On the Joint Optimization of Dispersion Matrices and Constellations for Near-Capacity Irregular Precoded Space-Time Shift Keying

Shinya Sugiura, *Senior Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Joint dispersion-matrix and constellation optimization algorithm is proposed, which is invoked for the recent space-time shift keying (STSK) scheme. More specifically, the theoretical gradients of the DCMC capacity with respect to both a dispersion-matrix set and to the modem constellations are derived, which allows a substantial reduction in the computational complexity required for maximizing the system's capacity. Furthermore, we also conceive a near-capacity irregular-precoded STSK (IR-PSTSK) architecture, which is designed with the aid of extrinsic information transfer (EXIT) charts, while invoking STSK subcodes, which are optimized by using the proposed algorithm.**

***Index Terms*—Constellation, dispersion matrix, EXIT chart, irregular codes, iterative detection, MIMO, space-time shift keying.**

I. Introduction

The recent space-time coding (STC) concept of space-time shift keying (STSK) [1–3] allows us to strike a flexible rate-versus-diversity tradeoff [4] in the context of multiple-input multiple-output (MIMO) schemes, while benefiting from a low-complexity near-optimal detector based on matched filtering [5]. More specifically, the STSK scheme relies on the unique encoding philosophy of activating one out of Q space-time dispersion matrices during each block transmission. The activated dispersion matrix then spreads the classic complex-valued symbols of phase-shift keying (PSK) or quadrature amplitude modulation (QAM). Hence, the basic design parameters of the STSK's transmitter are constituted by a dispersion-matrix set and complex-valued constellations. As mentioned in [2] and the references therein, the STSK family, including the space-shift keying (SSK) [6] and the spatial modulation (SM) schemes [7], was shown to outperform other classic MIMO arrangements in many practical scenarios.

S. Sugiura is with Toyota Central R&D Laboratories, Inc., Nagakute, Aichi, 480-1192, Japan (e-mail: sugiura@ieee.org).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

The work of S. Sugiura was partially supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Number 23760353). The financial support of the RC-UK under the auspices of the India-UK Advanced Technology Centre, as well as of the China-UK 4^{th} generation wireless systems project and that of the European Union under the auspices of the Optimix project, is gratefully acknowledged.

It was also shown in [5] that surprisingly the star-QAM aided STSK scheme outperforms its square-QAM aided counterpart, which is the other way round in classic modems. This is due to the fact that the STSK scheme's achievable performance is affected not only by the minimum distance of the constellations employed, but also by their absolute values, as clarified in [5]. However, the constellations' optimization for the STSK scheme is still an open issue.

A number of dispersion-matrix optimization algorithms have been developed both for linear dispersion codes (LDCs) [8–10] and for STSK [1, 11] under the assumption of employing specific constellations. For example, in [9] a dispersion-matrix set was designed for the uncoded LDC scheme, by minimizing its pairwise-error probability (PEP) with the aid of the gradient descent method. The prevalent approach is that of maximizing the discrete-input continuous-output memoryless channel's (DCMC) capacity [12], but the associated computational cost may become excessive, especially for high-throughput or high-diversity LDC and STSK arrangements [1, 10, 13].[1] More importantly, although it is vital to jointly optimize both the dispersion matrix set and the modem constellations for the sake of maximizing the achievable rate, previous studies only optimized a set of dispersion matrices.

By contrast, in order to attain both a near-capacity performance as well as a low decoding complexity, the turbo principle [14] was extended by invoking multiple concatenated codes, which allows us to replace the high-complexity optimum non-iterative detector by lower-complexity iterative detection. More specifically, in [17] a recursive inner code was invoked for the sake of maximizing the interleaver gain and for avoiding the formation of a bit-error ratio (BER) floor. In order to further reduce the discrepancy between the achievable performance and the capacity bound, the irregular channel codes of [18] were designed with the aid of extrinsic information transfer (EXIT) charts. More recently, in [19, 20] irregular inner- and outer codes were proposed in order to further reduce the area gap in EXIT chart between the inner-code's and outer-code's EXIT curves, since this area was

[1] In order to provide further insights, DCMC capacity provides a tight BER performance bound in the context of near-capacity iteratively-decoded systems, employing a powerful channel code, such as turbo [14, 15] and low-density parity check (LDPC) codes [16], as documented in [1, 13]. On the other hand, the above-mentioned PEP or BER minimization criterion, which is developed for uncoded systems, no more serves as an exact performance metric. Hence, it can be said that DCMC capacity based system optimization is especially useful for near-capacity transceivers, which can be operated in the low SNR regime.

shown to be proportional to the distance from capacity.

Against this back-cloth, the novel contributions of this paper are two-fold:

- *We propose a joint dispersion-matrix and constellation optimization algorithm, where the simplified conjugate gradient (SCG) algorithm is invoked for maximizing the DCMC capacity. In order to facilitate this in a practical manner, the theoretical gradients of the DCMC capacity with respect to both a dispersion-matrix set and to the modem constellations are derived, which results in a substantial reduction of the computational complexity required for the system's optimization.*[2]
- *Another new contribution is that a set of the optimized STSK component codes is then invoked for creating a near-capacity irregular precoded STSK (IR-PSTSK) architecture, which is designed with the aid of EXIT charts, while making the best use of the maximized-capacity STSK code sets. This enables us to attain a near-capacity performance over a wide range of signal-to-noise ratios (SNR).*

The remainder of this paper is organized as follows. Section II reviews the STSK scheme's model, the DCMC capacity and the error-bound. In Section III we propose the novel dispersion-matrix and constellation optimization scheme. Section IV provides EXIT-chart-aided design guidelines for our irregular-precoded STSK architecture, while the STSK scheme designed by the proposed optimization scheme is characterized in Section V. Finally, Section VI concludes this paper.

II. Review of Space-Time Shift Keying

In Sections II-A, II-B and II-C, we briefly review the STSK scheme's model as well as its capacity- and BER-limits, respectively, noting that their further detailed descriptions can be found for example in [1, 2].

