, the $(N_P \times N_T)$-element STSK transmission matrix created from the LDC transmission matrix of (96) may be expressed as [39]:

$$\mathbf{S} = \overline{\mathbf{A}}_q s^m, \quad (141)$$

where the first $\mathrm{BPS} = \log_2 M$ source information bits are assigned to modulate a single MPSK/QAM symbol $s^m = \mathbb{M}(m)$, while the following $\mathrm{BPS}_Q = \log_2 N_Q$ source information bits are assigned to select a single dispersion matrix $\overline{\mathbf{A}}_q$ among a total number of N_Q candidates. There are a total of $(I = N_Q M)$ STSK codewords $\{\mathbf{S}^i\}_{i=0}^{N_Q M-1}$, and the STSK throughput is given by $(R = \frac{\log_2 I}{N_P} = \frac{\mathrm{BPS}+\mathrm{BPS}_Q}{N_P})$, where the employment of N_P time slots is considered.

Similarly to the signal processing performed at the LDC receiver introduced in Sec. II-C3, the STSK receiver may firstly vectorizes the received MIMO signal model of (1) in order to form the received LDC signal model of (97), which is rewritten here for the sake of clarification:

$$\overline{\mathbf{Y}} = \overline{\mathbf{S}}\cdot\overline{\chi}\cdot\overline{\mathbf{H}} + \overline{\mathbf{V}}, \quad (142)$$

where the $N_P N_R$-element equivalent received signal row-vector $\overline{\mathbf{Y}} = \mathrm{rvec}(\mathbf{Y})$, the $(N_Q \times N_P N_T)$-element equivalent dispersion matrix $\overline{\chi}$, the $(N_P N_T \times N_P N_R)$-element equivalent fading matrix $\overline{\mathbf{H}} = \mathbf{I}_{N_P}\otimes\mathbf{H}$ and the $N_P N_R$-element equivalent

AWGN row-vector $\overline{\mathbf{V}} = \text{rvec}(\mathbf{V})$ are all exactly the same as those of the LDC in (98). However, the N_Q-element equivalent STSK input signal row-vector in (142) is given by:

$$\overline{\mathbf{S}} = [\underbrace{0 \cdots 0}_{q-1}, s^m, \underbrace{0 \cdots 0}_{N_Q-q}], \tag{143}$$

which is in the same form as the SM input signal vector as seen in Table 4. Therefore, according to the STSK received signal model of (142), a STSK(N_T,N_R,N_P,N_Q) scheme is equivalent to a SM system associated with N_Q TAs and $N_P N_R$ RAs, where the equivalent SM fading matrix is given by STSK's $\breve{\mathbf{H}} = \overline{\chi}\overline{\mathbf{H}}$, as defined in (142). As a result, all the SM detectors summarized in Sec. III-A may be invoked by the STSK receivers.

STSK generation guidelines

(1) Randomly generate a group of N_Q unitary matrices $\{\widetilde{\mathbf{A}}_q\}_{q=1}^{N_Q}$ of size $(N \times N)$, where we have $N = \max(N_T, N_P)$.
   a) If $N_P > N_T$ is required, the STSK dispersion matrices are given by taking the first N_T columns of the scaled unitary matrices as $\{\overline{\mathbf{A}}_q = \sqrt{\frac{N_P}{N_T}}\widetilde{\mathbf{A}}_q \begin{bmatrix} \mathbf{I}_{N_T} \\ \mathbf{0} \end{bmatrix}\}_{q=1}^{N_Q}$, where $\mathbf{0}$ is a $(N_P - N_T) \times N_T$-element all-zero matrix.
   b) If $N_P = N_T$ is required, the STSK dispersion matrices are directly given by $\{\overline{\mathbf{A}}_q = \widetilde{\mathbf{A}}_q\}_{q=1}^{N_Q}$.
   c) If $N_P < N_T$ is required, the STSK dispersion matrices are given by taking the first N_P rows of the scaled unitary matrices as $\{\overline{\mathbf{A}}_q = [\mathbf{I}_{N_P}, \mathbf{0}]\,\widetilde{\mathbf{A}}_q\}_{q=1}^{N_Q}$, where $\mathbf{0}$ is a $N_P \times (N_T - N_P)$-element all-zero matrix.

Since only a single dispersion matrix is activated, STSK loses the LDC's capacity advantage, which will be further discussed in Sec. III-D. Nonetheless, the generation of the STSK's dispersion matrices may still rely on populating them with the aid of a random search, and then the specific dispersion matrix set that minimize the PEP of (8) may be selected. As discussed before, the PEP union bound of (8) is minimized, when $\mathbf{\Delta} = (\mathbf{S}^i - \mathbf{S}^{\bar{i}})^H(\mathbf{S}^i - \mathbf{S}^{\bar{i}})$ is unitary, which is equivalent to the following requirements:

$$\overline{\mathbf{A}}_q^H \overline{\mathbf{A}}_q = \frac{N_P}{N_T}\mathbf{I}_{N_T}, \quad \forall q \in \{1, \cdots, N_Q\}, \tag{144a}$$

$$\overline{\mathbf{A}}_q^H \overline{\mathbf{A}}_{\bar{q}} = -\overline{\mathbf{A}}_{\bar{q}}^H \overline{\mathbf{A}}_q, \quad \forall q \neq \bar{q} \in \{1, \cdots, N_Q\}. \tag{144b}$$

We note that the first requirement of (144a) may be readily satisfied by directly generating the scaled unitary matrices for the case of $N_P \geq N_T$, while the second requirement of (144b) can only be approached by maximizing either the minimum determinant $\{\det(\mathbf{\Delta})\}_{\min}$ or the second metric $\sum_{\forall q \neq \bar{q}} \|\overline{\mathbf{A}}_q^H \overline{\mathbf{A}}_{\bar{q}} + \overline{\mathbf{A}}_{\bar{q}}^H \overline{\mathbf{A}}_q\|$ in (102) according to the suggestions in [37].

In more details, the generation of STSK may be summarized as:

(2) Rank criterion: for the resultant ($I = N_Q M$) STSK codewords $\{\mathbf{S}^i\}_{i=0}^{I-1}$ of (141), having a full rank should be guaranteed for all combinations of $\mathbf{\Delta} = (\mathbf{S}^i - \mathbf{S}^{\bar{i}})^H(\mathbf{S}^i - \mathbf{S}^{\bar{i}})$ as $\text{rank}(\mathbf{\Delta}) = \min(N_T, N_P)$.
(3) Determinant criterion: The minimum determinant among all combinations of $\mathbf{\Delta}$ is given by $\{\det(\mathbf{\Delta})\}_{\min}$. The related random search may be carried out by repeating Steps (1) and (2), while the chosen set should maximize $\{\det(\mathbf{\Delta})\}_{\min}$. For the sake of designing high-throughput STSK schemes, the chosen set may aim for maximizing $\sum_{\forall q \neq \bar{q}} \|\overline{\mathbf{A}}_q^H \overline{\mathbf{A}}_{\bar{q}} + \overline{\mathbf{A}}_{\bar{q}}^H \overline{\mathbf{A}}_q\|$ instead of determinant for the sake of faster random search termination.

When more than one dispersion matrices are allowed to be activated for the sake of achieving an increased throughput, the STSK scheme may be further developed to the concept of Generalized Space-Time Shift Keying (GSTSK), as presented in [151] and [254], where both STSK and LDC constitute special cases of GSTSK. Furthermore, since LDC was proposed for generalizing both V-BLAST and STBC, GSTSK may include virtually all MIMO schemes. In more details, the dispersion matrix of V-BLAST is given by (73), while the dispersion matrix design of STBC was discussed in Sec. II-B. The classic MIMO schemes of V-BLAST, STBC and LDC may all be deemed to be special GSTSK cases, which rely on activating all TAs. Moreover, SM may be considered to be a special case of STSK, where the SM dispersion matrices are given by:

$$\mathbf{A}_q = [\underbrace{0 \cdots 0}_{q-1}, 1, \underbrace{0 \cdots 0}_{N_T-q}], \quad \forall q \in \{1, \cdots, N_T\}. \tag{145}$$

Furthermore, in [152]–[155], SM was improved for the sake of achieving a transmit diversity gain by activating more than one TAs in order to convey STBC codewords, which can be readily subsumed by the framework of GSTSK according to the STBC dispersion matrix design of Sec. II-B. However, the orthogonal channels of STBC-aided transmit diversity were created either by employing the idealistic orthogonal shapping filters of [156]–[158] or the orthogonal frequency-hopping codes of [159]. These schemes no longer fit into the scope of GSTSK due to their additional hardware requirements. Moreover, when more than one classic modulated symbols are transmitted by GSTSK, the problem of IAI once again arises, unless orthogonal STBC codewords are transmitted. As a result, sub-optimal interference-rejecting receivers are proposed to be employed by the family of GSTSK receivers in [160] and [162], which are less consistent with the SM/STSK motivation of low-complexity ML receiver designs. However, it was suggested in [163] and [164] that the IAI may vanish, if the multiple activated TAs of the Generalized Spatial Modulation (GSM)

TABLE 5. Summary of hard-decision optimal SM detectors.

Hard-decision optimal SM Detectors	Complexity Order	Computational Complexity (real-valued multiplications)	Related Literatures
General ML aided MIMO detector	$O(N_T M)$	$(4N_T N_R + 2N_R)N_T M$	[12], [66], [165]
Simplified SM detector	$O(N_T M)$	$6N_T N_R + 6N_T M$	[134]
Reduced-scope-based SM detector	MPSK: $O(N_T M/4)$	$(6N_R + M/2 + 2)N_T$	[138], [139], [259]
	Square MQAM: $O(N_T\sqrt{M})$	$(6N_R + 2\sqrt{M} + 2)N_T$	
Hard-limiter-based SM detector	$O(2N_T)$	MPSK: $(6N_R + 9)N_T$	[137], [140]
		Square MQAM: $(6N_R+11)N_T$	

TABLE 6. Summary of hard-decision suboptimal SM detectors.

Hard-decision suboptimal SM Detectors	Complexity Order	Computational Complexity (real-valued multiplications)	Related Literatures
Maximum Ratio Combining (MRC)-based SM detector	$O(N_T + 1)$	MPSK: $4N_R + 2$	[27]
		Square MQAM: $4N_R + 4$	
Normalized-MRC-based SM detector	$O(N_T + 1)$	MPSK: $(6N_R + 2)N_T + 4$	[141]–[143]
		Square MQAM: $(6N_R + 2)N_T + 6$	
List-normalized-MRC-based SM detector	$O(N_T + 2N_{List})$	MPSK: $(6N_R + 2)N_T + 9N_{List}$	[141], [142], [144]
		Square MQAM: $(6N_R + 2)N_T + 11N_{List}$	
Unity-constellation-power-based SM detector	$O(N_T\widetilde{M} + 1)$	MPSK: $(6N_R + 2\widetilde{M} + 2)N_T + 4$	[145]
		Square MQAM: $(6N_R + 2\widetilde{M} + 2)N_T + 6$	
List-unity-constellation-power-based SM detector	$O(N_T\widetilde{M} + 2N_{List})$	MPSK: $(6N_R + 2\widetilde{M} + 2)N_T + 9N_{List}$	[146]
		Square MQAM: $(6N_R+2\widetilde{M}+2)N_T+11N_{List}$	
Distance-ordered-based SM detector	$O(2N_T + N_{List})$	MPSK: $(6N_R + 7)N_T + 5N_{List}$	[147]
		Square MQAM: $(6N_R + 9)N_T + 5N_{List}$	

opt for transmitting the same symbol. Further discussions on relaxing the GSM scheme's constraints concerning N_T may be found in [255]–[258].

Against this background, the concept of a GSM/GSTSK scheme that achieves an improved capacity without imposing IAI remains an open prospect, which we set aside for future work. In order to better prepare for this ambitious objective, we offer a discussion on the SM/STSK scheme's error probability and capacity in Sec. III-D, so that their strength and limitations may be better understood.

C. COMPARISON BETWEEN OPTIMAL AND SUBOPTIMAL DETECTORS

The optimal SM detectors and the suboptimal SM detectors introduced in Sec. III-A are summarized at a glance in Tables 5 and 6, respectively. The hard-decision optimal SM detectors of Table 5 were developed for reducing the general ML aided MIMO detection complexity, while maintaining their ML detection capability. By contrast, the hard-decision suboptimal SM detectors of Table 6 aim for improving the performance of the MRC-based SM detector of [27], which is the problematic TA activation index detection. The associated computational complexity in Tables 5 and 6 is summarized in terms of the total number of real-valued multiplications.

Fig. 29 portrays the performance comparison between the different hard-decision SM detectors, when they are invoked by the SM receivers and by the STSK receivers. It can be seen in Fig. 29(a) that both the normalized-MRC-based and list-normalized-MRC-based SM detectors exhibit an error floor for SM detection, when there is no receive diversity gain owing to having $N_R = 1$, but their performance improves as N_R increases. It may also be observed in Fig. 29 that unity-constellation-power-based SM detector performs better than normalized-MRC-based SM detector and that the list-based detectors such as the list-unity-constellation-power-based and list-normalized-MRC-based SM detectors outperform their respective unity-constellation-power-based and normalized-MRC-based SM counterparts. In general, all optimal SM detectors of Table 5 achieve the same ML performance, while all suboptimal SM detectors of Table 6 impose a performance loss on both SM and STSK in Fig. 29 in a conceptually similar manner to the MMSE detector's performance loss inflicted upon V-BLAST as seen in Fig. 25.

The detection complexity comparison of the different hard-decision SM detectors is presented in Fig. 30. First of all, compared to the simplified ML aided SM detector, the reduced-scope-based SM detector offers a substantial complexity reduction, which is as high as 84.3% for SM(N_T,1)-Square 16QAM ($1 \leq \text{BPS}_T \leq 4$) in Fig. 30(a), and is up to 93.8% for SM(4,1)-MPSK/QAM ($1 \leq \text{BPS} \leq 6$), as seen in Fig. 30(b). Moreover, the hard-limiter-based SM detector further provides a slightly lower complexity than the reduced-scope-based SM detector, when the number of modulation levels is as high as $M = 64$ in Fig. 30(b). For the representatives of suboptimal SM detectors, it can be seen in Fig. 30 that the normalized-MRC-based and list-normalized-MRC-based SM detectors do not show a significant complexity advantage. Considering their suboptimal performance quantified in Fig. 29, we may conclude that both the reduced-scope-based SM detector and the hard-limiter-based SM detector are more attractive candidates in terms of

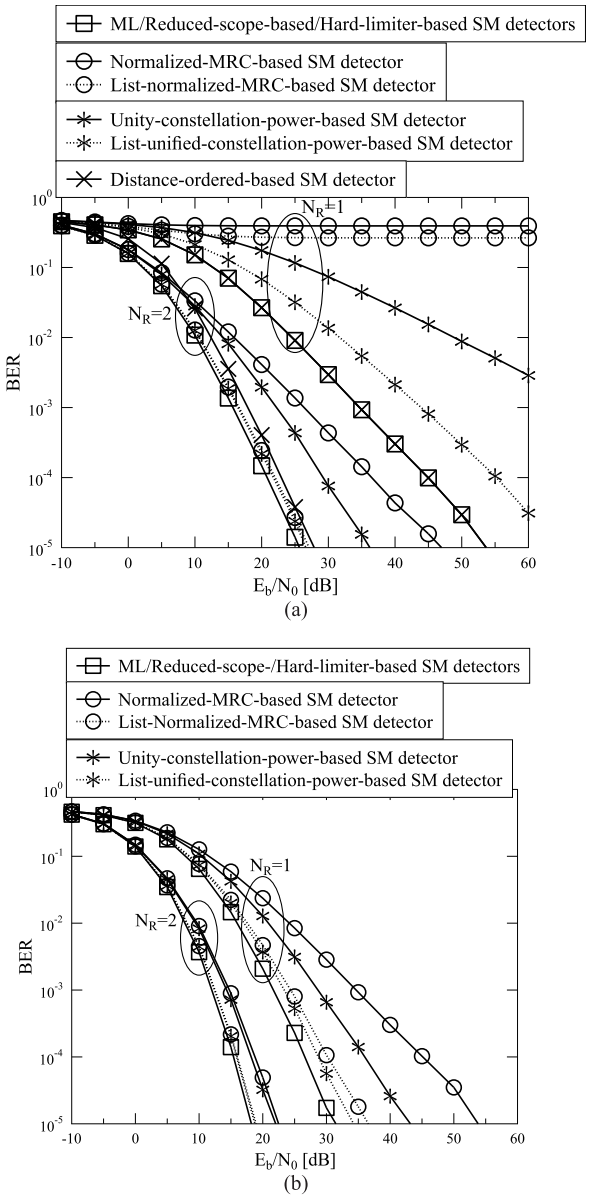


FIGURE 29. Performance comparison between different SM detectors summarized in Sec. III-A, when they are invoked by SM and STSK receivers. (a) SM(4,N_R)-Square 16QAM. (b) STSK(4,N_R,2,16)-Square 16QAM.

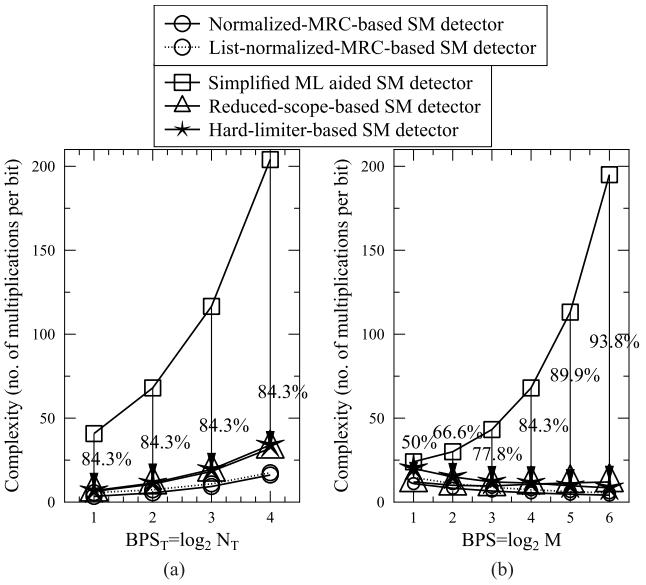


FIGURE 30. Complexity comparison between optimal and suboptimal hard-decision SM detectors. (a) SM(N_T,1)-Square 16QAM. (b) SM(4,1)-*M*PSK/Square *M*QAM.

