INVITED PAPER

Sixty Years of Coherent Versus Non-Coherent Tradeoffs and the Road From 5G to Wireless Futures

CHAO XU[1], (Senior Member, IEEE), **NAOKI ISHIKAWA**[2], (Member, IEEE),
RAKSHITH RAJASHEKAR[1], (Senior Member, IEEE),
SHINYA SUGIURA[3], (Senior Member, IEEE),
ROBERT G. MAUNDER[1], (Senior Member, IEEE),
ZHAOCHENG WANG[4], (Senior Member, IEEE),
LIE-LIANG YANG[1], (Fellow, IEEE),
AND LAJOS HANZO[1], (Fellow, IEEE)

[1]School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.
[2]Graduate School of Information Sciences, Hiroshima City University, Hiroshima 731-3194, Japan
[3]Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan
[4]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

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

ABSTRACT Sixty years of coherent versus non-coherent tradeoff as well as the twenty years of coherent versus non-coherent tradeoff in Multiple-Input Multiple-Output (MIMO) systems are surveyed. Furthermore, the advantages of adaptivity are discussed. More explicitly, in order to support the diverse communication requirements of different applications in a unified platform, the 5G New Radio (NR) offers unprecendented adaptivity, abeit at the cost of a substantial amount of signalling overhead that consumes both power and the valuable spectral resources. Striking a beneficial coherent versus non-coherent tradeoff is capable of reducing the pilot overheads of channel estimation, whilst relying on low-complexity detectors, especially in high-mobility scenarios. Furthermore, since energy-efficiency is of salient importance both in the operational and future networks, following the powerful Index Modulation (IM) pholosophy, we conceive a holistic adaptive pholosophy striking the most appropriate coherent/non-coherent, single-/multiple-antenna and diversity/multiplexing tradeoffs, where the number of RF chains, the Peak-to-Average Power Ratio (PAPR) of signal transmission and the maximum amount of interference tolerated by signal detection are all taken into account. We demonstrate that this intelligent tripple-fold adaptivity offers significant benefits in next-generation applications of mmWave and Terahertz solutions, in space-air-ground integrated networks, in full-duplex techniques and in other sophisticated channel coding assisted system designs, where powerful machine learning algorithms are expected to make autonomous decisions concerning the best mode of operation with minimal human intervention.

INDEX TERMS Coherent, channel estimation, non-coherent, differential encoding, sphere decoding, decision-feedback, multiple-input multiple-output, differential space-time modulation, energy-efficiency, index modulation, peak-to-average power ratio, 5G New Radio, millimeter wave, physical layer, link budget, 6G, Terahertz, space-air-ground, full-duplex, channel coding, machine learning, artificial intelligence.

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

I. INTRODUCTION

Historically speaking, a new generation of mobile communication system has been conceived roughly every decade.

Following the initial commercialization of the analog 1G communication systems in the 1980s, the 2G Global System of Mobile Communications (GSM) was launched in the early 1990s in Europe, which eventually spread right across the globe. The reasons behind the remarkable success of 2G GSM were mainly twofold. First of all, GSM [1]–[3] was the first ever digital mobile radio system supporting international/global roaming, where digital error-correction codes were employed in order to improve the signal reception over narrowband channels. Secondly, the important process of standardization by regulators encouraged market access and regulated competitions. The advances in Code Division Multiple Access (CDMA) heralded the 3G [4]–[7] era in the late 2000s, when the proliferation of smart phones with apps that demand ever-increasing Internet access motivated unprecedented investment by operators to exploit the valuable spectral resources. The spectral-efficient Orthogonal Frequency Division Multiplexing (OFDM), which divides the frequency-selective wideband channel into overlapping but orthogonal multi-carrier narrowband subchannels became the key enabler for the 4G Long-Term Evolution (LTE) [8]–[10] in the 2010s. At the time of writing, following the recent digital and analog circuit breakthroughs in processing massive-bandwidth millimeter Wave (mmWave) signals, the deployment of 5G New Radio (NR) [11]–[16] is well underway. In this work, we endeavor to survey the PHYsical (PHY) layer techniques of the 5G NR, with a particular emphasis on coherent versus non-coherent tradeoffs [17]–[21], and then our visions on how 6G may evolve over the next 10-15 years are also presented.

A communication system is referred to as being "coherent", when the Channel State Information (CSI) is known at the receiver, which has to rely on CSI estimation techniques [10], [22], [23]. These techniques generally rely on training/pilot symbols that often absorb a non-negligible percentage of the valuable spectral resources and power budget [24]–[30]. Moreover, accurate CSI estimation becomes increasingly challenging in high-mobility scenarios, which either substantially erodes the performance of coherent detection or requires doubling the pilot-overhead used for sampling the channel's fading envelope. By contrast, when the CSI is not known at the receiver, "non-coherent" detection is encountered, which relies on its own form of blind CSI estimation. This is achieved by exploiting the correlation between the consecutive received signals imposed by the channel's memory. One of the most popular optimization criterion is constituted by Maximum Likelihood Sequence Estimation (MLSE) that jointly performs channel estimation and data detection, which dates back to Kailath's seminal work [31] in 1960. It was later further developed into a "estimator-correlator" technique by Kailath [32] in 1969. The MLSE tentatively performs channel estimation for each possible combination of the data-carrying sequence, and then these combinations are compared in order to make a joint decision on the sequence. However, the MLSE complexity increases exponentially with the message length, hence its

TABLE 1. Nomenclature (Part I).

ADPSK	Absolute-amplitude Differential Phase Shift Keying
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASTSK	Asynchronous Space Time Shift Keying
ATM	Air Traffic Management
BICM	Bit-Interleaved Coded Modulation
BLAST	Bell Laboratories Layered Space-Time
BS	Base Station
CC	Convolutional Code
CDD	Conventional Differential Detection
CDMA	Code Division Multiple Access
CIR	Channel Impulse Response
cmWave	centimeter Wave
CP	Cyclic Prefix
CQI	Channel Quality Information
CRC	Cyclic Redundancy Check
CSI	Channel State Information
DAC	Digital-to-Analog Converter
DAPSK	Differential Amplitude Phase Shift Keying
DFDD	Decision-Feedback Differential Detection
DL	DownLink
DLDC	Differential Linear Dispersion Code
DM-RS	DeModulation Reference Signal
dmWave	decimeter Wave
DPSK	Differential Phase Shift Keying
DSM	Differential Spatial Modulation
DSTBC	Differential Space-Time Block Code
DSTM	Differential Space-Time Modulation
DSTSK	Differential Space-Time Shift Keying
eMBB	enhanced Mobile BroadBand
EXIT	EXtrinsic Information Transfer
FD	Full Duplex
FDD	Frquency Division Duplex
FDMA	Frequency Division Multiple Access
FSO	Free Space Optical
GEO	Geostationary Earth Orbit
GS	Groud Station
GSTSK	Generalized Space-Time Shift Keying
HARQ	Hybrid Automatic Repeat Request
IAI	Inter-Antenna Interference
IBFD	In-Band Full-Duplex
IM	Index Modulation
IoT	Internet-of-Things
IRCC	IrRegular Convolutional Code
ISI	Inter-Symbol Interference
LDC	Linear Dispersion Code
LDPC	Low-Density Parity-Check
LEO	Low Earth Orbit
LLR	Log Likelihood Ratio
LTE	Long-Term Evolution
GMSK	Gaussian Minimum Shift Keying

practical employment has remained limited. Against this background, Forney [33] proposed to invoke the Viterbi algorithm for MLSE in 1972, where the number of trellis states that determines the MLSE complexity grows exponentially only with the channel's memory rather than with the message length. In 1979, Morley and Snyder [34] demonstrated that the MLSE is capable of tackling any form of channel memory, regardless whether the memory is imposed by frequency-selective or by time-selective channels, such as correlated Rayleigh, Rician and lognormal fading.

In analogy to the channel's memory, the philosophy of "modulation with memory" was exploited by the classic Differential Phase Shift Keying (DPSK) concept developed from Lawton's work [35], [36] in 1959-1960.

TABLE 2. Nomenclature (Part II).

GSM	Global System for Mobile Communications
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MDD	Multicarrier Division Duplex
MEO	Medium Earth Orbit
MIMO	Multiple-Input Multiple-Output
MLSE	Maximum Likelihood Sequence Estimation
MMSE	Minimum Mean Squared Error
MSDD	Multiple-Symbol Differential Detection
MSDSD	Multiple-Symbol Differential Sphere Decoding
mmWave	millimeter Wave
mMTC	massive Machine Type Communication
NGN	Next Generation Network
NOMA	Non-Orthogonal Multiple Access
NR	New Radio
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PAPR	Peak-to-Average Power Ratio
PBCH	Physical Broadcast Channel
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PRACH	Physical Random Access Channel
PSAM	Pilot Symbol Assisted Modulation
PSS	Primary Synchronization Signal
PT-RS	Phase-Tracking RS
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QoS	Quality-of-Service
RA	Receive Antenna
RB	Resource Block
RE	Resource Element
RS	Reference Signal
SCS	Sub-Carrier Spacing
SISO	Single-Input Single-Output
SM	Spatial Modulation
SRS	Sounding Reference Signal
SS	Synchronization Signal
SSS	Secondary Synchronization Signal
STBC	Space-Time Block Code
STSK	Space-Time Shift Keying
TA	Transmit Antenna
TC	Turbo Code
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UAV	Unmanned Aerial Vehicle
UE	User Equipment
UL	UpLink
URLLC	Ultra-Reliable Low-Latency Communication

More explicitly, the DPSK transmitter maps the data-carrying symbols onto the phase changes between consecutive transmitted symbols. The low-complexity Conventional Differential Detection (CDD) recovers the source information by observing the phase change between every pair of consecutive received samples. This implies that CDD is a special case of MLSE, where the channel's memory is simply truncated to a window-length of $N_w = 2$. However, it was demonstrated by Cahn [37] in 1959 that the CDD aided DPSK scheme suffers from a 3 dB performance penalty compared to its coherent MPSK counterpart, when assuming idealistic perfect CSI estimation. Moreover, it was discovered by Bello and Nelin [38] in 1962 that an irreducible error floor is encountered, when the fading channel fluctuates rapidly. In order to mitigate this problem, the Multiple-Symbol

Differential Detection (MSDD) [39]–[42] has been developed since the 1990s, where an improved MLSE memory of $N_w \geq 2$ is observed, and a total number of $(N_w - 1) \geq 1$ data-carrying symbols are jointly detected. It was demonstrated that the MSDD is capable of reducing the 3 dB performance penalty both for transmission over AWGN channels and over slowly fluctuating fading channels [39]–[42]. Furthermore, the CDD's error floor experienced in rapidly fading channels may also be mitigated by the MSDD [41], [42].

Therefore, in the absence of the training/pilot overheads, the non-coherent techniques are still capable of achieving a comparable performance to their best-performing coherent counterparts. Nonetheless, similar to the MLSE, the MSDD complexity grows exponentially with N_w. In order to alleviate this problem, reduced-complexity MSDD solutions may be conceived relying on decision feedback [43]–[48], sphere decoding [18], [49]–[51] and trellis decoding [52]–[56], where the former two techniques are termed as Decision-Feedback Differential Detection (DFDD) and Multiple-Symbol Differential Sphere Decoding (MSDSD), respectively. Instead of jointly detecting $(N_w - 1)$ data-carrying symbols in MSDD, the linear DFDD aims for detecting only a single new symbol at a time, while the remaining $(N_w - 2)$ symbols are given by the previous DFDD decisions. It was confirmed in [46], [57] that the non-coherent DFDD is equivalent to the pilot-based CSI estimation of coherent detection relying on the same Minimum Mean Squared Error (MMSE) Wiener filter. The only difference is that instead of relying on the known pilot symbols in the coherent regime, the CSI estimation in the non-coherent DFDD operates based on the decisions of the previously detected symbols. As a result, the DFDD is prone to error propagation, which erodes the optimal MLSE/MSDD performance. By contrast, both trellis based and sphere decoder based solutions are capable of matching the MSDD performance at a reduced complexity. Specifically, the MSDSD complexity is substantially lower when provided with a sufficiently high SNR and/or *a priori* information by a channel decoder, thanks to the early termination of tree search in the face of large separation of the constellation points. By contrast, trellis decoding exhibits a constant complexity that is determined solely by the number of trellis states.

At the time of writing, high-Doppler scenarios are routinely encountered, hence it is of prime importance to examine the coherent versus non-coherent tradeoffs in the light of their sixty-year development. As a departure from the 4G LTE, the 5G NR [16] aims for offering three major modes of operation, namely the enhanced Mobile BroadBand (eMBB), Ultra-Reliable Low-Latency Communication (URLLC) and massive Machine Type Communication (mMTC) modes, as seen in Fig. 1. Moreover, we will also speculate on how 6G may evolve over the next 10-15 years. In a nutshell, we envision that the 6G system will exploit new resources of the electromagnetic spectrum [58]–[62], including the expanded space-air-ground dimensions [63]–[67]. Furthermore, the 6G system
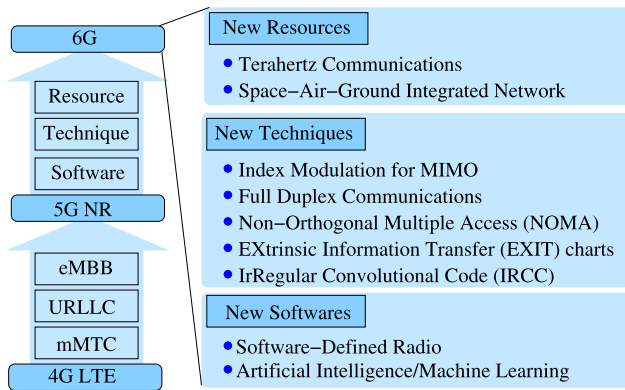
FIGURE 1. A vision of evolution to 6G.

might exploit Index Modulation (IM) aided Multiple-Input Multiple-Output (MIMO) techniques [68]–[79], Full-Duplex (FD) techniques [9], [80]–[87] and Non-Orthogonal Multiple Access (NOMA) [12], [88]–[90] for near-instantaneously optimizing the energy-efficiency, bandwidth-efficiency and power-efficiency, respectively. Moreover, enhanced channel coding arrangements such as IrRegular Convolutional Code (IRCC) [23], [91] and the powerful analysis tool of EXtrinsic Information Transfer (EXIT) charts [23], [91]–[93] are expected to play an active role in 6G scenarios. Finally, the state-of-the-art software-defined radio [94]–[97] and the machine learning tools relying on artificial intelligence [98]–[103] are envisioned to support the flexibility and autonomousity of 6G with minimal human intervention.

It is widely acknowledged that the majority of communication systems require some degree of coherent detection. Nonetheless, non-coherent detection is often beneficial in the following three scenarios: mitigating the pilot overhead, especially in systems having limited bandwidth, for low-complexity CDD in slowly-fluctuating channels as well as for supporting best-effort MSDD solutions in high-mobility scenarios. Moreover, the blind estimate of the CSI extracted by a non-coherent receiver can still be fed back to the transmitter for the sake of resource management of duplex scheduling, for spectrum allocation, for link budget adjustment, for Modulation and Coding Scheme (MCS) adaptivity and so on. Against this background, we survey the sixty-year history of coherent versus non-coherent tradeoffs, whilst identifying their potential applications in the future communication networks. More explicitly, the novel contributions of this paper are as follows:

1) First of all, we appraise the salient PHY layer techniques of the 5G NR standard [16], [104]–[112]. More explicitly, in order to support the *diverse* communication requirements in a *unified* platform, 5G NR is conceived to offer unprecendented adaptivity. This includes the scalable numerology concerning the Sub-Carrier Spacing (SCS) of OFDM, the dynamic Time Division Duplex (TDD) schedulling,

the flexible configurations of the control channels, the asynchronous and adaptive Hybrid Automatic Repeat Request (HARQ), etc. As a result, on one hand, the communication resources are flexibly tailored for different bandwidth, delay and Quality-of-Service (QoS) requirements of 5G applications. On the other hand, a substantial amount of overheads has to be invoked, which compromises the effective bandwidth efficiency in many critical communication tasks. Against this background, we identify the contributions of

(A) non-coherent detection in the inherent broadcast channels for initial access,

(B) blind detection in the broadcast and control channels as well as

(C) the benefit of non-coherent detection in high-mobility scenarios.

Moreover, we note that the non-coherent schemes are especially beneficial for broadcasting, paging and waking-up signals, where the pilot overhead and channel estimation latency of coherent detection are deemed excessive.

2) Secondly, in order to assess the pros and cons of the coherent versus non-coherent tradeoff, we look back into the sixty years of development history that covers the challenges in channel estimation, the non-coherent techniques as well as the bandwidth-efficient star QAM scheme. In particular, we show that the non-coherent detection problem even for Single-Input Single-Output (SISO) schemes is very much similar to the MIMO detection problem. More explicitly, the MSDD endeavors to make a joint decision based on N_w correlated symbols, while the MIMO ML detection jointly detects M interfering symbols transmitted by M Transmit Antennas (TAs). The MSDSD solutions and the sphere decoding aided MIMO detection obey the same tree-search philosophy. The DFDD invokes the same MMSE Wiener filter as employed for minimizing the Inter-Antenna Interference (IAI) in the state-of-the-art linearized MIMO detectors.

3) Thirdly, inspired by the success of MIMO techniques, we present a detailed review of the twenty years of coherent versus non-coherent tradeoff in MIMO systems. The three key MIMO tradeoffs of multiplexing/diversity tradeoff, performance/complexity tradeoff and single-/multiple-RF tradeoff are examined in both coherent and non-coherent scenarios. Most notably, since energy-efficiency is becoming increasingly important both in the operational and in the future networks, we pay special attention to the so-called 'infinite-cardinality problem' in the context of differential MIMO design. More explicitly, we prove mathematically that a range of conventional Differential Space-Time Modulation (DSTM) schemes that impose differential encoding on the MIMO transmit signal matrices [113]–[125] exhibit a

modulated signal constellation cardinality and Peak-to-Average Power Ratio (PAPR) tending to infinity over time, despite the fact that the practical Digital-to-Analog Converters (DACs) can only operate based on finite resolution, while the practical Power Amplifiers (PAs) only have limited dynamic range. The recently developed DSTM schemes [19], [20], [126]–[133] that follow the finite-cardinality low-PAPR design guidelines [19], [20] are summarized and compared. Moreover, the recent DSTM developments in massive MIMO setups [134]–[136] are also introduced.

4) Fourthly, our simulation results quantifying both the bandwidth-efficiency, as well as the power-efficiency and the link coverage once again confirm that the coherent and non-coherent schemes perform better in low- and high-Doppler scenarios, respectively. Furthermore, the upgrade from the single-TA to multiple-TA setups is shown to be more beneficial at near-unity channel coding rates, since the performance difference between the two setups is reduced in the face of a strong low-rate channel coding arrangement. Moreover, as expected, the diversity-oriented and multiplexing-oriented MIMO designs are shown to perform better in low- and high-throughput scenarios, respectively.

5) Finally, we elaborate on our vision concerning Next Generation Network (NGN). Compared to the technolgy-driven 1G/2G/3G/4G, the 5G is more resource and business driven, with a focus on exploiting new spectrum and the new URLLC and MTC applications. It may be anticipated that 6G may continue to follow a similar application-oriented trend. The potential benefits of the classic coherent versus non-coherent tradeoff as well as our recently developed finite-cardinality low-PAPR design [19], [20], [137], [138] may indeed influence NGNs. Just to name a few, firstly, when aiming for extremely high-rate beyond-100Gb/s transmission and reception [139]–[141], the hitherto unexploited mmWave and Terahertz bands [58]–[62] may be utilized for accommodating NGN serives. However, since the PA dynamic range is generally reduced as the carrier frequency is increased [58], the energy-efficient finite-cardinality designs of [19], [20], [137], [138] conceived for both coherent and non-coherent scenarios become promising candidates for NGNs. Furthermore, since the high-frequency channels' statistical characteristics change drastically depending on the weather-conditions for example, a certain degree of non-coherent design may improve the attainable degree of robustness against channel variations. Moreover, for the high-mobility vehicles including satellites in space, manned/unmanned aircraft in the air as well as high-speed trains and cars on the ground [63]–[67], relying on adaptive

coherent/non-coherent techniques is especially beneficial for increasing the reliability of the mission-critical control links. Finally, our holistic three-fold adaptivity of coherent/non-coherent, single-/multiple-TA and diversity/multiplexing tradeoffs offers intelligent performance improvements in the context of index modulation [68]–[79], full-duplex [9], [80]–[87] and sophisticated channel coded systems [23], [91]–[93], where machine learning algorithms [98]–[103] may be invoked for autonomous decisions concerning the most appropriate mode of operation in NGNs.
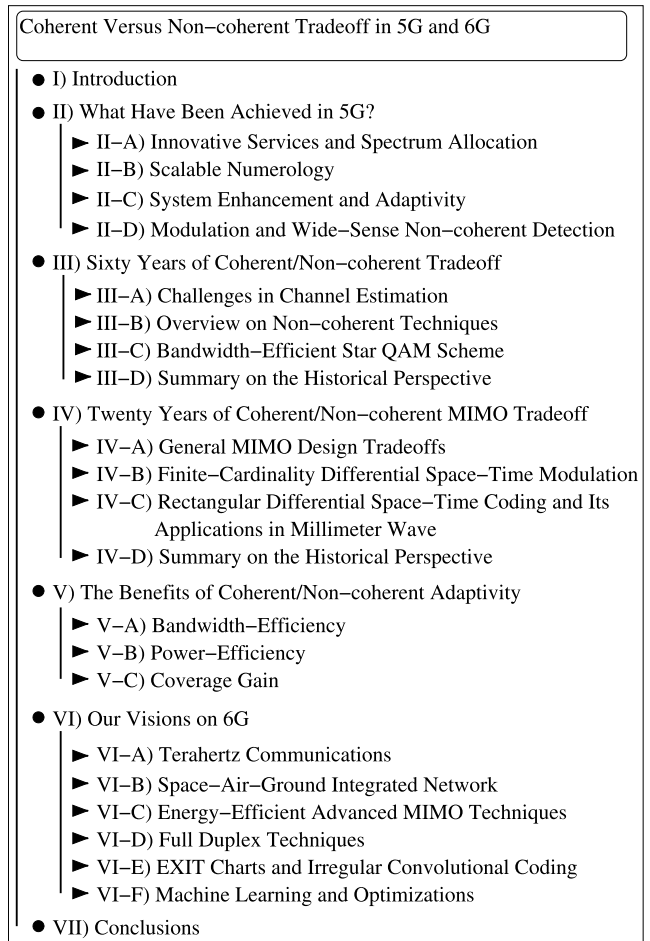
FIGURE 2. Structure of this paper.

The structure of this paper is portrayed by Fig. 2. We firstly review what has been achieved in 5G in Sec. II. Following this, the sixty-year history of coherent/non-coherent tradeoffs in SISO scenarios and the twenty-year history of coherent/non-coherent tradeoffs in MIMO systems are reviewed in Sec. III and Sec. IV, respectively. Our performance results on the bandwidth-efficiency, power-efficiency and link coverage are summarized in Sec. V. Furthermore, our NGN visions are elaborated on in Sec. VI. Finally, our conclusions are offered in Sec. VII.

II. WHAT HAS BEEN ACHIEVED IN 5G?
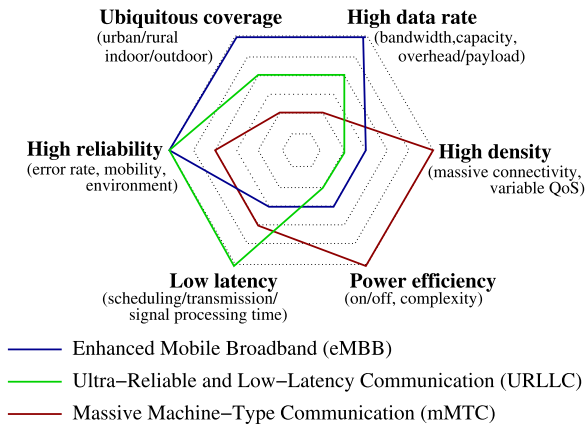### A. INNOVATIVE SERVICES AND SPECTRUM ALLOCATION

FIGURE 3. "Spider graph" of 5G New Radio (NR) services summarized according to [16].

The 5G NR is generally characterized into three operating modes conceived for specific types of services associated with different priorities, as portrayed by Fig. 3. Firstly, the enhanced Mobile BroadBand (eMBB) mode aims for substantially improving the users experience, which supports up to 1 Gbps indoor downlink data rate, while its ubiquitous coverage is assisted by its dual connection to both LTE and NR. Secondly, the Ultra-Reliable Low-Latency Communication (URLLC) mode supports critical communications that require a BLock Error Rate (BLER) below 10^{-5} and a maximum of 1 ms end-to-end latency. Thirdly, the massive Machine Type Communication (mMTC) mode extends the licensed cellular coverage to Internet-of-Things (IoT) devices, where power-efficient machine-to-machine communication helps to achieve ultra-long battery life of over 10 years for various IoT devices. The detailed NR service requirements can be found in [104].

Therefore, the distinction between 5G and the past generations is that 5G aims for supporting *diverse* communication requirements by a *unified* platform. This is achieved by its unprecedented flexibility and by its agile adaptivity of communication resources and techniques. Specifically, the spectrum allocations of 4G LTE and 5G NR are portrayed in Fig. 4 according to [105]–[107]. In this spirit, each country may authorize subsets of these recommandations for different operators. More explicitly, due to the spectrum scarcity and congestion, the 4G LTE leverages its Frquency Division Duplex (FDD) and Time Division Duplex (TDD) operations in the S-band, as shown in Fig. 4(a). By contrast, the 5G NR has two Frequency Ranges (FRs), which are the sub-6 GHz FR1 (0.45~6 GHz) of Fig. 4(b) and the millimeter Wave (mmWave) FR2 (24.25~52.6 GHz) of Fig. 4(c). We note that the boundaries between centimeter Wave (cmWave), mmWave and terahertz are not always consistent among regulation bodies such as the ITU and the IEEE. Strictly speaking, the mmWave band ranges from

30 GHz to 300 GHz, where the corresponding wavelength range is between 1 mm and 10 mm. The term of mmWave commonly refers to FR2 (24.25~52.6 GHz) in the 5G NR standards [105]–[107], which is adopted in this section. Our detailed discussions on the characteristic differences between cmWave, mmWave and terahertz will be elaborated on in our discussions related to future terahertz applications in Sec. VI-A. In a nutshell, the associated degree of freedom for the 5G NR to innovatively exploit the spectral resources is further elaborated on from the following three perspectives.

First of all, the highly-loaded UHF and L-band below 2 GHz are inherited by NR FR1 for the sake of wide, deep and ubiquitous coverage, where the FDD operations heavily rely on the disjoint narrow bands, as seen in Fig. 4(b). Secondly, in comparison to the LTE allocation of Fig. 4(a), more paired FDD spectrum slices and contiguous TDD bands are refarmed and licensed in the higher S-band and C-band between 2 GHz and 4 GHz in NR FR1 of Fig. 4(b). The antenna aperture decreases with the square of the carrier frequency, which is beneficial for reducing the physical size of User Equipments (UEs) and IoT devices. However, according to the Friis transmission equation, the path-loss also grows with the square of the carrier frequency, which has to be mitigated by high-gain beamforming. Meanwhile, in comparison to Fig. 4(a), there is a clear surge in TDD operations in this category, as seen in Fig. 4(b), where the Channel State Information (CSI) required at the transmitter for beamforming is typically assumed to be identical to that at the receiver, thanks to the channel's reciprocity.

Lastly but most importantly, the massive bandwidth in the previously unlicensed mmWave K-band and Ka-band has also been included in the NR FR2 of Fig. 4(c). There are three mmWave features that have previously been perceived as unfavorable but have been overcome in 5G:

(1) Firstly, the rain and atmospheric absorption siginificantly attenuated mmWave signals. This effect was shown to be much less severe at 28 GHz and 38 GHz [142], which has been taken into account in the spectrum selections seen in Fig. 4(c).

(2) Secondly, since the mmWave signals suffer from severe building penetration loss, the indoor and outdoor networks are separately deployed. Therefore, as long as the cell size is chosen to be sufficiently small, the interfering signals emanating from the neighbouring cells will be considerably attenuated, which becomes a beneficial effect in 5G mmWave scenarios.

(3) Thirdly, since the increased carrier frequency allows us to proportionally shrink the antenna aperture, a large array of mmWave antennas may be formed at a similar physical size as the antennas used in FR1 [143]. As a result, the aggregated antenna apertures may form narrow beams for the sake of mitigating the path loss of mmWave signals, which is deemed to be a key enabler of the 5G eMBB service.

We note that although high-gain beamforming-aided Base Stations (BS) and Mobile Stations (MS) play an essential
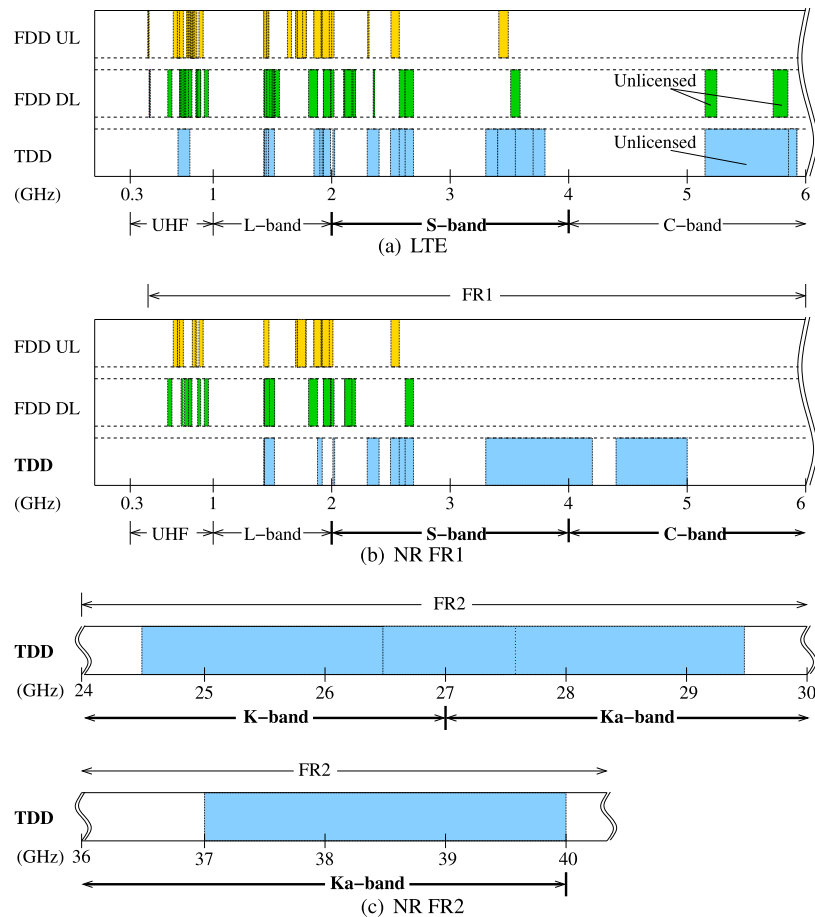
FIGURE 4. The spectrum allocations of (a) 4G Long-Term Evolution (LTE); (b) 5G New Radio (NR) Frequency Range 1 (FR1); (c) 5G New Radio (NR) Frequency Range 2 (FR2).

role in mmWave [142], the detrimental effects of beamforming pattern misalignment may substantially deteriorate the perfect-alignment-based system's performance. The classic solution in the microwave bands relies on circular polarization [144], which rotates the signal by simultaneously exciting a pair of orthogonal electric field vectors of equal amplitude, so that a signal component is always received. However, as the power is split between the pair, the employment of circular polarization requires an increased transmit power, which becomes a challengue to deliver in the mmWave band owing to the fact that the PA's dynamic range generally reduces upon increasing the frequency [58], [145]–[147]. Against this background, the state-of-the-art tri-orthogonal polarization technique that was originally proposed for achieving a diversity gain in rich-scattering microwave channels [148], [149] may be harnessed for the three-dimensional beamforming in the LoS-dominant mmWave channels [147], [150], [151]. More explicitly, it was demonstrated in [147], [151] that by using three polarized ports, at least two polarizations are always adequately aligned in any link direction. Nonetheless, the polarization diversity techniques suffer from the cross-polarization discrimination, which refers to the mean signal difference between co-polarized and cross-polarized branches [152]. This is reported to be more significant in line-of-sight [152]. Moreover, due to the insufficient isolation, the deleterious coupling of energy between polarized ports also leads to substantial capacity loss [149], [153]. For further discussions on these issues, interested readers might like to refer to [147], [151] for more details in this area.

The 5G pioneer bands in European countries [154], [155], for example, are identified to be in the UHF 700 MHz band that supports reliable long-range 30 MHz uplink/downlink connectivity. Furthermore, the S-band in the vicinity of 3.6 GHz supports up to 3 Gbps urban coverage, while the K/Ka-band 26 at GHz supports up to 10 Gbps small-cell hotspot coverage. Considering the example of the compelling Virtual Reality (VR) services [156], the early stage of 960×960 single-lens resolution video streams and the entry-level 1920×1920 single-lens resolution video typically require a data rate of 16 Mbps as well as 64 Mbps and a bandwidth of 25 MHz and 100 MHz, respectively. These two modes are mostly likely to be accommodated in the S-band and C-band of Fig. 4(b). At time of writing, the first commercialized VR streaming in 5G is commissioned in Japan for the 2020 Olympic games in Tokyo, where the bands 3.6~4.2 GHz and 4.4~4.9 GHz in the S-band and C-band of Fig. 4(b) are specifically licensed for the VR services.

B. SCALABLE NUMEROLOGY

The classic Orthogonal Frequency Division Multiplexing (OFDM) [9], [10], which divides the frequency-selective wideband channel into overlapping but orthogonal multi-carrier narrowband subchannels, constitutes the principal waveform of both LTE and NR. Specifically in 5G, the scalable numerology of the OFDM waveform is the key enabler of supporting the diverse service requirements of eMMB, URLLC and mMTC in a unified NR platform.

From the system's perspective, the Medium Access Control (MAC) layer assembles the data to be transmitted into transport blocks, which are passed down to the PHYsical (PHY) layer once per Transmission Time Interval (TTI), which is 1 ms in 4G LTE. In the PHY layer, the longest radio time unit is that of a frame of 10 ms, which is further partitioned into ten subframes of 1 ms duration. Furthermore, the frequency-time resource unit is defined by the Resource Block (RB), which is assigned 12 subcarriers in the frequency domain and 1 slot in the time domain. When the normal Cyclic Prefix (CP) overhead is configured, one slot is comprised of 7 and 14 OFDM symbols in LTE and in NR, respectively. Moreover, the smallest resource unit is the Resource Element (RE), which is constituted by a single subcarrier for an OFDM symbol duration. Therefore, in order to improve the bandwidth efficiency and reduce latency, the Sub-Carrier Spacing (SCS) in LTE should be scaled, which is the methodology adopted by NR, as demonstrated by Fig. 5.

More explicitly, first of all, the OFDM SCS in 4G LTE is generally a fixed 15 kHz, except for the even narrower SCS of 7.5 kHz for the VoIP application. By contrast, the OFDM SCS in 5G NR becomes scalable as shown by Fig. 5(a), where the resultant RE bandwidth becomes $SCS = 2^\mu \times 15$ kHz ranging from 15 kHz to 240 kHz. The parameter μ may assume $\mu = \{0, 1, 2\}$ in FR1 and $\mu = \{2, 3, 4\}$ in FR2. Accordingly, the OFDM duration of $T_{OFDM} = 1/SCS$ also varies ranging from 66.6 us to 4.17 us, as demonstrated by Fig. 5(c). Furthermore, the RB bandwidth and duration are evaluated by $BW_{RB} = 12 \times SCS$ and $T_{slot} = 2^\mu \times 1000$ us in Figs. 5(c) and 5(d), respectively, where the duration difference of $(T_{slot} - 14 \times T_{OFDM})$ is contributed to the CP overhead.

Therefore, a higher SCS results in both a higher RB bandwidth and a smaller slot duration. Consequently, a higher number of slots are conveyed in a fixed frame duration of 10 ms, which results in a higher rate of busty transmission. In the light of this development, the concept of BandWidth Part (BWP) has been adopted in 5G NR [16]. A BWP has a configurable transmission bandwidth of $BW = N_{RB}BW_{RB}$ that is defined by the number of RBs N_{RB} and by the configurable parameter μ. Carrier Aggregation (CA) may be invoked in order to aggregate a sufficiently high number of N_{RB} across non-contiguous carrier frequencies. A UE may be allowed to use up to four BWPs for multiplexing different services that have different data rate and latency requirements.

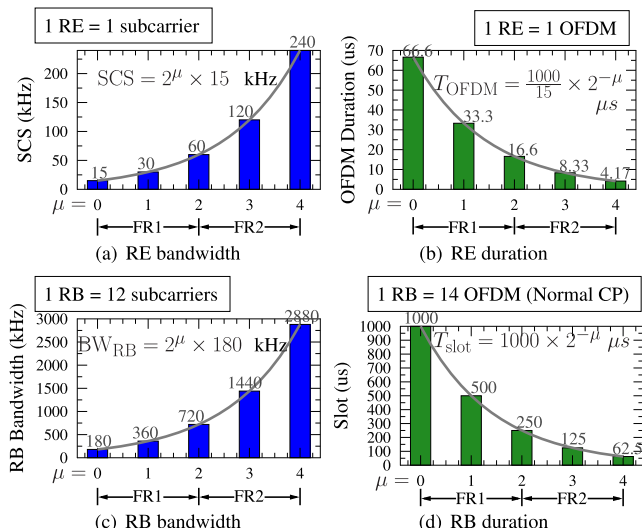In general, the CP duration must be higher than delay spread $T_{CP} \geq T_d$ in order to prevent Inter-Symbol

FIGURE 5. The scalable numerology on OFDM waveform and its effects on the resource allocations of (a) frequency-domain Sub-Carrier Spacing (SCS)/Resource Element (RE) bandwidth; (b) time-domain RE duration/OFDM duration; (c) frequency-domain Resource Block (RB) bandwidth; (d) time-domain RB duration/slot.
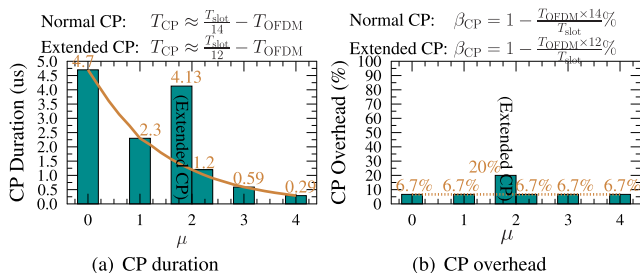
FIGURE 6. The scalable numerology on Cycle Prefix (CP) with respect to (a) CP duration; (b) CP overhead.

Interference (ISI). However, it is demonstrated in Fig. 6(b) that NR opts for a constant CP extension of $\beta_{CP} = 6.7\%$ for normal CP configuration, hence T_{CP} is reduced upon increasing μ in Fig. 6(a). The extended CP is only defined for $\mu = 2$, where 12 OFDM symbols are assigned to a slot. This once again confirms that when the SCS becomes larger, the transmission distance is intended to be smaller in 5G NR.

C. SYSTEM ENHANCEMENT AND ADAPTIVITY

3GPP defines a pair of 5G deloyment options, which are the Non-Stand Alone (NSA) and Stand Alone (SA) options [16]. The former supports both LTE and NR services, where the UE may rely on dual connectivity to both the LTE Base Station (BS) of eNB as well as the NR BS of gNB, while the SA architecture is completely operated by gNB. Therefore, both backward compatibility as well as distinctive upgrades are of essential importance to the success of 5G NR.