A. System Model

Consider the M-antenna-assisted STSK transmitter, having Q preassigned space-time dispersion matrices $\mathbf{A}_q \in \mathbb{C}^{M\times T}$ $(q = 1, \cdots, Q)$ as well as the $\mathcal{L}$-point complex-valued constellations s_l $(l = 1, \cdots, \mathcal{L})$. During each block interval, $B = \log_2(Q \cdot \mathcal{L}) = \log_2 Q + \log_2 \mathcal{L}$ information bits are input to the transmitter, where a single one out of the Q dispersion matrices $\mathbf{A}_q$ is activated according to the $\log_2 Q$ input bits, while the $\log_2 \mathcal{L}$ input bits are mapped to a symbol s_l. Finally, the transmitter arrives at the space-time matrix $\mathbf{S}_{q,l} = s_l \mathbf{A}_q \in \mathbb{C}^{M\times T}$, which is transmitted from each of the M transmit antenna elements (AEs) over T symbol durations.

The signals $\mathbf{Y} \in \mathbb{C}^{N\times T}$ received by the N receive AEs, may be expressed as

$$\mathbf{Y} = \mathbf{H}\mathbf{S}_{q.l} + \mathbf{V}, \tag{1}$$

[2]In this paper we mainly focus our attention on a new optimization algorithm of the STSK family, which is invoked in the context of near-capacity channel-encoded scenarios. For readers who are interested in the STSK scheme's performance comparisons with diverse other classic MIMO schemes, please refer to [2] and references therein.

where $\mathbf{H} \in \mathbb{C}^{N\times M}$ represents the Rayleigh fading channel coefficients, modeled by the complex-valued Gaussian distribution of $\mathcal{CN}(0, 1)$, while $\mathbf{V} \in \mathbb{C}^{N\times T}$ represents the associated noise components, obeying the complex-valued Gaussian distributions of $\mathcal{CN}(0, N_0)$, with N_0 being noise variance. By carrying out the vector-stacking operation $vec(\bullet)$ at both sides of Eq. (1), we arrive at a more tractable vectorial signal model as follows:

$$\bar{\mathbf{Y}} = \bar{\mathbf{H}}\chi\mathbf{K}_{q.l} + \bar{\mathbf{V}}, \tag{2}$$

where we have

$$\bar{\mathbf{Y}} = vec(\mathbf{Y}) \in \mathbb{C}^{NT\times 1}, \tag{3}$$

$$\bar{\mathbf{V}} = vec(\mathbf{V}) \in \mathbb{C}^{NT\times 1}, \tag{4}$$

$$\bar{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H} \in \mathbb{C}^{NT\times MT}, \tag{5}$$

$$\chi = [vec(\mathbf{A}_1), \cdots, vec(\mathbf{A}_Q)] \in \mathbb{C}^{MT\times Q}, \tag{6}$$

and

$$\mathbf{K}_{q,l} = [0, \cdots, 0, \underset{\uparrow}{s_l}, 0, \cdots, 0]^{\mathrm{T}} \in \mathbb{C}^{Q\times 1}. \tag{7}$$

$$q\text{th element}$$

Furthermore, $\otimes$ denotes the Kronecker product and $\mathbf{I}$ is the identity matrix, while χ is referred to as the dispersion character matrix (DCM) [13], which characterizes the set of dispersion matrices employed. In order to maintain a transmission power per symbol duration which is unity, the following constraint is imposed on the DCM:

$$[\chi]_{(q)}^{H} [\chi]_{(q)} = T \quad (q = 1, \cdots, Q), \tag{8}$$

where $[\bullet]_{(q)}$ denotes the qth column of $\bullet$.

Note that the above-mentioned equivalent signal vector $\mathbf{K}_{q,l}$ of Eq. (7) contains only one non-zero element in the qth position, hence the corresponding signals received at the STSK receiver are free from inter-element interference (IEI). This unique signal structure allows us to conceive the low-complexity detector of [5], which is capable approaching the optimal maximum likelihood (ML) detector's performance.

Similarly to the nomenclature used in [1, 2, 5], we employ the notation of 'STSK(M, N, T, Q)' in the rest of this paper, where again, the space-time (ST) matrix $\mathbf{S}_{q,l} = s_l \mathbf{A}_q$ is transmitted from the M AEs over T time slots.

B. Capacity of STSK

The classic continuous-input continuous-output memoryless channel (CCMC) capacity of general MIMOs may be formulated as [21]

$$C_{\mathrm{CCMC}} = \mathbb{E}\left[\log_2 \det\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{H}^H}{N_0}\right)\right], \tag{9}$$

where $\mathbb{E}[\bullet]$ is the expectation operation. Let us note that the CCMC capacity is not affected by the specific constellations employed, because it is based on the assumption of having a Gaussian input signal. Hence, this ergodic capacity cannot be used as the metric, which optimizes the constellation s_l $(l = 1, \cdots, \mathcal{L})$ of the STSK system.

On the other hand, the STSK scheme's DCMC capacity is given by [1]

$$C_{\mathrm{DCMC}} = \frac{1}{T}\left(B - \frac{1}{2^B}\sum_{b=1}^{2^B}\mathbb{E}\left[\log_2\left\{\sum_{b'=1}^{2^B}\exp\left(\Psi^{(b,b')}\right)\right\}\right]\right), \quad (10)$$

where we have

$$\Psi^{(b,b')} = \frac{-\left\|\bar{\mathbf{H}}\chi(\mathbf{K}_{g(b)} - \mathbf{K}_{g(b')}) + \bar{\mathbf{V}}\right\|^2 + \left\|\bar{\mathbf{V}}\right\|^2}{N_0}, \quad (11)$$

while $g(b)$ represents a pair of (q, l) associated with the index b ($1 \leq b \leq 2^B$) of the input bits. In contrast to the CCMC capacity, the DCMC capacity reflects the effects of classic constellations, as suggested by Eq. (10). However, its calculation requires the computationally expensive expectation evaluation. This implies that the direct random-search based maximization of the DCMC capacity is intractable for high-dimensional STSK scenarios. To overcome this limitation, we will derive the theoretical gradients of the DCMC capacity C_{DCMC} with respect to both the DCM χ and to the constellation s_l ($l = 1, \cdots, \mathcal{L}$) later in Section III and in the Appendix.