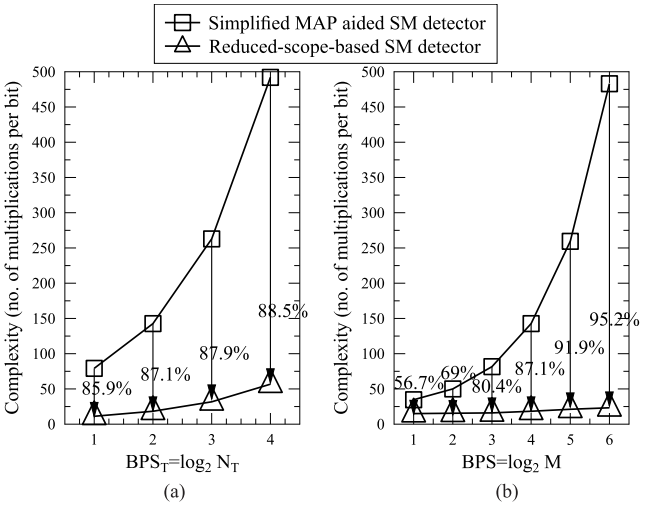


FIGURE 31. Complexity comparison between the soft-decision simplified MAP aided SM detector and the soft-decision reduced-scope-based SM detector. (a) SM(N_T,1)-Square 16QAM. (b) SM(4,1)-*M*PSK/QAM.

offering a substantially reduced detection complexity, while maintaining the optimum SM performance.

Fig. 31 further offers complexity comparison of the soft-decision SM detectors in terms of the total number of real-valued multiplications. Owing to the zeros in the SM transmitted symbol vector of (103), the SM probability metric estimation of (108) is already less computationally complex than the MIMO probability metric estimation of (14). Nonetheless, Fig. 31 evidences that the reduced-scope-based SM detector offers a further substantial complexity reduction compared to the simplified MAP aided SM detector, which is as high as 85.9% $\sim$ 88.5% for SM(N_T,1)-Square 16QAM employing different number of TAs from the set $N_T = \{2, 4, 8, 16\}$ in Fig. 31(a) and up to 56.7% $\sim$ 95.2% for SM(4,1) employing different *M*PSK/QAM constellations of $M = \{2, 4, 8, 16, 32, 64\}$ in Fig. 31(b). It is worth noting that the complexity reduction demonstrated in Fig. 31 is particularly beneficial for turbo detection assisted MIMO schemes, where the soft-decision MIMO detector is invoked several times in order to achieve the best possible performance promised by the MIMO capacity predictions.

D. ERROR PROBABILITY AND CAPACITY ANALYSIS FOR SM AND STSK

The theoretical average BER of virtually all MIMO schemes is characterized by (7), where each PEP is bounded by (8).

For the case of SM, the Pairwise Squared Euclidean Distance (PSED) that directly determines the PEP of (8) may be expressed as:

$$\left\|\mathbf{S}^{i}-\mathbf{S}^{\bar{i}}\right\|^{2}=\begin{cases}|s^{m}|^{2}+|s^{\bar{m}}|^{2}, & \text{Case 1: } v\neq\bar{v},\ m\neq\bar{m}\\ 2|s^{m}|^{2}, & \text{Case 2: } v\neq\bar{v},\ m=\bar{m}\\ |s^{m}-s^{\bar{m}}|^{2}, & \text{Case 3: } v=\bar{v},\ m\neq\bar{m},\end{cases} \tag{146}$$

where the SM codeword indices i and $\bar{i}$ represent the TA activation indices and the classic modulated symbol indices $\{v, m\}$ and $\{\bar{v}, \bar{m}\}$, respectively. The corresponding relationships are given by $(i = mN_T+v-1)$ and $(\bar{i} = \bar{m}N_T+\bar{v}-1)$. For the sake of comparison, the PSED of the V-BLAST scheme may be expressed in a similar form as:

$$\left\|\mathbf{S}^{i}-\mathbf{S}^{\bar{i}}\right\|^{2}=\sum_{v=1}^{N_T}|s_{v}^{i}-s_{v}^{\bar{i}}|^{2}. \tag{147}$$

For the case of V-BLAST, the worst case of the minimum PSED that may maximize the PEP bound of (8) occurs, when the two V-BLAST codeword vectors $\mathbf{S}^i$ and $\mathbf{S}^{\bar{i}}$ only differ in a single element, which corresponds to SM's Case 3 in (146). However, any SM codeword has a total number of $(N_T - 1)$ zeros, which means that any two SM vectors $\mathbf{S}^i$ and $\mathbf{S}^{\bar{i}}$ in (146) share at least $(N_T - 2)$ zero elements. However, V-BLAST's pairwise codewords are often different in more than two elements for $(N_T > 2)$.

Moreover, since the throughput of V-BLAST is given by $R = N_T$BPS while that of SM by $R = \text{BPS}_T + \text{BPS}$, the SM system has to employ a higher-order MPSK/QAM constellation in order to match the throughput of the V-BLAST system equipped with the same number of N_T TAs. For example, a V-BLAST($4,N_T$)-QPSK scheme has a throughput of $R = 8$, which requires the SM($4,N_R$) system to employ a 64QAM scheme. As a result, the SM's PSED of (146) is substantially degraded owing to both the reduced constellation point powers and the reduced Euclidean distances between the constellation points.

For these reasons, SM is unlikely to outperform V-BLAST at the same system throughput under the same hardware and software conditions, albeit SM has a potential low-complexity advantage. Indeed, this would only be possible for SM systems, under the employment of extra hardware for creating transmit diversity techniques [15], [158], [160], orthogonal shapping filters [15], [156], [158], or when aiming for a reduced SM throughput [161] or when using more complex ML aided SM detectors while opting for suboptimal LF aided V-BLAST detectors [27], [140], [141], [149], [152]. In order to clarify this matter, we will provide a discussion on the performance and complexity tradeoff between V-BLAST and SM in Sec. III-E, where the system requirements are the same for both of them.

It is also interesting to see in (146) that the SM's PEP experienced in some combinations is determined by constellation point power $\{|s^m|^2\}_{m=0}^{M-1}$, which is not the case for V-BLAST,

TABLE 7. Summary of the minimum constellation point distances $\min_{\forall\{s^m\neq s^{m'}\}}|s^m - s^{m'}|$ and the minimum constellation point powers $\min_{\forall s^m}|s^m|^2$ for 16-level and 64-level PSK/QAM constellations.

	$\min_{\forall\{s^m\neq s^{m'}\}}\lvert s^m-s^{m'}\rvert$	$\min_{\forall s^m}\lvert s^m\rvert^2$
Square 16QAM	0.4	0.2
Star 16QAM ($M_A = 2, \alpha = 2$)	0.234	0.4
16PSK	0.152	1
Square 64QAM	0.0952	0.0476
Star 64QAM ($M_A = 4, \alpha = 1.4$)	0.0425	0.279
64PSK	0.0096	1

as presented in (147). The same feature may also be observed in terms of the STSK's PSED, which may be expressed as:

$$\left\|\mathbf{S}^{i}-\mathbf{S}^{\bar{i}}\right\|^{2}=\begin{cases}\left\|\overline{\mathbf{A}}_{q}s^{m}-\overline{\mathbf{A}}_{\bar{q}}s^{\bar{m}}\right\|^{2}, & \text{Case 1: } q\neq\bar{q},\ m\neq\bar{m}\\ \left\|\overline{\mathbf{A}}_{q}-\overline{\mathbf{A}}_{\bar{q}}\right\|^{2}\cdot|s^{m}|^{2}, & \text{Case 2: } q\neq\bar{q},\ m=\bar{m}\\ \left\|\overline{\mathbf{A}}_{q}\right\|^{2}\cdot|s^{m}-s^{\bar{m}}|^{2}, & \text{Case 3: } q=\bar{q},\ m\neq\bar{m},\end{cases} \tag{148}$$

where the STSK codeword indices are formulated as $(i = mN_Q + q - 1)$ and $(\bar{i} = \bar{m}N_Q + \bar{q} - 1)$. It can be seen in both (146) and (148) that a higher value of $\min_{\forall s^m}|s^m|^2$ is required by Case 2, while a higher $\min_{\forall\{s^m\neq s^{m'}\}}|s^m - s^{m'}|$ value is required by Case 3 for both SM and STSK. These two requirements cannot be satisfied by the MPSK/QAM constellations at the same time. For example, the minimum constellation point distances and the minimum constellation point powers are summarized for 16-level and 64-level PSK/QAM constellations in Table 7, which demonstrates that Square QAM exhibits the highest minimum constellation point distance $\min_{\forall\{s^m\neq s^{m'}\}}|s^m - s^{m'}|$, but both PSK and Star QAM have a higher constellation point power $\min_{\forall s^m}|s^m|^2$.

Fig. 32 portrays the performance comparison between SM/STSK employing different MPSK/QAM constellations. It may be observed in Fig. 32 that for a lower modulation order of $M = 16$, SM(16,2) and STSK(4,2,2,16) employing 16PSK perform even better than their Square 16QAM and Star 16QAM based counterparts, which is an explicit benefit of 16PSK's dominant advantage of having a higher constellation point power as seen in Table 7. However, as the number of modulation levels is increased to $M = 64$, SM(16,2) and STSK(4,2,2,16) employing Star 64QAM perform the best, where the 64PSK's reduced constellation point distance of $\min_{\forall\{s^m\neq s^{m'}\}}|s^m - s^{m'}|$ shown in Table 7 severely degrades its performance in the concept of SM/STSK systems. We note that although MPSK and Star MQAM may exhibit a performance advantage of SM/STSK, the detection complexity for SM/STSK employing Square MQAM becomes the lowest, when the reduced-scope SM detector is employed, owing to the fact that the real and imaginary parts of the Square MQAM constellation may be visited separately.

Even though its complexity advantage is attractive, SM and STSK fail to achieve the full MIMO capacity of (4). In more

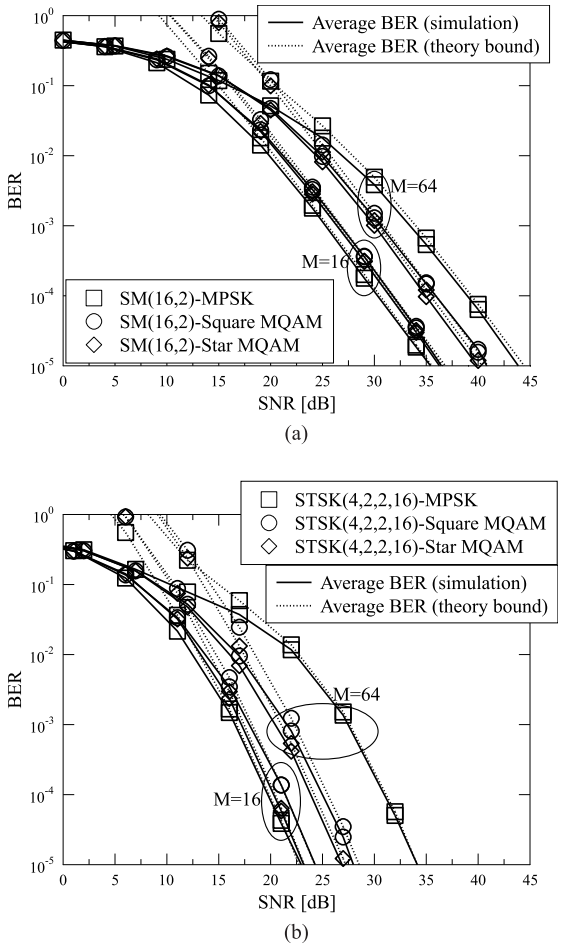


FIGURE 32. Performance comparison between SM/STSK employing different *M*PSK/QAM constellations. (a) SM employing *M*PSK/QAM. (b) STSK employing *M*PSK/QAM.

details, the SM's mutual information between the input and output signals may be formulated as [137], [140]:

$$C_{SM}^{CCMC}(SNR) = \max_{p(s^m),p(v)} I(\{s^m, v\}; \mathbf{Y}) = \max_{p(s^m)} I(s^m; \mathbf{Y}|v) + \max_{p(v)} I(v; \mathbf{Y}), \quad (149)$$

where the input signal vector $\mathbf{S}$ is given by (103), while the output signal vector $\mathbf{Y}$ is given by (1). The first term $\max_{p(s^m)} I(s^m; \mathbf{Y}|v)$ in (149) represents a SIMO system's capacity, which is maximized, when the input is assumed to be a Gaussian-distributed continuous signal. This may be expressed as:

$$C_{SM,1}^{CCMC}(SNR) = \frac{1}{N_T}\sum_{v=1}^{N_T} \log_2(1+\eta\|\mathbf{H}_{v,-}\|^2), \quad (150)$$

where the entropy of the AWGN variable is given by $H(\mathbf{Y}|\{s^m, v\}) = H(\mathbf{V}) = \log_2 \det\left[\pi e N_0 \mathbf{I}_{N_R}\right]$, while that of the Gaussian-distributed output signal by $H(\mathbf{Y}|v) = \log_2 \det\left[\pi e(\mathbf{H}_{v,-}^H \mathbf{H}_{v,-} + N_0\mathbf{I}_{N_R})\right]$. Furthermore, the second capacity term of (149) is also maximized by the Gaussian PDF of the output signal, which is given by $p(\mathbf{Y}|v) = \frac{1}{\det(\pi \mathbf{R}_{YY|v})} \exp\left[-\mathbf{Y}\mathbf{R}_{YY|v}^{-1}\mathbf{Y}^H\right]$, where provided that the v-th TA is active, the autocorrelation matrix of received signal $\mathbf{Y}$ is given by $\mathbf{R}_{YY|v} = \mathrm{E}\left(\mathbf{Y}^H\mathbf{Y}|v\right) = \mathbf{H}_{v,-}^H\mathbf{H}_{v,-} + N_0\mathbf{I}_{N_R}$. Therefore, the determinant term is given by $\det(\pi\mathbf{R}_{YY|v}) = \pi^{N_R}N_0^{N_R}\det(\mathbf{I}_{N_R} + \eta\mathbf{H}_{v,-}^H\mathbf{H}_{v,-}) = \pi^{N_R}N_0^{N_R}(1+\eta\kappa_v^2)$, where we have $\kappa_v^2 = \|\mathbf{H}_{v,-}\|^2$, as defined in (107). As a result, the second capacity term $\max_{p(v)} I(v; \mathbf{Y})$ of (149) may be further extended as:

$$C_{SM,2}^{CCMC}(SNR) = \max_{p(v)} \int\int p(\mathbf{Y}|v)p(v)\log_2\frac{p(\mathbf{Y}|v)}{p(\mathbf{Y})}dvd\mathbf{Y}, \quad (151)$$

where the average output signal PDF is given by $p(\mathbf{Y}) = \int p(\mathbf{Y}|v)p(v)dv$. Naturally, (151) is maximized, when the input PDF $p(v)$ is Gaussian. However, the TA activation index v is confined to the limited range of $(1 \le v \le N_T)$, which cannot be generalized by letting N_T tend to infinity. Therefore, we have to accept the fact that the TA activation index v can only be interpreted as a discrete input signal, and hence (151) is maximized for equiprobable sources of $\{p(v) = \frac{1}{N_T}\}_{v=1}^{N_T}$ as (152), as shown at the bottom of this page, where the statistically Gaussian output signal may be directly generated, given the sole input signal v as $\mathbf{Y} = \mathbf{H}_{v,-} + \mathbf{V}$, which was appropriately revised from (1).