The NR PHY layer CHannels (CHs) [108] are summarized in Fig. 7(a), which exhibits an improved symmetry between the UpLink (UL) and DownLink (DL) compared to LTE. More explicitly, first of all, the Physical Broadcast
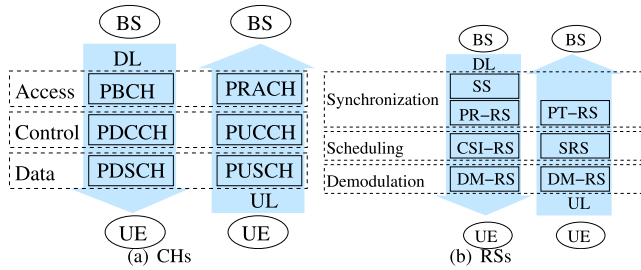
FIGURE 7. Summary on PHYsical (PHY) layer CHannels (CHs) and Reference Signals (RSs) in 5G New Radio (NR).

Channel (PBCH) and Physical Random Access Channel (PRACH) are responsible for initial access and synchronization, The typical functions of these channels support the synchronization and access requests of the UEs, for example. Secondly, the control links, such as the Physical Downlink Control Channel (PDCCH) and Physical Uplink Control Channel (PUCCH) assist in both data transmission as well as resource grant scheduling. Thirdly, Physical Downlink Shared Channel (PDSCH) and Physical Uplink Shared Channel(PUSCH) carry the payload data of transport blocks assembled in the MAC layer, which are protected by Cyclic Redundancy Check (CRC). Based on this, Hybrid Automatic Repeat Request (HARQ) is facilitated by the control entities in the MAC layer, which suports retransmissions upon erroneous signal reception. It is worth noting that three LTE CHs are no longer specified in NR, which are the Physical Multicast Channel (PMCH) for multi-cell synchronized broadcasting of the same signal, Physical Control Format Indicator Channel (PCFICH) conceived for assisting the UE in listening on the specific areas of the PDCCH, as well as the Physical HARQ Indicator Channel (PHICH) originally designed for conveying HARQ ACK/NACK flags concerning the PUSCH. We note that in NR, HARQ ACK/NACK acknowledgements are carried by the control CHs of PDCCH and PUCCH.

The 5G Reference Signals (RSs) [108] are summarized in Fig. 7(b). Firstly, the Synchronization Signal (SS) includes both Primary SS (PSS) and Secondary SS (SSS). More explicitly, PSS is responsible for the synchronization of cell identity, initial timing, carrier frequency and sampling clock, while SSS is in charge of further synchronization functions related to the OFDM frame format and beamforming. Moreover, since the phase noise typically increases with the oscillator carrier frequency, a new type of Phase-Tracking RS (PT-RS) has been introduced in NR. Secondly, Channel State Information (CSI) RS for DL and Sounding RS (SRS) for UL seen in Fig. 7(b) are not directly related to the coherent data detection. Instead, they generally operate based on the Channel Quality Information (CQI) simply estimated from the received signal strength, which assits in the grant-based scheduling of the space, time and frequency resources [109]. Thirdly, the DeModulation RS (DM-RS) acquires the detailed CSI including the amplitudes and phases of the Channel Impulse Responses (CIRs), so that coherent

detection may be performed at the receiver end. We note that LTE defines cell-specific RS for all UEs for coherently demodulating the PDCCH/PDSCH, while additional UE-specific RSs may be transmitted for demodulating the beamformed PDCCH/PDSCH. By contrast, DM-RSs in NR are all UE-specific.

In order to better appreciate the functionalities of the PHY CHs and RSs, three examples are offered in Fig. 8, which are related to the important concepts of contention-based random access, grant-based resource scheduling and beamforming, respectively. First of all, after reading the always-on PBCH and SS, the UE's random access of Fig. 8(a) commences with Message (Msg) 1 of the random access preamble carrying the UE's temporary signature. Following this, the BS replies by sending the random access response on the PDCCH/PHSCH as Msg 2, which informs the UE of the initial resource grant. Then the UE is instructed to begin its scheduled PUCCH/PUSCH transmissions as Msg 3. Finally, the BS transmits the UE contention resolution identity on PDSCH as Msg 4, which is protected by HARQ. Upon recovering this, the UE shall respond with the HARQ ACK flag transmitted on the PUCCH, which confirms the successful random access process [110].

The grant-based scheduling is exemplified in Fig. 8(b), where the transmission of SRS is for the sake of estimating the channel quality across the available bandwidth. Consequently, the BS is capable of scheduling the RBs that experience good channel conditions for data transmission. The CQI obtained from the SRS estimation also helps the BS to decide on the choice of beamforming, Modulation and Coding Scheme (MCS) as well as power and timing control. We note that in general the scheduling of frequency-, time- and spatial-resources can be near-instantaneously dynamic, semi-static or static, where the dynamic options achieve the maximum scheduling gain, while the static scheduling minimizes the control overhead.

For the sake of appropriately configuring the beamforming, the UE may monitor the link quality based on the SS and CSI-RS, where SS is always broadcast periodically. As a result, the UE performs candidate-beam detection, as seen in Fig. 8(c) by comparing the acquired signal strength to thresholds, which triggers beam reporting and enables DL beamforming [109]. We note that owing to the excessive directional antenna size in the S-band, LTE beamforming is only enabled on PDSCH [157]. By contrast, thanks to the substantially reduced mmWave antenna aperture of 5G NR, beamforming is enabled on both UE and gNB in NR, where SRS is utilized for UL beam management.

Finally, the key upgrades from LTE to NR are summarized in Table 3. First of all, given a fixed SCS of 15 kHz, LTE offers a maximum bandwidth of 20 MHz constituted by both $N_{RB} = 100$ RBs as well as additional guard-bands for alleviating the adjacent-band interference. We note that in LTE-Advanced, the maximum aggregated bandwidth is given by 100 MHz, comprising five component carriers [105]. By contrast, thanks to the scalable SCS, NR is
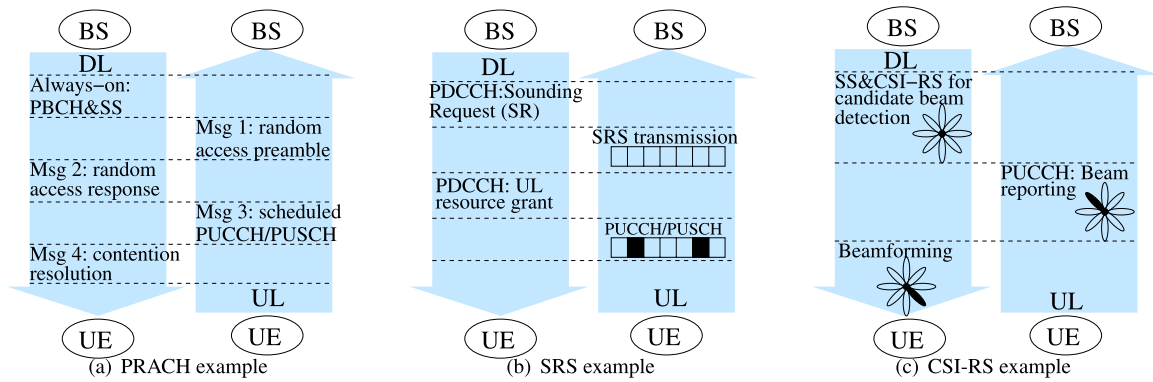
FIGURE 8. Examples of Physical Random Access Channel (PRACH) for contention-based random access, Sounding Reference Signal (SRS) for grant-based resource scheduling and Channel State Information Reference Signal (CSI-RS) for beamforming.

TABLE 3. Architecture comparison between 4G Long-Term Evolution (LTE) and 5G New Radio (NR).

	LTE	NR
SCS	15 kHz.	$2^{\mu} \times 15$ kHz. FR1: 15 kHz ($\mu=0$); 30 kHz ($\mu=1$); 60 kHz ($\mu=2$). FR2: 60 kHz ($\mu=2$); 120 kHz ($\mu=3$); 240 kHz ($\mu=4$).
Time Resource	1 frame=10 ms; 1 subframe=1 ms; 1 slot=7 OFDM symbols=0.5 ms (normal CP).	1 frame=10 ms; 1 subframe=1 ms; 1 slot=14 OFDM symbols=$2^{-\mu} \times 1$ ms (normal CP).
Frequency Resource	1 RB=12 subcarriers over 1 slot of 180 kHz bandwidth.	1 RB=12 subcarriers over 1 slot of $2^{\mu} \times 180$ kHz bandwidth.
Maximum Channel Bandwidth	20 MHz(100 RBs)	FR1: 50 MHz (270 RBs, $\mu=0$); 100 MHz (273 RBs, $\mu=1$); 100 MHz(135 RBs, $\mu=2$). FR2: 200 MHz(264 RBs, $\mu=2$); 400 MHz(264 RBs, $\mu=3$); 400 MHz($\mu=4$).
Waveform	UL: SC-FDMA; DL: CP-OFDM.	UL: DFT-S-OFDM/CP-OFDM; DL: CP-OFDM.
Coding	Data: Turbo code; Control: TBCC.	Data: LDPC; Control: Polar code.
Duplex	FDD, static TDD.	FDD, static/semi-static/dynamic TDD.
PUCCH	Located at bandwidth edges.	Flexible configurable.
PDCCH	Full carrier bandwidth.	Flexible configurable.
HARQ	UL: Synchronous and Non-Adaptive/Adaptive HARQ DL: Asynchronous and Adaptive HARQ	UL/DL: Asynchronous and Adaptive HARQ

capable of supporting at least 100 MHz per link for $\mu > 0$, while backward compatibility with LTE is assured by providing the choice of $\mu = 0$. Secondly, the Single-Carrier Frequency Division Multiple Access (SC-FDMA) signal of LTE is generated by Discrete Fourier Transform-Spread-OFDM (DFT-S-OFDM) in NR, as seen in Table 3. For DFT-S-OFDM, the transform precoding of DFT and zero padding is performed before the OFDM's IFFT stage. As a result, the unified OFDM interface is also used for the single-carrier waveform, which exhibits favourable oversampling and pulse-shaping characteristics, while its beneficially low Peak-to-Average Power Ratio (PAPR) is retained. Thirdly, in an effort to improve the performance and yet to reduce latency, the Turbo code and Tail-Biting Convolutional Code (TBCC) of LTE were replaced by Low-Density Parity-Check (LDPC) code and Polar code in NR for the data and control links, respectively. Fourthly, the scheduling in NR becomes more flexible, which include slot assignments to UL/DL duplex and bandwidth assignments to control/data links. Lastly, we note that HARQ in NR is always asynchronous and adaptive, where the retransmission does not have to occur at predefined instants, while the MCS can also be changed for each retransmission.

In summary, the unprecedented flexibility of 5G NR design offers substantial upgrades that are capable of adaptively accommodating diverse Quality-of-Service (QoS) requirements. Meanwhile, a substantial surge in control signalling overhead becomes inevitable. Against this background, we will further investigate the PHY modulation schemes used in NR in the next section in order to find techniques of reducing the overhead by the coherent versus non-coherent tradeoff.

D. MODULATION AND WIDE-SENSE NON-COHERENT DETECTION

The modulation schemes of 5G NR [108] are summarized in Table 4. First of all, the constant-envelope Zadoff-Chu (ZC) sequence is widely used as single-carrier waveform in the UL. More explicitly, the ZC sequence is constituted by a sequence of optimized phase rotations that exhibits zero auto-correlation and minimum cross-correlation. The inherent 0 dB PAPR in the discrete time domain also results in sufficiently low PAPR for the analog signal after pulse-shaping, which is beneficial for the Power Amplifiers (PAs) of the UEs, since no power-hungry linear class-A amplification is needed for low-PAPR signals. Moreover, the DFT or IDFT of

TABLE 4. List of modulation schemes for PHYsical (PHY) layer CHannels (CHs), Reference Signals (RSs), Synchronization Signals (SSs) for 5G New Radio (NR) UpLink (UL) and DownLink (DL).

UL	CH	PRACH	ZC
		PUCCH	Format 0: ZC; Format 1: ZC spread BPSK/QPSK; Format 2: QPSK OFDM; Format 3&4: $\pi/2$-BPSK/QPSK DFT-S-OFDM.
		PUSCH	$\pi/2$-BPSK/BPSK/QPSK/16QAM/64QAM/256QAM DFT-S-OFDM/OFDM.
	RS	PT-RS	PT-RS for PUSCH: QPSK modulated Gold.
		SRS	ZC
		DM-RS	DM-RS for PUCCH Format 1: ZC; Format 2: QPSK modulated Gold; Format 3&4: ZC. DM-RS for PUSCH: ZC for DFT-S-OFDM, QPSK modulated Gold for OFDM.
DL	CH	PBCH	QPSK OFDM
		PDCCH	QPSK OFDM
		PDSCH	QPSK/16QAM/64QAM/256QAM OFDM
	RS	PT-RS	PT-RS for PDSCH: QPSK modulated Gold.
		CSI-RS	QPSK modulated Gold.
		DM-RS	DM-RS for PBCH/PDCCH/PDSCH: QPSK modulated Gold.
	SS	PSS	BPSK modulated m-squence.
		SSS	BPSK modulated cyclic extended m-squence.

a ZC sequence is given by a weighted cyclically shifted version of itself, which preserves the single-carrier waveform. Secondly, the pseudo-random Gold codes associated with low cross-correlations are also widely invoked by the OFDM waveforms of the RSs of DL and UL, as shown in Table 4.

Regarding the MIMO aspects, although arrays of massive antenna elements are employed in NR, only up to four and eight MIMO-layers are supported for Single-User MIMO (SU-MIMO) in the UL and DL transmissions, respectively. Furthermore, when the single-carrier DFT-S-OFDM waveform is used in the UL, only a single transmission layer is activated. Moreover, multi-layer transmission is not supported for Multi-User MIMO (MU-MIMO) in NR. More explicitly, as proven in [158], the channel responses may become quasi-orthogonal as the number of BS antennas becomes much higher than the number of UEs. As a result, the UEs may share the same time-frequency resource, which mitigates the grant-based scheduling overhead. However, first of all, as the number of UEs becomes high in a cell, the ZC and Gold sequences seen in Table 4 may have to be frequently reused within each other's vicinity, which imposes detrimental pilot contamination. Secondly, the degree of freedom provided by large antenna arrays has to be exploited for mmWave beamforming, which limits the number of antennas available for supporting large-scale MU-MIMO systems. Thirdly, the power consumption imposed by supporting many MIMO layers is often prohibitive, especially in the predominately small cells of 5G NR. By contrast, the massive MIMO setup conceived for hybrid analog-digital mmWave beamforming always invokes substantially reduced numbers of transmit/receive RF chains compared to the number of transmit/receive antennas [159].

Finally, it is demonstrated by Table 4 that the PHY layer inevitably imposes substantial overheads, including the control CHs and RSs, which compromise both the bandwidth-efficiency and the latency in exchange for improved reliability. Furthermore, the RSs such as the DM-RSs assigned to the PUCCH and PDCCH become effectively the "overhead of overhead", where no user data is conveyed. Moreover, both the channel coding and the CRC of HARQ further increases the redundancies, where the lowest code rate in NR is as low as $30/1024 \approx 0.03$ [111]. In order to alleviate overheads in LTE and NR, substantial efforts have been invested in conceiving a wide-range of non-coherent detection techniques, including blind detection. We summarize these scenarios based on the 3GPP Technical Specification (TS) series 36/38 and [16], [157] as follows:

- *Non-coherent Detection for PSS:* In order to perform cell-search, the UE in NR assumes that the PSS, SSS and PBCH are transmitted consecutively, which forms a SS/PBCH block. However, only the PBCH is assigned with DM-RS. As a result, the UE may perform PSS non-coherent timing detection without *a priori* knowledge of the channel.

- *Non-coherent Detection for SSS:* After PSS detection, the UE becomes capable of deducing CSI for the sake of coherently detecting the forth-coming SSS. However, the interfering PSS received from the neighbouring BS may degrade the initial channel estimation. Therefore, the UE may still opt for invoking non-coherent detection for SSS.

- *Non-coherent Detection for PUCCH:* In LTE, once required, the second RS on the PUCCH format 2 may be used for carrying the HARQ ACK/NACK flag. Moreover, in NR, the additional DM-RSs are provided for the PUCCH format 3 and 4, which may still be reclaimed for CSI report. As a result, the resources available for DM-RS may become insufficient, especially in high-Doppler scenarios, where non-coherent detection may become more realistic.

- *Blind Detection for PDCCH:* In LTE, the PDCCH may be randomly distributed across the entire bandwidth for the sake of improving the frequency diversity. If the bitmap controlling the instantaneous frequency hopping is not available at the UE, the UE may perform blind detection and then act on the messages having correct CRC. Furthermore, the PDCCH in NR is transmitted in

COntrol REsource SETs (CORESETs) using either contiguous or distributed REs under a given SCS numerology, where the blind detection mode is also configured at the UE.

- *Blind Detection for PBCH:* Since the PBCH does not actively use all transmit antenna ports, the UE must blindly deduce the number of antennas activated for the PBCH. Furthermore, when the UE first attempts to access a cell by reading the system information on the PBCH, blind detection often resorts to observing several PBCH frames.

- *Blind Detection for ZC:* An erroneous cyclic shift is imposed, when high Doppler shift is encountered, which contaminates the ZC's zero auto-correlations. As a remedy, multiple windows of the ZC signals may be non-coherently combined for mitigating the timing uncertainty.

In summary, although the majority of the current communication systems tend to favour coherent detection, non-coherent detection techniques are capable of reducing the overheads, particularly for the DM-RS and PT-RS of Table 4. Moreover the non-coherent schemes are particularly beneficial for broadcasting, paging and wake-up signals, where the pilot overhead and channel estimation latency of coherent detection may be deemed excessive. Furthermore, we will proceed by demonstrating in the following sections that non-coherent detection techniques are capable of offering a more reliable coverage in high-mobility scenarios associated with high Doppler frequencies.

On a related note, the tradeoff between coherent detection and direct detection plays a similar role in optical fiber communications [160]–[162]. The conventional direct detection recovers the intensity of the optical signal by a low-complexity photodiode, without giving any cognizance to the phase of the signal. The linear transmission impairments such as chromatic dispersion constitute the key limitations of the direct detection. However, with the advent of breakthroughs in high-speed ADCs and digital signal processors, coherent optical detection has become capable of extracting the phase information of the optical signal. As a result, the linear impairments can be readily compensated by fast phase-rotation-based equalizers, which facilitates long-haul optical transmission. Furthermore, parallel to sophisticated radio frequency techniques, advanced complex-valued modulation formats such as high-order PSK and QAM may also be used for coherent optical detection, which has substantially improved the spectral efficiency attained. Nonetheless, as the associated SNR increases, the non-linear transmission impairments such as the Kerr effect arise, which calls for sophisticated digital nonlinearity mitigation techniques [162]–[164]. Hence, direct detection is often employed for short distances, while coherent detection may now be used for long-haul optical systems. Moreover, in order to mitigate the frequency mismatch between the transmitter's laser and the receiver's local oscillator, the aforementioned radio frequency technique of differential encoding and non-coherent detection is also often invoked in the coherent optical regime [165]–[167]. To avoid digression from our main focus on wireless cellular networks, interested readers are referred to [160]–[162], [168] for more details on coherent optical fiber communications.

III. SIXTY YEARS OF COHERET/ NON-COHERENT TRADEOFF
A. CHALLENGES IN CHANNEL ESTIMATION
The signal received at the output of a wireless communications channel typically contains a faded and noise-contaminated replica of the transmitted signal. In the classic AWGN channels, often a frequency offset of $\exp(j\theta)$ also corrupts the transmitted symbol s_n, so that the received symbol is given by $y_n = \exp(j\theta)s_n + v_n$, where v_n is the AWGN element. This frequency offset, which is the difference between the frequency of the received signal and the designed frequency of the receiver, is typically induced by numerous sources, such as the Doppler shift and the frequency difference of the heterodyning oscillators. Furthermore, in fading channels the frequency offset $\exp(j\theta)$ is replaced by a fading factor h_n, which may be a random variable, because the signals received from different paths may be superimposed either constructively or destructively. Against this background, a receiver is referred to as being "coherent", when the CSI is estimated prior to detection at the receiver. This requirement of CSI knowledge at the receiver has inspired the development of channel estimation techniques [10], [22], [23].

In order to better understand the challenges in channel estimation, let us briefly revisit the channel modelling fundamentals. The radio propagation mechanism is generally divided into the distance-related path-loss, shadowing fading imposed by blocking-induced attenuation by large objects and multipath fading. Both path-loss and shadow fading must be accounted for by the link budget during the broadcasting and random access phases introduced in Sec. II. Our example concerning the link budget is presented in Sec. V, where the minimum signal level required by the receiver has to be satisfied for establishing an adequate link based on the distance and antenna radiation parameters. Then the major factors determining the capacity and the integrity of the link established are the specific distribution of the multipath fading, the Doppler frequency of fading, which is proportional to the carrier frequency as well as the vehicular speed. The final pivotal propagation parameter is the length and shape of the Channel Impulse Response (CIR). On one hand, when all the reflected and diffracted mutlipath components arrive almost simultaneously within a symbol period, non-dispersive fading is encountered, which is represented by a dirac-delta CIR. Its frequency-domain channel transfer function then becomes near-constant across the entire signal bandwidth. By contrast, when the time-domain delay-spread of the multipath components exceeds the symbol period, detrimental inter-symbol interference arises. This CIR results in a frequency-dependent channel transfer function.

To elaborate a little further, the *coherence bandwidth* defines the range of frequency over which the frequency-domain channel transfer function is near-constant and this bandwidth is inversely propotional to the mean delay-spread. Therefore, when the signal bandwidth is lower than the coherence bandwidth, non-dispersive fading channels are encountered. The frequency-selectivity of the wideband 4G/5G is counteracted by using either multi-carrier OFDM or single-carrier DFT-S-OFDM associated with frequency-domain equalization as briefly touched upon in Sec. II.

To elaborate a little further on the Doppler effect, this is widely known from our daily lives. When for example an ambulance is approaching us and then receding from us, the pitch/frequency of its sirene gradually changes. This is because the propagation time of the sirene is shortened, when it is approaching, while it is extended, when it is receding from us. The same phenomenon influences the propgation of the inaudible RF signals. For example, for a high-speed train velocity of 500 km/h and an aircraft speed of 1080 km/h, the maximum Doppler frequency may reach thousands of Hertz [21], [64], [67]. Again, a higher vehicle speed leads to a faster fading fluctuation. In order to quantify this, the concept of *coherence time* can be used, which is inversely propotional to the maximum Doppler frequency, and it characterizes the time inverval over which fading channel envelope may be deemed near-constant. When the coherence time is substantially longer than the symbol period, the fading may remain constant over a block of symbol transmissions. In this case, CSI estimation may be carried out by transmitting training symbols that are known to the receiver at the beginning of the signal frame, as seen in Fig. 9. By contrast, when the coherence time is reduced due to increasing the vehicular speed, the fading changes more rapidly over time, hence the pilot symbols used for CSI estimation have to be inserted more frequently in the data frame, as seen in Fig. 9. For detailed discussions on channel models, the motivated readers might like to refer to [22], [169]–[172].
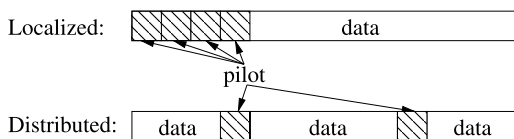

FIGURE 9. Arrangements of pilots for time-domain channel estimation of coherent detection.

We note that the multipath fading channels in LTE and NR typically experience correlation among consecutive transmissions in the time-domain, among contiguous subcarriers in frequency-domain and among adjacent antennas in spatial-domain. Thanks to this three-fold correlation, the predefined cell-specific and UE-specific RSs in LTE as well as DM-RS in NR are periodically transmitted, so that the receiver may deduce the CSI experienced by these pilots and then estimate the CSI for each data symbol by interpolation. Owing to the potentially excessive complexity of the optimal joint estimation, the three-dimensional channel estimation is often decoupled into their respective individual domains. The time-domain pilot-based estimation is exemplied by Fig. 9.

More explicitly, the arrangement of pilot symbols may be either localized or distributed, where the former is also referred to as the preamble-based training assisted signal transmission, while the latter is often termed as Pilot Symbol Assisted Modulation (PSAM). The preamble-based training assisted signal transmission of Fig. 9 is conceived for fading channels, which remain more-or-less time-invariant during a block of symbol periods. In this case, a sequence of training symbols that are known to both the transmitter and receiver may be transmitted before the sequence of data symbols, so that the receiver may first recover the CSI knowledge by observing the faded and noise-contaminated training symbols, and then a coherent demodulator may be invoked for detecting the following data symbols with the aid of the CSI estimate. The training sequence design was optimized to maximizing the capacity bound in [24]–[27]. In particular, in 2003, Hassibi and Hochwald [27] proposed a framework for appropriate time- and power-allocation between the training and data symbols based on the capacity lower bounds. They discovered that if the training symbols and the data symbols are assigned the same power, half of the channel's coherence time should be dedicated to training. However, if the power allocation is optimized by maximizing the capacity bounds, the number of neccessary training symbols may be set to the number of TAs. It was also observed in [27] that using training is optimal for obtaining accurate CSI estimation in the high SNR region in conjunction with a long coherence time, but channel estimation relying on training becomes suboptimal in the low-SNR region. It was demonstrated in [173], [174] that the Mean Squared Error (MSE) lower-bound of training is given by the Cramer-Rao Bound (CRB) of $CRB = N_0/N_{OW}$ in block fading channels, where N_{OW} represents the observation window length used for training symbols. This further confirms that channel estimation using training is challenging, when the noise power is high, i.e. the SNR is low.

When the channel's coherence time is reduced and becomes comparable to the symbol period, a so-called rapidly fading channel is encountered, where the fading factor is generally considered to be different for each consecutive transmitted symbol. In this scenario, a beneficial solution is to insert a pilot tone into the spectrum of the transmitted signal, so that an estimate of the fading factor may be extracted at the receiver. This technique was proposed by McGeehan and Bateman [28] in 1984 termed as the Transparent Tone-In-Band (TTIB) solution. The design challenges of TTIB include an increased PAPR, and the TTIB also has to carefully choose the position of the pilot tone in the signal spectrum. In order to avoid this problem, the PSAM technique was proposed by Moher and Lodge [29] in 1989 and then it was theoretically analysed by Cavers [30] in 1991. As portrayed in Fig. 9, the PSAM scheme periodically transmits pilot symbols inserted among data symbols, so that the receiver may interpolate both the correlated fading envolope and the phase by a Wiener filter that aims for achieving

the Minimum Mean Squared Error (MMSE) between the estimated CSI and the actual CSI. The PSAM scheme's MSE lower bound, which corresponds to the ideal case of the fading envelope and phase remaining near-constant over N_{OW} symbol observations, is given by $\sigma^2_{\text{MSE-LB}} = N_0/(N_0 + N_{\text{OW}})$, which is worse than that of the preamble-based training in the context of block fading. This channel estimation error expression implies that channel estimation is even more of a challenge in the low-SNR region, when rapid fading is encountered. Moreover, as the Doppler frequency is increased, the PSAM scheme's pilot spacing has to be reduced in order to sample the fading channels more frequently.[1] As a result, higher transmission power has to be dedicated to the pilot symbols, which is to the detriment of the data-carrying symbols.

As discussed in Sec. II-C, the demodulators are capable of operating at a relatively low SNR with the aid of powerful channel coding schemes that is close to the capacity limit, where accurate CSI knowledge is difficult to obtain. Having said that, researchers often rely on the idealized simplifying assumption of having perfect CSI knowledge even in rapidly fading channels at low SNRs. In practice, realistic channel estimation always imposes a performance penalty that is typically more severe in coded systems than in uncoded systems. This is because the soft-decision-aided coherent demodulators relying on imperfect CSI will produce erroneous soft-decisions in the form of Log Likelihood Ratios (LLRs), which deviate from the true probabilities. Hence, the resultant exaggerated LLR values may become more and more difficult to correct by the channel decoder, as the number of iterations increases, which is thoroughly investigated in [51].

B. OVERVIEW ON NON-COHERENT TECHNIQUES

By contrast, when the CSI is not explicitly estimated at the receiver, "non-coherent" detection is encountered. In order to eliminate channel estimation with the aid of non-coherent receivers, differentially encoded modulated schemes may be employed [22]. As an example, a simple Differential Phase Shift Keying (DPSK) transceiver is portrayed in Fig. 10. At the DPSK transmitter, the data-carrying MPSK symbol x_{n-1} is mapped onto the difference between the consecutive transmitted symbols as $s_n = x_{n-1}s_{n-1}$. In AWGN channels or in non-dispersive slow-fading channels, the received signal seen in Fig. 10 may be expressed as $y_n = s_nh_n + v_n$, where h_n and v_n refer to the CSI and AWGN, respectively. As a result, the DPSK receiver of Fig. 10 may be capable of eliminating the need for channel esimation by invoking the simple correlation operation of $z_{n-1} = y_ny^*_{n-1}$, which may be directly demapped to bits. Again, this low-complexity non-coherent detection, which detects a single symbol based on two observations was termed as CDD in Sec. I.

---

[1] We note that more frequently inserting pilots cannot improve the low-SNR-induced degradation of channel estimation described in the previous paragraph, because there is no correlation amongst the noise samples.
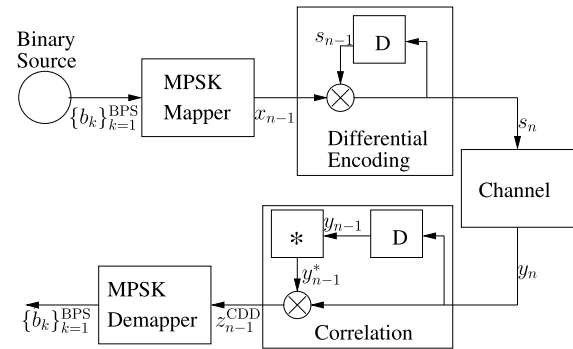

FIGURE 10. Schematic of a simple Differential Phase Shift Keying (DPSK) transceiver.

The pragmatic approach to improve the performance of the non-coherently detected DPSK scheme for the sake of approaching the optimum MLSE performance is to increase the number of observations. In 1979, it was proposed by Masamura *et al.* [175] that the DBPSK's pair of data-carrying bits detected from three consecutive observations at the non-coherent receiver may be treated as an information bit and a parity bit. As a result, detecting DBPSK as a channel code may improve its performance in AWGN channels. Furthermore, Samejima *et al.* [176] extended the work to M-level DPSK in [175] and to more than three observations. Following this, the groundbreaking MSDD concept was conceived for DPSK operating in AWGN channels by Wilson *et al.* [39] in 1989 and then theoretically analysed by Divsalar and Simon [40] in 1990. As portrayed by Fig. 11(a), the MSDD extends the CDD's observation window from two to $N_w \geq 2$ observations, where a total number of $(N_w - 1)$ data-carrying symbols are jointly detected. The MSDD may be viewed a special case of MLSE, where the channel memory is assumed to span over N_w transmission intervals. Moreover, for the sake of simplicity, the MSDD assumes having a message length of N_w, which is the same as the channel's memory, so that the employment of the Viterbi algorithm by the MLSE [33] may be avoided. Furthermore, the MSDD design conceived for DPSK was extended to Rayleigh fading channels by Ho and Fung [41] in 1992. As a further advance, the MSDD techniques developed for both DPSK and for Differential Quadrature Amplitude Modulation (DQAM) operating both in AWGN and Rayleigh fading channels were summarized by Divsalar and Simon [42] in 1994. It was demonstrated that the MSDD may be capable of reducing the 3 dB performance penalty both for AWGN channels and for slowly fluctuating fading channels [39]–[42]. Furthermore, the CDD's error floor experienced in rapidly fading channels may also be mitigated by the MSDD [41], [42]. However, the major disadvantage of the MSDD is that its detection complexity increases exponentially with N_w. Given an M-level DPSK scheme, the MSDD complexity order imposed by detecting a single symbol is given by $O\left[M^{(N_w-1)}/(N_w - 1)\right]$.

In order to further reduce the MSDD complexity, an efficient algorithm was conceived by Mackenthun [177]
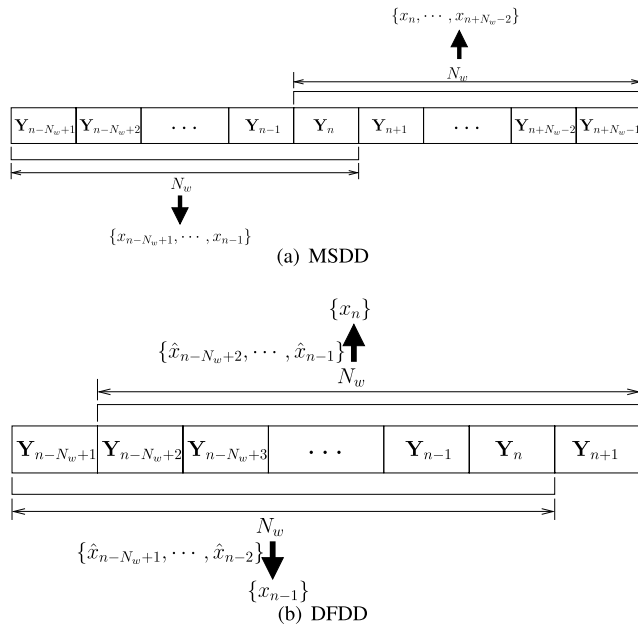
FIGURE 11. Schematics of MSDD and DFDD for non-coherent detection.

in 1994 for MSDD in AWGN channels, where a low detection complexity order of $O(\log N_w)$ was imposed by detecting a single symbol relying on tracing the phase changes over the N_w samples. This method may also be applied for reducing the MSDD complexity of DPSK operating in block fading channels. However, as discussed before, channel estimation for block fading channels is capable of maintaining a low estimation error that diminishes upon increasing both the SNR and the training window. Furthermore, it was also demonstrated by Chen et al. [178] that harnessing a Turbo code for assisting the channel estimation of DPSK in block fading channels may achieve a performance that is close to Shannon's capacity limit. By contrast, it is widely recognized that channel estimation remains a challenge for rapidly fading channels. Therefore, it is vitally important to implement MSDD in rapidly fading channels at an affordable complexity.

Similar to the MIMO's full-search-based ML detection, the exponentially increasing MSDD complexity is the result of jointly detecting all the $(N_w - 1)$ data-carrying symbols, as seen in Fig. 11(a). In order to separately - rather than jointly - consider the individual symbols, decision-feedback may be introduced, where a total of $(N_w - 2)$ data-carrying symbols are detected from the previous detection windows, while only a single unknown symbol has to be detected during the current detection window. This decision-feedback aided version of MSDD may be termed as Decision-Feedback Differential Detection (DFDD), which is portrayed in Fig. 11(b). The DFDD concept was originally proposed for DPSK operating in AWGN channels by Leib and Pasupathy [43] in 1988 and by Edbauer [44] in 1992, which were shown to be equivalent by Adachi and Sawahashi [45] in 1993. Leib [46] later confirmed in 1995 that the DFDD of [43] is equivalent

to the MSDD of [40] operating in decision-feedback mode. The DFDD designed for DPSK was further extended to Rayleigh fading channels by Schober et al. [47] in 1999, which was derived from the MSDD of [41], [42]. Furthermore, the DFDD designed for DPSK was developed to be able to both accept and to produce soft-bit decisions by Lampe et al. [48] in 2001, so that the DFDD may be invoked in turbo detection. Another form of DFDD may be derived from linear prediction, where a channel sample is predicted by a low-pass filter based on the previous observations and decisions, so that coherent detection may be performed for the current symbol. This idea was originally introduced by Svensson [179] in 1994, which was tailored for DQPSK operating in fading channels.

In 1999, Bin and Ho [180] further generalized the prediction-based DFDD to M-level DPSK, while Hoeher and Lodge [53] proposed the soft-decision-aided prediction-based DFDD for DPSK in fading channels. In 2000, Schober and Gerstacker [57] extended the prediction-based DFDD designed for DPSK to Ricean fading channels, and they proved that both MSDD-based DFDD [47], [48] and prediction-based DFDD employing the MMSE Wiener filter [53], [179], [180] are identical for DPSK operating both in AWGN and in Rayleigh fading channels. Furthermore, these two forms of DFDD also perform similarly in Ricean fading channels. The above DFDD solutions are capable of improving the CDD's performance by mitigating its error floor, when the fading fluctuates rapidly. Moreover, the DFDD complexity order is simply given by $O(M)$, because only a single symbol has to be detected at a time. However, the DFDD's imperfect decision feedback results in a performance loss compared to MSDD.

In order to retain the optimum MSDD performance, the concept of Multiple-Symbol Differential Sphere Detection (MSDSD) was proposed by Lampe et al. [49] in 2005, where the problem of optimizing the MSDD decision metric was transformed into a shortest-vector search problem [181], so that the SD may be invoked for MSDD. An example of the MSDSD conceived for DQPSK is portrayed in Fig. 12(a). More explicitly, assuming that the SD's initial radius is set to be sufficiently large, and the SD index v may start from $v = 1$, for each SD index v, only M MPSK candidates have to be examined for detecting a single symbol. When the best candidate is chosen, the SD index v may be increased and the search may be repeated until $v = N_w$ is reached, where a valid MSDSD output is found and the SD's radius may be updated accordingly. Then the SD index v may be reduced, so that the next-best candidate may be examined. If no more valid candidates can be found within the SD's radius, the SD index v may be further reduced until reaching $v = 1$, where the sphere decoding process is terminated.

As a result, the MSDSD complexity order imposed by detecting a single symbol is lower-bounded by $O(M)$. If the MSDSD is replaced by trellis decoding aided MSDD using the Viterbi algorithm as portrayed by Fig. 12(b), the same optimum MSDD decisions may be obtained. The trellis seen
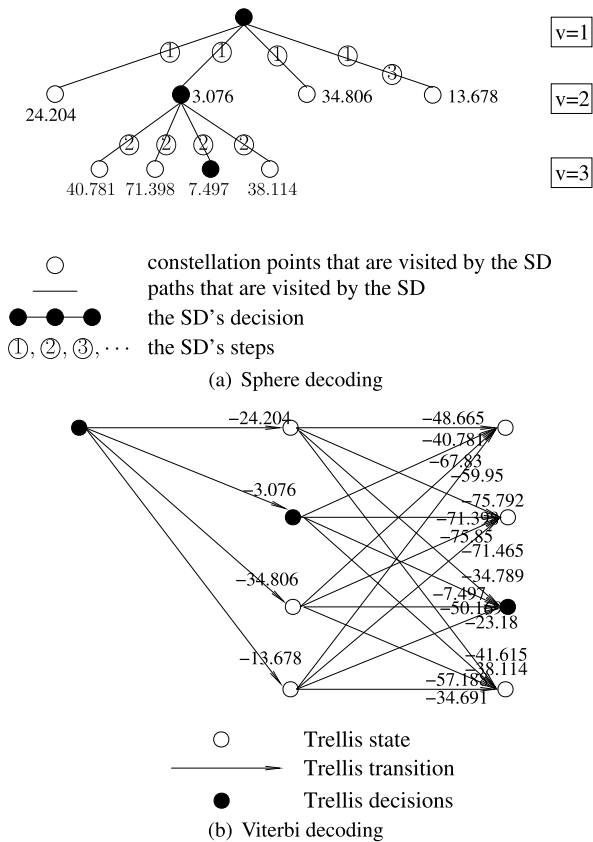
FIGURE 12. Example of Multiple-Symbol Differential Sphere Detection (MSDSD) conceived for DQPSK recorded at SNR=10 dB, where we have $N_w = 3$, and its corresponding trellis decoding aided MSDD using the Viterbi algorithm.

in Fig. 12(b) is drawn according to the differential encoder's memory [52], [53], as seen in Fig. 10 instead of the channel's memory [54]–[56], so that the number of trellis states may be kept to a minimum. In general, the trellis decoding process may span over much more than just two time symbols, as seen in Fig. 12(b). Considering that except for the first time symbol, there are always M trellis transition branches emerging from a particular current state, while there are always M transition braches merging into a particular next state, the average complexity order of the trellis-decoded MSDD imposed by detecting a single symbol is given by $O(M^2)$, which is higher than the MSDSD's lower bound. Inspired by the development of MSDSD, Pauli *et al.* [50] proposed the soft-decision-aided MSDSD concept for DPSK in 2006.