C. Unified PEP Upper Bound

Let us define the PEP as $P(\mathbf{S} \to \mathbf{S}')$, which indicates the probability that a codeword $\mathbf{S}$ is erroneously decoded as $\mathbf{S}'$. According to the block-based system model of our STSK scheme, which was formulated in Eq. (1), we arrive at the corresponding PEP conditioned on the channel matrix $\mathbf{H}$ as follows:

$$\begin{aligned} P(\mathbf{S} \to \mathbf{S}'|\mathbf{H}) &= \Pr\left(\|\mathbf{Y} - \mathbf{H}\mathbf{S}'\| - \|\mathbf{Y} - \mathbf{H}\mathbf{S}\| < 0\right) \\ &= Q\left[\sqrt{\frac{\|\mathbf{H}\Delta\|^2}{2N_0}}\right], \end{aligned} \quad (12)$$

where we have $\Delta = \mathbf{S} - \mathbf{S}'$ and $Q[\bullet]$ represents the integral form of the Q-function. Then, according to [22], the unconditional PEP may be obtained by averaging the conditional PEP over the legitimate range of potential channel components as:

$$P(\mathbf{S} \to \mathbf{S}') = \frac{1}{\pi}\int_0^{\pi/2}\prod_{m=1}^{M}\left(1 + \frac{\mu_m}{4N_0\sin^2\theta}\right)^{-N}\mathrm{d}\theta, \quad (13)$$

where μ_m is the mth eigenvalue of $\Delta\Delta^H$. Having arrived at the exact PEP expression, we can now compute the tight upper bound on the average BER, by summing the PEP over all error events corresponding to a given transmitted codeword $\mathbf{S}$ [21], which is given by

$$P_{\mathrm{e,bit}}(\mathbf{S}) \leq \frac{1}{B}\sum_{\mathbf{S}\neq\mathbf{S}'} d(\mathbf{S}, \mathbf{S}')P(\mathbf{S} \to \mathbf{S}'), \quad (14)$$

where $d(\mathbf{S}, \mathbf{S}')$ represents the Hamming distance between $\mathbf{S}$ and $\mathbf{S}'$. Finally, averaging over all the legitimate 2^B codewords $\mathbf{S}$, the upper bound of the averaged BER $\bar{P}_{\mathrm{e,bit}}$ is given by

$$\begin{aligned} \bar{P}_{\mathrm{e,bit}} &\leq \mathbb{E}\left[\frac{1}{B}\sum_{\mathbf{S}\neq\mathbf{S}'} d(\mathbf{S}, \mathbf{S}')P(\mathbf{S} \to \mathbf{S}')\right] \\ &= \frac{1}{B\cdot 2^B}\sum_{\mathbf{S}}\sum_{\mathbf{S}\neq\mathbf{S}'} d(\mathbf{S}, \mathbf{S}')P(\mathbf{S} \to \mathbf{S}'). \end{aligned} \quad (15)$$

Alternatively, Eq. (15) may be rewritten as

$$\bar{P}_{\mathrm{e,bit}} \leq \frac{1}{B\cdot 2^B}\sum_{i=1}^{B}\sum_{b=0}^{1}\sum_{\forall\mathbf{S}_{(i,b)}}\sum_{\forall\mathbf{S}_{(i,\bar{b})}} P\left(\mathbf{S}_{(i,b)} \to \mathbf{S}_{(i,\bar{b})}\right), \quad (16)$$

where $\mathbf{S}_{(i,b)}$ represents STSK signals, having the ith bit ($1 \leq i \leq B$) value of $b = \{0, 1\}$, while $\bar{b}$ is the inverse of the bit b. Note that Eq. (16) assumes the absence of *a priori* information.

III. DCM and Constellation Optimization

In this section, we propose the joint optimization of the dispersion-matrix and of the constellation sets.

Firstly, in order to have an initial set of $[\chi(1), \boldsymbol{s}(1)]$, we carry out a Monte Carlo (MC) simulation based random search. To be specific, N_{MC} number of $(\chi, \boldsymbol{s})$ sets are randomly generated and then the set exhibiting the maximum DCMC-capacity value is chosen as the initial set $[\chi(1), \boldsymbol{s}(1)]$.

Based on $[\chi(1), \boldsymbol{s}(1)]$, we maximize the STSK scheme's parameters with the aid of the SCG algorithm [23, 24] as follows:

1) *Initialization:* Set a step size of $\mu > 0$ and the termination scaler of $\beta > 0$; given the initial matrix of $\chi(1)$ and the constellation of $\boldsymbol{s}(1)$, set $\mathbf{D}(1) = \nabla_\chi C_{\mathrm{DCMC}}(\chi(1), \boldsymbol{s}(1)) \in \mathbb{C}^{MT\times Q}$, $\mathbf{G}(1) = \nabla_s C_{\mathrm{DCMC}}(\chi(1), \boldsymbol{s}(1)) \in \mathbb{C}^{\mathcal{L}\times 1}$ and $l = 1$.
2) *Loop:* If $\|(\nabla_\chi + \nabla_s)C_{\mathrm{DCMC}}(\chi(l), \boldsymbol{s}(l))\| < \beta$, goto *Stop*.

$$\chi(l+1) = \chi(l) + \mu\mathbf{D}(l) \quad (17)$$

$$[\chi(l+1)]_q = \sqrt{T}\cdot\frac{[\chi(l+1)]_q}{\left\|[\chi(l+1)]_q\right\|} \quad (q = 1, \cdots, Q) \quad (18)$$

$$\boldsymbol{s}(l+1) = \boldsymbol{s}(l) + \mu\mathbf{G}(l) \quad (19)$$

$$\boldsymbol{s}(l+1) = \sqrt{\mathcal{L}}\cdot\frac{\boldsymbol{s}(l+1)}{\|\boldsymbol{s}(l+1)\|} \quad (20)$$

$$\phi_l = \frac{\|\nabla_\chi C_{\mathrm{DCMC}}(\chi(l+1), \boldsymbol{s}(l+1))\|^2}{\|\nabla_\chi C_{\mathrm{DCMC}}(\chi(l), \boldsymbol{s}(l+1))\|^2} \quad (21)$$

$$\mathbf{D}(l+1) = \phi_l\mathbf{D}(l) + \nabla_\chi C_{\mathrm{DCMC}}(\chi(l+1), \boldsymbol{s}(l+1)) \quad (22)$$

$$\psi_l = \frac{\|\nabla_s C_{\mathrm{DCMC}}(\chi(l+1), \boldsymbol{s}(l+1))\|^2}{\|\nabla_s C_{\mathrm{DCMC}}(\chi(l), \boldsymbol{s}(l+1))\|^2} \quad (23)$$

$$\mathbf{G}(l+1) = \psi_l\mathbf{G}(l) + \nabla_s C_{\mathrm{DCMC}}(\chi(l+1), \boldsymbol{s}(l+1)) \quad (24)$$

$l = l + 1$, goto *Loop*.
3) *Stop:* $\chi(l)$ and $\boldsymbol{s}(l)$ are the solution.