For the case of STSK, the CCMC capacity may also be evaluated based on (149), where the equivalent fading channel of SM is given by $\overline{\mathbf{H}} = \overline{\chi}\overline{\mathbf{H}}$ according to the received signal vectorization of (142). Naturally, the STSK capacity has to be normalized by N_P owing to the employment of multiple time slots. Therefore, the SM's CCMC capacity of (149) may be revised for STSK as:

$$C_{STSK}^{CCMC}(SNR) = \max_{p(s^m),p(q)} \frac{1}{N_P} I(\{s^m, q\}; \overline{\mathbf{Y}}) = \max_{p(s^m)} \frac{1}{N_P} I(s^m; \overline{\mathbf{Y}}|q) + \max_{p(q)} \frac{1}{N_P} I(q; \overline{\mathbf{Y}}). \quad (153)$$

$$C_{SM,2}^{CCMC}(SNR) = \frac{1}{N_T}\sum_{v=1}^{N_T} \mathrm{E}\left\{\log_2\frac{p(\mathbf{Y}|v)}{p(\mathbf{Y})}\right\} = \frac{1}{N_T}\sum_{v=1}^{N_T} \mathrm{E}\left\{\log_2 \frac{\frac{N_T}{1+\eta\kappa_v^2}\exp\left[-\mathbf{Y}\left(\mathbf{H}_{v,-}^H\mathbf{H}_{v,-} + N_0\mathbf{I}_{N_R}\right)^{-1}\mathbf{Y}^H\right]}{\sum_{\bar{v}=1}^{N_T}\frac{1}{1+\eta\kappa_{\bar{v}}^2}\exp\left[-\mathbf{Y}\left(\mathbf{H}_{\bar{v},-}^H\mathbf{H}_{\bar{v},-} + N_0\mathbf{I}_{N_R}\right)^{-1}\mathbf{Y}^H\right]}\right\}, \quad (152)$$

The first part of (153) may be modified from (150) as:

$$C^{CCMC}_{STSK,1}(SNR) = \max_{p(s^m)} \frac{1}{N_P} H(\overline{\mathbf{Y}}|q) - \frac{1}{N_P} H(\overline{\mathbf{Y}}|\{s^m, q\}) \tag{154a}$$

$$= \frac{1}{N_P N_Q} \sum_{q=1}^{N_Q} \log_2 \det \times \left(\mathbf{I}_{N_P N_R} + \eta \overline{\mathbf{H}}^H \overline{\chi}^H_{q,-} \overline{\chi}_{q,-} \overline{\mathbf{H}}\right) \tag{154b}$$

$$= \frac{1}{N_P N_Q} \sum_{q=1}^{N_Q} \log_2 \left(1 + \eta \|\overline{\chi}_{q,-} \overline{\mathbf{H}}\|^2\right), \tag{154c}$$

where the related entropies are given by $H(\overline{\mathbf{Y}}|\{s^m, q\}) = H(\overline{\mathbf{V}}) = \log_2 \det\left[\pi e N_0 \mathbf{I}_{N_P N_R}\right]$ and $H(\overline{\mathbf{Y}}|q) = \log_2 \det\left[\pi e(\breve{\mathbf{H}}^H_{q,-}\breve{\mathbf{H}}_{q,-} + N_0 \mathbf{I}_{N_P N_R})\right]$ according to the STSK's equivalent received signal model of (142), while $\{\breve{\mathbf{H}}_{q,-}\}_{q=1}^{N_Q}$ and $\{\overline{\chi}_{q,-}\}_{q=1}^{N_Q}$ refer to the q-th row vectors obtained from $\breve{\mathbf{H}} = \overline{\chi}\overline{\mathbf{H}}$ and $\overline{\chi}$ of (142), respectively. Comparing (154b) to the LDC capacity of (99), it may be observed that the STSK capacity cannot reach the full MIMO capacity by forcing $\overline{\chi}^H_{q,-}\overline{\chi}_{q,-} = \mathbf{I}_{N_P N_T}$, because it requires that the elements of the dispersion matrix $\overline{\mathbf{A}}_q$ satisfy both $\{\{|\overline{A}_q^{t,v}|^2 = 1\}_{t=1}^{N_P}\}_{v=1}^{N_T}$ and $\{\{(\overline{A}_q^{\bar{t},\bar{v}})^* \overline{A}_q^{t,v} = 0\}_{\forall \bar{t} \neq t}\}_{\forall \bar{v} \neq v}$, which cannot be achieved. Comparing (154c) to (150), it may also be observed that the first term of the STSK capacity expression is smaller than that of the SM capacity. In more details, for the case of $N_P \geq N_T$, all STSK dispersion matrices may satisfy $\overline{\mathbf{A}}^H_q \overline{\mathbf{A}}_q = \frac{N_P}{N_T}\mathbf{I}_{N_T}$, as discussed in Sec. III-B. Hence (154c) may be further extended as:

$$C^{CCMC}_{STSK,1}(\eta) = \frac{1}{N_P N_Q} \sum_{q=1}^{N_Q} \log_2 \left(1 + \frac{N_P \eta}{N_T} \|\mathbf{H}\|^2\right)$$
$$= \frac{1}{N_P} C^{CCMC}_{SIMO}\left(\frac{N_P \eta}{N_T}, N_R N_P\right)$$
$$< C^{CCMC}_{SIMO}(\eta, N_R) = C^{CCMC}_{SM,1}(\eta, N_R), \tag{155}$$

where we have $\|\overline{\chi}_{q,-}\overline{\mathbf{H}}\|^2 = \|\overline{\mathbf{A}}_q \mathbf{H}\|^2$ according to (142). It is shown by (155) that the first term of the STSK capacity $C^{CCMC}_{STSK,1}(\eta)$ of (154) is equivalent to the SIMO system's capacity associated with the scaled SNR of $\frac{N_P \eta}{N_T}$ as well as with the increased number of $N_R N_P$ RAs, which is normalized over N_P channel uses. Therefore, the first STSK capacity term $C^{CCMC}_{STSK,1}(\eta)$ of (154) is smaller than the first SM capacity term $C^{CCMC}_{SM,1}(\eta)$ of (150), which equals to the SIMO system's capacity of $C^{CCMC}_{SIMO}(\eta, N_R) = \mathrm{E}\left[\log_2(1 + \eta \cdot \|\mathbf{H}\|^2)\right]$.

Finally, the second part of the STSK capacity of (153) may be obtained by modifying the SM's (151) and (152) according to (142) as (156), as shown at the bottom of the next page, where the STSK dispersion matrix selection is discretized similarly to the SM TA selection, and hence the mutual information $I(q; \overline{\mathbf{Y}})$ is maximized for the equiprobable source of $\{p(q) = \frac{1}{N_Q}\}_{q=1}^{N_Q}$, while we have $\{\kappa_q^2 = \|\breve{\mathbf{H}}_{q,-}\|^2\}$ for the STSK's equivalent received signal model of (142).

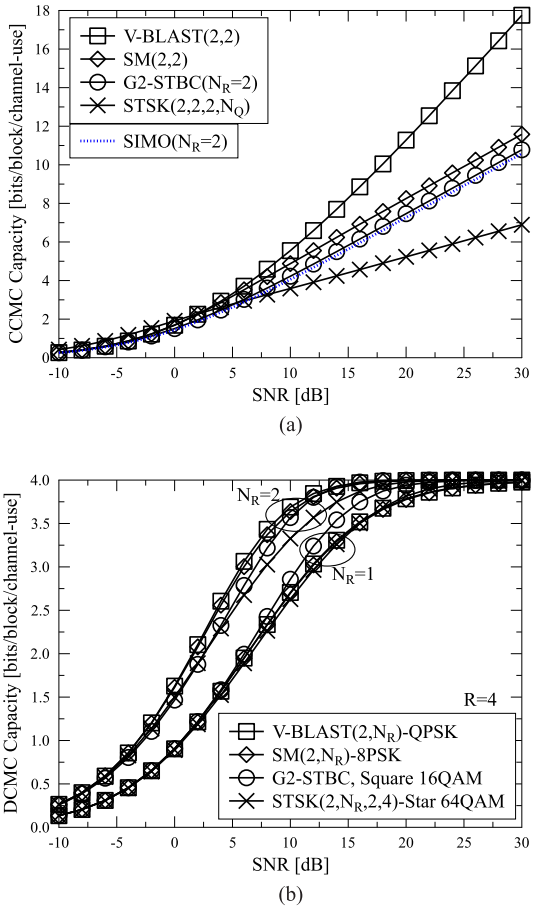


FIGURE 33. Capacity comparison between V-BLAST, SM, Alamouti's G2-STBC and STSK, where $N_T = 2$ TAs are employed. (a) CCMC Capacity. (b) DCMC Capacity.

Since the first STSK capacity term of (154) is lower than the SIMO capacity, while the second STSK capacity term of (156) saturates according to $\max_{SNR} C^{CCMC}_{STSK,2}(SNR) = \frac{\mathrm{BPS}_Q}{N_P}$, the overall STSK capacity of (153) is expected to be lower than the SIMO system capacity in the high-SNR region.

E. SUMMARY OF MIMO PERFORMANCE COMPARISONS

The multiplexing versus diversity tradeoffs associated with the classic MIMO schemes of V-BLAST, STBC and LDC were presented in Sec. II-D. In this section, we further enrich the MIMO performance comparisons by including the results of SM and STSK, where the performance versus complexity tradeoff are highlighted.

1) CAPACITY COMPARISON

As discussed in Sec. III-A, both SM and STSK constitute attractive design alternatives to multiplexing- and diversity-oriented MIMO schemes as a benefit of their lower detection complexity. Let us examine the capacity of SM and STSK first. Fig. 33(a) shows that SM(2,2) cannot achieve the

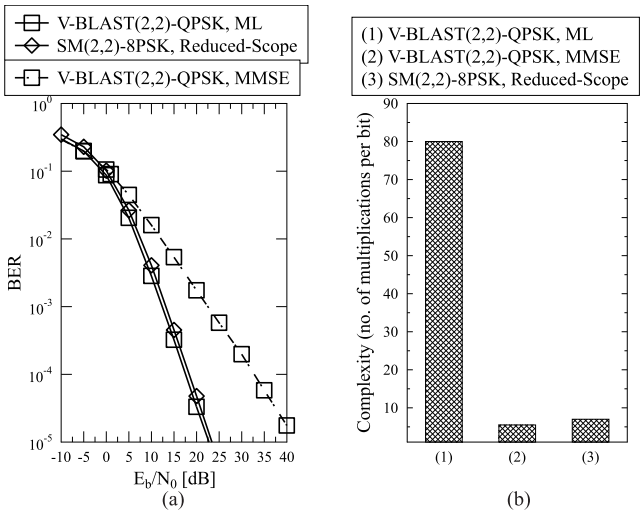


FIGURE 34. BER and Complexity comparison between hard-decision ML and MMSE aided V-BLAST(2,2)-QPSK as well as reduced-scope-based SM(2,2)-8PSK, which are associated with the same throughput of $R = 4$ bits/block/channel-use. (a) BER Comparison. (b) Complexity Comparison.

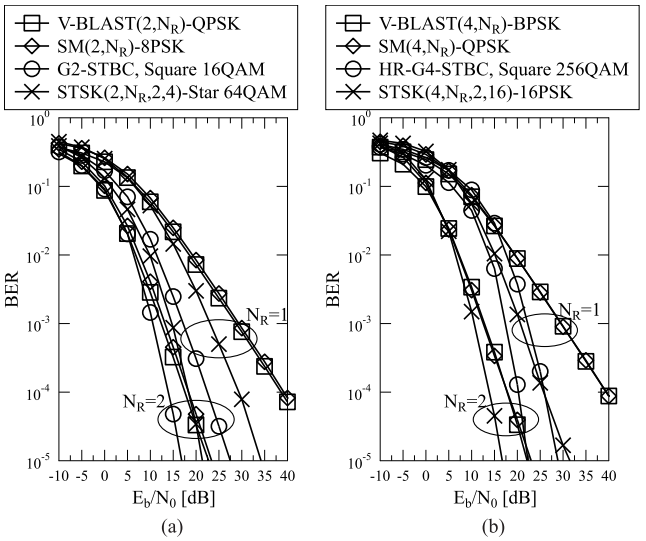


FIGURE 35. Performance comparison between V-BLAST, SM, STBC and STSK associated with the same throughput of $R = 4$ bits/block/channel-use. (a) $N_T = 2$. (b) $N_T = 4$.

V-BLAST(2,2)'s full MIMO capacity, but the SM capacity is evidently higher than that of both Alamouti's G2-STBC and of the SIMO system. By contrast, STSK(2,2,2,4) performs poorly in terms of CCMC capacity, as seen in Fig. 33(a), where the STSK capacity is seem to be even lower than the capacity of the SIMO system, as previously predicted in Sec. II-B3.

Fig. 33(b) demonstrates furthermore that Alamouti's G2-STBC approaches its full DCMC capacity at a lower SNR than the others for the case of $N_R = 1$. However, when $N_R = 2$ RAs are employed, both Alamouti's G2-STBC and STSK exhibit a lower DCMC capacity in the low SNR region, as evidenced by Fig. 33(b). This is because both the diversity-oriented schemes have to employ higher-order modulations in order to achieve the same throughput, as their multiplexing-oriented counterparts.

2) THE PERFORMANCE VERSUS COMPLEXITY TRADEOFF IN UNCODED MIMO SYSTEMS

First of all, a simple BER and complexity comparison between V-BLAST and SM is exemplified by Fig. 34. It can be seen in Fig. 34(a) that the performance difference between SM(2,2)-8PSK and V-BLAST(2,2)-QPSK is almost negligible compared to the performance degradation imposed by employing the low-complexity MMSE detector for V-BLAST(2,2)-QPSK. Fig. 37(a) further demonstrates that the reduced-scope-based SM(2,2)-8PSK detector exhibits a similarly low detection complexity to that of the linear MMSE aided V-BLAST(2,2)-QPSK detector, both of which are substantially lower than the ML aided V-BLAST(2,2)-QPSK detector.

Fig. 35 further extends the scope of our discussions by including the diversity-oriented MIMO schemes of STBC and STSK. Explicitly, it can be seen in Fig. 35(a) that when there is no receive diversity, SM(2,1)-8PSK performs slightly worse than V-BLAST(2,1)-QPSK, while STSK(2,1,2,4)-Star 64QAM has an improved performance as a benefit of its diversity gain, but Alamouti's G2-STBC employing Square 16QAM exhibits the best performance. However, Fig. 35(a) also shows that as N_R is increased, both the SM scheme and the V-BLAST scheme perform better at low SNRs, because the STSK and the G2-STBC arrangements have to employ high-order QAM in order to compensate for their throughput loss owing to utilizing N_P symbol periods. Similar trends may be observed in Fig. 35(b) for the case of $N_T = 4$. It is worth noting that the transmit diversity order of STSK(4,N_R,2,16)-16PSK is given by $\min(N_T, N_P) = 2$, which is lower than the full diversity order of HR-G4-STBC. The number of symbol periods N_P is flexibly adjustable for STSK, which results in a very flexible system design. More explicitly, the STSK associated with $N_P = N_T$ may achieve the full diversity order of $N_T N_R$, while a lower $N_P < N_T$ allows the STSK scheme to employ a lower-order modulation to be used for achieving the same system throughput, which may result in a better performance in the low SNR region.

$$C^{CCMC}_{STSK,2}(SNR) = \frac{1}{N_P N_Q} \sum_{q=1}^{N_Q} \mathrm{E}\left\{ \log_2 \frac{\frac{N_Q}{1+\eta\kappa_q^2} \exp\left[-\overline{\mathbf{Y}} \left(\breve{\mathbf{H}}^H_{q,-} \breve{\mathbf{H}}_{q,-} + N_0 \mathbf{I}_{N_P N_R} \right)^{-1} \overline{\mathbf{Y}}^H \right]}{\sum_{\tilde{q}=1}^{N_T} \frac{1}{1+\eta\kappa_{\tilde{q}}^2} \exp\left[-\overline{\mathbf{Y}} \left(\breve{\mathbf{H}}^H_{\tilde{q},-} \breve{\mathbf{H}}_{\tilde{q},-} + N_0 \mathbf{I}_{N_P N_R} \right)^{-1} \overline{\mathbf{Y}}^H \right]} \right\}, \tag{156}$$

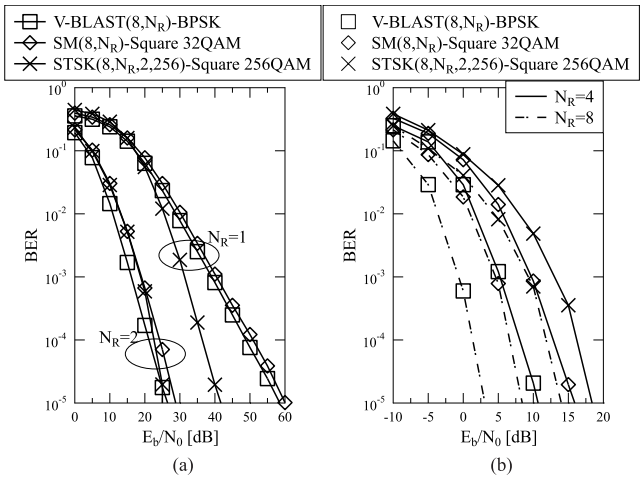


FIGURE 36. Performance comparison between V-BLAST, SM and STSK associated with a high number of TAs $N_T = 8$ and a higher throughput of $R = 8$ bits/block/channel-use. (a) $N_T = 8$, $N_R = \{1, 2\}$. (b) $N_T = 8$, $N_R = \{4, 8\}$.

Once again, comparing the results of Fig. 25 and Fig. 35, it is essential to note that the performance loss imposed by employing SM instead of V-BLAST is significantly lower than that of employing a low-complexity linear MMSE receiver for V-BLAST in Fig. 25. The same trend prevails, when STSK is compared to LDC in Fig. 25. Therefore, the ultimate benefit of the SM and STSK systems lies on their complexity advantage.

Let us now elaborate a little further on the performance of MIMO systems associated with a higher number of antennas. Fig. 36 shows that the transmit diversity gain obtained by STSK only becomes advantageous, when there is no receive diversity owing to using $N_R = 1$ RA. As the number of RAs increases to $N_R = 2$, $N_R = 4$ and even to $N_R = 8$, the performance of V-BLAST and SM becomes better and the performance difference between V-BLAST and SM is increased. This is because the high multiplexing gain of V-BLAST allows it to employ the low-order BPSK modulation for achieving the same throughput as SM and STSK employing higher-order modulation schemes. This important feature implies that altough STBC and STSK may be conceived for any arbitrary number of TAs, V-BLAST may be preferred for large-scale MIMO systems equipped with a large number of antennas at the base stations [15]–[18]. SM may act as an alternative to V-BLAST at the cost of a slightly degraded performance achieved at a substantially reduced detection complexity.