In summary, the links between the classic MIMO receivers and the non-coherent receivers introduced in this section are presented in Fig. 13, where the same performance-complexity tradeoff exists for both systems. In more details, both the MIMO's ML receiver and the non-coherent MSDD jointly detect multiple data-carrying symbols, so that the optimum performance may be achieved, albeit this imposes a high complexity. Both the MIMO's SD and the non-coherent MSDSD substantially reduce the original detection complexity by invoking a sphere decoder. Moreover, both the MIMO's

MMSE based linear receiver and the non-coherent DFDD opt for separating the multiple data streams by employing a Wiener filter, which results in a low detection complexity, but a degraded performance.

C. BANDWIDTH-EFFICIENT STAR QAM SCHEME

If a higher bandwidth-efficiency is pursued, the DQAM constellations may be considered. The early attempt to invoke differential encoding for QAM stems from Simon and Smith [182] in 1974, where they suggested that differential encoding may resolve the quadrant ambiguity of QAM transmission in the presence of a constant phase rotation in AWGN channels. This idea was implemented by Weber [183] in 1978, where part of the information bits were assigned to a differentially encoded phase. Furthermore, in 1982, Simon *et al.* [184] proposed a more general regime for the absolute-amplitude based DQAM scheme, where the transmitted phase was differentially encoded but the transmitted amplitude was the original data-carrying amplitude. This transmission regime [184], which was conceived for the maximum-minimum distance Square QAM constellation, was originally proposed to be detected non-coherently for recovering the data-carrying phase, but coherently recovering the data-carrying amplitude. However, this required that the channel amplitude had to be estimated at the receiver. In the sequel, it was discovered later by Lampe and Schober [185] in 2001 that the equivalent differential encoding of the absolute-amplitude DQAM may be represented by the revised expression of $s_n = \frac{1}{|s_{n-1}|}x_{n-1}s_{n-1}$, where the amplitude of the previous transmitted symbol $|s_{n-1}|$ is normalized, so that the amplitude of the next transmitted symbol $|s_n|$ becomes equal to the absolute amplitude of the data-carrying QAM symbol $|x_{n-1}|$. Lampe and Schober [185] adopted the absolute-amplitude DQAM regime for the Star QAM constellation, which may be termed as Absolute-amplitude Differential Phase Shift Keying (ADPSK). An example constellation diagram of the 16-ADPSK scheme's data-carrying symbols is portrayed in Fig. 14(a). For the ADPSK receiver, based on the revised differential encoding model of $s_n = \frac{1}{|s_{n-1}|}x_{n-1}s_{n-1}$, non-coherent detection may be invoked for both amplitude and phase detection. Considering CDD for ADPSK in block fading as an example, the received signal may be expressed as $y_n = s_n h_n + v_n = \frac{1}{|s_{n-1}|}x_{n-1}y_{n-1} + (v_n - \frac{1}{|s_{n-1}|}x_{n-1}v_{n-1})$, where $h_n = h_{n-1}$ and v_n refer to the block fading and AWGN factors, respectively. As a result, both the amplitude and phase of x_{n-1} may be recovered by the CDD with the aid of both the previous received sample y_{n-1} and the previous decision on $|s_{n-1}| = |x_{n-2}|$. It is worth noting that many QAM aided differential MIMO schemes [116], [118], [119], [124], [186] make use of this revised differential encoding model for dynamically constraining the transmitted symbol's amplitude.

As an alternative to ADPSK, the classic Differential Amplitude Phase Shift Keying (DAPSK) was proposed by Webb, Hanzo and Steele [187] in 1991. An example of
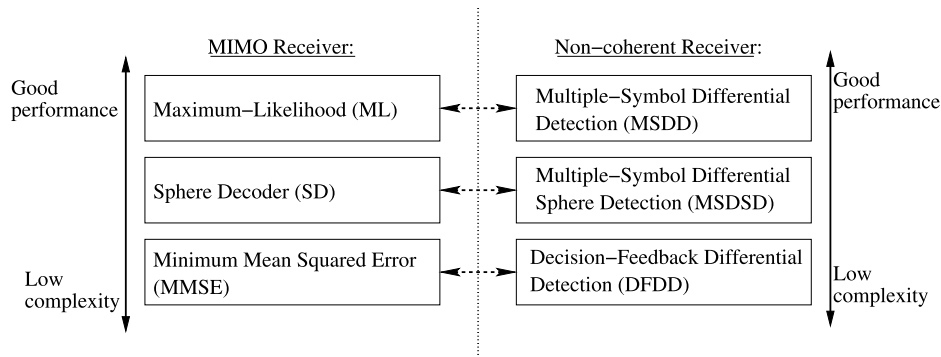
FIGURE 13. Links between generic MIMO detection schemes and generic non-coherent detection schemes.

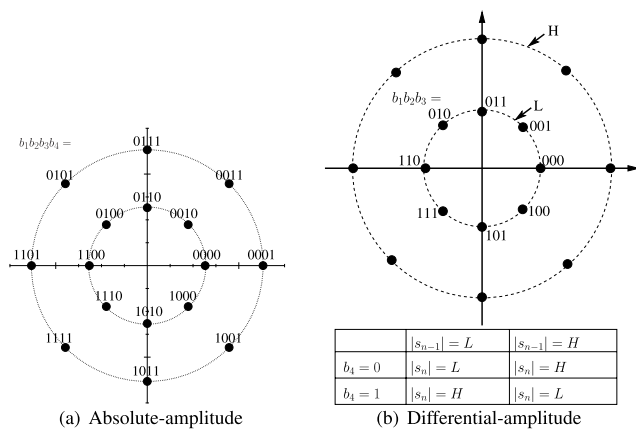(a) Absolute-amplitude   (b) Differential-amplitude

FIGURE 14. The constellation diagram of the absolute-amplitude scheme of 16-ADPSK and that of the differential-amplitude scheme of 16-DAPSK.

16-DAPSK is portrayed in Fig. 14(b). More explicitly, the 16-DAPSK scheme's first three bits are assigned to modulate the change in phase between two consecutive transmitted symbols, while the last bit is assigned to modulate the change in amplitude ring. It can be seen in Fig. 14(b) that in order to maintain a Star 16QAM constellation for the transmitted symbols, $b_4 = 0$ and $b_4 = 1$ determines whether the consecutive transmitted symbols' amplitude $|s_{n-1}|$ and $|s_n|$ should be retained or be toggled to its other legitimate value. In other words, if the DAPSK's differential encoding process is to be represented by $s_n = x_{n-1}s_{n-1}$, the transmitted symbols s_{n-1} and s_n are always drawn from the classic Star QAM constellation. However, the specific constellation diagram of the data-carrying symbol x_{n-1} depends on the previously transmitted amplitude $|s_{n-1}|$. Considering the 16-DAPSK scheme of Fig. 14(b) as an example, if we have $|s_{n-1}| = L$, then $b_4 = 0$ and $b_4 = 1$ for modulating $|x_{n-1}|$ should enable the transitions of $L \rightarrow L$ and $L \rightarrow H$, respectively. However, if we have $|s_{n-1}| = H$, then the modulation of $|x_{n-1}|$ should enable the different amplitude transitions of $H \rightarrow H$ and $H \rightarrow L$ for $b_4 = 0$ and $b_4 = 1$, respectively. Nonetheless, the phase of 16-DAPSK's data-carrying symbol x_{n-1} is still drawn from the original 8PSK constellation diagram. It was proposed in [187] that the

detection of DAPSK's data-carrying amplitude may be carried out by testing the amplitude change between consecutive received samples $|y_n|/|y_{n-1}|$, while the data-carrying phase may be detected by testing the phase change $\angle y_n - \angle y_{n-1}$. In summary, the fundamental difference between the ADPSK and the DAPSK is that the ADPSK aims for maintaining the mapping regime of a Star QAM constellation for the data-carrying symbols, while the DAPSK aims for maintaining a Star QAM constellation for the transmitted symbols.

The development of DAPSK has attracted substantial research interests. The performance comparison between coherent QAM and CDD aided DAPSK was provided by Adachi and Sawahashi [188] and also by Chow *et al.* [189] in 1992. In 1995, Rohling and Engels [190] proposed the application of DAPSK in digital terrestrial video broadcasting, where the authors compared the performance of the CDD aided DAPSK to that of the coherent QAM in the presence of realistic channel estimation errors. Despite the satisfactory performance of CDD aided DAPSK in AWGN and block fading channels, it was observed by the authors of [187] that the DAPSK performance degrades and eventually an error floor is formed, as the Doppler frequency is increased. It was suggested by the authors of [187] that both oversampling and channel coding may be invoked for mitigating this problem. We note that oversampling was also proposed for improving the performance of DPSK in rapidly fading channels [191], [192]. Moreover, it was observed by Chow *et al.* [193] in 1993 and then analysed by Chung [194] in 1997 that employing multiple RAs is capable of reducing the error floor of DAPSK in rapidly fading channels. Nonetheless, the optimum non-coherent detector of both the ADPSK and the DAPSK is the MSDD characterized by Divsalar and Simon [42] in 1994, which is capable of improving the performance of DQAM in different channel scenarios.

In order to mitigate the MSDD's exponentially increasing complexity as a function of the window-width, the prediction-based DFDD was proposed for 16-DAPSK operating in AWGN channels by Adachi and Sawahashi [195] in 1996, and then its amplitude detection was improved by Wei and Lin [196] in 1998. The MSDD-based DFDD was also proposed for 16-DAPSK for transmission over

AWGN channels by Schober *et al.* [197] in 1998. Following this, the prediction-based DFDD designed for DAPSK communicating in fading channels was proposed by Gerstacker *et al.* [198] in 1999 and later it was improved by Schober *et al.* [199]. Notably, the prediction-based DFDD specifically conceived for ADPSK operating in fading channels was proposed by Lampe *et al.* [185] in 2001.

Inspired by the near-capacity performance achieved by turbo detection, soft-decision-aided DQAM detection has also been developed throughout the last two decades. May *et al.* [200] proposed Trellis decoded DQAM using the Viterbi algorithm in 1998, and then Fischer *et al.* [201] invoked MSDD for DQAM in both multilevel coding and in Bit-Interleaved Coded Modulation (BICM) in 2001. Moreover, Ishibashi *et al.* [202] proposed the low-complexity soft-decision-aided CDD concept conceived for DAPSK operating in Rayleigh fading channels, where the amplitude and phase are separately detected. However, no iteration was invoked between the channel decoder and the DQAM detector in these contributions. Liang *et al.* [203] proposed to employ the CDD for DAPSK relying on turbo detection in 2011, where the amplitude and phase are jointly detected. This soft-decision-aided CDD conceived for DAPSK was further streamlined by Xu *et al.* [204] in 2013, where the authors also discovered that completely separately detecting the DAPSK's amplitude and phase may impose a performance loss, which is more substantial in coded systems. Furthermore, in 2012, Wang and Hanzo [205] proposed a new arrangement for soft-decision-aided DAPSK detection. More explicitly, Wang and Hanzo [205] proposed to invoke MSDD and MSDSD for detecting the DAPSK's amplitudes and phases, which may be referred to as Multiple Symbol Differential Amplitude Detection (MSDAD) and Multiple Symbol Differential Phase Sphere Detection (MSDPSD), respectively. Given that the data-carrying amplitudes and phases are correlated, MSDAD and MSDPSD are capable of iteratively exchanging their decisions in order to achieve a near-optimum MSDD performance for DAPSK with the aid of *a priori* information gleaned from the channel decoder.

Nonetheless, the major challenge of DQAM detection is that unlike DPSK, the transmitted symbol's amplitudes do not form a unitary matrix for MSDD, MSDSD and DFDD. More explicitly, the MSDD [42] relies on the knowledge of channel correlation, which is determined by the Doppler frequency and the noise power. For DPSK, the transmitted phases may form a unitary matrix, which may be separated from the channel correlation matrix, so that a lower triangular matrix that is created by decomposition from the inverse of the channel's correlation matrix may be utilized in the context of sphere decoding [49], [50]. However, for DQAM, the transmitted symbol-amplitudes cannot form a unitary matrix, hence the symbol-amplitude-dependent channel correlation matrix only becomes known, when all the symbol-amplitudes are detected. This is the reason why Wang [205] invoked MSDD for amplitude detection. As a result, without the assistance of channel coding, Wang's solution [205] may introduce an error floor for uncoded DAPSK schemes, because the MSDAD and the MSDPSD may exchange erroneous decisions. Furthermore, the DFDD conceived for DQAM operating in fading channels and documented in the existing literature [185], [198], [199] relies on the same constant channel correlation matrix that was derived for DPSK, which implies that these DFDD solutions are sub-optimal and they are not equivalent to the decision-feedback version of MSDD.

In order to mitigate this problem, a dynamic MSDSD algorithm was conceived for DQAM by Xu *et al.* [206] in 2016, where the optimal estimation of a submatrix of the channel correlation matrix is carried out at each step of SD, so that the holistic channel correlation matrix may be recovered, once the SD is terminated. Moreover, the associated DFDD solution becomes equivalent to the optimal MSDD of [42] operating in the decision-feedback mode, which substantially outperforms the previous solutions of [185], [198], [199]. Furthermore, the corresponding soft-decision aided MSDSD and DFDD schemes were proposed in [207], where the options of hard-decision-directed and soft-decision-directed SD algorithms were devised, depending on the specific treatments of the first symbol-amplitude within the MSDSD window. Additional reduced-complexity detection algorithms are also conceived by exploiting the symmetry of the Star QAM constellations.

D. SUMMARY ON THE HISTORICAL PERSPECTIVE
This section is summarized by a list of major contributions on non-coherent detection in Table 5. We now proceed by introducing non-coherent MIMO detection in the following sections.

IV. TWENTY YEARS OF COHERET/NON-COHERENT MIMO TRADEOFF
A. GENERAL MIMO DESIGN TRADEOFFS
The three key MIMO design tradeoffs are portrayed in Fig. 15. First of all, the multiplexing/diversity tradeoff constitutes the most salient MIMO design tradeoff since the inventions of Foschini's Bell Laboratories Layered Space-Time (BLAST) [208] in 1996 and Alamout's Space-Time Block Code (STBC) [209] in 1998. On one hand, the family of BLAST schemes [208], [210], [211] effectively use M Transmit Antennas (TAs) to transmit M indepdent data streams. The resultant ergodic capacity may grow linearly, rather than logarithmically, with the number of antennas [212]. However, the simultaneous transmission of multi-stream signals imposes Inter-Antenna Interference (IAI), which results in the ML detection complexity growing exponentially with the number of TAs. In order to alleviate this problem, a variety of BLAST detectors including the potentially excessive-complexity ML scheme [23], [91], SD [213]–[217], the popular MMSE arrangement [91], [218], [219] as well as a variety of decision-feedback techniques [220]–[222] have been devised in order to strike a performance/complexity tradeoff in Fig. 15, which was also previously illustrated in Fig. 13.

TABLE 5. Summary of major contributions on non-coherent detection.

Year	Author(s)	Topic	Contribution
1959-1960	Lawton [35], [36]	DPSK	Proposed that the constant phase rotation in AWGN channels may be mitigated by the simple DPSK transceiver.
1959	Cahn [37]	DPSK	Demonstrated that the CDD aided DPSK scheme suffers from a 3 dB performance penalty in AWGN channels compared to its coherent PSK assuming the idealistic perfect estimation of the frequency offset phase.
1960	Kailath [31]	MLSE	Proposed to jointly perform channel estimation and data detection, where channel estimation was tentatively performed for each possible combination of the data-carrying sequence, and then these combinations were compared in order to make a final decision on the data-carrying sequence.
1962	Bello and Nelin [38]	DPSK	Discovered that an irreducible error floor occurs, when the CDD that was originally designed for AWGN channels is employed in the rapidly fluctuating fading channels.
1972	Forney [33]	MLSE	Proposed to invoke the Viterbi algorithm for MLSE, which relies on the early termination of the candidate evaluation process by ignoring those candidates that are unlikely to be the maximum likelihood solution.
1979	Morley and Snyder [34]	MLSE	Proposed that with the aid of Viterbi algorithm, the MLSE complexity grows exponentially with the channel memory that imposed onto the data transmission in the form of Inter-Symbol Interference (ISI).
1982	Simon *et al.* [184]	DQAM	Proposed the absolute-amplitude DQAM scheme based on Square QAM constellation, where the transmitted phase was differentially encoded but the transmitted data-carrying amplitude was the original data-carrying amplitude.
1988	Leib and Pasupathy [43]	DFDD (DPSK)	Proposed DFDD for DPSK operating in AWGN channels, which was later shown in [46] to be equivalent to the MSDD of [40] operating in decision-feedback mode.
1990	Divsalar and Simon [40]	MSDD (DPSK)	Proposed MSDD for DPSK operating in AWGN channels, which partially compensated the 3 dB performance penalty by observing more than two received samples.
1991	Webb, Hanzo and Steele [187]	DAPSK	Proposed DAPSK which differentially encodes both amplitude and phase, where the Star QAM constellation is maintained for all transmitted symbols.
1992	Ho and Fung [41]	MSDD (DPSK)	Proposed MSDD for DPSK operating in Rayleigh fading channels, which mitigated the error floor in rapidly fluctuating fading channels.
1994	Divsalar and Simon [42]	MSDD (DPSK DQAM)	Summarized the MSDD techniques developed both for DPSK and for DQAM operating both in AWGN channels and in fading channels.
1994	Mackenthun [177]	MSDD (DPSK)	Proposed an efficient algorithm for MSDD aided DPSK operating in AWGN channels.
1996	Adachi *et al.* [195]	DFDD (DAPSK)	Proposed prediction-based DFDD for 16-DAPSK operating in AWGN channels.
1998	Schober *et al.* [197]	DFDD (DAPSK)	Proposed MSDD-based DFDD for 16-DAPSK for transmission over AWGN channels.
1999	Schober *et al.* [47]	DFDD (DPSK)	Proposed DFDD for DPSK operating in Rayleigh fading channels, which is derived from the MSDD of [41], [42].
1999	Bin and Ho [180]	DFDD (DPSK)	Proposed prediction-based DFDD for DPSK operating in Rayleigh fading channels, where a channel sample is predicted based on the previous decisions, so that coherent detection may be performed for the current symbol.
1999	Hoeher and Lodge [53]	DFDD (DPSK)	Proposed soft-decision-aided prediction-based DFDD for DPSK operating in fading channels, so that DFDD may be invoked in turbo detection.
2000	Schober and Gerstacker [57]	DFDD (DPSK)	Proved that both MSDD-based DFDD [47], [48] and prediction-based DFDD employing the MMSE Wiener filter [53], [179], [180] are identical for DPSK operating both in AWGN and in Rayleigh channels.
2001	Schober *et al.* [199]	DFDD (DAPSK)	Proposed prediction-based DFDD for 16-DAPSK communicating in fading channels.
2001	Lampe *et al.* [48]	DFDD (DPSK)	Proposed soft-decision-aided MSDD-based DFDD for DPSK operating in Rayleigh fading channels, where the DFDD was involved in turbo detection.
2001	Lampe and Schober [185]	DFDD (ADPSK)	Proposed the absolute-amplitude DQAM scheme based on the Star QAM constellation, and proposed prediction-based DFDD conceived for this ADPSK scheme.
2005	Lampe *et al.* [49]	MSDSD (DPSK)	Proposed MSDSD for DPSK operating in Rayleigh fading channels, which invoked a sphere decoder for MSDD.
2005	Ishibashi *et al.* [202]	DAPSK	Proposed the low-complexity soft-decision-aided CDD conceived for DAPSK in Rayleigh fading channels, where the amplitude and phase are separately detected.
2006	Pauli *et al.* [50]	MSDSD (DPSK)	Proposed soft-decision-aided MSDSD for DPSK communicating in Rayleigh fading channels, so that MSDSD may be invoked in turbo detection.
2011	Liang *et al.* [203]	DAPSK	Proposed to invoke the CDD for DAPSK in turbo detection in 2011, where the amplitude and phase are jointly detected.
2012	Wang and Hanzo [205]	MSDSD (DAPSK)	Proposed to invoke MSDD and MSDSD for detecting the DAPSK's amplitudes and phases, which may be referred to as MSDAD and MSDPSD, respectively. The pair are capable of iteratively exchanging decisions in order to achieve a near-optimum MSDD performance in coded systems.
2013	Xu *et al.* [204]	DAPSK	Streamlined the soft-decision-aided CDD of [203], and discovered that completely separately detecting DAPSK's amplitude and phase may impose a performance loss, which is more substantial in coded systems.
2016	Xu *et al.* [206], [207]	MSDSD /DFDD (DAPSK /ADPSK)	Proposed dynamic MSDSD/DFDD algorithms that facilitate the estimation of the QAM-amplitude-dependent channel correlation matrix. The hard-decision and soft-decision MSDSD solutions achieve the optimal MSDD performance at substantially reduced complexities. Moreover, the associated DFDD solutions also outperform the LPDD solutions in [185], [198], [199] that rely on the constant channel correlation matrix derived for DPSK

On the other hand, for the family of orthogonal STBCs [209], [223]–[226], the spatial resources of mutliple TAs have been exploited for the sake of achieving a diversity gain, where multiple replicas of the modulated symbols are transmitted by multiple TAs over multiple symbol periods. Thanks to the orthogonality of this design, the multiple data
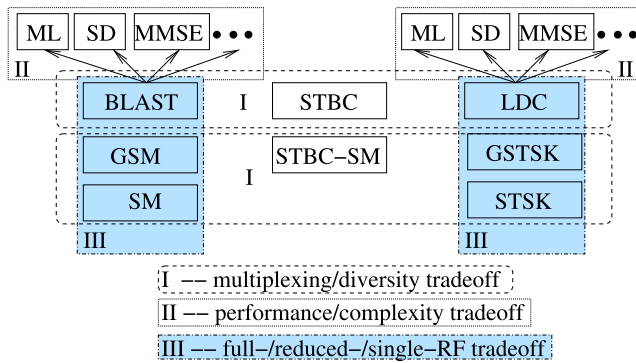
FIGURE 15. The three key MIMO tradeoffs of multiplexing/diversity tradeoff, performance/complexity tradeoff, full-/reduced-/single-RF tradeoff.

(a) Spatial Modulation (SM)

(b) Space-Time Shift Keying (STSK)

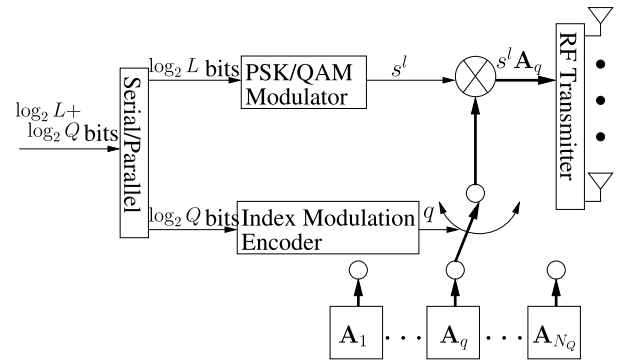**FIGURE 16.** Schematics of Spatial Modulation (SM) and Space-Time Shift Keying (STSK).

streams may be decoupled at the receiver, which eliminates IAI. However, owing to the replica transmission, the MIMO bandwidth-efficiency is undermined by this STBC design. In order to improve the STBC throughput, the family of Quasi-Orthogonal (QO) STBCs [227]–[233] have been conceived, where the problem of IAI is encountered again, because the orthogonality is compromised.

In the landmark paper [234] by Zheng and Tse published in 2003, it was shown that both multiplexing and diversity gains can be simultaneously obtained. Inspired by this, the family of Linear Dispersion Codes (LDCs) [235], [236] that are capable of achieving both the full BLAST-throughput and the full STBC-diversity have been extensively developed. More explicitly, according to the LDC representation of [236], the MIMO signals transmitted by M TAs over T symbols periods are modelled as $\mathbf{S} = \sum_{q=1}^{Q} s_q \mathbf{A}_q$, where a total of Q modulated symbols are dispersed by Q number of $(T \times M)$-element dispersion matrices $\{\mathbf{A}_q\}_{q=1}^{Q}$. These dispersion matrices that maximize the diversity gain may be obtained by random generation [235], [236], gradient search [235], [237], as well as divison algebra [238]–[243]. Notably, the so-called Golden code [239] as well as its generic extension termed as perfect STBC [240]–[243] are capable of always achieving a better performance than both V-BLAST and STBC, when several Receive Antennas (RAs) ($N > 1$) are used. However, on one hand, the effect of IAI is encountered again, which results in the same performance/complexity tradeoffs upon employing MIMO detectors, as those shown in Fig. 15. On the other hand, owing to the fact that the modulated symbols multiplied by dispersion matrices are superimposed before pulse-shaping, the PAPR of the LDC signals is substantially increased. The issues of the potentially escalating constellation cardinality and PAPR will be further discussed in the next section.

As the number of antennas employed both at the BS and the UE grows, it becomes excessively complex to assign a dedicated RF chain to all antenna elements, as required by the conventional MIMO schemes of BLAST, STBC and LDC. To mitigate this problem, the Index Modulation (IM) philosophy has been applied for creating a variety of single-RF
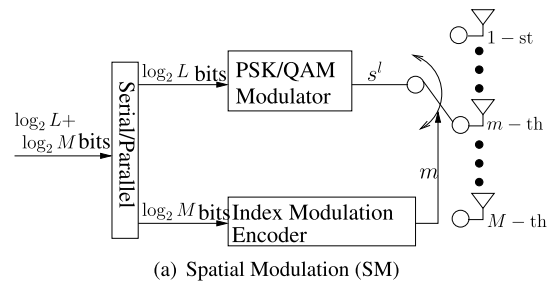
and reduced-RF MIMO schemes [68]–[79]. Specifically, the concept of Spatial Modulation (SM) was proposed by Song *et al.* [70] in 2004, which is an extension of the scheme proposed in 2001 by Chau and Yu [71]. The schematic of SM shown in Fig. 16(a) has attracted substantial attention ever since the explicit demonstration of the SM's wide-ranging advantages by Mesleh *et al.* [72] in 2008. More explicitly, Fig. 16(a) shows that two blocks of source bits are separately assigned for modulating a LPSK/QAM symbol and an activation index, so that only a single one out of M TAs is activated. In order to achieve a beneficial diversity gain, the concept of Space-Time Shift Keying (STSK) was proposed by Sugiura *et al.* [123] in 2010, where the IM bits are assigned to activate a single one out of Q LDC dispersion matrices, as seen in Fig. 16(b). By configuring the dispersion matrices to be sparse, the Asynchronous STSK (ASTSK) arrangement conceived in [123] also guarantees single-RF transmission. As a result, the SM and its diversity-assisted counterparts impose a substantially reduced power consumption when relying only on a single RF chain, regardless of the total number of antennas available. Moreover, the Inter-Antenna Synchronization (IAS) is avoided and the IAI is also eliminated, which facilitates low-complexity single-stream-based ML detection at the receiver [244], [245]. Nonetheless, the bandwidth-efficiency of SM and STSK is only increased logarithmically with M and Q, which cannot compete with the bandwidth-efficiency of V-BLAST and LDC. As a remedy, the reduced-RF Generalized SM (GSM) and Generalized STSK (GSTSK) schemes [75], [246]–[249] were conceived, but the problems of IAS and IAI resurface once again. A special case of Space-Time Block Coded Spatial Modulation

(STBC-SM) was conceived by Basar et al. [73], where a reduced number M_A out of M TAs is activated to convey the STBC signal matrix. More comprehensive tutorial materials on IM may be found in [68], [77]–[79].

The energy-efficiency of both single-RF SM schemes and of the family of full-RF MIMO schemes is evalutated in [250], [251]. More explicitly, the energy-efficiency is characterized in [250], [251] by the modulation-independent Shannon capacity divided by the power consumption at the BS in the DL. It is demonstrated in [251] that despite its typically lower data rate, SM achieves a potentially better energy-efficiency than its full-RF MIMO counterparts, provided that the radiated power of the single RF chain is not driven into saturation. By contrast, the V-BLAST scheme achieves a higher overall data rate and also a better energy-efficiency, when the total radiated power approaches its overall maximum. It is worth noting that under the current LTE/NR frameworks, only the classic OFDM waveform is invoked for the DL, as seen in Tables 3 and 4. However, the family of single-RF and reduced-RF MIMO schemes has to rely on single-carrier transmission, because the IFFT-based multi-carrier OFDM processing spreads the signal and hence activates all the RF chains [69], [77], [252]. Therefore, we may argue that the SM scheme may become particularly suitable for upgrading the single-carrer DFT-S-OFDM aided UL, where currently only a single TA is activated, as discussed in Sec. II-D. Similarly, in the IoT scenarios, a variety of unlicensed networks including Bluetooth, Zigbee and IEEE 802.11ah as well as 5G IoT enablers of Long Range (LoRa) and NarrowBand IoT (NB-IoT) [253], [254] currently all employ a single TA. All of these scenarios offer compelling opportunities both for SM and for its diversity-aided counterparts, which are capable of exploiting the MIMO's multiplexing and/or diversity gains without the extra cost of additional RF chains.

More explicitly, from the PA's perspective, the tradeoff between single-RF, reduced-RF and full-RF schemes is characterized in Fig. 17. On one hand, even at a zero PA output power of $P_{out} = 0$, the transmitter architecture continues to consume non-negligible power, which may even be as high as 50% of the PA's input power P_{in} [255]. The associated transmit power independent power dissipation grows linearly with the number of activated RF chains, as seen in Fig. 17. Secondly, the PA of the single-RF system has to deliver M times higher output power P_{out} than its counterparts operating in the full-RF mode, hence the transmit power dependent term exhibits a reversed trend with the number of RF chains in Fig. 17. In summary, a reasonable rule of thumb is that when a lower-gain PA is employed, the RF-drive power is of a higher proportion in P_{in} [256], hence reducing the number of RF chains may become even more beneficial. This is particularly suitable both for the power-hungry UL handset as well as for the increasingly dense Device-to-Device (D2D) links both in the operational and in future networks. Furthermore, based on the different requirements on both the output power P_{out} and PA efficiency $\eta = P_{out}/P_{in}$, the different commercially
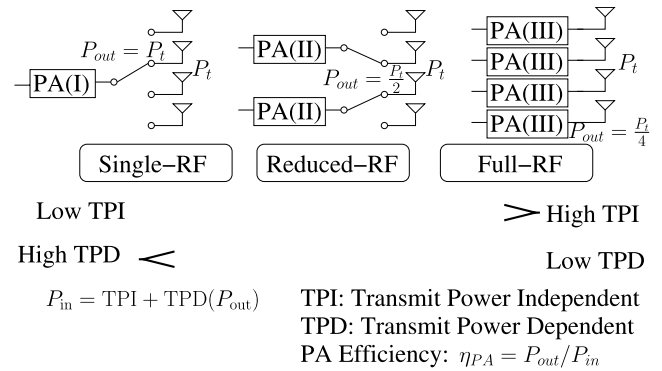
FIGURE 17. The tradeoff between single-RF, reduced-RF and full-RF schemes from the PA's perspective.

available PAs should be chosen accordingly, as demonstrated in [145], [146]. Otherwise, the lower transmit power $P_{out} = P_t/M$ of full-RF schemes may impose a low efficiency η, which leads to a substantial heat dissipation. In summary, when a SIMO scheme is upgraded to a MIMO arrangement, the single-RF SM and its diversity counterparts are particularly suitable, as the same choice of PA may be reused.

Nonetheless, it is worth noting that the SM schemes require a new hardware device in form of an RF switch, which has already been widely used in a variety of PA structures. Specifically, the classic Doherty technique [145], [256] is constituted by a main class-B PA and an auxiliary class-C PA, where the latter is only switched on for high signal amplitudes. The associated transistors can be turned on and off at an ultrasonic rate [257], which can be convenient utlized as the RF switch for the SM structrure of Fig. 16.

B. FINITE-CARDINALITY DIFFERENTIAL SPACE-TIME MODULATION (DSTM)

Since the pilot overhead of coherent detection inevitably grows with the number of antennas [27], the family of Differential Space-Time Modulation (DSTM) schemes that dispenses with high-complexity channel estimation has attracted substantial research interests. Following the success of Alamouti's G2 STBC [209], Differential STBC (DSTBC) using the same signal structure was proposed by Tarokh and Jafarkhani [113] in 2000, which was further extended to use both multiple TAs [114], [115] and star/square QAM constellations [116]–[118]. The DSTM's ($T \times M$)-element signal matrix $\mathbf{S}_n$ is obtained by the matrix-based differential encoding of $\mathbf{S}_n = \mathbf{X}_{n-1}\mathbf{S}_{n-1}$, where the ($T \times T$)-element data-signal matrix $\mathbf{X}_{n-1}$ carries source information. It was demonstrated both by Hochwald and Sweldens [127] as well as by Hughes [126] in 2000 that $\mathbf{X}_{n-1}$ may be designed to be unitary, so that the transmitted signals matrix $\mathbf{S}_n$ retains orthogonal rows, and the average signal power remains the same over differential encoding in the time-domain.

Following the unitary DSTM design philosophy, the family of Differential Group Codes (DGCs) was proposed in [126], [127], where the signal matrices including $\mathbf{S}_n$ and $\mathbf{X}_{n-1}$ form a finite group under multiplication. More

explicitly, a group is formed by a set of signal matrices $\{\mathbf{X}^i\}_{i=1}^I \in \mathcal{X}$ and an operation of matrix multiplication, which must comply with the following group axioms [258]:

(1) Closure: for all $\mathbf{X}^i$ and $\mathbf{X}^{i'}$ in $\mathcal{X}$, the result of operation $\mathbf{X}^i\mathbf{X}^{i'}$ is also in $\mathcal{X}$.
(2) Associativity: for all $\mathbf{X}^i$, $\mathbf{X}^{i'}$ and $\mathbf{X}^{i''}$ in $\mathcal{X}$, we have $(\mathbf{X}^i\mathbf{X}^{i'})\mathbf{X}^{i''} = \mathbf{X}^i(\mathbf{X}^{i'}\mathbf{X}^{i''})$.
(3) Identity element: the identity matrix is in $\mathcal{X}$.
(4) Inverse element: for all $\mathbf{X}^i$ in $\mathcal{X}$, the inverse $(\mathbf{X}^i)^{-1}$ is also in $\mathcal{X}$.

As a result, despite the matrix-based differential encoding operation of $\mathbf{S}_n = \mathbf{X}_{n-1}\mathbf{S}_{n-1}$, the cardinality of the DGC's transmitted signals does not tend to infinity, which is in contrast to the trend encountered by the family of DSTBC [113]–[118]. Nonetheless, the available choices of full-diversity finite-group DGCs remain limited, which are explicitly summarized in [128], [129]. In order to improve the DGC's throughput, a variety of non-group based and infinite group based solutions² were further conceived in [259]–[263].

The non-coherent counterpart of LDC, namely Differential Linear Dispersion Code (DLDC) was proposed by Hassibi and Hochwald [119] in 2002, which was further optimized in [120]–[122] for attaining an improved performance. More explicitly, in order to obtain unitary signal matrices, the Cayley transform is invoked to convert the Hermitian matrix $\widetilde{\mathbf{X}} = \sum_{q=1}^Q x_q \widetilde{\mathbf{A}}_q$ to unitary, where Q number of modulated symbols $\{x_q\}_{q=1}^Q$ are dispersed by $\{\widetilde{\mathbf{A}}_q\}_{q=1}^Q$ that are generated to be Hermitian matrices. This is due to the fact that the summation of unitary matrices does not remain unitary, but the summation of Hermitian matrices always leads to a Hermitian matrix. It is worth noting that due to the associated non-linear mapping, both DGCs and DLDCs exhibit detection complexities that grow exponentially with the throughput. As a remedy, the polynomial-complexity lattice decoding and sphere decoding arrangements were devised for DGC and DLDC in [264] and [119], respectively, which however impose a performance loss.

In order to mitigate this excessive complexity, Differential STSK (DSTSK) was conceived by Sugiura *et al.* [123] in 2010, which was extended to QAM constellations in [124], while the DLDC's Cayley transform was eliminated in [125]. More explicitly, the DSTSK's unitary data-carrying matrix is given by $\mathbf{X} = x^l \mathbf{A}_q$, where both the modulated symbol index l and the dispersion matrix activation index q carry source bits. Moreover, in the absence of the DLDC's matrix summation, the DSTSK's dispersion matices $\{\mathbf{A}_q\}_{q=1}^Q$ are directly generated to be unitary, which dispenses with the DLDC's Cayley transform. Furthermore, inspired by the success of SM, the Differential SM (DSM) concept was proposed by Bian *et al.* [130], [131] in 2013, which was further extended to the employment of star QAM in [132], [133]. Moreover,

a beneficial transmit diversity was also achieved by the single-RF DSM schemes in [265], [266].

It is worth noting that although the majority of DSTM schemes generally rely on unitary matrix design, it was demonstrated by Bhatnagar *et al.* [267] that non-unitary MIMO schemes - including QO-STBCs [227]–[233], generic LDCs [120], [235], [236], [238], [268], [269], Golden code [239] and perfect STBCs [240]–[243]- may also be invoked by the matrix-based differential encoding operation of $\mathbf{S}_n = \mathbf{X}_{n-1}\mathbf{S}_{n-1}$ in conjunction with appropriate power normalization. As a result, the full benefits of multiplexing and diversity gains of coherent MIMO schemes may also be exploited by the family of DSTM arrangements.

SISO DPSK:	$s_n = x_{n-1}s_{n-1}$

DSTBC (M=2): $\mathbf{S}_n = \mathbf{X}_{n-1}\mathbf{S}_{n-1}$

$$\begin{bmatrix} s_{n,1} & s_{n,2} \\ -s_{n,2}^* & s_{n,1}^* \end{bmatrix} = \begin{bmatrix} x_{n-1,1} & x_{n-1,2} \\ -x_{n-1,2}^* & x_{n-1,1}^* \end{bmatrix} \begin{bmatrix} s_{n-1,1} & s_{n-1,2} \\ -s_{n-1,2}^* & s_{n-1,1}^* \end{bmatrix}$$

$$s_{n,1} = x_{n-1,1}s_{n-1,1} - x_{n-1,2}s_{n-1,2}^*$$
$$s_{n,2} = x_{n-1,1}s_{n-1,2} + x_{n-1,2}s_{n-1,1}^*$$

DGC−cyclic (M=2):

$$\begin{bmatrix} s_{n,1} & 0 \\ 0 & s_{n,2} \end{bmatrix} = \begin{bmatrix} x_{n-1,1} & 0 \\ 0 & x_{n-1,2} \end{bmatrix} \begin{bmatrix} s_{n-1,1} & 0 \\ 0 & s_{n-1,2} \end{bmatrix}$$

$$s_{n,1} = x_{n-1,1}s_{n-1,1}$$
$$s_{n,2} = x_{n-1,2}s_{n-1,2}$$

FIGURE 18. Differential encoding of the SISO DPSK, DSTBC ($M = 2$) using PSK and DGC-cyclic schemes ($M = 2$). The DSTBC and DGC schemes are further explained by Example 1 and Example 4, respectively.

However, in contrast to DPSK and DAPSK/ADPSK introduced in Sec. III, which always retain the same PSK and star QAM constellations for the transmitted signals after differential encoding, the matrix multiplications in $\mathbf{S}_n = \mathbf{X}_{n-1}\mathbf{S}_{n-1}$ lead to the so-called infinite-cardinality problem, which is analysed in great detail in [19]. Specifically, Fig. 18 exemplifies that compared to DPSK, the DSTBC's differential encoding involves signal additions/subtractions, which result in the indistinguishable constellation diagram seen in Fig. 19. As a result, both the cardinality and PAPR of the DSTBC signals tends to infinity, despite the fact that the practical DACs can only operate based on finite resolution, while the practical PAs only have limited dynamic range. In order to formally state this problem, we offer the following propositions.

