Differential Space-Time Coding Dispensing with Channel-Estimation Approaches the Performance of Its Coherent Counterpart in the Open-Loop Massive MIMO-OFDM Downlink

Naoki Ishikawa, *Member, IEEE*, Rakshith Rajashekar, *Senior Member, IEEE*, Chao Xu, *Member, IEEE*, Shinya Sugiura, *Senior Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—In this paper, we propose a simple yet powerful mapping scheme that converts any conventional square-matrix-based differential space-time coding (DSTC) into a nonsquare-matrix-based DSTC. This allows DSTC schemes to be used practically in open-loop large-scale multiple-input multiple-output scenarios. Our proposed scheme may be viewed as the differential counterpart of coherent spatial modulation (SM), of the generalized SM, of Bell laboratories layered space-time architecture, and of subcarrier-index modulation. The fundamental impediment of the existing DSTC schemes is the excessive complexity imposed by the unitary constraint. Specifically, the transmission rate of conventional DSTC schemes decays as the number of transmit antennas increases. Our proposed scheme eliminates this impediment and thus achieves a significantly higher transmission rate. We introduce four novel construction methods for the nonsquare codewords, some of which include an arbitrary number of nonzero elements in each codeword column. Our analysis shows that the proposed encoding technique reduces the complexity of both the inverse Fourier transform and of the detection processes. Our proposed scheme is shown to approach the performance of its coherent counterpart for low-mobility scenarios, where the number of transmit antennas is increased up to 256.**

***Index Terms*—differential modulation, differential space-time block codes, spatial modulation, differential spatial modulation, massive MIMO, OFDM.**

I. Introduction

THE massive multiple-input multiple-output (MIMO) concept has attracted intense research interests as a benefit of its attractive features [1, 2]. Commonly, time-division duplex is assumed to exploit the channel's reciprocity [1, 3]. A large number of antenna elements can be accommodated at a base station for attaining a high transmit beamforming (BF) gain, which is achieved for example by low-complexity conjugate downlink BF [1]. This BF operation requires the knowledge of the downlink channel coefficients, which are estimated from the uplink channel coefficients. Assuming that we use orthogonal frequency division multiplexing (OFDM), the channel estimation complexity increases linearly with the number of subcarriers.

N. Ishikawa is with the Graduate School of Information Sciences, Hiroshima City University, Japan (e-mail: naoki@ishikawa.cc). R. Rajashekar, C. Xu and L. Hanzo are with the School of ECS, University of Southampton, UK (e-mail: rmr1u14@soton.ac.uk, cx1g08@ecs.soton.ac.uk, lh@ecs.soton.ac.uk). S. Sugiura is with the Institute of Industrial Science, University of Tokyo, Meguro-ku, Tokyo 153-8505, Japan (sugiura@ieee.org).

Spatial modulation (SM) is a compelling MIMO scheme that has found favor since 2001 [4]. In contrast to well-known spatial multiplexing based Bell laboratories layered space-time architecture (BLAST), the SM scheme conveys implicit information by activating an information-dependent antenna for the entire set of multiple transmit antennas (TAs). This unique antenna-hopping structure has been shown to be capable of achieving reliable communication at a low complexity [5–7]. The SM principle has also been generalized to multiple activated antennas and to multiple time slots, which are referred to as generalized SM (GSM) [8] and asynchronous space-time shift keying (ASTSK) [9]. Later, SM-oriented on-off signaling was also applied to the time- and frequency-domains [10, 11]. Even its low-complexity architecture relying on a low number of radio frequency (RF) chains, the SM concept was also investigated in open-loop large-scale MIMO systems [12–14], in which no channel state information (CSI) is needed at the transmitter for attaining a BF gain.

Differential space-time coding (DSTC) has been investigated since 2000 [15–17]. In parallel to the development of coherent SM [4, 8], its differential counterpart has been studied [18–23] within the DSTC framework, in which neither transmitter nor receiver needs CSI. The differential counterpart of the SM scheme has been termed as differential spatial modulation (DSM) [18, 20], which relies on sophisticated square-shaped sparse-matrix multiplications termed as unitary differential encoding [15, 16]. Thanks to this sparse SM-oriented structure, the DSM scheme imposes a low complexity both at the transmitter and at the receiver. Furthermore, this structure avoids the infinite-cardinality problem raised in [22], namely that the resultant constellation has an infinite cardinality.[1] Explicitly, since the DSTC scheme relies on unitary matrix multiplications, the resultant constellation may have an infinite number of arbitrary phases, which hence requires high-resolution power-thirsty digital-to-analog converters (DACs) at the transmitter.

In massive MIMO scenarios, a pertinent problem is *pilot contamination* [1]. The simultaneous reuse of uplink pilot symbols in the same frequency-slot induces inter-cell interferences. This problem becomes severe as the number of

[1]This problem was later termed as the unbounded differential constellation size issue in [23].

users increases. Additionally, the estimated channel matrix may not be accurate, which is caused both by the channel and low-cost hardware [24]. Accurate channel tracking requires periodic pilot insertions, which reduces the effective throughput. Furthermore, the high-resolution DACs required by a large number of TAs impose high static and dynamic energy dissipations [25].

The finite-cardinality DSTC schemes of [22, 23] may solve all the above pilot-related problems. However, the conventional DSTC is unsuitable for large-scale scenarios, because the conventional schemes have to rely on a large number of square matrices. For example, let us consider a scenario with $M = 4$ TAs and $R = 4$ [bits/symbol]. In this case, the conventional DSTC scheme has to prepare $2^{4\cdot4} = 65536$ 4×4 unitary matrices both at the transmitter and the receiver, which is impractical, and yet it has been overlooked in the existing studies. Furthermore, the complexity of the inverse fast Fourier transform (IFFT) based modulator becomes excessive for large-scale MIMO-OFDM scenarios. Specifically, the IFFT complexity increases linearly with the number of TAs M and subcarriers $N_{\rm sc}$ increase, according to $M \cdot N_{\rm sc} \log_2(N_{\rm sc})$.

Against this background, the novel contributions of this paper are as follows.

1) We propose a low-complexity mapping method for converting an arbitrary $M \times M$ square-matrix-based DSTC (S-DSTC) scheme into an $M \times T$ nonsquare-matrix-based DSTC (N-DSTC) scheme, which is capable of operating in massive MIMO-OFDM scenarios. This conversion increases the transmission rate of the conventional DSTC from R to $R \cdot M/T$ [bits/symbol],[2] which is achieved by a set of appropriately designed bases.
2) We propose four novel construction methods for these bases that strike a flexible rate-diversity tradeoff. The number of nonzero elements P in each codeword column can be adjusted within the range of $1 \leq P \leq M$. Thus, the special cases of our proposed scheme are equivalent to the differential counterparts of the conventional coherent SM [4], of GSM [8], of BLAST [26], of ASTSK [9] and of subcarrier index modulation [11, 27, 28]. The number of nonzero elements P determines the computational complexity both at the transmitter and the receiver.
3) By beneficially designing our projection basis set, the resultant constellation has a finite cardinality [22, 23]. Thus, low-cost and low-resolution DACs can be used at the base station. This simplified RF chain reduces the power consumption and the cost, which are especially significant for large-scale scenarios.
4) Our proposed scheme is capable of reducing the rather high $N_{\rm sc}$–point IFFT complexity order from $M \cdot N_{\rm sc} \log_2(N_{\rm sc})$ to $P \cdot N_{\rm sc} \log_2(N_{\rm sc})$ on average, where the low-complexity sparse FFT algorithm of [29] is invoked.
5) We develop a sub-optimal detector for our projection-based N-DSTC, the complexity of which is similar to that of the conventional coherent SM or of the BLAST scheme, and it is much lower than that of the conventional S-DSTC scheme. This detector relies on a forgetting factor and reconstructs the full-rank $M \times M$ unitary matrix from the rank-deficient $M \times T$ matrix. Since the probability that a number of blocks do not contain CSI for all TAs is always nonzero, this paper uses a recursive decision-feedback detector instead of the conventional twin-block or multiple-block detector.
6) We demonstrate that the proposed scheme is capable of achieving beneficial performance gains over the conventional S-DSTC schemes in terms of complexity, reliability and capacity, for both narrowband and wideband scenarios. The gains are especially significant in large-scale low-mobility scenarios. The conventional S-DSTC schemes are unsuitable for massive MIMO-OFDM scenarios due to their exponentially escalating complexity.

The proposed N-DSTC scheme subsumes the scheme proposed in [30], which only supports the conventional SM scheme, in which only one nonzero element is in each codeword column, i.e., $P = 1$. The conventional scheme of [30] results in an infinite cardinality constellation. Additionally, only narrowband channels were considered in [30], while OFDM was designed for dispersive wireless systems. Moreover, our proposed scheme subsumes the space-time block coded RDSM scheme proposed in [31], which is an extension of the conventional RDSM scheme [30] having $T = 2$.

The remainder of this paper is organized as follows. Section II describes the system model assumed entire the paper. Section III reviews the conventional S-DSTC schemes, while Section IV introduces the proposed N-DSTC scheme. Section V presents our analysis and design guidelines for the proposed scheme. In Section VI, we compare the achievable performances of the proposed scheme and of conventional schemes. Finally, Section VII concludes this paper.

II. System Model

In this paper, the number of transmit antennas (TAs) is denoted by M, and the number of receive antennas (RAs) is denoted by N. The number of OFDM data symbols in the frequency domain is represented by $N_{\rm sc}$, while the guard interval length in the time domain is represented by T_g. We consider $N_{\rm tap}$ number of delayed taps and assume that the delay spread is shorter than the guard interval T_g. We also consider the carrier frequency offset (CFO) ϵ, which is caused by synchronization errors and Doppler shifts. The CFO is not negligible in high-Doppler scenarios, where accurate channel tracking becomes a challenging task.

A. Narrowband Scenario

In narrowband scenarios, we assume that the channel coefficients $\mathbf{H}(i) \in \mathbb{C}^{N\times M}$ obey frequency-flat Rayleigh fading, which may be described by a complex Gaussian distribution with mean zero and variance one, i.e., $\mathcal{CN}(0, 1)$. The received signal block $\mathbf{Y}(i) \in \mathbb{C}^{N\times T}$ at a block index i is given by

$$\mathbf{Y}(i) = \mathbf{H}(i)\mathbf{S}(i) + \mathbf{V}(i), \tag{1}$$

[2]In this paper, the transmission rate is defined by the number of bits transmitted within a symbol duration in uncoded scenarios. In practice, the achievable capacity is more important, which will be presented in Section VI.

where $\mathbf{S}(i)$ is an $M \times T$ space-time codeword, and the thermal noise $\mathbf{V}(i)$ is assumed to be additive white Gaussian noise (AWGN) $\mathcal{CN}(0, \sigma_v^2)$. Thus, the corresponding signal-to-noise ratio (SNR) is defined by $\mathrm{SNR}^{\mathrm{coh}} = 1/\sigma_v^2$, where coherent detection is assumed.

B. Wideband Scenario

In wideband scenarios, we use the frequency-selective channel model proposed in [32]. Here, we omit the block index i for simplicity. For N_{sc} data subcarriers, N_{sc} number of space-time codewords $\left\{\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \cdots, \mathbf{S}^{(N_{\mathrm{sc}})}\right\}$ are generated. At the subcarrier index j $(1 \le j \le N_{\mathrm{sc}})$, the corresponding received signal block $\mathbf{Y}^{(j)} \in \mathbb{C}^{N \times T}$ is represented as [32]

$$\mathbf{Y}^{(j)} = \mathbf{H}^{(j)}\mathbf{S}^{(j)} + \mathbf{N}^{(j)}\left(\mathbf{S}^{(1)}, \cdots, \mathbf{S}^{(N_{\mathrm{sc}})}\right) + \mathbf{V}^{(j)}. \quad (2)$$

Here, $\mathbf{H}^{(j)}$ denotes the diagonal channel matrix, while $\mathbf{N}^{(j)}(\cdot) \in \mathbb{C}^{N \times T}$ denotes the inter-carrier interference (ICI) depending on the CFO ϵ and the codewords $\mathbf{S}^{(j')}$ $(j \ne j')$. When assuming that the CFO is equal to $\epsilon = 0.0$, the Frobenius norm of the interference $\left\|\mathbf{N}^{(j)}(\cdot)\right\|_{\mathrm{F}}^2$ becomes zero [32]. The CFO can be relative accurately estimated based on the received signals, but its residual mean squared error is still not negligible even at high SNRs [32, 33]. We assume that the multipath intensity profile obeys the exponential distribution of $\exp(l/N_{\mathrm{tap}})$ $(1 \le l \le N_{\mathrm{tap}})$ [32, 34]. Additionally, the discrete channel impulse responses are generated according to the Jakes channel model [35], which vary rapidly when the normalized Doppler frequency $F_d T_s$ is high. In Eq. (2), the thermal noise $\mathbf{V}^{(j)}$ is assumed to obey AWGN $\mathcal{CN}(0, \sigma_v^2)$.