In order to offer a quantitative complexity comparison between the conventional MIMO receivers designed for V-BLAST and STBC as well as the low-complexity optimal MIMO receivers conceived for SM and STSK, it is assumed that the fading channels do not change for a sufficiently long period of time, so that the MMSE filters taps of (37) adjusted for V-BLAST are not required to be updated frequently, while the fading channel powers $\{\kappa_v\}_{v=1}^{N_T}$ estimated by the SM detectors of Sec. III-A and the STSK scheme's equivalent fading

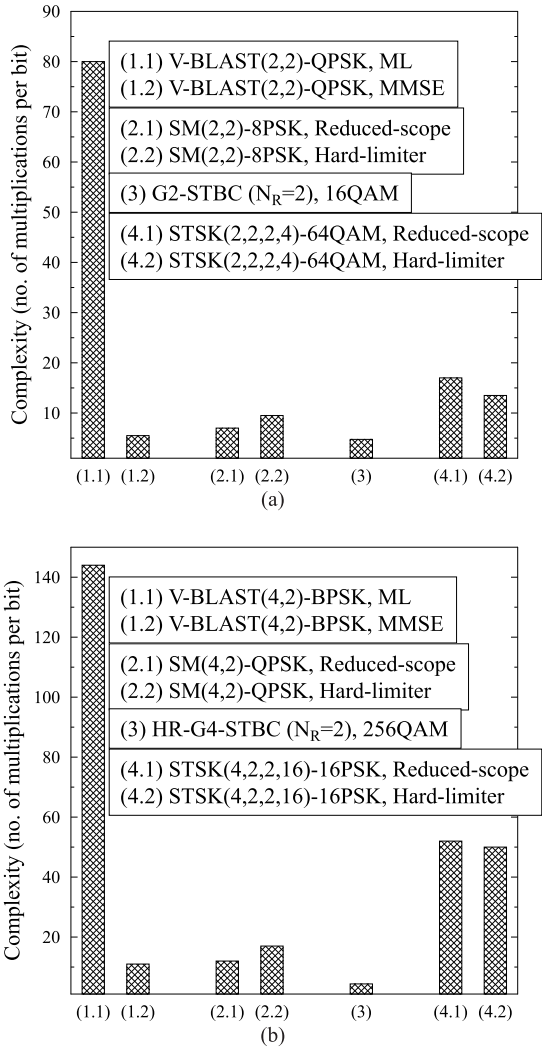


FIGURE 37. Complexity comparison between V-BLAST, SM, STBC and STSK associated with the same throughput of $R = 4$ bits/block/channel-use, where the fading channels are assumed to be constant, so that the same operations do not have to be repeated by the MIMO receivers. (a) $N_T = 2$, $N_R = 2$. (b) $N_T = 4$, $N_R = 2$.

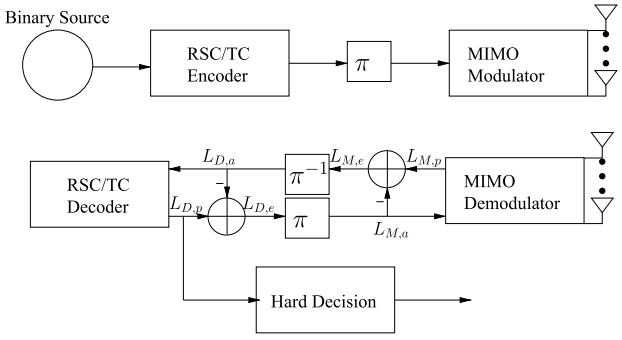


FIGURE 38. The schematic of a RSC/TC coded MIMO system.

matrix $\breve{\mathbf{H}} = \overline{\chi}\overline{\mathbf{H}}$ of (142) may also remain unchanged. Under this condition, it can be seen in Fig. 37 that the orthogonal STBC achieves the lowest detection complexity for both $N_T = 2$ and $N_T = 4$. It is also evidenced by Fig. 37 that

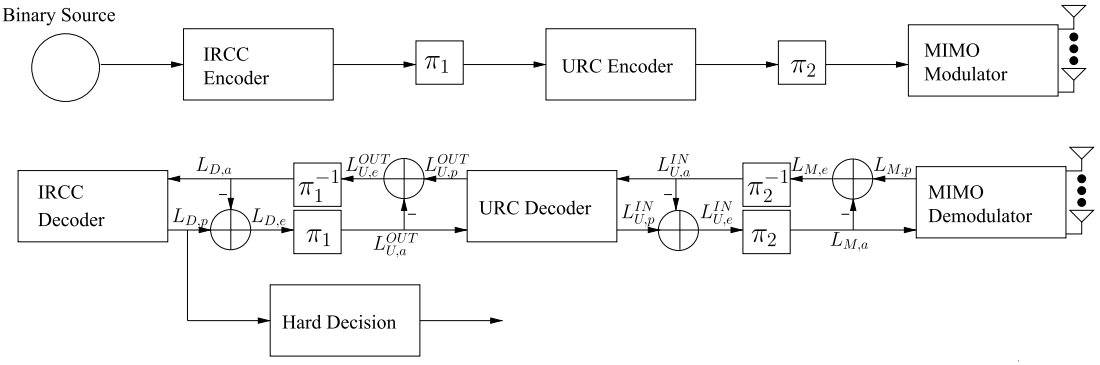


FIGURE 39. The schematic of a IRCC and URC coded MIMO system.

TABLE 8. System parameters.

Channel Coding	RSC coded MIMO (Schematics Fig. 38) TC coded MIMO (Schematics Fig. 38) IRCC-URC coded MIMO (Schematics Fig. 39)
RSC coded MIMO	$IR_{RSC-MIMO}$ iterations between soft-decision RSC decoder and soft-decision MIMO detector
TC coded MIMO	IR_{TC} iterations within soft-decision TC decoder $IR_{TC-MIMO}$ iterations between soft-decision TC decoder and soft-decision MIMO detector
IRCC-URC coded MIMO	$IR_{URC-MIMO}$ iterations between soft-decision URC decoder and soft-decision MIMO detector $I_{IRCC-\{URC-MIMO\}}$ iterations between soft-decision IRCC and the amalgamated soft-decision URC-MIMO decoder.
Frame length	1 000 000 bits

the ML MIMO detector designed for V-BLAST exhibits the highest complexity, while the linear MMSE receiver successfully mitigates this complexity problem, at the cost of an eroded performance as seen in Fig. 25. Against this background, the SM detectors are capable of offering a complexity that is slightly higher than that of the MMSE receiver of V-BLAST, but still substantially lower than that of the ML MIMO detector of V-BLAST, as demonstrated by Fig. 37. Let us recall from Sec. III-B that the STSK receivers require extra signal processing, before being able to invoke the SM detectors. Therefore, it is shown by Fig. 37 that the STSK detection complexity is higher than the SM detection complexity. Nonetheless, considering that STSK is capable of offering a diversity gain for SM, as demonstrated by Fig. 35, the employment of STSK is beneficial, because its detection complexity is considerably lower than that of both the V-BLAST and the LDC receivers invoking the ML MIMO detector, as evidenced by Fig. 37.

3) THE PERFORMANCE VERSUS COMPLEXITY TRADEOFF IN CODED MIMO SYSTEMS

In this section, we compare diverse MIMO schemes in the context of a variety of coded systems, where the simulation parameters are summerized in Table 8. We note that the reduced-scope-based soft-decision SM detector of Sec. III-A8 is employed by both SM and STSK in this section. In order to overcome these limitations of the conventional MIMO systems, as discussed in Sec. II-A9, it is desirable

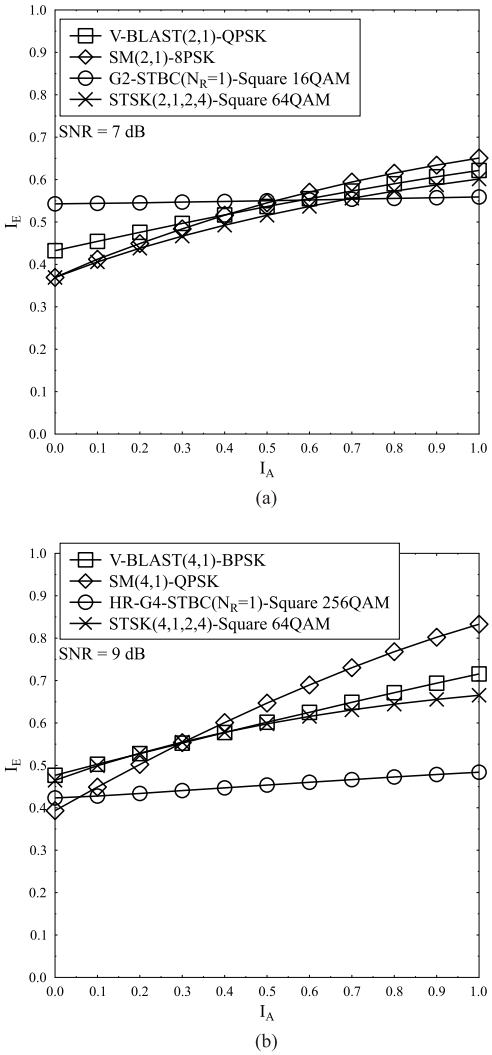


FIGURE 40. EXIT chart comparison of V-BLAST, SM, STBC and STSK associated with the same system throughput of $R_c R = 2$ bits/block/channel use. (a) $N_T = 2$. (b) $N_T = 4$.

for SM to invoke its optimum detector at a reduced detection complexity. Let us firstly examine the performance of V-BLAST and SM together with STBC and STSK in the

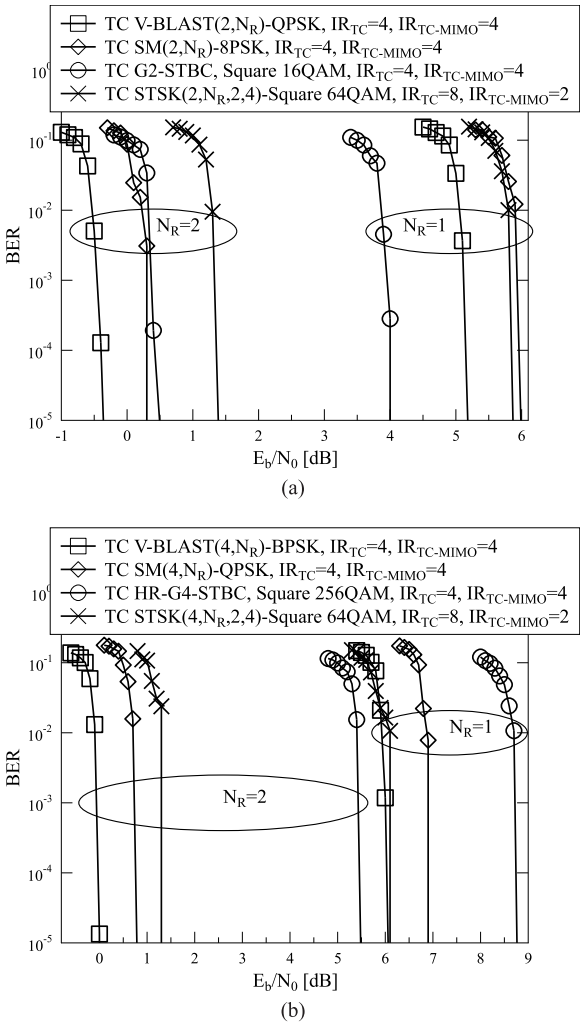


FIGURE 41. BER performance comparison between half-rate TC V-BLAST, SM, STBC and STSK associated with the same system throughput of $R_c R = 2$ bits/block/channel use. (a) $N_T = 2$. (b) $N_T = 4$.

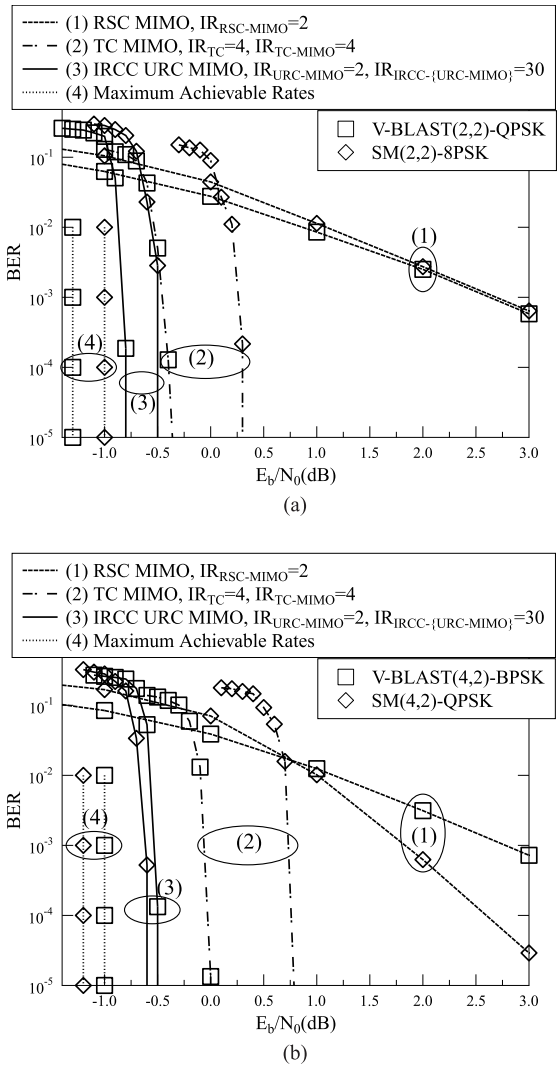


FIGURE 42. BER performance comparison between V-BLAST and SM in the context of RSC, TC and IRCC-URC coded systems associated with the same system throughput of $R_c R = 2$ bits/block/channel use. (a) $N_T = 2$. (b) $N_T = 4$.

context of coded systems with the aid of the EXIT charts of Fig. 40 and the BER performance curves of Fig. 41. It can be seen in Fig. 40 that the STBC's orthogonal design results in a near-horizontal EXIT curve, similarly to a classic SISO scheme. By contrast, the V-BLAST, SM and STSK schemes exhibit a considerable iteration gain. For this reason, the number of iterations is set to $IR_{TC} = 4$ and $IR_{TC-MIMO} = 4$ for the TC coded V-BLAST, SM and STSK systems, while $IR_{TC} = 8$ and $IR_{TC-MIMO} = 2$ are used for the TC coded STBC systems. The BER performance of Fig. 41 shows that when there is no receive diversity, SM(2,1)-8PSK and SM(4,1)-QPSK perform worse than their respective V-BLAST counterparts of V-BLAST(2,1)-QPSK and V-BLAST(4,1)-BPSK in the context of the TC coded MIMO systems, but the STSK(2,1,2,4)-Square 64QAM and STSK(4,1,2,4)-Square 64QAM schemes offer a performance improvement over their SM counterparts, as a benefit of STSK's transmit diversity gain. It can be seen in Fig. 41(a) that the G2 STBC employing Square 16QAM exhibits the best performance amongst the TC coded MIMO systems, when we have $N_R = 1$. However, when all MIMO schemes benefit from a receive diversity gain owing to $N_R = 2$, V-BLAST(2,2)-QPSK performs the best in Fig. 41(a), SM(2,2)-8PSK performs similarly to G2 STBC, while STSK(2,2,2,4)-Square 64QAM performs the worst. This is because the diversity-oriented STBC and STSK schemes have to employ higher order QAM arrangements in order to compensate for their throughput loss. Hence STBC and STSK generally perform worse at low SNRs than their V-BLAST and SM counterparts employing lower order MPSK/QAM, when all of them benefit from a diversity gain owing to employing $N_R > 1$. For the same reason, it may be observed in Fig. 41(b) that V-BLAST(4,2)-BPSK performs the best amongst the TC MIMO systems, while SM(4,2)-QPSK performs better than its STSK(4,2,2,4)-Square 64QAM counterpart. Furthermore, it can be seen that the TC half-rate

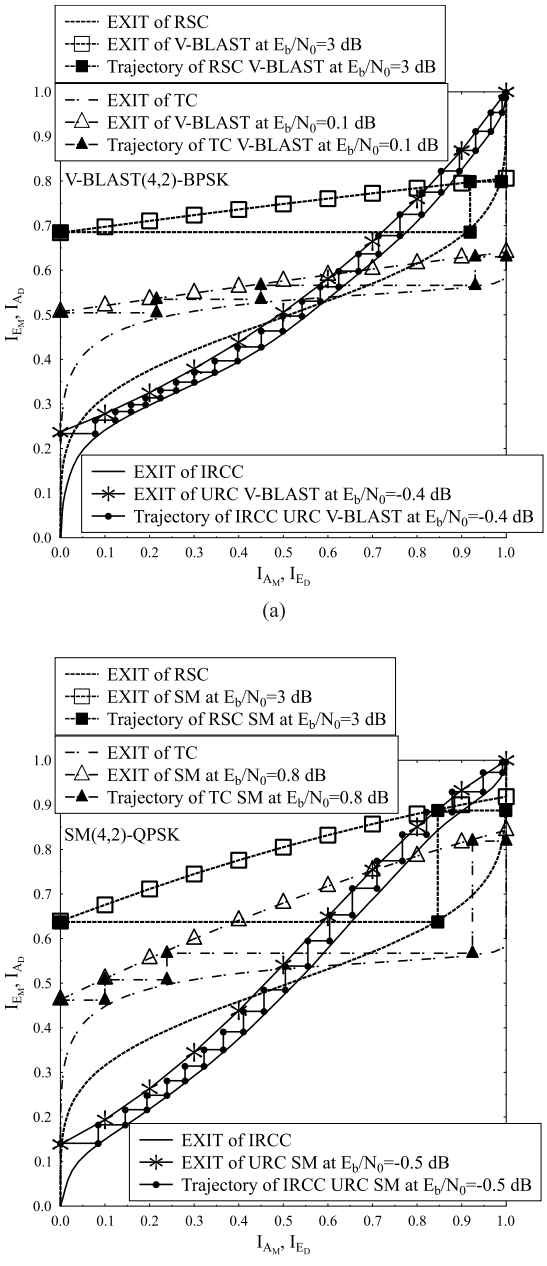


FIGURE 43. Decoding trajectories recorded for V-BLAST and SM in the context of RSC, TC and IRCC-URC coded systems associated with the same system throughput of $R_cR = 2$ bits/block/channel use. (a) V-BLAST(4,2)-BPSK. (b) SM(4,2)-QPSK.