Proposition 1: For the DSTBC scheme using Alamouti's signal structure [23], [113], [115] and LPSK signalling ($L \geq 8$), the theoretical cardinality of the transmitted signal set $\mathbf{S}_n = \mathbf{X}_{n-1}\mathbf{S}_{n-1}$ is given by:

$$\mathcal{L}_n = \frac{\mathcal{L}_{n-1}(\mathcal{L}_{n-1} - 1)}{4}, \quad \text{for } n > 2,$$

where $\mathcal{L}_n$ denotes the cardinality of the signals in the set $\mathbf{S}_n$ associated with time index n.

Please refer to Appendix A for the proof.

Proposition 2: For the DSTBC scheme using Alamouti's signal structure in [23], [113], [115] and LPSK signalling

---

²We note that non-group refers to the set that does not comply with the group axioms, while the term infinite group refers to a group that has infinite number of elements.

TABLE 6. Summary of the DSTM schemes.

	Transmitter complexity		Receiver complexity		Performance		
	Finite-cardinality?	Single-RF?	Single-stream ML detection?	Detection complexity order	Transmit diversity?	Diversity gain maximization?	Throughput
DSTBC [113]–[115]	×	×	✓	$\mathcal{O}(Q)$	✓	×	$R = \frac{Q \log_2 L}{T}$
DLDC [119], [263]	×	×	×	$\mathcal{O}(I = 2^{RT})$	✓	✓	$R = \frac{Q \log_2 L}{T}$
DSTSK [123]–[125]	×	×	✓	$\mathcal{O}(Q)$	✓	✓	$R = \frac{\log_2 LQ}{T}$
DGC [126]–[129]	✓	✓	×	$\mathcal{O}(I = 2^{RT})$	✓	✓	$R = \frac{\log_2 L}{T}$
DSM [130]–[133]	✓	✓	✓	$\mathcal{O}(2^{\lfloor \log_2 T! \rfloor})$	×	×	$R = \frac{T \log_2 L + \lfloor T! \rfloor}{T}$
DSTBC-ISK [19]	✓	✓	✓	$\mathcal{O}(T)$	✓	×	$R = \frac{\log_2 LT}{T}$
DSTSK-TAST [20]	✓	✓	✓	$\mathcal{O}(QT)$	✓	✓	$R = \frac{\log_2 LQT}{T}$

(a) Constellation for **X**   (b) Constellation for **S**

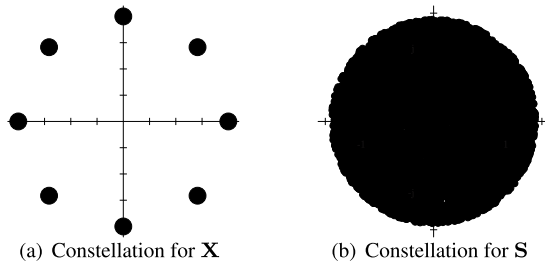**FIGURE 19. Constellation diagrams for DSTBC signals in X and S of Fig. 18, where (M = 2) TAs and 8PSK signals are used.**

($L \geq 8$), the theoretical PAPR of the transmitted signals in $\mathbf{S}_n = \mathbf{X}_{n-1}\mathbf{S}_{n-1}$ is given by:

$$\text{PAPR}_n = \text{PAPR}_{n-1} + 3\text{dB}, \quad \text{for } n > 2,$$

where PAPR_n denotes the PAPR of signals in $\mathbf{S}_n$ associated with time index n and expressed in dB.

Please refer to Appendix B for the proof. By contrast, due to the sparsity of the signal matrix design, the DGC scheme exemplified in Fig. 18 is still capable of retaining the same PSK constellation for all transmitted signals. Against this background, the finite-cardinality design is summarized in [19] as:

Finite-Cardinality Low-PAPR Design: In order to arrive at a finite-cardinality transmit-signal set that retains the same low-PAPR as the intended modulated PSK/QAM symbols, it is sufficient to ensure that the signal matrices belong to a specific type of sparse matrices that have only a single non-zero element in each row and column, where the non-zero elements assume the equi-spaced phases and/or ring-amplitudes of the PSK and star QAM constellations.

The DSTM schemes including DSTBC [113]–[118], DLDC [119]–[122] and DSTSK [123]–[125] all suffer from the same infinite-cardinality problem, as summarized in Table 6. The low-complexity transceiver features of finite-cardinality low-PAPR signal transmission and the single-stream ML detection capability are highlighted in Table 6, where only the three recently-developed schemes of DSM [130]–[133], DSTBC using Index Shift Keying (DSTBC-ISK) [19] as well as DSTSK using the Treaded Algebraic Space-Time (TAST) arrangement of [20] satisfy the above-mentioned low-complexity transceiver requirements. Specifically, DSM has a throughput of

$R = \frac{T \log_2 L + \lfloor T! \rfloor}{T}$, which is higher than that of the SIMO systems. Moreover, thanks to the STBC structure, DSTBC-ISK achieves an attractive performance at low SNR associated with low throughputs, while retaining the lowest detection complexity. Furthermore, DSTSK-TAST improves the diversity/throughput tradeoff with the aid of its beneficial dispersion matrix design, hence DSTSK-TAST is capable of achieving a persistent performance advantage over DSM for a wide range of SNRs, throughputs and MIMO setups.

More explicitly, the DSM scheme modulates a total number of M L-PSK/QAM symbols $\{x^{l_q}\}_{q=1}^M$ by $M \log_2 L$ bits, and their permutation locations in the signal matrix is determined by the permutation index $\bar{m}$ modulated by $\lfloor \log_2 M! \rfloor$ bits. As exemplified by Table 7, the resultant signal matrix is given by $\mathbf{X} = \sum_{q=1}^M x^{l_q} \mathbf{A}_{\bar{m},q}$, where $\{\mathbf{A}_{\bar{m},q}\}_{q=1}^M$ represent the permutation-based dispersion matrices. As a result, first of all, the DSM using PSK and star QAM constellations [130]–[133] complies with the finite-cardinality design principle, where the transmitted signals retain their original PSK or star QAM format under the matrix-based differential encoding operation of $\mathbf{S}_n = \mathbf{X}_{n-1}\mathbf{S}_{n-1}$. Secondly, owing to the linear mapping of the unitary DSM matrix $\mathbf{X} = \sum_{q=1}^M x^{l_q} \mathbf{A}_{\bar{m},q}$, which follows the same orthogonality principles as STBC, the single-stream ML detection may also be applied to DSM as demonstrated in [19], where the ML detection complexity does not increase with constellation size. Thirdly, the DSM achieves a throughput of $R = \frac{T \log_2 L + \lfloor T! \rfloor}{T}$, which is higher than that of its SIMO counterparts. However, in the absence of transmit diversity, the DSM does not perform well, especially when a small number N of receive antennas is used.

As a classic single-RF finite-cardinality DSTM scheme, the cyclic DGCs of [126]–[129] construct the data-carrying signal matrix by $\mathbf{X}^l = \mathbf{G}_c^l$, where $\mathbf{G}_c = \text{diag}([w_L^{u_1}, w_L^{u_2}, \cdots, w_L^{u_T}])$ and we have $w_L = \exp(j\frac{2\pi}{L})$. The integer phase-rotation parameters $\mathbf{u} = [u_1, u_2, \cdots, u_T]$ are chosen for the sake of maximizing the diversity product of $\Lambda_p = \min_{l \neq 0} [\prod_{t=1}^T |\sin(\frac{\pi u_t l}{L})|]^{\frac{1}{T}}$. For example, for the MIMO setup of $M = T = 2$ and $L = 8$, the legitimate parameter combinations are $\mathbf{u} = [1, 1]$ and $\mathbf{u} = [1, 3]$, which lead to $\Lambda_p = 0.3827$ and $\Lambda_p = 0.5946$. All the other combinations of $\mathbf{u}$ would either result in the same $\Lambda_p = \{0.3827, 0.5946\}$ or $\Lambda_p = 0$. Therefore, the set $\mathbf{u} = [1, 3]$ is

TABLE 7. Example of DSM ($M = T = 2$) using BPSK.

Input bits	BPSK symbols	Permutation index	Dispersion matrices	Data-carrying matrix
000	$x^{l_1}=1, x^{l_2}=1$	$\bar{m}=1$	$\mathbf{A}_{1,1}=\begin{bmatrix}1&0\\0&0\end{bmatrix}, \mathbf{A}_{1,2}=\begin{bmatrix}0&0\\0&1\end{bmatrix}$	$\mathbf{X}=\sum_{q=1}^{2} x^{l_q}\mathbf{A}_{\bar{m},q}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$
001	$x^{l_1}=1, x^{l_2}=1$	$\bar{m}=2$	$\mathbf{A}_{2,1}=\begin{bmatrix}0&1\\0&0\end{bmatrix}, \mathbf{A}_{2,2}=\begin{bmatrix}0&0\\1&0\end{bmatrix}$	$\mathbf{X}=\sum_{q=1}^{2} x^{l_q}\mathbf{A}_{\bar{m},q}=\begin{bmatrix}0&1\\1&0\end{bmatrix}$
010	$x^{l_1}=1, x^{l_2}=-1$	$\bar{m}=1$	$\mathbf{A}_{1,1}=\begin{bmatrix}1&0\\0&0\end{bmatrix}, \mathbf{A}_{1,2}=\begin{bmatrix}0&0\\0&1\end{bmatrix}$	$\mathbf{X}=\sum_{q=1}^{2} x^{l_q}\mathbf{A}_{\bar{m},q}=\begin{bmatrix}1&0\\0&-1\end{bmatrix}$
011	$x^{l_1}=1, x^{l_2}=-1$	$\bar{m}=2$	$\mathbf{A}_{2,1}=\begin{bmatrix}0&1\\0&0\end{bmatrix}, \mathbf{A}_{2,2}=\begin{bmatrix}0&0\\1&0\end{bmatrix}$	$\mathbf{X}=\sum_{q=1}^{2} x^{l_q}\mathbf{A}_{\bar{m},q}=\begin{bmatrix}0&1\\-1&0\end{bmatrix}$
100	$x^{l_1}=-1, x^{l_2}=1$	$\bar{m}=1$	$\mathbf{A}_{1,1}=\begin{bmatrix}1&0\\0&0\end{bmatrix}, \mathbf{A}_{1,2}=\begin{bmatrix}0&0\\0&1\end{bmatrix}$	$\mathbf{X}=\sum_{q=1}^{2} x^{l_q}\mathbf{A}_{\bar{m},q}=\begin{bmatrix}-1&0\\0&1\end{bmatrix}$
101	$x^{l_1}=-1, x^{l_2}=1$	$\bar{m}=2$	$\mathbf{A}_{2,1}=\begin{bmatrix}0&1\\0&0\end{bmatrix}, \mathbf{A}_{2,2}=\begin{bmatrix}0&0\\1&0\end{bmatrix}$	$\mathbf{X}=\sum_{q=1}^{2} x^{l_q}\mathbf{A}_{\bar{m},q}=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$
110	$x^{l_1}=-1, x^{l_2}=-1$	$\bar{m}=1$	$\mathbf{A}_{1,1}=\begin{bmatrix}1&0\\0&0\end{bmatrix}, \mathbf{A}_{1,2}=\begin{bmatrix}0&0\\0&1\end{bmatrix}$	$\mathbf{X}=\sum_{q=1}^{2} x^{l_q}\mathbf{A}_{\bar{m},q}=\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$
111	$x^{l_1}=-1, x^{l_2}=-1$	$\bar{m}=2$	$\mathbf{A}_{2,1}=\begin{bmatrix}0&1\\0&0\end{bmatrix}, \mathbf{A}_{2,2}=\begin{bmatrix}0&0\\1&0\end{bmatrix}$	$\mathbf{X}=\sum_{q=1}^{2} x^{l_q}\mathbf{A}_{\bar{m},q}=\begin{bmatrix}0&-1\\-1&0\end{bmatrix}$

TABLE 8. Example of DGC-cyclic ($M = T = 2$, $L = 8$) at $R = 1.5$.

$$\mathbf{X}^0=\begin{bmatrix}1&0\\0&1\end{bmatrix}, \quad \mathbf{X}^1=\begin{bmatrix}w_8&0\\0&w_8^3\end{bmatrix}, \quad \mathbf{X}^2=\begin{bmatrix}w_8^2&0\\0&w_8^6\end{bmatrix}, \quad \mathbf{X}^3=\begin{bmatrix}w_8^3&0\\0&w_8^1\end{bmatrix},$$
$$\mathbf{X}^4=\begin{bmatrix}w_8^4&0\\0&w_8^4\end{bmatrix}, \quad \mathbf{X}^5=\begin{bmatrix}w_8^5&0\\0&w_8^7\end{bmatrix}, \quad \mathbf{X}^6=\begin{bmatrix}w_8^6&0\\0&w_8^2\end{bmatrix}, \quad \mathbf{X}^7=\begin{bmatrix}w_8^7&0\\0&w_8^5\end{bmatrix}$$

chosen, and the resultant signal matrices are given by $\{\mathbf{X}^l = \text{diag}([w_8^l, w_8^{3l}])\}_{l=0}^{7}$, which are detailed in Table 8. Despite the careful optimization of the parameters for different MIMO setups, owing to the non-linear mapping, the DGC detection complexity increases exponentially with the throughput R. As a remedy, the polynomial-complexity lattice decoding regime was invoked for DGC in [264], which however may impose a performance loss.

In order to improve the performance of DSM without unduly increasing its transceiver complexity, the infinite-cardinality DSTBCs of [113]–[115] were converted to the finite-cardinality design of DSTBC-ISK [19], which retains both the diversity gains and a low receiver complexity. More explicitly, as exemplified by Table 9, Alamouti's G2 structure of $G_2([x_1, x_2]) = \begin{bmatrix}x_1 & x_2\\ -x_2^* & x_1^*\end{bmatrix}$ is revised to be either $G_2([x, 0]) = \begin{bmatrix}x & 0\\ 0 & x^*\end{bmatrix}$ or $G_2([0, x]) = \begin{bmatrix}0 & x\\ -x^* & 0\end{bmatrix}$, where $\log_2 Q = 1$ bit is assigned to the ISK position, while $\log_2 L$ bits are assigned to modulate the classic PSK or star QAM symbol x. For $M = T > 2$, the QO-STBC signal structure is invoked as $X = G_T^{QO}(\mathbf{x})$, where only a single non-zero PSK or star QAM symbol is modulated in the Q-element signal vector $\mathbf{x}$. As a result, first of all, the finite-cardinality design principle is satisfied. Secondly, owing to the sparse signal vector $\mathbf{x}$, even the QO signal structure becomes linear and unitary, which results in low-complexity single-stream based ML detection. Thirdly, the STBC's beneficial diversity gain is retained at the cost of a reduced throughput of $R = \frac{\log_2 LT}{T}$.

In order to further improve the diversity/throughput trade-off, the DSTSK-TAST scheme was conceived in [20] based on both the DSTSK [123]–[125] and TAST [120] signal structures. The DSTSK-TAST constructs the signal matrix as $\mathbf{X} = x^l \mathbf{A}_{\tau,q}$, where the dispersion matrices are given by

TABLE 9. Example of DSTBC-ISK ($M = T = 2$) using QPSK.

Input bits	ISK index	QPSK symbol	Data-carrying matrix
000	$q = 1$	$x^0 = 1$	$\mathbf{X}=G_2([1,0])=\begin{bmatrix}1&0\\0&1\end{bmatrix}$
001	$q = 1$	$x^1 = j$	$\mathbf{X}=G_2([j,0])=\begin{bmatrix}j&0\\0&-j\end{bmatrix}$
010	$q = 1$	$x^2 = -j$	$\mathbf{X}=G_2([-j,0])=\begin{bmatrix}-j&0\\0&j\end{bmatrix}$
011	$q = 1$	$x^3 = -1$	$\mathbf{X}=G_2([-1,0])=\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$
100	$q = 2$	$x^0 = 1$	$\mathbf{X}=G_2([0,1])=\begin{bmatrix}0&1\\-1&0\end{bmatrix}$
101	$q = 2$	$x^1 = j$	$\mathbf{X}=G_2([0,j])=\begin{bmatrix}0&j\\j&0\end{bmatrix}$
110	$q = 2$	$x^2 = -j$	$\mathbf{X}=G_2([0,-j])=\begin{bmatrix}0&-j\\-j&0\end{bmatrix}$
111	$q = 2$	$x^3 = -1$	$\mathbf{X}=G_2([0,-1])=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$
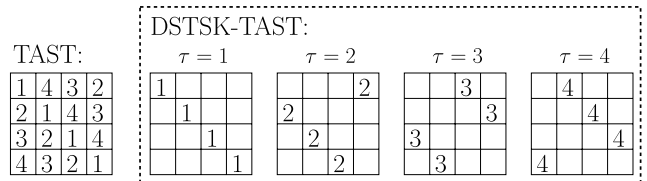
FIGURE 20. Examples of the layer-switching index of the DSTSK-TAST signal matrix associated with $T = 4$.

$$\{\{\mathbf{A}_{\tau,q} = \phi_r^{\tau-1}\text{diag}([w_{L_{DM}}^{(q-1)u_1}, \cdots, w_{L_{DM}}^{(q-1)u_T}])\mathbf{G}_r^{\tau-1}\}_{\tau=1}^T\}_{q=1}^Q.$$

More explicitly, the so-called layer-switching matrix is given by $\mathbf{G}_r = \begin{bmatrix}0 & \cdots & 0 & 1\\ 1 & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & 1 & 0\end{bmatrix}$ and the associated phase rotation is given by $\{\phi_r^{\tau-1} = w_{L_r}^{\tau-1} = \exp(j\frac{2\pi(\tau-1)}{L_r})\}_{\tau=1}^T$. As a result, the TAST arrangement partitions the $(T \times T)$-element signal space into T non-overlapping layers, where the layer-switching index τ carries $\log_2 T$ information bits. Moreover, the integer parameters $\{u_t\}_{t=1}^T$ are introduced to rotate the phases of the dispersion elements $\{\{w_{L_{DM}}^{(q-1)u_t} = \exp(j\frac{2\pi(q-1)u_t}{L_{DM}})\}_{t=1}^T\}_{q=1}^Q$, so that the diversity gains are actively maximized for different

TABLE 10. Example of DSTSK-TAST($M = T = 2$) using $Q = 2$, $L = 2$, $L_{DM} = 4$, $L_r = 4$ and $\mathbf{u} = [1, 3]$.

Input bits	BPSK symbol	Index τ	Index q	Dispersion matrix	Data-carrying matrix
000	$x^0 = 1$	$\tau = 1$	$q = 1$	$\mathbf{A}_{1,1} = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$	$\mathbf{X} = x^0 \mathbf{A}_{1,1} = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$
001	$x^0 = 1$	$\tau = 1$	$q = 2$	$\mathbf{A}_{1,2} = \begin{matrix} j & 0 \\ 0 & -j \end{matrix}$	$\mathbf{X} = x^0 \mathbf{A}_{1,2} = \begin{matrix} j & 0 \\ 0 & -j \end{matrix}$
010	$x^0 = 1$	$\tau = 2$	$q = 1$	$\mathbf{A}_{2,1} = \begin{matrix} 0 & j \\ j & 0 \end{matrix}$	$\mathbf{X} = x^0 \mathbf{A}_{2,1} = \begin{matrix} 0 & j \\ j & 0 \end{matrix}$
011	$x^0 = 1$	$\tau = 2$	$q = 2$	$\mathbf{A}_{2,2} = \begin{matrix} 0 & -1 \\ 1 & 0 \end{matrix}$	$\mathbf{X} = x^0 \mathbf{A}_{2,2} = \begin{matrix} 0 & -1 \\ 1 & 0 \end{matrix}$
100	$x^1 = -1$	$\tau = 1$	$q = 1$	$\mathbf{A}_{1,1} = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$	$\mathbf{X} = x^1 \mathbf{A}_{1,1} = \begin{matrix} -1 & 0 \\ 0 & -1 \end{matrix}$
101	$x^1 = -1$	$\tau = 1$	$q = 2$	$\mathbf{A}_{1,2} = \begin{matrix} j & 0 \\ 0 & -j \end{matrix}$	$\mathbf{X} = x^1 \mathbf{A}_{1,2} = \begin{matrix} -j & 0 \\ 0 & j \end{matrix}$
110	$x^1 = -1$	$\tau = 2$	$q = 1$	$\mathbf{A}_{2,1} = \begin{matrix} 0 & j \\ j & 0 \end{matrix}$	$\mathbf{X} = x^1 \mathbf{A}_{2,1} = \begin{matrix} 0 & -j \\ -j & 0 \end{matrix}$
111	$x^1 = -1$	$\tau = 2$	$q = 2$	$\mathbf{A}_{2,2} = \begin{matrix} 0 & -1 \\ 1 & 0 \end{matrix}$	$\mathbf{X} = x^1 \mathbf{A}_{2,2} = \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}$

MIMO setups. The activation index q carries an additional $\log_2 Q$ source information bits. The resultant DSTSK-TAST signal matrices are exemplified in Table 10. In summary, owing to the phase rotations and the sparse matrix design, the DSTSK-TAST maintains the attractive finite-cardinality low-PAPR signal transmission features. Secondly, thanks to the DSTSK signal structure, the single-stream ML detection complexity is retained. Thirdly, the diversity gains are always actively maximized by manipulating the parameters of $\{u_t\}_{t=1}^{T}$. We note that a DSTSK-TAST scheme is uniquely specified by the integer parameters of $(Q, L, L_{DM}, L_r, \mathbf{u} = [u_1, \cdots, u_T])$, where the transceiver no longer has to store the complex-valued signal matrices. Moreover, the search space for $\{u_t\}_{t=1}^{T}$ is strictly finite, which implies that the full-search for the globally optimal diversity gain is feasible at low throughputs, while random search may still be performed at high throughputs, since the optimization is no longer sensitive to the values of $\{u_t\}_{t=1}^{T}$ at high throughputs.

In order to quantify the performance of the recently-developed finite-cardinality DSTM schemes, first of all, it is demonstrated by Fig. 21 that compared to the best-performing cyclic DGC [126]–[129], DSTBC-ISK and DSTSK-TAST achieve higher diversity gains at low and high throughputs, respectively. Secondly, our complexity comparisons are offered in Fig. 22, which evidences that all the three schemes of DSM, DSTBC-ISK, DSTSK-TAST may employ the single-stream-based ML detector that exhibits the same level of complexity as the DPSK and star QAM detectors. Compared to the matrix-based DGC detection, the complexity reduction achieved by employing the single-stream ML detector become even more substantial as the throughput grows, as demonstrated by Fig. 22. Finally, DSTSK-TAST and DSTBC-ISK are compared to DSM for $M = T = \{2, 4\}$ and $N = \{1, 2, 4, 8, 16\}$ both at low and high throughputs of $R = \{2.0, 5.0\}$ in Figs. 23(a)-(d), which once again confirms that DSTSK-TAST is capable of achieving a persistent performance advantage over a wide range of MIMO throughputs and setups.
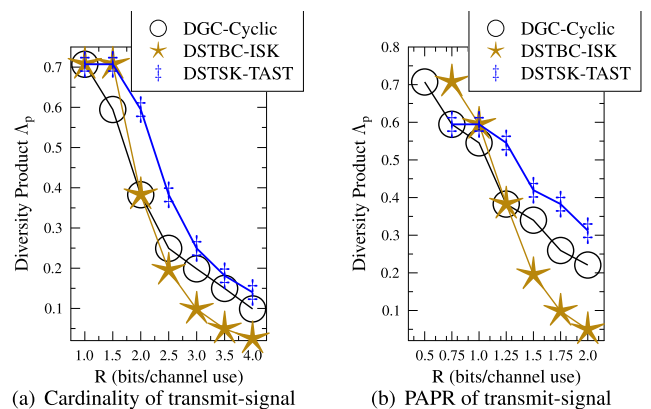
FIGURE 21. Comparison of diversity products of finite-cardinality DSTM schemes.

C. RECTANGULAR DIFFERENTIAL SPACE-TIME CODING AND ITS APPLICATIONS IN MILLIMETER WAVE

The family of DSTM schems introduced in Sections IV-A and IV-B is capable of reducing the detection complexity, avoiding the infinite cardinality problem, whilst maximizing the achievable diversity gain. However, these solutions have to satisfy the ($M = T$) constraint, which results in a square-shaped unitary matrix $\mathbf{X}$. Thanks to this unitary constraint, the low-complexity non-coherent CDD performs well, despite its low overhead, which is negligible compared to that of its coherent counterpart.

The square-matrix-based DSTM schemes incur two major issues, namely a reduced throughput and the absence of beamforming. Firstly of all, as demonstrated in Table 6, the throughput definition of $R = \frac{\log_2 I}{T}$ contains $T = M$ in its denominator. This implies that the effective throughput decreases upon increasing the number of transmit antennas M. Secondly, due to the absence of CSI knowledge, neither transmit nor receive beamforming can be performed both at transmitter and receiver. As a result, the conventional DSTM schemes are not suitable for large-scale or millimeter wave MIMO scenarios.
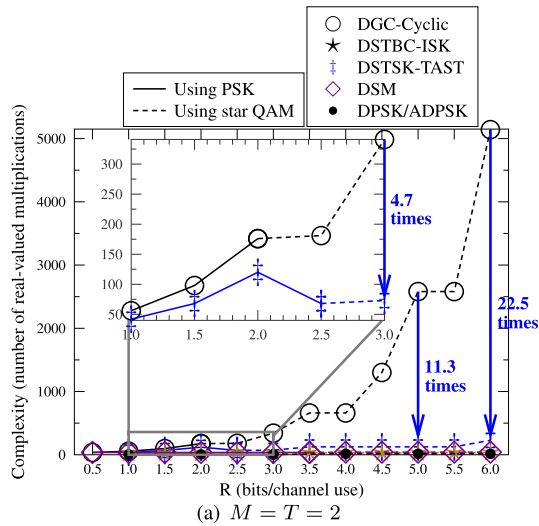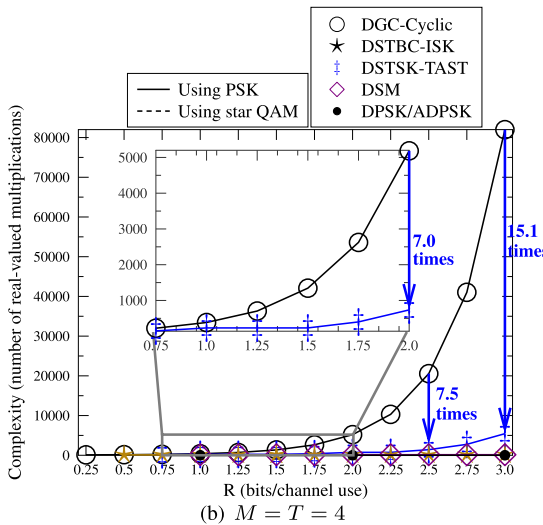
FIGURE 22. Complexity comparison for finite-cardinality DSTM schemes associated with $M = T = \{2, 4\}$ and $N = 1$, where the DSTSK-TAST's complexity reduction compared to DGC-cyclic are marked by blue arrows.

In order to tackle these two challenges, Ishikawa and Sugiura proposed a simple mapping method in 2017 [134], which converts a square-matrix-based DSTM scheme into a non-square-matrix-based DSTM arrangement. As shown in Fig. 24, this method multiplies $\mathbf{S}_n$ by an $M \times T$ basis $\mathbf{E}_1$ and generates an $M \times T$ matrix $\mathbf{S}_n\mathbf{E}_1$. The basis $\mathbf{E}_1$ can be constructed by a unitary matrix [135] such as the DFT matrix of

$$[\mathbf{E}_1 \ \mathbf{E}_2 \ \cdots \mathbf{E}_{M/T}]$$

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

(1)

where we have $\omega = \exp(-2\pi j/M)$. The CDD reference symbol of the identity matrix is replaced by the basis set $[\mathbf{E}_1 \ \mathbf{E}_2 \ \cdots \mathbf{E}_{M/T}]$.

Let us consider a specific example for the DSTSK-TAST scheme exemplified in Table 10, where we have $M = 2$ and $R = 1.5$. The basis $\mathbf{E}_1$ is defined by $[1 \ 1]^T$. As a result, the data-carrying vectors are generated by $\mathbf{XE}_1$ as follows:

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

which are similar to the QPSK-aided BLAST transmission. Finally, the transmission rate is increased from $R = 1.5$ to $R' = 1.5 \cdot M/T = 3.0$.

The most prominent advantage of the non-square DSTM is its increased throughput. This (M/T)-fold improvement is crucial for high-rate communications. Additionally, the performance of non-square DSTM was investigated in millimeter wave scenarios in [136], where only analog BF was considered instead of the CSI-aided hybrid BF. It was demonstrated in [136] that the non-square DSTM was capable of tracking high-mobility millimeter wave channel variation, while reducing the detection complexity and satisfying the finite-cardinality criterion.

D. SUMMARY ON THE HISTORICAL PERSPECTIVE
This section is summarized by a list of major contributions on non-coherent MIMO detection in Tables 11-12.

V. THE BENEFITS OF COHERENT/ NON-COHERENT ADAPTIVITY
In this section, we investigate the benefits of coherent/non-coherent adaptivity from three different perspective, including their bandwidth-efficiency in Sec. V-A, their power-efficiency in Sec. V-B as well as their coverage gain in Sec. V-C.

A. BANDWIDTH-EFFICIENCY
According to the classic Shannon-Hartley law, the channel capacity is given by $C = B \cdot \mathcal{I}(X; Y)$ (b/s), where B (Hz) denotes the channel bandwidth, while the mutual information $\mathcal{I}(X; Y)$ (b/s/Hz) is maximized for Gaussian-distributed continuous-input and continuous-output variables X and Y, respectively. When the discrete-valued input variable X is considered in realistic systems, the Discrete-input Continuous-output Memoryless Channel (DCMC) capacity [23], [91], [282] plotted in Fig. 25 is encountered. The DCMC capacity portrayed in Fig. 25 is bounded by the so-called effective throughput R^e. For the non-coherent scheme that eliminates the pilot overhead, we have $R^e = R_m$, where R_m represents the throughput. By contrast, the effective throughput of the coherent scheme relying on pilot-based CSI estimation is given by $R^e = (1 - f_p)R_m$, where f_p denotes the pilot percentage. According to the Nyquist Theorem, the CSI estimation has to satisfy $f_p \geq 2f_d$, which naturally
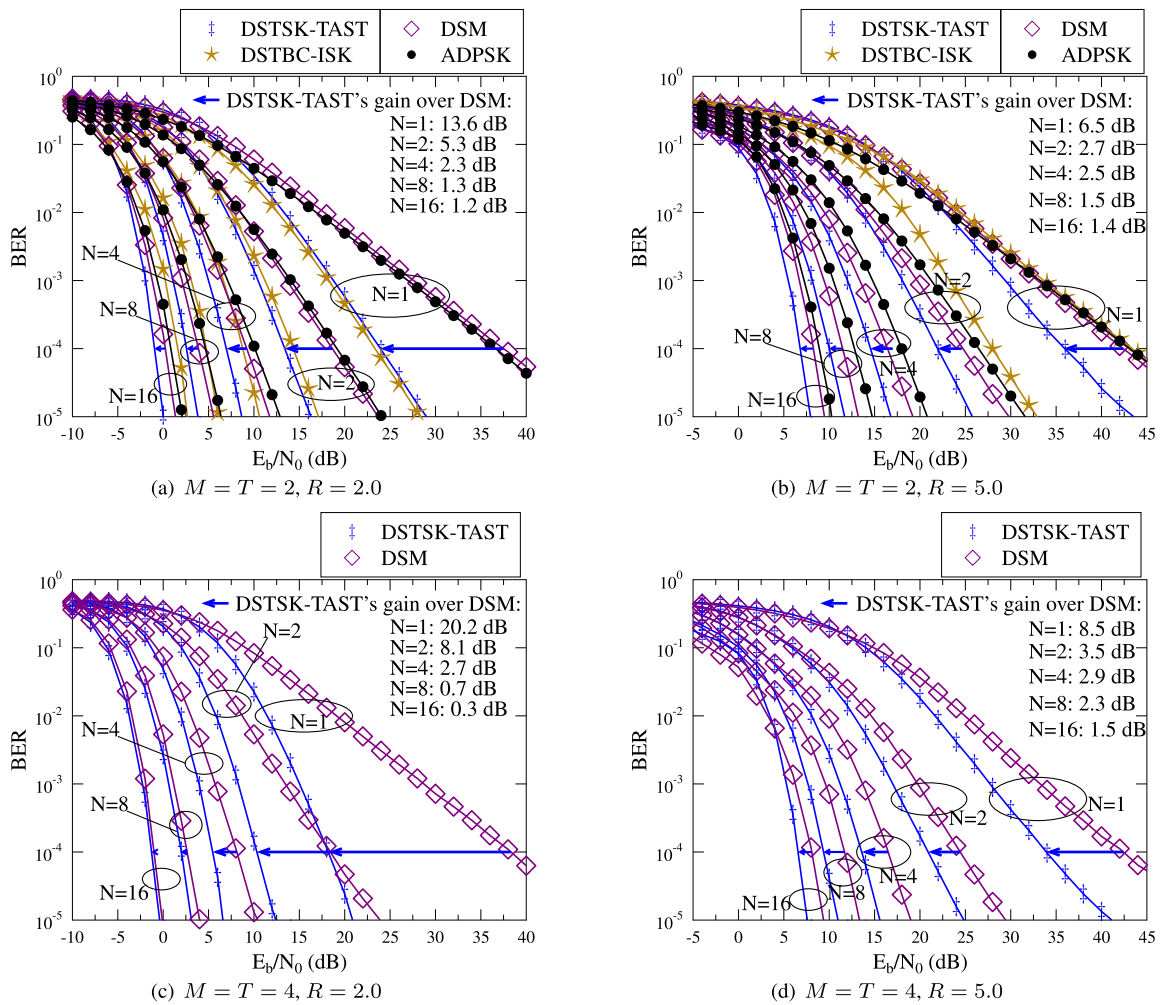
FIGURE 23. Performance comparison between the DSTSK-TAST, DSTBC-ISK and DSM arrangements for $M = T = \{2, 4\}$ and $N = \{1, 2, 4, 8, 16\}$ at $R = \{2.0, 5.0\}$ (adapted from [20]). The diversity gains of DSTSK-TAST over DSM at BER = 10^{-4} are marked by blue arrows and also listed in figures.
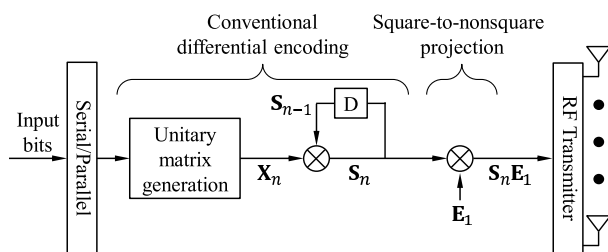
FIGURE 24. Schematic of the non-square DSTM transmitter.

requires more pilots as f_d grows. Furthermore, it was shown in [21], [283] that f_p has to be substantially higher than $2f_d$ for the sake of achieving a satisfactory performance, because the pilot-symbols used for sampling the complex-valued channel envelope are contaminated by the receiver's noise. Therefore, the pilot percentages of $f_p = 0.05$ and $f_p = 0.1$ are chosen for the low-mobility and high-mobility scenarios of $f_d = 0.001$ and $f_d = 0.03$, respectively, as presented in Fig. 25.

As a result, Fig. 25 evidences that due to the pilot overhead and CSI estimation error, the coherent scheme suffers from a capacity loss, which becomes more substantial as f_d

increases. By contrast, the noncoherent scheme is capable of achieving a higher capacity in the high-mobility scenarios, as demonstrated by Fig. 25. This implies that in order to achieve a target data rate in a realistic wireless network, the coherent scheme requires a higher bandwidth than that required by its non-coherent counterpart, so that the capacity loss of using the pilot overhead may be compensated. Let us consider an Unmanned Aerial Vehicle (UAV) example, where the 3GPP recommendation for its Command & Control (C&C) data rate is given by 100 kb/s [64], [112]. For the noncoherent scheme of DQPSK $R^e = R_m = 2$ b/s/Hz associated with a half-rate channel coding scheme of $R_c = 0.5$, a clean single-user link bandwidth of $B = 100$ kHz is sufficient to support the target data rate, when neglecting Nyquist-filtering. By contrast, in order to compensate for the capacity loss as demonstrated by Fig. 25, the coherent QPSK schemes require the increased bandwidth of $B = 100/(1-f_p)$ kHz, which are given by $B = 105.3$ kHz and $B = 111.1$ kHz for the cases of using $f_p = 0.05$ and $f_p = 0.1$, respectively.

Fig. 26 further investigates the tradeoffs between the single-TA DPSK as well as the multiple-TA DSM and

TABLE 11. Summary of major contributions on non-coherent MIMO detection (Part I). The acronyms of FC, SS and DG refer to the design features of finite-cardinality, single-stream-based ML detection and diversity gain.