Here, $\|\bullet\|$ represents the Frobenius norm, while we have

$$\nabla_\chi = \left[\begin{array}{ccc} \frac{\partial}{\partial[\chi]_1} & \cdots & \frac{\partial}{\partial[\chi]_Q} \end{array}\right], \quad \nabla_s = \left[\begin{array}{ccc} \frac{\partial}{\partial s_1} & \cdots & \frac{\partial}{\partial s_{\mathcal{L}}} \end{array}\right]^{\mathrm{T}}. \quad (25)$$

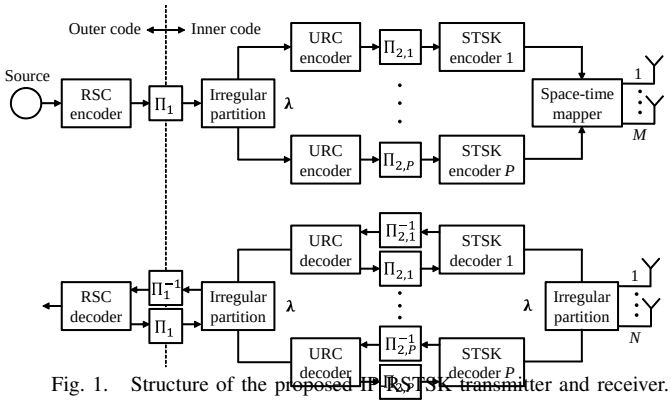

Fig. 1. Structure of the proposed IR-PSTSK transmitter and receiver.

In order to implement Eqs. (21)–(24) of the above-mentioned SCG algorithm, the gradient calculations of $\nabla_\chi C_{\rm DCMC}$ and $\nabla_s C_{\rm DCMC}$ are needed. To avoid the computationally intensive problem associated with their perturbation-based calculations, we derive the theoretical values in the Appendix.

Since typically a significantly higher step size μ is used in the conjugate gradient algorithm in comparison to the steepest gradient algorithm, the convergence of the SCG is expected to be more rapid. Additionally, as mentioned in [25], it is beneficial to periodically reset the values of ϕ_l and ψ_l either to zeros or to their negative counterparts for the sake of avoiding convergence to a local optimum.

IV. EXIT-Chart Aided IR-PSTSK Design

In this section, we introduce our IR-PSTSK architecture, which is composed of multiple STSK subcodes designed with the aid of our optimization algorithm of Section III.

A. IR-PSTSK Transceiver

Fig. 1 shows the structure of our three-stage-concatenated recursive systematic convolutional (RSC) coded and unity-rate-convolutional (URC) coded IR-PSTSK structure, including P different STSK component codes, each optimized by the algorithm of Section III, in order to maximize the associated DCMC capacity. We note that since IR-PSTSK consists of linear combinations of STSK components codes, each component code, optimized with the aid of our proposed algorithm of Section III, contributes to the direct improvement of the resultant IR-PSTSK scheme's DCMC capacity.

The transmitter firstly channel-encodes the source information bits by the RSC code, which are then interleaved by the first interleaver Π_1. The interleaved bits are divided into P parallel bit sequences with the aid of a weighting coefficient vector $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_P]$, where we have the relationship of $\sum_{p=1}^{P} \lambda_p = 1$. The pth bit-sequence is encoded by the URC code and the encoded bits are interleaved again by the interleaver $\Pi_{2,p}$. The interleaved bits are then mapped to STSK symbols. Finally, the P different STSK signal streams are combined, which are transmitted by the M transmit antennas.

At the receiver of Fig. 1, the $(2P+1)$ soft-input soft-output (SISO) decoders iteratively exchange their extrinsic information. More specifically, $I_{\rm in}$ number of inner iterations are carried out between each STSK decoder and the associated URC decoder, per outer iteration. Therefore, the total number of iterations becomes $(I_{\rm in} \cdot I_{\rm out})$, where $I_{\rm out}$ is the number of outer iterations.

Provided that the normalized transmission rate of each STSK subcode is R_p $(p = 1, \cdots, P)$, then we arrive at the IR-PSTSK's normalized transmission rate in the form of

$$R = \frac{R_{\rm CC}}{\sum_{p=1}^{P} \lambda_p / R_p} \ [\text{bits/symbol}], \tag{26}$$

where $R_{\rm CC}$ represents the rate of the RSC code.

B. EXIT Chart Based Design

In order to realize our near-capacity IR-PSTSK architecture, we determine the weighting coefficients $\boldsymbol{\lambda}$ with the aid of EXIT charts [13]. As shown in [20], the average EXIT function of an irregular code having P component codes may be approximated by their linear combinations. Hence, considering the pth inner-code's EXIT function at the SNR of interest ρ to be $\Gamma_p(I_{\rm A}, \rho)$, we arrive at the resultant IR-PSTSK's inner-EXIT function $\Gamma_{\rm IR-PSTSK}(I_{\rm A}, \rho)$ in the form of

$$\Gamma_{\rm IR-PSTSK}(I_{\rm A}, \rho) = \sum_{p=1}^{P} \lambda_p \Gamma_p(I_{\rm A}, \rho). \tag{27}$$

From Eq. (27), the weighting coefficients $\boldsymbol{\lambda}$ are designed with the aid of exhaustive search, using the step size of τ, so as to match the inner-code's EXIT curve to the outer-code's EXIT curve as closely as possible.[3]

V. Performance Results

In this section, we provide our simulation results in order to characterize the proposed SCG-based optimization algorithm as well as the achievable performance of our IR-PSTSK scheme.