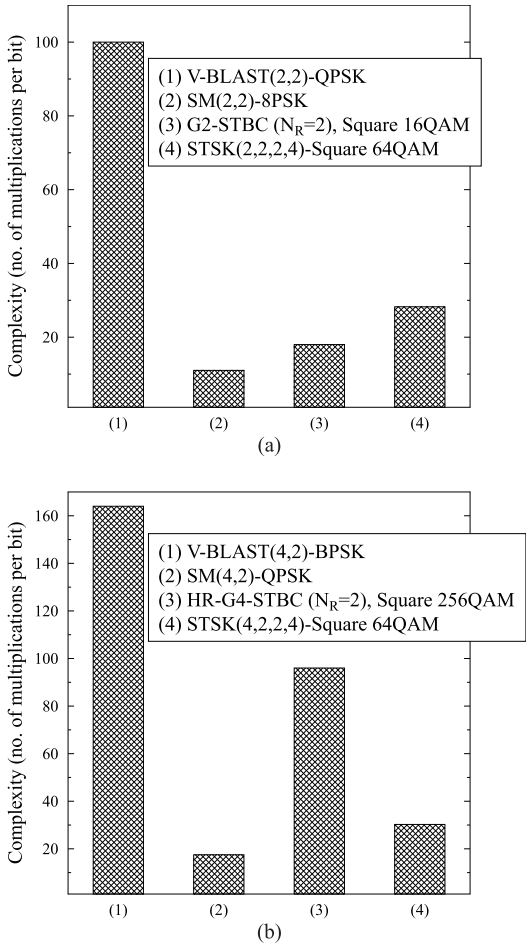


FIGURE 44. Complexity comparison between the soft-decision detectors conceived for V-BLAST, SM, STBC and STSK associated with the same throughput of $R = 4$ bits/block/channel use, where the fading channel envelope is assumed to be constant for the duration of a channel use. (a) $N_T = 2$, $N_R = 2$. (b) $N_T = 4$, $N_R = 2$.

G4 STBC arrangement performs the worst for both $N_R = 1$ and $N_R = 2$ in Fig. 41(b), because it has to employ a high-order 256QAM scheme in order to provide the same system throughput, and its diversity advantage exhibited at high SNRs is eroded in channel coded systems operating at relatively low SNRs.

Although the EXIT charts of Fig. 40 predict a similar detection capability for V-BLAST and SM, the BER performance of Fig. 41 demonstrates that SM performs worse than V-BLAST by about 0.8 dB in TC MIMO systems associated with the same throughput of $R_cR = 2$ bits/block/channel use. As seen in Fig. 43, TC associated with $IR_{TC} = 4$ exhibits a horizontal EXIT curve, which does not match well with the steep EXIT curves of V-BLAST and SM. In order to provide a more thorough comparison, Fig. 42 shows the BER performance of V-BLAST and SM in the context of RSC, TC and IRCC-URC coded systems, while the corresponding decoding trajectories are recorded in Fig. 43. It can be seen in Fig. 42(a) that RSC coded SM(2,2)-8PSK performs very close to RSC coded V-BLAST(2,2)-QPSK, while the performance difference between the IRCC-URC coded SM(2,2)-8PSK and IRCC-URC coded V-BLAST(2,2)-QPSK arrangements is only 0.3 dB. Furthermore, Fig. 42(b) shows that RSC coded SM(4,2)-QPSK outperforms RSC coded V-BLAST(4,2)-BPSK. This is because SM(4,2)-QPSK exhibits a higher iteration gain than V-BLAST(4,2)-BPSK, as demonstrated by Figs. 40 and 43, which benefits its performance in the context of RSC coded systems, because the EXIT curve of the RSC decoder is much steeper than that of the TC decoder. For the same reason, IRCC-URC coded

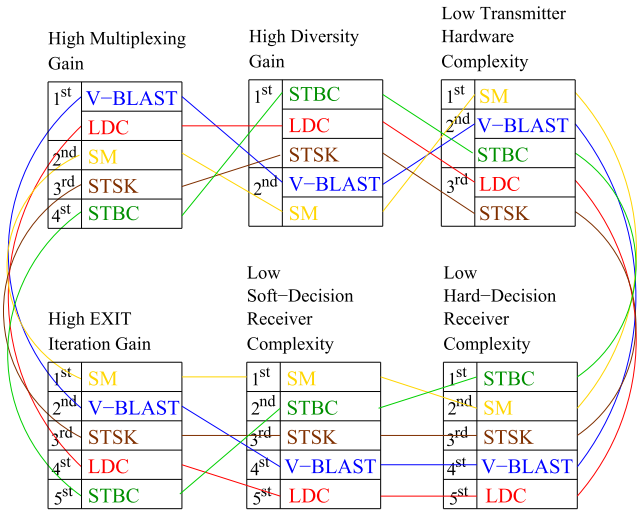


FIGURE 45. Comparison between MIMO schemes of different categories.

SM(4,2)-QPSK also slightly outperforms IRCC-URC coded V-BLAST(4,2)-BPSK, as evidenced by Fig. 42(b). In summary, we may conclude that SM is capable of achieving a comparable performance to V-BLAST in coded systems, provided that the appropriate channel coding schemes are selected.

Fig. 44 further compares the computational complexities of different soft-decision MIMO detectors. It can be seen in Fig. 44 that both the STBC and the STSK schemes exhibit a lower detection complexity than the conventional MIMO detector, but the SM detectors offer the lowest detection complexity in the context of coded MIMO systems. We note that in terms of the overall system complexity, the coded STBC system is the best, because the soft-decision STBC detector has to be invoked for a lower number of times, albeit only, because it benefits to a lesser extent from the *a priori* information, which suggests to limit the number of iterations. Nonetheless, for both cases of $N_T = 2$ and $N_T = 4$ in Fig. 44, the SM detection complexity is as low as 10% of the V-BLAST detection complexity, which offers a substantial reduction of both the signal processing complexity and the delay.

IV. CONCLUSIONS

A. SUMMARY

In this treatise, two key tradeoffs of MIMO systems were analysed. More explicitly, the first era of MIMO system design was fueled by the multiplexing-diversity tradeoff, where the associated V-BLAST, STBC and LDC schemes were introduced in Sec. II. By contrast, the second era of MIMO system design was predominantly motivated by the performance-complexity tradeoff. The corresponding SM and STSK arrangements were introduced in Sec. III. The MIMO schemes of different categories are compared in Fig. 45, where the 'tradeoff links' explicitly show that there is no dominant MIMO scheme that can be advantageous in all categories. The MIMO system design hinges on a delicate balance of complex issues that have to take into account wide-ranging factors in different system scenarios. Nonetheless, a conclusive summary of the advantages of different MIMO schemes is offered in Table 9 based on the discussions in this treatise.

TABLE 9. A brief summary of the advantages of differet MIMO schemes.

Full MIMO capacity:	V-BLAST, LDC
Full multiplexing gain ($\overline{R} = \frac{R}{\log_2 M} = N_T$):	V-BLAST, LDC
High multiplexing gain ($\overline{R} = \frac{R}{\log_2 M} > 1$):	SM, STSK
Full diversity gain ($\overline{D} = N_T N_R$):	STBC, LDC, STSK
Low PEP at high SNR region:	STBC, LDC, STSK
Low PEP at low SNR region (especially when $N_R > 1$ RAs and/or channel coding are used):	High throughput schemes employing low-level MPSK/QAM
High EXIT iteration gain:	V-BLAST, LDC, SM, STSK
Low ML/MAP receiver signal processing complexity:	STBC, SM, STSK
Low transmitter hardware complexity:	SM
Flexible MIMO systems design:	LDC, STSK

More explicitly, in terms of the MIMO's multiplexing feature seen in Table 9, both V-BLAST and LDC achieve the full MIMO capacity of (4), provided that the LDC's parameters satisfy $N_Q \geq N_T N_P$. Both V-BLAST and LDC are capable of achieving the full multiplexing gain of ($\overline{R} = \frac{R}{BPS} = N_T$), which leads to a system throughput that is N_T times higher than that of a SISO and SIMO system. Moreover, the SM and STSK arrangements may also achieve a multiplexing gain, hence their throughput is higher than that of their SISO, SIMO and STBC counterparts, but remains lower than that of V-BLAST.

In terms of the MIMO's diversity feature seen in Table 9, the three MIMO schemes of STBC, LDC and STSK are capable of attaining the full diversity order, which minimizes the PEP of (8) according to its rank criterion. The diversity-oriented MIMO schemes of STBC, LDC and STSK also aim for minimizing the PEP at high SNRs according to the determinant criterion. However, the full transmit diversity order of $\min\{N_T, N_P\}$ requires a high transmission duration of $N_P = N_T$. Consequently, the diversity-oriented MIMO schemes have to employ high-order MPSK/QAM in order to compensate for the time-expansion-induced throughput loss, which erodes their performance at low SNRs. As a result, when multiple RAs are used and especially when channel coding is applied, the high-throughput multiplexing-oriented MIMO schemes of V-BLAST and SM tend to perform better in the low-SNR region, which was extensively analysed in Sec. III-E.

In coded scenarios, all the four MIMO schemes of V-BLAST, LDC, SM and STSK are capable of producing an improved iteration gain, when the soft-decision MIMO detectors exchange extrinsic information with the channel decoders using turbo detection. The MIMO schemes exhibit

different advantages in the context of different coded systems, as shown in Sec. III-E3. For example, the SM schemes have a high iteration gain, and tend to perform best in RSC coded or in IRCC-URC coded systems. By contrast, the near-horizontal EXIT curves of the STBC is better matched to the EXIT curves of the TC, which results in good performance at low system complexity in TC coded systems. It is worth noting that when the channel coding arrangements are appropriately selected, all MIMO schemes may perform closer to their capacity limits. Therefore, the careful complexity profiling plays a salient role in coded MIMO system design.

For example, the reduced-scope-based SM detectors may performance similarly to the ML and MAP aided V-BLAST detectors in uncoded and coded systems, as seen in Secs. III-E2 and III-E3, respectively. Moreover, the reduced-scope-based SM detection complexity is comparable to the linear MMSE aided V-BLAST detection complexity, which is substantially lower than the ML/MAP aided V-BLAST detection complexity, despite the fact that the MMSE aided V-BLAST detectors impose a significant performance loss and they are generally designed for uplink MIMO systems associated with $N_T \leq N_R$. Similarly, both STBC and STSK may also rely on ML/MAP aided MIMO detection at a substantially reduced complexity, as summarized in Table 9.

Another important advantage of the SM scheme seen in Table 9 is its low transmitter hardware complexity, where only a single RF chain is activated. Moreover, the dispersion matrix generation of the LDC and STSK offers a more flexible MIMO system design, where the parameters of N_T, N_R, N_P and N_Q may be readily adjusted to any particular system requirement. However, it is worth noting that since the LDC and STSK matrices are randomly generated and optimized according to the PEP and to the DCMC capacity, the transmitted symbols are no longer drawn from the classic MPSK/QAM constellations, which imposes further strain on the MIMO transmitter's hardware complexity.

Furthermore, in this treatise, we offered a comprehensive survey of soft-decision MIMO detectors. Moreover, when both the channel's output signal and the *a priori* LLRs gleaned from the channel decoder are taken into account in (14), the soft-decision MAP aided MIMO detectors of Sec. II-A2 have to evaluate and compare all the MIMO combinations. As a result, the soft-decision MIMO detection may contribute a substantial fraction of the total complexity in coded systems. Against this background, we also highlighted the state-of-the-art reduced-complexity algorithms in Secs. II-A8 and III-A8, which are invoked by a wide range of soft-decision MIMO detectors, including the SD and LF aided V-BLAST and LDC detectors, the linear STBC detectors as well as the reduced-scope-based SM and STSK detectors. The rationale of the reduced-complexity algorithms is portrayed by Fig. 46.

More explicitly, the first step in Fig. 46 is to simplify the probability metric of (14). As a result, the soft-decision SD aided V-BLAST of Sec. II-A4, the LF aided V-BLAST of Sec. II-A7 and the linear STBC detection of Sec. II-B2

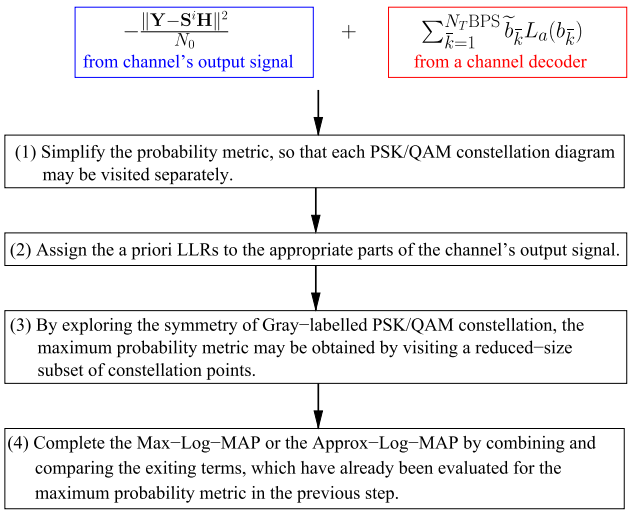


FIGURE 46. The schematic of the reduced-complexity soft-decision MIMO detection algorithm.

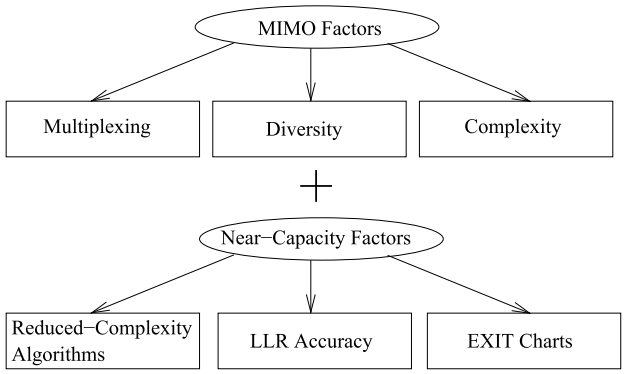


FIGURE 47. Design Guidelines for Near-capacity MIMO systems.

may separate the MIMO data streams, while the soft-decision SM detection of Sec. III-A8 may obtain N_T normalized matched filter output signals that correspond to the N_T separate MPSK/QAM candidates. In this way, each MPSK/QAM constellation diagram may be visited separately. The second step in Fig. 46 is to assign *a priori* LLRs to the appropriate parts of the channel's output signal, so that in the third step, the maximum probability metric may be directly obtained by visiting a reduced number of constellation points, thanks to the symmetry provided by Gray-labelling. In the end, the Max-Log-MAP algorithm of (15) and the Approx-Log-MAP algorithm of (16) may be completed based on the evaluations, which have already been calculated in the previous steps. The resultant complexity reduction was shown to be substantial in Secs. II-A8 and III-C, which is benefical especially when the soft-decision MIMO detectors are invoked several times for turbo detection in coded systems.

B. DESIGN GUIDELINES

Based on this treatise, we simplify and summarize the design guidelines for near-capacity MIMO systems as seen in Fig. 47, which are detailed as follows:

1) MIMO FACTOR OF MULTIPLEXING

For practical realizations, the multiplexing-oriented MIMO schemes of V-BLAST and LDC exhibit the advantage of a high throughput, which is N_T times higher than that of the SISO and SIMO arrangements. Therefore, the V-BLAST and LDC may employ lower-order PSK/QAM schemes, when they are compared to other MIMO schemes associated with the same throughput. The employment of low-order PSK/QAM results in a low PEP in the low-SNR region, which becomes more important, when multiple RAs ($N_R > 1$) are used and also when channel coding is applied. The MIMO schemes of SM and STSK may also exhibit a modest throughput gain over their SISO, SIMO and STBC counterparts.

2) MIMO FACTOR OF DIVERSITY

Both transmit diversity and receive diversity are benefical in both uncoded and coded systems. Again, the diversity-oriented MIMO schemes of STBC, LDC and STSK achieve the full transmit diversity order at the cost of requiring several time-slots. As a result, the STBC and STSK generally have to employ higher-order PSK/QAM schemes in order to compensate for their multiple-slot-induced throughput loss, which leads to a degraded performance in the low-SNR region. In summary, in terms of the performance of coded systems, which may rely on multiple RAs for achieving a useful receive diversity gain, the STBC and STSK schemes having the benefit of transmit diversity gain are no longer preferred over their higher-throughput MIMO counterparts.

3) MIMO FACTOR OF COMPLEXITY

In terms of the transmitter's hardware complexity, the SM is particularly advantageous, because only a single RF chain is required. By contrast, LDC and STSK impose a high hardware complexity on the MIMO transmitters, because the transmitted signals obtained from randomly generated dispersion matrices are no longer drawn from the classic PSK/QAM constellations.

In terms of the receiver's signal processing complexity, the STBC, SM and STSK arrangements do not suffer from the problem of IAI. This feature allows them to invoke low-complexity linear detectors without imposing any performance loss on the generic ML and MAP MIMO detectors in uncoded and coded systems, respectively. By contrast, when the SD and LF are invoked for V-BLAST and LDC, a performance-complexity tradeoff is encountered. Moreover, the SD and LF aided V-BLAST detectors are generally designed for uplink MIMO systems associated with $N_T \leq N_R$. For the rank-deficient MIMO systems associated with $N_T > N_R$, which are often encountered in realistic wireless communication systems, extra signal detection efforts are required. Furthermore, the suboptimal non-MAP soft-decision detectors are not particularly suitable for coded systems, because they tend to produce unreliable LLRs, which may mislead turbo detection. In summary, the low-complexity soft-decision STBC, SM and STSK detectors are the advantageous choices in coded systems.

4) NEAR-CAPACITY FACTOR OF EXIT CHARTS

The employment of EXIT charts is particularly beneficial for analysing the performance of coded MIMO systems, since they help to choose the appropriate channel coding arrangements for different MIMO schemes. For example, the SM schemes that have a high iteration gain may perform better in RSC coded or in IRCC-URC coded systems, as indicated by their EXIT charts. By contrast, the STBCs that have near-horizontal EXIT curves perform better in TC coded systems.

5) NEAR-CAPACITY FACTOR OF LLR ACCURACY

The LLR accuracy test constitutes an important tool in examining the reliability of the extrinsic LLRs produced by the soft-decision MIMO detectors. More explicitly, the extrinsic LLR definition of $L_e = \ln \frac{p(L_e|b=1)}{p(L_e|b=0)} = \ln \frac{p(b=1|L_e)}{p(b=0|L_e)}$ is supposed to be statistically true because of the consistency condition of $p(L_e|b = 1) = p(L_e|b = 0)e^{L_e}$, where the probabilities of $p(b = 1|L_e)$ and $p(b = 0|L_e)$ may be evaluated based on the histograms of the extrinsic LLRs produced by the soft-decision MIMO detector under investigation. A severe deviation from $L_e = \ln \frac{p(b=1|L_e)}{p(b=0|L_e)}$ implies that the soft-decision detector tested may produce large LLR values that deviate from the true probabilities of $\ln \frac{p(b=1|L_e)}{p(b=0|L_e)}$. These unreliable LLRs may mislead the turbo detection, because they may become more and more difficult to correct after a few iterations. In general, the generic soft-decision MAP aided MIMO detectors and their reduced-complexity variants used for STBC, SM and STSK may always guarantee to produce reliable LLRs, hence these options are preferred in coded MIMO systems.