III. Conventional Square DSTC

The conventional S-DSTC schemes commonly rely on sophisticated $M \times M$ unitary matrix multiplications [15, 16]. This structure enables the receiver to estimate the original space-time codeword without knowing the channel coefficients.

A. Unitary Encoding

At a block index i $(i \ge 0)$, a space-time codeword $\mathbf{S}(i) \in \mathbb{C}^{M \times M}$ is transmitted from M antennas over M time slots, which satisfies the power constraint of $\|\mathbf{S}(i)\|_{\mathrm{F}}^2 / M = 1$. In this paper, we introduce the diagonal unitary coding (DUC) [36] and the algebraic DSM (ADSM) encoding [23] principles, both of which achieve the maximum theoretically attainable diversity order despite their simple structures.

1) DUC [36]: The B-length input bit sequence is mapped to 2^B number of square matrices, each of which is defined by [36]

$$\mathbf{X}(i) = \mathrm{diag}\left[\exp\left(\jmath\frac{2\pi b}{2^B}u_1\right), \cdots, \exp\left(\jmath\frac{2\pi b}{2^B}u_M\right)\right] \in \mathbb{C}^{M \times M} \quad (3)$$

for $b = 0,\ 1,\ \cdots,\ 2^B - 1$, where $\jmath$ denotes the imaginary number, i.e., $\jmath^2 = -1$. Here, the M number of diversity-maximizing factors $0 < u_1 \le \cdots \le u_M \le 2^B/2 \in \mathbb{Z}$ are designed so as to maximize the diversity product of [36]:

$$\min_{b \in \{1, \cdots, 2^B - 1\}} \left| \prod_{m=1}^{M} \sin\left(\frac{\pi b u_m}{2^B}\right) \right|^{\frac{1}{M}}. \quad (4)$$

2) ADSM [23]: The B-length input bit sequence is partitioned into two sequences: $B_1 = \log_2(M)$ and $B_2 = \log_2(L)$ [bits]. The B_1 [bits] are conveyed by selecting a specific DM out of M DMs $\mathbf{A}_1, \cdots, \mathbf{A}_M \in \mathbb{C}^{M \times M}$, which are given by [23]:

$$\{\mathbf{A}_1, \cdots, \mathbf{A}_M\} = \left\{\mathbf{I}_M, \mathbf{M}, \mathbf{M}^2, \cdots, \mathbf{M}^{M-1}\right\}. \quad (5)$$

Here, the permutation matrix $\mathbf{M}$ is defined by

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & \cdots & 0 & e^{\jmath u_1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \in \mathbb{C}^{M \times M} \quad (6)$$

and the single factor is determined by $u_1 = 2\pi/L$ to maximize the coding gain [23]. The selected DM is represented as $\mathbf{A}_m$. Next, the input B_2 [bits] are mapped to an L–PSK symbol $s(i) \in \mathbb{C}$. Then, a data matrix $\mathbf{X}(i) = s(i)\mathbf{A}_m$ is generated.

After the data matrix $\mathbf{X}(i)$ is generated by the DUC or ADSM scheme, the corresponding space-time codeword is differentially encoded as follows [15, 37]:

$$\mathbf{S}(i) = \begin{cases} \mathbf{I}_M & (i = 0) \\ \mathbf{S}(i-1)\mathbf{X}(i) & (i > 0) \end{cases}. \quad (7)$$

Later, we denote the number of nonzero elements in each column of $\mathbf{S}(i)$ as N_c. For example, both the DUC and the ADSM schemes have $N_c = 1$. The 2×2 differential orthogonal space-time block coding (DOSTC) scheme has $N_c = 2$, which was proposed by Alamouti in [38].

B. Noncoherent ML Detection

The receiver performs independent ML detection for each of the jth subcarrier. Here, we omit the subcarrier index j for simplicity. Assuming that we have $\mathbf{H}(i) = \mathbf{H}(i-1)$ and there is no ICI, the received signal block $\mathbf{Y}(i)$ of Eq. (1) may be transformed into

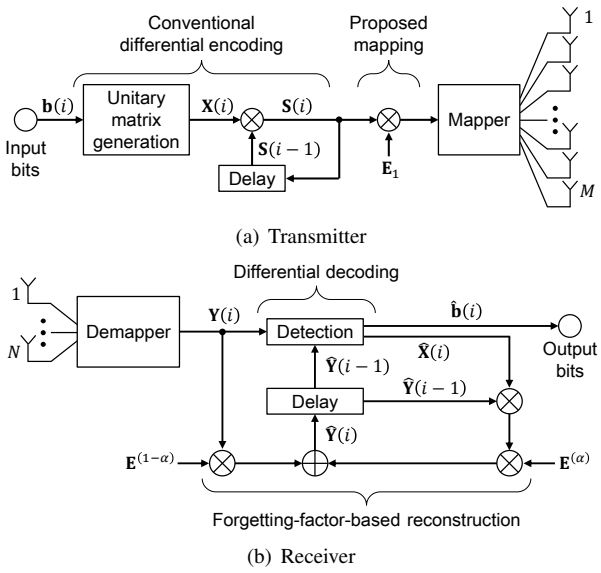
$$\begin{aligned} \mathbf{Y}(i) &= \mathbf{H}\mathbf{S}(i-1)\mathbf{X}(i) + \mathbf{V}(i) \\ &= \left(\mathbf{Y}(i-1) - \mathbf{V}(i-1)\right)\mathbf{X}(i) + \mathbf{V}(i) \\ &= \mathbf{Y}(i-1)\mathbf{X}(i) + \underbrace{\mathbf{V}(i) - \mathbf{V}(i-1)\mathbf{X}(i)}_{\text{AWGN terms}}. \end{aligned} \quad (8)$$

Thus, the ML detection for $i > 0$ is performed by:

$$\hat{\mathbf{X}}(i) = \arg\min_{\mathbf{X}} \|\mathbf{Y}(i) - \mathbf{Y}(i-1)\mathbf{X}(i)\|_{\mathrm{F}}^2. \quad (9)$$

This detection imposes the well-known 3.01 [dB] performance loss, because Eq. (8) contains doubled noise. The effective SNR of the S-DSTC is given by

$$\begin{aligned} \mathrm{SNR}^{\mathrm{diff}} &= \frac{1}{\underbrace{\sigma_v^2}_{\mathbf{V}(i)} + \underbrace{\sigma_v^2}_{\mathbf{V}(i-1)\mathbf{X}(i)}} = \frac{1}{2}\mathrm{SNR}^{\mathrm{coh}} \\ &= \mathrm{SNR}^{\mathrm{coh}} - 3.01\ [\mathrm{dB}], \end{aligned} \quad (10)$$



(a) Transmitter

(b) Receiver

Fig. 1. Schematic of the proposed system.

which is lower than that of its coherent counterpart $\mathrm{SNR}^{\mathrm{coh}}$.

IV. Proposed Nonsquare DSTC

We propose a novel N-DSTC scheme, which is developed from the conventional S-DSTC. Our proposed scheme invokes a simple projection at the transmitter, which trims away a part of the information from $\mathbf{S}(i)$, and performs a forgetting-factor-based reconstruction at the receiver, which recovers the trimmed information. A set of designed bases $\mathbf{E}_1, \cdots, \mathbf{E}_{M/T} \in \mathbb{C}^{M\times T}$ plays a key role in both the projection and reconstruction procedures. For simplicity, we omit the subcarrier index j and ignore the ICI in this section, but the proposed scheme is evaluated in MIMO-OFDM scenarios in Section VI.

A. Square-to-Nonsquare Projection

Fig. 1(a) shows the transmitter schematic of our proposed scheme. As shown in Fig. 1(a), the transmitter firstly generates an $M\times M$ space-time codeword $\mathbf{S}(i)=\mathbf{S}(i-1)\mathbf{X}(i)$ and then it is multiplied by a basis $\mathbf{E}_1 \in \mathbb{C}^{M\times T}$. Since the multiplied codeword $\mathbf{S}(i)\mathbf{E}_1$ is an $M\times T$ complex-valued matrix, the number of time slots is reduced from M to T. Hence, the transmission rate is increased from R to $M/T\cdot R$.

In order to meet the power constraint at the transmitter and to enable matrix reconstruction at the receiver, the basis $\mathbf{E}_k \in \mathbb{C}^{M\times T}$ for $1\le k\ne k'\le M/T$ has to satisfy the following constraints:

- Power constraint: $\|\mathbf{E}_k\|_{\mathrm{F}}^2 = T$,
- Orthogonality: $\mathbf{E}_k^{\mathrm{H}}\mathbf{E}_k = \mathbf{I}_T$ and $\mathbf{E}_k^{\mathrm{H}}\mathbf{E}_{k'} = \mathbf{0}_T$, and
- Reconstructability: $\sum_{k=1}^{M/T}\mathbf{E}_k\mathbf{E}_k^{\mathrm{H}} = \mathbf{I}_M$.

In general, the component sub-matrices taken from a unitary matrix $\mathbf{U}_M \in \mathbb{C}^{M\times M}$ satisfy all the above constraints. Thus, we construct a set of bases as $\begin{bmatrix}\mathbf{E}_1 & \mathbf{E}_2 & \cdots & \mathbf{E}_{M/T}\end{bmatrix} = \mathbf{U}_M$, which is also defined by

$$\mathbf{E}_k = \mathbf{U}_M^{(((k-1)\cdot T+1)\,:\,k\cdot T)}, \tag{11}$$

where $\mathbf{U}_M^{(j\,:\,j')}$ represents the jth to j'th columns of $\mathbf{U}_M$. Next, we introduce four types of basis construction methods.

1) Random Basis: The bases $\mathbf{E}_1 \cdots \mathbf{E}_{M/T}$ may be constituted by a random unitary matrix. By generating a random Hermitian matrix $\mathbf{R}\in\mathbb{C}^{M\times M}$, which only has to satisfy $\mathbf{R}=\mathbf{R}^{\mathrm{H}}$, the corresponding unitary matrix is calculated by the Cayley transform as follows: [39]

$$\mathbf{U}_M = (\mathbf{I}_M - j\mathbf{R})(\mathbf{I}_M + j\mathbf{R})^{-1}. \tag{12}$$

The performance of a random basis set depends on its seed. Thus, we have to design it by exploiting the rank- and determinant-criterion or the constrained capacity [40] for the associated codewords $\mathbf{X}(i)\mathbf{E}_1$. This design process requires time-consuming exhaustive search, while the following sparse, dense, and hybrid basis sets do not.

2) Sparse Basis: Sparse bases are constructed by the submatrices of the identity matrix, i.e., $\mathbf{U}_M=\mathbf{I}_M$. For example, if we consider the $(M,T)=(4,2)$ case, the identity-matrix-aided bases $\mathbf{E}_1$ and $\mathbf{E}_2\in\mathbb{C}^{4\times 2}$ are calculated by

$$\{\mathbf{E}_1,\mathbf{E}_2\} = \left\{\begin{bmatrix}\mathbf{I}_T\\ \mathbf{0}_T\end{bmatrix}, \begin{bmatrix}\mathbf{0}_T\\ \mathbf{I}_T\end{bmatrix}\right\} = \left\{\begin{bmatrix}1&0\\0&1\\0&0\\0&0\end{bmatrix}, \begin{bmatrix}0&0\\0&0\\1&0\\0&1\end{bmatrix}\right\}. \tag{13}$$

The sparse basis contains a single nonzero element in each column, while the following dense basis contains M number of nonzero elements in each column.