Firstly, Fig. 2 shows the convergence characteristics of our optimization algorithm proposed in Section III. More specifically, we considered the following four STSK scenarios: $(\mathcal{L} = 4)$-STSK$(4, 4, 4, 4)$, $(\mathcal{L} = 8)$-STSK$(4, 4, 1, 4)$, $(\mathcal{L} = 16)$-STSK$(4, 4, 3, 16)$ and $(\mathcal{L} = 16)$-STSK$(4, 4, 1, 8)$, which were characterized at SNR = 0 dB. Observe in Fig. 2 that in each scenario the DCMC capacity is monotonically increased upon increasing the number of iterations. This is due to the explicit benefit of employing the theoretical gradient values derived in the Appendix. Note that 'non-smooth' points of the convergence curves correspond to the ones, where the values of ϕ_l and ψ_l were reset to zero in order to circumvent convergence to a local optimum, as described in Section III. Additionally, the initial capacity points of Fig. 2, which were designed with the aid of MC simulations, correspond to the conventional performance results. Hence, our iterative

[3]The total search-space of our IR-PSTSK scheme's exhaustive search is $\prod_{i=1}^{P}(\alpha - i + 1)$, where α denotes the number of quantization levels of each weighting coefficient $0 \leq \lambda_i \leq 1$.The detailed optimization process of weighting coefficients can also be found in Chapter 7 of [13].

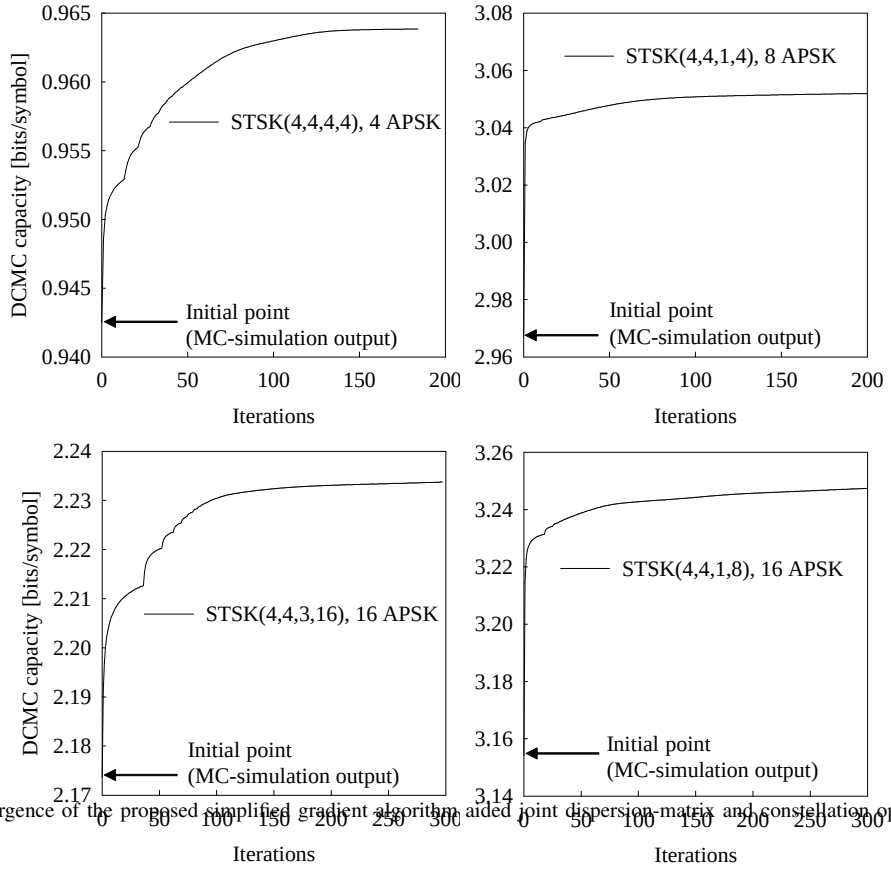

Fig. 2. Convergence of the proposed simplified gradient algorithm aided joint dispersion-matrix and constellation optimization at SNR = 0 dB.

optimization schemes exhibited clear performance advantages over the conventional random-search algorithm.[4]

Additionally, in Fig. 3 we investigated the effect of step size μ on the convergence speed, while varying μ 0.1 to 5. Here, we considered the $(\mathcal{L} = 8)$-STSK$(2, 2,$ arrangement. It can be seen for Fig. 3 that regardles the value of μ, each SCG optimization converged to a function value of 2.62, while exhibiting different converg speeds. Similarly, our additional extensive simulations sh that the converged cost-function value remains unaffecte the choice of the step size μ.

Next, we designed our IR-PSTSK scheme by consid $P = 8$ different STSK component codes, which are l in Table I. To be specific, the component-code set exh different transmission rates, ranging from $R = 1$ bits/symbol. Then, the weighting coefficients $\boldsymbol{\lambda}$ were desi with the aid of EXIT charts as well as using an exhau search by employing the step size of $\tau = 0.01$. N specifically, the gap area between the IR-PSTSK scheme's inner code's EXIT curve and the half-rate RSC(2.3.2) outer code's EXIT curve was minimized at the SNR of -4.8 dB, as shown in Fig. 4. Here, the area between the inner and outer

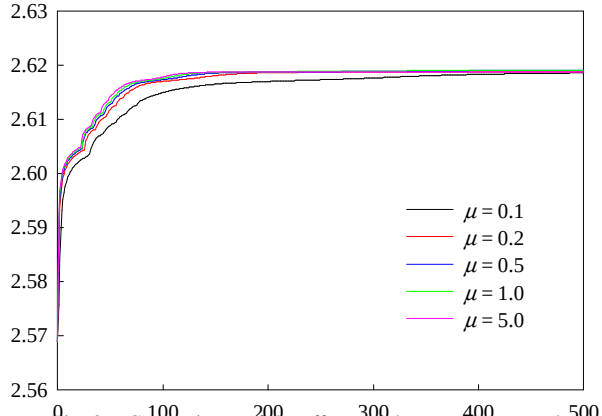

Fig. 3. Comparisons of the effects of the step size μ on the convergence behavior, while considering the 8-APSK aided STSK$(2, 2, 2, 8)$ scheme recorded at SNR = 4 dB.

codes' EXIT curves was as low as 0.0144 and the weighting coefficients $\boldsymbol{\lambda}$ of the $P = 8$ components were given by $\boldsymbol{\lambda} = [0.08\ 0.54\ 0.05\ 0.00\ 0.04\ 0.13\ 0.00\ 0.16]$. The inner code rate of IR-PSTSK and the resultant normalized transmission rate were 1.92 and 0.96 bits/symbol, respectively. In this

[4]Here, we invoked Monte Carlo simulations in order to attain the initial condition of our SCG-based optimization. However, several other random search algorithms, such as genetic algorithms (GA), may be readily employed for calculating good initial parameters in a more efficient manner.