6) NEAR-CAPACITY FACTOR OF REDUCED-COMPLEXITY ALGORITHMS

The generic soft-decision MAP aided MIMO detectors are preferred in turbo detection, because they are capable of producing reliable LLRs. However, the soft-decision MIMO detection may contribute a substantial fraction of the total complexity in coded systems, because all the MIMO combinations have to be examined, when both the channel's output signal and the *a priori* LLRs gleaned from the channel decoder are taken into account. Against this background, the reduced-complexity soft-decision MIMO detection algorithms may be carried out in two stages. First of all, the multiple data streams have to be separated, so that each individual PSK/QAM constellation may be visited separately. In this way, the conventional matrix-by-matrix-based signal processing, which directly deals with the MIMO signal matrices is transformed into symbol-by-symbol-based detection. Secondly, by exploring the symmetry provided by Gray-labelled constellations, the number of constellation points visited by the soft-decision detectors may be reduced. As a result, the symbol-by-symbol-based detection may be further simplified to bit-by-bit-based detection, where the uncorrelated groups of bits representing a reduced subset of constellation points are detected separately.

C. FUTURE RESEARCH

1) REDUCED-COMPLEXITY DESIGN APPLIED TO GENERALIZED SPATIAL MODULATION

As discussed in Sec. III-D, one of the major disadvantages of SM is its CCMC capacity loss compared to V-BLAST, which is explicitly demonstrated by Fig. 33(a). As the recent developments in the millimeter-wave band [13]–[15] allows us to accommodate a high number of antennas, especially at the base stations [15]–[18], the V-BLAST's full MIMO CCMC capacity may increase linearly with the number of antennas, as specified by (3). By contrast, the SM's CCMC capacity of (149) can only increase logarithmically with the number of antennas, because the maximum mutual information between the antenna activation index and the received signal formulated by (151)-(152) is upper-bounded by $\log_2 N_T$.

Therefore, the antenna activation procedure portrayed by the SM schematic of Fig. 12 may be modified in order to convey more information bits. From a historic perspective, the fractional-bit based SM proposed in [255] allows the transmitter to employ any arbitrary number of antennas N_T instead of requiring N_T being a power of 2. Specifically, when N_T is not a power of 2, the antenna activation index may carry a variable numbers of bits, where some antenna index candidates are encoded by $(\lfloor \log_2 N_T \rfloor)$ bits, while other candidates are encoded by $(\lfloor \log_2 N_T \rfloor + 1)$ bits. However, the variable number of bits assigned for antenna activation may lead to an error propagation problem, when the antenna index and the modulation index are detected separately at the SM receiver. As a remedy, a bit-padding method was introduced in [256], where an extra bit was attached at the end of the short codewords so that all antenna activation index candidates may convey the same number of $(\lfloor \log_2 N_T \rfloor + 1)$ bits.

The earliest effort to assign more bits to the antenna activation procedure is constituted by the Generalized Space-Shift Keying (GSSK) philosophy [135], where more than one transmit antennas are activated. More explicitly, when a total number of n_t out of N_T transmit antennas are activated, the total number of possible combinations is given by the binomial coefficient of $\overline{U}_T = \binom{N_T}{n_t}$. As a result, the total number of bits that can be conveyed by GSSK is given by $\text{BPS}_T = \lfloor \log_2 \overline{U}_T \rfloor = \log_2 U_T$, where there are a total of $U_T = 2^{\text{BPS}_T}$ GSSK codewords. Therefore, SSK [136] constitutes a special case of GSSK, where we have $n_t = 1$ and $U_T = N_T$. Furthermore, the Generalized Spatial Modulation (GSM) [164] may rely on the same antenna activation procedure as GSSK, while the same MPSK/QAM symbol conveying $BPS = \log_2 M$ bits is transmitted by all the activated antennas, so that GSM still retains the advantage of no inter-antenna interference. As a result, the SM of Sec. III also becomes a special case of the GSM associated with $n_t = 1$ and $U_T = N_T$. Moreover, the total number of bits assigned to the antenna activation procedure was further improved by the Hamming code aided techniques of [257].

The GSM proposed in [160] opted for using the n_t out of N_T activated antennas to transmit n_t different MPSK/QAM symbols, so that the total number of bits conveyed by the GSM scheme may be increased to $\text{BPS}_T + n_t\text{BPS}$, where we have $\text{BPS}_T = \lfloor \log_2 [\binom{N_T}{n_t}] \rfloor$. Similarly, the GSTSK scheme proposed in [151] and [254] also activates n_q out of N_Q dispersion matrices to spread a total of n_q different MPSK/QAM symbols. Naturally, the GSM of [160] has a higher capacity than the GSM of [164], [258]. However, the inter-antenna interference problem arises again both for the GSM of [160] and also for the GSTSK of [151], [254]. As a remedy, sub-optimal reduced-interference receivers were proposed for the GSM and GSTSK receivers in [160] and [162], which are not consistent with the SM/STSK motivation of low-complexity single-stream ML receiver design. Against this background, the preferred GSM arrangement for future research is the one conceived in [164] and [258], where the n_t activated antennas convey the same MPSK/QAM symbol and hence there is once again no IAI. Ideally, the GSM and the corresponding GSTSK schemes should invoke the low-complexity SM detectors of Sec. III.

There are also a pair of hard-decision suboptimal GSM detectors in the literature [163], [260] which may achieve a near-optimum performance in uncoded GSM systems. Moreover, it is demonstrated in [163] and [258] that GSM may outperfom SM, but the simulation results of [164] indicate the opposite. Therefore, the detailed capacity and performance comparison of SM and GSM - especially in the context of coded systems - is still aside for future work. Moreover, the above reduced-complexity design guidelines are also applicable to soft-decision GSM detectors.

2) REDUCED-COMPLEXITY DESIGN APPLIED TO DIFFERENTIAL MIMO SCHEMES

Many recently developed communication systems demonstrated growing interest in employing Differential Space-Time Modulation (DSTM), which facilitates sophisticated signal processing both in the spatial and in the temporal dimensions, while the high-complexity requirement of accurate channel estimation is eliminated. For example, the cooperative communication systems of [261]–[264] opted for employing single-element mobile stations cooperatively sharing their antennas, so that a Virtual Antenna Array (VAA) may be formed for MIMO transmission, where the distributed antennas typically experience uncorrelated fading. As a result, it becomes unrealistic for the relays and the destinations to estimate the channel of all VAA links, hence the employment of DSTM may be perferred.

The first DSTM scheme was proposed by Tarokh and Jafarkhani [265] in 1999, where the Differential STBC (DSTBC) based on the G2 STBC structure employing MPSK signalling was conceived. More explicitly, the DSTBC of [265] proposed to employ the G2 STBC codeword for both the data-carrying matrix and the transmission matrix, where MPSK signalling was used for all the transmitted symbols. In order to also retain the MPSK signalling for the data-carrying symbols in $\mathbf{X}_{n-1}$, Hughes [266] proposed the family of group codes in 2000, where however a throughput loss was

encountered. In 2001, Jafarkhani and Tarokh [267] further extended the DSTBC [265] to the case of employing multiple transmit antennas based on the general STBC structure of Sec. II-B5. Furthermore, the Differential LDC (DLDC) philosophy was proposed by Hassibi and Hochwald [268] in 2002, where the Cayley transform was invoked in order to guarantee that the data-carrying matrix $\mathbf{X}_{n-1}$ is always unitary. In 2003 and 2004, Hwang *et al.* [269] and Nam *et al.* [270] proposed to employ QAM for the DSTBCs of [265] and [267], respectively. Moreover, Wang *et al.* [271] suggested in 2005 that high-rate DLDCs may be generated with the aid of a gradient-ascent method. Oggier and Hassibi [272] suggested in 2007 that the high rate DLDCs may also be constructed based on division algebra, so that the Cayley codes may be expressed in closed-form. As mentioned before, the DSTSK scheme was proposed by Sugiura *et al.* [39] in 2010 together with the proposal of the STSK, while the DLDC's Cayley transformed was eliminated by Xu *et al.* [259] in 2011. In 2013, the concept of Differential Spatial Modulation (DSM) was proposed by Bian *et al.* [273], [274] based on the DSTM concept, which was then further improved by Ishikawa and Sugiura [275] in 2014. The Star-QAM aided DSM was proposed by Martin [276] in 2015. Naturally, the multiplexing-diversity tradeoff and the performance-complexity tradeoff also exist in the context of the DSTM schemes of DSTBC, DLDC, DSTSK and DSM.

The noncoherent receivers conceived for SISO/SIMO schemes including Multiple-Symbol Differential Detection (MSDD), Multiple-Symbol Differential Sphere Detection (MSDSD) and Decision-Feedback Differential Detection (DFDD) have also been developed for the DSTM. More explicitly, the MSDD and the DFDD were firstly developed for DSTM by Schober and Lampe [277] in 2002. Furthermore, Pauli and Lampe [192] proposed the MSDSD concept for DSTM employing MPSK in 2007. In 2011, the MSDSD conceived for DSTBC employing MQAM was developed by Xu *et al.* [278], and then the reduced-complexity MSDSD conceived for DSTSK was proposed by Xu *et al.* [259]. Furthermore, notably, the two-dimensional MSDSD aided Differential Space-Frequency Modulation (DSFM) was developed for OFDM systems by Pauli *et al.* [194] in 2008. Moreover, the DFDD and MSDSD aided DSTM were also conceived for multi-user scenarios by Cheung and Schober [279] as well as by Wang and Hanzo [280] in 2006 and 2011, respectively.

Further research efforts invested in noncoherent DSTM detection may be deemed to be three-fold. First of all, the aforementioned noncoherent receivers all rely on hard-bit decisions, hence a thorough study of soft-decision-aided noncoherent detectors conceived for coded DSTM is still awaited. Second, it was noted in [259] and [278] that owing to the associated matrix-based signal processing, the noncoherent DSTM detectors generally exhibit a higher complexity than their noncoherent DPSK detector counterparts. Therefore, a systematic reduced-complexity design is needed for noncoherent DSTM detection. Thirdly, the recent development of massive MIMOs [16], [17] demonstrated the interesting result that the linear MMSE MIMO detector may become near-optimum, when the number of antennas is increased substantially. Since accurate channel estimation may become a challenge in massive MIMO systems, the employment of DSTM may be preferred. Against this background, one may predict that the optimum MSDD/MSDSD may become hardly affordable, when a massive number of antennas are employed. Hence the DFDD philosophy may become the preferred choice.

REFERENCES

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2. Geneva, Switzerland, May 1993, pp. 1064–1070.

[2] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[3] S. Benedetto and G. Montorsi, "Serial concatenation of block and convolutional codes," *Electron. Lett.*, vol. 32, no. 10, pp. 887–888, May 1996.

[4] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[5] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 32, no. 18, p. 1645, Aug. 1996.

[6] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[7] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[8] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 169–171, Nov. 1997.

[9] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 715–724, Apr. 1999.

[10] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2. Jun. 1999, pp. 858–863.

[11] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8 PSK signaling," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.

[12] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation*. New York, NY, USA: Wiley, May 2009.

[13] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[14] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, Jr., "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[15] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proc. IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.

[16] 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.

[17] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[18] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, pp. 186–195, Feb. 2014.

[19] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.

[20] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.
[21] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.
[22] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.
[23] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications—A key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.
[24] D. Gore, R. W. Heath, and A. Paulraj, "On performance of the zero forcing receiver in presence of transmit correlation," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2002, p. 159.
[25] D. Wubben, R. Bohnke, V. Kuhn, and K. D. Kammeyer, "MMSE extension of V-BLAST based on sorted QR decomposition," in *Proc. IEEE 58th Veh. Technol. Conf. (VTC-Fall)*, vol. 1. Oct. 2003, pp. 508–512.
[26] S. Song, Y. Yang, Q. Xionq, K. Xie, B. J. Jeong, and B. Jiao, "A channel hopping technique I: Theoretical studies on band efficiency and capacity," in *Proc. Int. Conf. Commun., Circuits Syst. (ICCCAS)*, vol. 1. Jun. 2004, pp. 229–233.
[27] 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.
[28] P. Yang, M. Di Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design guidelines for spatial modulation," *IEEE Commun. Surveys Tuts.*, vol. 17, pp. 6–26, no. 1, pp. 6–26, 1st Quart., 2015.
[29] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[30] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
[31] H. E. Gamal and A. R. Hammons, "A new approach to layered space-time coding and signal processing," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2321–2334, Sep. 2001.
[32] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Syst., Electron.*, Oct. 1998, pp. 295–300.
[33] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.
[34] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
[35] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.
[36] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, Jul. 2002.
[37] S. Sandhu and A. Paulraj, "Unified design of linear space-time block codes," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 2. Nov. 2001, pp. 1073–1077.
[38] R. W. Heath and A. J. Paulraj, "Linear dispersion codes for MIMO systems based on frame theory," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2429–2441, Oct. 2002.
[39] 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.
[40] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[41] R. W. Hamming, "Error detecting and error correcting codes," *Bell Syst. Tech. J.*, vol. 29, no. 2, pp. 147–160, Apr. 1950.
[42] P. Elias, "Coding for noisy channels," *Proc. Inst. Radio Eng.*, vol. 43, no. 3, p. 356, 1955.
[43] A. Hocquenghem, "Codes correcteurs derreurs," *Chiffres*, vol. 2, pp. 147–156, Sep. 1959.
[44] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Inf. Control*, vol. 3, no. 1, pp. 68–79, Mar. 1960.
[45] R. C. Bose and D. K. Ray-Chaudhuri, "Further results on error correcting binary group codes," *Inf. Control*, vol. 3, pp. 279–290, Sep. 1960.
[46] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.
[47] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
[48] J. Wolf, "Efficient maximum likelihood decoding of linear block codes using a trellis," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 76–80, Jan. 1978.
[49] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.
[50] W. Koch and A. Baier, "Optimum and sub-optimum detection of coded data disturbed by time-varying intersymbol interference," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 3. Dec. 1990, pp. 1679–1684.
[51] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2. Jun. 1995, pp. 1009–1013.
[52] Y. Yasuda, K. Kashiki, and Y. Hirata, "High-rate punctured convolutional codes for soft decision Viterbi decoding," *IEEE Trans. Commun.*, vol. 32, no. 3, pp. 315–319, Mar. 1984.
[53] G. D. Forney, *Concatenated Codes*. Cambridge, MA, USA: MIT Press, 1966.
[54] P. Elias, "Error-free Coding," *Trans. IRE Prof. Group Inf. Theory*, vol. 4, no. 4, pp. 29–37, Sep. 1954.
[55] A. Morello, G. Montorosi, and M. Visintin, "Convolutional and Trellis coded modulations concatenated with block codes for digital HDTV," in *Proc. Int. Workshop Digit. Commun.*, Tirennia, Italy, Sep. 1993, pp. 237–250.
[56] E. C. Posner, L. L. Rauch, and B. D. Madsen, "Voyager mission telecommunication firsts," *IEEE Commun. Mag.*, vol. 28, no. 9, pp. 22–27, Sep. 1990.
[57] G. Battail, M. Decouvelaere, and P. Godlewski, "Replication decoding," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 332–345, May 1979.
[58] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in *Proc. IEEE Global Telecommun. Conf. Exhibit.*, vol. 3. Nov. 1989, pp. 1680–1686.
[59] J. Lodge, P. Hoeher, and J. Hagenauer, "The decoding of multidimentional codes using separable MAP 'filters,'" in *Proc. 16th Biennial Symp. Commun.*, May 1992, pp. 343–346.
[60] J. Lodge, R. Young, P. Hoeher, and J. Hagenauer, "Separable MAP 'filters' for the decoding of product and concatenated codes," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 3. pp. 1740–1745, Geneva, May 1993.
[61] G. D. Forney, "Coset codes II: Binary lattices and related codes," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1152–1187, Sep. 1988.
[62] R. Pyndiah, A. Glavieux, A. Picart, and S. Jacq, "Near optimum decoding of product codes," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 1994, pp. 339–343.
[63] R. M. Pyndiah, "Near-optimum decoding of product codes: Block turbo codes," *IEEE Trans. Commun.*, vol. 46, no. 8, pp. 1003–1010, Aug. 1998.
[64] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.
[65] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.
[66] J. G. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 1995.
[67] B. Sklar, *Digital Communications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.
[68] P. Mecklenburg, W. Pehlert, and D. Sullivan, "Correction of errors in multilevel Gray coded data," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, pp. 336–340, May 1973.
[69] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.
[70] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.
[71] A. R. Calderbank and N. J. A. Sloane, "Four-dimensional modulation with an eight-state trellis code," *AT T Tech. J.*, vol. 64, no. 5, pp. 1005–1018, May 1985.
[72] A. Calderbank and N. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inf. Theory*, vol. 33, no. 2, pp. 177–195, Mar. 1987.

[73] L.-F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inf. Theory*, vol. 33, no. 4, pp. 483–501, Jul. 1987.

[74] J. G. Forney, R. G. Gallager, G. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, Sep. 1984.

[75] G. J. Pottie and D. P. Taylor, "Multilevel codes based on partitioning," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 87–98, Jan. 1989.

[76] A. R. Calderbank, "Multilevel codes and multistage decoding," *IEEE Trans. Commun.*, vol. 37, no. 3, pp. 222–229, Mar. 1989.

[77] S. Sayegh, "A class of optimum block codes in signal space," *IEEE Trans. Commun.*, vol. 34, no. 10, pp. 1043–1045, Oct. 1986.

[78] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channels: Performance criteria," *IEEE Trans. Commun.*, vol. 36, no. 9, pp. 1004–1012, Sep. 1988.