3) Dense Basis: We construct this dense basis set by incorporating the flexible discrete Fourier transform (DFT) unitary matrix.[3] The DFT basis is constructed by the well-known DFT matrix $\mathbf{W}_M\in\mathbb{C}^{M\times M}$, which is given by

$$\mathbf{W}_M = \frac{1}{\sqrt{M}}\begin{bmatrix}1&1&1&\cdots&1\\1&\omega&\omega^2&\cdots&\omega^{M-1}\\1&\omega^2&\omega^4&\cdots&\omega^{2(M-1)}\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&\omega^{M-1}&\omega^{2(M-1)}&\cdots&\omega^{(M-1)(M-1)}\end{bmatrix}, \tag{14}$$

where we have $\omega=\exp(-2\pi j/M)$. Then, based on the DFT matrix $\mathbf{W}_M$, the DFT basis set is calculated by replacing $\mathbf{U}_M$ in Eq. (11) with $\mathbf{U}_M=\mathbf{W}_M$. For example, if we consider the $(M,T)=(4,2)$ case, the DFT matrix is given by

$$\mathbf{W}_4 = \frac{1}{2}\begin{bmatrix}1&1&1&1\\1&-j&-1&j\\1&-1&1&-1\\1&j&-1&-j\end{bmatrix}. \tag{15}$$

Then, the bases $\mathbf{E}_1$ and $\mathbf{E}_2\in\mathbb{C}^{4\times 2}$ are calculated by

$$\{\mathbf{E}_1,\mathbf{E}_2\} = \frac{1}{2}\left\{\begin{bmatrix}1&1\\1&-j\\1&-1\\1&j\end{bmatrix}, \begin{bmatrix}1&1\\-1&j\\1&-1\\-1&-j\end{bmatrix}\right\}. \tag{16}$$

Due to this flexibility, later we will only use DFT-matrix-aided bases for dense bases.

[3]Here, it is also possible to use the Hadamard matrix [41], which is real-valued and thus reduces complexity. However, the use of Hadamard matrix limits the number of TAs to $M=1,2,4,8,12,16,20,24,28,32\cdots$ [42].

4) Hybrid Basis: We also introduce a hybrid construction method, which combines the sparse and dense bases constructions. Here, the number of nonzero elements in each column is generalized to an integer N_b, where we have the constraint of $1 \leq N_b \leq M$. Then, the hybrid basis set is generated upon replacing $\mathbf{U}_M$ in Eq. (11) by

$$\mathbf{U}_M = \mathrm{bdiag}\underbrace{[\mathbf{W}_{N_b}, \cdots, \mathbf{W}_{N_b}]}_{M/N_b \text{ repetition}} \in \mathbb{C}^{M\times M}. \tag{17}$$

Note that M/N_b must be an integer, i.e., $M/N_b = \lfloor M/N_b \rfloor$. The hybrid basis subsumes both the sparse and dense basis sets. Specifically, the sparse basis can be represented by the hybrid basis having $N_b = 1$, while the dense basis can be represented by the hybrid basis having $N_b = M$.

For example, if we consider the $(M, T, N_b) = (4, 1, 2)$ case, the associated unitary matrix is formulated as

$$\mathbf{U}_4 = \frac{1}{\sqrt{2}}\mathrm{bdiag}\left[\mathbf{W}_2, \mathbf{W}_2\right] = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \tag{18}$$

and the corresponding bases are

$$\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4\} = \frac{1}{\sqrt{2}}\left\{\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\-1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\\-1\end{bmatrix}\right\}. \tag{19}$$

When assuming that we have the $M \times M$ ADSM codewords having $L = 2$, the $M \times 1$ projected counterparts are exemplified by

$$\frac{1}{\sqrt{2}}\left\{\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix}, \begin{bmatrix}-1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}-1\\-1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\-1\\-1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\0\\0\\-1\end{bmatrix}\right\},$$

which are similar to the GSM [8] or subcarrier-index modulation (SIM) [11] codewords.

B. Symbol Transmissions and Receptions

The conventional S-DSTC scheme transmits a reference symbol $\mathbf{I}_M$ over M time slots, while our proposed scheme transmits the reference symbols of $\left\{\mathbf{E}_1, \cdots, \mathbf{E}_{M/T}\right\}$ over M time slots. Thus, during the block index $1 \leq i \leq M/T$, the received symbol block is expressed as

$$\mathbf{Y}(i) = \mathbf{H}(i)\mathbf{E}_i + \mathbf{V}(i). \tag{20}$$

The data matrix $\mathbf{X}(i)$ for $i > M/T$ is differentially encoded according to Eq. (7), and then the space-time codeword $\mathbf{S}(i)$ is mapped onto an $M \times T$ rectangular matrix as follows:

$$\mathbf{Y}(i) = \mathbf{H}(i)\mathbf{S}(i)\mathbf{E}_1 + \mathbf{V}(i). \tag{21}$$

This data transmission is carried out over $i = M/T + 1, \cdots, W/T$ blocks. Later, we denote the number of data symbols as $N_d = W/T - M/T$. The reference insertion ratio is calculated by $\eta = M/T/(W/T) = M/W$. Thus, the effective transmission rate of our proposed system is reduced from R to $R^{\text{eff}} = (1-\eta)\cdot R$. Each column of $\mathbf{S}(i)\mathbf{E}_1$ contains $P = \max(N_b, N_c)$ number of nonzero elements. In narrowband scenarios, P becomes the same as the number of transmit RF chains required, while in MIMO-OFDM scenarios, the number of RF chains required is equal to $\min(P \cdot N_{\text{sc}}, M)$.

C. Noncoherent Detection

Fig. 1(b) shows the schematic of our proposed receiver, which performs independent detection for each subcarrier index. This detection procedure attempts to reconstruct the $N \times M$ matrix $\mathbf{H}(i)\mathbf{S}(i)$ from the $N \times T$ matrices $\mathbf{H}(i)\mathbf{S}(i)\mathbf{E}_1$ and $\mathbf{H}(i-1)\mathbf{S}(i-1)\hat{\mathbf{X}}(i)\mathbf{E}_1$, where the blending ratio is determined by a forgetting factor $\alpha \in \mathbb{R}$ $(0 \leq \alpha \leq 1)$. This reconstruction is a key feature of the proposed scheme. In the noncoherent detection process of the conventional S-DSTC scheme, $\mathbf{Y}(i) \approx \mathbf{H}(i)\mathbf{S}(i) \in \mathbb{C}^{N\times M}$ is available at the receiver without imposing any additional complexity, which contains the CSI of M TAs. By contrast, our proposed scheme maps the full-rank codeword $\mathbf{S}(i)$ onto the rank-deficient codeword $\mathbf{S}(i)\mathbf{E}_1$. Thus, we have to rely on a recursive and decision-feedback reconstruction method for $\hat{\mathbf{Y}}(i) \approx \mathbf{H}(i)\mathbf{S}(i) \in \mathbb{C}^{N\times M}$ based on $\mathbf{Y}(i) \approx \mathbf{H}(i)\mathbf{S}(i)\mathbf{E}_1 \in \mathbb{C}^{N\times T}$, which cannot provide sufficient CSI. Here, we propose a sub-optimal noncoherent detector as follows:

$$\hat{\mathbf{X}}(i) = \arg\min_{\mathbf{X}} \left\| \mathbf{Y}(i) - \hat{\mathbf{Y}}(i-1)\mathbf{X}\mathbf{E}_1 \right\|_{\text{F}}^2, \tag{22}$$

where the key component $\hat{\mathbf{Y}}(i)$ is defined by

$$\hat{\mathbf{Y}}(i) = \begin{cases} \sum_{k=1}^{M/T} \mathbf{Y}(k)\mathbf{E}_k^{\text{H}} & \left(i = \frac{M}{T}\right) \\ \mathbf{Y}(i)\mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}(i-1)\hat{\mathbf{X}}(i)\mathbf{E}^{(\alpha)} & \left(i > \frac{M}{T}\right) \end{cases}. \tag{23}$$

In Eq. (23), the constant matrices $\mathbf{E}^{(\alpha)}$ and $\mathbf{E}^{(1-\alpha)}$ are given by

$$\begin{cases} \mathbf{E}^{(\alpha)} & = \alpha\mathbf{E}_1\mathbf{E}_1^{\text{H}} + \sum_{k=2}^{M/T} \mathbf{E}_k\mathbf{E}_k^{\text{H}} \\ \mathbf{E}^{(1-\alpha)} & = (1-\alpha)\mathbf{E}_1^{\text{H}} \end{cases}. \tag{24}$$

At the block index $i' = M/T$, we have $\hat{\mathbf{Y}}(i') = \sum_{k=1}^{i'} \mathbf{Y}(k)\mathbf{E}_k^{\text{H}} = \mathbf{H}\left(\sum_{k=1}^{i'} \mathbf{E}_k\mathbf{E}_k^{\text{H}}\right) = \mathbf{H}\mathbf{I}_M = \mathbf{H}$, assuming that the channel matrix $\mathbf{H}$ is constant and we have $\mathbf{Y}(i) = \mathbf{H}\mathbf{E}_i$ for $i = 1, \cdots, i'$. Then, at the block index $i'+1$, we have $\mathbf{Y}(i'+1) = \mathbf{H}\mathbf{S}(i'+1)\mathbf{E}_1$, and $\hat{\mathbf{Y}}(i'+1)$ is calculated by

$$\begin{aligned} \hat{\mathbf{Y}}\left(i'+1\right) &= \mathbf{H}\mathbf{S}\left(i'+1\right)\mathbf{E}_1\mathbf{E}^{(1-\alpha)} + \mathbf{H}\mathbf{S}\left(i'\right)\hat{\mathbf{X}}\left(i'+1\right)\mathbf{E}^{(\alpha)} \\ &= \mathbf{H}\mathbf{S}\left(i'+1\right), \end{aligned}$$

since we have the relationships of $\mathbf{S}\left(i'\right)\hat{\mathbf{X}}\left(i'+1\right) = \mathbf{S}(i'+1)$ and $\mathbf{E}_1\mathbf{E}^{(1-\alpha)} + \mathbf{E}^{(\alpha)} = \mathbf{I}_M$. Hence, at high SNRs, $\hat{\mathbf{Y}}(i)$ in Eq. (23) approaches $\mathbf{H}(i)\mathbf{S}(i)$. Note that $\mathbf{E}^{(\alpha)}$ in Eq. (24) can also be defined by $\mathbf{E}^{(\alpha)} = \mathbf{I}_M - \mathbf{E}_1\mathbf{E}^{(1-\alpha)}$. The design guidelines for the forgetting factor α of Eq. (24) will be presented in Section V.

D. Relationships with Conventional Schemes

Our proposed scheme is capable of subsuming the existing square DSTC [15] and the rectangular DSM [30].

1) Square DSTC: The proposed scheme having $M = T$, $\mathbf{E}_1 = \mathbf{I}_M$ and $\alpha = 0$ is equivalent to the conventional S-DSTC encoding, including the differential STSK [43] and DSM [18]. In this case, the proposed noncoherent detector of Eq. (22) becomes equivalent to the conventional detector of Eq. (9).

2) Conventional Rectangular DSM: Our proposed scheme having DSM codewords and the space basis is equivalent to the conventional rectangular DSM scheme. When assuming the employment of the sparse basis, the definition of $\hat{\mathbf{Y}}(i)$ in Eq. (23) is rewritten as

$$\begin{aligned}\hat{\mathbf{Y}}(i) = \hat{\mathbf{Y}}(i-1)\hat{\mathbf{X}}(i)\mathrm{bdiag}\left[\alpha\mathbf{I}_T\ \mathbf{I}_T\ \cdots\ \mathbf{I}_T\right] \\ + \mathbf{Y}(i)\left[(1-\alpha)\mathbf{I}_T\ \mathbf{0}_T\ \cdots \mathbf{0}_T\right],\end{aligned}$$

which is the same as the conventional $\hat{\mathbf{Y}}(i)$ definition of Eq. (14) in [30].

V. Analysis and Design Guidelines

Let us now analyze the proposed scheme in terms of its effective SNR at the receiver and the cardinality of the resultant constellation at the transmitter. Then, we provide the complexity analysis of both the transmitter and of the receiver. We demonstrate that the special case of our scheme results in a finite-cardinality, which leads to the reduction of the transmitter's complexity. Finally, we propose design guidelines for the forgetting factor so as to maximize the effective SNR.

