NONCOHERENT MIMO SIGNALING FOR BLOCK-FADING CHANNELS

Approaches and Challenges

Ramy H. Gohary and Halim Yanikomeroglu

The majority of existing wireless communication systems relies on training-based signaling, i.e., schemes that rely on accurate channel estimates for detecting the transmitted symbols. Training requires its own hardware and algorithms and consumes a significant portion of the channel coherence interval, which is typically short in vehicular communications. The drawbacks of training can be alleviated by using noncoherent signaling schemes, which dispense with the training phase altogether. We will focus on a particular class of such schemes, i.e., Grassmannian signaling. This scheme is optimal for high signal-to-noise ratio (SNR) communication over noncoherent block-fading channels, which are likely to arise in many future networks.

Training-based signaling requires the receiver to obtain reliable channel estimates prior to detecting the transmitted symbols [1]. However, when the number of wireless nodes becomes large, as in the prospective paradigms of machine-type communications (MTCs) and the Internet of Things (IoT), acquiring channel estimates constitutes a formidable task, especially for multiple-input, multiple-output (MIMO) systems. For these paradigms, a viable alternative is offered by noncoherent signaling, which does not require the receiver to have access to channel estimates.

In comparison with their training-based counterparts, noncoherent schemes feature several advantages. First, they dispense with the hardware and the algorithms required for learning the channel. To appreciate this, we note that, for the urban microcell (UMi) long-term evolution (LTE) scenario of a user terminal operating at 1.8 GHz and traveling at 25 km/h, the channel coherence time is about eight symbol durations of 71.4 μs each. Using a four-antenna 3rd Generation Partnership Project training-based system to communicate in this scenario consumes 14.3% of the resources for training [2]; a cost that can be avoided by noncoherent signaling. Second, noncoherent schemes offer the potential of achieving higher spectral efficiencies. We will elaborate on this aspect later in this article. Third, noncoherent signaling alleviates one of the key problems that arise in massive MIMO cellular systems. In these systems, each base station (BS) is equipped with a large number of antennas, whereas the users are each equipped with a small number of antennas. When training-based schemes are used in the uplink of these systems, multiple neighboring cells will reuse the same training pilots. The received signals corresponding to these pilots interfere, resulting in pilot contamination. Hence, using noncoherent signaling in

Digital Object Identifier 10.1109/MVT.2018.2866306
Date of publication: 4 January 2019

such systems can provide superior performance to training-based ones. Indeed, applications of noncoherent signaling in massive MIMO has been shown to provide promising results [3]. Noncoherent signaling schemes can be classified into differential and block schemes.

Similar to differential phase shift-keying (PSK) [4], [5], differential MIMO signaling is desirable when the fading process of the channel is continuous with no abrupt temporal variations and when channel estimation and tracking are considered cumbersome. The premise of differential schemes is that the transmitted messages are encoded in the consecutive transitions between signaling blocks. Such transitions can be detected reliably at the receiver provided that the channel between blocks remains essentially constant [4]. Similar to their coherent counterparts, differential MIMO signaling schemes can be deployed in multiuser communication scenarios and can be used in conjunction with effective encoding techniques, e.g., bit-interleaved coded modulation [4], [6]. To improve performance, differential MIMO can be used with efficient multiple-symbol detection schemes that use trellis, sphere, and decision feedback decoding.

In contrast with their differential counterparts, block-signaling schemes are suitable for signaling over block-fading channels, i.e., channels that undergo abrupt variations between blocks, like those in which consecutive packets take different paths to arrive at the receiver. In such cases, the channel remains constant for a coherence time of T symbol durations. Then it takes on an independent realization. This model does not accurately represent slowly varying mobile communication scenarios, such as those arising in communication systems mounted on vehicles traveling at relatively low speeds. However, such a model is likely to arise in future vehicle-to-vehicle and airborne communication networks, which are expected to be exceedingly agile (to deal with high-mobility and rapid channel and demand variations) and heterogeneous (featuring ultradense small cells and assisted by multiple antenna relays). The prospective heterogeneity and density of deployment imply that, in many cases, there will be no predetermined path between a source and the network. Rather, each packet will find its way to a network node, possibly through multiple hops, see, e.g., Figure 1, wherein BSs communicate with a bullet train through mobile airborne and ground relays mounted on slower vehicles. In this case, the channels observed by individual packets are statistically independent, especially when beam-steering techniques are used by the relaying nodes. The coherence time in this case may not be known a priori but can be estimated using the statistical model of the channel, which depends on the agility of the beam-steering mechanism and the speed of the relaying nodes. Block-fading channels may arise in air-to-air and air-to-ground airborne communications, for instance when BSs aboard drones

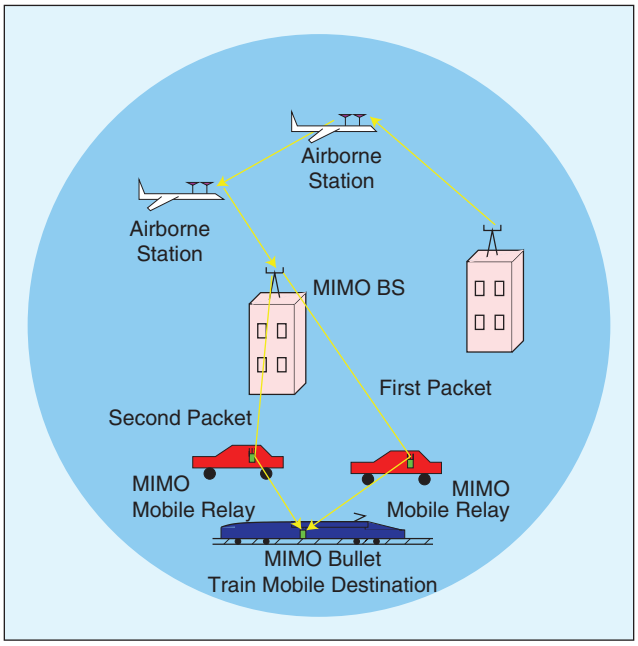

FIGURE 1 In future MIMO vehicular communication networks, each packet may see a different channel. The channel between airborne and ground stations may exhibit high Doppler frequencies and short coherence times.

or commercial or military aircrafts travel at high speeds yielding high Doppler frequencies.

Block-fading channels also arise in IoT systems, wherein communications are likely to be bursty, implying independent channels being observed over consecutive bursts and the coherence time is a deterministic value that corresponds to the burst duration. Other instances of block-fading channels arise in spatial modulation systems wherein a user receives packets from distinct subsets of a large antenna array, resulting in a deterministic coherence time equal to the symbol duration [7].

To achieve spectrally efficient communication over noncoherent block-fading channels, the transmitted signals must be adapted to the operating SNR. At low SNRs, the optimal signals assume an impulse-like structure, whereby the transmitter emits a short-duration pulse from one of its antennas and becomes dormant until it collects sufficient energy for transmitting the next pulse [8]. At high SNRs, the Shannon-award-winning paper [9] showed that the optimal signaling technique possesses a so-called Grassmannian structure, which is a special case of the generic noncoherent block-signaling structure discovered in [10]. At moderate-to-high SNRs, the optimal signaling structure is not known. However, it was shown in [11], that in this region Grassmannian signaling is close to optimal.

The benefits of block noncoherent signaling comes at the expense of several challenges that must be overcome to become an attractive alternative to its training-based counterpart. We will later allude to these challenges and

to approaches for their mitigation. We believe that the future of noncoherent signaling is not to replace its training-based counterpart but, rather, to coexist with it to serve in instances where training is deemed too cumbersome and wasteful of resources.

We now elaborate on the spectral advantage of block noncoherent signaling over its training-based counterpart for communicating over block-fading channels. Let us consider a channel that is static for a finite coherence interval. In this case, using training to acquire channel information prior to detection is one of potentially many signaling strategies. However, the optimal strategy remains to be one that achieves the capacity of the original noncoherent channel, i.e., the channel whose information is not revealed to the receiver. Any signaling strategy other than the optimal, including the training-based one, is suboptimal and only achieves a lower bound on capacity [1], [9].

Let us now consider the case when the coherence interval becomes infinite. The noncoherent channel capacity approaches the ideal coherent one in which the coefficients are perfectly known a priori with the cost of training ignored, i.e., the two capacities become identical. This is because the portion of time required for training becomes negligible as the coherence intervals increases. Hence, it can be seen that the advantage of noncoherent signaling is more pronounced when the channel coefficients vary rapidly and abruptly with time [1], [9], [12]. In conclusion, for channels with short coherence intervals like the ones that arise in vehicular communications, Grassmannian signaling offers valuable performance advantages over training-based schemes [13], and these advantages diminish as the coherence time increases. In addition to its higher spectral efficiency, Grassmannian signaling was shown in [14] to yield significant performance advantages in high-mobility communication systems with large numbers of transmit and receive antennas.

Noncoherent Grassmannian MIMO Signaling for Block-Fading Channels

Block-fading channels arise in many communication scenarios. However, finding signaling schemes that enable the highest rate to be noncoherently communicated over these channels remains an open problem. For simplicity of exposition, we focus herein on spatially white channels, i.e., channels with no spatial correlation, e.g., see [15] for results on such channels. In those scenarios, the noncoherent capacity can be achieved by input signals that have the form of the product of an isotropically distributed (i.d.) $T\times M$ unitary matrix and an $M\times M$ diagonal matrix with nonnegative entries, where T is the block length and M is the number of transmit antennas [10]. [Note that isotropicality signifies uniform statistical distribution on a (hyper) sphere, e.g., PSK symbols on a circle.] Although the optimal distribution of the unitary component of the input signal is known, the corresponding distribution of the diagonal component is not known except for the asymptotic cases of low and high SNRs, as detailed below.

We now consider the case of high SNRs when the block size, T, exceeds the total number of transmit and receive antennas. The case when T is less than this value is considered in [16]. At high SNRs, the channel capacity is achieved when the diagonal component is fixed, implying that all information is encoded in the tall unitary matrix and none in the diagonal one [9]. For the unitary matrix, we note that its columns are orthogonal unit vectors that represent the subspace in which these vectors lie. Since each subspace is spanned by infinitely many bases, it can be seen that two unitary matrices either span the same subspace or two distinct subspaces. The set of distinct subspaces form a geometric object known as the *Grassmann manifold*, where a manifold is a smooth (differentiable) surface that resembles a Euclidean space in the neighborhood of each point. (Examples of common manifolds include the sphere and the torus; a cube is not a smooth manifold.) The Grassmann manifold can therefore be represented by the set of tall unitary matrices that span distinct subspaces. The real Grassmann manifold with $T = 3$ and $M = 1$ can be visualized by Boy's surface depicted in Figure 2.

Historical Note

The Grassmann manifold is named after Hermann Günther Grassmann, born in Stettin, Prussia (now Poland), in 1809. Although Grassmann did not receive formal university training in mathematics, he made revolutionary discoveries that gave birth to the, then new, notions of exterior algebra and linear subspaces. Most of Grassmann's discoveries were too advanced for his time and their depth and implications were only appreciated after his death in 1887.

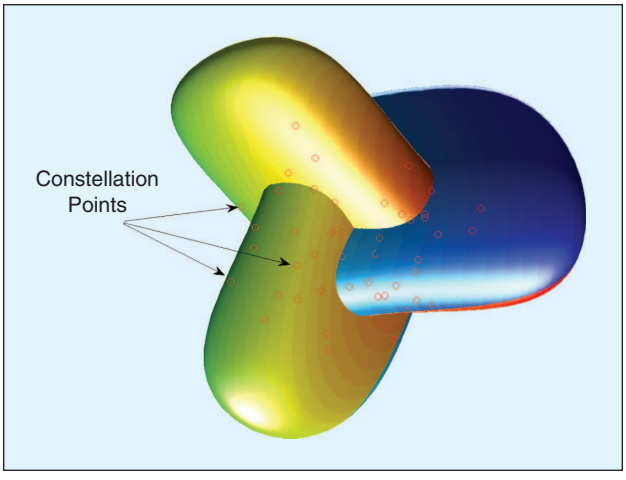

FIGURE 2 Boy's surface represents the real Grassmann manifold with $T = 3$ and $M = 1$ [17].

When the block size T exceeds the total number of transmit and receive antennas, the high-SNR capacity-optimal distribution of the unitary component of the input signal is i.d. on the Grassmann manifold, i.e., the set of unitary matrices spanning distinct subspaces [9]. For instance, an i.d. constellation on the Grassmann manifold with $T = 3$ and $M = 1$ corresponds to a set of uniformly distributed points on Boy's surface; see the marked dots in Figure 2, where one can appreciate that designing i.d. Grassmannian constellations can be more involved than designing constellations on more conventional surfaces, e.g., spheres.

The key feature that enables Grassmannian signals to achieve the high-SNR noncoherent capacity follows from the construction of these signals. In particular, each of these signals represents a basis that spans a particular subspace. When transmitted over the channel, the basis is scaled and rotated, but the subspace spanned by this basis is preserved. For instance, consider the case with block-size $T = 3$ and $M = 2$ transmit and receive antennas, with the transmitted signal

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

and the channel matrix being

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}.$$

We note that $\mathbf{X}$ spans the xy plane but has a zero component in the z dimension. From a Grassmannian perspective, this $\mathbf{X}$ is equivalent to any unitary matrix spanning the same subspace, e.g.,

$$\tilde{\mathbf{X}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

In contrast, $\mathbf{X}$ is not equivalent to

$$\hat{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

because the latter spans the xy plane. In the absence of noise, i.e., at asymptotically high SNRs, the received signal corresponding to $\mathbf{X}$ is

$$\mathbf{Y} = \mathbf{X}\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ 0 & 0 \end{bmatrix}.$$

We note that $\mathbf{Y}$ spans the same xy plane as $\mathbf{X}$ and has a zero component in the z dimension, irrespective of the channel matrix. An illustration of this fact is provided in Figure 3.

Several approaches for designing Grassmannian constellations are available. For instance, one approach uses an exponential transformation to map coherent space–time block codes (STBCs) to tall unitary matrices spanning distinct subspaces [18]. Other approaches that use direct optimization techniques on the Grassmann manifold are provided in [19]. When used for noncoherent signaling, the constellations generated by those techniques offer significant performance gains over their training-based counterparts. For instance, the performance of an uncoded 4,096-point 4×2 Grassmannian constellation designed using the technique in [19] is compared in Figure 4 with the performance of a training-based scheme that uses the Golden code with an eight–quadrature amplitude modulation (QAM) constellation when the block size is $T = 4$. The training-based scheme proceeds in two phases: a training phase followed by a communication phase. In the training phase, pilot symbols are transmitted with one from each antenna at consecutive time slots. Using its knowledge of the pilots, the receiver generates an estimate of the channel matrix. In the communication phase, an STBC, e.g., Alamouti or Golden, is used to transmit the desired message. The receiver then

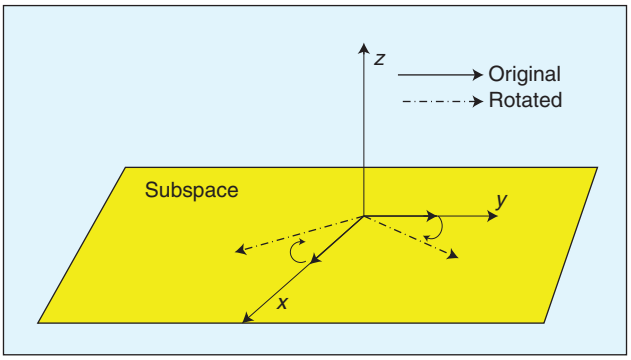

FIGURE 3 The channel rotates and scales the bases but preserves the subspace spanned by the signal matrix [9].

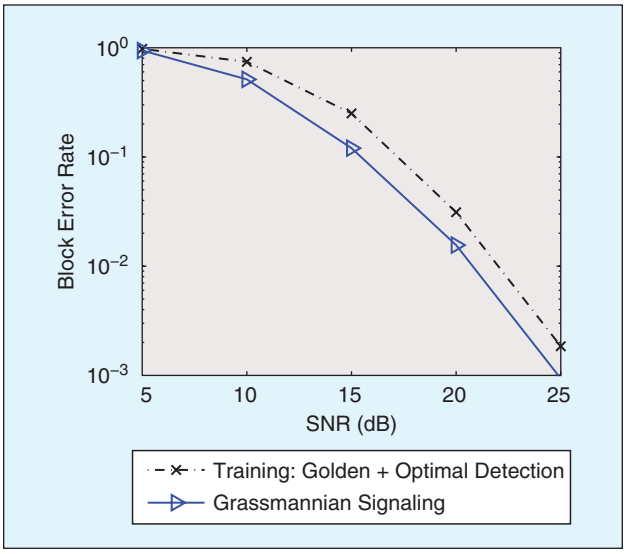

FIGURE 4 The performance of noncoherent Grassmannian and training-based signaling with the Golden code [19].

uses the potentially inaccurate channel estimate generated in the training phase to coherently detect the STBC transmitted in the communication phase.

The training-based scheme under consideration uses the Golden code, which, to date, is the best available 2×2 STBC. The coherence time is equally split between the training phase and the communication phase, which is optimal when the channel exhibits no temporal correlation across blocks [1]. When such a correlation exists, as in the aforementioned UMi LTE communication scenario, the cost required for tracking the channel can be reduced using channel interpolation techniques, and when the channel coefficients vary within each block, the analysis in [14] indicates an increased advantage of Grassmannian over training-based signaling.

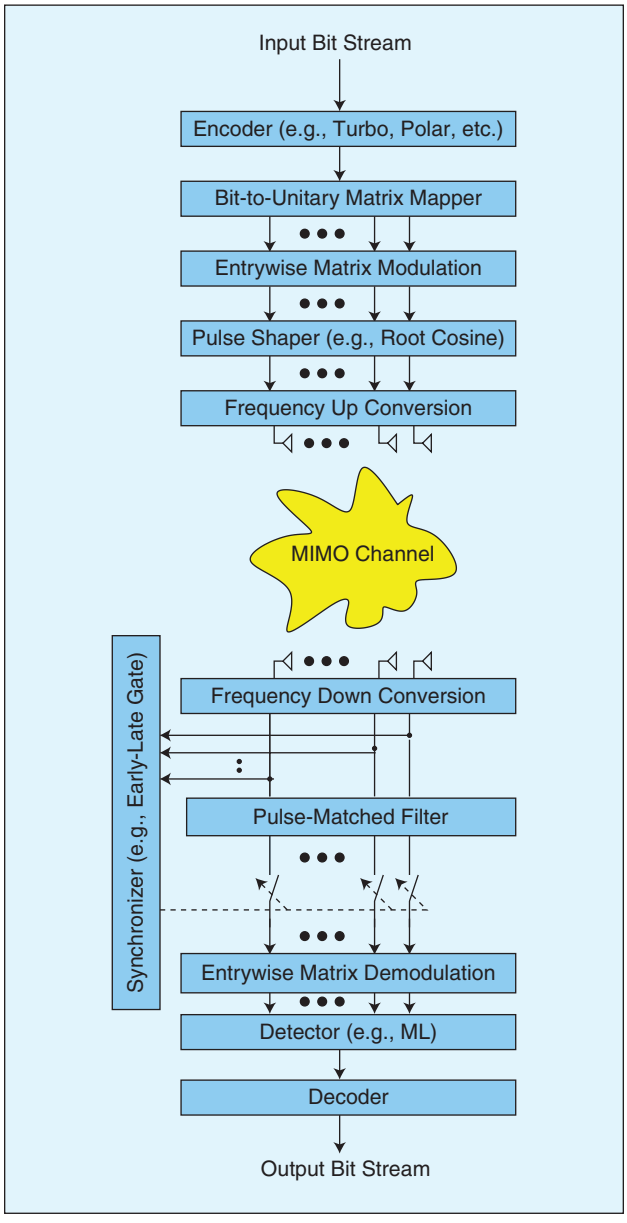

FIGURE 5 The potential embodiment of a noncoherent system. ML: maximum likelihood.

With both schemes operating at an overall spectral efficiency of 3 b/s/Hz, Figure 4 shows that the Grassmannian scheme offers an SNR advantage of up to 2 dB. This advantage is greater for systems operating at higher rates and with more receive antennas. A few comments on this figure are in order, the ratio $T/M = 2$. If the number of transmit antennas increases with T so that T/M remains fixed, then theoretical results suggest that a performance analogous to the one depicted in this figure is expected. In a complementary fashion, when M is fixed, the performance of training-based signaling approaches that of noncoherent signaling as T increases. However, for large T, the design of practical Grassmannian constellations is generally difficult, rendering training-based signaling more desirable. However, it is worth noting that training-based schemes are available only for specific values of M and T; for large values of T, new coherent schemes need to be developed. For instance, the Golden code is, to date, the best available scheme for coherent MIMO systems with two transmit and two receive antennas with a normalized coherence interval of $T = 2$. No similar code is available for the case with $T = 3$, let alone arbitrary values of T. Similar difficulties arise in systems with arbitrary numbers of transmit and receive antennas.

Transition from Theory to Practice

In conventional communications, complex scalars are used to carry information over the channel. In Grassmannian signaling, this task is performed by tall unitary matrices, which are loaded with binary information through a mapper that associates tall unitary matrices with distinct binary strings of a particular length. The length of such a string and the corresponding number of matrices determine the rate at which data are transmitted (Figure 5). Despite the lack of visual symmetries, this process resembles the way in which QAM and PSK are labeled in standard coherent systems, wherein each QAM or PSK symbol is assigned a binary label, Gray or otherwise, that represents information to be transmitted. In contrast, in Grassmannian signaling, the matrix symbol is treated as a single entity to which a binary label is assigned. To transmit a unitary matrix, during each time slot, each antenna element emits a sinusoidal waveform with the amplitude and phase of the corresponding complex scalar entry in the matrix. Each waveform is then shaped and received by root-raised-cosine filters. For maximal received power, sampling at the output of each matched filter must be synchronized with the symbol duration. This can be accomplished by using standard synchronizers, including pilot-free early-late gate ones. In other words, similar to standard coherent communication systems, synchronizing samples with the symbol durations in noncoherent systems is necessary and can

be accomplished without invoking pilot signaling. For detection, the ML technique, or a simplified version thereof [19], is performed by searching for the unitary matrix in the noncoherent constellation that maximizes the likelihood function.

From this discussion, it can be seen that, although noncoherent signaling dispenses with pilot symbols for learning the channel coefficients, it does not dispense with synchronized sampling at the matched filter output. A potential embodiment of a noncoherent communication system is depicted in Figure 5.

Practical Implementation Challenges

In coherent communications, constellations are structured as linear combinations of scalar QAM and PSK symbols and their complex conjugates. In contrast, in noncoherent communications, constellations are unitarily structured, which implies that these constellations cannot be readily expressed as linear combinations of scalar constellations. This lack of decomposability results in implementation challenges including systematic design, labeling, and detection. We elaborate on these challenges in the following sections.

Systematic Design and Parameterization

Designing multidimensional constellations whether for training-based or noncoherent communications usually involves numerical optimization. The performance of such constellations is usually superior to their scalar counterparts but at the expense of higher dimensionalities and cardinalities. Grassmannian constellations are no exception. To gain insight, we note that, in conventional communications, the signals are designed to be at a maximum distance under various constraints. For instance, in PSK, the signals are constrained to lie on the unit circle, leading the points of the optimal constellation to be uniformly distributed on that circle. This is, in a sense, analogous to the case of designing Grassmannian symbols, but in a matrix form. The unit circle corresponds to the set of tall unitary matrices and the uniformly distributed phases in the PSK constellation correspond to uniformly distributed points on that set, which is a curved non-Euclidean space (Figure 2). Differential MIMO signaling offers another example in which the signaling matrices are uniformly distributed on the set of unitary matrices. However, in differential MIMO, the matrices are square, and their size is determined solely by the number of transmit antennas. This feature renders the task of designing constellations for differential MIMO easier than their Grassmannian counterparts. However, such matrices are not suitable for communicating over the block-fading channels considered herein.

The design of unitary constellations directly is generally difficult not only because of the structure of curved spaces but also because of the number of matrices that is usually required in these designs, especially for large values of T. To elaborate, we note that the goal of Grassmannian signaling is to achieve the high-SNR capacity of the noncoherent channel. For instance, in the UMi LTE communication scenario, the coherence time is about eight symbol durations, which implies that the channel capacity is maximized when the number of transmit and receive antennas is $M = 4$ [9]. At an SNR of 30 dB, this capacity is about 20 b/s/Hz. Hence, operating at half the capacity, requires $2^{80} \approx 1.21 \times 10^{24}$ unitary matrices!

For large values of T, training-based schemes perform well and for small values of T, the difficulty of designing Grassmannian constellations can be effectively mitigated by using parameterization techniques which enable families of constellations to be generated with a small number of scalar variables. See, e.g., [18] for a methodology that uses the exponential transformation to map standard STBCs to unitary matrices. Another approach for designing structured Grassmannian constellations has been proposed in [13]. This approach uses a technique in which a Grassmannian constellation is obtained by projecting a Euclidean constellation on the Grassmann manifold. Such techniques will facilitate designing Grassmannian constellations of cardinalities and dimensions that suit practical implementation in various propagation environments. Fortunately, designing Grassmannian constellations is only performed offline, and hence, the computational cost incurred in this process is unrecurrent and transparent to the end user.

Labeling

In conventional communications, bits are mapped to PSK and QAM channel symbols, a process that is usually referred to as labeling. Conventional communication systems use variations of standard Gray labeling, which is preferable in turbo and low-density parity-check coding, and set-partitioning labeling, which is preferable in Trellis-coded modulation and polar coding. In PSK and QAM, labeling can be performed using algebraic methods and is facilitated by visual symmetries but in multidimensional constellations, including Grassmannian ones, visual symmetries are not available and labeling is performed using algebraic techniques.

Finding optimal labelings for matrix symbols is computationally prohibitive for constellations with large cardinalities. To address this problem, effective techniques for providing quasi Gray labels and quasi set-partitioning have been developed in [20] and [19], respectively. And while current labeling techniques rely on the distances between points, they do not make explicit use of the geometric properties of the Grassmannian surface on which the unitary matrices lie. Invoking such properties can reveal symmetries that facilitate efficient labeling. For instance, on the Grassmann manifold, the maximum distance occurs between matrices that span orthogonal

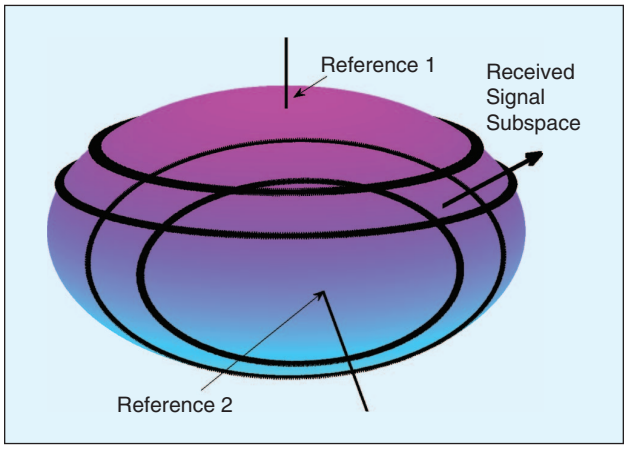

FIGURE 6 The RS detector evaluates the likelihood function for the constellation points that lie within a strap from the subspace spanned by the received signal. The detection region in this figure is the intersection of the two straps.

subspaces. This property can be used to halve the size of the labeling problems. It is expected that invoking other symmetries can further simplify the labeling task.

Detection

Similar to training-based signaling, the optimal ML decoding rule for Grassmannian signaling is mathematically well defined, see, e.g., [10]. In particular, as in training-based communications, the full ML detector in Grassmannian noncoherent MIMO signaling evaluates the likelihood function for every symbol in the constellation and decides on the one that maximizes this function. Hence, the two communication schemes have comparable complexity for optimal detection. In particular, the number of likelihood evaluations required for optimal detection of Grassmannian and training-based communications is 2^R, where R is the transmission rate. However, the main difference between training-based and Grassmannian communications is that, in the former, the symbols are constructed from scalar constellations, whereas, in the latter, the symbols represent linear subspaces.

Full ML detection is generally computationally expensive in coherent and noncoherent scenarios when the constellations have large cardinalities and do not possess apparent symmetries, e.g., constellations generated through numerical optimization. In coherent scenarios the difficulty is alleviated by an efficient sphere detection technique that exploits the linear structure of the transmitted signals to search for the point that maximizes the likelihood function within a prescribed sphere around the received signal vector. The radius of this sphere provides a tradeoff between detection performance and complexity; a larger sphere implies more candidate constellation points and better average performance.

In noncoherent communications, detection is slightly more involved because therein, the signaling matrices are designed directly, with no known structure, apart from unitarity. To alleviate this difficulty, an effective reduced-search (RS) noncoherent detector for Grassmannian constellations was developed in [19]. The philosophy of this detector mimics that of the sphere detector in that it limits the search to the neighborhood of the subspace spanned by the received signal. A pictorial illustration of the idea of this detector is provided in Figure 6; the Grassmann manifold is represented by a sphere. Two arbitrary points are chosen to be references. When a signal is received, its subspace and the reference points are used to construct two straps that contain this subspace. Using a lookup table, only those constellation points that lie at the intersection of the two straps are examined against the likelihood metric. Analogous to sphere detectors in coherent communications, in the RS detector, the number of reference points and the width of each strap offer a tradeoff between performance and computational complexity. Proper adjustment of these parameters enables the RS detector to perform close to ML detection but with a significantly less computational cost. To appreciate this, in Figure 7(a) and (b), we compare the performance and the number of likelihood evaluations invoked by the ML and the RS detectors when one, ten, and 15 reference points are used to detect a 1,024-point Grassmannian constellation with $T = 4$ and $M = 2$. From Figure 7(a), it can be seen that, with 15 reference points, the RS detector yields a performance comparable to that of ML but with a computational saving of about 97% at an SNR of 30 dB.

In contrast with uncoded systems, in coded ones, the task of detecting Grassmannian constellations can be coupled with that of decoding. Since joint detection and decoding is often overly complicated, the complexity of this task can be mitigated by using iterative detection and decoding. Such a technique was used in [20] with a 4,096-point 4×2 Grassmannian constellation with an outer turbo code. A comparison between Grassmannian and training-based signaling with Golden code and optimal detection in a turbo-coded system is shown in Figure 8.

Figure 8 shows that Grassmannian signaling provides an SNR gain of about 0.7 dB at a bit error rate of 10^{-3}, which confirms the performance advantage of Grassmannian signaling over its training-based counterpart in coded communication systems.

Finally, we note that detection complexity depends on the structure of the constellations points: stringent structures are easier to detect but might compromise performance. Hence, mitigating detection complexity can be accomplished either directly at the receiver using the RS philosophy, or indirectly if the constellations were designed in such a way that facilitates detection, e.g., [18].

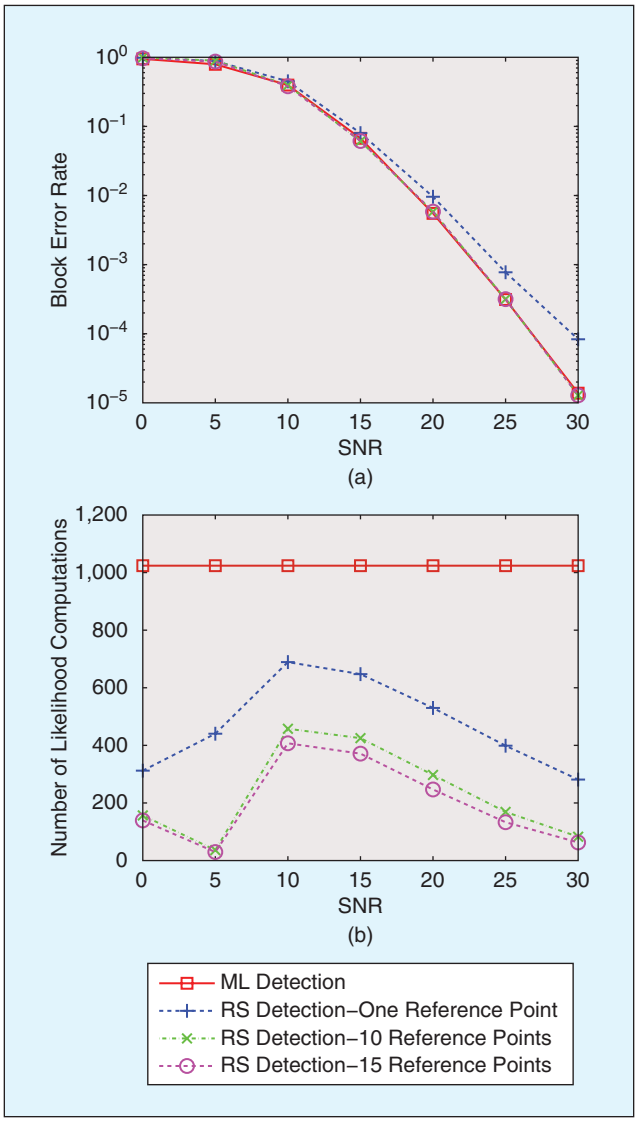

FIGURE 7 The performance and computational advantages of the RS detector. (a) The performance comparison and (b) the complexity comparison.

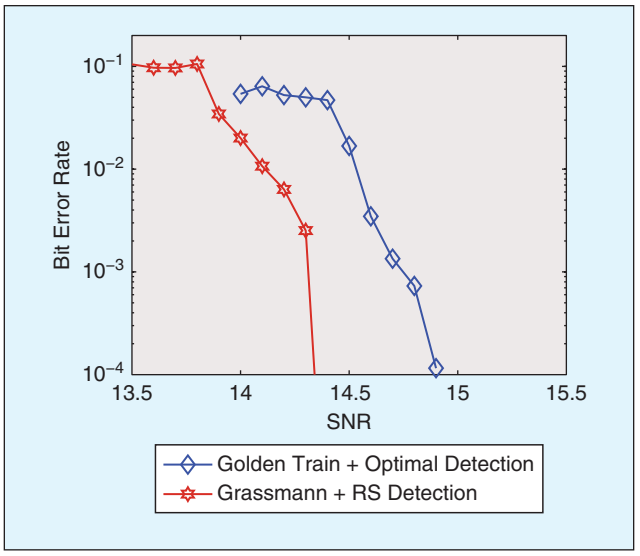

FIGURE 8 The performance of Grassmannian and training-based signaling in a turbo-coded system with iterative detection and decoding.

Peak-to-Average Power Ratio

One of the challenges that must be overcome for Grassmannian signaling to be more conducive to practical implementation is the peak-to-average power ratio (PAPR), which directly affects the dynamic range and the cost of the power amplifier used at the transmitter. The PAPR depends on the uniformity of the signal amplitudes transmitted from each antenna.

Being elementwise nonuniform, Grassmannian constellations are likely to exhibit high PAPRs. Fortunately, such constellations offer an opportunity to minimize the PAPR. To see that, we note that Grassmannian constellation points are invariant under right multiplication by square unitary matrices. Hence, each constellation point has an equivalent representation wherein the original $T \times M$ constellation point is right multiplied by a square $M \times M$ unitary matrix. Hence, optimizing these unitary matrices will enable the PAPR to be minimized. The impact of this property on minimizing PAPR has yet to be explored.

Conclusions

This article provided a descriptive characterization of the mathematical concepts that underlie Grassmannian noncoherent signaling, which suits communication over MIMO block-fading channels. We highlighted key challenges that must be overcome for it to become an attractive alternative for high-SNR communication in emerging complexity-constrained paradigms, e.g., MTCs and the IoT. The challenges that face noncoherent communications are not insignificant, but their future is promising.

Acknowledgments

We would like to acknowledge the valuable feedback provided by the anonymous reviewers.

Author Information

Ramy H. Gohary (Ramy.Gohary@carleton.ca) is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada. His research interests include the design of embedded systems; applications of embedded systems in machine-to-machine communications; the Internet of Things; cross-layer design of wireless networks; analysis and design of multiple-input, multiple-output wireless communication systems; applications of optimization and geometry in signal processing and communications; and information theoretic aspects of multiuser communication systems. He is a Senior Member of the IEEE.

Halim Yanikomeroglu (halim@sce.carleton.ca) is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada. His research interests include wireless technologies with a special emphasis on wireless networks including nonterrestrial networks. He is a Fellow of the IEEE.

References

[1] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[2] T. Abe. (2009, Dec. 17–18). 3GPP self-evaluation methodology and results. 3GPP LTE-Advanced Evaluation Workshop. [Online]. Available: http://www.3gpp.org/ftp/workshop/2009-12-17_ITU-R_IMT-Adv_eval/docs/pdf/REV-090007%20SelfEvalulation%20assumption.pdf

[3] Y. Han and J. Lee, "Grassmannian training for massive MIMO cellular networks," in *Proc. IEEE 2016 50th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2016, pp. 193–197.

[4] V. M. Baeza, A. G. Armada, W. Zhang, M. El-Hajjar, and L. Hanzo, "A noncoherent multiuser large-scale SIMO system relying on M-ary DPSK and BICM-ID," *IEEE Trans. Veh. Technol.*, vol. 67, pp. 1809–1814, Feb. 2018.

[5] Y. Jing and B. Hassibi, "Unitary space-time modulation via Cayley transform," *IEEE Trans. Signal Process.*, vol. 51, pp. 2891–2904, Nov. 2003.

[6] V. M. Baeza, A. G. Armada, M. El-Hajjar, and L. Hanzo, "Performance of a non-coherent massive SIMO M-DPSK system," in *Proc. IEEE Vehicular Technology Conf.*, Toronto, Canada, Sept. 2017, pp. 1–5.

[7] R. Y. Mesleh, H. Haas, S. Sinanović, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol*, vol. 57, pp. 2228–2241, July 2008.

[8] S. G. Srinivasan and M. K. Varanasi, "Optimal constellations for the low-SNR noncoherent MIMO block Rayleigh-fading channel," *IEEE Trans. Inf. Theory*, vol. 55, pp. 776–796, Feb. 2009.

[9] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, pp. 359–383. Feb. 2002.

[10] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[11] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1473–1484, June 2002.

[12] Y. M. M. Fouad, R. H. Gohary, H. Yanikomeroglu, J. Cabrejas, D. Calabuig, S. Roger, and J. F. Monserrat, "Time-frequency Grassmannian signalling for MIMO frequency-flat broadband systems," *IEEE Commun. Lett.*, vol. 19, pp. 475–478, Mar. 2015.

[13] N. Khac-Hoang, A. Decurninge, M. Guillaud, and S. Yang, "Design and analysis of a practical codebook for non-coherent communications," in *Proc. IEEE 2016 51st Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 2017, pp. 1237–1241.

[14] J. Cabrejas, S. Roger, D. Calabuig, Y. M. M. Fouad, R. H. Gohary, J. F. Monserrat, and H. Yanikomeroglu, "Non-coherent open-loop MIMO communications over temporally-correlated channels," *IEEE Access*, vol. 4, pp. 6161–6170, Jan. 2016.

[15] R. H. Gohary and H. Yanikomeroglu, "The ergodic high SNR capacity of the spatially-correlated non-coherent MIMO channel within an SNR-independent gap," in *Proc. IEEE Information Theory Workshop*, Jeju, South Korea, Oct. 2015, pp. 234–238.

[16] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 117–132, Feb. 2013.

[17] F. Apéry, "La surface de Boy," *Adv. Math.*, vol. 61, no. 3, pp. 185–266, Sept. 1986.

[18] I. Kammoun, A. M. Cipriano, and J.-C. Belfiore, "Non-coherent codes over the Grassmannian," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3657–3667, Oct. 2007.

[19] R. H. Gohary and T. N. Davidson, "Non-coherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, pp. 1176–1205, Mar. 2009.

[20] G. W. Colman, R. H. Gohary, M. A. El-Azizy, T. N. Davidson, and T. J. Willink, "Quasi-Gray labelling for Grassmannian constellations," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 626–636, Feb. 2011.

VT