[79] M. K. Simon and D. Divsalar, "The performance of trellis coded multilevel DPSK on a fading mobile satellite channel," *IEEE Trans. Veh. Technol.*, vol. 37, no. 2, pp. 78–91, May 1988.

[80] N. Seshadri and C.-E. W. Sundberg, "Multilevel trellis coded modulations for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 41, no. 9, pp. 1300–1310, Sep. 1993.

[81] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.

[82] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[83] J. Hagenauer, "The turbo principle: Tutorial introduction and state of the art," in *Proc. Int. Symp. Turbo Codes Rel. Topics*, 1997, pp. 1–11.

[84] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2449–2457, Sep. 1995.

[85] C. Douillard *et al.*, "Iterative correction of intersymbol interference: Turbo equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, 1995.

[86] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.

[87] M. Tuchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using *a priori* information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.

[88] M. Moher, "An iterative multiuser decoder for near-capacity communications," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 870–880, Jul. 1998.

[89] M. C. Reed, C. B. Schlegel, P. D. Alexander, and J. A. Asenstorfer, "Iterative multiuser detection for CDMA with FEC: Near-single-user performance," *IEEE Trans. Commun.*, vol. 46, no. 12, pp. 1693–1699, Dec. 1998.

[90] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.

[91] M. L. Honig, G. K. Woodward, and Y. Sun, "Adaptive iterative multiuser decision feedback detection," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 477–485, Mar. 2004.

[92] M. Sellathurai and S. Haykin, "Turbo-BLAST: Performance evaluation in correlated Rayleigh-fading environment," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 340–349, Apr. 2003.

[93] C. Studer, S. Fateh, and D. Seethaler, "ASIC implementation of soft-input soft-output MIMO detection using MMSE parallel interference cancellation," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1754–1765, Jul. 2011.

[94] M. Moher, "Decoding via cross-entropy minimization," in *Proc. IEEE Houston GLOBECOM*, vol. 2. Nov. 1993, pp. 809–813.

[95] M. Moher and T. A. Gulliver, "Cross-entropy and iterative decoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 3097–3104, Nov. 1998.

[96] S. ten Brink, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, no. 10, pp. 806–808, May 1999.

[97] S.-Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 657–670, Feb. 2001.

[98] H. E. Gamal and A. R. Hammons, "Analyzing the turbo decoder using the Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 671–686, Feb. 2001.

[99] M. Tuchler, "Design of serially concatenated systems depending on the block length," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 209–218, Feb. 2004.

[100] S. ten Brink, "Iterative decoding trajectories of parallel concatenated codes," in *Proc. 3rd IEEE/ITG Conf. Sour. Channel Coding*, Jan. 2000, pp. 75–80.

[101] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[102] I. Land, P. Hoeher, and S. Gligorevic, "Computation of symbol-wise mutual information in transmission systems with log APP decoders and application to EXIT charts," in *Proc. Int. ITG Conf. Sour. Channel Coding*, 2004, pp. 195–202.

[103] J. Kliewer, S. X. Ng, and L. Hanzo, "Efficient computation of EXIT functions for nonbinary iterative decoding," *IEEE Trans. Commun.*, vol. 54, no. 12, pp. 2133–2136, Dec. 2006.

[104] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[105] L. Hanzo, R. G. Maunder, J. Wang, and L. Yang, *Efficient Computation of EXIT Functions for Nonbinary Iterative Decoding*. New York, NY, USA: Wiley, 2011.

[106] D. Divsalar, S. Dolinar, and F. Pollara, "Serial concatenated Trellis coded modulation with rate-1 inner code," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 2. Nov. 2000, pp. 777–782.

[107] A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," U.S. Patent 5 345 599, Sep. 6, 1994.

[108] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[109] E. Telatar, "Capacity of multiantenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[110] Y. C. Yoon, R. Kohno, and H. Imai, "A spread-spectrum multiaccess system with cochannel interference cancellation for multipath fading channels," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, pp. 1067–1075, Sep. 1993.

[111] A. L. C. Hui and K. Letaief, "Successive interference cancellation for multiuser asynchronous DS/CDMA detectors in multipath fading links," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 384–391, Mar. 1998.

[112] C. A. Belfiore and J. H. Park, "Decision feedback equalization," *Proc. IEEE*, vol. 67, no. 8, pp. 1143–1156, Aug. 1979.

[113] J. Salz, "Optimum mean-square decision feedback equalization," *Bell System Tech. J., The*, vol. 52, no. 8, pp. 1341–1373, Oct. 1973.

[114] D. D. Falconer and G. J. Foschini, "Theory of minimum mean-square-error QAM systems employing decision feedback equalization," *Bell System Tech. J., The*, vol. 52, no. 10, pp. 1821–1849, Dec. 1973.

[115] O. Damen, A. Chkeif, and J. C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161–163, May 2000.

[116] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[117] H. Vikalo, B. Hassibi, and T. Kailath, "Iterative decoding for MIMO channels via modified sphere decoding," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2299–2311, Nov. 2004.

[118] C. Studer, A. Burg, and H. Bolcskei, "Soft-output sphere decoding: Algorithms and VLSI implementation," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 290–300, Feb. 2008.

[119] C. Studer and H. Bolcskei, "Soft-input soft-output single tree-search sphere decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4827–4842, Sep. 2010.

[120] A. V. Geramita and J. Seberry, *Orthogonal Designs: Quadratic Forms Hadamard Matrices*. New York, NY, USA: Marcel Dekker, 1979.

[121] J. Radon, "Lineare scharen orthogonaler matrizen," *Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg*, vol. 1. Berlin, Germany: Springer, 1922.

[122] G. Ganesan and P. Stoica, "Space-time diversity using orthogonal and amicable orthogonal designs," *Wireless Pers. Commun.*, vol. 18, no. 2, pp. 165–178, 2001.

[123] G. Ganesan and P. Stoica, "Space-time block codes: A maximum SNR approach," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1650–1656, May 2001.

[124] G. Ganesan and P. Stoica, "Differential modulation using space-time block codes," in *Proc. 35th Asilomar Conf. Signals, Syst. Comput.*, vol. 1. Nov. 2001, pp. 236–240.

[125] S. Sandhu and A. Paulraj, "Space-time block codes: A capacity perspective," *IEEE Commun. Lett.*, vol. 4, no. 12, pp. 384–386, Dec. 2000.
[126] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.
[127] C. B. Papadias and G. J. Foschini, "Capacity-approaching space-time codes for systems employing four transmitter antennas," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 726–732, Mar. 2003.
[128] B. A. Sethuraman, B. S. Rajan, and V. Shashidhar, "Full-diversity, high-rate space-time block codes from division algebras," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2596–2616, Oct. 2003.
[129] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H. F. Lu, "Explicit space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3869–3884, Sep. 2006.
[130] M. O. Damen, A. Tewfik, and J. C. Belflore, "A construction of a space-time code based on number theory," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 753–760, Mar. 2002.
[131] H. E. Gamal and M. O. Damen, "Universal space-time coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.
[132] F. Oggier, G. Rekaya, J. C. Belfiore, and E. Viterbo, "Perfect space-time block codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 3885–3902, Sep. 2006.
[133] Y. A. Chau and S. H. Yu, "Space modulation on wireless fading channels," in *Proc. IEEE VTS 54th Veh. Technol. Conf. (VTCFall)*, vol. 3. Oct. 2001, pp. 1668–1671.
[134] 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.
[135] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Generalized space shift keying modulation for MIMO channels," in *Proc. IEEE 19th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Cannes, France, Sep. 2008, pp. 1–5.
[136] 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.
[137] 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.
[138] C. Xu, S. Sugiura, S. X. Ng, and L. Hanzo, "Reduced-complexity soft-decision aided space-time shift keying," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 547–550, Oct. 2011.
[139] C. Xu, S. Sugiura, S. X. Ng, and L. Hanzo, "Spatial modulation and space-time shift keying: Optimal performance at a reduced detection complexity," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 206–216, Jan. 2013.
[140] R. Rajashekar, K. V. S. Hari, and L. Hanzo, "Reduced-complexity ML detection and capacity-optimized training for spatial modulation systems," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 112–125, Jan. 2014.
[141] M. Guo, C. Jia, and Y. Shen, "Detection algorithm for spatial modulation system under unconstrained channel," in *Proc. 12th IEEE Int. Conf. Commun. Technol. (ICCT)*, Nanjing, China, Nov. 2010, pp. 458–461.
[142] N. R. Naidoo, H. J. Xu, and T. A.-M. Quazi, "Spatial modulation: Optimal detector asymptotic performance and multiple-stage detection," *IET Commun.*, vol. 5, no. 1, pp. 1368–1376, Jul. 2011.
[143] J. Wang, S. Jia, and J. Song, "Signal vector based detection scheme for spatial modulation," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 19–21, Jan. 2012.
[144] J. Zheng, "Signal vector based list detection for spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 265–267, Aug. 2012.
[145] S. Sugiura, C. Xu, S. X. Ng, and L. Hanzo, "Reduced-complexity coherent versus non-coherent QAM-aided space-time shift keying," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3090–3101, Nov. 2011.
[146] P. Yang, Y. Xiao, L. Li, Q. Tang, and S. Li, "An improved matched-filter based detection algorithm for space-time shift keying systems," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 271–274, May 2012.
[147] Q. Tang, Y. Xiao, P. Yang, Q. Yu, and S. Li, "A new low-complexity near-ML detection algorithm for spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 90–93, Feb. 2013.
[148] A. Younis, M. Di Renzo, R. Mesleh, and H. Haas, "Sphere decoding for spatial modulation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011, pp. 1–6.
[149] A. Younis, S. Sinanovic, M. Di Renzo, R. Mesleh, and H. Haas, "Generalised sphere decoding for spatial modulation," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2805–2815, Jul. 2013.
[150] A. Younis, R. Mesleh, H. Haas, and P. M. Grant, "Reduced complexity sphere decoder for spatial modulation detection receivers," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–5.
[151] 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.
[152] 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.
[153] R. Rajashekar and K. V. S. Hari, "Modulation diversity for spatial modulation using complex interleaved orthogonal design," in *Proc. IEEE Region Conf.*, Nov. 2012, pp. 1–6.
[154] D. Yang, C. Xu, L. L. Yang, and L. Hanzo, "Transmit-diversity-assisted space-shift keying for colocated and distributed/cooperative MIMO elements," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2864–2869, Jul. 2011.
[155] M. T. Le, V. D. Ngo, H. A. Mai, X. N. Tran, and M. Di Renzo, "Spatially modulated orthogonal space-time block codes with non-vanishing determinants," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 85–99, Jan. 2014.
[156] M. Di Renzo and H. Haas, "Space shift keying (SSK̊) 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.
[157] D. M. Renzo, D. D. Leonardis, F. Graziosi, and H. Haas, "Space shift keying (SSK̊) MIMO with practical channel estimates," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 998–1012, Apr. 2012.
[158] M. D. Renzo and H. Haas, "On transmit diversity for spatial modulation MIMO: Impact of spatial constellation diagram and shaping filters at the transmitter," *IEEE Trans. Veh. Technol.*, vol. 62, no. 6, pp. 2507–2531, Jul. 2013.
[159] L. L. Yang, "Signal detection in antenna-hopping space-division multiple-access systems with space-shift keying modulation," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 351–366, Jan. 2012.
[160] J. Wang, S. Jia, and J. Song, "Generalised 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.
[161] R. 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.
[162] S. Sugiura, C. Xu, S. X. Ng, and L. Hanzo, "Reduced-complexity iterative-detection-aided generalized space-time shift keying," *IEEE Trans. Veh. Technol.*, vol. 61, no. 8, pp. 3656–3664, Oct. 2012.
[163] J. Fu, C. Hou, W. Xiang, L. Yan, and Y. Hou, "Generalised spatial modulation with multiple active transmit antennas," in *Proc. IEEE GLOBECOM Workshops (GLOBECOME)*, Dec. 2010, pp. 839–844.
[164] A. Younis, N. Serafimovski, R. Mesleh, and H. Haas, "Generalised spatial modulation," in *Proc. The 44th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2010, pp. 1498–1502.
[165] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[166] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.
[167] S. X. Ng and L. Hanzo, "On the MIMO channel capacity of multi-dimensional signal sets," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 528–536, Mar. 2006.
[168] M. I. Irshid and I. S. Salous, "Bit error probability for coherent M-ary PSK systems," *IEEE Trans. Commun.*, vol. 39, no. 3, pp. 349–352, Mar. 1991.
[169] 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.
[170] S. Sugiura and L. Hanzo, "On the joint optimization of dispersion matrices and constellations for near-capacity irregular precoded space-time shift keying," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 380–387, Jan. 2013.
[171] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[172] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 199–207, Feb. 1999.

[173] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 527–537, Apr. 1999.

[174] S. ten Brink, J. Speidel, and R.-H. Han, "Iterative demapping for QPSK modulation," *Electron. Lett.*, vol. 34, no. 15, pp. 1459–1460, Jul. 1998.

[175] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 1. Nov. 1998, pp. 579–584.

[176] A. Stefanov and T. M. Duman, "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 958–968, May 2001.

[177] L. Hanzo, T. Liew, B. Yeap, R. Tee, and S. X. Ng, *Turbo Coding, Turbo Equalisation Space-Time Coding: EXIT-Chart-Aided Near-Capacity Designs for Wireless Channels*. Hoboken, NJ, USA: Wiley, 2011.

[178] M. O. Damen, K. Abed-Meraim, and S. Burykh, "Iterative QR detection for BLAST," *Wireless Pers. Commun.*, vol. 19, no. 3, pp. 179–191, 2001.

[179] M. O. Damen, K. Abed-Meraim, and J. C. Belfiore, "A generalized lattice decoder for asymmetrical space-time communication architecture," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5. Oct. 2000, pp. 2581–2584.

[180] M. O. Damen, K. Abed-Meraim, and J. C. Belfiore, "Generalised sphere decoder for asymmetrical space-time communication architecture," *Electron. Lett.*, vol. 36, no. 2, pp. 166–167, Jan. 2000.

[181] L. Hanzo, M. El-Hajjar, and O. Alamri, "Near-capacity wireless transceivers and cooperative communications in the MIMO era: Evolution of standards, waveform design, and future perspectives," *Proc. IEEE*, vol. 99, no. 8, pp. 1343–1385, Aug. 2011.

[182] Q. Li *et al.*, "MIMO techniques in WiMAX and LTE: A feature overview," *IEEE Commun. Mag.*, vol. 48, no. 5, pp. 86–92, May 2010.

[183] M. Pohst, "On the computation of lattice vectors of minimal length, successive minima and reduced bases with applications," *ACM Sigsam Bull.*, vol. 15, no. 1, pp. 37–44, 1981.

[184] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, no. 170, pp. 463–471, Apr. 1985.

[185] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Program.*, vol. 66, no. 1, pp. 181–191, 1994.

[186] K. W. Wong, C. Y. Tsui, R. S. K. Cheng, and W. H. Mow, "A VLSI architecture of a K-best lattice decoding algorithm for MIMO channels," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, vol. 3. May 2002, pp. 273–276.

[187] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-Best sphere decoding for MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 491–503, Mar. 2006.

[188] M. Wenk, M. Zellweger, A. Burg, N. Felber, and W. Fichtner, "K-best MIMO detection VLSI architectures achieving up to 424 Mbps," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2006, p. 1154.

[189] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bolcskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE J. Solid-State Circuits*, vol. 40, no. 7, pp. 1566–1577, Jul. 2005.

[190] L. Lampe, R. Schober, V. Pauli, and C. Windpassinger, "Multiple-symbol differential sphere decoding," *IEEE Trans. Commun.*, vol. 53, no. 12, pp. 1981–1985, Dec. 2005.

[191] V. Pauli, L. Lampe, and R. Schober, "'Turbo DPSK' using soft multiple-symbol differential sphere decoding," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1385–1398, Apr. 2006.

[192] V. Pauli and L. Lampe, "On the complexity of sphere decoding for differential detection," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1595–1603, Apr. 2007.

[193] V. Pauli and L. Lampe, "Tree-search multiple-symbol differential decoding for unitary space-time modulation," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1567–1576, Aug. 2007.

[194] V. Pauli, L. Lampe, and J. Huber, "Differential space-frequency modulation and fast 2-D multiple-symbol differential detection for MIMO-OFDM," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 297–310, Jan. 2008.

[195] Q. Liu and L. Yang, "A novel method for initial radius selection of sphere decoding," in *Proc. IEEE 60th Veh. Technol. Conf. (VTC-Fall)*, vol. 2. Sep. 2004, pp. 1280–1283.

[196] S. Baro, J. Hagenauer, and M. Witzke, "Iterative detection of MIMO transmission using a list-sequential (LISS) detector," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4. May 2003, pp. 2653–2657.

[197] C. Xu, X. Zuo, S. Ng, R. Maunder, and L. Hanzo, "Reduced-complexity soft-decision multiple-symbol differential sphere detection," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3275–3289, Sep. 2015.

[198] J. Jalden and B. Ottersten, "Parallel implementation of a soft output sphere decoder," in *Proc. 39th Asilomar Conf. Signals, Syst. Comput. Conf. Rec.*, Oct. 2005, pp. 581–585.

[199] E. Telatar, "Capacity of multiantenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[200] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, Aug. 2005.

[201] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[202] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[203] G. D. Golden, C. J. Foschini, R. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, no. 1, pp. 14–16, Jan. 1999.

[204] S. Loyka and F. Gagnon, "Performance analysis of the V-BLAST algorithm: An analytical approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1326–1337, Jul. 2004.

[205] D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K. D. Kammeyer, "Efficient algorithm for decoding layered space-time codes," *Electron. Lett.*, vol. 37, no. 22, pp. 1348–1350, Oct. 2001.

[206] A. Benjebbour, H. Murata, and S. Yoshida, "Comparison of ordered successive receivers for space-time transmission," in *Proc. IEEE VTS 54th Veh. Technol. Conf. (VTC-Fall)*, vol. 4. Oct. 2001, pp. 2053–2057.