TABLE I
A SET OF STSK COMPONENT CODES ($P = 8$) FOR $M = N = 4$ TRANSMIT AND RECEIVE ANTENNAS

Index	M	N	T	Q	$\mathcal{L}$	D	Rate	Weight λ
#1	4	4	4	4	4	16	1.0	0.0
#2	4	4	4	8	8	16	1.5	0.5
#3	4	4	4	16	16	16	2.0	0.0
#4	4	4	3	16	16	12	2.7	0.0
#5	4	4	2	16	16	8	4.0	0.0
#6	4	4	1	4	8	4	5.0	0.1
#7	4	4	1	8	16	4	7.0	0.0
#8	4	4	1	16	16	4	8.0	0.1

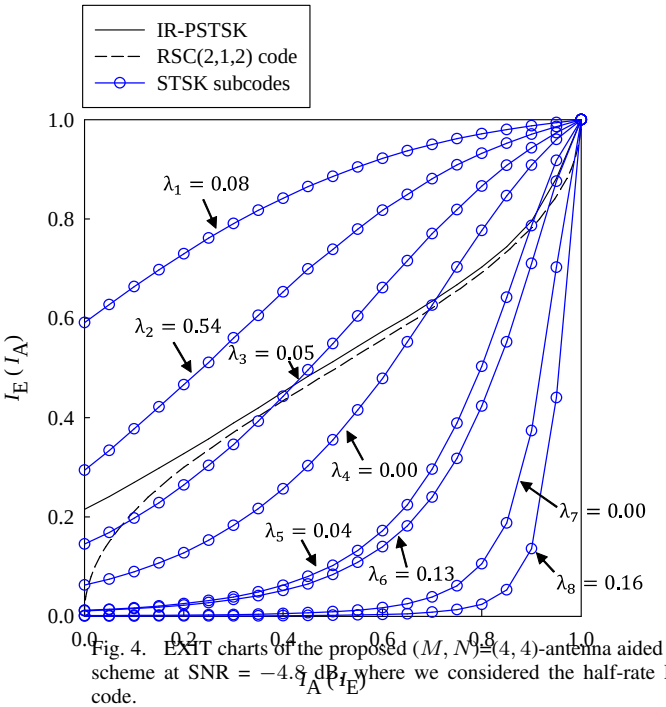

Fig. 4. EXIT charts of the proposed $(M, N)=(4, 4)$-antenna aided IR-PSTSK scheme at SNR = -4.8 dB, where we considered the half-rate RSC(2,1,2) code.

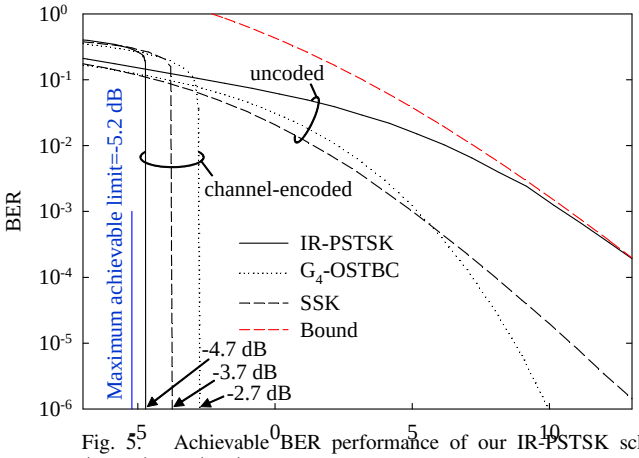

Fig. 5. Achievable BER performance of our IR-PSTSK scheme having $(M, N) = (4, 4)$ antenna elements, compared with the G_4 OSTBC and SSK schemes, employing the three-stage RSC-coded and URC-precoded arrangement of Fig. 1. Here, the BER curves of the uncoded counterpart of each scheme was also calculated, while plotting the associated theoretical BER bound of our uncoded IR-PSTSK scheme.

paper, we assumed the employment of the high-complexity MAP algorithm for the STSK detector at the receiver. On the other hand, it is possible to employ the reduced-complexity STSK detector of [5], by imposing additional constraint of the symmetric property over I- and Q-axes on the optimization of the APSK constellations.

Considering that a narrow open tunnel appeared between the two EXIT curves and that the inner code's EXIT curve reached the point of perfect convergence at $(I_A, I_E) = (1, 1)$, it is predicted that a vanishingly low BER may be achieved with the aid of a sufficiently high number of iterations. Similarly, we may be able to optimize our near-capacity IR-PSTSK scheme over a wide range of SNRs. Additionally, we note that It is also readily possible to employ irregular channel coding (IRCC) scheme with our IR-PSTSK scheme, similarly to [20], which enables further flexible matching of inner and outer EXIT curves. However, the performance improvement is considered to be marginal for the scenario of Fig. 4, since the gap area is already significantly small.

In order to provide further insights, in Fig. 5 the BER curve of our IR-PSTSK of Fig. 4 was compared to those of other MIMO benchmark schemes, such as the 16-QAM aided G_4-orthogonal space-time block coding (OSTBC) scheme [26] and the SSK scheme [6], each exhibiting the normalized transmission rate of 1 bits/symbol. Here, a low-complexity symbol-based SISO MAP detector [27] was invoked for our IR-PSTSK and SSK schemes, while the optimal SISO MAP detector [13] was used for the OSTBC scheme. The maximum number of inner and outer iterations was set to $I_{\text{in}} = 70$ and $I_{\text{out}} = 2$, respectively and the interleaver length was 200 000 bits per frame. For comparison, we also characterized the uncoded counterparts of the above-mentioned schemes, where the optimal ML detector was employed.[5] Observe in Fig. 5 that in channel-encoded scenarios, the proposed IR-PSTSK scheme attained an infinitesimally low BER at SNR = -4.7 dB as expected from the EXIT charts of Fig. 4, which outperformed the other two channel-encoded benchmarkers, exhibiting BER cliffs at SNRs of -3.7 dB and -2.7 dB, respectively. By contrast, in uncoded scenarios each benchmark scheme, i.e. the uncoded OSTBC and SSK scheme, exhibited a better performance than our IR-PSTSK scheme. This suggests that the classic STCs do not provide a near-capacity performance in practical channel-encoded scenarios, while our IR-PSTSK scheme has the explicit benefit of facilitating a flexible EXIT-chart-based system design.