A. Effective SNR Analysis

We derive the effective SNR for the proposed scheme, in a similar manner as the conventional square differential case of Eq. (10). We assume that $\mathbf{H}(i)$ remains constant over W time slots and all the data symbols are perfectly decoded, i.e., $\hat{\mathbf{X}}(i) = \mathbf{X}(i)$.[4] Then, Eq. (23) can be transformed into $\hat{\mathbf{Y}}(i) = \mathbf{H}\mathbf{S}(i) + \hat{\mathbf{V}}(i)$, where we have

$$\hat{\mathbf{V}}(i) = \begin{cases} \sum_{k=1}^{M/T} \mathbf{V}(k)\mathbf{E}_k^{\mathrm{H}} & (i = M/T) \\ \mathbf{V}(i)\mathbf{E}^{(1-\alpha)} + \hat{\mathbf{V}}(i-1)\mathbf{X}(i)\mathbf{E}^{(\alpha)} & (i > M/T) \end{cases}. \tag{25}$$

Substituting $\hat{\mathbf{Y}}(i) = \mathbf{H}\mathbf{S}(i) + \hat{\mathbf{V}}(i)$ into Eq. (1) yields

$$\begin{aligned}\mathbf{Y}(i) &= \mathbf{H}\mathbf{S}(i)\mathbf{E}_1 + \mathbf{V}(i) \\ &= \mathbf{H}\mathbf{S}(i-1)\mathbf{X}(i)\mathbf{E}_1 + \mathbf{V}(i) \\ &= \left(\hat{\mathbf{Y}}(i-1) - \hat{\mathbf{V}}(i-1)\right)\mathbf{X}(i)\mathbf{E}_1 + \mathbf{V}(i) \\ &= \hat{\mathbf{Y}}(i-1)\mathbf{X}(i)\mathbf{E}_1 + \underbrace{\mathbf{V}(i) - \hat{\mathbf{V}}(i-1)\mathbf{X}(i)\mathbf{E}_1}_{\text{AWGN terms}}. \end{aligned} \tag{26}$$

[4]This assumption is equivalent to the case where the SNR is sufficiently high.

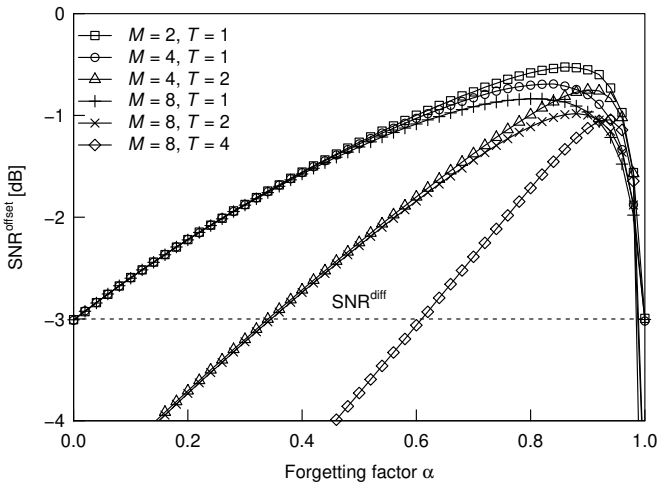


Fig. 2. Forgetting factor α and $\mathrm{SNR}^{\mathrm{offset}}$ relationship, where the number of TAs was varied from $M = 2$ to 8. The number of RAs was $N = 1$. The normalized frame length was fixed to $W/T = 10^2$. We employed the dense bases.

Provided that we have $M = T$ and $\mathbf{E}_1 = \mathbf{I}_M$, Eq. (26) can be rewritten as

$$\begin{aligned}\mathbf{Y}(i) &= \hat{\mathbf{Y}}(i-1)\mathbf{X}(i)\mathbf{E}_1 - \hat{\mathbf{V}}(i-1)\mathbf{X}(i)\mathbf{E}_1 + \mathbf{V}(i) \\ &= \mathbf{Y}(i-1)\mathbf{X}(i) + \underbrace{\mathbf{V}(i) - \mathbf{V}(i-1)\mathbf{X}(i)}_{\text{AWGN terms}},\end{aligned}$$

which becomes the same as Eq. (8). Based on Eq. (26), the effective SNR of the proposed scheme is calculated by

$$\begin{aligned}\mathrm{SNR}^{\mathrm{pro}} &= \frac{1}{\underbrace{\sigma_v^2}_{\mathbf{V}(i)} + \underbrace{\sigma_{v'}^2}_{\hat{\mathbf{V}}(i-1)\mathbf{X}(i)\mathbf{E}_1}} = \frac{1}{\sigma_v^2} \cdot \frac{1}{1 + \sigma_{v'}^2/\sigma_v^2} \\ &= \mathrm{SNR}^{\mathrm{coh}} + \mathrm{SNR}^{\mathrm{offset}}\ [\mathrm{dB}],\end{aligned} \tag{27}$$

where we have

$$\mathrm{SNR}^{\mathrm{offset}} = 10\log_{10}\left(\frac{1}{1 + \sigma_{v'}^2/\sigma_v^2}\right), \tag{28}$$

and

$$\begin{aligned}\sigma_{v'}^2 &= \frac{1}{NT} \cdot \frac{T}{W-M} \sum_{i=M/T+1}^{W/T} \mathrm{E}\left[\left\|\hat{\mathbf{V}}(i-1)\mathbf{X}(i)\mathbf{E}_1\right\|_{\mathrm{F}}^2\right] \\ &= \frac{1}{N} \cdot \frac{1}{W-M} \sum_{i=M/T+1}^{W/T} \mathrm{E}\left[\left\|\hat{\mathbf{V}}(i-1)\mathbf{E}_1\right\|_{\mathrm{F}}^2\right].\end{aligned} \tag{29}$$

Fig. 2 shows the relationship between the forgetting factor α and $\mathrm{SNR}^{\mathrm{offset}}$, where we had $\mathrm{SNR}^{\mathrm{coh}} = 0$ [dB]. The number of TAs was set to $M = 2, 4$, and 8, while the number of time slots in each codeword was set to $1 \le T \le M$. We checked that this comparison was independent of the number of RAs N. It was shown in Fig. 2 that the forgetting factor α had a significant effect on $\mathrm{SNR}^{\mathrm{offset}}$. When $\mathrm{SNR}^{\mathrm{offset}}$ is higher than $\mathrm{SNR}^{\mathrm{diff}} = -3.01$ [dB], our proposed scheme performs better than the conventional S-DSTC scheme. The design guidelines for α based on $\mathrm{SNR}^{\mathrm{offset}}$ will be described in Section V-D.

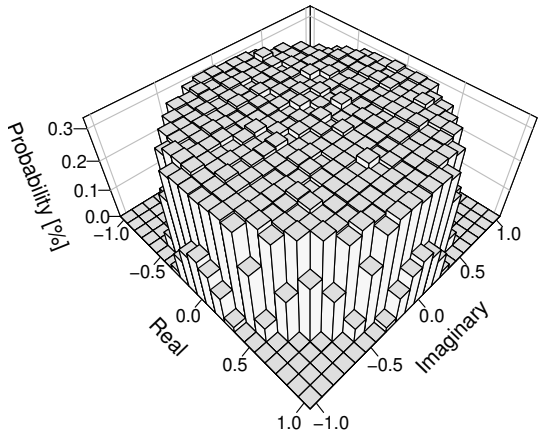


Fig. 3. The histogram of S-DOSTC constellation for the $(M, L) = (2, 8)$ case, where the 10^5 number of random codewords were generated.

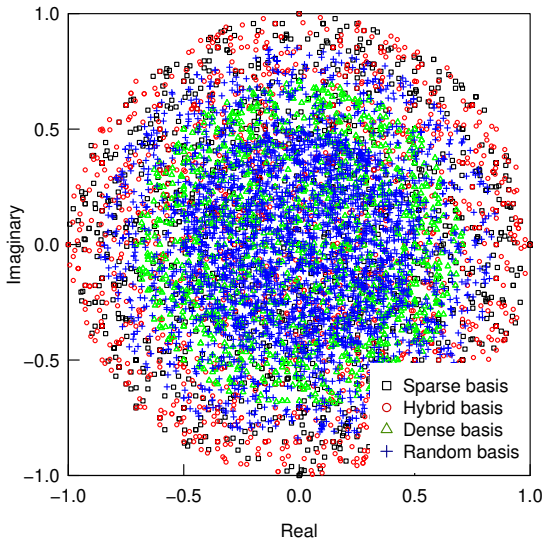
B. Constellation Analysis

Since the differential encoding operation of $\mathbf{S}(i) = \mathbf{S}(i-1)\mathbf{X}(i)$ includes complex multiplications and additions, the resultant symbols may differ from the conventional APSK constellation. Fig. 3 shows an illustrative example for the 2×2 DOSTC symbols [38] having $L = 8$. As shown in Fig. 3, the probability distribution of the DOSTC symbols obeys a uniform distribution over a circle. Thus, the transmitter has to be equipped with multiple high-resolution and power-thirsty DACs for accurately representing an infinite number of near-continuous input symbols, instead of discrete input symbols.[5] This issue was identified in [22, 23] and was referred to as the infinite-cardinality constellation. The ADSM scheme introduced in this paper is capable of producing a finite-cardinality constellation, provided that the ADSM codewords are suitably designed. The number of symbols to be transmitted under multiplications is limited to $N_s = L + 1$, which includes the zero symbol.

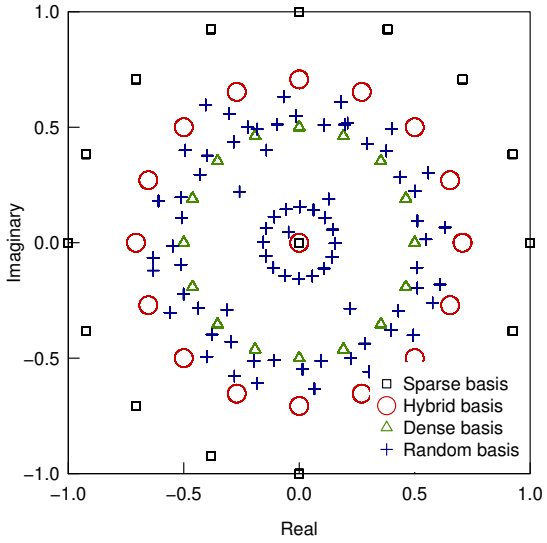
Our proposed scheme does not reintroduce the infinite-cardinality problem because the square-to-nonsquare projection is applied to the transmitted signal matrix directly. Specifically, our proposed scheme projects an $M \times M$ square DSTC codeword onto an $M \times T$ nonsquare codeword. The projected $M \times T$ codeword is not used at the next differential encoding process. Thus, this projection does not incur the infinite-cardinality problem when relying on the DSTC scheme that does not incur the problem.

The resultant constellation cardinality depends on the type of basis employed. More specifically, if we use the sparse or hybrid basis sets, the resultant cardinality is kept the same as that of the conventional DSTC, i.e., $N'_s = N_s$. In case we use the dense basis set, the resultant codeword does not include zero symbols and thus we have $N'_s = N_s - 1$ number of symbols. If we use the random basis set, the resultant constellations size is increased from N_s to $N'_s = M \cdot (N_s + M - 1)$, which is still limited to a finite number. Fig. 4 shows constellation examples for the nonsquare-matrix-based DOSTC and ADSM,

[5]This relationship is similar to the capacity comparison formulated by the continuous input and the discrete input.



(a) N-DOSTC



(b) N-ADSM

Fig. 4. Constellation examples for the $(M, T, R) = (4, 1, 6.0)$ case, where the 800 number of random codewords were generated.

where the 4×4 matrices are projected onto 4×1 matrices. We considered all the four basis sets introduced in Section IV-A, namely the sparse, hybrid, dense, and random basis sets. The results in Fig. 4 show that the constellation size of the N-DOSTC is exponentially increased as the block index is increased. By contrast, our proposed scheme was capable of keeping the constellation size to a finite number. The random basis set generated an increased number of input symbols, i.e., $N'_s = M \cdot (N_s + M - 1) = 4 \cdot (16 + 4 - 1) = 76$, which was still finite and was much smaller than for the N-DOSTC case of Fig. 4(a).

C. Complexity Analysis

We next evaluate the computational complexity of our proposed scheme in terms of the IFFT calculation at the transmitter and the detection process at the receiver. Specifically,

the computational complexity is quantified in terms of the number of real-valued multiplications [9], which corresponds to the number of multipliers on a module [44]. For example, the multiplication of two complex values requires four real-valued multiplications. The complex addition cost is negligible against the multiplication cost [45].

1) Transmitter Complexity: At the transmitter, we focus our attention on the number of real-valued multiplications in the IFFT calculation. The Cooley-Tukey algorithm [46] reduces the inverse Fourier transform complexity from $\Omega(M \cdot N_{\rm sc}^2)$ to $\Omega\left[M \cdot N_{\rm sc} \log_2(N_{\rm sc})\right]$ by exploiting the efficient butterfly diagram.[6] However, this reduced complexity still becomes a large burden for large-scale MIMO-OFDM systems, because it linearly increases upon increasing the number of TAs.

Against this problem, our proposed scheme significantly reduces the IFFT complexity from $\Omega\left[M \cdot N_{\rm sc} \log_2(N_{\rm sc})\right]$ to

$$\Omega\left[P \cdot N_{\rm sc} \log_2\left(N_{\rm sc}\right)\right], \tag{30}$$

where the sparse FFT algorithm is invoked [29].[7] Specifically, each subcarrier data frame per antenna includes $\sum_{q=0}^{N_{\rm sc}} \binom{N_{\rm sc}}{q} \left(\frac{P}{M}\right)^q \left(1-\frac{P}{M}\right)^{(N_{\rm sc}-q)} \cdot q = \frac{P}{M} N_{\rm sc}$ number of nonzero elements on average, where $P = \max\left(N_b, N_c\right)$ represents the number of nonzero elements in a codeword column. Eq. (30) does not include the number of TAs M. Thus, our proposed scheme exhibits a high scalability against the increase in M, and is suitable for large-scale MIMO-OFDM systems.[8]

2) Receiver Complexity: We count the number of real-valued multiplications at the detection process of Eq. (22), where $\mathbf{E}^{(\alpha)}$, $\mathbf{E}^{(1-\alpha)}$, and $\mathbf{X}(i)\mathbf{E}_1$ are calculated in advance. The decoding complexity per time slot is calculated for our proposed scheme by

$$\begin{aligned} &2^{RT+1}\left[2\max\left(N_b, N_c\right)+1\right]N \\ &+\frac{M}{T}\left(cN_b+4N_c\right)N + cN_bN, \end{aligned} \tag{31}$$

where we have $c=2$ for real-valued bases and $c=4$ for complex-valued bases. Since our proposed scheme is equivalent to the conventional schemes for some specific setups, as described in Section IV-D, Eq. (31) is also valid for the conventional schemes. Eq. (22) is lower bounded by $\Omega\left(2^{RT}PN\right)$. Thus, the number of time slots in a codeword T has a dominant effect on the detection process. The conventional S-DSTC exhibits the complexity of $\Omega\left(2^{RM}MN\right)$, which can be reduced to $\Omega\left(2^{R}N\right)$ by our proposed scheme. This $1/\left(2^{M}M\right)$ reduction is significant for a large-scale MIMO system.

Fig. 5 shows the computational complexity comparisons both at the transmitter and the receiver. We varied the number of TAs from $M=2^1$ to 2^{10}, while the number of RAs was fixed to $N=1$. We considered $N_{\rm sc}=16$ subcarriers for the IFFT calculation. It was shown in Fig. 5(a) that the IFFT complexities of the conventional MIMO-OFDM and of the proposed dense basis set having $N_b=M$ were identical. The IFFT complexities of the sparse and hybrid basis sets having $N_b=1$ or 2 remained constant upon increasing the number of TAs. Next, as shown in Fig. 5(b), the decoding complexities of the conventional square-matrix-based family were inevitably high. By contrast, our sparse-DSTC-based scheme having sparse or dense basis sets exhibited similar complexities as the conventional coherent SM or BLAST schemes. Hence, both Fig. 5(a) and (b) implied that our proposed scheme was capable of achieving low-complexity encoding and decoding, which were significantly lower than that of the conventional square-matrix-based schemes. In summary, Table I characterizes the constellation and complexity analyses of Sections V-B and V-C.

[6]We use Donald Knuth's big Omega $\Omega(\cdot)$ notation [47], which represents an asymptotic lower bound.

[7]We considered the worst subcarrier-activation case. In general, the calculation cost can be further reduced by the algorithm proposed in [48].

[8]Note that the sparsity in the frequency domain does not contribute to the FFT process at the receiver, because the time-domain symbols have a dense structure.

D. Forgetting Factor Design

As shown in Fig. 2, the design of the forgetting factor α has a dominant effect on the achievable performance. The conventional design criterion for α is given by [30]

$$\tilde{\alpha}^{(1)} = \arg\min_{\alpha} J_{\rm MSE}, \tag{32}$$

where $J_{\rm MSE}$ is defined as

$$J_{\rm MSE} = \frac{1}{NT}\frac{T}{W-M}\sum_{i=M/T+1}^{W/T} \mathrm{E}\left[\left\|\hat{\mathbf{Y}}(i) - \mathbf{H}(i)\mathbf{S}(i)\right\|_{\rm F}^2\right]. \tag{33}$$

This criterion supports both time-invariant and time-variant channels, because it contains the channel matrix $\mathbf{H}(i)$. Obviously, it is a time-consuming task to obtain the average of Eq. (33) due to the random generations of $\mathbf{X}(i)$, $\mathbf{H}(i)$, and $\mathbf{V}(i)$. Thus, we develop a new design criterion that maximizes $\mathrm{SNR}^{\rm offset}$ of Eq. (28) as follows:

$$\begin{aligned} \tilde{\alpha}^{(2)} &= \arg\max_{\alpha} \mathrm{SNR}^{\rm offset} \\ &= \arg\min_{\alpha} \sum_{i=M/T+1}^{W/T} \mathrm{E}\left[\left\|\hat{\mathbf{V}}(i)\right\|_{\rm F}^2\right]. \end{aligned} \tag{34}$$

Note that Eq. (34) only supports time-invariant scenarios. In Eq. (34), we still have to generate random $\mathbf{X}(i)$ and $\mathbf{V}(i)$. For further simplification, we propose the following optimistic criterion:

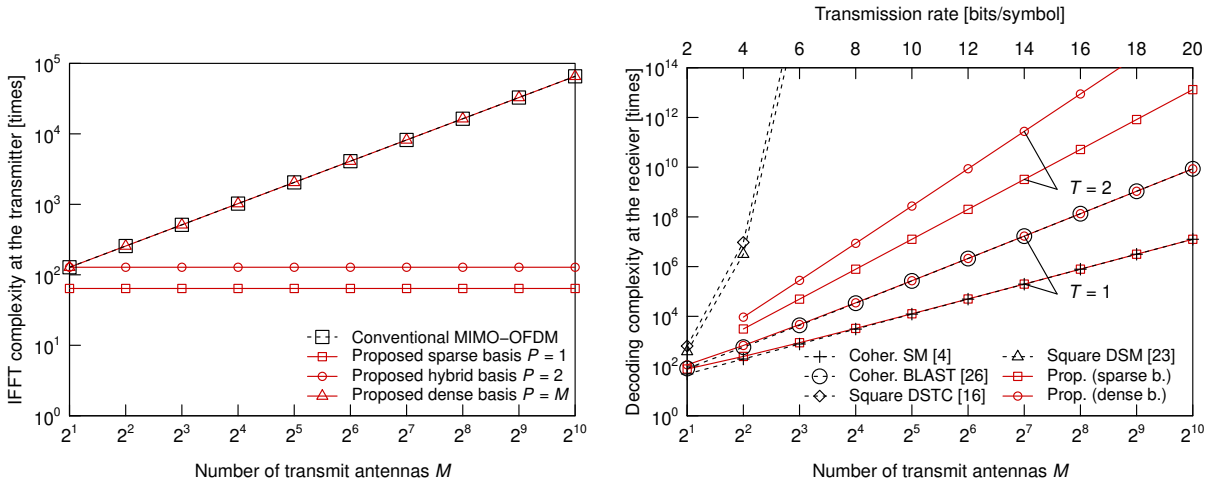
$$\tilde{\alpha}^{(3)} = \arg\min_{\alpha} \sum_{i=M/T+1}^{W/T} \mathrm{E}\left[\left\|\hat{\mathbf{V}}'(i)\right\|_{\rm F}^2\right], \tag{35}$$

where we have

$$\hat{\mathbf{V}}'(i) = \begin{cases} \mathbf{V}'(i) & (i = M/T) \\ \mathbf{V}'(i)\mathbf{E}^{(1-\alpha)} + \hat{\mathbf{V}}'(i-1)\mathbf{M}\mathbf{E}^{(\alpha)} & (i > M/T) \end{cases}. \tag{36}$$

In Eq. (36), $\mathbf{M}$ is generated by Eq. (6) having $u_1=\pi$. Additionally, $\mathbf{V}'(i) \in \mathbb{C}^{N\times T}$ obeys $\mathcal{CN}(0,1)$, which is equivalent to the SNR of 0 [dB].[9] Eq. (35) implies that the interaction

[9]The SNR does not affect the forgetting factor design.



(a) Transmitter complexity. (b) Receiver complexity.

Fig. 5. The computational complexities of the IFFT at transmitter and the decoding process at receiver. The number of TAs was varied from $M = 2^1$ to 2^{10}, while the number of RAs was $N = 1$. The number of subcarriers was $N_{\rm sc} = 16$.

TABLE I
FUNDAMENTAL CHARACTERISTICS OF OUR PROPOSED SCHEME.

	Sparse DSTC $N_c = 1$		Dense DSTC $N_c = M$	
	P	Complexity	P	Complexity
Without basis $M = T$	1	$2^{RM+1}3N + 8N \geq \Omega(2^{RM}N)$	M	$2^{RM+1}(2M+1)N + 4(M+1)N \geq \Omega(2^{RM}MN)$
Sparse basis $N_b = 1$	1	$2^{RT+1}3N + 6\frac{M}{T}N + 2N \geq \Omega(2^{RT}N)$	M	$2^{RT+1}(2M+1)N + 2\frac{M}{T}(2M+1)N + 2N \geq \Omega(2^{RT}MN)$
Hybrid basis $N_b = 2$	2	$2^{RT+1}5N + 8\frac{M}{T}N + 4N \geq \Omega(2^{RT}N)$	M	$2^{RT+1}(2M+1)N + 4\frac{M}{T}(M+1)N + 4N \geq \Omega(2^{RT}MN)$
Dense basis $N_b = M$	M	$2^{RT+1}(2M+1)N + 2\frac{M}{T}(M+1)N + 2MN \geq \Omega(2^{RT}MN)$	M	$2^{RT+1}(2M+1)N + 6\frac{M^2}{T}N + 2MN \geq \Omega(2^{RT}MN)$

between the AWGN terms $\mathbf{V}'(i)$ and the constant matrices $\left(\mathbf{E}^{(1-\alpha)}, \mathbf{E}^{(\alpha)}\right)$ determines the optimal forgetting factor in terms of $\mathrm{SNR}^{\rm offset}$. In our simulations, we checked that the gap between $\tilde{\alpha}^{(1)}$ and $\tilde{\alpha}^{(3)}$ was small for time-invariant channels, when the number of data symbols N_d is sufficiently high, such as the case in which the reference insertion ratio is below 10%.

Fig. 6 shows the relationship between the number of data symbols N_d and the forgetting factor according to Eq. (35). Here, we only considered the sparse and dense bases for simple illustration. We varied the number of TAs from $M = 2$ to 64. It was shown in Fig. 6 that the forgetting factor $\tilde{\alpha}^{(3)}$ for the sparse basis decreased upon increasing the number of TAs M. For large-scale MIMO scenarios, the optimal forgetting factor for the sparse basis set tends to become small because each received block contains a limited channel information. By contrast, for the $M \geq 8$ cases, the corresponding forgetting factor for the dense basis set was unchanged. This was because the dense basis aided codewords include the sufficient channel information associated with all the TAs. In Section VI, we will investigate the effect of the codeword sparsity on the achievable performance.

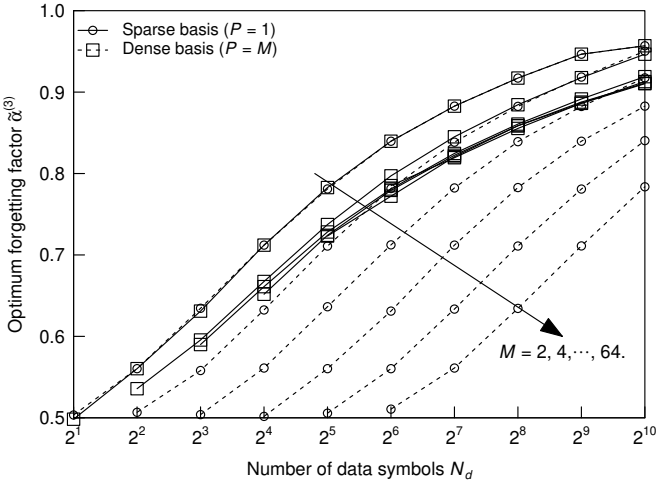
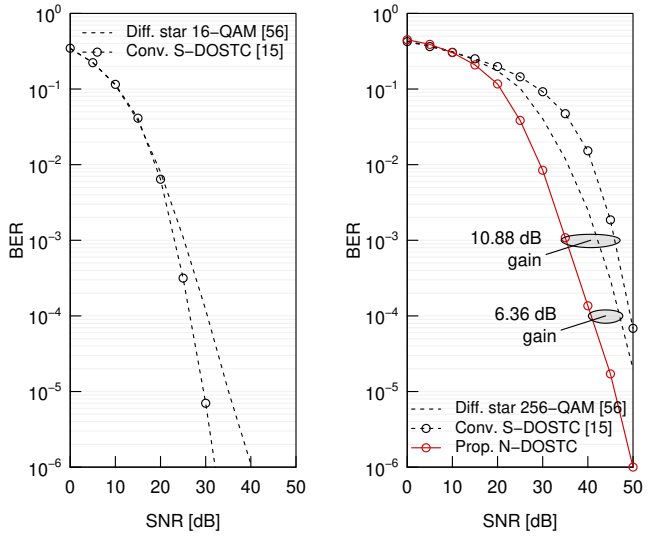


Fig. 6. Relationship between N_d and $\tilde{\alpha}^{(3)}$, which maximizes the effective SNR, where both the sparse and dense bases were considered. The number of TAs was set to $M = 2,\ 4,\ 8,\ 16,\ 32$, and 64. The numbers of RAs and time slots in a codeword were $N = 1$ and $T = 1$, respectively.

VI. PERFORMANCE RESULTS

In this section, we compared the conventional S-DSTC and the proposed N-DSTC schemes both in terms of their bit error ratio (BER) and constrained average mutual information (AMI) [49]. Here, we considered both the ideal narrowband and the realistic wideband scenarios. Ideally, we



(a) S-DOSTC ($R = 4.0$) (b) N-DOSTC ($R = 8.0$)

Fig. 7. BER comparisons between the S-DOSTC and its nonsquare counterpart, where the $(M, N) = (2, 2)$ scenario was considered and the reference insertion ratio was $\eta = 5\%$. We used the sparse basis and the forgetting factor of $\tilde{\alpha}^{(3)} = 0.796$.

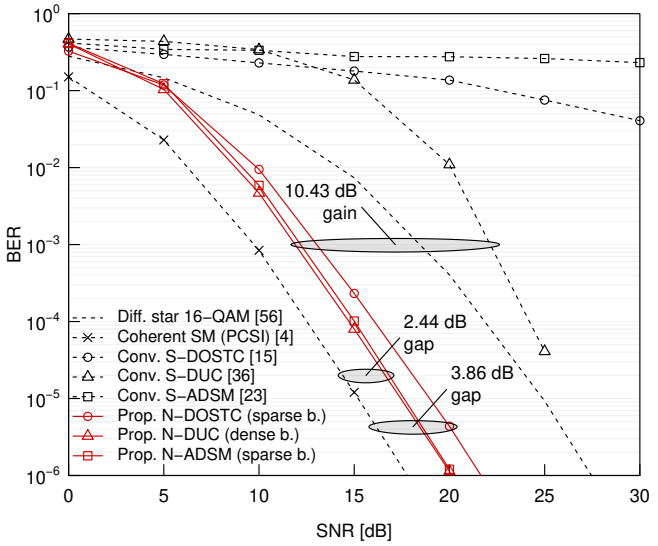


Fig. 8. BER comparison between the square and nonsquare schemes, where the $(M, N, R) = (4, 4, 4.0)$ scenario was considered and the reference insertion ratio was $\eta = 5\%$. We used the sparse basis having the forgetting factor of $\tilde{\alpha}^{(3)} = 0.797$, and the dense basis having the forgetting factor of $\tilde{\alpha}^{(3)} = 0.810$.

should compare the proposed scheme to the family of state-of-the-art training-based channel estimation schemes [50–52] or the S-DSTC scheme designed for time-varying channels [22, 23, 35, 53–55]. All the existing studies have indicated that the coherent counterpart of a non-coherent scheme having perfect channel state information (PCSI) can be regarded as a performance upper-bound. Thus, in our simulations, we only considered ideal reference schemes, such as the coherent counterparts of the SM, the BLAST, and the S-DSTC schemes having PCSI.

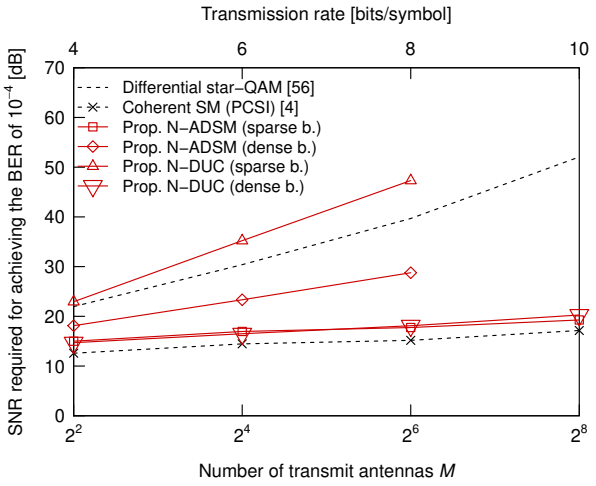
A. Ideal Narrowband Scenario

First, we evaluated our proposed scheme in ideal narrowband scenarios, where the quasi-static channel coefficients obeyed frequency-flat Rayleigh fading. This ideal model does not include CFOs and ICIs, as detailed in Section II.

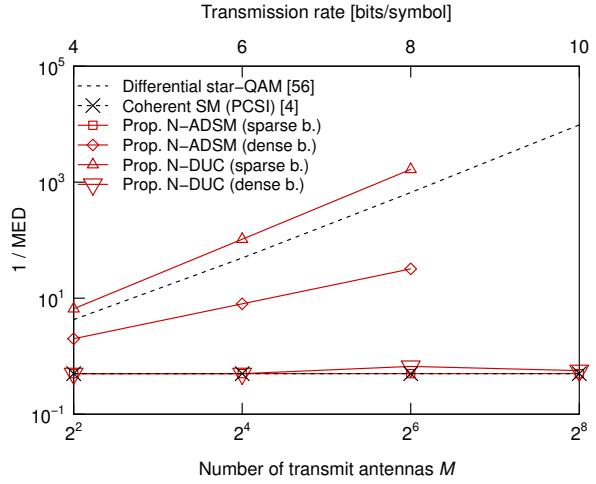
Fig. 7 shows the BER comparisons between the differential star-QAM [56] and the DOSTC schemes. The transmission rate of the original S-DOSTC using a 16–PSK scheme was $R = 4.0$ [bits/symbol] in Fig. 7(a), while that of its projected counterpart was increased from $R = 4.0$ to $R = 8.0$ [bits/symbol] in Fig. 7(b). Here, the 2×2 matrices were projected onto 2×1 matrices by multiplying the sparse basis $\mathbf{E}_1 \in \mathbb{C}^{2\times 1}$. We plotted the BER curve of the conventional S-DOSTC using a 256–PSK in Fig. 7(b) for reference. It was shown in both Figs. 7(a) and (b) that the S-DOSTC scheme achieved a transmit diversity order of two. In Fig. 7(b), the proposed N-DOSTC scheme based on the S-DOSTC scheme of Fig. 7(a) achieved a better BER than the differential star-QAM scheme. Here, the corresponding transmit diversity order was one, since the number of its time slots was reduced to $T = 1$ from $T = 2$. In Fig. 7(b), the N-DOSTC scheme achieved 6.36 [dB] gain at BER $= 10^{-4}$ over the differential star 256–QAM scheme. Note that the conventional S-DOSTC scheme performed worse than the single-stream low-complexity differential scheme. This performance loss is commonly observed for high-rate scenarios.

Fig. 8 shows our BER comparisons between the square and the nonsquare DSTC schemes, where we considered DOSTC, DUC, and ADSM encoding. The numbers of transmit and RAs were $M = N = 4$. The transmission rate was $R = 4.0$ [bits/symbol] for all the schemes considered. The BER curves of the conventional differential star-QAM and of the coherent SM having PCSI were depicted as benchmarks. As shown in Fig. 8, the BER of the conventional S-DSTC schemes was worse than that of the differential star-QAM scheme, which is a low-complexity single-stream scheme. By contrast, our proposed N-DUC scheme having $P = 1$ basis performed the best among all the schemes considered, where the gain obtained by the nonsquare architecture was 10.43 [dB]. Additionally, the proposed N-ADSM scheme achieved a similar performance to the N-DUC scheme. In the following simulations, we only consider the N-DUC and N-ADSM, both of which performs better than the N-DOSTC scheme. The SNR gap between the coherent SM and the N-DUC was 2.43 [dB], which was lower than the well-known 3.01 [dB] gap. This was achieved by following our design guideline for the forgetting factor $\tilde{\alpha}^{(3)}$. As shown in Fig. 2, the effective SNR of our proposed scheme may become better than that of the S-DSTC scheme.

Fig. 9(a) shows the effective SNR comparison at BER $= 10^{-4}$, where the number of TAs was $M = 4, 16, 64, 256$, and the corresponding transmission rate was $R = 4, 6, 8, 10$ [bits/symbol]. Here, we were unable to consider the S-DSTC schemes such as S-DUC and S-ADSM. For example, for the $(M, R) = (16, 6)$ case, the S-DUC scheme has to prepare an excessive number of $L = 2^{16\times 6} \approx 7.92 \times 10^{28}$ 16×16 codewords. For the $(M, R) = (64, 8)$ case, the number of



(a) Effective SNR.



(b) Reciprocal of MED.

Fig. 9. Effective SNR and MED comparisons in large-scale MIMO scenarios, where the number of RAs was $N = 4$ and the reference insertion ratio was $\eta = 5\%$. The number of TAs was varied from $M = 4$ to $M = 256$, while the corresponding transmission rate was $R = 4, 6, 8$, and 10. We used both the sparse and dense basis sets.

codewords increases to $L = 2^{64\times 8} \approx 1.34 \times 10^{154}$, which cannot be implemented. As shown in Fig. 9(a), both the sparse-basis-aided N-ADSM and the dense-basis-aided DUC schemes achieved a similar performance to the coherent SM having perfect CSI. By contrast, the sparse-basis-aided N-DUC achieved worse performance than the differential star-QAM. This performance loss can be predicted by the minimum Euclidean distance (MED) between the projected codewords $\mathbf{X}(i)\mathbf{E}_1$. Fig. 9(b) shows the MED comparisons associated with Fig. 9(a). It was observed from Fig. 9(b) that the reciprocal of MED was correlated with the effective SNR without any exception. Thus, the performance of random, sparse, dense and hybrid bases can be estimated by their MED in advance of transmissions. This MED metric is known as the rank- and determinant-criterion, which is further detailed in [40].

Fig. 10 compares the achievable diversity of the proposed N-DUC scheme, where the numbers of transmit and RAs were $M = 16$ and $N = 1$. The number of time slots of a codeword

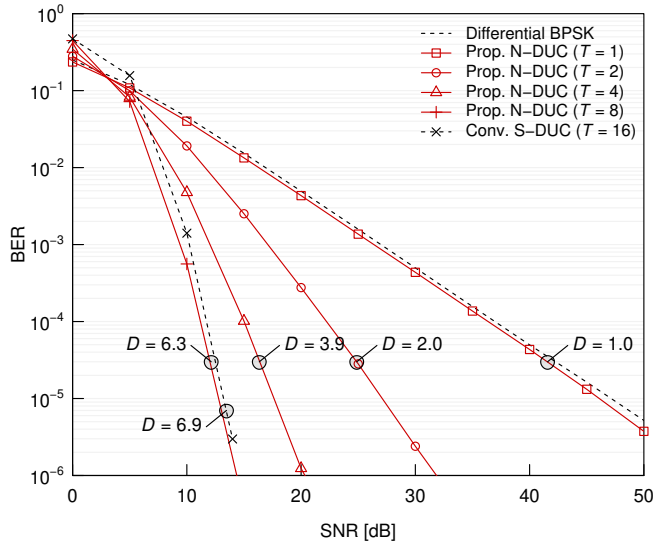


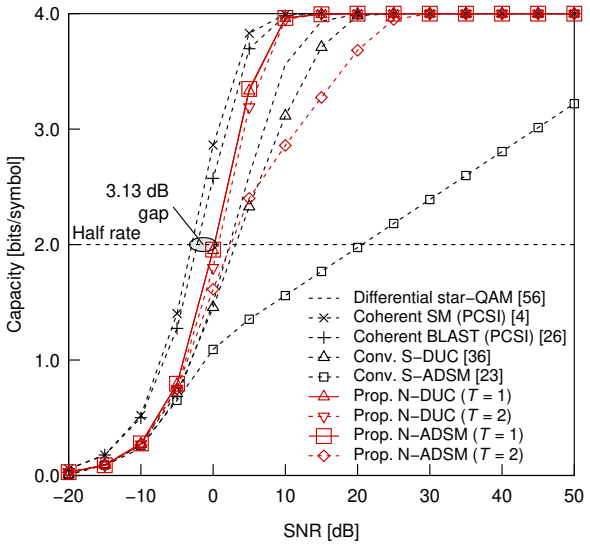
Fig. 10. Achievable diversity comparisons for the $(M, N, R) = (16, 1, 1.0)$ scenario, where the reference insertion ratio was $\eta = 5\%$. The number of time slots in a codeword was varied from $T = 1$ to 16. We used the dense basis having $\tilde{\alpha}^{(3)} = 0.855$.

was set to $T = 1, 2, 4, 8$, and 16. Note that the N-DUC scheme having $T = 1$ is equivalent to the differential PSK signaling[10], and the N-DUC scheme having $T = 16$ is equivalent to the conventional S-DUC scheme. It was shown in Fig. 10 that for the $1 \leq T \leq 4$ cases, the proposed N-DUC scheme achieved the maximum diversity order of $D = T$. For the $T = 8$ case, the N-DUC scheme's diversity order was $D = 6.3$, which was lower than the maximum of $T = 8.0$. In the BER $\leq 10^{-6}$ region, this diversity order may increase to 8.0, which was predicted by the RDC [40].

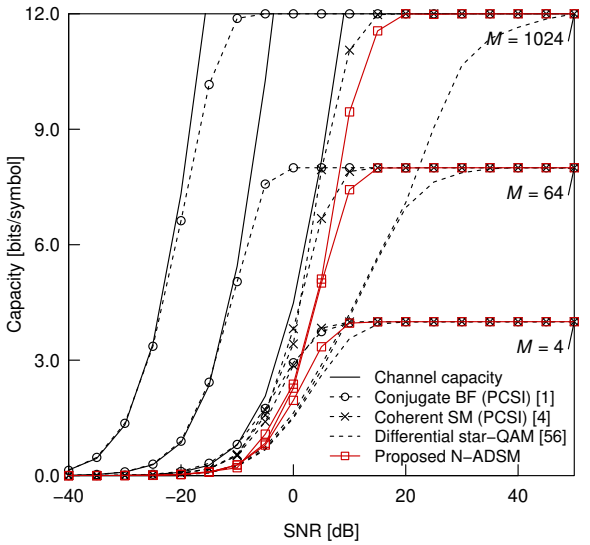
Fig. 11 compares the constrained AMI of the proposed N-DSTC and of the conventional S-DSTC schemes, where the DUC and ADSM encoding were considered. Specifically, the constrained AMI of a differential scheme was calculated by assuming that the associated AMI is $10\log_{10}(2) \approx 3.01$-dB lower than that of its coherent counterpart [57].[11] The number of TAs was set to $M = 4$, 64 and 1024, while the number of RAs was fixed to $N = 4$. The corresponding transmission rate was $R = 4$, 8, and 12. In Fig. 11(a), we considered the coherent SM, BLAST and the differential star-QAM as benchmarks. We used the dense basis set for N-DUC and the sparse basis set for N-ADSM. It was shown in Fig. 11(a) that our proposed N-DUC and N-ADSM schemes having $T = 1$ achieved the same performance, which was about 3.0 dB worse than that of the coherent SM scheme having perfect CSI. Here, the differential star-QAM scheme had a higher AMI than both the proposed N-ADSM scheme having $T = 2$ and the square-matrix-based scheme, because both of the better schemes were specifically designed for maximizing the diversity gain, not the unconstrained AMI. In Fig. 11(b), the number of TAs was increased from $M = 4$

[10]In Fig. 10, the N-DUC scheme having $T = 1$ performed better than the differential BPSK signaling with the aid of our forgetting-factor-aided successive detector.

[11]An exact AMI formulation is found in [22].

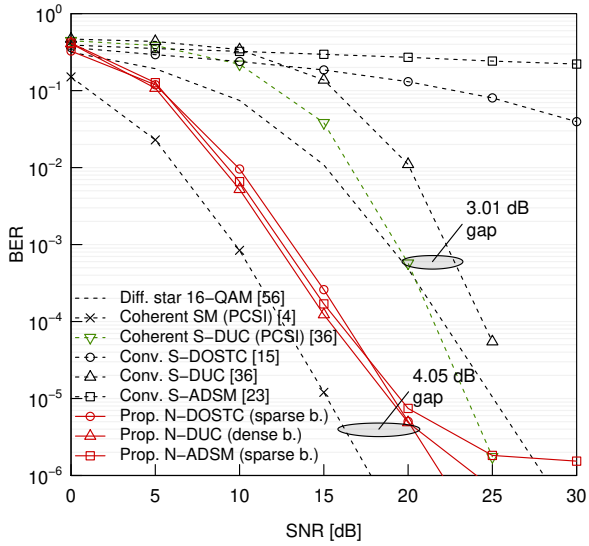


(a) $M = 4$ TAs.

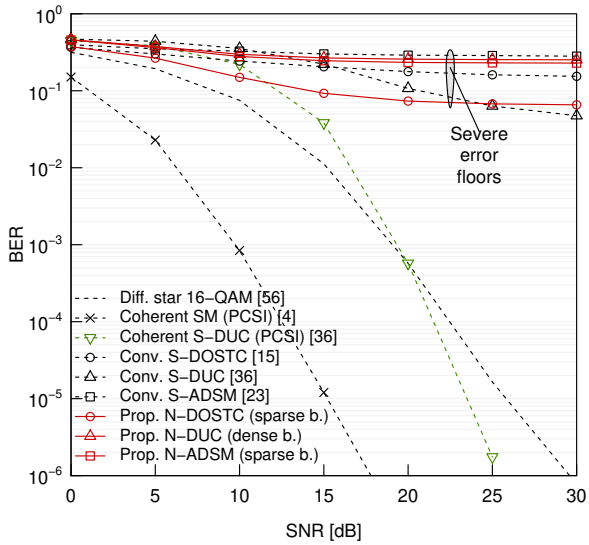


(b) $M = 4, 16$, and 64 TAs.

Fig. 11. Constrained AMI comparisons between the S- and N-DSTC schemes. The channel capacity, the conjugate BF and the differential star-QAM were considered as benchmarks.



(a) $F_dT_s = 10^{-3}$.



(b) $F_dT_s = 10^{-2}$.

Fig. 12. BER comparison between the square and nonsquare schemes, where the $(M, N, R) = (4, 4, 4.0)$ scenario was considered and other parameters were the same as those used in Fig. 8. We considered the normalized Doppler frequencies of $F_dT_s = 10^{-3}$ and 10^{-2}.

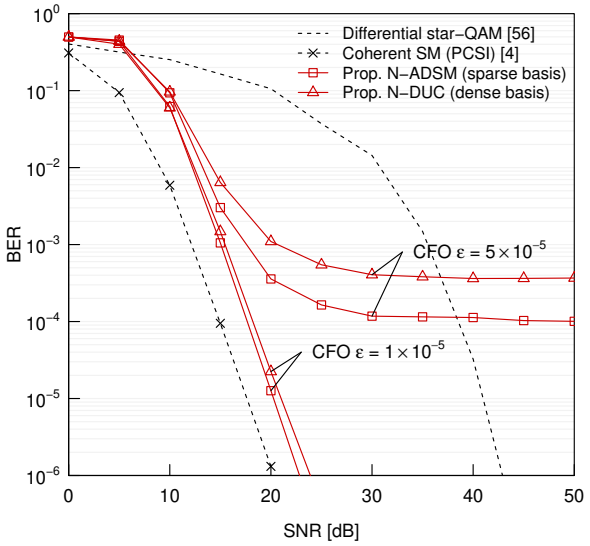
to 64. Here, we considered the conjugate BF having $N = 4$ independent symbol streams as a benchmark. It was shown in Fig. 11(b) that the conjugate BF scheme achieved nearly the same performance as the unconstrained channel capacity with the aid of the BF gain obtained by a large number of TAs. Upon increasing the number of TAs, the gap between the coherent SM and our proposed N-ADSM schemes was kept to about 3.0 [dB]. Our proposed scheme was capable of achieving the best constrained AMI among the differential schemes that do not require accurate channel estimation.

B. Frequency-Selective Wideband Scenario

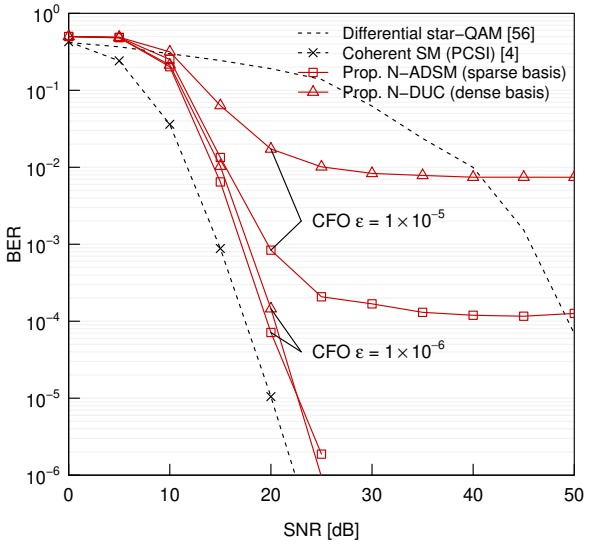
Next, we evaluated our proposed scheme in high-mobility MIMO-OFDM scenarios, where the CFO due to synchronization errors and Doppler shifts was not negligible. In our simulations, we assumed that the number of delayed taps was $N_{\rm tap} = 10$ [11], the number of data subcarriers allocated to a single user was $N_{\rm sc} = 16$, and the reference insertion ratio was $\eta = 5\%$.

Fig. 12 shows the BER comparisons between the square and the nonsquare DSTC schemes, where we considered the same scenario as those used in Fig. 8. The BER curves of the coherent SM and S-DUC schemes having PCSI were plotted as references. Different from Fig. 8, we considered the normalized Doppler frequencies of $F_dT_s = 10^{-3}$ and 10^{-2} [55]. As shown in Fig. 12(a), our proposed scheme exhibited a similar trend to Fig. 8 and the performance losses induced by Doppler shifts were negligible in the $F_dT_s = 10^{-3}$ scenario.[12] By contrast, in the $F_dT_s = 10^{-2}$ scenario, our

[12]In our additional simulations, we observed that the BER achieved at high SNRs remained unchanged upon decreasing the reference insertion ratio η.



(a) $(M, N, R) = (64, 4, 8.0)$.



(b) $(M, N, R) = (256, 4, 10.0)$.

Fig. 13. BER comparisons for the $M = 64$ and 256 scenarios, where the reference insertion ratio was $\eta = 5\%$. We considered the MIMO-OFDM channel model having CFO $\epsilon = 1.0 \times 10^{-6}$, $\epsilon = 1.0 \times 10^{-5}$, and $\epsilon = 5.0 \times 10^{-5}$.

scheme exhibited error floors which were higher than the BER curve of the simple differential star 16–QAM scheme. Here, it is worth noting that the proposed scheme outperformed the conventional S-DUC scheme having PCSI in the $F_d T_s = 10^{-3}$ scenario. This implies that the proposed scheme is especially beneficial for low-Doppler scenarios.

Fig. 13(a) shows the BER comparisons between the sparse-basis-aided N-ADSM having $\tilde{\alpha}^{(3)} = 0.801$ and the dense-basis-aided N-DUC having $\tilde{\alpha}^{(3)} = 0.919$,[13] where CFOs of $\epsilon = 1.0 \times 10^{-5}$ and 5.0×10^{-5} [32] were considered. The BER curve of the coherent SM scheme having perfect CSI was added as a benchmark. It was shown in Fig. 13(a) that our proposed scheme was capable of approaching the coherent performance even when we considered the CFO of $\epsilon = 1.0 \times 10^{-5}$. By contrast, both the N-ADSM and N-DUC schemes exhibited high error floors above $\text{BER} \geq 10^{-4}$ when we had the CFO of $\epsilon = 5.0 \times 10^{-5}$, which can be mitigated by powerful channel coding schemes such as turbo codes [40].

Fig. 13(b) shows the BER comparisons between the N-ADSM having $\tilde{\alpha}^{(3)} = 0.794$ and the N-DUC having $\tilde{\alpha}^{(3)} = 0.973$, where CFOs of $\epsilon = 1.0 \times 10^{-6}$ and 1.0×10^{-5} were considered. In contrast to Fig. 13(a), we increased the number of TAs from $M = 64$ to $M = 256$, which was a large-scale scenario. It was shown in Fig. 13(b) that our proposed scheme achieved a near-coherent performance for the CFO $\epsilon = 1.0 \times 10^{-6}$ case. By contrast, for the CFO $\epsilon = 1.0 \times 10^{-5}$ case, both the proposed N-ADSM and the N-DUC schemes exhibited high error floors. Specifically, the error floor of N-ADSM was $\text{BER} = 1.0 \times 10^{-4}$, while that of N-DUC was $\text{BER} = 7.0 \times 10^{-3}$. Hence, the N-ADSM scheme has a higher robustness against the CFO-induced ICIs, with the aid of its codeword sparsity.

[13] Strictly, we have to use $\tilde{\alpha}^{(1)}$ for time-varying scenarios. To circumvent its complexity, we used $\tilde{\alpha}^{(3)}$ for simplicity.

VII. Conclusion

In this paper, we proposed a simple projection method that converts the conventional S-DSTC scheme into the N-DSTC scheme. The proposed mapping method increases the transmission rate from R to $R \cdot M/T$ [bits/symbol] and reduces both the IFFT calculation and the detection complexities. The proposed scheme can be regarded as a differential counterpart of the conventional SM, GSM, BLAST, ASTSK, or SIM scheme. Since all the proposed random, sparse, dense and hybrid bases achieve different performances, we provided both theoretical and numerical analysis, which assist in the design of the proposed scheme. Our proposed scheme achieved a similar complexity, BER and AMI to the conventional coherent counterpart, where both narrowband and wideband channel models were considered. Specifically, the sparse-basis-aided ADSM scheme achieved the best BER in the face of ICI. By contrast, in the high-mobility scenarios, our scheme exhibited severe error-floors, which will be investigated in our future study.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[2] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2015.

[3] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1461–1476, 2016.

[4] R. Y. Mesleh, H. Haas, S. Sinanovic, C. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.

[5] N. Ishikawa, S. Sugiura, and L. Hanzo, "50 years of permutation, spatial and index modulation: From classic RF to visible light communications and data storage," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1905–1938, 2018.

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

[7] M. I. Kadir, S. Sugiura, S. Chen, and L. Hanzo, "Unified MIMO-multicarrier designs: A space–time shift keying approach," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 550–579, 2015.

[8] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Generalized space shift keying modulation for MIMO channels," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Cannes, France, Sept. 15-18, 2008.

[9] S. Sugiura, S. Chen, and L. Hanzo, "A universal space-time architecture for multiple-antenna aided systems," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, pp. 401–420, 2012.

[10] T. Ishihara and S. Sugiura, "Faster-than-nyquist signaling with index modulation," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 630–633, 2017.

[11] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Transactions of Signal Processing*, vol. 61, no. 22, pp. 5536–5549, 2013.

[12] D. A. Basnayaka, M. Di Renzo, and H. Haas, "Massive but few active MIMO," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 9, pp. 6861–6877, 2016.

[13] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 2156–2168, 2015.

[14] L. He, J. Wang, J. Song, and L. Hanzo, "On the multi-user multi-cell massive spatial modulation uplink: How many antennas for each user?" *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1437–1451, 2017.

[15] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 7, pp. 1169–1174, 2000.

[16] B. L. Hughes, "Differential space-time modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567–2578, 2000.

[17] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.

[18] Y. Bian, M. Wen, X. Cheng, H. Poor, and B. Jiao, "A differential scheme for spatial modulation," in *IEEE Global Communications Conference*, Atlanta, GA, USA, Dec. 9-13, 2013.

[19] N. Ishikawa and S. Sugiura, "Unified differential spatial modulation," *IEEE Wireless Communications Letters*, vol. 3, no. 4, pp. 337–340, 2014.

[20] Y. Bian, X. Cheng, M. Wen, L. Yang, H. V. Poor, and B. Jiao, "Differential spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 3262–3268, 2015.

[21] R. Rajashekar, N. Ishikawa, S. Sugiura, K. V. S. Hari, and L. Hanzo, "Full-diversity dispersion matrices from algebraic field extensions for differential spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 385–394, 2017.

[22] C. Xu, R. Rajashekar, N. Ishikawa, S. Sugiura, and L. Hanzo, "Single-RF index shift keying aided differential space-time block coding," *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 773–788, 2018.

[23] R. Rajashekar, C. Xu, N. Ishikawa, S. Sugiura, K. V. S. Hari, and L. Hanzo, "Algebraic differential spatial modulation is capable of approaching the performance of its coherent counterpart," *IEEE Transactions on Communications, in press*, vol. 65, no. 10, pp. 4260–4273, 2017.

[24] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, 2014.

[25] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2387–2399, 2016.

[26] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proceedings of the International Symposium on Signals, Systems, and Electronics*, Pisa, Italy, Oct. 2, 1998.

[27] N. Ishikawa, S. Sugiura, and L. Hanzo, "Subcarrier-index modulation aided OFDM – will it work?" *IEEE Access*, vol. 4, pp. 2580–2593, 2016.

[28] S. Althunibat, R. Mesleh, and E. Basar, "Differential subcarrier index modulation," *IEEE Transactions on Vehicular Technology, in press*.

[29] H. Hassanieh, P. Indyk, D. Katabi, and E. Price, "Nearly optimal sparse Fourier transform," in *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, New York, USA, 2012.

[30] N. Ishikawa and S. Sugiura, "Rectangular differential spatial modulation for open-loop noncoherent massive-MIMO downlink," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1908–1920, 2017.

[31] C. Wu, Y. Xiao, L. Xiao, P. Yang, and X. Lei, "Space-time block coded rectangular differential spatial modulation," in *IEEE International Conference on Communications*, Kansas City, MO, USA, May 20-24, 2018.

[32] Y. Yao and G. Giannakis, "Blind carrier frequency offset estimation in SISO, MIMO, and multiuser OFDM systems," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 173–183, 2005.

[33] P. Sun, M. Morelli, and L. Zhang, "Carrier frequency offset tracking in the IEEE 802.16e OFDMA uplink," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3613–3619, 2010.

[34] Y. S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 49, no. 8, pp. 1375–1387, 2001.

[35] R. Schober, S. Member, W. H. Gerstacker, and J. B. Huber, "Decision-feedback differential detection of MDPSK for flat Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, no. 7, pp. 1025–1035, 1999.

[36] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Transactions on Communications*, vol. 48, no. 12, pp. 2041–2052, 2000.

[37] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 1962–1973, 2000.

[38] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[39] B. Hassibi and B. Hochwald, "Cayley differential unitary space-time codes," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1485–1503, 2002.

[40] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-capacity multi-functional MIMO systems*. John Wiley & Sons, Ltd, 2009.

[41] F. Adachi, M. Sawahashi, and K. Okawa, "Tree-structured generation of orthogonal spreading codes with different lengths for forward link of DS-CDMA mobile radio," *Electronics Letters*, vol. 33, no. 1, pp. 27–28, 1997.

[42] N. J. Sloane, "A library of hadamard matrices," 1999, http://www.research. att. com/˜ njas/hadamard.

[43] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: a dispersion matrix approach," *IEEE Transactions on Communications*, vol. 58, no. 11, pp. 3219–3230, 2010.

[44] E. Cavus and B. Daneshrad, "A very low-complexity space-time block decoder (STBD) ASIC for wireless systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 1, pp. 60–69, 2006.

[45] R. P. Brent and P. Zimmermann, *Modern Computer Arithmetic*. Cambridge University Press, 2010.

[46] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.

[47] D. E. Knuth, "Big Omicron and big Omega and big Theta," *ACM SIGACT News*, vol. 8, no. 2, pp. 18–24, 1976.

[48] B. Ghazi, H. Hassanieh, P. Indyk, D. Katabi, E. Price, and L. Shi, "Sample-optimal average-case sparse Fourier transform in two dimensions," in *Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 2013.

[49] 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," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1343–1385, 2011.

[50] R. Rajashekar, K. V. S. Hari, and L. Hanzo, "Reduced-complexity ML detection and capacity-optimized training for spatial modulation systems," *IEEE Transactions on Communications*, vol. 62, no. 1, pp. 112–125, 2014.

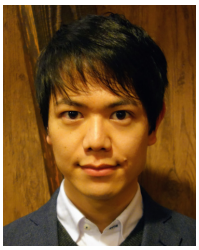
[51] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 540–553, 2017.

[52] E. P. Simon, L. Ros, H. Hijazi, J. Fang, D. P. Gaillot, and M. Berbineau, "Joint carrier frequency offset and fast time-varying channel estimation for MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 955–965, 2011.

[53] R. Schober and L. H. J. Lampe, "Noncoherent receivers for differential space-time modulation," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 768–777, 2002.

[54] P. A. Martin, "Differential spatial modulation for APSK in time-varying fading channels," *IEEE Communications Letters*, vol. 19, no. 7, pp. 1261–1264, 2015.

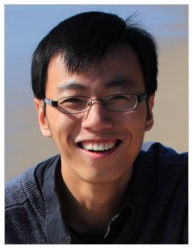
[55] L. Wang, L. Li, C. Xu, D. Liang, S. X. Ng, and L. Hanzo, "Multiple-symbol joint signal processing for differentially encoded single- and multi-carrier communications: Principles, designs and applications," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 689–712, 2014.
[56] R. Webb, W.T. Hanzo, L. Steele, "Bandwidth efficient QAM schemes for Rayleigh fading channels," *IEE Proceedings*, vol. 138, no. 3, pp. 169–175, 1991.
[57] H. V. Nguyen, C. Xu, S. X. Ng, and L. Hanzo, "Near-capacity wireless system design principles," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1806–1833, 2015.



Naoki Ishikawa (S'13–M'17) was born in Kanagawa, Japan, in 1991. He received the B.E., M.E., and Ph.D. degrees from the Tokyo University of Agriculture and Technology, Tokyo, Japan, in 2014, 2015, and 2017, respectively. From June 2015 to September 2015, he was an academic visitor with the School of Electronics and Computer Science, University of Southampton, UK. From April 2016 to March 2017, he was a research fellow of the Japan Society for the Promotion of Science. From April 2017, he has been an assistant professor in the Graduate School of Information Sciences, Hiroshima City University, Japan.

He was certified as an Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS 2017. He also received eight domestic awards, including the Yasujiro Niwa Outstanding Paper Award from Tokyo Denki University in 2018, the Telecom System Technology student Award (honorable mention) from Telecommunications Advancement Foundation of Japan in 2014, the Outstanding Paper Award for Young C&C Researchers from NEC C&C Foundation in 2014, and the Young Researcher's Encouragement Award from the IEEE VTS Japan Chapter in 2014.

Rakshith Rajashekar (M'14–SM'17) received the B.E. degree in electrical communication engineering from Visvesvaraya Technological University, Karnataka, India, in 2007, the Ph.D. degree from the Department of Electrical Communication Engineering, Indian Institute of Science, India, in 2014. He is currently a Research Fellow at the University of Southampton, U.K. He was a Senior Scientist with Broadcom Communications from 2014 to 2015 and a Systems Engineer with Accord Software & Systems, Bangalore, India, from 2007 to 2009. His research interests include diversity schemes in MIMO systems, millimeter wave communication, and differential communication, with a focus on space-time signal processing and coding.



Chao Xu (S'09–M'14) received a B.Eng. degree from Beijing University of Posts and Telecommunications, China, and a BSc(Eng) with First Class Honours from Queen Mary, University of London, UK, through a Sino-UK joint degree program in 2008, both in Telecommunications Enginneering with Management. He obtained a MSc degree with distinction in Radio Frequency Communication Systems and a Ph.D. degree in Wireless Communications from the University of Southampton, UK in 2009 and 2015, respectively. He is currently a post-doctoral researcher working at Southampton Wireless Group, University of Southampton, UK. His research interests include reduced-complexity MIMO design, noncoherent detection, extrinsic-information-transfer-chart-aided turbo detection, and cooperative communications. He was awarded the Best M.Sc. Student in Broadband and Mobile Communication Networks by the IEEE Communications Society (United Kingdom and Republic of Ireland Chapter) in 2009. He also received 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. From 2013 to 2018, he was an Associate Professor with the Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology, Tokyo, Japan. Since 2018, he has been an Associate Professor with the Institute of Industrial Science, University of Tokyo, Tokyo, Japan, where he heads the Wireless Communications Research Group. His research has covered a range of areas in wireless communications, networking, signal processing, and antenna technology. He authored or coauthored over 60 IEEE journal papers.

Dr. Sugiura 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 IEEE COMMUNICATIONS LETTERS in 2013 and 2014, and that of IEEE TRANSACTIONS ON COMMUNICATIONS in 2018.

Lajos Hanzo (http://www-mobile.ecs.soton.ac.uk) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 111 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1700+ research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE ComSoc and VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk Lajos has 34 000+ citations.