Year	Author(s)	Topic	Contribution
1999	Marzetta *et al.* [270]	Non-coherent MIMO	Without invoking differential encoding on the MIMO signals, the authors proposed to generate the $(T \times M)$-element transmitted signal matrix by a $(T \times T)$-element isotropically distributed unitary matrix multiplied by a $(T \times M)$-element diagonal, real and non-negative random matrix. The authors showed that the resultant non-coherent MIMO capacity approaches to the shannon capacity with perfect channel knowledge as T grows to infinity.
2000	Hochwald and Marzetta [271]	Non-coherent MIMO	Proposed the so-called Unitary Space-Time Modulation (USTM) based on the theory of [270], where the $(T \times M)$-element transmitted signal matrix are designed to have orthogonal columns. The non-coherent detection is operated by a simple matrix-based correlation, which performs well when using $T >> M$ or at high SNR.
2000	Hochwald *et al.* [272]	Non-coherent MIMO	Proposed systematic design of the USTM constellations, where the $(T \times T)$-element isotropically distributed unitary matrix is formed by a diagonal signal matrix using L-PSK signalling. In summary, USTM eliminates differential encoding, but a substantial throughput loss is also induced owing to the requirement of $T > M$.
2000	Tarokh and Jafarkhani [113]	Full-RF DSTM(SS,DG)	Proposed Differential Space-Time Block Code (DSTBC) based on Alamouti's scheme [209]. DSTBC retains both the full STBC's diversity gain and the single-stream-based ML detection complexity. However, due to the matrix-based differential encoding, DSTBC suffers from the infinite-cardinality problem, where the cardinality of transmitted signals grows infinite over time. Meanwhile, the PAPR is also substantially increased.
2000	Hughes [126], Hochwald Sweldens [127]	Single-RF DSTM(FC,DG)	Proposed Differential Group Code (DGC) as the MIMO extension of the DPSK scheme. The DGC's diagonal or anti-diagonal signal matrices form a finite group under multiplication, where the signal elements are given by exponentiation functions of an L-PSK symbol. As a result, there is no constellation extension and the same L-PSK constellation is always retained for transmitted signals. Meanwhile, the diversity gain is aimed to be maximized but the detection complexity grows exponentially with the throughput.
2001	Jafarkhani [114]	Full-RF DSTM(SS,DG)	Extended the DSTBC using $M = 2$ TAs of [113] to the case of $M = 4$ TAs. However, the orthogonality can only be applied to real-valued constellation in this work.
2001	Shokrollahi *et al.* [129]	Single-/Reduced-/Full-RF DSTM (DG)	Proposed the family of DGCs based on the fixed-point-free group theory, where the PSK cardinality is not limited to a power of a prime number $L = 2^\iota$. A fixed-point-free group can be represented by a group of unitary matrices with full diversity. The authors also conceived a range of Reduced-/Full-RF non-group codes that achieves full diversity but exhibits constellation extensions.
2001	Tao *et al.* [117]	DSTM(SS,DG)	Proposed the power normalization methodology for the DSTBC schemes of [113], [114] using QAM, which was later analysed in [118].
2002	Ganesan and Stoica [115]	Full-RF DSTM(SS,DG)	Proposed DSTBC based on the Amicable Orthogonal (AO) STBC structure [224]–[226], which retains the minimum delay of $T = M = 2^\iota$ for any integer ι, while an improved throughput of $R = \frac{(\iota+1)\log_2 L}{T}$ is achieved.
2002	X. G. Xia [116]	Full-RF DSTM(SS,DG)	Proposed to invoke the star QAM constellation for the DSTBC design of [113].
2002	Hassibi and Hochwald [119]	Full-RF DSTM (DG)	Proposed Differential Linear Dispersion Code (DLDC). The Q number of modulated symbols $\{x_q\}_{q=1}^Q$ are dispersed by space-time hermitian matrices $\{\widetilde{\mathbf{A}}_q\}_{q=1}^Q$ as $\widetilde{\mathbf{X}} = \sum_{q=1}^Q x_q \widetilde{\mathbf{A}}_q$, which is then converted to a unitary matrix by the Cayley transform. This is due to the fact that the summation of unitary matrices does not remain unitary, but the summation of Hermitian matrices always leads to a Hermitian matrix. DLDC suffers from the infinite-cardinality problem, and the associated detection complexity grows exponentially with the throughput.
2002	Schober *et al.* [273]	DFDD	Proposed the hard-decision DFDD conceived for the generic DSTM over Rayleigh fading channels.
2002	Liang and Xia [259]	Full-RF DSTM (DG)	Defined the performance metrics of diversity sum and diversity product, which are the Euclidean and determinant metrics that determine the error performance in low and high SNR regions, respectively. Based on these criteria, the authors also proposed a new parametric code that achieves good performance but suffers from the infinite-cardinality problem and a detection complexity that grows exponentially with the throughput.
2003	Hughes [128]	Single-RF DSTM(FC,DG)	Demonstrated that for the PSK cardinality being a power of a prime number $L = 2^\iota$, all full-diversity DGCs are either cylic or dicyclic, which invoke the diagonal or anti-diagonal signal matrices, respectively.
2007	Pauli *et al.* [274]	MSDSD	Proposed the hard-decision MSDSD conceived for the generic DSTM over Rayleigh fading channels.
2007	Yuen *et al.* [275]	Full-RF DSTM (DG)	Proposed the DSTBC relying on the Qusi-Orthogonal (QO) STBC [227]–[233] structure, where joint mapping was conceived in order to form unitary signal matrices.
2009	Bhatnagar *et al.* [267]	Full-RF DSTM (DG)	Demonstrated that non-unitary MIMO schemes - including QO-STBCs [227]–[233], generic LDCs [120], [235], [236], [238], [268], [269], Golden code [239] and perfect STBCs [240]–[243]- may also be invoked by the matrix-based differential encoding in conjunction with appropriate power normalization. However, the bespoke non-coherent detection has to deal with high-complexity evaluations of matrix inversions.
2010	Sugiura *et al.* [123]	Full-RF DSTM(SS,DG)	Proposed Differential Space-Time Shift Keying (DSTSK), where only a single DLDC's dispersion matrix in [119] is activated for dispersing a single modulated symbol.
2011	Xu *et al.* [125]	MSDSD	Eliminated DLDC's Cayley transform [119] in DSTSK [123] and conceived reduced-complexity hard-decision MSDSD solutions for DSTSK.
2011	Xu *et al.* [276]	MSDSD	Proposed hard-decision MSDSD for DSTBC using QAM, where the variable channel correlation matrix is formed by hypothesized combinations of transmitted symbols.
2013	Bian *et al.* [130]	Single-RF DSTM(FC,SS)	Proposed Differential Spatial Modulation (DSM), which was further generalized in [131]. DSM modulates a total number of M L-PSK symbols $\{x^{l_q}\}_{q=1}^M$ by $M \log_2 L$ bits, and their permutation locations in the signal matrix is determined by $\lfloor \log_2 M! \rfloor$ bits. As a result, DSM achieves a higher throughput than DPSK without encountering the infinite-cardinality problem, but it does not perform well when N is small. Moreover, its single-stream ML detection is devised in [19], which does not impose any performance loss.

TABLE 12. Summary of major contributions on non-coherent MIMO detection (Part II). The acronyms of FC, SS and DG refer to the design features of finite-cardinality, single-stream-based ML detection and diversity gain.

Year	Author(s)	Topic	Contribution
2014	Ishikawa *et al.* [265]	Single-RF DSTM(SS,DG)	Introduced a transmit diversity gain to DSM [130], [131] by designing sparse dispersion matrices for the DSTSK of [123].
2015	Martin [132]	Single-RF DSTM(FC,SS)	Proposed to invoke star QAM constellations for the DSM scheme [130], [131].
2017	Rajashekar *et al.* [266]	Single-RF DSTM(SS,DG)	Proposed Field Extension based DSM (FE-DSM) as a diversity-aided DSM scheme. FE-DSM was refined in [277] in order to generate a finite-cardinality signal set, which became a special case of the DGC $G_{m,r}$ of [129] associated with $r = 1$.
2017	Ishikawa *et al.* [134]	Single-RF DSTM(DG)	Proposed a rectangular DSM scheme for open-loop massive MIMO scenarios, where the $M = 1024$ TAs was considered.
2018	Xu *et al.* [19]	Single-RF DSTM (FC,SS,DG)	Proposed DSTBC using Index Shift Keying (DSTBC-ISK), which transformed the infinite-cardinality DSTBCs of [113]–[115] to finite-cardinality sets by activating only a single modulated symbol in the STBC signal structure. Both the transmit diversity gains and the low-complexity receiver are retained. This scheme is the first new full-diversity group code that is beyond the family of fixed-point-free group codes summarized in [128], [129].
2018	Ishikawa *et al.* [135]	Reduced-RF DSTM(FC,DG)	Generalized the rectangular concept of [134] to support an arbitrary number of activated RF chains and analyzed the effects of forgetting factor design.
2019	Xu *et al.* [21]	MSDSD/DFDD	Proposed hard-decision and soft-decision aided MSDSD and DFDD for the generic DSTM transmitting over Ricean channels. This scheme improved the existing solutions of [278]–[281] by taking into account the following high-mobility aeronautical Ricean features: (1) high frequency offset on the strong LOS; (2) high Doppler frequency; (3) a random channel phase rotation; (4) Angle of Departure (AoD) and Angle of Arrival (AoA).
2019	Xu *et al.* [20]	Single-RF DSTM (FC,SS,DG)	Proposed DSTSK using Treaded Algebraic Space-Time (DSTSK-TAST) arrangement. First of all, following the TAST methodology, the $(T \times T)$-element signal space is partitioned into T non-overlapping layers, where the activation of a single layer carries $\log_2 T$ bits. Secondly, following the DSTSK design, the dispersion matrix activation index q and modulated symbol index l further conveys $\log_2 Q$ and $\log_2 L$ bits, respectively. Thirdly, the elements in the sparse dispersion matrices are all drawn from a PSK constellation, and they are actively rotated in order to maximize the diversity gain for different MIMO setups. As a result, as DSTBC-ISK [19] performs better at low throughputs, DSTSK-TAST are shown to achieve a persistent performance advantage over DSM [130], [131] for a wide range of MIMO throughputs and setups.
–	Ishikawa *et al.* [136]	Reduced-RF DSTM(FC)	Proposed a Differential Generalized Spatial Modulation (GSM) scheme for high-mobility millimeter wave communications. The forgetting factor of [134], [135] was automatically adjusted based on received signals.

FIGURE 25. DCMC capacity comparison of QPSK relying on pilot-based CSI estimation and DQPSK employing MSDD over Ricean fading channels ($K = 0$ dB, $f_d = \{0.001, 0.03\}$).

FIGURE 26. DCMC capacity comparison of noncoherent scheme of single-TA DPSK as well as multiple-TA DSM and DSTBC-ISK employing MSDD over Ricean fading channels ($K = 0$ dB, $f_d = 0.03$).

DSTBC-ISK schemes. First of all, Fig. 26(a) evidences that the diversity-oriented DSTBC-ISK scheme performs best at the throughputs of $R^e = R_m = \{1.0, 2.0\}$ b/s/Hz. Secondly, as summarized in Table 6, the DSM scheme achieves a higher throughput, which results in a better performance at the high throughput of $R^e = R_m = 3.0$ b/s/Hz, as evidenced by

Fig. 26(a). Thirdly, when a channel coding scheme of rate R_c is considered, the performance of the resultant coded system is portrayed in Fig. 26(b). More explicitly, it is demonstrated by Fig. 26(b) that the capacity differences between the single-TA and multiple-TA schemes are small, when a strong channel coding scheme associated with $R_c = 0.5$ is

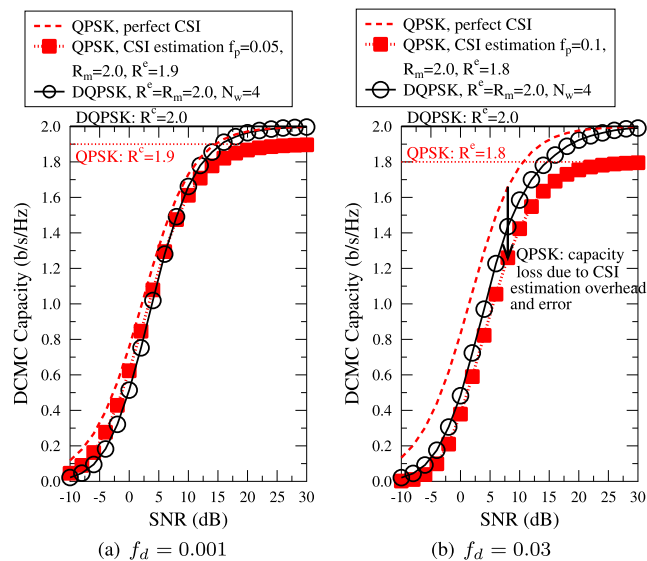**FIGURE 27.** Power-Efficiency comparison of QPSK relying on pilot-based CSI estimation and DQPSK employing MSDD over Ricean fading channels, where we have $R_m = 2.0$ and $R_c = 0.5$.
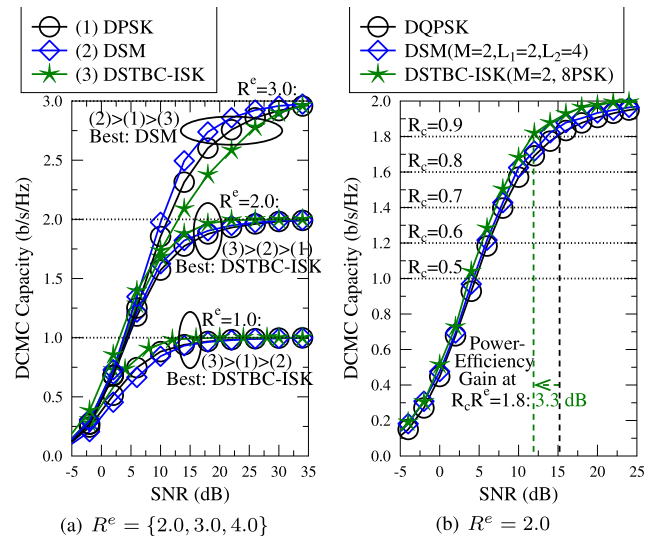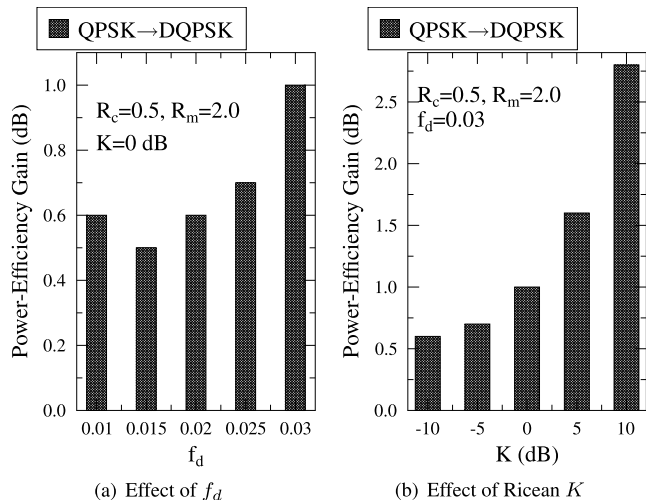

FIGURE 28. Power-Efficiency comparison of noncoherent scheme of single-TA DPSK as well as multiple-TA DSM and DSTBC-ISK employing MSDD over Ricean fading channels ($K = 0$ dB, $f_d = \{0.001, 0.03\}$).

applied. However, the capacity advantage of the multiple-TA DSTBC-ISK over its single-TA DPSK counterpart is substantially improved, as the channel coding rate R_c increases. For example, at $R_c = 0.9$, an SNR difference of 3.3 dB is observed.

B. POWER-EFFICIENCY

Based on the DCMC capacity results of Sec. V-A, we may conclude that it is beneficial to

(1) adaptively switch between coherent and non-coherent schemes based on f_d;

(2) adaptively switch between single-TA and multiple-TA schemes based on R_c;

(3) adaptively switch between high-diversity and high-throughput multiple-TA schemes based on R_m.

This is confirmed by the results of our three-fold adaptivity proposed in [21] based on EXIT charts. In this section, we proceed to quantify the power-efficiency gains achieved by the three-fold adaptivity.

The evaluation of power-efficiency is exemplied by Fig. 26(b), where the SNRs required by the different schemes for achieving a target system throughput of $R_c R^e$ are compared. The power-efficiency gains achieved by switching between coherent QPSK and non-coherent DQPSK schemes are summarized in Fig. 27. It is confirmed by Fig. 27(a) that an improved power-efficiency gain is achieved by invoking DQPSK as f_d increases, where a 1 dB gain is recorded at $f_d = 0.03$. Furthermore, Fig. 27 demonstrates that the power-efficiency gain of invoking DQPSK over QPSK further improves as the Ricean K factor increases. This once again confirms the advantage of switching to the non-coherent scheme in high-mobility scenarios, regardless of the LoS strength.

Fig. 28(a) confirms that the power-efficiency gain of switching from single-TA DPSK to multiple-TA DSTBC-ISK

improves as the channel coding rate R_c increases, where a substantial 3.3 dB gain is recorded at $R_c = 0.9$. Moreover, Fig. 28(b) further demonstrates the power-efficiency gains achieved by invoking the high-diversity DSTBC-ISK and the high-throughput DSM at low and high modulation throughputs, respectively. In summary, the three-fold adaptivity is considered beneficial, especially for the cell-edge users, where the 3.0 dB SNR gain may substantially improve the link quality. Following this, in the next section, we continue to investigate the beneficial impact of the power-efficiency gains on the coverage extension/quality across over the 5G spectrum.

C. COVERAGE GAIN

In order to evaluate the link budget, the generic statistical path loss model may be expressed in dB in the following form [159], [284], [285]:

$$PL(d)[dB] = \alpha + 10\beta \log_{10}(d) + \xi,$$

where d is the distance between the transmitter and the receiver, α and β are linear model parameters, and ξ is a log-normally distributed parameter that takes into account the shadowing effect. This subsumes the free-space isotropic path loss based on the Friis' Law of $PL(d)[dB] = 20 \log_{10}(4\pi/\lambda) + 20 \log_{10}(d)$ as a special case associated with $\alpha = 20 \log_{10}(4\pi/\lambda)$ and $\beta = 2$. Upon considering the situation that both the distance-independent paramters α and ξ are fixed in a link budget design, the 1 dB power-efficiency gain of switching from QPSK to DQPSK at $K = 0$ dB and $f_d = 0.03$ in Fig. 27(a) may be translated into a modest 1.12 meters link-extension gain.

However, let us now elaborate further by considering the link-extension portrayed in Fig. 29, where we assume that the maximum path loss is fixed to 76 dB. This corresponds to a routine setting, where the transmit power and

(a) Coverage
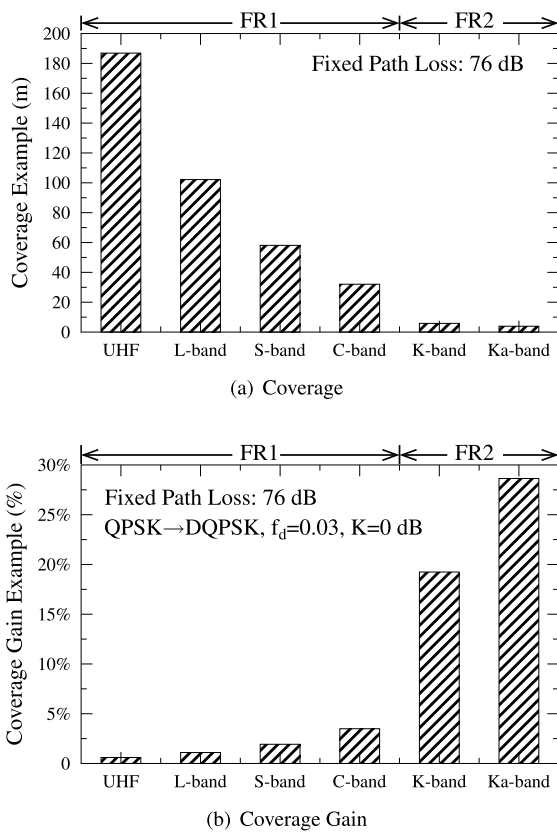
(b) Coverage Gain

FIGURE 29. Example of coverage over 5G spectrum of Fig. 4, where the maximum allowable path loss is fixed to be 76 dB. The coverage gain is evaluated based on the power-efficiency gain of Fig. 27, where we have $K = 0$ dB and $f_d = 0.03$.

the receiver sensitivity are 20 dBm and -104 dBm, respectively, while a 48 dB transmit and receive beamforming gain is provided. However, according to Friis' Law, increasing the carrier frequency by a factor of 10 adds 20 dB extra path loss, regardless of the distance. Moreover, a range of additional factors may contribute to a further reduced link budget at mmWave frequencies. This includes the atmospheric loss, rain loss and foliage loss, which are observed to be 0.06 dB/km, 3.46 dB/km and 11.34 dB/km at the center frequency of 28 GHz, respectively, according to the ITU and FCC recommendations [286]–[288]. Therefore, maintaining the link budget at mmWave frequencies may require a variety of measures including a higher Effective Isotropic Radiated Power (EIRP) allowance at the transmitter, a reduced receiver sensitivity as well as an improved beamforming gain, as discussed in Sec. II. Nonetheless, in order to simplify the coverage comparison in this section, we opt for fixing the path loss to 76 dB over the 5G spectrum, where the additional path loss is assumed to be mitigated by the improved beamforming gains.

As expected, it is demonstrated by Fig. 29(a) that the link coverage decreases as the carrier frequency of the designate spectrum increases. The exemplified center frequencies of microwave FR1 UHF, L-band, S-band and C-band are given

by 0.806 GHz, 1.4745 GHz, 2.593 GHz and 4.7 GHz, respectively, as seen in Fig. 29(a), while the mmWave FR2 K-band and Ka-band examples have 25.875 GHz and 38.5 GHz, respectively. It is also confirmed by Fig. 29(b) that the aforementioned 1.12 meters coverage gain obtained by switching from QPSK to DQPSK constitutes a higher coverage percentage especially in the mmWave FR2 band, which verifies the benefit of the adaptivity in the 5G cellular network.

VI. WHAT WILL 6G ADOPT?

A. TERAHERTZ COMMUNICATIONS

As acutely predicted by Edholm's law [139], the exponentially increasing mobile data rate has been doubled every 18 months over the last 30 years. It is envisioned that the major milestone of 100 Gb/s may be achieved at some stage after 2020 [140], which is right on time to herald the new era of 6G experiments both in academia and in industry. The recent hardware developments in indium phosphide and silicon photonics make high-speed transceiver designs operating at 400 Gb/s a reality [141]. Against this background, the following question arises in this early stage of 6G research: where do we find sufficient bandwidth for accommodating the escalating thirst for data rate?

To answer this question, we once again examine the electromagnetic spectrum in Table 13, where the human health aspects and the availability for wireless communications are of primary concern. First of all, the photon energy seen in Table 13 is evaluated by $E = \frac{hc}{\lambda}$ eV, where $h = 6.62607015 \times 10^{34}$ eV $\cdot s$, $c = 3 \times 10^8$ m/s and λ denote the Planck constant, the speed of light and the wavelength, respectively, while the energy metric eV refers to the electronvolt. Beyond the visible light frequency of 800 THz, harmful ionizing radiation is generally encountered in the Ultraviolet, X-ray and Gamma ray bands, if the photon energy exceeds the 12 eV threshold, beyond which electrons are removed from atoms/molecules [58], [289]. The associated high penetration depth makes these frequencies only suitable for medical applications, as indicated in Table 13, but harmless wireless communications are not feasible. By contrast, the lower frequencies associated with $E < 12$ eV are known to have low penetration depth into human body, where the only major health concern is owing to thermal effects that may be avoided by regulating the maximum transmission power. Furthermore, visible light communication has to mitigate the ubiquitous non-data-carrying interferences from both sunlight and other light-producing devices, while the transmission power is limited by the eye-safety considerations. Therefore, the frequencies below the visible light band above 400 THz remain attractive choices for communication, as highlighted in Table 13.

Secondly, it can be readily seen in Table 13 that the large bandwidth of the terahertz band remains hitherto unclaimed territory for wireless communication networks, which attracts increasing interests for future 6G applications [58]–[60]. Nonetheless, it is well understood that part of the terahertz

TABLE 13. Summary of the electromagnetic spectrum for communications (with highlights in safety and availability).

	Radio	Microwave	millimeter Wave (mmWave)	Terahertz	Visible Light	Beyond Visible Light
Frequency	3 Hz to 0.3 GHz	0.3 to 300 GHz	30 to 300 GHz	0.1 to 30 THz	430 to 750 THz	>750 THz
Wavelength	10^5 km to 1 m	1 m to 1 mm	10 mm to 1 mm	3 mm to 10 μm	700 nm to 400 nm	<400 nm
Photon Energy	12.4 feV to 1.24 μeV	1.24 μeV to 1.24 meV	0.124 meV to 1.24 meV	41.3 meV to 0.124 eV	1.77 eV to 3.1 eV	>3.1 eV
Safety	Non-Ionizing	Non-Ionizing, Thermal-Safety	Non-Ionizing, Thermal-Safety	Non-Ionizing, Thermal-Safety	Non-Ionizing, Eye-Safety	Ionizing Radiation
Devices	Electronic	Electronic	Electronic	Electronic, Photonic	Photonic	Photonic
Usage	Radio, Television, Aeronautical	Terrestrial, Aeronautical, Satellite	Terrestrial, Satellite	Astronomy	Visible Light Communication	Medical Applications
Cellular	1G	2G/3G/4G/5G	5G	Unlicensed	Unlicensed	Unlicensed

band has extraordinary astronomy interests, as 98% of the photons emitted since the Big Bang fall into the far-infrared spectrum of 0.6 to 7.5 THz [290]. A variety of radio telescopes and satellite-based passive earth-monitoring devices operate in the terahertz band, which do not actively transmit signals, but passively receive low-power signals using highly sensitive receivers. The potential conflicts between astronomy and communications at terahertz frequencies may be deemed insignificant, as the astronomy devices are generally located in very remote areas. Nonetheless, in summary, the health issues and regulations of large-scale terahertz communications still require further scrutiny before its potential exploitation for 6G in 10–15 years time.

Following this, Table 14 summarizes the terahertz propagation features and the potential enabling techniques. We note that the boundaries of centimeter Wave (cmWave), mmWave and terahertz bands are not always consistent among regulation bodies such as the ITU and the IEEE, because the specific bandwidth of interest may vary depending on the field of study. In order to investigate the propagation implications both on the link budget and on hardware requirements, five examples are presented in Table 14. First of all, when the Path Loss (PL) is fixed in the link budget, based on the Friis's Law of PL$= 20 \log_{10}(4\pi d/\lambda)$ dB, Example (I) confirms that the coverage distance d decreases as the frequency increases from cmWave to mmWave and terahertz. Similarly, when the distance d is fixed in Example (II), the PL increases quadratically with the frequency. However, the antenna aperture also shrinks as the frequency increases according to $A_e = \frac{G_e \lambda^2}{4\pi}$. As a result, Example (III) demonstrates that when the antenna aperture A_e is fixed at the transmitter, the effective antenna gain $G_e = 10 \log_1 0(4\pi A_e/\lambda^2)$ dB also increases quadratically with the frequency, which effectively compensates for the PL upon increasing the frequency. Moreover, Example (IV) shows that when the antenna gain G_e is fixed, the antenna size decreases dramatically upon increasing the frequency from the cmWave to mmWave and to the terahertz band. Explicitly, at terahertz it is comparable to a sharp pencil. Similarly, when the aperture A_e is fixed in Example (V), the angular resolution of $\Delta_e = \frac{\lambda}{\sqrt{A_e}}$ decreases in line with the wavelength λ from cmWave to mmWave and

to terahertz. Quantitatively, the wavelength of say 3 GHz is 10 cm, of 30 GHz is 1 cm and of 300 GHz is 1 mm.

Despite some similarities shared with the mmWave band, a range of distinct terahertz challenges have to be taken into account by radically new enabling techniques. First of all, the remaining unlicensed bands become scarce in the cmWave and mmWave band, where the available contiguous bandwidth in C-band, V-band and E-band is at best 8.6 GHz or less [291], as presented in Table 14. By contrast, the entire 10 THz bandwidth in the terahertz band are available for use. Secondly, for the frequencies below 0.1 THz, the atmospheric absorption is mainly contributed by the oxygen molecules, which is already much higher in the mmWave band than in the cmWave band. By contrast, the molecule absorption in terahertz is dominated by water vapor, where the water vapor resonant peaks are reported to be over 60 dB higher than those of the oxygen absorption peaks [292]. This implies that only some windows segmented by the absorption peaks are feasible for communication in the terahertz band. Furthermore, the water absorption in the terahertz band is highly dependent on distance. Almost the entire 10 THz band is available for communication distances under 1 meter, where the absorption is negligible [61]. Moreover, it is observed in [61] that within the four bandwidth windows of 0.38 ∼ 0.44 THz, 0.45 ∼ 0.52 THz, 0.62 ∼ 0.72 THz and 0.77 ∼ 0.92 THz, the water absorption is below 10dB/km, which becomes comparable to the atmospheric absorption in the mmWave band [142].

Thirdly, the beamforming in the cmWave 4G LTE band is generally performed at the transmitter side of the downlink, where a wide beam created by the BS may be shared by multiple MS receivers [157]. As the number of antennas increases substantially in the mmWave band, both the transmitter and the receiver are capable of creating narrow beams. Therefore, it becomes impractical to employ the typical full-RF based cmWave setup, where each antenna is connected to a dedicated RF chain. Instead, typically the hybrid architecture that employs analog phase shifters in conjunction with fewer RF chains is routinely assumed in the mmWave band [159]. Furthermore, Table 14 demonstrates that highly directional pencil beamforming is required in the terahertz band. On one hand, the current mmWave analog phase

TABLE 14. Comparison between centimeter Wave (cmWave), millimeter Wave (mmWave) and terahertz.

	cmWave	mmWave	terahertz
Boundaries (of interest)	$f = 3$ GHz~ 30 GHz $\lambda = 10$ cm~ 1 cm	$f = 30$ GHz~ 100 GHz $\lambda = 10$ mm~ 3 mm	$f = 0.1$ THz~ 10 THz $\lambda = 3$ mm$\sim 30\,\mu$m
Example (I) on Distance	Fixed PL=76 dB=$20\log_{10}(\frac{4\pi d}{\lambda})$: $d = 50.2$ m~ 5.02 m	Fixed PL=76 dB=$20\log_{10}(\frac{4\pi d}{\lambda})$: $d = 5.02$ m~ 1.5 m	Fixed PL=76 dB=$20\log_{10}(\frac{4\pi d}{\lambda})$: $d = 1.5$ m~ 0.015 m
Example (II) on PL	Fixed $d = 15$ m: PL= 65.5 dB~ 85.5 dB	Fixed $d = 15$ m: PL= 85.5 dB~ 96 dB	Fixed $d = 15$ m: PL= 96 dB~ 136 dB
Example (III) on Tx Gain	Fixed apertureA_e=80 cm^2=$\frac{G_e\lambda^2}{4\pi}$: $G_e = 10$ dB~ 30 dB	Fixed aperture A_e=80 cm^2=$\frac{G_e\lambda^2}{4\pi}$: $G_e = 30$ dB~ 40.5 dB	Fixed aperture A_e=80 cm^2=$\frac{G_e\lambda^2}{4\pi}$: $G_e = 40.5$ dB~ 80.5 dB
Example (IV) on Aperture	Fixed gain $G_e = 30$ dB: $A_e = 79.6$ dm$^2 \sim 80$ cm^2	Fixed gain $G_e = 30$ dB: $A_e = 80$ cm$^2 \sim 7.16$ cm^2	Fixed gain $G_e = 30$ dB: $A_e = 7.16$ cm$^2 \sim 0.07$ mm^2
Example (V) on Angle	Fixed apertureA_e=80 cm^2: $\Delta_e = \frac{\lambda}{\sqrt{A_e}} = 64^o \sim 6.4^o$	Fixed apertureA_e=80 cm^2: $\Delta_e = \frac{\lambda}{\sqrt{A_e}} = 6.4^o \sim 1.9^o$	Fixed apertureA_e=80 cm^2: $\Delta_e = \frac{\lambda}{\sqrt{A_e}} = 1.9^o \sim 0.02^o$
Available Windows	C-band: 5.725 to 5.875 GHz	V-band: 57 to 64 GHz E-band: 76 to 81 GHz	Entire 10 THz bandwidth subject to distance and absorption windows
Propagation Loss	Dominated by path loss.	Increased atmospheric absorption, scattering and penetration loss.	Similar to mmWave but water absorption becomes dominate.
Beamforming	Tx wide beams in 4G LTE	Tx&Rx narrow beams in 5G NR	Tx&Rx pencil beams
RF Chains	Fully-loaded	Fully-Connected	Array of Subarrays
Fading Characteristics	Rich multipath, frequency-selective	Sparse multipath, frequency-selective	Stronger LoS, sparser multipath, near-flat

shifters are not sufficiently accurate, hence the recently developed graphene-based waveguide [62] constitutes a suitable choice for terahertz beamforming. On the other hand, the so-called subarrays based architecture is more energy-efficient in the terahertz band [60], where each RF chain is only connected to a subset of antennas, hence only driving a fraction of the antennas.

Finally, it is also indicated in Table 14 that the employment of OFDM may no longer be necessary. More explicitly, the terahertz channels are more LoS-dominated and even sparser in multipath than the mmWave channels. The large available channel bandwidth leads to nearly frequency-flat fading associated with limited frequency selectivity, because due to its high path loss no reflected paths occuer. Moreover, from the PA's point of view, the IBO that can be provided by the PA generally decreases, as the carrier frequency increases [58]. This implies that processing the OFDM signals associated with high PAPR may become increasingly challenging in the terahertz band.

In summary, terahertz research is still in its infancy. Considering its short wavelength and high absorption, its indoor application is more likely to emerge. For example, the air in an office environment is mainly constituted by 78.1% nitrogen, 20.9% oxygen and 0.1-10% water vapor [292]. Therefore, the dominant water absorption phenomenon of the terahertz band may be less severe in indoor scenario. Furthermore, considering the large available bandwidth, the normalized Doppler frequency remains small even at relatively high velocity. In this context, we note that the coherent versus non-coherent tradeoff may be addressed in the terahertz band by bearing in mind the following three aspects. Firstly, the terahertz channels change drastically depending on the propagation medium, the distance and the environment. This leaves room for beneficial non-coherent designs that improve the robustness against channel variations. Secondly, as the PA operating at terahertz frequencies tends to reduce the IBO, the low-PAPR finite-cardinality DSTM schemes introduced in Sec. IV-B may become particularly favourable. Thirdly, the rectangular DSTC introduced in Sec. IV-C exhibits an advantageous performance with the aid of mmWave beamforming. Its extension in the terahertz band using the subarray based architecture awaits further exploration.

B. SPACE-AIR-GROUND INTEGRATED NETWORK

The success of 6G is likely to be dependent on three major aspects: new technologies, solid business model and strong market demand. In recent years, numerous infrastrature and commercial initiatives have endeavoured to extend the wireless broadband dividend to a wider range of areas and population. To name a few, Google's Project Loon uses High-Altitude Platform (HAP) balloons in the stratosphere that offer LTE services to remote areas [64]. The "Other 3 Billion" (O3b) initiative deploys Medium Earth Orbit (MEO) satellites to provide coverage for 3 billion people in developing countries [65]. The 3GPP recommends to offer LTE services to Low-Altitude Platform (LAP) UAVs flying up to 300 m above ground, which may act as aerial relays for the sake of extending the cellular coverage [112]. Moreover, billions of people are travelling in airplanes, trains and cruise ships, where the wireless access demand cannot be satified by the terrestrial cellular network alone [63], [66], [67]. All these recent developments inspire new technologies, business models and stimulate the demand for the future 6G, which is envisioned to integrate the space, air and ground networks, as portrayed in Fig. 30.

First of all, the satellite network in space is constituted by three layers of satellites, namely Geostationary Earth Orbit (GEO), Medium Earth Orbit (MEO) and Low Earth

Space	Altitude	GEO: 35786 km, MEO: 2000−35786 km, LEO: 180−2000 km
	Delay (S−G)	GEO: 0.12 s, MEO: 6.7−120 ms, LEO: 0.6−6.7 ms
	Mobility	GEO: stationary, MEO:2−24 hr orbital period, LEO: 2 hr orbital period (28080 km/h)
	Enabler	Backhaul: laser links, Access: dmWave, cmWave, mmWave.
	SWAP	Bulky steerable antennas, expensive links.
Air	Altitude	HAP: Above 18 km in stratosphere, Civil: 9−12km cruising, LAP: below 120m without licensing
	Mobility	Hypersoic UAV: >>1192km/h, Civil:up to 1080km/h cruising, High maneuver dynamics especially for UAV
	Enabler	radio links: ATC control, dmWave: ATC surveillance/data, mmWave: LAP UAV, laser links: HAP
	SWAP	Increasing demand but bandwidth−hungry. Size−limited and power−hungry especially for LAP UAV.
Ground	Mobility	Train: up to 603km/h (Japan 2015) Car: up to 447.19km/h (2017)
	Enabler	4G/5G: dmWave, cmWave, mmWave.
	SWAP	Bandwidth−Hungry due to regulations. Size−limited and power−hungry for handsets.

GEO: Geostationary Earth Obit   MEO: Medium Earth Obit   LEO: Low Earth Obit   S−G: Satellite to Ground
SWAP: Size, Weight And Power   HAP: High Altitude Platform   LAP: Low Altitude Platform   UAV: Unmanned Aerial Vehicle

FIGURE 30. Schematic of space-air-ground integrated network.

Orbit (LEO) satellites, where both the MEO and LEO satellites orbit at high speed, as summarized in Fig. 30. Owing to their high altitude, the delay of a satellite network is dominated by the propagation delay. For example, considering the GEO at the fixed 35786 km altitude, the one-way delay to the Groud Station (GS) is about 120 ms, which is deemed excessive even for interactive conventional telephony, let alone many low-latency applications. By contrast, a MEO satellite at a distance of say 3000 km imposes a propagation delay of 10 ms, while the LEO-satellite's delay may be below 1 ms. Generally speaking, the overall commmunication latency has to take into account all the transmission, propagation, processing and queuing delays. In the end, the human eyes can perceive visual delays below 10 ms, which would perturb lip-synchronized video telephony. The touch interations of modern smart devices require a more rigorous 1 ms latency [293]. Therefore, the satellite network is more suitable for offloading delay-tolerant but bandwidth-hungry tele-traffics. Communications between satellites are often supported by Free Space Optical (FSO) laser links, which can rely on wide unlicensed bandwidth, as seen in Table 13, but the access technologies for the space-ground communication often rely on cmWave and mmWave carriers, since the laser beams cannot propagate through cloud, fog and rain in the earth's atmosphere.

The past few years have witnessed a soaring demand for improving air-ground communication. Specifically, over a million commercial UAVs had been sold by 2015, which substantially outnumbered the civil and military applications. As seen in Fig. 30, the HAP UAVs and civil aviations generally travel at much higher speed than ships and cars. This imposes substantial challenges, when invoking conventional terrestrial techniques for air-ground communications. Moreover, the critical air-ground link may become blocked by the body of aircraft during maneuvers, which imposes grave airframe fading [294]. This effect is especially detrimental to the UAVs that do not have humans on-board, which are often deployed to perform more dramatic maneuvers for mission-critical applications [64]. Another major limitation of the LAP UAV is its limited Size, Weight And Power (SWAP), as explicitly indicated in Fig. 30. Moreover, as the terrestrial network continues to evolve towards 6G, the ever-growing beyond-100Gb/s data rate, the ubiquitous coverage required by rural and remote areas and the seemless connections to high-speed vehicles are also identified as major challenges for terrestrial communications in Fig. 30.

Against this background, the coherent versus non-coherent tradeoff may play an increasingly important role in the space-air-ground integrated network in the context of the following aspects. First of all, MEO and LEO satellites in space, manned and unmanned aircrafts in the air as well as trains and cars on the ground all travel at record-breaking high velocities. The high Doppler frequencies encountered may be better dealt with by non-coherent schemes. Secondly, the heterogeneous space-air-ground network is expected to deliver more frequent handovers among segment networks having different protocols and interfaces. This inevitably results in escalating overheads, which may be alleviated by non-coherent solutions. As discussed in Sec. II-D, the 5G network relies on non-coherent detection for the broadcast channels, where no channel knowledge is available at the beginning of cell-search. Finally, it is particularly beneficial

to employ the adaptive coherent/non-coherent technique for the control links of the high-velocity air/ground vehicles, which are normally assigned rather limited bandwidths. For example, the control link data rates for HAP fix-wing UAVs and LAP rotary-ring UAVs are estimated to be about 10 kbps and 100 kbps according to the aviation Air Traffic Management (ATM) [295] and 3GPP recommendations [112], respectively. Moreover, the ATM for civil aviation mainly relies on the crowded radio and decimeter Wave (dmWave) spectrum below 3GHz [21]. As a result, apart from the fact that the accuracy of channel estimation is severely eroded by the excessive normalized Doppler frequencies, the pilot overhead required for channel estimation would take up a higher percentage of the valuable bandwidth in low-rate control links, which may be avoided by invoking non-coherent schemes.

Looking back to the 3G era, when communication companies paid billions of dollars worldwide for licensed spectrum, they may now gain more by investing into the infrastructure and commercial integration with the existing satellite and aeronautical operators in the forthcoming 6G era. Compared to the technolgy-driven 1G-4G era, the current 5G era is more resource and business driven in terms of new spectrum and new URLLC and MTC applications. We anticipate that 6G may continue to follow the trend of better exploiting resources in both the spectral domain and in the space-air-ground domain, while new business models based on the current experiences may become vitally important for their future success.
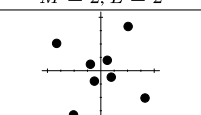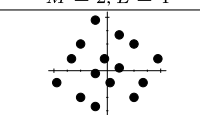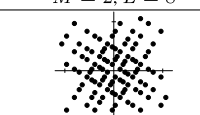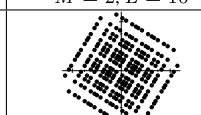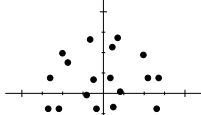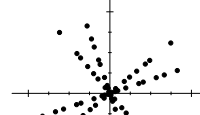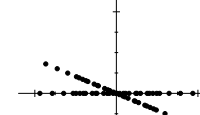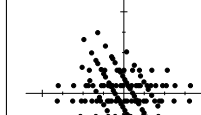
C. ENERGY-EFFICIENT INDEX MODULATION FOR ADVANCED MIMO DESIGN

Let us firstly once again clarify the three efficiency metrics that lie at the heart of MIMO design: bandwidth-efficiency, power-efficiency and energy-efficiency. *Bandwidth-efficiency* is directly measured by the effective system throughput, which is the data rate that can be delivered over a given bandwidth. For example, the multiplexing-oriented V-BLAST scheme [208], [210], [211] has a high throughput of $R = M \log_2 L$ that grows linearly both with the number of TAs M and with the number of bits per symbol $\log_2 L$, hence V-BLAST may be classified as a bandwidth-efficient MIMO design. *Power-efficiency* refers to the signal transmission power required for achieving specific target performance. For example, the diversity-oriented STBCs [209], [223]–[226] are capable of achieving a better performance in rich-scattering fading environments. As a result, for achieving a target DCMC capacity or BER performance, the SNR required by STBC is typically lower than that required by V-BLAST at low throughputs, especially when a low number of RAs (N) is used. Therefore, the employment of STBC is considered power-efficient in these scenarios. *Energy-efficiency* takes into account the energy consumption of supporting the hardware and the computational complexity of signal processing at both the transmitter and the

receiver. The emerging SM scheme [70]–[72] reduces the number of RF chains required at the transmitter from M to 1, while the inter-antenna synchronization is also eliminated. Moreover, owing to the absence of IAI, low-complexity single-stream based ML detection may be employed at the receiver [244], [245]. Therefore, the principal motivation of the SM design is to improve the energy-efficiency of MIMO systems. Nonetheless, as portrayed by Fig. 17 of Sec. IV-A, the reduction in the number of RF chains also results in an increase in the PA output power. For example, the PA of the single-RF SM scheme has to deliever M times higher output power $P_{out} = P_t$ than its counterparts operating in the full-RF V-BLAST scheme having $P_{out} = P_t/M$. Consequently, there exists an energy-efficiency tradeoff between single-RF, reduced-RF and full-RF MIMO designs. A reasonable rule-of-thumb is that when a low-gain PA is employed, reducing the number of RF chains may become more beneficial, as the circuit-drive power constitutes a higher proportion of the PA power consumption [138]. Regardless of the number of activated RF chains, from the PA's perspective, the actual signal constellation and the associated PAPR predetermine the power consumption of each active RF chain. To this end, the constant-envelope modulation schemes are generally considered to be the most energy-efficient.

As discussed in Sec. IV-A, the family of perfect STBCs [240]–[243] including the Golden code [239] are capable of concurrently achieving both the V-BLAST's multiplexing gain and the STBC's diversity gain, hence they are superior in terms of both bandwidth- and power-efficiency over a wide range of MIMO setups. However, despite the fact that the Golden code was once included in the IEEE 802.16e-2005 WiMAX standard [296], the judicious question arises, why the perfect STBCs are not employed in mainstream wireless networks? The primary reason is their degraded energy-efficiency, which is exemplied in Table. 15. More explicitly, the perfect STBC partitions the ($M \times M$) space-time signal matrix into M layers, where each layer disperses a total of M modulated QAM/HEX [240] symbols. For example, the mathematical description of the Golden code is portrayed in Fig. 31(a), where a total of $M^2 = 4$ modulated symbols $\{s_m\}_{m=1}^4$ are dispersed across $M = 2$ layers in the (2×2) signal matrix space. As a result, the effective throughput is given by $R = M \log_2 L$, which is the same as that of V-BLAST. Moreover, the employment of the Golden ratio $\theta = \frac{1+\sqrt{5}}{2}$ ensures a high diversity gain that does not diminish as the throughput R increases. However, as portrayed by Table. 15, the signal superpositions in the perfect STBC design result in excessive PAPRs during signal transmission, while the receiver has to deal with the increased IAI $= M^2$. In summary, compared to V-BLAST and STBC, the family of perfect STBCs impose substantial extra energy consumptions both at the transmitter and at the receiver, which can hardly be justified in the light of their modest performance gain in realistic communication systems.

TABLE 15. Constellation diagrams and energy-efficiency measures of Golden code [239] and perfect STBCs [240]–[243].

$M = 2, L = 2$	$M = 2, L = 4$	$M = 2, L = 8$	$M = 2, L = 16$
PAPR = 2.77 dB	PAPR = 2.77 dB	PAPR = 4.6 dB	PAPR = 5.3 dB
Equivalent IAI = 4	Equivalent IAI = 4	Equivalent IAI = 4	Equivalent IAI = 4
$M = 3, L = 2$	$M = 4, L = 2$	$M = 5, L = 2$	$M = 6, L = 2$
PAPR = 1.71 dB	PAPR = 5.75 dB	PAPR = 6.43 dB	PAPR = 6.08 dB
Equivalent IAI = 9	Equivalent IAI = 16	Equivalent IAI = 25	Equivalent IAI = 36

Arguably, the future 6G network may further improve the energy-efficiency even possibly to the detriment of both bandwidth- and power-efficiency for the following reasons. First of all, the trend of modern circuitry development indicates that the PA's dynamic range generally decreases as the carrier frequency increases [58]. Therefore, as the enabling technolgies move up from dmWave to cmWave, mmWave and even terahertz, having a low PAPR becomes increasingly important. Secondly, in the era of beyond-100Gbps data rate, the processing of high-bandwidth signals becomes progressively more power-thirsty, which urgently calls for more energy-efficient designs. Finally, with the aid of strong channel coding arrangements, the gaps between different MIMO performances is reduced, since they all operate in the near-Shannon-capacity region [23], [91], [244]. In summary, the most energy-efficient design is more likely to stand out as the promising candidate for next-generation networks.

Against this background, we have recently conceived a novel Finite-Cardinality Generalized STSK (FC-GSTSK) [137] scheme based on the TAST and IM philosophies, as portrayed in Fig. 36(b). First of all, following the IM philosophy, the signal superpositions within each TAST layer are replaced by the data-carrying dispersion matrix activation, where both the activation index of the matrix and the classic modulated symbol carry source information. Secondly, all the dispersion elements are taken from an L_{DM}-ary PSK constellation, where we have [$w_{L_{DM}} = \exp(j\frac{2\pi}{L_{DM}})$] in Fig. 36(b), while the signals from the second layer are rotated by $\phi_r = w_{L_r} = \exp(j\frac{2\pi}{L_r})$. Thirdly, the elements in the FC-GSTSK's dispersion matrices are actively rotated by the integer parameters $\{u_1, u_2\}$ seen in Fig. 36(b) in order to maximize the diversity gains. As a result, firstly, the constant-envelope PSK modulation is retained for signal transmission, when $\{s^{l_1}, s^{l_2}\}$ are modulated as L-PSK symbols. The bandwidth-efficient star QAM constellation may also be invoked without substantially increasing the PAPR in the same way, as in the family of STBCs. Secondly, the IM design eliminates the inter-layer IAI in Fig. 36(b),

so that the total IAIs that the receiver has to deal with is reduced to the same level of IAI = M, as that of V-BLAST. Thirdly, the FC-GSTSK throughput is given by $R = \log_2 L + \log_2 \overline{Q}$, where the cardinality of the dispersion matrix set $\overline{Q}$ is adjustable without having to impose any hardware cost in terms of having more TAs. Finally, all the integer parameters of $\{L_{DM}, L_r, u_1, u_2\}$ seen in Fig. 36(b) are always actively adjusted in order to optimize the diversity gains for different MIMO setups.

It is unprecedented in the two decades of MIMO system design that as the new FC-GSTSK arrangement [137] achieves both the multiplexing gain and the diversity gain, the PAPR of signal transmission and the equivalent IAI of signal detection are reduced to the same level as that of V-BLAST. Furthermore, together with its coherent STSK counterpart [138] of the DSTSK schemes discussed in Sec. IV-B, a holistic coherent versus non-coherent MIMO tradeoff family is unveiled for next-generation wireless networks. First of all, the FC-GSTSK of [137] is a full-RF coherent MIMO scheme, which may also be reconfigured into its reduced-RF version as shown in [137]. The FC-GSTSK scheme achieves the best performance in the family of finite-cardinality MIMO designs at the cost of the same IAI = M as V-BLAST. Secondly, the CE-STSK of [138] using the classic PSK constellation is a single-RF and constant-envelope coherent MIMO scheme, which remains IAI-free just as the orthogonal STBCs, but the STSK throughput is lower than that of the FC-GSTSK [137]. Thirdly, the DSTBC-ISK [19] and DSTSK-TAST [20] arrangements introduced in Sec. IV-B are all non-coherent MIMO schemes that exhibit the compelling benefits of being of single-RF, constant-envelope (for PSK signalling) and IAI-free nature. Finally, considering the realistic energy-efficiency tradeoff in terms of the RF chains, all the single-RF STSK and DSTSK signal matrices can be transformed into their full-RF or reduced-RF counterparts upon simply multiplying them by a normalized Hadamard matrix or by a reduced-size normalized Hadamard matrix, respectively, as demonstrated
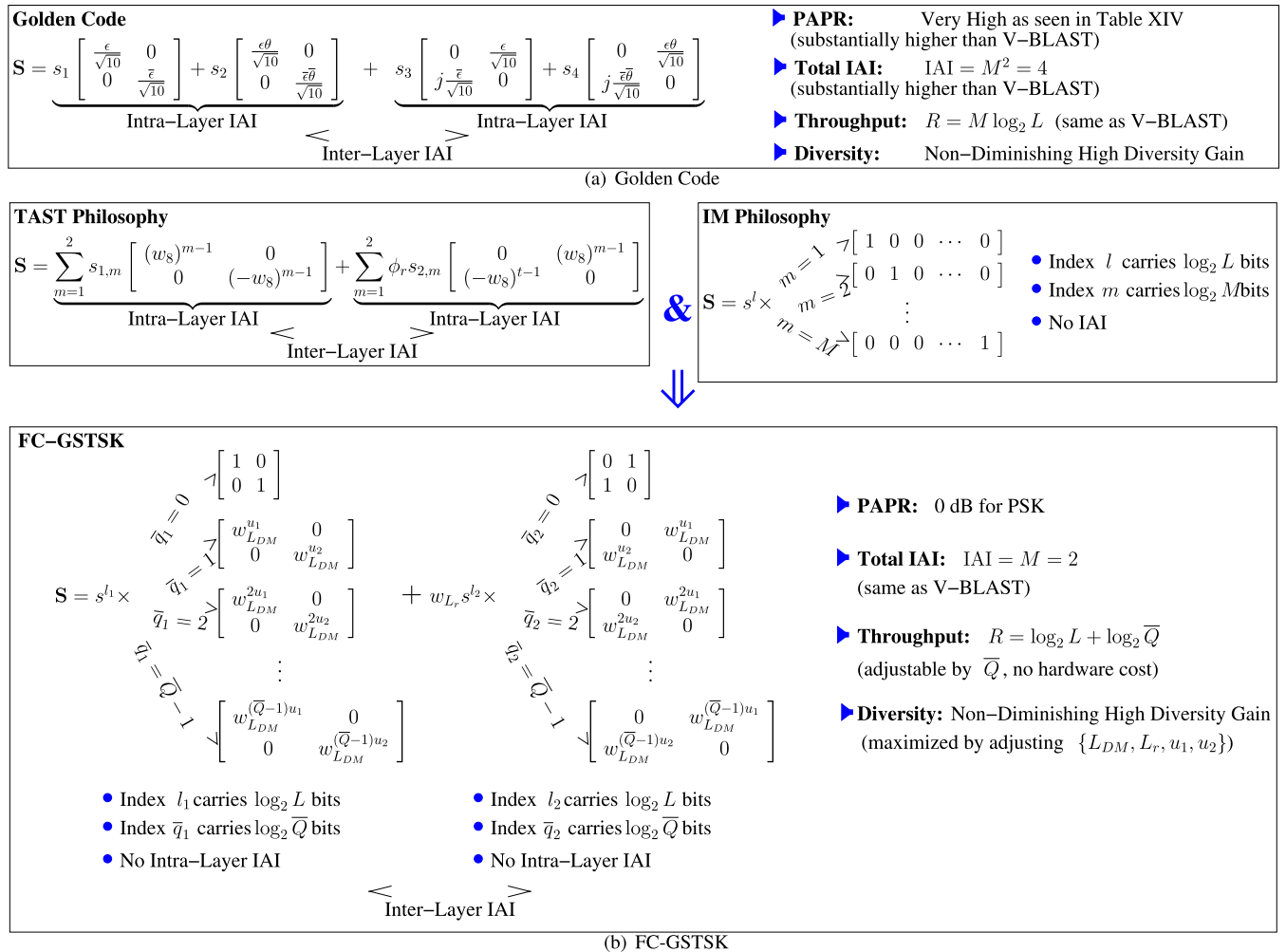
Golden Code

$$\mathbf{S} = s_1 \underbrace{\begin{bmatrix} \frac{\epsilon}{\sqrt{10}} & 0 \\ 0 & \frac{\bar{\epsilon}}{\sqrt{10}} \end{bmatrix} + s_2 \begin{bmatrix} \frac{\epsilon\theta}{\sqrt{10}} & 0 \\ 0 & \frac{\bar{\epsilon}\bar{\theta}}{\sqrt{10}} \end{bmatrix}}_{\text{Intra−Layer IAI}} + \underbrace{s_3 \begin{bmatrix} 0 & \frac{\epsilon}{\sqrt{10}} \\ j\frac{\bar{\epsilon}}{\sqrt{10}} & 0 \end{bmatrix} + s_4 \begin{bmatrix} 0 & \frac{\epsilon\theta}{\sqrt{10}} \\ j\frac{\bar{\epsilon}\bar{\theta}}{\sqrt{10}} & 0 \end{bmatrix}}_{\text{Intra−Layer IAI}}$$

$$\underbrace{}_{\text{Inter−Layer IAI}}$$

▶ **PAPR:** Very High as seen in Table XIV (substantially higher than V−BLAST)

▶ **Total IAI:** IAI $= M^2 = 4$ (substantially higher than V−BLAST)

▶ **Throughput:** $R = M \log_2 L$ (same as V−BLAST)

▶ **Diversity:** Non−Diminishing High Diversity Gain

(a) Golden Code

TAST Philosophy

$$\mathbf{S} = \underbrace{\sum_{m=1}^{2} s_{1,m} \begin{bmatrix} (w_8)^{m-1} & 0 \\ 0 & (-w_8)^{m-1} \end{bmatrix}}_{\text{Intra−Layer IAI}} + \underbrace{\sum_{m=1}^{2} \phi_r s_{2,m} \begin{bmatrix} 0 & (w_8)^{m-1} \\ (-w_8)^{t-1} & 0 \end{bmatrix}}_{\text{Intra−Layer IAI}}$$

$$\underbrace{}_{\text{Inter−Layer IAI}}$$

&

IM Philosophy

$$\mathbf{S} = s^l \times \begin{matrix} m=1 \nearrow \\ m=2 \searrow \\ \vdots \\ m=M \searrow \end{matrix} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

- Index l carries $\log_2 L$ bits
- Index m carries $\log_2 M$ bits
- No IAI

$$\Downarrow$$

FC−GSTSK

$$\mathbf{S} = s^{l_1} \times \begin{matrix} \bar{q}_1 = 0 \nearrow \\ \bar{q}_1 = 1 \nearrow \\ \bar{q}_1 = 2 \searrow \\ \vdots \\ \bar{q}_1 = \overline{Q}-1 \searrow \end{matrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} w_{L_{DM}}^{u_1} & 0 \\ 0 & w_{L_{DM}}^{u_2} \end{bmatrix}, \begin{bmatrix} w_{L_{DM}}^{2u_1} & 0 \\ 0 & w_{L_{DM}}^{2u_2} \end{bmatrix}, \begin{bmatrix} w_{L_{DM}}^{(\overline{Q}-1)u_1} & 0 \\ 0 & w_{L_{DM}}^{(\overline{Q}-1)u_2} \end{bmatrix} + w_{L_r} s^{l_2} \times \begin{matrix} \bar{q}_2 = 0 \nearrow \\ \bar{q}_2 = 1 \nearrow \\ \bar{q}_2 = 2 \searrow \\ \vdots \\ \bar{q}_2 = \overline{Q}-1 \searrow \end{matrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & w_{L_{DM}}^{u_1} \\ w_{L_{DM}}^{u_2} & 0 \end{bmatrix}, \begin{bmatrix} 0 & w_{L_{DM}}^{2u_1} \\ w_{L_{DM}}^{2u_2} & 0 \end{bmatrix}, \begin{bmatrix} 0 & w_{L_{DM}}^{(\overline{Q}-1)u_1} \\ w_{L_{DM}}^{(\overline{Q}-1)u_2} & 0 \end{bmatrix}$$

- Index l_1 carries $\log_2 L$ bits
- Index $\bar{q}_1$ carries $\log_2 \overline{Q}$ bits
- No Intra−Layer IAI

- Index l_2 carries $\log_2 L$ bits
- Index $\bar{q}_2$ carries $\log_2 \overline{Q}$ bits
- No Intra−Layer IAI

$$\underbrace{}_{\text{Inter−Layer IAI}}$$

▶ **PAPR:** 0 dB for PSK

▶ **Total IAI:** IAI $= M = 2$ (same as V−BLAST)

▶ **Throughput:** $R = \log_2 L + \log_2 \overline{Q}$ (adjustable by $\overline{Q}$, no hardware cost)

▶ **Diversity:** Non−Diminishing High Diversity Gain (maximized by adjusting $\{L_{DM}, L_r, u_1, u_2\}$)

(b) FC-GSTSK

FIGURE 31. Schematic descriptions of Golden code [239] and the recently developed FC-GSTSK [137], where $M = 2$ TAs are used. The parameters for Golden code are the fixed $\theta = \frac{1+\sqrt{5}}{2}$, $\epsilon = 1 + j - j\theta$, $\bar{\theta} = \frac{1-\sqrt{5}}{2}$ and $\bar{\epsilon} = 1 + j - j\bar{\theta}$. The parameters for FC-GSTSK are the adjustable $\{L_{DM}, L_r, u_1, u_2\}$, while the indices $\{l_1, l_2, \bar{q}_1, \bar{q}_2\}$ carries source information. The notations are $w_{L_{DM}} = \exp(j\frac{2\pi}{L_{DM}})$ and $\phi_r = w_{L_r} = \exp(j\frac{2\pi}{L_r})$.

in [138]. Specifically, for the example of using $M = 2$ TAs, the single-RF setup is transformed into the full-RF setup by:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{s_1}{\sqrt{2}} & \frac{s_1}{\sqrt{2}} \\ \frac{s_2}{\sqrt{2}} & -\frac{s_2}{\sqrt{2}} \end{bmatrix},$$

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 0 & s_1 \\ s_2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{s_1}{\sqrt{2}} & -\frac{s_1}{\sqrt{2}} \\ \frac{s_2}{\sqrt{2}} & \frac{s_2}{\sqrt{2}} \end{bmatrix},$$

where evidently the PAPR of signal transmission remains unaffected. Furthermore, the normalized Hadamard matrix $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ is unitary, whose Hermitian transpose is simply itself. Therefore, once the received signal matrix is multiplied by the same normalized Hadamard matrix, coherent/non-coherent STSK detection can be invoked without encountering any IAI.

D. FULL DUPLEX TECHNIQUES

As seen in Fig. 4, the 5G NR employs both FDD and TDD in the dmWave band, while TDD is the sole access technology in both the cmWave and mmWave bands. Considering that the electromagnetic fields of the transmitted and received signals are independent of each other, the Full-Duplex (FD) techniques that simultaneously transmit and receive signals are potentially capable of doubling the spectrum value. The schematics of TDD, FDD as well as the FD techniques of In-Band Full Duplex (IBFD) [80]–[84] and Multicarrier Division Duplex (MDD) [9], [85]–[87] are portrayed in Fig. 32, while their pros and cons are summarized in Table 16. More explicitly, the TDD separates the UL and DL in the time domain, which requires a guard interval during which the UL/DL switching transients subside. Thanks to the combined UL and DL spectrum, the TDD scheme may benefit from a frequency diversity gain, where the signal transmission is dynamically scheduled to that specific frequency band, which is associated with good near-instantaneous channel

TABLE 16. Comparison of duplex technologies.

Duplex	Advantages	Disadvantages
TDD	(1) combined UL and DL spectrum. (2) dynamic scheduling for frequency diversity. (3) support asymmetric traffic. (4) CSIT available without feedback.	(1) require guard interval. (2) stringent time synchronization. (3) high delay and outdated CSIT. (4) hidden terminal problem.
FDD	(1) long coverage. (2) no hidden terminal problem. (3) instantaneous HARQ.	(1) require guard band. (2) paired UL and DL bandwidths. (3) require feedback for CSIT
IBFD	(1) improved bandwidth-efficiency. (2) CSIT available without feedback. (3) no hidden terminal problem. (4) instantaneous HARQ.	(1) SI in both analog and digital domains. (2) stringent time synchronization. (3) small improvement for asymmetric traffic.
MDD	(1) higher bandwidth-efficiency than FDD/TDD. (2) free of SI in digital domain. (3) reduced OFDM PAPR. (4) dynamic scheduling for frequency diversity. (5) support asymmetric traffic. (6) CSIT available without feedback. (7) no hidden terminal problem. (8) instantaneous HARQ.	(1) SI in analog domain. (2) stringent time synchronization. (3) reduced bandwidth-efficiency compared to IBFD.
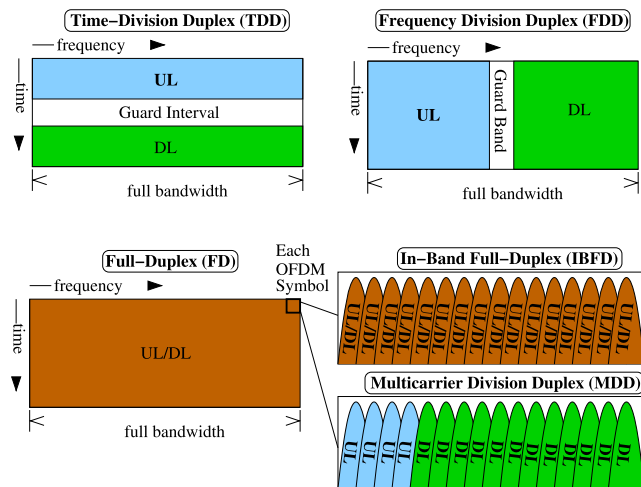
FIGURE 32. Schematic descriptions of Time Division Duplex (TDD), Frequency Division Duplex (FDD), In-Band Full-Duplex (IBFD) and Multicarrier Division Duplex (MDD).

conditions. The asymmetric UL/DL traffic may be approprately accommodated by assigning asymmetric time slot durations. Moreover, due to channel reciprocity, the Channel State Information at Transmitter (CSIT) becomes available without having to use feedback, which is especially benefical for the beamforming techniques required for cmWave and mmWave carriers.

Nonetheless, the guard interval of the TDD mode is often configured to be as high as 1/14 for LTE-A, 12.5% for IEEE 802.11ad, 20% for IEEE 802.11ac, which constitute substantial waste of the time-domain resources, when no useful information is transmitted. Moreover, the stringent time synchronization requirements substantially reduce the TDD coverage. Furthermore, the delay in TDD may defer the activation of HARQ in the MAC layer, while the CSIT may also become outdated. The hidden terminal problem arises, when a new node falsely detects an idle state of the access

point, which is in fact during the TDD reception state for another node that is hiden to the newcomer. By contrast, the out-of-band full-duplex scheme of FDD may alleviate the TDD shortcomings of this wasteful guard interval, stringent time synchronization, delayed HARQ, outdated CSIT and the hidden terminal problem. However, the FDD scheme is inflexible in terms of supporting asymmetric UL/DL traffic, because having simultaneously idle fixed and paired frequency bands is becoming increasingly scarce. As a result, the FDD mode is more suitable for long-range, symmetric and low-rate coverage, which cannot satisfy the increasing data rate demands of next generation wireless networks.

The IBFD mode [80]–[84] aims for exploiting the time- and frequency-resources to the fullest extent. However, the problem of Self-Interference (SI) constitutes a stumbling block in FD applications, where the high-power transmit signal detrimentally interferes with the low-power desired received signal at the FD node. Intuitively, the FD node knows its own transmitted signal, hence after estimating the leakage-path's impulse-response between the transmitter and receiver the SI can be reconstructed and substracted from the lower-power received signal. Explicitly, the series of analog components including the DAC, oscillator and PA in the transmit RF chain may impose both linear and non-linear signal distortions, which has to be accurately estimated before SI subtraction. More importantly, the SI may be up to billions of times stronger than the desired received signal (30∼120 dB [80]–[84]), which saturates the limited dynamic range of the finite-precision ADC. Considering the example that SI is 40dB (= 10 000 times) stronger than the received signal, while the rule-of-thumb ADC dynamic range of $6.02 \cdot Q$ dB is given by 48.2 dB for $Q = 8$ resolution bits, the effective ADC resolution for the desired signal is reduced to merely $(48.2 - 40)/6.02 \approx 1.36$ bits, which inevitably results in a substantially increased quantization noise inflicted upon the desired signal. In general, the SI in the IBFD scheme has to be mitigated in propagation, both by transmit and receive

antenna separation, as well as by analog and digital domain cancellation. More explicitly, first of all, the TA and RA may be physically separated, where the SI is attenuated by the natural propagation loss and penetration loss. As the second line of defense, the attenuated SI is modelled in the analog domain and subtracted from the received signal by the analog circuitry before ADC. Finally, the residual SI after ADC is estimated and mitigated in the digial domain. The associated hardware and software problems require further research before IBFD implementations find their way into commercial networks.

The recently developed MDD concept [9], [85]–[87] portrayed in Fig. 32 assigns the orthogonal OFDM subcarriers to the separate UL and DL channels. As a result, the MDD scheme is capable of mitigating many of the aforementioned drawbacks of TDD, FDD and IBFD. First of all, the TDD's guard interval and FDD's guard band are mitigated in the MDD mode, which innovatively exploits the OFDM interface. Secondly, no SI contaminates the signal detection in the digital domain, albeit the dynamic range of the ADC used for signal reception still has to be addressed. Thirdly, owing to the fact that only a subset of OFDM subcarriers are activated, the PAPR of signal transmission is substantially reduced, especially for the low-traffic UL when it is assigned a smaller number of subcarriers. Fourthly, thanks to the combined overall spectrum, the TDD advantages of frequency diversity and flexibly supporting asymmetric traffic are retained for the MDD mode. Lastly, owing to the simultaneous signal transmission and reception, the beneficial features of FDD and of IBFD including the availability of the CSIT without feedback, encountering no hidden terminal problems and no delay in HARQ are preserved by the MDD mode.

Having said this, the FD techniques may be deemed more promising for next-generation networks that rely on mmWave and terahertz carriers for the following reasons. First of all, as the antenna aperture decreases in the mmWave and terahertz bands, which is seen in Table 14 of Sec. VI-A, it becomes more feasible to achieve a relative high antenna separation compared to the wavelength in support of the FD operation, which imposes a higher propogation loss, hence mitigating the SI. Secondly, considering that the mmWave and terahertz signals typically suffer from high material-dependent penetration loss, blocking the SI by physical obstacles such as tinted glass or a metal shell may become much easier. Thirdly, the small mmWave and terahertz cell radius is favourable to the time synchronization requirements of both the TDD and FD modes.

Furthermore, as discussed in Sec. VI-B, the control links of high-mobility vehicles, aircraft and satellites are often assigned small but clean uncontended and uninterfered bandwidth. The new generation of "full-duplex in the sky" concept may effectively improve the commercial value of the limited spectrum assigned to the emerging space-air-ground control links, while the coherent/non-coherent adaptivity, the single-/multiple-TA adaptivity as well as the

multiplexing/diversity adaptivity discussed in Sec. V may further enhance the network's resilience against time-variant impairments.

E. EXIT CHARTS AND IRREGULAR CONVOLUTIONAL CODING

Sophisticated channel coding schemes [297] play an essential role both in the operational and future communication systems. By imposing redundancies on the mutually-independent source bits, the channel coded bits become highly correlated, so that errors occurring during signal transmission may be corrected by the channel decoder at the receiver. It is widely recognized that the bit-to-symbol mapping of the modulation schemes may impose correlation among the input bits [23], [91], [92], [298], [299]. Therefore, in order to closely approach the capacity limit, the soft-decision demodulator and the channel decoder may exchange extrinsic soft-information using turbo detection, as portrayed in Fig. 33. The popular coding schemes are the Convolutional Code (CC) of the IEEE 802.11a/n/ac/ah standard family, the Turbo Code (TC) of 4G LTE, the LDPC code of 5G NR and IrRegular Convolutional Code (IRCC) [23], [91].
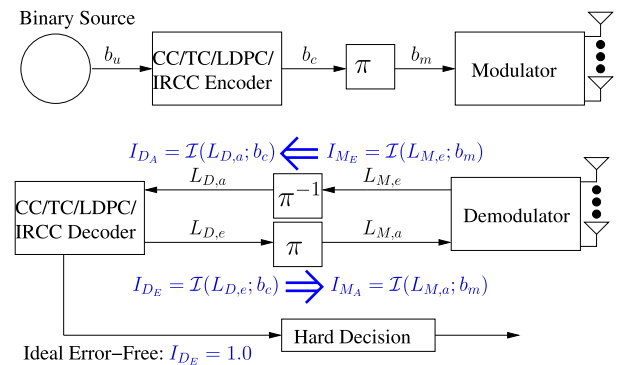
FIGURE 33. Schematic of channel coding aided modulation scheme and its two-stage turbo decoder. The channel coding arrangements include Convolutional Code (CC) of IEEE 802.11 family, Turbo Code (TC) of 4G LTE, Low-Density Parity-Check (LDPC) code of 5G NR and IrRegular Convolutional Code (IRCC) for the next-generation network.

At the transmitter, the source bits b_u are mapped to the coded bits b_c by the channel encoder and then interleaved to create the stream b_m before the bit-to-symbol mapping action of the modulator, as shown in Fig. 33. The two-stage turbo receiver portrayed in Fig. 33 is constituted by the demodulator and the channel decoder. Both of them accept and produce soft-bit decisions in the form of Log-Likelihood Ratios (LLRs). More explicitly, the extrinsic LLRs $L_{M,e}$ produced by the demodulator are de-interleaved and fed into the channel decoder as the *a priori* LLRs $L_{D,a}$. Similarly, the extrinsic LLRs $L_{D,e}$ produced by the channel decoder are interleaved and then fed into the demodulator the *a priori* LLRs $L_{M,a}$. The mutual information between the source bit at the transmitter and the corresponding LLR at the receiver
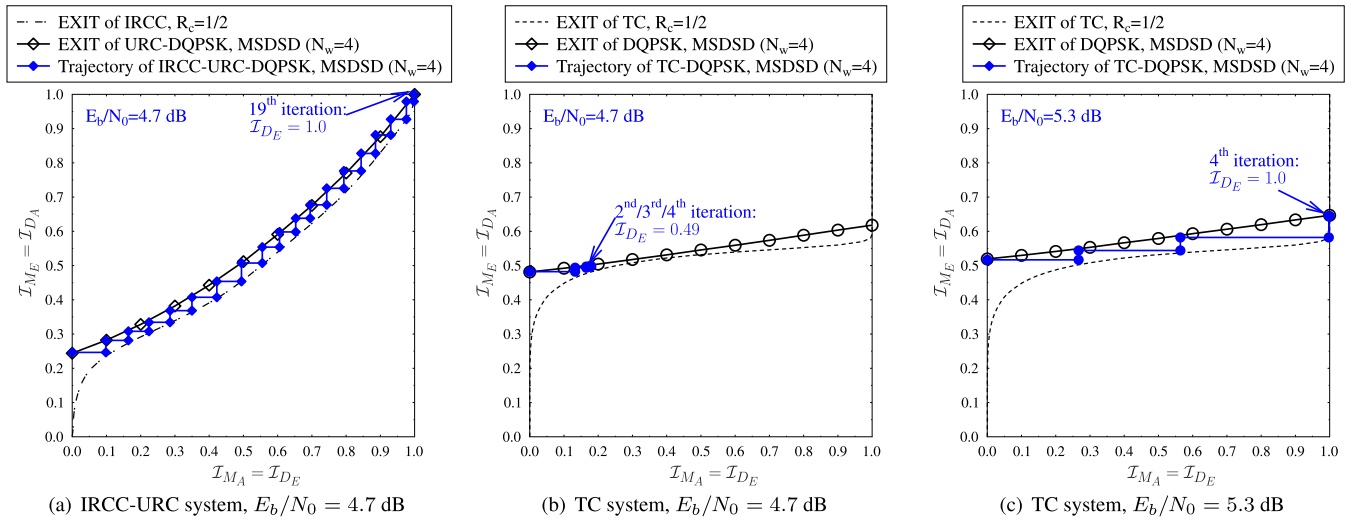
FIGURE 34. EXIT charts and decoding trajectories of MSDSD ($N_W = 4$) aided DQPSK over Ricean fading channels ($K = 0$ dB) associated with $N = 1$ in TC and IRCC-URC coded systems.

is evaluated from: [23], [91]–[93]

$$\mathcal{I}(L; b) = \frac{1}{2} \sum_{b=1,0} \int_{-\infty}^{\infty} p(L|b) \log_2 \frac{2p(L|b)}{\sum_{b'=1,0} p(L|b')} dL.$$

As a result, the extrinsic information produced by the demodulator becomes the *a priori* information of the channel decoder $\mathcal{I}_{M_E} = \mathcal{I}(L_{M,e}; b_m) = \mathcal{I}_{D_A} = \mathcal{I}(L_{D,a}; b_c)$, while the extrinsic information produced by the channel decoder becomes the *a priori* information of the demodulator $\mathcal{I}_{D_E} = \mathcal{I}(L_{D,e}; b_c) = \mathcal{I}_{M_A} = \mathcal{I}(L_{M,a}; b_m)$, as demonstrated in Fig. 33.

The powerful analytical tool of EXtrinsic Information Transfer (EXIT) charts [23], [91]–[93] visualizes the convergence behaviour of the receiver in Fig. 33, which is exemplified for the TC and IRCC arrangements in Fig. 34. Without delving into the specifics of creating an EXIT chart, suffice to say that in order to achieve decoding convergence to a vanishingly low BER, the following two requirements have to be satisfied in Fig. 34:

(I) the area under the EXIT curve of the outer channel code of Fig. 33 (TC/IRCC) decoder is higher than that of the inner component constituted by the demodulator;

(II) the only intersection point between the EXIT curves of the outer channel (TC/IRCC) decoder and the inner demodulator should be at $\mathcal{I}_{D_E} = 1.0$.

Based on this, the powerful IRCC seen in Fig. 34(a) is constituted by the weighted superposition of $P = 17$ component CCs that satisfy [23], [91], [300]:

$$\sum_{i=1}^{P} \alpha_i = 1, \quad R_c = \sum_{i=1}^{P} \alpha_i R_c^i,$$

where $\{\alpha_i\}_{i=1}^{P}$ are the weighting coefficients, while $\{R_c^i\}_{i=1}^{P}$ denote the coding rates of the component CCs. The appropriate coefficients $\{\alpha_i\}_{i=1}^{P}$ are chosen to minimize the area for achieving decoding convergence, as exemplified by Fig. 34(a). We note that a low-complexity Unity Rate Code (URC) is introduced as an intermediate code in the

IRCC coded system, where a free distance of two in the resultant concatenated scheme is the sufficient and neccessary condition for achieving an infinitesimally low BER [23], [91], [301], [302]. At the receiver, the URC decoder and the demodulator are amalgamated as the inner "demodultor" block of Fig. 33 in the same way as shown in [21].

In summary, Fig. 34(a) shows that the decoding convergence of $\mathcal{I}_{D_E} = 1.0$ is achieved by the IRCC system at $E_b/N_0 = 4.7$ dB. By contrast, at the same $E_b/N_0 = 4.7$ dB, the TC scheme can only reach the erroneous $\mathcal{I}_{D_E} = 0.49$, where further increasing the decoding iterations does not offer any performance improvement. As a result, Fig. 34(c) evidences that a higher $E_b/N_0 = 5.3$ dB is required for achieving $\mathcal{I}_{D_E} = 1.0$ by the TC system. Nonetheless, we note that the power-efficiency gain of the IRCC system is achieved at the cost of employing a higher number of iterations. For the IRCC system using $\mathrm{IR}_{\mathrm{URC-DQPSK}} = 2$ inner iterations and $\mathrm{IR}_{\mathrm{IRCC-\{URC-DQPSK\}}} = 20$ outer iterations, the total number of turbo detection iterations is given by 40. By contrast, for the TC system using $\mathrm{IR}_{\mathrm{TC}} = 4$ inner iterations and $\mathrm{IR}_{\{\mathrm{TC-DQPSK}\}} = 4$ outer iterations, the total number of system iterations is only 16, which is substantially smaller than that of the IRCC system.

Let us once again compare the performance of the coherent QPSK relying on pilot-based CSI estimation and the non-coherent MSDSD ($N_w = 4$) aided DQPSK in the channel coded scenarios, which are reproduced from [21]. First of all, Fig. 35 evidences that the IRCC coded arrangements are capable of achieving power-efficiency gains of up to 1dB over the TC coded counterparts. Secondly, Fig. 35(a) demonstrates that at the low $f_d = 0.001$, the coherent scheme outperforms its non-coherent counterpart in both TC and IRCC coded scenarios. However, Fig. 35(b) shows a reversal of the trend, where the non-coherent scheme performs better in the high-mobility scenario of $f_d = 0.03$. This is due to the following reasons:
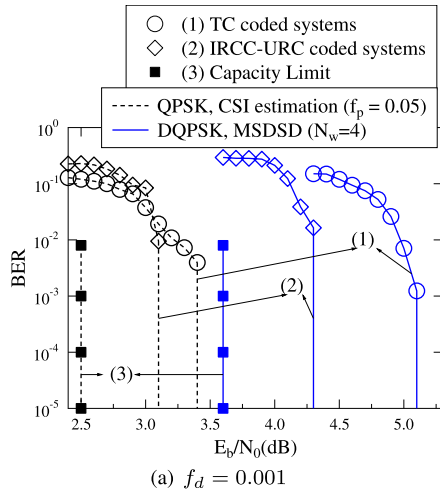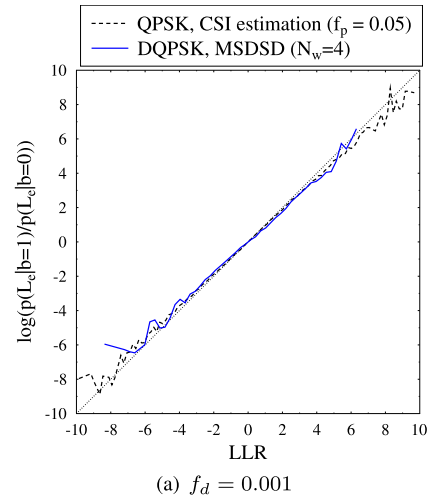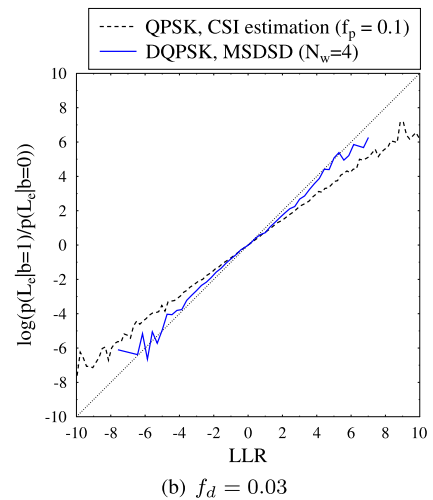
(a) $f_d = 0.001$
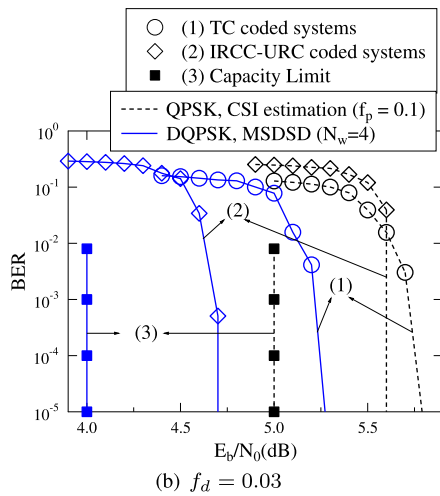
(b) $f_d = 0.03$

FIGURE 35. BER performance comparison betweem coherent QPSK versus non-coherent MSDSD ($N_w = 4$) aided DQPSK over Ricean fading channels ($K = 0$ dB) associated with $N = 1$ in TC and IRCC-URC coded systems. In TC coded systems, the numbers of iterations are ($IR_{TC} = 16$, $IR_{TC-QPSK} = 1$) and ($IR_{TC} = 4$, $IR_{\{TC-DQPSK\}} = 4$). In IRCC-URC coded systems, we have ($IR_{URC-QPSK} = 1$, $IR_{IRCC-\{URC-QPSK\}} = 40$) and ($IR_{URC-DQPSK} = 2$, $IR_{IRCC-\{URC-DQPSK\}} = 20$).

(1) For coherent detection, the accuracy of the pilot-based CSI estimation degrades as f_d increases, which results in the eroded LLR accuracy seen in Fig. 36. More explicitly, given LLR definition of $L_e = \ln \frac{p(L_e|b=1)}{p(L_e|b=0)}$, the LLR calculated by the decoder based on this equation is supposed to result in a diagonal line in Fig. 36. However, we observe in the figure that this is not satisfied by the coherent detector represented by the dashed line at $f_d = 0.03$ owing to the less accurate channel estimation. As the number decoding iterations increases for the turbo receiver seen in Fig. 33, the performance of the coherent scheme degrades further in the channel coded scenarios as seen in Fig. 35(b).

(2) By contrast, the MSDSD (solid line) dispensing with CSI estimation remains capable of producing more reliable LLRs, as evidenced by Fig. 36. As a result, with the assistance of channel coding, the MSDSD

(a) $f_d = 0.001$

(b) $f_d = 0.03$

FIGURE 36. LLR accuracy test of coherent QPSK versus non-coherent MSDSD ($N_w = 4$) aided DQPSK over Ricean fading channels ($K = 0$ dB) associated with $N = 1$ and recorded at $E_b/N_0 = 1$ dB.

outperforms its coherent counterpart at $f_d = 0.03$ in Fig. 35(b).

F. MACHINE LEARNING AND OPTIMIZATIONS

Given our access to wider spectral resources facilitated by the advances in mmWave enabling techniques and heterogenous space/air/ground networks, including energy-efficient MIMO, full-duplex, NOMA and sophisticated channel coding arrangements, we have a large selection of options to choose from for next-generation wireless networks. In fact, the design-space of parameters has become excessive for evaluating all combinations of system parameters by exaustive search for achieving the holistic optimization of the system. Sophisticated reduced-search-scope techniques are required for searching through a fraction of options, whilst achieving a near-optimal performance. In recent years, it has been increasingly accepted that Artificial Intelligence (AI) is capable of making better decisions than humans. The stunning victory of Google's AlphaGo against the world champion, Lee Sedol, surprised both Go players and AI
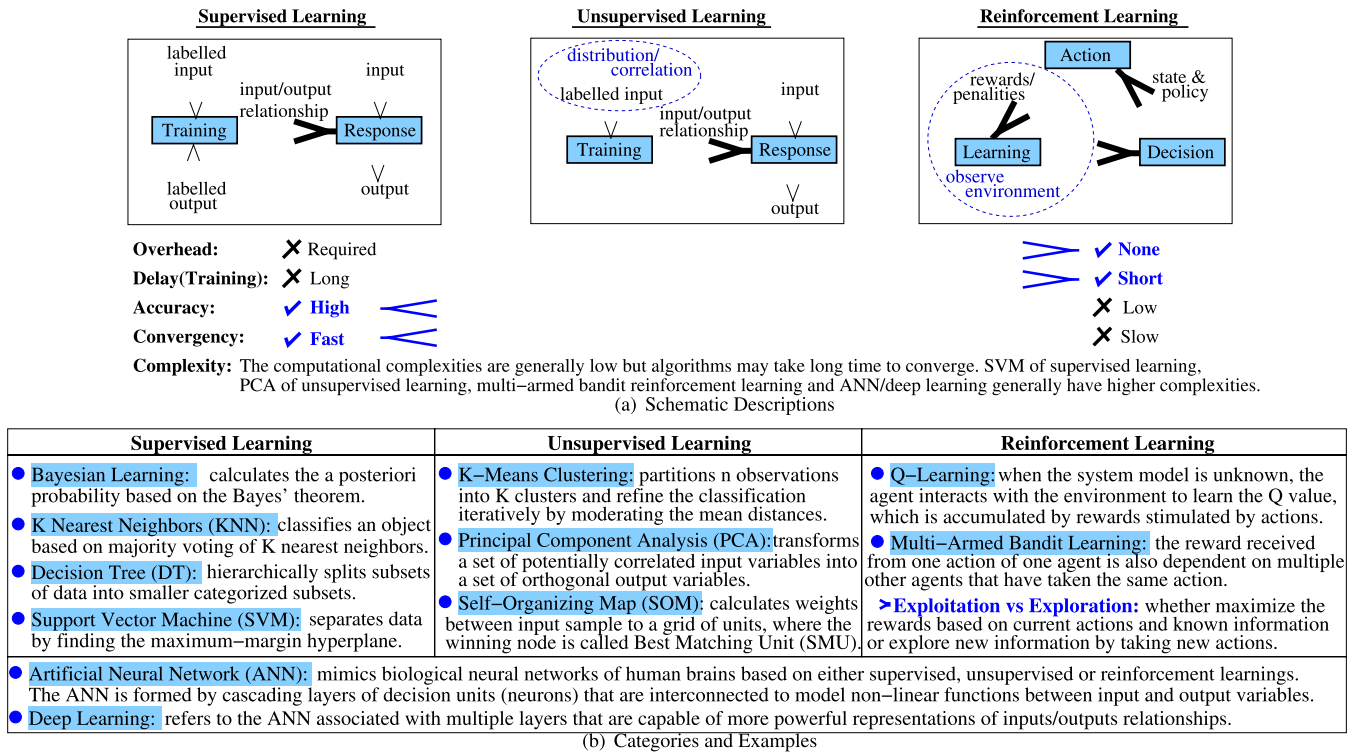
FIGURE 37. Categories, schematics, characteristics and examples of machine learning algorithms.

engineers for its unexpected winning move that no human would have opted for.

As a prospering field of AI, the machine learning concept was conceived in the 1950s [303] as an autonomous analysis tool that solves problems without explicit programming. Having been consistently improved over the last few decades, machine learning algorithms are envisioned to gain a much wider scale of commercial deployment in the intelligent 6G networks. More explicitly, machine learning techniques may be broadly categorized into supervised learning, unsupervised learning and reinforcement learning, as portrayed in Fig. 37(a).

To expound a little further on these three classes, we commence by considering the *supervised learning*. Explicitly, the algorithm is provided with a set of labelled traning input samples and their desired output samples, so that the underlying relationship function may be learned by the intelligent agent. Then the agent proceeds to operate on-line in the target environment, where new outputs are produced with given new inputs. The family of state-of-the-art supervised learning algorithms is portrayed in Fig. 37(b), andfurther elaborated on as follows [98]–[103]:

(A.1) *Bayesian Learning:* Based on Bayes' theorem, the *a posteriori* probability of the target hypothesis is evaluated based on the *a prior* knowledge of the associated event probability and distribution.

(A.2) *K Nearest Neighbors (KNN):* An object is categoized into a specific class based on the majority voting of its K nearest neighbors. In contrast to Bayesian learning,

no assumptions have to be made concerning the data distribution, yet a low computational complexity is retained.

(A.3) *Decision Tree (DT):* The labels of data are decided by a learning tree, which hierarchically splits subsets into smaller categorized subsets. The DT algorithm imposes the specific requirement that the data set has to be linearly separable by a hyperplane.

(A.4) *Support Vector Machine (SVM):* The training data are transformed into a higher dimensional space, where they become linearly separable by a hyperplane. Based on this, the SVM learns in order to find the maximum-margin hyperplane, which produces the largest separation between different classes. The SVM is capable of solving unconstrained non-convex optimization problems at the cost of a higher computational complexity than other less sophisticated supervised learning algorithms.

Secondly, in contrast to supervised learning, in *unsupervised learning*, no lablled output training samples are provided. Instead, the agent endeavors to learn the hidden structure based on the correlations and distributions of the input samples, as portrayed in Fig. 37(a). More explicitly, the unsupervised learning algorithms exemplified in Fig. 37(b) are further detailed as follows [98]–[103]:

(B.1) *K-Means Clustering:* The algorithm partitions n data points into K clusters, where each data point is assigned to a cluster based on the Euclidean distance.

The algorithm iteratively refines the assignment by updating the mean of each cluster, until a predefined maximum distance-sum is reached.

(B.2) *Principal Component Analysis (PCA):* The algorithm transforms a large set of potentially correlated variables into a smaller set of uncorrelated variables, which are referred to as the principal components. Hence the PCA may be viewed as a dimension-reduction tool capable of transforming a large ensemble of data into a reduced-size data representation, so that the amount of data to be processed and transmitted may be substantially reduced.

(B.3) *Self-Organizing Map (SOM):* In contrast to neural networks that construct decision units in sequential layers, the SOM is constituted by a grid of prototype neuron units. The SOM evaluates the weights between the input sample and all possible neurons. The closest neuron is chosen as the Best Matching Unit (BMU), and the input sample is assigned to that category accordingly.

Thirdly, when the system model is unknown, the *reinforcement learning* technique may be invoked, where the agent continously interacts with the environment by taking tentative actions based on the associated states and policies. As portrayed in Fig. 37(a), a reward/penalty mechanism is used by the agent to learn the consequences of its actions, so that the agent may be trained to make better decisions over time. The specific example of reinforcement learning seen in Fig. 37(b) are [98]–[103]:

(C.1) *Q-Learning:* When both the system model and the state transition probability are unknown, the Q-learning algorithm aims for finding an optimal action policy for a Markov decision process. Explicitly, a legitimate action of the agent is associated with a state and a policy. Moreover, the rewards gleaned from a series of actions are accumulated as the Q value. The agent always activates the specific action associated with the largest Q value, which can be trained to become more reliable over time.

(C.2) *Multi-Armed Bandit Learning:* When multiple agents are deployed, the reward received after a specific action of a particular agent is also dependent on those of the other agents who activate the same action. Since the agents have no priori knowledge of the system model, there is a crucial trade-off between exploitation and exploration, where the algorithm has to decide whether to maximize the rewards based on the current actions and known information or to explore the potential benefits of new information by taking new actions.

Moreover, as shown in Fig. 37(b), the *Artificial Neural Network (ANN)* that mimics the operation of the biological neural networks of the human brain may operate based on supervised, unsupervised and reinforcement learning. More explicitly, the ANN consists of cascading layers of decision units, which are also known as neurons. The neurons of consecutive layers are interconnected by non-linear functions

that are in charge of evaluating the weights between the input/output variables and the intermediate neuron nodes. Thanks to the multiple-input multiple-output neural network feature, the ANNs are capable of executing multiple tasks at the same time. Moreover, the ANN associated with a higher number of neuron-layers is often termed as *Deep Learning*, which has enhanced representation capability.

The machine learning techniques investigated in communications can be categorized into two major classes, which are the regression problems of predicting a value and the classification problems of deciding on a category, property or cluster. The major challenges include accurate data representation and optimization modelling, as well as finding the most suitable algorithm for the specific target application considered. More explicitly, as portrayed in Fig. 37(a), supervised learning imposes a training overhead and long delay dedicated to off-line training, which may be more suitable for well-trained scenarios such as localization, mobility prediction, handover management, spectrum and power allocation, etc. By contrast, the reinforcement learning dispenses with the need for labelled training samples and hence it can act promptly, as shown in Fig. 37(a). However, the convergence of reinforcement learning may be slow, and their performance in multi-agent non-Markovian systems may become particularly unsatisfactory [102]. As a result, reinforcement learning is mostly applied in self-optimization scenarios [99]. Moreover, as indicated in Fig. 37(a), most machine learning algorithms are designed to have low computational complexity, where the power consumption tends to be governed by their long convergence time. Nonetheless, as shown in Fig. 37(a), the SVM of supervised learning, the PCA of unsupervised learning, multi-armed bandit technique of reinforcement learning as well as the ANN and deep learning algorithms are generally considered to be computationally more complex. Hence the better ones are not particularly suitable for IoT and wireless sensor networks. By contrast, the classification and dimension-reduction learning techniques of KNN, DT and K-means clustering are better candidates for reducing the data processing time and data transmission volume in low-complexity IoT and sensor scenarios [100], [101].

Finally, Fig. 38 summarizes a range of potential cost functions and techniques that may be beneficially invoked for solving challenging multi-component optimization problems in next-generation wireless networks. We have previously demonstrated in Sec. II that the 5G systems offer hitherto unprecedented flexibility and adaptivity in diverse communication scenarios. Since 6G further extends this scope to unexplored domains, it may be envisioned that more sophisticated techniques will be developed to fulfil the future demands, where machine learning will play an increasingly essential role in complex multi-objective Pareto-optimization problems. The different cost functions of Fig. 38 are often intricately inter-linked. For example, the BER typically cannot be reduced without increasing the received power, complexity or latency. The ultimate solution is to determine the so-called Pareto-front of the multi-objective optimization problem
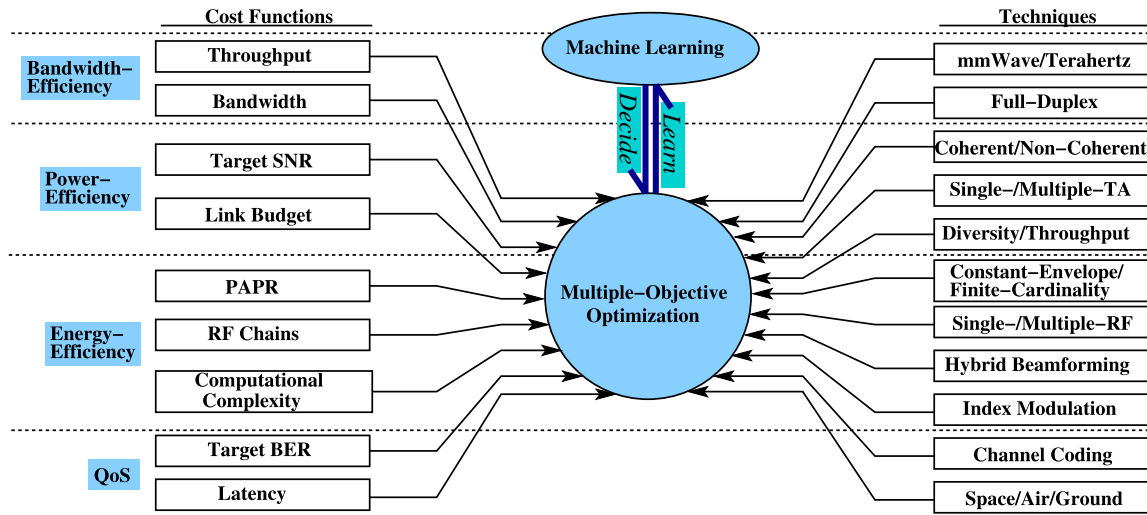
FIGURE 38. Summary on cost functions and techniques for learning and optimizing the next-generation wireless networks.

considered [304]–[306], as exemplified in Fig. 39. More explicitly, the Pareto-front contains all optimal solutions, which may include the minimized BER and latency as well as the best bandwidth-/power-/energy-efficiencies. In other words, the Pareto-front contains all optimal solutions of a specific problem and none of the cost function can be further improved without compromising at least one of the others.

Fig. 39 portrays a pair of different approaches conceived for solving a pair of optimization problems. On one hand, the two objective function components may be equally important in many scenarios, such as meeting the dual objectives of having the minimum BER and latency as part of the 5G URLLC. Therefore, the joint optimization problem of Fig. 39(a) is preferred, where Objectives 1 and 2 are optimized at the same time. By contrast, sometimes Objective 1 has a higher priority than Objective 2. For example, at the BS, the reliable coverage is generally more important than having the minimal power consumption. As a result, Fig. 39(b) demonstrates that Objective 1 may be prioritized over Objective 2, where the first optimization provides solutions closer to optimum, while the subsequent second optimization has to tolerate a suboptimal result.

In order to better illustrate this effect, let us consider the MIMO design tradeoffs previously discussed in Sec. IV-A. There are four formulated cost functions to be maximized in the generic MIMO design problem [19], [20], [137], [138]. Firstly, the diversity product Λ_p representing the diversity gain is defined as the normalized determinant term in the pairwise error probability function. Secondly, the diversity sum Λ_s is defined as the normalized Euclidean distance, which indicates the separation between the nearest constellation points. Moreover, the average diversity product $\overline{\Lambda}_p$ and the average diversity sum $\overline{\Lambda}_s$ are calculated by averaging over all bit-mappings according to the BER formula. In summary, we note that the four objective functions exhibit a tradeoff relationships, since they cannot be all maximized at the same time. Moreover, the four objective
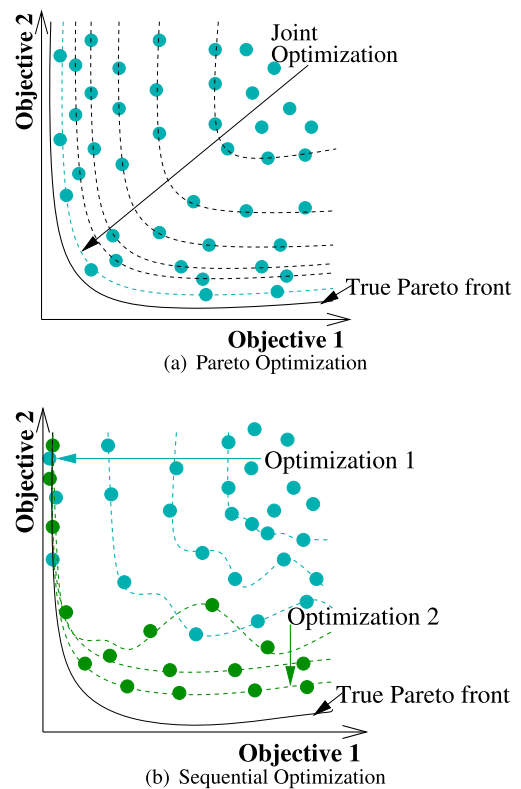

(a) Pareto Optimization

(b) Sequential Optimization

FIGURE 39. Examples of two objective optimization problems.

functions are not equally important for different MIMO design scenarios.

Let us consider the FC-GSTSK scheme of Sec. VI-C as an example. According to the schematics portrayed in Fig. 36(b), for a specific MIMO setup scenario that has M TAs, T transmission time slots and a throughput of R b/s/Hz, the best FC-GSTSK candidate is obtained by adjusting the integer parameters of $\{\overline{Q}, L, L_{DM}, L_r, \mathbf{u}\}$, which becomes excessively complex for four optimization objectives. As exemplified by Fig. 40, we firstly have to
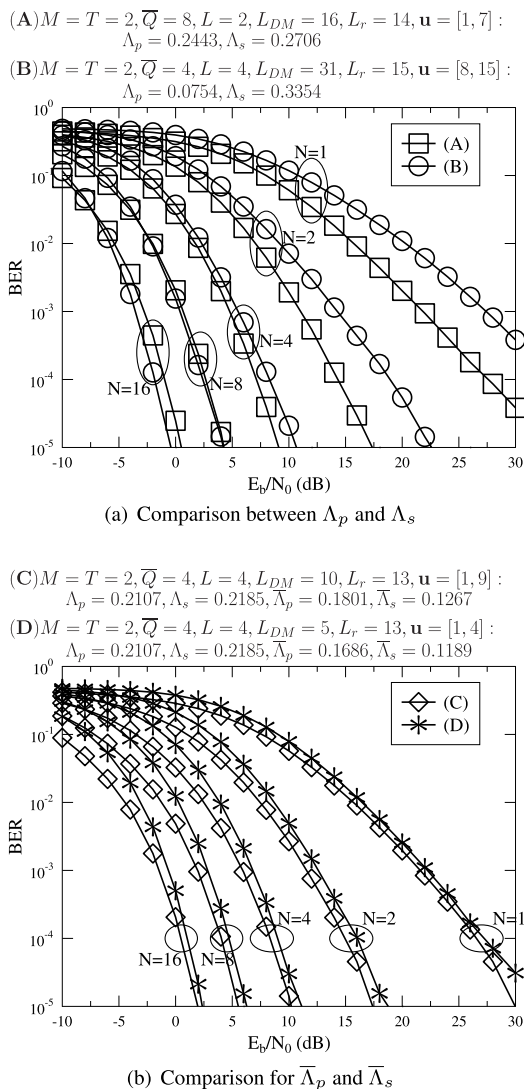
(A)$M = T = 2, \overline{Q} = 8, L = 2, L_{DM} = 16, L_r = 14, \mathbf{u} = [1, 7]$:
$\Lambda_p = 0.2443, \Lambda_s = 0.2706$

(B)$M = T = 2, \overline{Q} = 4, L = 4, L_{DM} = 31, L_r = 15, \mathbf{u} = [8, 15]$:
$\Lambda_p = 0.0754, \Lambda_s = 0.3354$

(a) Comparison between Λ_p and Λ_s

(C)$M = T = 2, \overline{Q} = 4, L = 4, L_{DM} = 10, L_r = 13, \mathbf{u} = [1, 9]$:
$\Lambda_p = 0.2107, \Lambda_s = 0.2185, \overline{\Lambda}_p = 0.1801, \overline{\Lambda}_s = 0.1267$

(D)$M = T = 2, \overline{Q} = 4, L = 4, L_{DM} = 5, L_r = 13, \mathbf{u} = [1, 4]$:
$\Lambda_p = 0.2107, \Lambda_s = 0.2185, \overline{\Lambda}_p = 0.1686, \overline{\Lambda}_s = 0.1189$

(b) Comparison for $\overline{\Lambda}_p$ and $\overline{\Lambda}_s$

FIGURE 40. Performance comparison between four FC-GSTSK($M = T = 2$, $R = 4$) associated with different parameters (adapted from [137]).

decide on the priorities of the four cost function components $\{\Lambda_p, \Lambda_s, \overline{\Lambda}_p, \overline{\Lambda}_s\}$. More explicitly, Fig. 40(a) shows that arrangement (A) has a larger Λ_p, which performs better for small number of RAs $N = \{1, 2, 4\}$. By contrast, arrangement (B) has a larger Λ_s, which results in a superior performance for large $N = \{8, 16\}$. Furthermore, Fig. 40(b) evidences that despite having the same Λ_p and Λ_s, the arrangement (C) associated with larger $\overline{\Lambda}_p$ and $\overline{\Lambda}_s$ perform better than arrangement (D). This once again confirms that the objective function having the four components of $\{\Lambda_p, \Lambda_s, \overline{\Lambda}_p, \overline{\Lambda}_s\}$ cannot achieve their maxima at the same time, and they have different priorities for different MIMO setup.

There are a total number of 4! $= 24$ possible priority arrangements for the four-tuple $\{\Lambda_p, \Lambda_s, \overline{\Lambda}_p, \overline{\Lambda}_s\}$. In order to solve this, in [137], we proposed to firstly provide training samples by generating random parameter combinations of

$\{\overline{Q}, L, L_{\mathrm{DM}}, L_r, \mathbf{u}\}$ and then observed their BER results. The analysis is characterized by the following linear regression equation:

$$SNR_t = \kappa_0 + \kappa_1 \Lambda_p + \kappa_2 \Lambda_s + \kappa_3 \overline{\Lambda}_p + \kappa_4 \overline{\Lambda}_s,$$

where for each example construction of $\{\overline{Q}, L, L_{\mathrm{DM}}, L_r, \mathbf{u}\}$, the associated four-tupples $\{\Lambda_p, \Lambda_s, \overline{\Lambda}_p, \overline{\Lambda}_s\}$ are evaluated as sample inputs of the regression model, while the sample output SNR_t is recorded at the target BER $= 10^{-4}$. The regression coefficients $\{\kappa_0, \cdots, \kappa_4\}$ can be automatically determined by the multiple regression analysis, and gradually descending priorities are assigned to the four parameters of the objective function according to their descending coefficient values. Following this, the related optimization may be carried out one by one according to their priorities, as exemplied by Fig. 39(b).

Finally, we note that thanks to the intensive development of machine learning, high-performance Graphic Processing Units (GPUs) are available at low cost, such as Nvidia TITAN RTX and GeForce RTX 2080Ti [307], which facilitate TFLOPS-order of mathematical calculations. As a result, the recent machine learning framework is expected to be capable of completing multi-objective optimization tasks in a shorter duration than conventional approaches, even if low-cost GPUs are used. Moreover, the learning process requires a high-performance GPU, while the inference process requires a relatively simple computational resource that can be implemented within a wireless module. Motivated readers may refer to [307] for an up-to-date tool-kit of GPU-aided machine learning algorithms that aim for solving diverse optimization problems in wireless Index Modulation (IM) applications.

VII. CONCLUSION

In this contribution, we commenced by reviewing the landscape of 5G NR standards, where the importance of the coherent versus non-coherent tradeoff was analysed. We have surveyed the sixty-years history of coherent versus non-coherent tradeoffs, where the myth that non-coherent techniques completely dispense with channel estimation was dispelled. Moreover, we also studied the twenty-years history of coherent versus non-coherent tradeoffs in MIMO systems, where the energy-efficiency issues including the number of RF chains, the PAPR of signal transmission and the IAI of signal detection were highlighted in the context of our design guidelines. Finally, we demonstrated that our adaptive design offers significant benefits in next-generation mmWave and Terahertz communication, space-air-ground integrated networks, in full-duplex techniques and in the enhanced channel coding assisted system design, where machine learning algorithms are expected to make autonomous decisions concerning the specific choice of the near-instantaneously adaptivie mode of operation.

APPENDIXES
APPENDIX A
PROOF OF PROPOSITION 1

Proof: As exemplified by Fig. 18, the differential encoding starts from the identity matrix as $\mathbf{S}_1 = \mathbf{I}_2$ for $n = 1$. The elements in Alamouti's signal matrix are always modulated as the normalized LPSK symbols as $\{x_{n-1,1}, x_{n-1,2}\} \in \{x^l = \frac{1}{\sqrt{2}} \exp(j\frac{2\pi}{L}l)\}_{l=0}^{L-1}$. As a result, for $n = 2$, the constellation set for $\{s_{n,1}, s_{n,2} \in \{s^l\}_{l=0}^{\mathcal{L}_n - 1}\}_{n=2}$ is the same as $\{x_{n-1,1}, x_{n-1,2}\}$, hence we have $\mathcal{L}_2 = L$. As the time index n further increases, assuming the simple example of $x_{n-1,1} = x_{n-1,2} = \frac{1}{\sqrt{2}}$, the differential encoding seen in Fig. 18 leads to new constellation points of $\frac{1}{\sqrt{2}}(s^l + s^{l'})$ for $l \neq l'$. Owing to the associated constellation symmetry, the four operations of $\frac{1}{\sqrt{2}}(s^l + s^{l'})$, $\frac{1}{\sqrt{2}}(s^l - s^{l'})$, $\frac{1}{\sqrt{2}}[s^l + (s^{l'})^*]$ and $\frac{1}{\sqrt{2}}[s^l - (s^{l'})^*]$ result in the same constellation set over all possible $l \neq l'$. Moreover, we note that $\{x^l\}_{l=0}^{L-1}$ remains a subset of $\{s^l\}_{l=0}^{\mathcal{L}_n - 1}$, hence without the assumption of $x_{n-1,1} = x_{n-1,2} = \frac{1}{\sqrt{2}}$, the operations seen in Fig. 18 always lead to the same set $\{s^l\}_{l=0}^{\mathcal{L}_n - 1}$. Therefore, in summary, the cardinality is given by $\mathcal{L}_n = \frac{\mathcal{L}_{n-1}(\mathcal{L}_{n-1} - 1)}{4}$, which tends to infinity as the time index n increases. $\qquad \square$

APPENDIX B
PROOF OF PROPOSITION 2

Proof: The average signal power of $\{|s_{n,1}|^2 = |s_{n,2}|^2 = 1/2\}_{\forall n}$ is normalized by the number of antennas. As discussed for Proposition 1, the constellation set for $\{s^l\}_{l=0}^{\mathcal{L}_2 - 1}$ is the same as $\{x^l\}_{l=0}^{L-1}$, hence we have the same peak power of $1/2$ and the PAPR of $\text{PAPR}_2 = 0$ dB for $n = 2$. Furthermore, for $n = 3$, the peak-power of a symbol becomes $s_n^{peak} = 1$, which corresponds the case of $x_{n-1,1} = x_{n-1,2} = s_{n-1,1} = s_{n-1,2} = \frac{1}{\sqrt{2}}$ in Fig. 18. The resutant PAPR is given by $\text{PAPR}_3 = 3$ dB for $n = 3$. As the symbol index n increases, the peak-power of a symbol is always in the form of $s_n^{peak} = \sqrt{2}s_{n-1}^{peak}$, which is given by $x_{n-1,1} = x_{n-1,2} = \frac{1}{\sqrt{2}}$ and $s_{n-1,1} = s_{n-1,2} = s_{n-1}^{peak}$. This implies that the theoretical peak signal power doubles when n is increased by 1, as given by $|s_n^{peak}|^2 = 2|s_{n-1}^{peak}|^2$. Therefore, in summary, we have $\text{PAPR}_n = \text{PAPR}_{n-1} + 3$ dB. We note that the probability of having $\pm s_n^{peak}$ in simulation is given by $2/\mathcal{L}_n$, which also decreases dramatically with n as proven by Proposition 1. Thus the theory PAPR is not fully reflected by Fig. 3 in [19]. $\qquad \square$

REFERENCES

[1] L. Hanzo and R. Steele, "The pan-European mobile radio system: Part I," *Eur. Trans. Telecommun.*, vol. 5, no. 2, pp. 245–260, 1994.

[2] L. Hanzo and R. Steele, "The Pan-European mobile radio system: Part II," *Eur. Trans. Telecommun.*, vol. 5, no. 2, pp. 261–276, 1994.

[3] R. Steele and L. Hanzo, *Mobile Radio Communications*. Hoboken, NJ, USA: Wiley, 1999.

[4] L. Hanzo, C. H. Wong, and M. S. Yee, *Adaptive Wireless Transceivers: Turbo-Coded, Turbo-Equalized and Space-Time Coded TDMA, CDMA and OFDM Systems*. Hoboken, NJ, USA: Wiley, 2002.

[5] L. Hanzo, M. Munster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. Hoboken, NJ, USA: Wiley, 2003.

[6] L. Hanzo, L.-L. Yang, E. L. Kuan, and K. Yen, *Single and Multi-Carrier DS-CDMA: Multi-User Detection, Space-Time Spreading, Synchronisation, Networking and Standards*. Hoboken, NJ, USA: Wiley, 2004.

[7] L. Hanzo, J. S. Blogh, and S. Ni, *3G, HSPA and FDD Versus TDD Networking*. Hoboken, NJ, USA: Wiley, 2008.

[8] L. Hanzo and T. Keller, *OFDM and MC-CDMA: A Primer*. Hoboken, NJ, USA: Wiley, 2007.

[9] L.-L. Yang, *Multicarrier Communications*. Hoboken, NJ, USA: Wiley, 2009.

[10] L. Hanzo, Y. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent Versus Non-Coherent and Cooperative Turbo Transceivers*. Hoboken, NJ, USA: Wiley, 2011.

[11] Z. Zhang, X. Wang, K. Long, A. V. Vasilakos, and L. Hanzo, "Large-scale MIMO-based wireless backhaul in 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 58–66, Oct. 2015.

[12] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[13] L. Wan, Z. Guo, Y. Wu, W. Bi, J. Yuan, M. Elkashlan, and L. Hanzo, "4G/5G spectrum sharing: Efficient 5G deployment to serve enhanced mobile broadband and Internet of Things applications," *IEEE Veh. Technol. Mag.*, vol. 13, no. 4, pp. 28–39, Dec. 2018.

[14] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.

[15] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart., 2018.

[16] *Summary of Rel-15 Work Items*, document 21.915, 3GPP. 2019. [Online]. Available: http://www.3gpp.org/release-15

[17] L. Wang and L. Hanzo, "Dispensing with channel estimation: Differentially modulated cooperative wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 3, pp. 836–857, 3rd Quart., 2012.

[18] 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 Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 689–712, 2nd Quart., 2014.

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

[20] C. Xu, P. Zhang, R. Rajashekar, N. Ishikawa, S. Sugiura, L. Wang, and L. Hanzo, "Finite-cardinality single-RF differential space-time modulation for improving the diversity-throughput tradeoff," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 318–335, Jan. 2019.

[21] C. Xu, J. Zhang, T. Bai, P. Botsinis, R. G. Maunder, R. Zhang, and L. Hanzo, "Adaptive coherent/non-coherent single/multiple-antenna aided channel coded ground-to-air aeronautical communication," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1099–1116, Feb. 2019.

[22] L. Hanzo, S. X. Ng, T. Keller, and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. Hoboken, NJ, USA: Wiley, 2004.

[23] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation*. Hoboken, NJ, USA: Wiley, 2009.

[24] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[25] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2338–2353, Aug. 2002.

[26] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[27] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[28] J. P. McGeehan and A. J. Bateman, "Phase-locked transparent tone-in-band (TTIB): A new spectrum configuration particularly suited to the transmission of data over SSB mobile radio networks," *IEEE Trans. Commun.*, vol. COM-32, no. 1, pp. 81–87, Jan. 1984.

[29] M. L. Moher and J. H. Lodge, "TCMP-a modulation and coding strategy for Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 9, pp. 1347–1355, Dec. 1989.

[30] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[31] T. Kailath, "Correlation detection of signals perturbed by a random channel," *Proc. IRE*, vol. 6, no. 3, pp. 361–366, Jun. 1960.

[32] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 350–361, May 1969.

[33] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 363–378, May 1972.

[34] R. Morley and D. Snyder, "Maximum likelihood sequence estimation for randomly dispersive channels," *IEEE Trans. Commun.*, vol. COM-27, no. 6, pp. 833–839, Jun. 1979.

[35] J. G. Lawton, "Theoretical error rates of differentially coherent binary and Kineplex data transmission systems," *Proc. IRE*, vol. 47, no. 2, pp. 333–334, 1959.

[36] J. G. Lawton, "Investigation of digital data communication systems," Cornell Aeronaut. Lab., Buffalo, NY, USA, Tech. Rep. CAL-UA-1420-S-1, Jan. 1961.

[37] C. R. Cahn, "Performance of digital phase-modulation communication systems," *Proc. IRE*, vol. 7, no. 1, pp. 3–6, May 1959.

[38] P. Bello and B. D. Nelin, "The influence of fading spectrum on the binary error probabilites of incoherent and differentially coherent matched filter recievers," *Proc. IRE*, vol. 10, no. 2, pp. 160–168, Jun. 1962.

[39] S. G. Wilson, J. Freebersyser, and C. Marshall, "Multi-symbol detection of M-DPSK," in *Proc. IEEE Global Telecommun. Conf. Exhibit. (GLOBECOM)*, Dallas, TX, USA, vol. 3, Nov. 1989, pp. 1692–1697.

[40] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 300–308, Mar. 1990.

[41] P. Ho and D. Fung, "Error performance of multiple-symbol differential detection of PSK signals transmitted over correlated Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 40, no. 10, pp. 1566–1569, Oct. 1992.

[42] D. Divsalar and M. K. Simon, "Maximum-likelihood differential detection of uncoded and trellis coded amplitude phase modulation over AWGN and fading channels-metrics and performance," *IEEE Trans. Commun.*, vol. 42, no. 1, pp. 76–89, Jan. 1994.

[43] H. Leib and S. Pasupathy, "The phase of a vector perturbed by Gaussian noise and differentially coherent receivers," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 6, pp. 1491–1501, Nov. 1988.

[44] F. Edbauer, "Bit error rate of binary and quaternary DPSK signals with multiple differential feedback detection," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 457–460, Mar. 1992.

[45] F. Adachi and M. Sawahashi, "Decision feedback multiple-symbol differential detection for M-ary DPSK," *Electron. Lett.*, vol. 29, no. 15, pp. 1385–1387, Jul. 1993.

[46] H. Leib, "Data-aided noncoherent demodulation of DPSK," *IEEE Trans. Commun.*, vol. 43, nos. 2–4, pp. 722–725, Feb. 1995.

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

[48] L. H.-J. Lampe and R. Schober, "Iterative decision-feedback differential demodulation of bit-interleaved coded MDPSK for flat Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1176–1184, Jul. 2001.

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

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

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

[52] F. Adachi and M. Sawahashi, "Viterbi-decoding differential detection of DPSK," *Electron. Lett.*, vol. 28, no. 23, pp. 2196–2198, Nov. 1992.

[53] P. Hoeher and J. Lodge, "'Turbo DPSK': Iterative differential PSK demodulation and channel decoding," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 837–843, Jun. 1999.

[54] G. M. Vitetta and D. P. Taylor, "Viterbi decoding of differentially encoded PSK signals transmitted over Rayleigh frequency-flat fading channels," *IEEE Trans. Commun.*, vol. 43, nos. 2–4, pp. 1256–1259, Feb. 1995.

[55] F. Adachi, "Reduced state transition Viterbi differential detection of M-ary DPSK signals," *Electron. Lett.*, vol. 32, no. 12, pp. 1064–1066, Jun. 1996.

[56] G. Colavolpe and R. Raheli, "Noncoherent sequence detection," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1376–1385, Sep. 1999.

[57] R. Schober and W. H. Gerstacker, "Decision-feedback differential detection based on linear prediction for MDPSK signals transmitted over Ricean fading channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 391–402, Mar. 2000.

[58] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.

[59] A. A. Boulogeorgos, A. Alexiou, T. Merkle, C. Schubert, R. Elschner, A. Katsiotis, P. Stavrianos, D. Kritharidis, P. Chartsias, J. Kokkoniemi, M. Juntti, J. Lehtomaki, A. Teixeira, and F. Rodrigues, "Terahertz technologies to deliver optical network quality of experience in wireless systems beyond 5G," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 144–151, Jun. 2018.

[60] C. Lin and G. Y. Li, "Terahertz communications: An array-of-subarrays solution," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 124–131, Dec. 2016.

[61] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Phys. Commun.*, vol. 12, pp. 16–32, Sep. 2014.

[62] P. Y. Chen, C. Argyropoulos, and A. Alù, "Terahertz antenna phase shifters using integrally-gated graphene transmission-lines," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1528–1537, Apr. 2013.

[63] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714–2741, 4th Quart., 2018.

[64] C. Xu, T. Bai, J. Zhang, R. Rajashekar, R. G. Maunder, Z. Wang, and L. Hanzo, "Adaptive coherent/non-coherent spatial modulation aided unmanned aircraft systems," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 170–177, Aug. 2019.

[65] S. H. Blumenthal, "Medium earth orbit Ka band satellite communications system," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2013, pp. 273–277.

[66] J. Zhang, T. Chen, S. Zhong, J. Wang, W. Zhang, X. Zuo, R. G. Maunder, and L. Hanzo, "Aeronautical ad hoc networking for the Internet-above-the-clouds," *Proc. IEEE*, vol. 107, no. 5, pp. 868–911, May 2019.

[67] J. Wu and P. Fan, "A survey on high mobility wireless communications: Challenges, opportunities and solutions," *IEEE Access*, vol. 4, pp. 450–476, 2016.

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

[69] P. Yang, Y. Xiao, Y. L. Guan, K. V. S. Hari, A. Chockalingam, S. Sugiura, H. Haas, M. D. Renzo, C. Masouros, Z. Liu, L. Xiao, S. Li, and L. Hanzo, "Single-carrier SM-MIMO: A promising design for broadband large-scale antenna systems," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1687–1716, 3rd Quart., 2016.

[70] S. Song, Y. Yang, Q. Xionq, K. Xie, B. J. Jeong, and B. Jiao, "A channel hopping technique I: Theoretical studies on band efficiency and capacity," in *Proc. Int. Conf. Commun., Circuits Syst. (ICCCAS)*, Chengdu, China, vol. 1, Jun. 2004, pp. 229–233.

[71] Y. A. Chau and S. H. Yu, "Space modulation on wireless fading channels," in *Proc. IEEE VTS 54th Veh. Technol. Conf. (VTC Fall)*, Atlantic City, NJ, USA, vol. 3, Oct. 2001, pp. 1668–1671.

[72] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, Jul. 2008.

[73] E. Başar, U. Aygolu, E. Panayirci, and H. V. Poor, "Space-time block coded spatial modulation," *IEEE Trans. Commun.*, vol. 59, no. 3, pp. 823–832, Mar. 2011.

[74] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Generalized space shift keying modulation for MIMO channels," in *Proc. IEEE 19th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Cannes, France, Sep. 2008, pp. 1–5.

[75] J. Wang, S. Jia, and J. Song, "Generalised spatial modulation system with multiple active transmit antennas and low complexity detection scheme," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1605–1615, Apr. 2012.

[76] Y. Xiao, Z. Yang, L. Dan, P. Yang, L. Yin, and W. Xiang, "Low-complexity signal detection for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 403–406, Mar. 2014.

[77] S. Sugiura, T. Ishihara, and M. Nakao, "State-of-the-art design of index modulation in the space, time, and frequency domains: Benefits and fundamental limitations," *IEEE Access*, vol. 5, pp. 21774–21790, 2017.

[78] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693–16746, 2017.

[79] 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 Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1905–1938, 3rd Quart., 2018.

[80] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[81] Z. Zhang, X. Chai, K. Long, A. V. Vasilakos, and L. Hanzo, "Full duplex techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 128–137, May 2015.

[82] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[83] G. Liu, F. R. Yu, H. Ji, V. C. M. Leung, and X. Li, "In-band full-duplex relaying: A survey, research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 500–524, 2nd Quart., 2015.

[84] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2017–2046, 4th Quart., 2015.

[85] L.-L. Yang, "Multicarrier-division duplexing (MDD): A duplexing scheme whose time has come," *IEEE COMSOC MMTC E-Lett.*, vol. 10, no. 1, pp. 35–38, Jan. 2015.

[86] W. Qiu, P. Tarasak, and H. Minn, "Orthogonal multicarrier division multiple access for multipoint-to-multipoint networks," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3841–3853, Sep. 2013.

[87] R. Rajashekar, C. Xu, N. Ishikawa, L.-L. Yang, and L. Hanzo, "Multi-carrier division duplex aided millimeter wave communications," *IEEE Access*, vol. 7, pp. 100719–100732, 2019.

[88] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1–5.

[89] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[90] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[91] C. Xu, S. Sugiura, S. X. Ng, P. Zhang, L. Wang, and L. Hanzo, "Two decades of MIMO design tradeoffs and reduced-complexity MIMO detection in near-capacity systems," *IEEE Access*, vol. 5, pp. 18564–18632, 2017.

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

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

[94] A. M. Akhtar, X. Wang, and L. Hanzo, "Synergistic spectrum sharing in 5G HetNets: A harmonized SDN-enabled approach," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 40–47, Jan. 2016.

[95] K. Liang, G. Liu, L. Zhao, X. Chu, S. Wang, and L. Hanzo, "Performance analysis of cellular radio access networks relying on control- and user-plane separation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7241–7245, Jul. 2019.

[96] J. Yang, B. Yang, S. Chen, Y. Zhang, Y. Zhang, and L. Hanzo, "Dynamic resource allocation for streaming scalable videos in SDN-aided dense small-cell networks," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2114–2129, Mar. 2019.

[97] H. Zhang, N. Liu, K. Long, J. Cheng, V. C. M. Leung, and L. Hanzo, "Energy efficient subchannel and power allocation for software-defined heterogeneous VLC and RF networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 658–670, Mar. 2018.

[98] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[99] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392–2431, 4th Quart., 2017.

[100] T. Park, N. Abuzainab, and W. Saad, "Learning how to communicate in the Internet of Things: Finite resources and heterogeneity," *IEEE Access*, vol. 4, pp. 7063–7073, 2016.

[101] M. A. Alsheikh, S. Lin, D. Niyato, and H. P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1996–2018, 4th Quart., 2014.

[102] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1136–1159, 3rd Quart., 2013.

[103] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.

[104] *Service Requirements for the 5G System*, document 22.261, 3GPP, 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/22_series/22.261/

[105] *User Equipment (UE) Radio Transmission and Reception*, document 36.101, 3GPP, 2019. http://www.3gpp.org/ftp//Specs/archive/36_series/36.101/

[106] *User Equipment (UE) Radio Transmission and Reception Part1: Range 1 Standalone*, document, 3GPP, 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.101-1/

[107] *User Equipment (UE) Radio Transmission and Reception Part2: Range 2 Standalone*, document, 3GPP, 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.101-2/

[108] *NR Physical Channels and Modulation (Release 15)*, document 38.211, 3GPP, 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.211/

[109] *Requirements for Support of Radio Resource Management*, document 38.133, 3GPP, 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.133/

[110] *Physical Layer Procedures for Control*, document 38.213, 3GPP, 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.213/

[111] *Physical Layer Procedures for Data*, document 38.214, 3GPP, 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38_series/38.214/

[112] *Technical Specification Group Radio Access Network; Study on Enhanced LTE Support for Aerial Vehicles (Release 15)*, document 36.777, 3GPP, Jan. 2019. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/36_series/36.777/

[113] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1169–1174, Jul. 2000.

[114] H. Jafarkhani and V. Tarokh, "Multiple transmit antenna differential detection from generalized orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2626–2631, Sep. 2001.

[115] G. Ganesan and P. Stoica, "Differential modulation using space-time block codes," *IEEE Signal Process. Lett.*, vol. 9, no. 2, pp. 57–60, Feb. 2002.

[116] X.-G. Xia, "Differentially en/decoded orthogonal space-time block codes with APSK signals," *IEEE Commun. Lett.*, vol. 6, no. 4, pp. 150–152, Apr. 2002.

[117] M. Tao and R. S. Cheng, "Differential space-time block codes," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 2, Nov. 2001, pp. 1098–1102.

[118] C.-S. Hwang, S. H. Nam, J. Chung, and V. Tarokh, "Differential space time block codes using nonconstant modulus constellations," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2955–2964, Nov. 2003.

[119] B. Hassibi and B. M. Hochwald, "Cayley differential unitary space-time codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1485–1503, Jun. 2002.

[120] H. E. Gamal and M. O. Damen, "Universal space-time coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.

[121] J. Wang, X. Wang, and M. Madihian, "Design of minimum error-rate Cayley differential unitary space-time codes," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1779–1787, Sep. 2005.

[122] F. Oggier and B. Hassibi, "Algebraic Cayley differential space-time codes," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1911–1919, May 2007.

[123] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3219–3230, Nov. 2010.

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

[125] C. Xu, S. Sugiura, S. X. Ng, and L. Hanzo, "Reduced-complexity non-coherently detected differential space-time shift keying," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 153–156, Mar. 2011.

[126] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[127] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[128] B. L. Hughes, "Optimal space-time constellations from groups," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 401–410, Feb. 2003.

[129] A. Shokrollahi, B. Hassibi, B. M. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2335–2367, Sep. 2001.

[130] Y. Bian, M. Wen, X. Cheng, H. Poor, and B. Jiao, "A differential scheme for spatial modulation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3925–3930.

[131] Y. Bian, X. Cheng, M. Wen, L. Yang, H. V. Poor, and B. Jiao, "Differential spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 64, no. 7, pp. 3262–3268, Jul. 2015.

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

[133] J. Liu, L. Dan, P. Yang, L. Xiao, F. Yu, and Y. Xiao, "High-rate APSK-aided differential spatial modulation: Design method and performance analysis," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 168–171, Jan. 2017.

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

[135] N. Ishikawa, R. Rajashekar, C. Xu, S. Sugiura, and L. Hanzo, "Differential space-time coding dispensing with channel-estimation approaches the performance of its coherent counterpart in the open-loop massive MIMO-OFDM downlink," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6190–6204, Dec. 2018.

[136] N. Ishikawa, R. Rajashekar, C. Xu, M. El-Hajjar, S. Sugiura, L.-L. Yang, and L. Hanzo, "Differential-detection aided large-scale generalized spatial modulation is capable of operating in high-mobility millimeter-wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1360–1374, Oct. 2019.

[137] C. Xu, P. Zhang, R. Rajashekar, N. Ishikawa, S. Sugiura, Z. Wang, and L. Hanzo, "'Near-perfect' finite-cardinality generalized space-time shift keying," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2146–2164, Sep. 2019.

[138] C. Xu, T. Bai, J. Zhang, R. G. Maunder, S. Sugiura, Z. Wang, and L. Hanzo, "Constant-envelope space-time shift keying," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1387–1402, Oct. 2019.

[139] S. Cherry, "Edholm's law of bandwidth," *IEEE Spectr.*, vol. 41, no. 7, pp. 58–60, Jul. 2004.

[140] T. Kürner and S. Priebe, "Towards THz communications—Status in research, standardization and regulation," *J. Infr., Millim., THz Waves*, vol. 35, no. 1, pp. 53–62, Jan. 2014.

[141] K. Roberts, Q. Zhuge, I. Monga, S. Gareau, and C. Laperle, "Beyond 100 Gb/s: Capacity, flexibility, and network optimization," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 9, no. 4, pp. C12–C23, Apr. 2017.

[142] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[143] W. Roh, J. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[144] B. Y. Toh, R. Cahill, and V. F. Fusco, "Understanding and measuring circular polarization," *IEEE Trans. Educ.*, vol. 46, no. 3, pp. 313–318, Aug. 2003.

[145] J. Joung, C. K. Ho, K. Adachi, and S. Sun, "A survey on power-amplifier-centric techniques for spectrum- and energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 315–333, 1st Quart., 2015.

[146] J. Joung, C. K. Ho, and S. Sun, "Spectral efficiency and energy efficiency of OFDM systems: Impact of power amplifiers and countermeasures," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 208–220, Feb. 2014.

[147] N. P. Lawrence, B. W. Ng, H. J. Hansen, and D. Abbott, "5G terrestrial networks: Mobility and coverage—Solution in three dimensions," *IEEE Access*, vol. 5, pp. 8064–8093, 2017.

[148] M. R. Andrews, P. P. Motra, and R. de Carvalho, "Tripling the capacity of wireless communications using electromagnetic polarization," *Nature*, vol. 409, pp. 316–318, Jan. 2001.

[149] C. Y. Chiu, J. B. Yan, and R. D. Murch, "Compact three-port orthogonally polarized MIMO antennas," *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 619–622, 2007.

[150] N. P. Lawrence, C. Fumeaux, and D. Abbott, "Planar triorthogonal diversity slot antenna," *IEEE Trans. Antennas Propag.*, vol. 65, no. 3, pp. 1416–1421, Mar. 2017.

[151] N. P. Lawrence, B. W. H. Ng, H. J. Hansen, and D. Abbott. *Analysis of Millimetre-Wave Polarization Diverse Multiple-Input Multiple-Output Capacity*. Accessed: Dec. 10, 2019. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rsos.150322

[152] A. M. D. Turkmani, A. Arowojolu, P. A. Jefford, and C. J. Kellett, "An experimental evaluation of the performance of two-branch space and polarization diversity schemes at 1800 MHz," *IEEE Trans. Veh. Technol.*, vol. 44, no. 2, pp. 318–326, May 1995.

[153] J. W. Wallace and M. A. Jensen, "Mutual coupling in MIMO wireless systems: A rigorous network theory analysis," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1317–1325, Jul. 2004.

[154] Ofcom. *Update on 5G Spectrum in the U.K.* Accessed: Dec. 10, 2019. [Online]. Available: https://www.ofcom.org.uk/_data/assets/pdf_file/0021/97023/5G-update-08022017.pdf

[155] Qualcomm. *Spectrum for 4G and 5G*. Accessed: Dec. 10, 2019. [Online]. Available: https://www.qualcomm.com/media/documents/files/spectrum-for-4g-and-5g.pdf

[156] Ofcom. *Whitepaper on the VR-Oriented Bearer Network Requirement*. Accessed: Dec. 10, 2019. [Online]. Available: https://www.huawei.com/en/press-events/news/2016/11/whitepaper-vr-oriented-b%earer-network-requirements

[157] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution: From Theory to Practice*. Hoboken, NJ, USA: Wiley, 2011.

[158] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[159] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[160] M. S. Faruk and S. J. Savory, "Digital signal processing for coherent transceivers employing multilevel formats," *J. Lightw. Technol.*, vol. 35, no. 5, pp. 1125–1141, Mar. 1, 2017.

[161] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *J. Lightw. Technol.*, vol. 34, no. 1, pp. 157–179, Jun. 1, 2016.

[162] J. Zhang, P. Lei, S. Hu, M. Zhu, Z. Yu, B. Xu, and K. Qiu, "Functional-link neural network for nonlinear equalizer in coherent optical fiber communications," *IEEE Access*, vol. 7, pp. 149900–149907, 2019.

[163] R. Dar and P. J. Winzer, "Nonlinear interference mitigation: Methods and potential gain," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 903–930, Feb. 15, 2017.

[164] X. Liu, A. R. Chraplyvy, P. J. Winzer, R. W. Tkach, and S. Chandrasekhar, "Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit," *Nature Photon.*, vol. 7, pp. 560–568, Jul. 2013.

[165] A. Leven, N. Kaneda, U. V. Koc, and Y. K. Chen, "Frequency estimation in intradyne reception," *IEEE Photon. Technol. Lett.*, vol. 19, no. 6, pp. 366–368, Mar. 15, 2007.

[166] X. Zhou, L. E. Nelson, P. Magill, R. Isaac, B. Zhu, D. W. Peckham, P. I. Borel, and K. Carlson, "High spectral efficiency 400 Gb/s transmission using PDM time-domain hybrid 32–64 QAM and training-assisted carrier recovery," *J. Lightw. Technol.*, vol. 31, no. 7, pp. 999–1005, Feb. 15, 2013.

[167] X. Zhou and L. Nelson, "Advanced DSP for 400 Gb/s and beyond optical networks," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2716–2725, Aug. 15, 2014.

[168] M. A. Esmail, A. M. Ragheb, H. A. Fathallah, M. Altamimi, and S. A. Alshebeili, "5G-28 GHz signal transmission over hybrid all-optical FSO/RF link in dusty weather conditions," *IEEE Access*, vol. 7, pp. 24404–24410, 2019.

[169] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[170] J. G. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 1995.

[171] B. Sklar, *Digital Communications*. Upper Saddle River, NJ, USA: Prentice-Hall, 1988.

[172] M. Schwartz, R. B. William, and S. Stein, *Communication Systems and Techniques*. Hoboken, NJ, USA: Wiley, 1995.

[173] E. D. Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. 1st IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun.*, Paris, France, Apr. 1997, pp. 129–132.

[174] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about training: Guidelines derived using the affine precoding framework and the CRB," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 940–954, Mar. 2006.

[175] T. Masamura, S. Samejima, Y. Morihiro, and H. Fuketa, "Differential detection of MSK with nonredundant error correction," *IEEE Trans. Commun.*, vol. COM-27, no. 6, pp. 912–918, Jun. 1979.

[176] S. Samejima, K. Enomoto, and Y. Watanabe, "Differential PSK system with nonredundant error correction," *IEEE J. Sel. Areas Commun.*, vol. SAC-1, no. 1, pp. 74–81, Jan. 1983.

[177] K. M. Mackenthun, "A fast algorithm for multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 42, nos. 2–4, pp. 1471–1474, Feb./Apr. 1994.

[178] R.-R. Chen, R. Koetter, U. Madhow, and D. Agrawal, "Joint noncoherent demodulation and decoding for the block fading channel: A practical framework for approaching Shannon capacity," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1676–1689, Oct. 2003.

[179] A. Svensson, "Coherent detector based on linear prediction and decision feedback for DQPSK," *Electron. Lett.*, vol. 30, no. 20, pp. 1642–1643, Sep. 1994.

[180] L. Bin and P. Ho, "Data-aided linear prediction receiver for coherent DPSK and CPM transmitted over Rayleigh flat-fading channels," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1229–1236, Jul. 1999.

[181] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

[182] M. Simon and J. Smith, "Carrier synchronization and detection of QASK signal sets," *IEEE Trans. Commun.*, vol. 22, no. 2, pp. 98–106, Feb. 1974.

[183] W. Weber, "Differential encoding for multiple amplitude and phase shift keying systems," *IEEE Trans. Commun.*, vol. COM-26, no. 3, pp. 385–391, Mar. 1978.

[184] M. K. Simon, G. K. Huth, and A. Polydoros, "Differentially coherent detection of QASK for frequency-hopping systems—Part I: Performance in the presence of a Gaussian noise environment," *IEEE Trans. Commun.*, vol. COM-30, no. 1, pp. 158–164, Jan. 1982.

[185] L. H. J. Lampe and R. Schober, "Low-complexity iterative demodulation for noncoherent coded transmission over Ricean-fading channels," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1481–1496, Nov. 2001.

[186] S. H. Nam, C. S. Hwang, J. Chung, and V. Tarokh, "Differential space time block codes using QAM for four transmit antennas," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2, Jul. 2004, pp. 952–956.

[187] W. T. Webb, L. Hanzo, and R. Steele, "Bandwidth efficient QAM schemes for Rayleigh fading channels," *IEE Proc. I-Commun., Speech Vis.*, vol. 138, no. 3, pp. 169–175, Jun. 1991.

[188] F. Adachi and M. Sawahashi, "Performance analysis of various 16 level modulation schemes under Rayleigh fading," *Electron. Lett.*, vol. 28, no. 17, pp. 1579–1581, Aug. 1992.

[189] Y. C. Chow, A. R. Nix, and J. P. McGeehan, "Analysis of 16-APSK modulation in AWGN and Rayleigh fading channel," *Electron. Lett.*, vol. 28, no. 17, pp. 1608–1610, Aug. 1992.

[190] H. Rohling and V. Engels, "Differential amplitude phase shift keying (DAPSK)—A new modulation method for DTVB," in *Proc. IET Int. Broadcast. Conv. (IBC)*, Sep. 1995, pp. 102–108.

[191] W. C. Dam and D. P. Taylor, "An adaptive maximum likelihood receiver for correlated Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 42, no. 9, pp. 2684–2692, Sep. 1994.

[192] G. M. Vitetta and D. P. Taylor, "Maximum likelihood decoding of uncoded and coded PSK signal sequences transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 43, no. 11, pp. 2750–2758, Nov. 1995.

[193] Y. C. Chow, A. R. Nix, and J. P. McGeehan, "Diversity improvement for 16-DAPSK in Rayleigh fading channel," *Electron. Lett.*, vol. 29, no. 4, pp. 387–389, Feb. 1993.

[194] C.-D. Chung, "Differentially amplitude and phase-encoded QAM for the correlated Rayleigh-fading channel with diversity reception," *IEEE Trans. Commun.*, vol. 45, no. 3, pp. 309–321, Mar. 1997.

[195] F. Adachi and M. Sawahashi, "Decision feedback differential detection of differentially encoded 16APSK signals," *IEEE Trans. Commun.*, vol. 44, no. 4, pp. 416–418, Apr. 1996.

[196] R. Y. Wei and M. C. Lin, "Modified decision feedback differential detection for differentially encoded 16 APSK signals," *Electron. Lett.*, vol. 34, no. 4, pp. 336–337, Feb. 1998.

[197] R. Schober, W. H. Gerstacker, and J. B. Huber, "Decision-feedback differential detection scheme for 16-DAPSK," *Electron. Lett.*, vol. 34, no. 19, pp. 1812–1813, Sep. 1998.

[198] W. H. Gerstacker, R. Schober, and J. B. Huber, "Decision-feedback differential detection for 16DAPSK transmitted over Rician fading channels," in *Proc. IEEE VTS 50th Veh. Technol. Conf. (VTC-Fall).*, Amsterdam, The Netherlands, vol. 5, Sep. 1999, pp. 2515–2519.

[199] R. Schober, W. H. Gerstacker, and J. B. Huber, "Decision-feedback differential detection based on linear prediction for 16DAPSK signals transmitted over flat Ricean fading channels," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1339–1342, Aug. 2001.

[200] T. May, H. Rohling, and V. Engels, "Performance analysis of Viterbi decoding for 64-DAPSK and 64-QAM modulated OFDM signals," *IEEE Trans. Commun.*, vol. 46, no. 2, pp. 182–190, Feb. 1998.

[201] R. F. H. Fischer, L. H. J. Lampe, and S. H. Muller-Weinfurtner, "Coded modulation for noncoherent reception with application to OFDM," *IEEE Trans. Veh. Technol.*, vol. 50, no. 4, pp. 910–919, Jul. 2001.

[202] K. Ishibashi, H. Ochiai, and R. Kohno, "Low-complexity bit-interleaved coded DAPSK for Rayleigh-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1728–1738, Sep. 2005.

[203] D. Liang, S. X. Ng, and L. Hanzo, "Soft-decision star-QAM aided BICM-ID," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 169–172, Jan. 2011.

[204] C. Xu, D. Liang, S. X. Ng, and L. Hanzo, "Reduced-complexity noncoherent soft-decision-aided DAPSK dispensing with channel estimation," *IEEE Trans. Veh. Technol.*, vol. 62, no. 6, pp. 2633–2643, Jul. 2013.

[205] L. Wang and L. Hanzo, "Low-complexity near-optimum multiple-symbol differential detection of DAPSK based on iterative amplitude/phase processing," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 894–900, Jan. 2012.

[206] C. Xu, S. X. Ng, and L. Hanzo, "Multiple-symbol differential sphere detection and decision-feedback differential detection conceived for differential QAM," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8345–8360, Oct. 2016.

[207] C. Xu, L. Wang, S. X. Ng, and L. Hanzo, "Soft-decision multiple-symbol differential sphere detection and decision-feedback differential detection for differential QAM dispensing with channel estimation in the face of rapidly fading channels," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4408–4425, Jun. 2016.

[208] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Feb. 1996.

[209] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. SAC-16, no. 8, pp. 1451–1458, Oct. 1998.

[210] H. E. Gamal and A. R. Hammons, "A new approach to layered space-time coding and signal processing," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2321–2334, Sep. 2001.

[211] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. Int. Symp. Signals, Syst., Electron. (ISSSE)*, Pisa, Italy, Sep. 1998, pp. 295–300.

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

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

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

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

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

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

[218] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.

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

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

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

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

[223] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

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

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

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

[227] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.

[228] C. B. Papadias and G. J. Foschini, "A space-time coding approach for systems employing four transmit antennas," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Salt Lake City, UT, USA, vol. 4, May 2001, pp. 2481–2484.

[229] C. B. Papadias and G. J. Foschini, "Capacity-approaching space-time codes for systems employing four transmitter antennas," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 726–732, Mar. 2003.

[230] N. Sharma and C. B. Papadias, "Improved quasi-orthogonal codes through constellation rotation," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 332–335, Mar. 2003.

[231] W. Su and X.-G. Xia, "Signal constellations for quasi-orthogonal space-time block codes with full diversity," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2331–2347, Oct. 2004.

[232] H.-F. Lu and P. V. Kumar, "A unified construction of space-time codes with optimal rate-diversity tradeoff," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1709–1730, May 2005.

[233] C. Yuen, Y. L. Guan, and T. T. Tjhung, "Quasi-orthogonal STBC with minimum decoding complexity," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2089–2094, Sep. 2005.

[234] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[235] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, Jul. 2002.

[236] R. W. Heath, Jr., and A. Paulraj, "Linear dispersion codes for MIMO systems based on frame theory," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2429–2441, Oct. 2002.

[237] R. H. Gohary and T. N. Davidson, "Design of linear dispersion codes: Asymptotic guidelines and their implementation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2892–2906, Nov. 2005.

[238] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu, "Explicit space–time codes achieving the diversity–multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3869–3884, Sep. 2006.

[239] J. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: A 2×2 full-rate space-time code with nonvanishing determinants," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1432–1436, Apr. 2005.

[240] F. Oggier, G. Rekaya, J.-C. Belfiore, and E. Viterbo, "Perfect space-time block codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3885–3902, Sep. 2006.

[241] P. Elia, B. A. Sethuraman, and P. V. Kumar, "Perfect space–time codes for any number of antennas," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 3853–3868, Nov. 2007.

[242] K. P. Srinath and B. S. Rajan, "Improved perfect space-time block codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7927–7935, Dec. 2013.

[243] K. P. Srinath and B. S. Rajan, "Fast-decodable MIDO codes with large coding gain," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 992–1007, Feb. 2014.

[244] C. Xu, S. Sugiura, S. X. Ng, and L. Hanzo, "Spatial modulation and space-time shift keying: Optimal performance at a reduced detection complexity," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 206–216, Jan. 2013.

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

[246] J. Fu, C. Hou, W. Xiang, L. Yan, and Y. Hou, "Generalised spatial modulation with multiple active transmit antennas," in *Proc. IEEE GLOBECOM Workshops*, Miami, FL, USA, Dec. 2010, pp. 839–844.

[247] A. Younis, N. Serafimovski, R. Mesleh, and H. Haas, "Generalized spatial modulation," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2010, pp. 1498–1502.

[248] S. Sugiura, S. Chen, and L. Hanzo, "Generalized space-time shift keying designed for flexible diversity-, multiplexing- and complexity-tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1144–1153, Apr. 2011.

[249] S. Sugiura, C. Xu, S. X. Ng, and L. Hanzo, "Reduced-complexity iterative-detection-aided generalized space-time shift keying," *IEEE Trans. Veh. Technol.*, vol. 61, no. 8, pp. 3656–3664, Oct. 2012.

[250] A. Stavridis, S. Sinanovic, M. Di Renzo, H. Haas, and P. Grant, "An energy saving base station employing spatial modulation," in *Proc. IEEE 17th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Barcelona, Spain, Aug. 2012, pp. 231–235.

[251] A. Stavridis, S. Sinanovic, M. Di Renzo, and H. Haas, "Energy evaluation of spatial modulation at a multi-antenna base station," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–5.

[252] S. Sugiura and L. Hanzo, "Single-RF spatial modulation requires single-carrier transmission: Frequency-domain turbo equalization for dispersive channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4870–4875, Oct. 2015.

[253] L. Da Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[254] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of Things in the 5G era: Enablers, architecture, and business models," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016.

[255] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. V. Der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, Nov. 2010.

[256] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 3, pp. 814–826, Mar. 2002.

[257] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, 4th Quart., 2011.

[258] I. Herstein, *Topics in Algebra*. Hoboken, NJ, USA: Wiley, 1975.

[259] X.-B. Liang and X.-G. Xia, "Unitary signal constellations for differential space-time modulation with two transmit antennas: Parametric codes, optimal designs, and bounds," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2291–2322, Aug. 2002.

[260] Y. Jing and B. Hassibi, "Design of fully diverse multiple-antenna codes based on Sp(2)," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2639–2656, Nov. 2004.

[261] Y. Jing and B. Hassibi, "Three-transmit-antenna space-time codes based on SU(3)," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3688–3702, Oct. 2005.

[262] J. Abarbanel, A. Averbuch, S. Rosset, and J. Zlotnick, "Unitary nongroup STBC from cyclic algebras," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3903–3912, Sep. 2006.

[263] F. Oggier, "Cyclic algebras for noncoherent differential space–time coding," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3053–3065, Sep. 2007.

[264] K. L. Clarkson, W. Sweldens, and A. Zheng, "Fast multiple-antenna differential decoding," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 253–261, Feb. 2001.

[265] N. Ishikawa and S. Sugiura, "Unified differential spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 337–340, Aug. 2014.

[266] 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 Trans. Veh. Technol.*, vol. 66, no. 1, pp. 385–394, Jan. 2017.

[267] M. R. Bhatnagar, A. Hjorungnes, and L. Song, "Differential coding for non-orthogonal space-time block codes with non-unitary constellations over arbitrarily correlated Rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3985–3995, Aug. 2009.

[268] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore, "Diagonal algebraic space-time block codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 628–636, Mar. 2002.

[269] B. A. Sethuraman, B. S. Rajan, and V. Shashidhar, "Full-diversity, high-rate space-time block codes from division algebras," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2596–2616, Oct. 2003.

[270] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

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

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

[273] R. Schober and L. H. J. Lampe, "Noncoherent receivers for differential space-time modulation," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 768–777, May 2002.

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

[275] C. Yuen, Y. L. Guan, and T. T. Tjhung, "Unitary differential space–time modulation with joint modulation," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3937–3944, Nov. 2007.

[276] C. Xu, L. Wang, S. X. Ng, and L. Hanzo, "Multiple-symbol differential sphere detection aided differential space-time block codes using QAM constellations," *IEEE Signal Process. Lett.*, vol. 18, no. 9, pp. 497–500, Sep. 2011.

[277] 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 Trans. Commun.*, vol. 65, no. 10, pp. 4260–4273, Oct. 2017.

[278] H. Sui and J. R. Zeidler, "Demodulation and performance analysis of differential unitary space-time modulation in time-varying Rician channels," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2006, pp. 1–5.

[279] T. Cui and C. Tellambura, "On multiple symbol detection for diagonal DUSTM over Ricean channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1146–1151, Apr. 2008.

[280] R. Schober and L. H. J. Lampe, "Differential modulation diversity," *IEEE Trans. Veh. Technol.*, vol. 51, no. 6, pp. 1431–1444, Nov. 2002.

[281] L. H. J. Lampe and R. Schober, "Bit-interleaved coded differential space-time modulation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1429–1439, Sep. 2002.

[282] S. X. Ng and L. Hanzo, "On the MIMO channel capacity of multidimensional signal sets," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 528–536, Mar. 2006.

[283] N. Sun and J. Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1462–1470, Mar. 2014.

[284] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks—With a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.

[285] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 870–913, 2nd Quart., 2018.

[286] *Attenuation by Atmospheric Gases*, document Rec. P.676-10, ITU-R, 2013.

[287] *Specific Attenuation Model for Rain for Use in Prediction Methods*, document Rec. P.838-3, ITU-R, 2005.

[288] *Millimeter Wave Propagation: Spectrum Management Implications*, Bullet 70, FCC, Washington, DC, USA, 2007.

[289] T. Wu, T. S. Rappaport, and C. M. Collins, "Safe for generations to come: Considerations of safety for millimeter waves in wireless communications," *IEEE Microw. Mag.*, vol. 16, no. 2, pp. 65–84, Mar. 2015.

[290] P. H. Siegel, "Terahertz technology," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 3, pp. 910–928, Mar. 2002.

[291] A. J. Seeds, H. Shams, M. J. Fice, and C. C. Renaud, "Terahertz photonics for wireless communications," *J. Lightw. Technol.*, vol. 33, no. 3, pp. 579–587, Feb. 1, 2015.

[292] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, Oct. 2011.

[293] H. Sun, Z. Zhang, R. Q. Hu, and Y. Qian, "Wearable communications in 5G: Challenges and enabling technologies," *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 100–109, Sep. 2018.

[294] R. Sun, D. W. Matolak, and W. Rayess, "Air-ground channel characterization for unmanned aircraft systems—Part IV: Airframe shadowing," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7643–7652, Sep. 2017.

[295] "Communications operating concept and requirements (COCR) for the future radio system," EUROCONTROL/FAA Future Commun. Study Oper. Concepts Requirements Team, Tech. Rep. 1.0.

[296] M. O. Sinnokrot and J. R. Barry, "Fast maximum-likelihood decoding of the Golden code," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 26–31, Jan. 2010.

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

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

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

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

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

[302] H. M. Tullberg and P. H. Siegel, "Serial concatenated TCM with an inner accumulate code—Part I: Maximum-likelihood analysis," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 64–73, Jan. 2005.

[303] T. O. Ayodele, *Introduction to Machine Learning*. Rijeka, Croatia: InTech Open, 2010.

[304] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidisciplinary Optim.*, vol. 26, no. 6, pp. 369–395, Apr. 2004.

[305] C. A. Coello Coello, "Evolutionary multi-objective optimization: A historical view of the field," *IEEE Comput. Intell. Mag.*, vol. 1, no. 1, pp. 28–36, Feb. 2006.

[306] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, and L. Hanzo, "A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 550–586, 1st Quart., 2017.

[307] N. Ishikawa, "IMToolkit: An open-source index modulation toolkit for reproducible research based on massively parallel algorithms," *IEEE Access*, vol. 7, pp. 93830–93846, 2019.

CHAO XU (S'09–M'14–SM'19) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, China, and the B.Sc. (Eng.) degree (Hons.) from the Queen Mary University of London, U.K., through a Sino-U.K. joint degree program, in 2008, both in telecommunications engineering with management, and the M.Sc. degree (Hons.) in radio frequency communication systems and the Ph.D. degree in wireless communications from the University of Southampton, U.K., in 2009 and 2015, respectively. He is currently a Research Fellow with the Next Generation Wireless Research Group, University of Southampton. His research interests include index modulation, full-duplex, reduced-complexity MIMO design, noncoherent detection, and turbo detection. He received 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 the 2012 Chinese Government Award for Outstanding Self-Financed Student Abroad and the 2017 Dean's Award from the Faculty of Physical Sciences and Engineering, University of Southampton.

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, U.K. From April 2016 to March 2017, he was a Research Fellow of the Japan Society for the Promotion of Science. Since April 2017, he has been an Assistant Professor with the Graduate School of Information Sciences, Hiroshima City University, Japan.

Dr. Ishikawa 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, and the Outstanding Paper Award for Young C&C Researchers from NEC C&C Foundation, in 2014. He was certified as an Exemplary Reviewer of the IEEE Transactions on Communications, in 2017.

RAKSHITH RAJASHEKAR (M'14–SM'17) received the B.E. degree in electrical communication engineering from Visvesvaraya Technological University, Belgaum, India, in 2007, and the Ph.D. degree from the Department of Electrical Communication Engineering, Indian Institute of Science (IISc), India, in 2014. He is currently a Research Fellow with the University of Southampton, U.K.. Before joining the University of Southampton, he worked at Accord Software and Systems, Bengaluru, India, as a Systems Engineer, from 2007 to 2009, and as a Senior Scientist with Broadcom Communications, Bengaluru, from 2014 to 2015. His research interests include antenna selection in MIMO systems, differential communication, millimeter wave communication, and communication between drones with a focus on space-time signal processing and coding. He was a recipient of the Special Recognition Award from Broadcom Communications and the Dean's Award from the University of Southampton, U.K., for excellence in research. He has received the Best Reviewer Award from the IEEE Transactions on Wireless Communications, the IEEE Transactions on Communications, and the IEEE Wireless Communications Letters.

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 the 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, The University of Tokyo, Tokyo, where he heads the Wireless Communications Research Group. He authored or coauthored over 60 IEEE journal articles. His research interests include wireless communications, networking, signal processing, and antenna technology.

Dr. Sugiura was a recipient of a number of awards, including the Fifth Yasuharu Suematsu Award, in 2019, the Sixth RIEC Award from the Foundation for the Promotion of Electrical Communication, in 2016, the Young Scientists' Prize by the Minister of Education, Culture, Sports, Science and Technology of Japan, in 2016, the 14th Funai Information Technology Award (First Prize) from the Funai Foundation, in 2015, the 28th Telecom System Technology Award from the Telecommunications Advancement Foundation, in 2013, the Sixth IEEE Communications Society Asia-Pacific Outstanding Young Researcher Award, in 2011, the 13th Ericsson Young Scientist Award, in 2011, and the 2008 IEEE Antennas and Propagation Society Japan Chapter Young Engineer Award. He was also certified as an Exemplary Reviewer of the IEEE Communications Letters, in 2013 and 2014, and the IEEE Transactions on Communications, in 2018.

ROBERT G. MAUNDER (S'12) received the B.Eng. degree (Hons.) in electronic engineering, in July 2003, and the Ph.D. degree in telecommunications, in December 2007. He has been studying at the School of Electronics and Computer Science, University of Southampton, U.K., since October 2000. He began a lectureship, in November 2007, and was promoted to Associate Professor, in March 2013, and to Professor, in August 2017. He was a Chartered Engineer of IET, in November 2013, and a Fellow of IET, in January 2017. His research interests include joint source/channel coding and the holistic design of algorithms and hardware implementations for wireless communications. He has published around 100 IEEE articles in these areas. He is the Founder and the CTO of AccelerComm Ltd, which is commercializing his research as soft-IP.

ZHAOCHENG WANG received the B.S., M.S., and Ph.D. degrees from Tsinghua University, in 1991, 1993, and 1996, respectively.

From 1996 to 1997, he was a Postdoctoral Fellow with Nanyang Technological University, Singapore. From 1997 to 1999, he was a Research Engineer/Senior Engineer with OKI Techno Centre Pte. Ltd., Singapore. From 1999 to 2009, he was a Senior Engineer/Principal Engineer with Sony Deutschland GmbH, Germany. Since 2009, he has been a Professor with the Department of Electronic Engineering, Tsinghua University, where he is currently the Director of the Broadband Communication Key Laboratory, Beijing National Research Center for Information Science and Technology (BNRist). He holds 34 U.S./E.U. granted patents (23 of them as the first inventor). Several granted U.S./E.U. patents have been accepted by a plurality of international standards, such as IEEE802.11n, DVB-T2, IEEE802.15.3c, and ECMA387. He authored or coauthored two books, which have been selected by the IEEE Series on Digital and Mobile Communication and published by Wiley-IEEE Press. He has published more than 150 peer-reviewed journal articles. His research interests include wireless communications, millimeter wave communications, and optical wireless communications.

Dr. Wang is a Fellow of the Institution of Engineering and Technology. He received the ICC2013 Best Paper Award, the OECC2015 Best Student Paper Award, the 2016 IEEE Scott Helt Memorial Award (Best Paper Award of IEEE Transactions on Broadcasting), the 2016 IET Premium Award (Best Paper Award of Electronics Letters), the 2016 National Award for Science and Technology Progress (First Prize), the ICC2017 Best Paper Award, and the 2018 IEEE ComSoc Asia-Pacific Outstanding Paper Award. He was the Signal Processing for Optical Wireless Communications Symposium Co-Chair of the IEEE GlobalSIP2015, the Optical Wireless Communications Symposium Co-Chair of OECC2015, and the Wireless Communications Symposium Co-Chair of IEEE ICC 2013. He was an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2011 to 2015, and an Associate Editor of the IEEE COMMUNICATIONS LETTERS, from 2013 to 2016.

LIE-LIANG YANG (M'98–SM'02–F'16) received the B.Eng. degree in communications engineering from Shanghai TieDao University, Shanghai, China, in 1988, and the M.Eng. and Ph.D. degrees in communications and electronics from Northern (Beijing) Jiaotong University, Beijing, China, in 1991 and 1997, respectively. From June 1997 to December 1997, he was a Visiting Scientist with the Institute of Radio Engineering and Electronics, Academy of Sciences of the Czech Republic. Since December 1997, he has been with the University of Southampton, U.K., where he is currently a Professor of wireless communications with the School of Electronics and Computer Science. He has published over 370 research articles in journals and conference proceedings, authored/coauthored three books, and published several book chapters. His research interests include wireless communications, wireless networks and signal processing for wireless communications, and molecular communications and nano-networks. He is a Fellow of IET. He served as an Associate Editor to the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and *Journal of Communications and Networks* (JCN). He is currently an Associate Editor to IEEE ACCESS and a Subject Editor to the IET *Electronics Letters*. He was a Distinguished Lecturer of the IEEE VTS.

LAJOS HANZO (M'91–SM'92–F'04) received the degree in electronics, in 1976, and the Ph.D. degree, in 1983. During his 40-year career in telecommunications, he has held various research and academic positions in Hungary, Germany. and U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, U.K., where he holds the Chair in telecommunications. He has successfully supervised about 100 Ph.D. students, coauthored 20 John Wiley/IEEE Press books on mobile radio communications, totaling in excess of 10 000 pages, published more than 1899 research entries at IEEE Xplore, acted as TPC and the General Chair of IEEE conferences, presented keynote lectures, and has been awarded a number of distinctions. He is currently 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), U.K., 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 a FREng, FIET, Fellow of EURASIP, and DSc. In 2009, he received an honorary doctorate by the Technical University of Budapest, while in 2015 by the University of Edinburgh. From 2008 to 2012, he was the Editor-in-Chief of the IEEE Press and a Chaired Professor at Tsinghua University, Beijing. His research is funded by the European Research Council's Senior Research Fellow Grant. He is a Governor of the IEEE VTS.

• • •