[5] In order to validate the calculated BERs we also plotted the theoretical bound, which was derived in Section II-C. More specifically, by considering the ith STSK subcode $\bar{P}_i$ $(i = 1, \cdots, P)$ of Eq. (15), the theoretical BER bound of our IR-PSTSK scheme is given by $\bar{P}_{\text{IR-PSTSK}} = \sum_{i=1}^{P} \lambda_i \bar{P}_i$.

VI. Conclusions

In this paper, we proposed the joint optimization of the STSK dispersion matrices and constellations, which was invoked for designing a near-capacity IR-PSTSK scheme. More specifically, our derivation of the theoretical gradient of the DCMC capacity facilitates an efficient SCG-algorithm based optimization. Furthermore, the IR-PSTSK scheme, which is constituted by the STSK subcodes optimized with the aid of our algorithm, was found to attain a near-capacity performance.

Appendix
Gradient Derivation

In this appendix, we derive the theoretical gradient matrix of the STSK scheme's DCMC capacity $\nabla_\chi C_{\mathrm{DCMC}} \in \mathbb{C}^{MT\times Q}$ with respect to the equivalent DCM χ, where we define a $(MT \times Q)$-element gradient matrix of $\nabla_\chi = [\nabla_1, \cdots, \nabla_Q]$ and $\nabla_q = \partial/\partial[\chi]_q$. Then, the qth vector of the gradient matrix $\nabla_q C_{\mathrm{DCMC}} \in \mathbb{C}^{MT\times 1}$ may be written as

$$\nabla_q C_{\mathrm{DCMC}} = \frac{\partial C_{\mathrm{DCMC}}}{\partial [\chi]_q} \tag{28}$$

$$= -\frac{1}{T\cdot 2^B}\nabla_q \left(\sum_{b=1}^{2^B} \mathbb{E}\left[f(b)\right]\right) \tag{29}$$

with the relationship of

$$f(b) = \log_2\left[\sum_{b'=1}^{2^B} \exp\left(\Psi^{(b,b')}\right)\right]. \tag{30}$$

Here, let us assume that the expectation operation is simplified to the average operation. Then, Eq. (29) may be rewritten as

$$\nabla_q C_{\mathrm{DCMC}} = -\frac{1}{T\cdot 2^B}\mathbb{E}\left[\sum_{b=1}^{2^B} \nabla_q f(b)\right]. \tag{31}$$

Furthermore, $\nabla_q f(b)$ of Eq. (31) can be expressed as

$$\nabla_q f(b) = \nabla_q \log_2\left[\sum_{b'=1}^{2^B} \exp\left(\Psi^{(b,b')}\right)\right] \tag{32}$$

$$= \frac{\nabla_q\left[\sum_{b'=1}^{2^B} \exp\left(\Psi^{(b,b')}\right)\right]}{(\ln 2)\sum_{b'=1}^{2^B} \exp\left(\Psi^{(b,b')}\right)} \tag{33}$$

$$= \frac{\sum_{b'=1}^{2^B} \exp\left(\Psi^{(b,b')}\right)\nabla_q \Psi^{(b,b')}}{(\ln 2)\sum_{b'=1}^{2^B} \exp\left(\Psi^{(b,b')}\right)}. \tag{34}$$

Moreover, $\nabla_q \Psi^{(b,b')}$ of Eq. (34) may be calculated from Eq. (11) as follows:

$$\nabla_q \Psi^{(b,b')} = -\frac{1}{N_0}\nabla_q\left\{\bar{\mathbf{H}}\chi(\mathbf{K}_{g(b)} - \mathbf{K}_{g(b')}) + \bar{\mathbf{V}}\right\}^H \times \left\{\bar{\mathbf{H}}\chi(\mathbf{K}_{g(b)} - \mathbf{K}_{g(b')}) + \bar{\mathbf{V}}\right\} \tag{35}$$

$$= -\frac{2}{N_0}\left[\mathbf{K}^*_{g(b)} - \mathbf{K}^*_{g(b')}\right]_{(q)}\bar{\mathbf{H}}^H \times \left\{\bar{\mathbf{H}}\chi(\mathbf{K}_{g(b)} - \mathbf{K}_{g(b')}) + \bar{\mathbf{V}}\right\}, \tag{36}$$

where $[\bullet]_{(q)}$ denotes the qth column. Finally, we arrive at $\nabla_\chi C_{\mathrm{DCMC}} = [\nabla_1 C_{\mathrm{DCMC}}, \cdots, \nabla_Q C_{\mathrm{DCMC}}]$ from Eqs. (31), (34) and (36).

In order to expound a little further, the derivation of the above-mentioned gradient matrix $\nabla_\chi C_{\mathrm{DCMC}}$ allows us to significantly reduce the computational complexity, under the assumption of the SCG algorithm of Section III. More specifically, the perturbation-based approach requires $(2MTQ+1)$ number of DCMC-capacity calculations in order to obtain a single gradient matrix $\nabla_\chi C_{\mathrm{DCMC}}$, while only the complexity equivalent to a single DCMC-capacity calculation is necessary for our theoretical gradient matrix. [6]

Similarly to Eqs. (31)–(36), we may also be able to derive the gradient vector of symbol constellations $\nabla_s C_{\mathrm{DCMC}} = [\partial/\partial s_1, \cdots, \partial/\partial s_{\mathcal{L}}]^T C_{\mathrm{DCMC}} \in \mathbb{C}^{\mathcal{L}\times 1}$, which is formulated as

$$\nabla_s C_{\mathrm{DCMC}} = -\frac{1}{T\cdot 2^B}\mathbb{E}\left[\sum_{b=1}^{2^B} \frac{\sum_{b'=1}^{2^B} \exp\left(\Psi^{(b,b')}\right)\nabla_s \Psi^{(b,b')}}{(\ln 2)\sum_{b'=1}^{2^B} \exp\left(\Psi^{(b,b')}\right)}\right], \tag{37}$$

where we have the relationship of

$$\frac{\partial \Psi^{(b,b')}}{\partial s_{\tilde{l}}} = -\frac{2}{N_0}\left(\delta_{(\tilde{l},l)}\mathbf{H}\left[\lambda\right]_{(q)} - \delta_{(\tilde{l},l')}\mathbf{H}\left[\lambda\right]_{(q')}\right)^H \times \left\{\bar{\mathbf{H}}\chi(\mathbf{K}_{q,l} - \mathbf{K}_{q',l'}) + \bar{\mathbf{V}}\right\} \quad (1 \le \tilde{l} \le \mathcal{L}), \tag{38}$$

$$\delta_{(l_1,l_2)} = \begin{cases} 1 & (l_1 = l_2) \\ 0 & (l_1 \neq l_2) \end{cases}, \tag{39}$$

while assuming that $g(b)$ and $g(b')$ correspond to the associated sets of (q,l) and (q',l'), respectively.