[207] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 509–519, Apr. 1990.

[208] R. Kohno, H. Imai, M. Hatori, and S. Pasupathy, "Combinations of an adaptive array antenna and a canceller of interference for direct-sequence spread-spectrum multiple-access system," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 4, pp. 675–682, May 1990.

[209] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, vol. 34, no. 10, pp. 124–136, Oct. 1996.

[210] S. Verdu, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[211] J. W. Choi, A. C. Singer, J. Lee, and N. I. Cho, "Improved linear soft-input soft-output detection via soft feedback successive interference cancellation," *IEEE Trans. Commun.*, vol. 58, no. 3, pp. 986–996, Mar. 2010.

[212] A. Tomasoni, M. Ferrari, D. Gatti, F. Osnato, and S. Bellini, "A low complexity turbo MMSE receiver for W-LAN MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 9. Jun. 2006, pp. 4119–4124.

[213] F. Tosato and P. Bisaglia, "Simplified soft-output demapper for binary interleaved COFDM with application to HIPERLAN/2," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2. May 2002, pp. 664–668.

[214] K. Hyun and D. Yoon, "Bit metric generation for Gray coded QAM signals," *Commun., IEE Proceedings-*, vol. 152, no. 6, pp. 1134–1138, Dec. 2005.

[215] L. Wang, D. Xu, and X. Zhang, "Low complexity bit metric generation for PAM signals based on nonlinear function," *Electron. Lett.*, vol. 47, no. 17, pp. 966–967, Aug. 2011.

[216] C. Xu, D. Liang, S. Sugiura, S. X. Ng, and L. Hanzo, "Reduced-complexity approx-log-MAP and max-log-MAP soft PSK/QAM detection algorithms," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1415–1425, Apr. 2013.

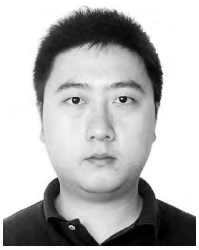
[217] (2009). *Coexistence Between IMT-2000 PCDMA-DS and IMT-2000 OFDMA TDD WMAN in the 2 500-2 690 MHZ Band Operating in Adjacent Bands in the Same Area*. [Online]. Available: http://www.itu.int/pub/R-REP-M.2146-2009
[218] Qualcomm. (2012). *IEEE 802.11ac: The Next Evolution of Wi-Fi Standards*. [Online]. Available: http://www.qualcomm.com/media/documents/files/ieee802-11ac-the-next-evolution-of-wi-fi.pdf
[219] S. ten Brink, "Design of serially concatenated codes based on iterative decoding convergence," in *Proc. 2nd Int. Symp. Turbo Codes Rel. Topics*, Sep. 2000, pp. 319–322.
[220] M. El-Hajjar and L. Hanzo, "EXIT charts for system design and analysis," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 127–153, Feb. 2014.
[221] H. Nguyen, C. Xu, S. Ng, and L. Hanzo, "Near-capacity wireless system design principles," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1806–1833, 4th Quart., 2015.
[222] F. Brannstrom, L. K. Rasmussen, and A. J. Grant, "Convergence analysis and optimal scheduling for multiple concatenated codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3354–3364, Sep. 2005.
[223] W. N. Venables and B. D. Ripley, *Modern Applied Statistics With SPLUS*. New York, NY, USA: Springer-verlag, 1994.
[224] I. Land, P. Hoeher, and U. Sorger, "Log-likelihood values and Monte Carlo simulation: Some fundamental results," in *Proc. Int. Symp. Turbo Codes*, Sep. 2000, pp. 43–46.
[225] H. M. Tullberg and P. H. Siegel, "Serial concatenated TCM with an inner accumulate code-part I: Maximum-likelihood analysis," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 64–73, Jan. 2005.
[226] A. Ashikhmin, G. Kramer, and S. ten Brink, "Code rate and the area under extrinsic information transfer curves," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2002, p. 115.
[227] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 891–907, May 2001.
[228] L. Kong, S. X. Ng, R. Tee, R. G. Maunder, and L. Hanzo, "Reduced-complexity near-capacity downlink iteratively decoded generalized multi-layer space-time coding using irregular convolutional codes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 684–695, Feb. 2010.
[229] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded Modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 7, no. 12, pp. 593–595, Dec. 2003.
[230] F. Simoens, H. Wymeersch, H. Bruneel, and M. Moeneclaey, "Multidimensional mapping for bit-interleaved coded modulation with BPSK/QPSK signaling," *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 453–455, May 2005.
[231] S. Pfletschinger and F. Sanzi, "Error floor removal for bit-interleaved coded modulation with iterative detection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3174–3181, Nov. 2006.
[232] N. S. Muhammad and J. Speidel, "Joint optimization of signal constellation bit labeling for bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 9, no. 9, pp. 775–777, Sep. 2005.
[233] F. Schreckenbach and G. Bauch, "Bit interleaved coded irregular modulation," *Eur. Trans. Telecommun.*, vol. 17, no. 2, pp. 269–282, 2006.
[234] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.
[235] A. Leon-Garcia, *Error Floor Removal for Bit-Interleaved Coded Modulation With Iterative Detection*, 2nd ed. Reading, MA, USA: Addison-Wesley, 1994.
[236] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[237] O. Tirkkonen and A. Hottinen, "Square-matrix embeddable space-time block codes for complex signal constellations," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 384–395, Feb. 2002.
[238] L. Hanzo, S. X. Ng, T. Keller, and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. Hoboken, NJ, USA: Wiley, 2004.
[239] C. Papadias, "On the spectral efficiency of space-time spreading schemes for multiple antenna CDMA systems," in *Proc. 33rd Asilomar Conf. Signals, Syst., Comput.*, vol. 1. Oct. 1999, pp. 639–643.
[240] W. Su and X. G. Xia, "On space-time block codes from complex orthogonal designs," *Wireless Pers. Commun.*, vol. 25, no. 1, pp. 1–26, 2003.
[241] X.-B. Liang, "Orthogonal designs with maximal rates," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2468–2503, Oct. 2003.
[242] W. Su and X.-G. Xia, "Signal constellations for quasi-orthogonal space-time block codes with full diversity," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2331–2347, Oct. 2004.
[243] C. B. Papadias and G. J. Foschini, "A space-time coding approach for systems employing four transmit antennas," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4. May 2001, pp. 2481–2484.
[244] N. Sharma and C. B. Papadias, "Improved quasi-orthogonal codes through constellation rotation," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 332–335, Mar. 2003.
[245] H.-F. Lu and P. V. Kumar, "A unified construction of space-time codes with optimal rate-diversity tradeoff," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1709–1730, May 2005.
[246] C. Yuen, Y. L. Guan, and T. T. Tjhung, "Quasi-orthogonal STBC with minimum decoding complexity," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2089–2094, Sep. 2005.
[247] N. Wu and L. Hanzo, "Near-capacity irregular-convolutional-coding-aided irregular precoded linear dispersion codes," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2863–2871, Jul. 2009.
[248] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "New trellis code design for spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2670–2680, Aug. 2011.
[249] R. Mesleh, I. Stepan, H. Haas, and P. M. Grant, "On the performance of Trellis coded spatial modulation," in *Proc. Int. ITG Workshop Smart Antennas*, Feb. 2009, pp. 235–241.
[250] 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, 2009.
[251] C. Vladeanu, A. MarÅčian, A. F. PĂČun, and S. E. Assad, "A new ML detector for trellis-coded spatial modulation using hard and soft estimates," in *Proc. 10th Int. Symp. Electron. Telecommun. (ISETC)*, Nov. 2012, pp. 143–146.
[252] H. Xu, "Simplified maximum likelihood-based detection schemes for Mary quadrature amplitude modulation spatial modulation," *IET Commun.*, vol. 6, pp. 1356–1363, Jul. 2012.
[253] H. Xu, "Simple low-complexity detection schemes for M-ary quadrature amplitude modulation spatial modulation," *IET Commun.*, vol. 6, no. 17, pp. 2840–2847, Nov. 2012.
[254] S. Sugiura, S. Chen, and L. Hanzo, "A unified MIMO architecture subsuming space shift keying, OSTBC, BLAST and LDC," in *Proc. IEEE 72nd Veh. Technol. Conf. Fall (VTC-Fall)*, Sep. 2010, pp. 1–5.
[255] N. Serafimovski, M. D. 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.
[256] Y. Yang and S. Aissa, "Bit-padding information guided channel hopping," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 163–165, Feb. 2011.
[257] R. Y. Chang, S. J. Lin, and W. H. Chung, "New space shift keying modulation with Hamming code-aided constellation design," *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 2–5, Feb. 2012.
[258] J. M. Luna-Rivera, D. U. Campos-Delgado, and M. G. Gonzalez-Perez, "Constellation design for spatial modulation," in *Proc. Iberoamerican Conf. Electron. Eng. Comput. Sci.*, 2013, pp. 71–78.
[259] C. Xu, S. Sugiura, S. X. Ng, and L. Hanzo, "Reduced-complexity non-coherently detected differential space-time shift keying," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 153–156, Mar. 2011.
[260] T. Datta and A. Chockalingam, "On generalized spatial modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 2716–2721.
[261] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
[262] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.
[263] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part II. Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.
[264] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
[265] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1169–1174, Jul. 2000.

[266] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.
[267] H. Jafarkhani and V. Tarokh, "Multiple transmit antenna differential detection from generalized orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2626–2631, Sep. 2001.
[268] B. Hassibi and B. M. Hochwald, "Cayley differential unitary space-time codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1485–1503, Jun. 2002.
[269] C.-S. Hwang, S. H. Nam, J. Chung, and V. Tarokh, "Differential space time block codes using nonconstant modulus constellations," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2955–2964, Nov. 2003.
[270] S. H. Nam, C. S. Hwang, J. Chung, and V. Tarokh, "Differential space time block codes using QAM for four transmit antennas," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2. Jul. 2004, pp. 952–956.
[271] J. Wang, X. Wang, and M. Madihian, "Design of minimum error-rate Cayley differential unitary space-time codes," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1779–1787, Sep. 2005.
[272] F. Oggier and B. Hassibi, "Algebraic Cayley differential space-time codes," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1911–1919, May 2007.
[273] Y. Bian, M. Wen, X. Cheng, H. Poor, and B. Jiao, "A differential scheme for spatial modulation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3925–3930.
[274] Y. Bian, X. Cheng, M. Wen, L. Yang, H. Poor, and B. Jiao, "Differential spatial modulation," *IEEE Trans. Veh. Tech.*, vol. 64, no. 7, pp. 3262–3268, Jul. 2015.
[275] N. Ishikawa and S. Sugiura, "Unified differential spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 337–340, Aug. 2014.
[276] P. Martin, "Differential spatial modulation for APSK in time-varying fading channels," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1261–1264, Jul. 2015.
[277] R. Schober and L. H. J. Lampe, "Noncoherent receivers for differential space-time modulation," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 768–777, May 2002.
[278] C. Xu, L. Wang, S. X. Ng, and L. Hanzo, "Multiple-symbol differential sphere detection aided differential space-time block codes using QAM constellations," *IEEE Signal Process. Lett.*, vol. 18, no. 9, pp. 497–500, Sep. 2011.
[279] S. K. Cheung and R. Schober, "Differential spatial multiplexing," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2127–2135, Aug. 2006.
[280] L. Wang and L. Hanzo, "Differential interference suppression for SDMA-OFDM based on joint multiple-symbol filtering and detection," *IEEE Trans. Veh. Tech.*, vol. 60, no. 11, pp. 4656–4662, Nov. 2011.

CHAO XU (S'09–M'14) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, China, the B.Sc. (Eng) degree (Hons.) from Queen Mary University of London, U.K., through a Sino-U.K. Joint Degree Program in 2008, both in telecommunications engineering with management, and the M.Sc. degree (Hons.) in radio frequency communication systems and the Ph.D. degree in wireless communications from the University of Southampton, U.K., in 2009 and 2015, respectively. He is currently a Post-Doctoral Researcher with the Southampton Wireless Group, University of Southampton. His research interests include reduced-complexity MIMO design, noncoherent detection, extrinsic information-transfer-chart-aided turbo detection, and cooperative communications. He received the Best M.Sc. Student in broadband and mobile communication networks from the IEEE Communications Society (United Kingdom and Republic of Ireland Chapter) in 2009 and the 2012 Chinese Government Award for Outstanding Self-Financed Student Abroad.

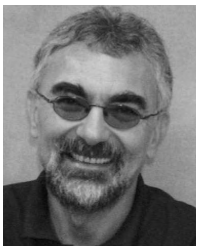
SHINYA SUGIURA (M'06–SM'12) received the B.S. and M.S. degrees in aeronautics and astronautics from Kyoto University, Kyoto, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in electronics and electrical engineering from the University of Southampton, Southampton, U.K., in 2010. From 2004 to 2012, he was a Research Scientist with Toyota Central Research and Development Laboratories, Inc., Aichi, Japan. Since 2013, he has been an Associate Professor with the Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology, Tokyo, Japan, where he heads the Wireless Communications Research Group. He has authored or co-authored over 80 refereed research publications, including 47 IEEE journal and magazine papers. His research interests include wireless communications, networking, signal processing, and antenna technology. He was a recipient of a number of awards, including the Sixth RIEC Award from the Foundation for the Promotion of Electrical Communication in 2016, the Young Scientists' Prize by the Minister of Education, Culture, Sports, Science and Technology of Japan in 2016, the 14th Funai Information Technology Award (First Prize) from the Funai Foundation in 2015, the 28th Telecom System Technology Award from the Telecommunications Advancement Foundation in 2013, the Sixth IEEE Communications Society Asia-Pacific Outstanding Young Researcher Award in 2011, the 13th Ericsson Young Scientist Award in 2011, and the 2008 IEEE Antennas and Propagation Society Japan Chapter Young Engineer Award. He was also certified as an Exemplary Reviewer of the IEEE COMMUNICATIONS LETTERS in 2013 and 2014.

SOON XIN NG (S'99–M'03–SM'08) received the B.Eng. degree (Hons.) in electronic engineering and the Ph.D. degree in telecommunications from the University of Southampton, Southampton, U.K., in 1999 and 2002, respectively. From 2003 to 2006, he was a Post-Doctoral Research Fellow and he was involved in collaborative European research projects known as SCOUT, NEWCOM, and PHOENIX. Since 2006, he has been a member of Academic Staff with the School of Electronics and Computer Science, University of Southampton. He is involved in the OPTIMIX and CONCERTO European projects and the IU-ATC and UC4G projects. He is currently an Associate Professor in telecommunications with the University of Southampton. His research interests include adaptive coded modulation, coded modulation, channel coding, space-time coding, joint source and channel coding, iterative detection, OFDM, MIMO, cooperative communications, distributed coding, quantum error correction codes and joint wireless-and-optical-fiber communications. He has authored over 200 papers and co-authored two John Wiley/IEEE Press books in this field. He is a Chartered Engineer and a fellow of the Higher Education Academy, U.K.



PEICHANG ZHANG received the B.Eng. degree (Hons.) in electronic engineering from the University of Central Lancashire, Preston, U.K., in 2009, and the M.Sc. and Ph.D. degree in wireless communications from the University of Southampton, Southampton, U.K., in 2010 and 2015, respectively. He is currently with the College of Information and Engineering, Shenzhen University, China. His research interests include antenna selection, coherent and non-coherent detection, iterative detection, and channel estimation.

LI WANG (S'09–M'10) was born in Chengdu, China, in 1982. He received the Ph.D. degree from the University of Southampton, Southampton, U.K., in 2010. From 2010 to 2012, he was a Senior Research Fellow with the School of Electronics and Computer Science, University of Southampton. He was involved in a number of projects, such as those from UK's EPSRC, Mobile VCE and Indian-U.K. Advanced Technology Center. In 2012, he joined the Research and Development Center, Huawei Technologies, Stockholm, Sweden, where he is currently a Principle Engineer with the Wireless Network Algorithm Laboratory. He has authored over 40 research papers in IEEE/IET journals and conferences, and also co-authored one JohnWiley/IEEE Press book. He has wide research interests in both radio transmission technology and radio resource management areas for future wireless communication technologies and networks, including PHY layer modeling, advanced iterative receiver design, noncoherent transmission techniques, link adaptation, power control, scheduling, cross-layer cross module system design, CoMP, massive MIMO, mmWave and communication system intelligentization. He has been conducting pioneering cross-discipline researches to build next-generation communication systems with artificial intelligence. Upon his significant contributions in this field, he was a recipient of the Huawei Individual Contribution Award in 2015 and the Furture Star Award in 2017.



LAJOS HANZO (M'91–SM'92–F'04) received the D.Sc. degree in electronics in 1976 and the Ph.D. degree in 1983. During his 40-year career in telecommunications, he has held various research and academic posts in Hungary, Germany, and U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, U.K. He has successfully supervised 111 Ph.D. students. In 2016, he was admitted to the Hungarian Academy of Science. He is currently the Chair in telecommunications with the University of Southampton. He is directing a 50-strong academic research team, where he is involved in a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council, U.K., the European Research Council's Advanced Fellow Grant, and the Royal Society's Wolfson Research Merit Award. He is also an enthusiastic supporter of industrial and academic liaison and offers a range of industrial courses. He has co-authored 18 John Wiley/IEEE Press books on mobile radio communications totaling in excess of 10 000 pages, published over 1681 research contributions on IEEE Xplore. He was an FREng, an FIET, and a fellow of the EURASIP. In 2009, he received the honorary doctorate from the Technical University of Budapest and in 2015 from The University of Edinburgh. He received a number of distinctions. He has over 30 000 citations and an h-index of 68. He served as the TPC chair and the general chair of the IEEE conferences, and presented keynote lectures. He is also a Governor of the IEEE VTS. From 2008 to 2012, he was an Editor-in-Chief of the IEEE Press and a Chaired Professor with Tsinghua University, Beijing.