References

[1] 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, Nov. 2010.

[2] ——, "A universal space-time architecture for multiple-antenna aided systems," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, pp. 401–420, Second Quarter, 2012.

[3] ——, "Generalized space-time shift keying designed for flexible diversity-, multiplexing- and complexity-tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1144–1153, Apr. 2011.

[4] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.

[5] S. Sugiura, C. Xu, S. X. Ng, and L. Hanzo, "Reduced-complexity coherent versus non-coherent QAM-aided space-time shift keying," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3090–3101, Nov. 2011.

[6] J. Jeganathan, A. Ghrayeb, L. Szczecinski, and A. Ceron, "Space shift keying modulation for MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3692–3703, 2009.

[7] R. Y. Mesleh, H. Haas, S. Sinanovic, C. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2242, 2008.

[8] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1804–1824, 2002.

[6] For example, for the ($\mathcal{L} = 64$)-point APSK modulated STSK$(4, 4, 4, 64)$ scheme the derived theoretical gradient matrix allows us to attain approximately 2000 times lower computational complexity per $\nabla_\chi C_{\mathrm{DCMC}}$ calculation than the perturbation-based empirical solution.

[9] J. Wang, X. Wang, and M. Madihian, "On the optimum design of space-time linear-dispersion codes," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2928–2938, 2005.
[10] M. Jiang and L. Hanzo, "Unitary linear dispersion code design and optimization for MIMO communication systems," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 497–500, 2010.
[11] S. Sugiura, "Dispersion matrix optimization for space-time shift keying," *IEEE Communications Letters*, vol. 15, no. 11, pp. 1152–1155, Nov. 2011.
[12] S. X. Ng and L. Hanzo, "On the MIMO channel capacity of multidimensional signal sets," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 528–536, 2006.
[13] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation*. John Wiley and IEEE Press, 2009.
[14] L. Hanzo, T. Liew, B. Yeap, R. Y. S. Tee, and S. X. Ng, *Turbo Coding, Turbo Equalisation, and Space-Time Coding for Transmission over Fading Channels*. John Wiley and IEEE Press, 2011.
[15] S. Sugiura, S. Chen, and L. Hanzo, "MIMO-aided near-capacity turbo transceivers: Taxonomy and performance versus complexity," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, pp. 421–442, Second Quarter, 2012.
[16] S. Lin and D. Costello, *Error Control Coding: Fundamentals and Applications*. Prentice Hall, 2004.
[17] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 909–926, 1998.
[18] M. Tüchler, "Design of serially concatenated systems depending on the block length," *IEEE Transactions on Communications*, vol. 52, no. 2, pp. 209–218, 2004.
[19] R. G. Maunder and L. Hanzo, "Near-capacity irregular variable length coding and irregular unity rate coding," *IEEE Transactions on Wireless Communications*, vol. 8, no. 11, pp. 5500–5507, 2009.
[20] N. Wu and L. Hanzo, "Near-capacity irregular-convolutional-coding-aided irregular precoded linear dispersion codes," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2863–2871, 2009.
[21] J. Proakis, *Digital communications*. McGraw-Hill, New York, 2001.
[22] M. K. Simon and M. S. Alouini, *Digital communication over fading channels*. Wiley-IEEE Press, 2005.
[23] M. Bazaraa, H. Sherali, and C. Shetty, *Nonlinear programming: theory and algorithms*. New York: Wiley, 1993.
[24] S. Chen, A. Samingan, B. Mulgrew, and L. Hanzo, "Adaptive minimum-BER linear multiuser detection for DS-CDMA signals in multipath channels," *IEEE Transactions on Signal Processing*, vol. 49, no. 6, pp. 1240–1247, 2001.
[25] S. Chen, N. Ahmad, and L. Hanzo, "Adaptive minimum bit-error rate beamforming," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 341–348, 2005.
[26] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE Journal on selected areas in communications*, vol. 17, no. 3, pp. 451–460, 1999.
[27] S. Sugiura, C. Xu, S. X. Ng, and L. Hanzo, "Reduced-complexity iterative-detection aided generalized space-time shift keying," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 10, pp. 3656–3664, Oct. 2012.

Shinya Sugiura (M'06-SM'12) received the B.S. and M.S. degrees in aeronautics and astronautics from the Kyoto University, Kyoto, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in mobile communications from the University of Southampton, Southampton, UK, in 2010. Since 2004, he has been with Toyota Central R&D Laboratories, Inc., Japan, where his research has covered a range of areas in wireless communications, networking, signal processing and antenna design.

Dr. Sugiura authored/co-authored more than 50 refereed research publications, including 25 IEEE journal and magazine papers. He was awarded a number of distinctions, most recently the 2011 IEEE Communications Society Asia-Pacific Outstanding Young Researcher Award, the 2011 Ericsson Young Scientist Award and the 2008 IEEE Antennas and Propagation Society Japan Chapter Young Engineer Award.

Lajos Hanzo (M'91-SM'92-F'04) FREng, FIEEE, FIET, DSc received his degree in electronics in 1976 and his doctorate in 1983. During his 34-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 co-authored 20 John Wiley - IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1300+ research papers and book chapters at IEEE Xplore, acted as TPC Chair of IEEE conferences, presented keynote lectures and been awarded a number of distinctions. Currently he is directing an 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 IST Programme and the Mobile Virtual Centre of Excellence (VCE), UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also an IEEE Distinguished Lecturer as well as a Governor of both the IEEE ComSoc and the VTS. During 2008 - 2012 he served as the Editor-in-Chief of the IEEE Press and as Chaired Prof. also at Tsinghua University, Beijing. He holds the European Research Council's Advance Fellow grant. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk