When Probabilistic Shaping Realizes Improper Signaling for Hardware Distortion Mitigation

When Probabilistic Shaping Realizes Improper Signaling for Hardware Distortion Mitigation

Sidrah Javed, *Student Member, IEEE,* Ahmed Elzanaty, *Member, IEEE,* Osama Amin, *Senior Member, IEEE,* Basem Shihada, *Senior Member, IEEE,* and Mohamed-Slim Alouini, *Fellow, IEEE*

Abstract

Hardware distortions (HWDs) render drastic effects on the performance of communication systems. They are recently proven to bear asymmetric signatures; and hence can be efficiently mitigated using improper Gaussian signaling (IGS), thanks to its additional design degrees of freedom. Discrete asymmetric signaling (AS) can practically realize the IGS by shaping the signals' geometry or probability. In this paper, we adopt the probabilistic shaping (PS) instead of uniform symbols to mitigate the impact of HWDs and derive the optimal maximum a posterior detector. Then, we design the symbols' probabilities to minimize the error rate performance while accommodating the improper nature of HWD. Although the design problem is a non-convex optimization problem, we simplified it using successive convex programming and propose an iterative algorithm. We further present a hybrid shaping (HS) design to gain the combined benefits of both PS and geometric shaping (GS). Finally, extensive numerical results and Monte Carlo (MC) simulations highlight the superiority of the proposed PS over conventional uniform constellation and GS. Both PS and HS achieve substantial improvements over the traditional uniform constellation and GS with up to one order magnitude in error probability and throughput.

Index Terms

Hardware distortion, asymmetric signaling, error probability analysis, improper discrete constellations, improper Gaussian noise, non-uniform priors, optimal detector.

I. Introduction

Exponentially rising demands of high data rates and reliable communications given the limited power and bandwidth resources impose enormous challenges on the next generation of wireless communication systems [1], [2]. Various research contributions propose new configurations and novel techniques to address these challenges [3], [4]. Nonetheless, the performance of

S. Javed, A. Elzanaty, O. Amin, B.Shihada and M.-S. Alouini are with Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia. E-mail: {sidrah.javed, ahmed.elzanaty, osama.amin, basem.shihada, slim.alouini}@kaust.edu.sa

TABLE I: List of Abbreviations

OFC	Optical fiber communications	SNR	Signal-to-noise ratio	FSO	Free-space optics
IGS	Improper Gaussian signaling	PDF	Probability density function	r.v.	Random variable
PGS	Proper Gaussian signaling	EbNo	Energy per bit per noise ratio	MC	Monte Carlo
HWD	Hardware distortion	AWGN	Additive white Gaussian noise	TB	Theoretical bound
NS	No-shaping	QAM	Quadrature amplitude modulation	RF	Radio frequency
GS	Geometric shaping	PSK	Phase shift keying	PEP	Pairwise error probability
PS	Probabilistic shaping	DoF	Degrees of freedom	BER	Bit error rate
HS	Hybrid shaping	ML	Maximum likelihood	DM	Distribution matching
AS	Asymmetric signaling	MAP	Maximum a posterior	UB	Upper bound
KKT	Karush Kuhn Tucker	SCP	Successive convex programming	LB	Lower bound

such systems can be highly degraded by the hardware imperfections in radio frequency (RF) transceivers [5]. Such imperfections give rise to additive signal distortions emerging from the phase noise, mismatched local oscillator, imperfect high power amplifier/low noise amplifier, non-linear amplitude-to-amplitude and amplitude-to-phase transfer [6]–[8]. Various contributions emphasized the distinct improper behavior of these hardware distortions (HWDs) [9]–[12], which requires effective compensation techniques to meet the performance demands.

The improper Gaussian signaling (IGS) is proven as an effective scheme to mitigate the deteriorating effects due to the existence of improper noise or interference in wireless communication systems. More precisely, IGS is a generalized complex signaling that allows the signal components to be correlated and/or to have unequal power, as opposed to proper Gaussian signaling [13]. IGS offers an additional degrees of freedom (DoF) in signaling design characterized by the circularity coefficient [14]. Several studies highlight the significance of IGS to improve the system performance under improper interference [15]–[18]. Recent studies quantified the impact of IGS in improving the ergodic rate and outage performance by effectively dampening the improper noise effects in multi-antenna or multi-nodal system settings [19]–[21], IGS benefits can also be reaped in various full-duplex/half-duplex relay settings by effectively compensating the residual self-interference, inter-relay interference and/or HWD [11], [20], [22].

A. Background

Despite the overwhelming benefits of IGS, it is practically infeasible owing to the high detection complexity and unbounded peak-to-average power ratio [2], [23]. This motivated the researchers to design some equivalent finite and discrete asymmetric signaling (AS) schemes for practical implementation. Improper discrete constellation, or AS, entails redesigning the symmetric discrete signal constellation to convert it into an asymmetric signal [2]. Several studies focused on geometric shaping (GS) as a possible designing scheme to improve system performance. GS transforms equally spaced symbols to unequally spaced symbols (due to correlated and/or unequal power distribution between quadrature components of the symbols) in a distinct geometric envelop such as ellipse [24], parallelogram [23], [25] or some irregular envelop [26]. A family of improper discrete constellations generated by widely linear processing of a square M-ary quadrature amplitude modulation (QAM) depict parallelogram envelop [23]. Similarly, GS based on optimal translation and rotation also yields parallelogram envelop [25]. However, conditioned on high signal-to-noise ratio (SNR) and higher order QAM, the optimal constellation is the intersection of the hexagonal lattice/packing with an ellipse where the eccentricity determines the circularity coefficient [24]. GS has emerged as a competent player to reduce shaping loss and improve reception at lower signal-to-noise ratios in terrestrial broadcast systems [27], [28]. GS parameters can be designed for diverse objectives such as capacity maximization [23], bit error rate (BER) reduction [25], and symbol error probability minimization [24]. Although the asymmetric discrete family of constellations is practical, they exhibit two types of loss, i.e., shaping loss and packing loss in approaching IGS theoretical limits [23].

B. Motivation and Related Work

Most of the efforts to close the gap between practical AS and theoretical IGS are concentrated around GS with a limited focus on probabilistic shaping (PS), as another way to implement AS for HWD. Given a fixed number of symbols and the symbol locations, an asymmetric constellation can be obtained by adjusting the symbol probabilities [29]. We aim to find the optimal probabilities of the symbols that can improve the system performance under improper HWD. Once the optimal probabilities are computed, the symbols can be probabilistically shaped using a distribution matching (DM), which maps uniformly distributed input bit string into M-QAM/phase shift keying (PSK) symbols with the target distribution. Numerous DM techniques are proposed for rate adaptation such as constant composition DM [30], sphere shaping and shell mapping [31], [32], adaptive arithmetic DM [33], and Trellis-Based [34] and compressive DM [35] based on syndrome encoding [36] and compressive sensing [37], respectively.

PS-based schemes have been employed to enhance the system performance in optical fiber communications (OFC) and free-space optics (FSO). In OFC, multiple transformations are presented to approach Gaussian channel capacity using PS including prefix codes [38], [39], many-to-one mappings combined with a turbo code [40], distribution matching [41] and cut-and-paste method [42]. Furthermore, multidimensional coded modulation format with hybrid probabilistic and geometric constellation shaping can effectively compensate non-linearity and approach Shannon limits in OFC [43]. Coded modulation scheme with PS aims to solve the shaping gap and coarse mode granularity problems [44]. Interested reader can read the classic work [45] for the design guidelines of AS in the coherent Gaussian channel with equal signal energies and unequal a priori probabilities. Probabilistic amplitude shaping is another concept that can only be used for symmetric constellation with coherent modulation, which greatly limits its application [46]. For FSO, a practical and capacity achieving PS scheme with adaptive coding modulation is proposed with intensity modulation/direct detection [47].

The concept of PS is widely employed in the OFC and FSO systems. However, it is quite novel in wireless communication systems and few studies have contributed in this domain [48]–[51]. For example, enumerative amplitude shaping is proposed as a constellation shaping scheme for IEEE 802.11 which renders Gaussian distribution on the constituent constellation [48]. Moreover, a modification to 5G new radio polar code with higher order modulation through probabilistic shaping can improve the error correction performance for additive white Gaussian noise (AWGN) channels [49]. Similarly, PS has been proposed to maximize the mutual information between transmit and receive signals for non-linear distortion effects in AWGN channels [50]. In addition, rotated-QAM based PS scheme is also analyzed for ergodic constellation constrained capacity maximization rendering closed-form pairwise error probability (PEP). The GS is implemented using rotation and then probability mass function is separately designed to attain a hybrid effect [51]. A very recent study has proposed joint PS/GS design using training neural networks for joint optimization of bit-wise mutual information in bit-interleaved coded modulation systems [52]. Nevertheless, the literature for hardware distorted communications with PS as well as hybrid shaping (HS) is still lacking. To the best of authors' knowledge, these asymmetric schemes have not been used to enhance the error performance or to realize the IGS for wireless communication systems with improper HWDs.

C. Contributions

In this paper, we propose PS as a method to realize improper signaling, which is beneficial in mitigating the impact of HWD on the BER performance. Motivated by IGS's theoretical results in various scenarios [2] and the issues associated with GS, such as high shaping gap and coarse granularity, we adopt PS to realize the IGS scheme and combat HWD to assure reliable communications. In the following, we summarize the main contributions as:

- We derive the optimal maximum a posterior (MAP) detector for a discrete AS and carry out BER analysis for the adopted HWD communication system.
- We design the probabilistic shaped AS under power and rate constraints for hardware distorted system and propose adaptive algorithm that tune the symbol probabilities for PS to minimize the BER performance.
- We further suggest a hybrid shaped AS scheme that reaps benefits of both PS and GS and present an adaptive algorithm that tune both signal probability and shaping parameters.

D. Paper Organization and Notation

The rest of the paper is organized as: Section II describes HWD model and optimal receiver for the adopted system. In section III, we present the error probability analysis using the union bound on pairwise error probability and derive BER for generalized M-ary modulation scheme. Next, we propose PS design using successive convex programming (SCP) algorithm and some toy examples for comprehensive illustration in section IV. Later, HS parameterization and design along with the respective MAP and error probability analysis is carried out in section V, followed by the numerical results in Section VI and the conclusion in Section VII.

Notations: In this paper, $|a|$ and a^* represent the absolute and complex conjugate of a scalar complex number a. The probability of an event A is defined as $\Pr(A)$. The notations $f_z(z)$ and $f_{z|y}(z|y)$ denote the probability density function (PDF) and conditional PDF of a random variable (r.v.) z given y. The operator $\mathbb{E}[.]$ denotes the expected value. Considering a r.v. Λ, the real/in-phase and imaginary/quadrature-phase components of Λ are denoted as Λ_{I} and Λ_{Q}, respectively. Moreover, $f'(x)$ denotes the first order derivative of $f(x)$ with respect to x. Additionally, $\mathcal{Z}^+$ represents a set of positive integers. $\mathbf{v} = [v_I \quad v_Q]^{\mathrm{T}}$ is the real-composite vector representation of the complex number $v = v_I + i\, v_Q$. Furthermore, $x^{(k)}$ and $\mathbf{p}^{(k)}$ represent the instance values of the variable x and vector $\mathbf{p}$, respectively, in the k^{th} iteration of an algorithm.

II. System Description

Impropriety incorporation is crucial for the systems dealing with improper signals, noise, or interference. Such characterization helps in meticulous system modeling, accurate performance analysis, and optimum signaling design. We direct readers to [2, eq. 4-7] and [2, Def. 1a, 6-9] for some preliminaries of statistical impropriety characterization. This will help to comprehend the impropriety concepts in the adopted system model with HWD. Then, the transceiver HWD model is described, and the optimal receiver is derived.

A. Transceiver Hardware Distortion Model

Consider a single-link wireless communication system suffering from various hardware impairments. The non-linear transfer functions of various transmitter RF stages, such as digital-to-analog converter, band-pass filter and high power amplifier result in accumulative additive distortion noise $\eta_{\mathrm{t}} \sim \mathcal{CN}(0, \kappa_{\mathrm{t}}, \tilde{\kappa}_{\mathrm{t}})$, where $|\tilde{\kappa}_{\mathrm{t}}| \leq \kappa_{\mathrm{t}}$ [5], [53].[1] These distortions raise the noise floor of the transmitted signal $x_{\mathrm{tx}} = x_m + \eta_{\mathrm{t}}$, where x_m is the single-carrier band-pass modulated signal taken from M-ary QAM or M-ary PSK constellation with a probability mass function $p_m \triangleq p_X(x_m)$ rendering the transmission probability of symbol x_m, and $\mathbf{p} \triangleq [p_1, p_2, \cdots, p_M]$. Let us define the set that includes all possible symbol distributions as

$$\mathbb{S} = \left\{ \mathbf{p} : \mathbf{p} = [p_1, p_2, \cdots, p_M], \sum_{j=1}^{M} p_j = 1, p_j \geq 0, \, \forall j \in \{1, 2, \cdots, M\} \right\}.$$

The transmitted signal further undergoes a slowly varying flat Rayleigh fading channel $g \sim \mathcal{CN}(0, \lambda, 0)$. Moreover, the receiver further induces an additive distortion η_{r}, resulting from the non-linear transfer function of low noise amplifier, band-pass filters, image rejection low pass filter, analog-to-digital converter. It is important to highlight that the receiver distortions are in addition to the conventional thermal noise at the receiver.

$$y = \sqrt{\alpha}\, g \; (x_m + \eta_{\mathrm{t}}) + \eta_{\mathrm{r}} + w; \quad m \in \{1, 2, \cdots, M\}, \tag{1}$$

where α is the transmitted power. The AWGN w and receiver HWD η_{r} are distributed as $w \sim \mathcal{CN}(0, \sigma_w^2, 0)$ and $\eta_{\mathrm{r}} \sim \mathcal{CN}(0, \alpha |g|^2 \kappa_{\mathrm{r}}, \alpha g^2 \tilde{\kappa}_{\mathrm{r}})$. The additive Gaussian distortion model for the aggregate residual RF distortions is backed by various theoretical investigations and measurement results, see, e.g., [6]–[8], [53]–[58] and references therein. This can also be motivated analytically by the central limit theorem. Furthermore, the improper nature of these additive distortions is motivated by the widely linear transformation (WLT) caused by the imbalance between in-phase and quadrature-phase branches during the up-conversion and down-conversion phases [59].

Lemma 1 (Aggregate effect of transceiver distortions [5], [19])**.** *For the following generalized received signal model*

$$y = \sqrt{\alpha} g x_m + z; \quad m \in \{1, 2, \ldots, M\}, \tag{2}$$

[1] The additive distortion noise is the equivalent aggregate baseband model of accumulative HWDs from various RF blocks [5].

where $z \triangleq \sqrt{\alpha} g\eta + w$ is the aggregate interference, accumulating the effect of both transmitter and receiver distortions $\eta \sim \mathcal{CN}(0, \kappa, \tilde{\kappa})$ with variance $\kappa = \kappa_{\rm t} + \kappa_{\rm r}$, pseudo-variance $\tilde{\kappa} = \tilde{\kappa}_{\rm t} + \tilde{\kappa}_{\rm r}$, and circularity coefficient $C_\eta = |\tilde{\kappa}|/\kappa$ in addition to the thermal noise w. Thus, z is the r.v. which can be modeled as an improper noise, i.e., $z \sim \mathcal{CN}\left(0, \upsilon, \tilde{\upsilon}\right)$, where $\upsilon = \alpha |g|^2 \kappa + \sigma_w^2$ and $\tilde{\upsilon} = \alpha g^2 \tilde{\kappa}$. The degree of improperness of this aggregate interference can be categorized using circularity quotient $C_z = \tilde{\upsilon}/\upsilon$.

It is important to note that (2) reduces to the conventional signal model $y = \sqrt{\alpha} g x_m + w$ in case of ideal hardware, i.e., $\kappa = 0$, which is induced by imposing $\kappa_{\rm t} = \kappa_{\rm r} = 0$ and also $\tilde{\kappa} = 0$ [2]. Exploiting the relation between the υ, $\tilde{\upsilon}$ and the variances $\sigma_I^2 = E\{z_I^2\}$, $\sigma_Q^2 = E\{z_Q^2\}$ along with their mutual correlation $r_{z_I z_Q} = E\left\{z_I z_Q\right\}$, we get $\upsilon = E\left\{|z|^2\right\} = \sigma_I^2 + \sigma_Q^2$ and $\tilde{\upsilon} = E\left\{z^2\right\} = \sigma_I^2 - \sigma_Q^2 + i2r_{z_I z_Q}$. Their inter relation enables us to evaluate the variance of $z_{\rm I}$ and $z_{\rm Q}$ as given in (3) and (4), respectively, as well as

$$\sigma_I^2 = \frac{\upsilon + \tilde{\upsilon}_I}{2} = \frac{\alpha |g|^2 \kappa + \sigma_w^2 + \alpha \Re\left(g^2 \tilde{\kappa}\right)}{2}, \tag{3}$$

$$\sigma_Q^2 = \frac{\upsilon - \tilde{\upsilon}_I}{2} = \frac{\alpha |g|^2 \kappa + \sigma_w^2 - \alpha \Re\left(g^2 \tilde{\kappa}\right)}{2}, \tag{4}$$

Furthermore, the non-zero pseudo-variance $\tilde{\upsilon} = \tilde{\sigma}_z^2$ motivates us to evaluate the correlation between $z_{\rm I}$ and $z_{\rm Q}$, i.e., correlation coefficient ρ_z using $r_{z_I z_Q}$ as

$$\rho_z = \frac{r_{z_I z_Q}}{\sigma_I \sigma_Q} = \frac{0.5 \tilde{\upsilon}_Q}{\sigma_I \sigma_Q} = \frac{\alpha \Im\left(g^2 \tilde{\kappa}\right)}{\sqrt{\left(\alpha |g|^2 \kappa + \sigma_w^2\right)^2 - \left(\alpha \Re\left(g^2 \tilde{\kappa}\right)\right)^2}}. \tag{5}$$

HWD can leave drastic effects on the system performance as they raise the noise floor. Although, the entropy loss of improper noise is less than the proper noise but it is difficult to tackle. It requires some meticulously designed improper signaling like IGS for effective mitigation. However, IGS is difficult to implement because of the unbounded peak-to-average power ratio and high detection complexity [2], [23]. Therefore, researchers resort to the finite discrete AS schemes obtained by GS.

We propose PS as another way to realize AS in order to effectively dampen the deteriorating effects of improper HWD. PS aims to design non-uniform symbol probabilities for a higher order QAM to minimize BER offering more degrees of freedom and adaptive rates. In the following section, we carry out the error probability analysis of the adopted system which lays foundation for the proposed PS design.

B. Optimal Receiver

Conventional systems with Gaussian interference employ least-complex receivers with either minimum Euclidean or maximum likelihood detectors. However, such receivers cannot accommodate the unequal symbol probabilities and improper noise. Therefore, the optimal detection in the presented scenario can only be achieved by the MAP detector at the expense of increased receiver complexity. Considering the improper Gaussian HWD and the non-uniform priors of the constellation symbols, the optimal MAP detection is given by

$$\hat{m}_{\mathrm{PS}} = \underset{1\leq m\leq M}{\arg\max} \quad p_X(x_m) f_{Y_{\mathrm{I}},Y_{\mathrm{Q}}|X,g}\left(y_{\mathrm{I}}, y_{\mathrm{Q}}|x_m, g\right), \tag{6}$$

where $f_{Y_{\mathrm{I}},Y_{\mathrm{Q}}|X,g}\left(y_{\mathrm{I}}, y_{\mathrm{Q}}|x_m, g\right)$ is the conditional Gaussian PDF of y representing maximum likelihood (ML) function given x_m and g [2] derived using [2, eq. 43], as expressed in (7)

$$f_{Y_{\mathrm{I}},Y_{\mathrm{Q}}|X,g}\left(y_{\mathrm{I}}, y_{\mathrm{Q}}|x_m, g\right)=\frac{1}{2\pi\sigma_{\mathrm{I}}\sigma_{\mathrm{Q}}\sqrt{1-\rho_z^2}}\exp\left\{\frac{-1}{2\left(1-\rho_z^2\right)}\left[\begin{array}{c}\frac{\left(y_{\mathrm{I}}-\sqrt{\alpha}\Re(gx_m)\right)^2}{\sigma_{\mathrm{I}}^2}+\frac{\left(y_{\mathrm{Q}}-\sqrt{\alpha}\Im(gx_m)\right)^2}{\sigma_{\mathrm{Q}}^2}+\\-\frac{2\rho_z\left(y_{\mathrm{I}}-\sqrt{\alpha}\Re(gx_m)\right)\left(y_{\mathrm{Q}}-\sqrt{\alpha}\Im(gx_m)\right)}{\sigma_{\mathrm{I}}\sigma_{\mathrm{Q}}}\end{array}\right]\right\}. \tag{7}$$

III. Error Probability Analysis

Considering the non-uniform priors and improper noise, the error probability analysis is carried out based on the optimal MAP detector presented in Section II. Symbol error probability P_{s} is the accumulated error probability of all symbols with respect to their prior probabilities and is given as

$$P_{\mathrm{s}} = \sum_{m=1}^{M} p_m \Pr\left(e|x_m\right), \tag{8}$$

where $\Pr\left(e|x_m\right)$ is the probability of an error event given symbol x_m was transmitted. In order to yield a tractable and simplified analysis especially for higher order modulation schemes, P_{s} can be upper bounded as

$$P_{\mathrm{s}} \leq \sum_{m=1}^{M} \sum_{\substack{n=1\\n\neq m}}^{M} p_m P_{mn}, \tag{9}$$

where, P_{mn} is the PEP, which represents the probability of deciding x_n given x_m was transmitted, ignoring all the other symbols in the constellation [60]. The PEP can be evaluated using the MAP rule in (6) as

$$P_{mn} = \Pr\left\{p_m\, f_{Y_{\mathrm{I}},Y_{\mathrm{Q}}|X,g}\left(y_{\mathrm{I}}, y_{\mathrm{Q}}|x_m, g\right) \leq p_n f_{Y_{\mathrm{I}},Y_{\mathrm{Q}}|X,g}\left(y_{\mathrm{I}}, y_{\mathrm{Q}}|x_n, g\right)\right\}. \tag{10}$$

[2]This work assumes M-QAM modulation in a coherent system. Therefore, channel state information is first estimated at the receiver and then shared with the transmitter which is used in both optimization and coherent detection stages.

By substituting the conditional probability from (7) in (10) and after some mathematical simplifications, the PEP can be written as in (11),

$$P_{mn}=\Pr\left\{2\left(1-\rho_z^2\right)\ln\left(\frac{p_m}{p_n}\right)\leq\left[\begin{array}{l}\frac{\left(y_{\mathrm{I}}-\sqrt{\alpha}\Re(gx_m)\right)^2-\left(y_{\mathrm{I}}-\sqrt{\alpha}\Re(gx_n)\right)^2}{\sigma_{\mathrm{I}}^2}+\frac{\left(y_{\mathrm{Q}}-\sqrt{\alpha}\Im(gx_m)\right)^2-\left(y_{\mathrm{Q}}-\sqrt{\alpha}\Im(gx_n)\right)^2}{\sigma_{\mathrm{Q}}^2}+\\+\frac{2\rho_z\left(y_{\mathrm{I}}-\sqrt{\alpha}\Re(gx_n)\right)\left(y_{\mathrm{Q}}-\sqrt{\alpha}\Im(gx_n)\right)-2\rho_z\left(y_{\mathrm{I}}-\sqrt{\alpha}\Re(gx_m)\right)\left(y_{\mathrm{Q}}-\sqrt{\alpha}\Im(gx_m)\right)}{\sigma_{\mathrm{I}}\sigma_{\mathrm{Q}}}\end{array}\right]\right\}. \tag{11}$$

Now, we find the in-phase and quadrature-phase components of the received signal y for a given transmitted symbol x_m as follows

$$y_{\mathrm{I}}=\sqrt{\alpha}\Re\left(gx_m\right)+z_I, \tag{12}$$

$$y_{\mathrm{Q}}=\sqrt{\alpha}\Im\left(gx_m\right)+z_Q, \tag{13}$$

respectively. Then, we substitute y_{I} and y_{Q} in (11), which can be further simplified obtaining,

$$P_{mn}=\Pr\left\{\psi\geq 2\left(1-\rho_z^2\right)\ln\left(\frac{p_m}{p_n}\right)+\alpha\gamma_{mn}\right\}, \tag{14}$$

Let $\xi_{mn}=g\,d_{mn}=g\left(x_m-x_n\right)$, with given channel coefficient g and d_{mn} representing the distance between the m^{th} and n^{th} symbol. Then, γ_{mn} in (14) can be defined as

$$\gamma_{mn}\triangleq\frac{\xi_{mn\,I}^{\;2}}{\sigma_{\mathrm{I}}^2}+\frac{\xi_{mn\,Q}^{\;2}}{\sigma_{\mathrm{Q}}^2}-\frac{2\rho_z\xi_{mn\,I}\xi_{mn\,Q}}{\sigma_{\mathrm{I}}\sigma_{\mathrm{Q}}}, \tag{15}$$

where $\xi_{mn\,I}$ and $\xi_{mn\,Q}$ are the real and imaginary components of ξ_{mn}. Moreover, ψ in (14) is obtained by the superposition of z_I and z_Q as

$$\psi=2\sqrt{\alpha}\rho_z\left[\left(\frac{\xi_{mn\,Q}}{\sigma_{\mathrm{I}}\sigma_{\mathrm{Q}}}-\frac{\xi_{mn\,I}}{\rho_z\sigma_{\mathrm{I}}^2}\right)z_I+\left(\frac{\xi_{mn\,I}}{\sigma_{\mathrm{I}}\sigma_{\mathrm{Q}}}-\frac{\xi_{mn\,Q}}{\rho_z\sigma_{\mathrm{Q}}^2}\right)z_Q\right]. \tag{16}$$

Clearly, ψ is another zero mean Gaussian random variable with variance σ_ψ^2 expressed as

$$\sigma_\psi^2=4\left(1-\rho_z^2\right)\alpha\gamma_{mn}. \tag{17}$$

Conclusively, P_{mn} is the complementary cumulative distribution function of ψ and is given as

$$P_{mn}=\mathcal{Q}\left(\frac{2\left(1-\rho_z^2\right)\ln\left(\frac{p_m}{p_n}\right)+\alpha\gamma_{mn}}{2\sqrt{\left(1-\rho_z^2\right)\alpha\gamma_{mn}}}\right). \tag{18}$$

Substituting the PEP derived in (18) to (9) along with the Gray coding assumption yields the following theoretical upper-bound on BER

$$P_{\mathrm{b}}\leq\mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}\triangleq\frac{1}{\log_2(M)}\sum_{m=1}^{M}\sum_{\substack{n=1\\n\neq m}}^{M}p_m\mathcal{Q}\left(\beta_{mn}\ln\left(\frac{p_m}{p_n}\right)+\frac{1}{2\beta_{mn}}\right), \tag{19}$$

where $\beta_{mn} \triangleq \sqrt{1-\rho_z^2}/\sqrt{\alpha\gamma_{mn}}$. The BER expression depends on the constellation size, prior probabilities of the symbols, power budget, mutual distances between the transmitted and received erroneous symbols under Rayleigh fading, and HWD statistical characteristics. Interestingly, for an ideal system employing PS without HWDs, the BER bound in (19) reduces to

$$P_{\rm b} \leq \mathrm{P}_{\rm b}^{\rm UB} \triangleq \frac{1}{\log_2(M)} \sum_{m=1}^{M} \sum_{\substack{n=1 \\ n\neq m}}^{M} p_m Q\left(\frac{\sigma_w^2 ln\left(\frac{p_m}{p_n}\right) + \alpha |g\, d_{mn}|^2}{\sqrt{2\alpha\sigma_w^2 |g\, d_{mn}|^2}} \right) \tag{20}$$

which can be derived from (19) for $\kappa = \tilde{\kappa} = 0$. This will render uncorrelated proper/symmetric noise with $\sigma_I^2 = \frac{\sigma_w^2}{2}$, $\sigma_Q^2 = \frac{\sigma_w^2}{2}$ and $\rho_z = 0$. We can use this as a benchmark to quantify the performance loss caused by the HWDs.

In contrast to the monotonically decreasing BER for the ideal systems (20), the BER in (19) saturates after a specific SNR in the hardware-distorted transceivers. In this regard, we carry out the asymptotic analysis of the bit error probability to quantify the error floor as high SNR. Let us set

$$\Upsilon = \lim_{\alpha\to\infty} \left(1-\rho_z^2\right) \triangleq 1 - \frac{\left(\Im\left(g^2\tilde{\kappa}\right)\right)^2}{\left(|g|^4\kappa^2\right) - \left(\Re\left(g^2\tilde{\kappa}\right)\right)^2}, \tag{21}$$

and

$$\lim_{\alpha\to\infty} \alpha\gamma_{\rm mn} = {\xi_{mn}}_I^2 \left(\lim_{\alpha\to\infty} \frac{\sigma_{\rm I}^2}{\alpha}\right)^{-1} + {\xi_{mn}}_Q^2 \left(\lim_{\alpha\to\infty} \frac{\sigma_{\rm Q}^2}{\alpha}\right)^{-1} - 2{\xi_{mn}}_I {\xi_{mn}}_Q \left(\lim_{\alpha\to\infty} \frac{\sigma_{\rm I}\sigma_{\rm Q}}{\alpha\rho_z}\right)^{-1}. \tag{22}$$

Thus, using (3)-(5) to evaluate the limit in (22) and further substituting (21) and (26) to obtain $\lim_{\alpha\to\infty} \beta_{\rm mn}$ renders the following error floor

$$\lim_{\alpha\to\infty} P_{\rm b} \leq \frac{1}{\log_2(M)} \sum_{m=1}^{M} \sum_{\substack{n=1 \\ n\neq m}}^{M} p_m Q\left(\frac{2\Upsilon \ln\left(\frac{p_m}{p_n}\right) + \left(\frac{2\Re(gd_{mn})^2}{|g|^2\kappa+\Re(g^2\tilde{\kappa})} + \frac{2\Im(gd_{mn})^2}{|g|^2\kappa-\Re(g^2\tilde{\kappa})} - 2\frac{\Re(gd_{mn})\Im(gd_{mn})\Im\left(g^2\tilde{\kappa}\right)}{\left(|g|^2\kappa\right)^2-(\Re(g^2\tilde{\kappa}))^2} \right)}{\sqrt{4\Upsilon\left(\frac{2\Re(gd_{mn})^2}{|g|^2\kappa+\Re(g^2\tilde{\kappa})} + \frac{2\Im(gd_{mn})^2}{|g|^2\kappa-\Re(g^2\tilde{\kappa})} - 2\frac{\Re(gd_{mn})\Im(gd_{mn})\Im(g^2\tilde{\kappa})}{\left(|g|^2\kappa\right)^2-(\Re(g^2\tilde{\kappa}))^2} \right)}} \right). \tag{23}$$

We can see that the error floor depends on the adopted M-ary constellation, channel coefficient, HWD statistical characteristics, and symbol probabilities.

IV. Proposed Probabilistic Signaling Design

We aim to design the non-uniform symbol probabilities, which minimize the BER of the adopted system suffering from HWD. The optimization is carried out given power and rate constraints. The rate of the conventional QAM with uniform symbol probabilities and modulation order $M_{\rm u}$ is fixed, i.e., $R = \log_2(M_{\rm u})$. However, we seek the maximum benefits of PS by allowing a higher-order modulation with $M_{\mathrm{nu}} > M_{\mathrm{u}}$, where M_{nu}, which is also fixed, is the modulation order of the constellation with non-uniform probabilities $\mathbf{p}$. Thus, the rate of this scheme can be designed such that $R \triangleq \mathrm{H}(\mathbf{p}) \geq \log_2(M_{\mathrm{u}})$, rendering more design flexibility and hence is capable of reducing the BER. PS is capable of changing the transmission rate by changing the symbol distribution for a fixed modulation order, unlike uniform signaling, which needs to change the modulation scheme's order to change the rate for uncoded communications.

After designing the symbol probabilities, we can implement PS by using distribution matching at the transmitter to map uniformly distributed input bits to $M_{\mathrm{nu}}-$QAM/PSK symbols [30]–[35]. For instance in constant composition DM, the designed distribution is quantized with a step size of $1/n_p$, where n_p is the frame length. The target PMF p is approximated by an n_p-type distribution, i.e., $\tilde{p} \sim [z_1/n_p, z_2/n_p, \ldots, z_{M_{\mathrm{nu}}}/n_p]$, where z_j is a positive integer representing the number of times at which the j^{th} symbol appears in the frame. The quantization error between the p and $\tilde{p}$ is negligible for asymptotically large n_p. Moreover, a large n_p allows us to realize very small symbol probabilities with the least possible symbol appearing just once in the entire frame yielding a minimum resolution of $1/n_p$.

At the receiver, the symbols can be detected using the proposed MAP detector (6) that incorporates the prior symbol distribution. Finally, the estimated PS symbols are inverted back to the uniformly distributed bits using distribution dematching. Arithmetic codes can be used to achieve low complexity invertible mapping between the uniform bits and PS symbols and vice versa [30].

In the following, we formulate the PS design problem and propose an algorithm to obtain the non-uniform symbol probabilities followed by some toy examples.

A. Problem Formulation

In this section, we consider a fixed symmetric $M_{\mathrm{nu}}-$ QAM/PSK constellation. The vector $\mathbf{p} \triangleq [p_1, p_2, \ldots, p_{M_{\mathrm{nu}}}]$ represents the probabilities of the PS shaped symbols. The non-uniform probabilities are designed to minimize the upper bound on the BER derived in (19). In particular, we formulate the problem as

$$\mathbf{P1}: \underset{\mathbf{p}\in\mathbb{S}}{\text{minimize}} \quad \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}(\mathbf{p}) \tag{24a}$$

$$\text{subject to} \quad \sum_{m=1}^{M_{\mathrm{nu}}} |x_m|^2 p_m \leq 1, \tag{24b}$$

$$\mathrm{H}(\mathbf{p}) \geq \log_2(M_{\mathrm{u}}), \tag{24c}$$

where (24b) and (24c) represent the average power and rate constraints, respectively, and $\mathrm{H}(\mathbf{p})$ is the source entropy, which represents the transmit rate in terms of bits/symbol and is defined as

$$\mathrm{H}(\mathbf{p}) \triangleq \sum_{m=1}^{M_{\mathrm{nu}}} -p_m \log_2(p_m). \tag{25}$$

Since communication systems with higher transmission rate and lower BER are always preferable, the constraint (24c) assures that the PS has at least the same rate of the uniform scheme. Interestingly, the concave nature of the entropy function renders a convex constraint in $\mathbf{p}$. On the contrary, if we consider designing the two systems to have exactly the same rate, i.e, $\mathrm{H}(\mathbf{p}) = \log_2(M_{\mathrm{u}})$ instead of (24c), this would yield not only non-convexity, but also may cause degradation in performance by shrinking the feasible area.

Another observation for the rate constraint is that M_{nu} should be larger than or equal M_{nu} so that (24c) is satisfied. For $M_{\mathrm{nu}} = M_{\mathrm{u}}$, the distribution should be uniform to satisfy the rate constraint because uniform signaling has the largest entropy. Therefore, $M_{\mathrm{nu}} > M_{\mathrm{u}}$ is the preferred choice to attain non-uniform probabilities which can render significant performance gains. Therefore, the idea is to minimize the BER by properly designing a higher order non-uniformly distributed M_{nu}-QAM/PSK with the same energy and at least the same rate as those of a lower order uniformly distributed $M_{\mathrm{u}}-$QAM/PSK.

B. Optimization Framework

The optimization problem **P1** (24) is a non-convex optimization problem owing to the non-convex objective function even though all the constraints are convex. Therefore, we propose successive convex approximation approach to tackle it. We begin by approximating $\mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}(\mathbf{p})$ with its first order Taylor series approximation. First order Taylor series approximation of a function $f(x)$ around a point $x^{(k)}$ is given as

$$\tilde{f}\left(x, x^{(k)}\right) \approx f\left(x^{(k)}\right) + \nabla_x f\left(x^{(k)}\right)\left(x - x^{(k)}\right). \tag{26}$$

Thus, we need to compute $\nabla_{\mathbf{p}} \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}$ and evaluate it at $\mathbf{p}^{(k)}$ to compute $\tilde{\mathrm{P}}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p}, \mathbf{p}^{(k)}\right)$.

$$\nabla_{\mathbf{p}} \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}} = \left[\frac{\partial \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}}{\partial p_1} \quad \frac{\partial \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}}{\partial p_2} \quad \dots \quad \frac{\partial \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}}{\partial p_{M_{\mathrm{nu}}}}\right]. \tag{27}$$

In order to compute $\partial \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}} / \partial p_t$, we rewrite (19) as

$$\mathrm{P}_{\mathrm{b}}^{\mathrm{UB}} = \frac{1}{\log_2(M_{\mathrm{nu}})} \sum_{m=1}^{M_{\mathrm{nu}}} \sum_{\substack{n=1 \\ n \neq m}}^{M_{\mathrm{nu}}} p_m \int_{\Omega_{mn}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du, \tag{28}$$

Algorithm 1 Successive Convex Programming

Initialize $i \leftarrow 0$, $\epsilon \leftarrow \infty$ and **Set** tolerance δ
Choose feasible starting point $\mathbf{p}^{(i)}$
while $\epsilon \geq \delta$ **do**
  Evaluate $\tilde{\mathrm{P}}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p}, \mathbf{p}^{(i)}\right)$
  Solve **P1a** and obtain $\mathbf{p}$ using $\mathbf{p}^{(i)}$
  Update $\mathbf{p}^{(i+1)} \leftarrow \mathbf{p}$, $\epsilon \leftarrow \left\|\mathbf{p}^{(i+1)} - \mathbf{p}^{(i)}\right\|$, and $i \leftarrow i+1$
end while
$\mathbf{p}^{*} \leftarrow \mathbf{p}^{i+1}$
$\mathrm{P}_{\mathrm{b}}^{*} \leq \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathcal{P}^{*}\right)$

where

$$\Omega_{mn} = \beta_{mn} \ln\left(\frac{p_m}{p_n}\right) + \frac{1}{2\beta_{mn}}. \tag{29}$$

From (28) and by applying the Leibniz integral rule, we get

$$\frac{\partial \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}}{\partial p_t} \leq \frac{1}{\log_2\left(M_{\mathrm{nu}}\right)} \sum_{\substack{n=1,\\ n\neq t,\\ m=t}}^{M_{\mathrm{nu}}} \left(\mathcal{Q}\left(\Omega_{mn}\right) - \frac{\beta_{mn}}{\sqrt{2\pi}} e^{-\frac{\Omega_{mn}^2}{2}}\right) + \frac{1}{\log_2\left(M_{\mathrm{nu}}\right)} \sum_{\substack{m=1,\\ m\neq t,\\ n=t}}^{M_{\mathrm{nu}}} \frac{\beta_{mn} p_m}{\sqrt{2\pi} p_n} e^{-\frac{\Omega_{mn}^2}{2}}. \tag{30}$$

Now, $\mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}$ can be approximated from (26), (27), and (30) using first order Taylor series expansion around an initial probability vector $\mathbf{p}^{(k)}$ as

$$\tilde{\mathrm{P}}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p}, \mathbf{p}^{(k)}\right) \triangleq \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p}^{(k)}\right) + \nabla_{\mathbf{p}} \mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p}^{(k)}\right)\left(\mathbf{p} - \mathbf{p}^{(k)}\right). \tag{31}$$

Successive convex programming minimizes **P1** by iteratively solving its convex approximation **P1a** as presented in Algorithm 1.

$$\begin{aligned} \mathbf{P1a}: \underset{\mathbf{p} \in \mathbb{S}}{\text{minimize}} \quad & \tilde{\mathrm{P}}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p}, \mathbf{p}^{(k)}\right) && \text{(32a)} \\ \text{subject to} \quad & \sum_{m=1}^{M_{\mathrm{nu}}} \left|x_m\right|^2 p_m \leq 1, && \text{(32b)} \\ & \mathrm{H}(\mathbf{p}) \geq \log_2\left(M_{\mathrm{u}}\right), && \text{(32c)} \end{aligned}$$

It begins with the initiation of counter i, stopping criteria ϵ and the stopping threshold δ. Secondly, we choose some feasible PMF set $\mathbf{p}^{(i)} \in \mathbb{S}$ which satisfies the constraints (24b) and (24c). This initial feasible point can be chosen in two different ways:

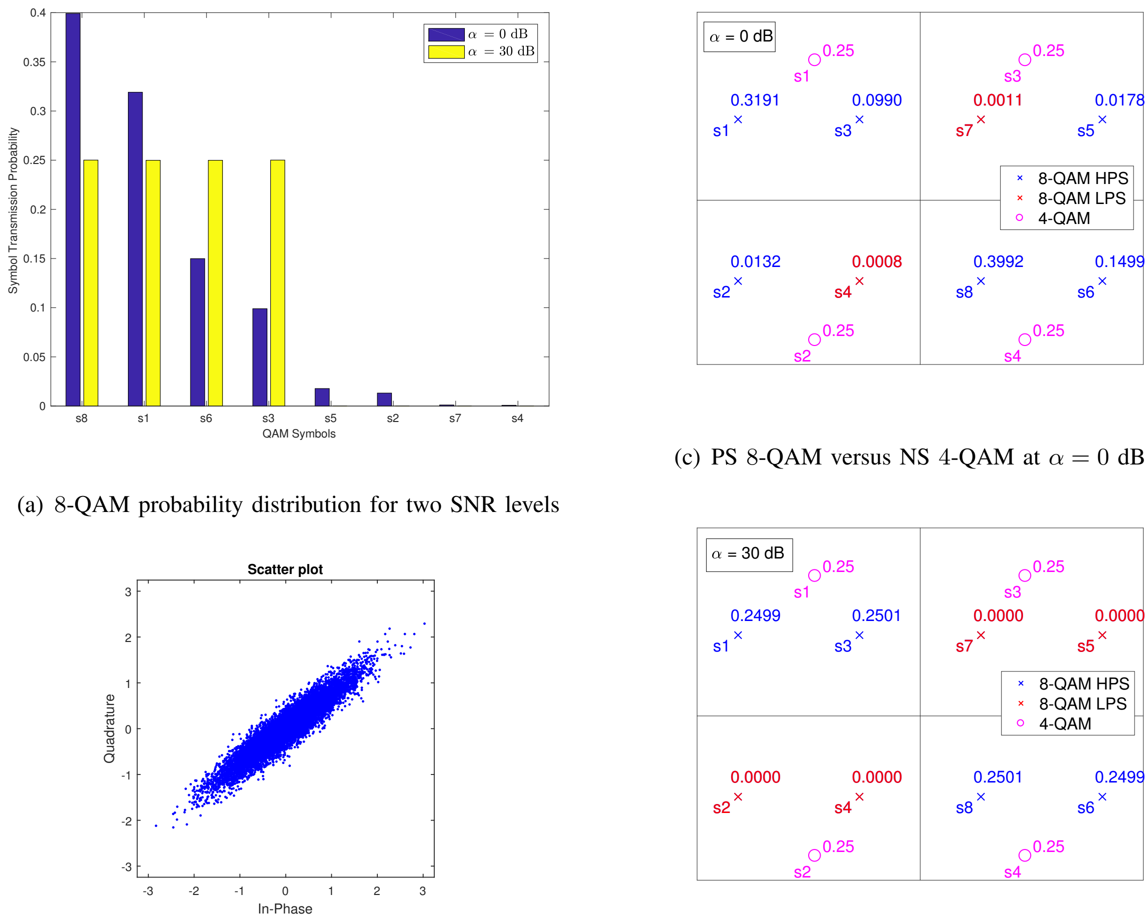
1) We can initiate with a uniform constellation with equi-probable M_{nu} symbols. This selection will render $\mathrm{H}(\mathbf{p}) = \log_2(M_{\mathrm{nu}})$ satisfying (24c) with strict inequality. Furthermore, the constellation is normalized to satisfy (24b) with strict equality.
2) We can start with a feasible M_{nu} constellation with any M_{u} symbols having probabilities $1/M_{\mathrm{u}}$ and the rest $M_{\mathrm{nu}} - M_{\mathrm{u}}$ symbols with zero probability of transmission. This choice will render $\mathrm{H}(\mathbf{p}) = \log_2(M_{\mathrm{u}})$ satisfying (24c) with strict equality. However, we can choose any M_{u} symbols from M_{nu} normalized constellation which yield average power less than 1, satisfying (24b).

The while loop starts by evaluating the approximation $\tilde{\mathrm{P}}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p}, \mathbf{p}^{(i)}\right)$ around $\mathbf{p}^{(i)}$. The convex problem **P1a** is solved using the Karush Kuhn Tucker (KKT) conditions derived in Appendix A to obtain the optimal probabilities for **P1a** [61]. The solution obtained in this iteration is updated as $\mathbf{p}^{(i+1)}$ and is used to evaluate the stopping criteria $\epsilon \leftarrow \left\|\mathbf{p}^{(i+1)} - \mathbf{p}^{(i)}\right\|$ as shown in Algorithm 1. The loop ends when the change in two subsequent solution parameters in terms of the ℓ_2 norm is less than a predefined threshold δ. Once the stopping criteria is attained, the solution parameters $\mathbf{p}^{(*)}$ are guaranteed to render a BER $\mathrm{P}_{\mathrm{b}}^{*}$ which will be lower than the bound $\mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}(\mathcal{P}^{*})$.

C. Toy Examples

A comprehensive illustration of probabilistically shaped $M_{\mathrm{nu}} = 8$-QAM with aggregate HWD power $\kappa = 0.99$ and a 2 bits/symbol rate constraint, corresponding to $M_{\mathrm{u}} = 4$, is presented in Fig. 1. The relation between optimal prior probabilities and two different SNR values is displayed in Fig. 1(a) when the system is subjected to highly improper distortion noise as shown in Fig. 1(b). Clearly in Fig. 1(a), the probability distribution is quite non-uniform for lower SNR level such as $\alpha = 0$ dB. However, it starts adopting uniform distribution of 0.25 for four of it's symbols, while zero probabilities for the remaining four symbols.

It is interesting to visualize the corresponding symbol constellations for both $\alpha = 0$ dB and $\alpha = 30$ dB. For $\alpha = 0$ dB, probabilistically shaped 8-QAM designates six symbols with significant transmission probabilities as highly probable symbol (HPS) whereas renders two symbols as least probable symbols (LPS) as depicted in Fig. 1(c). On the other hand, PS at $\alpha = 30$ dB only resorts to transmitting four of it's symbols, i.e., s1, s3, s6, and s8 (HPS) and discards the rest as depicted in Fig. 1(d). Notably, this technique assigns lowest probabilities to the symbols which are mostly affected by the highly improper noise in first and third quadrant. It is important to emphasize that the distortion power is proportional to the transmit power. This



(a) 8-QAM probability distribution for two SNR levels

(b) Improper HWD (κ=0.99)

(c) PS 8-QAM versus NS 4-QAM at $\alpha = 0$ dB

(d) PS 8-QAM versus NS 4-QAM at $\alpha = 30$ dB

Fig. 1: 8-QAM probability distribution at Rate = 2 bits/symbol.

strengthens distortions at high SNR and leads to the negligible transmission probabilities for the highly affected symbols. Hence, it is capable of achieving lower BER while maintaining 2 bits/symbol rate for a fair comparison with traditional 4-QAM using the same power budget.

V. Hybrid Shaping where Conventional meets State-of -the-Art

In this section, we increase the AS design flexibility by allowing joint GS and PS, which we call it here HS, to improve the underlying communication system performance further. Throughout the design procedure, HS transforms the equally spaced uniformly distributed QAM/PSK symbols to unequally spaced symbols in a geometric envelope with non-uniform prior distribution. Thus, HS aims to optimize the symbol probabilities (i.e., PS) and some spatial shaping parameters for the constellation (i.e., GS).

A. Hybrid Shaping Parameterization

Apart from the non-uniform priors, consider the asymmetric transmit symbol $\mathbf{v}_m = [v_{mI} \quad v_{mQ}]^{\mathrm{T}}$ resulting from the GS on the conventional baseband symmetric M-QAM/M-PSK symbol $\mathbf{x}_m = [x_{mI} \quad x_{mQ}]^{\mathrm{T}}$ as $\mathbf{v}_m = \mathbf{A}\mathbf{R}\mathbf{x}_m$, where

$$\mathbf{A}(\zeta) = \begin{bmatrix} \sqrt{1+\zeta} & 0 \\ 0 & \sqrt{1-\zeta} \end{bmatrix}, \tag{33}$$

with translation parameter $\zeta \in (0, 1)$. Furthermore, the rotation is given by

$$\mathbf{R}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \tag{34}$$

with rotation angle $\theta \in (0, \mu\,\pi/2)$ for some constant μ. Uniformly distributed symmetric M-QAM constellation has a rotation symmetry of $n\,\pi/2, n \in \mathcal{Z}^+$ rendering $\mu = n$ to be good choice for GS. However, non-uniformly distributed M-QAM constellation can only be rotationally symmetric after $2n\pi$, thus $\mu = 4n$ is suitable for HS. This technique renders non-uniformly spaced symbols in a parallelogram envelop. It is important to highlight that this transformation preserves the power requirement. Power invariance of the rotation is a well known fact in the literature [60]. However, the wisdom behind the structure of $\mathbf{A}(\zeta)$ is unfolded in the following theorem.

Remark 1. *GS parameterization using translation matrix $\mathbf{A}(\zeta)$ preserves the power invariance of a complex random variable and inculcates asymmetry/improperness with the circularity coefficient ζ.*

Proof. The proof is presented in Appendix B. Furthermore, the generalization of the same concept to the symmetric discrete constellations such as M-QAM and M-PSK is also described in Appendix B. □

B. Optimal Receiver

The optimal receiver for hybrid shaped AS is also a MAP detector as derived in (6), but with a modified reference constellation v_m in place of x_m for all $m \in \{1, 2, \cdots, M_{\mathrm{nu}}\}$. More precisely, the detected symbol, $\hat{m}_{\mathrm{HS}}$, is the one that maximizes the posterior distribution, i.e.,

$$\hat{m}_{\mathrm{HS}} = \underset{1 \le m \le M_{\mathrm{nu}}}{\arg\max} \quad p_V(v_m) f_{Y_{\mathrm{I}}, Y_{\mathrm{Q}}|V, g}(y_{\mathrm{I}}, y_{\mathrm{Q}}|v_m, g), \tag{35}$$

where, $f_{Y_{\mathrm{I}}, Y_{\mathrm{Q}}|V, g}(y_{\mathrm{I}}, y_{\mathrm{Q}}|v_m, g)$ is similar to (7) by replacing all appearances of x_m with v_m for all $m \in \{1, 2, \cdots, M_{\mathrm{nu}}\}$. It is worth noting that this MAP detector includes all the HS parameters.

At first, it includes non-uniform prior probabilities of the symbols $p_V(v_m)$ in the detection process unlike conventional ML detector. Next, it requires an updated reference constellation v_m to incorporate the GS parameters for appropriate detection.

C. Error Probability

HS follows the same BER bound as derived in (19) but with modified γ_{mn}. It can now be written using the following quadratic formulation as a function of ζ and θ.

$$\gamma_{mn}(\zeta,\theta) = \mathbf{x}_{mn}^{\mathrm{T}} \mathbf{R}(\theta)^{\mathrm{T}} \mathbf{A}(\zeta)^{\mathrm{T}} \mathbf{G} \mathbf{A}(\zeta) \mathbf{R}(\theta) \mathbf{x}_{mn}, \tag{36}$$

where $\mathbf{x}_{mn}$ is the real composite vector form of $\xi_{mn} = g d_{mn}$ given by

$$\mathbf{x}_{mn} = \begin{bmatrix} \xi_{mn\mathrm{I}} & \xi_{mn\mathrm{Q}} \end{bmatrix}^{T}, \tag{37}$$

and $\mathbf{G}$ contains the statistical characteristics of the aggregate noise including in-phase noise variance, quadrature-phase noise variance, and the correlation between these components.

$$\mathbf{G} = \begin{bmatrix} \frac{1}{\sigma_I^2} & \frac{-\rho_z}{\sigma_I \sigma_Q} \\ \frac{-\rho_z}{\sigma_I \sigma_Q} & \frac{1}{\sigma_Q^2} \end{bmatrix}. \tag{38}$$

Thus, the BER of HS can be upper bounded as

$$\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}}(\mathbf{p},\zeta,\theta) = \frac{1}{\log_2(M_{\mathrm{nu}})} \sum_{m=1}^{M_{\mathrm{nu}}} \sum_{\substack{n=1 \\ n\neq m}}^{M_{\mathrm{nu}}} p_m \mathcal{Q}\left(\frac{\sqrt{1-\rho_z^2}}{\sqrt{\alpha\gamma_{mn}(\zeta,\theta)}} \ln\left(\frac{p_m}{p_n}\right) + \frac{\sqrt{\alpha\gamma_{mn}(\zeta,\theta)}}{2\sqrt{1-\rho_z^2}} \right). \tag{39}$$

D. Problem Formulation

HS targets the joint design of PS PMF $\mathbf{p}$ and GS parameters involving translation ζ and rotation θ parameter to minimize the BER bound given in (39).

$$\mathbf{P2}: \underset{\substack{\mathbf{p}\in\mathbb{S}, 0\leq\zeta\leq 1, \\ 0\leq\theta\leq 2\pi}}{\text{minimize}} \quad \mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}}(\mathbf{p},\zeta,\theta) \tag{40a}$$

$$\text{subject to} \quad \sum_{m=1}^{M_{\mathrm{nu}}} |v_m|^2 p_m \leq 1, \tag{40b}$$

$$\mathrm{H}(\mathbf{p}) \geq \log_2(M_{\mathrm{u}}), \tag{40c}$$

where the average power constraint (24b) is updated as (40b) to account for the possible change in the power of the symbols by geometrically shaping the constellation. However, the proposed rate constraint (40c) remains intact. Additionally, there are some boundary constraints on ζ and θ, respectively.

Intuitively, it is quite difficult to tackle this non-convex multimodal joint optimization problem. Therefore, we resort to the alternate optimization of PS parameters $(\mathbf{p})$ and GS parameters (ζ, θ) using sub-problems **P2a** and **P2b**, respectively. Problem **P2a** designs the PS parameters for some given ζ and θ. It is quite similar to the problem **P1** and thus, can be solved using Algorithm 1.

$$\mathbf{P2a} : \underset{\mathbf{p}\in\mathbb{S}}{\text{minimize}} \quad \mathrm{P}^{\mathrm{UB}}_{\mathrm{b,HS}}\left(\mathbf{p}, \zeta, \theta\right) \tag{41a}$$

$$\text{subject to} \quad (40\text{b}), (40\text{c}). \tag{41b}$$

On the other hand, the GS optimization problem designs ζ and θ for fixed symbol probabilities $\mathbf{p}$, given as

$$\mathbf{P2b} : \underset{\substack{0\leq\zeta\leq 1,\\ 0\leq\theta\leq 2\pi}}{\text{minimize}} \; \mathrm{P}^{\mathrm{UB}}_{\mathrm{b,HS}}\left(\mathbf{p}, \zeta, \theta\right). \tag{42}$$

The optimization problem **P2b** is a multimodal non-convex problem which is hard to tackled even by the SCP approach as employed in Section IV. The difficulty arises due to the absence of any constraints which restrict the feasibility region. The feasibility space enclosed by the boundary constraints is highly insufficient to serve our purpose. Therefore, we can approximate the solution using any of the following two methods

- Trust region reflective method: This method defines a trust region around a specific initial point and then approximate the function within that region. The convex approximation is the first order Taylor series approximation using the gradient. It begins by minimizing convex approximation of the function to obtain a solution. This solution is the perturbation in the initial point rendering a new point which should minimize the original function. Otherwise, we need to shrunk the trust region and repeat the process. Reflections are used to increase the step size while satisfying box constraints. After each iteration, we receive a new point which renders a lower objective function than the initial point. This iterative approach leads us to a local minimum and stops when some specified stopping criterion are met [62], [63].
- Gradient descent: This method is a relatively faster approach to tackle the problem at hand. It is owing to the fact that it does not involve any approximation and underlying optimization. It begins with an initial point and keeps updating the point in the descent direction using the gradients and a step size until it reaches a local solution or satisfies some stopping criterion [61].

Interestingly, both of these methods require the gradients of $\mathrm{P}^{\mathrm{UB}}_{\mathrm{b,HS}}\left(\mathbf{p}, \zeta, \theta\right)$ with respect to ζ and θ. Gradients are used either to approximate the function with it's first order Taylor series

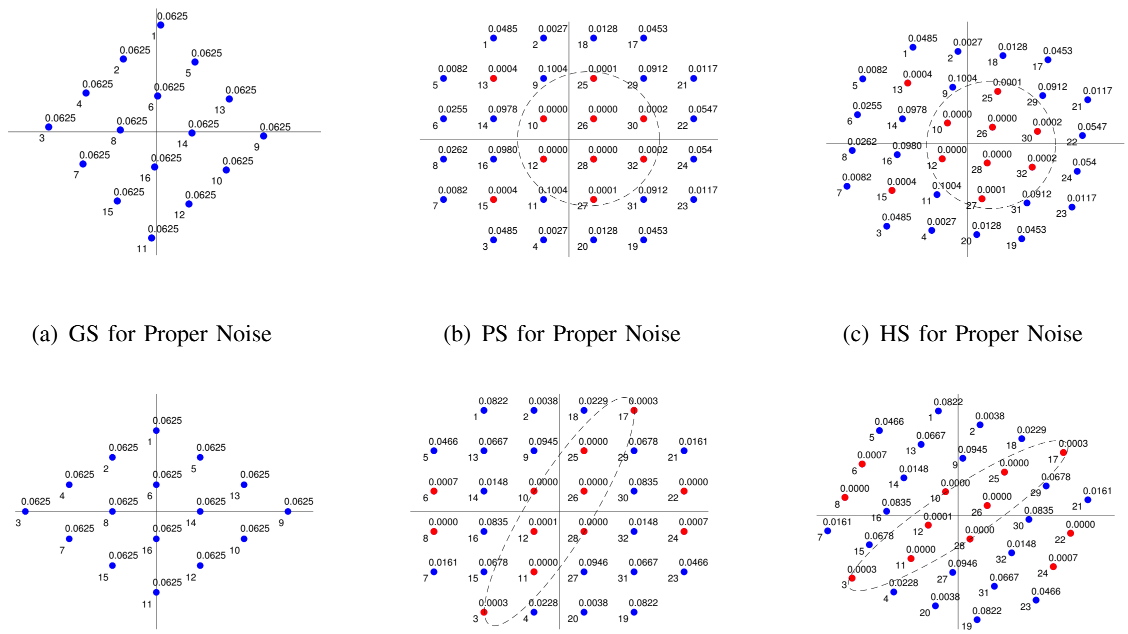
Algorithm 2 Alternate Optimization

1: **Initialize** $j \leftarrow 0$, $\epsilon \leftarrow \infty$ and **Set** tolerance δ
2: **Choose** feasible starting points $\mathbf{p}^{(j)}$, $\zeta^{(j)}$, and $\theta^{(j)}$.
3: **Evaluate** $\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}(j)}\left(\mathbf{p}^{(j)}, \zeta^{(j)}, \theta^{(j)}\right)$.
4: **while** $\epsilon \geq \delta$ **do**
5: Solve **P2a** using Algorithm 1 with starting point $\mathbf{p}^{(j)}$ and given $\zeta^{(j)}$, $\theta^{(j)}$ to obtain $\mathbf{p}^{(j*)}$
6: Solve **P2b** with starting points $\zeta^{(j)}$,$\theta^{(j)}$ and given $\mathbf{p}^{(j*)}$ to obtain $\zeta^{(j*)}$, $\theta^{(j*)}$
7: $\mathbf{p}^{(j+1)} \leftarrow \mathbf{p}^{(j*)}$, $\zeta^{(j+1)} \leftarrow \zeta^{(j*)}$, and $\theta^{(j+1)} \leftarrow \theta^{(j*)}$
8: Evaluate $\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}(j+1)}\left(\mathbf{p}^{(j+1)}, \zeta^{(j+1)}, \theta^{(j+1)}\right)$.
9: Update $\epsilon \leftarrow \left\|\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}(j+1)} - \mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}(j)}\right\|$ and $j \leftarrow j+1$
10: **end while**
11: Solution parameters: $\mathbf{p}^{*} \leftarrow \mathbf{p}^{j+1}$, $\zeta^{*} \leftarrow \zeta^{j+1}$, $\theta^{*} \leftarrow \theta^{j+1}$
12: Objective function: $\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}*} \leftarrow \mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}(j+1)}$
13: Consequence: $\mathrm{P}_{\mathrm{b,HS}}^{*} \leq \mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}*}$

approximation within a trust region or to find the next point in the descent direction. The gradients are evaluated and presented in Appendix C.

E. Proposed Algorithm

The joint optimization problem **P2** can be tackled using the alternate optimization algorithm as presented in Algorithm 2. It solves the sub problems **P2a** and **P2b** alternately and iteratively. It begins with some starting feasible points $\mathbf{p}^{(j)}$, $\zeta^{(j)}$, and $\theta^{(j)}$ and evaluates $\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}(j)}\left(\mathbf{p}^{(j)}, \zeta^{(j)}, \theta^{(j)}\right)$ as a benchmark. The alternate optimization begins by solving **P2a** to minimize $\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}}$ with respect to $\mathbf{p}$ given a pair of ζ and θ. It is achieved by replacing all entries of x_m with $v_m = \mathbf{AR}x_m\ \forall m$. $\mathbf{p}^{(j*)}$ is obtained using the framework provided in Algorithm 1 which solves **P1a** iteratively. Then, the optimum $\mathbf{p}^{(j*)}$ is used as a given PMF to obtain the pair $\zeta^{(j*)}$ and $\theta^{(j*)}$ by solving **P2b**. These optimum parameter values are updated to attain next initial points. Moreover, $\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}(j+1)}\left(\mathbf{p}^{(j+1)}, \zeta^{(j+1)}, \theta^{(j+1)}\right)$ is also evaluated to compare the decrease in objective function. The norm of this difference is stored in ϵ and the process is repeated until this value drops below a preset threshold δ. Eventually, the solution parameters are updated in $(\mathbf{p}^{*}, \zeta^{*}, \theta^{*})$ which yield the minimized BER upper bound $\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}*}$ using HS. Therefore, these HS parameters are capable of rendering a BER $\mathrm{P}_{\mathrm{b,HS}}^{*}$ lower than the bound $\mathrm{P}_{\mathrm{b,HS}}^{\mathrm{UB}*}$. Numerical evaluations reveal that the stopping criteria is mostly met in just one iteration. Interestingly, the sequential order of step 5



(a) GS for Proper Noise (b) PS for Proper Noise (c) HS for Proper Noise

(d) GS for Improper Noise (e) PS for Improper Noise (f) HS for Improper Noise

Fig. 2: Different asymmetric signaling designs for proper and improper aggregate interference

and 6 in Algorithm 2 need to be chosen carefully as PS-then-GS is not equivalent to GS-then-PS. Simulation results reveal that PS demonstrates better performance than GS at significant HWD levels whereas GS outperforms PS for negligible HWD levels (for some M_{nu}-QAM). Therefore, it is imperative to decide the order of shaping for HS as per the system HWD level. Notably, the first shaping method is dominant and it is given as the input for the second one for a further refinement. Therefore, the shaping order in the form PS-then-GS is the preferred choice for significant HWD levels and vice versa.

HS can be implemented by choosing the transmit symbols for the translated and rotated signal constellation, i.e., $v_m = \mathbf{A}\left(\zeta^*\right)\mathbf{R}\left(\theta^*\right)x_m$. Furthermore, the symbols are transmitted according to the optimized $\mathbf{p}^*$ where ζ^*, θ^* and $\mathbf{p}^*$ are designed using Algorithm 2. Upon reception, they are detected using the MAP detector as presented in (35).

F. Illustrative Example

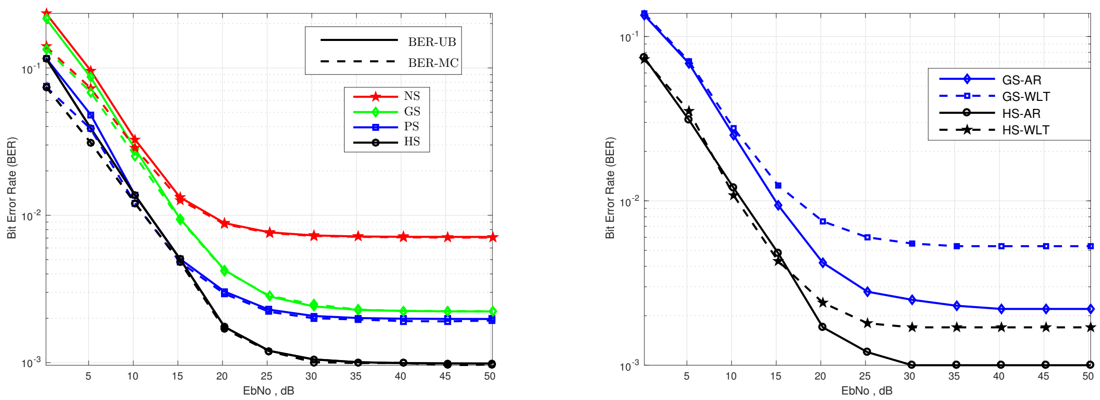
We present a comprehensive example to highlight the design of various distinct asymmetric constellations for a fixed rate of 4 bits/symbol. The red color visualizes constellation symbols with quite low probabilities. Fig. 2 represents different AS schemes assuming either proper

$C_\eta = 0$ or highly improper $C_\eta = 0.9$ HWDs, respectively. We use 16-QAM for GS whereas 32-QAM for both PS and HS. The shaping parameters are designed/optimized for a system suffering from high HWD, i.e., $\kappa = 0.99$ at 30dB SNR. Fig. 2(a) and 2(d) illustrate equally prior geometrically shaped constellation symbols in the presence of proper and maximally improper noise, respectively. Fig. 2(a) is a mere rotation of the original 16-QAM in the presence of proper HWDs whereas Fig. 2(d) also inculcates the translation rendering a squeezed parallelogram envelop in vertical axis. Next, probabilistic shaped constellations are presented in Fig. 2(b) and 2(e) for proper and maximally improper distortions, respectively. Evidently, the formation of red symbols around the origin transforms from a symmetric circle in Fig. 2(b) to an ellipse in 1st and 3rd quadrant in Fig. 2(e) corresponding to the respective symmetric and asymmetric noise. This reveals the reason behind superior performance of PS as it is capable of assigning negligible transmission probabilities to the symbols which are mostly affected by the aggregate noise as per it's proper/improper characteristics. Furthermore, this probabilistic shaped constellation undergoes GS to obtain hybrid shaped QAM constellation as shown in Fig. 2(c) and 2(f) under proper and maximally improper noise, respectively. This transformation allows the constellation to align itself as per the underlying noise characteristics and further improves the system performance.

VI. Numerical Results

Numerical evaluations of the adopted HWD system are carried out to study the drastic effects of hardware imperfections and the effectiveness of the mitigation strategies. The performance of the proposed asymmetric transmission schemes PS and HS is quantified as opposed to the benchmark no-shaping (NS) and conventional GS, with varying energy per bit per noise ratio (EbNo) and HWD levels. EbNo is obtained by normalizing SNR with the transmission rate. Moreover, GS can be implemented by transmitting symbols from a reshaped constellation $\mathbf{v}_m = \mathbf{A}\left(\zeta^*\right)\mathbf{R}\left(\theta^*\right)\mathbf{x}_m$, where ζ^* and θ^* can be obtained by solving **P2a** given uniform prior distribution. Upon reception, they are detected using the ML detector which is the simplified form of optimal MAP detector (35) given uniform prior probabilities. This ML detector considers the reshaped constellation symbols $\mathbf{v}_m$ as the reference to detect the received symbols.

For most of the numerical evaluations we assume Gray coded square QAM constellations of order $M_{\rm u} = 8$, i.e., $R = \log_2(M_{\rm u})$, for NS and GS as benchmarks. For PS and HS we employ $M_{\rm nu} = 32$-QAM with rate at least as high as that of GS, i.e., $R \geq \log_2(M_{\rm u})$. Moreover, we consider practical HWD values for the transmitter $\kappa_t = 0.01$ and receiver $\kappa_r = 0.12$. The pseudo-variances are derived from the $\tilde{\kappa}_{\rm tI} = \kappa_t/4$, $\tilde{\kappa}_{\rm rI} = \kappa_r/4$, and correlation coefficient



(a) BER performance of AS schemes

(b) Comparison of two GS Schemes

Fig. 3: BER Performance for a range of EbNo with $\kappa = 0.13$ in AWGN channel.

$\rho_\eta = 0.9$. Intuitively, AWGN channel assumes $g = 1$ and circularly symmetric Rayleigh fading channel is generated using $\lambda = 1$. Furthermore, the transmission EbNo is taken as 30 dB. The aforementioned values of the parameters are used throughout the numerical results, unless specified otherwise.

First, we evaluate the performance of various AS schemes for a range of EbNo from 0 dB to 50 dB in an AWGN channel as shown in Fig. 3(a). We employ M_{u}-QAM for NS and GS whereas M_{nu}-QAM for PS and HS. The BER upper bound (BER-UB) of PS and HS are given by (19) and (39), respectively, whereas the BER-UB of NS and GS are derived from (19) and (39) by assuming uniform distribution, respectively. The BER performance improves with increasing EbNo till 30 dB and then undergoes saturation owing to the presence of HWD. Further increase in bit energy also results in an increase in the distortion variance, as the system experiences an error floor which can be deduced from (23). Evidently, the proper/symmetric QAM is suboptimal and the BER performance is significantly improved using AS. Conventional GS is not beneficial at lower EbNo values, but it significantly improves the performance for higher EbNo values pertaining to the increased symbol space [25]. On the other hand, the proposed PS is capable of minimizing the BER for the entire range of EbNo. Substantial gains can be achieved by taking another step forward and employing HS. Therefore, we can safely conclude that the best performance can be achieved using PS for EbNo ≤ 15 dB and HS for EbNo ≥ 15 dB. At 20 dB, the BER reductions for GS, PS, and HS schemes with respect to unshaped constellation are approximately 52.22%, 66.67%, 80%, respectively. The numerical results in

Fig. 3(a) depict close accordance between the derived BER-UBs and the corresponding Monte Carlo (MC) performance of the various transmission schemes.

For the same system settings, we compare two different parameterization techniques to achieve asymmetric GS, which is a building block of HS, in Fig. 3(b). The GS-AR scheme represents our proposed GS scheme based on the optimal translation A and rotation R. This scheme induces a power imbalance between in-phase and quadrature components instead of their mutual correlation [25]. We compare this GS scheme with the well known WLT scheme referred as GS-WLT. We use the similar parameterization as adopted in [23] for our BER minimization problem and numerically solve the resultant non-convex optimization problem [3]. The comparison of GS schemes has been extended to hybrid shaping: where HS-AR and HS-WLT apply the proposed PS scheme to determine non-uniform probability distribution but respective GS techniques. Evidently, the candidate schemes perform equally good at low EbNo values but our proposed AR scheme outperforms WLT scheme in both GS and HS for relatively higher EbNo values.

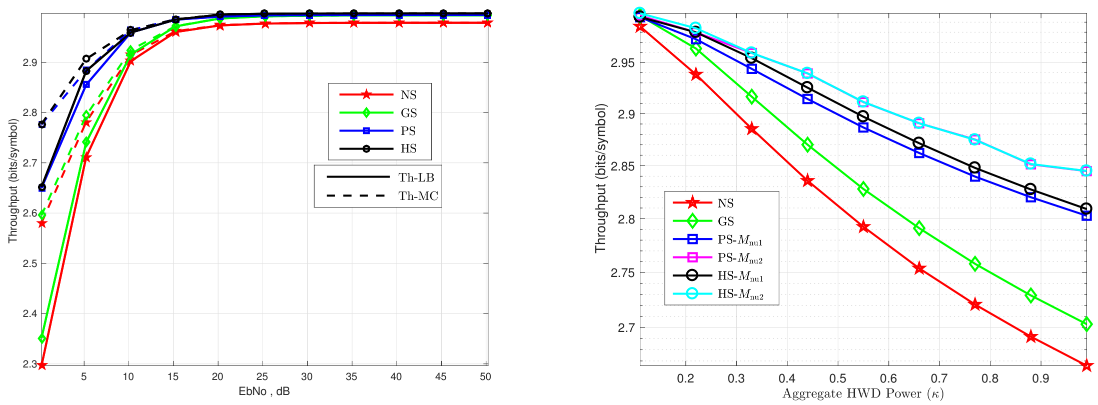
Given the simulation settings as in Fig. 3(a), we analyze system throughput (correctly received bits/symbol) for a range of EbNo values where the lower bound on system throughput can be obtained as

$$\mathcal{T}^{\mathrm{LB}}(\mathbf{p}) = \left[1 - \mathrm{P}_b^{\mathrm{UB}}(\mathbf{p})\right] \mathrm{H}(\mathbf{p}) \tag{43}$$

The throughput lower-bound (Th-LB) of all the transmission schemes can be calculated using their respective BER-UBs in (43). Fig. 4(a) validates the derived Th-LBs using MC simulations. It further depicts negligible throughput gain of GS over NS but noticeable throughput improvement using PS or HS. For instance, 1.5%, 6% and 7% percentage increase in throughput can be observed using GS, PS, and HS over NS at EbNo $= 5$ dB. At very low SNR, all the schemes depict unsatisfactory performance, as the required transmission rate can be higher than the maximum achievable rate which is related to the channel capacity. For moderate SNR, the throughput gain of the proposed schemes is quite substantial; nevertheless, it undergoes saturation when EbNo ≥ 20 dB. Interestingly, PS/HS saturates at 3 bits/symbol following rate fairness constraint with negligible BER whereas other schemes saturate below 3 bits/symbol depicting significant BER even though the entropy of 8-QAM with uniform distribution is $\log_2(8) = 3$.

Later, we analyze the behavior of various AS schemes with increasing distortion levels and their impact on the system throughput at EbNo $= 30$ dB. Fig. 4(b) compares the throughput performance of M_{u}-QAM NS and GS with $M_{\mathrm{nu1}} = 16$-QAM PS and HS as well as with

[3] We omit the derivation and implementation details of GS-WLT and HS-WLT due to the limited space.



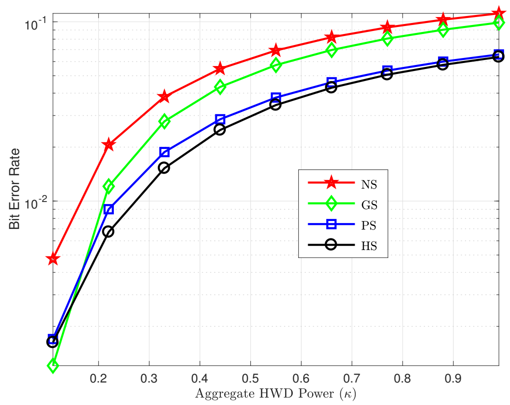
(a) Throughput versus EbNo

(b) Throughput versus HWD levels

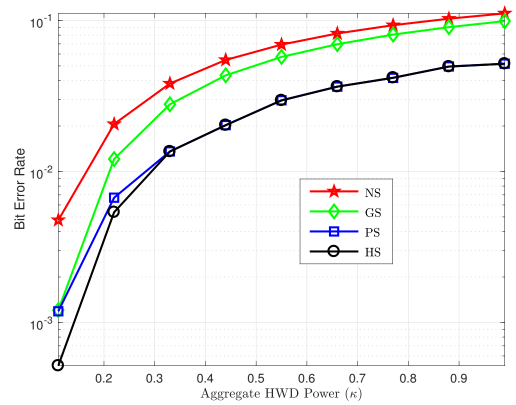
Fig. 4: Throughput Performance in AWGN channel.

$M_{\text{nu2}} = 32$-QAM PS and HS. System throughput decreases almost linearly with increasing HWD for all forms of signaling but with different slopes. NS demonstrates the steepest slope with increasing HWD and all the other AS schemes render gradual slopes. Quantitative analysis shows the slopes of -0.55, -0.41, -0.28, and -0.24 using NS, GS, 16-QAM PS/HS, and 32-QAM PS/HS, respectively, with increasing HWD. Therefore, PS and HS present the most favorable results as compared to the GS. Their performance can be even improved by increasing the modulation order. Another important observation is the overlapping response of PS and HS especially for higher ordered QAM, which suffices PS and revokes the need of HS to perform even better.

A similar analysis is undertaken to study the impact of increasing HWD on the system BER performance in an AWGN channel. We assume 8-QAM for NS and GS whereas 16-QAM for PS and HS as depicted in Fig. 5(a). Expectedly, the BER increases with increasing HWD levels and AS based systems achieve lower BER by efficiently mitigating the drastic HWD effects. Undoubtedly, the NS scheme suffers the most, but GS helps to decrease the BER to some extent. Further compensation can be achieved using the proposed PS and HS. Surprisingly, GS outperforms PS and HS at the lowest HWD values, e.g., $\kappa = 0.11$, in Fig. 5(a) but PS/HS maintain their superiority for $\kappa \geq 0.17$. Interestingly, PS/HS are still capable of outperforming GS even for the lowest HWD levels pertaining to their rate adaptation capability and added DoF using 32-QAM as highlighted in Fig. 5(b). We can observe enhanced mitigation offered by the 32-QAM PS/HS as compared to the 16-QAM PS/HS due to the added DoF. Evidently, there is a



(a) Using 16-QAM



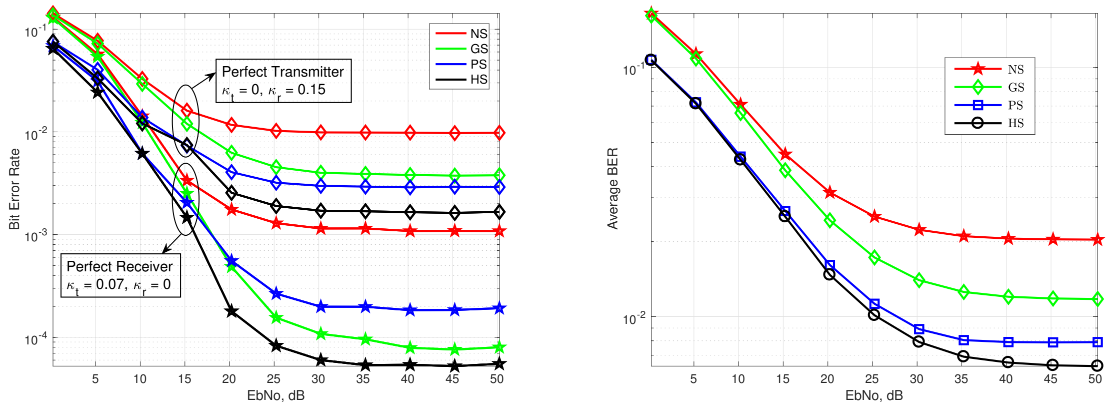
(b) Using 32-QAM

Fig. 5: HWD mitigation for PS and HS at EbNo $= 30$ dB in an AWGN channel.

trade-off between increased complexity and performance gain, which must be taken into account while choosing M_{nu} as per the system capability. For instance, we observe BER compensation of 66% and 77.5% using 32-QAM PS and HS, respectively, whereas BER compensation of 55% and 65% using 16-QAM PS and HS, respectively, at $\kappa = 0.22$ HWD level.

Another simulation example depicts the performance of the discussed AS schemes over a range of EbNo for two distinct scenarios of perfect receiver and perfect transmitter as presented in Fig. 6(a). Perfect receiver system as the name specifies includes ideal zero-distortion receiver but imperfect transmitter with $\kappa_t = 0.07$ whereas perfect transmitter system involves ideal zero-distortion transmitter but imperfect receiver with $\kappa_r = 0.15$. Note that the lower value of κ_t relative to κ_r is due to the fact that transmitters employ sensitive equipment to exhibit low distortions because the transmitter distortions are far more drastic than the receiver distortions. Interestingly, GS outperforms PS at EbNo > 15 dB for the perfect receiver case as opposed to EbNo < 15 dB where PS is still a better choice. HS outperforms both of them irrespective of the EbNo range classification. At such low HWD level, the BER percentage reduction of 81.82%, 90.91%, 94.55% is observed using PS, GS, and HS at 30 dB EbNo. Regarding the perfect transmitter scenario, GS and PS reverse the trend for higher EbNo level. Now the PS clearly outperforms GS for the entire range of EbNo and the HS marks its superiority over both of these schemes. At 0.15 HWD level, the EbNo gain of 8 dB, 12 dB, and 13 dB are estimated using GS, PS, and HS to attain the BER of 10^{-2}.

Finally, the average (ergodic) BER performance of the adopted system with $\kappa = 0.22$ HWD



(a) BER for Perfect Receiver and Perfect Transmitter

(b) Average BER in a Rayleigh fading channel.

Fig. 6: System BER performance with varying EbNo levels

level is evaluated over a Rayleigh fading channel for a range of EbNo values as given in Fig. 6(b). Evidently, the AS schemes preserve their BER trends and order. Clearly, average BER decreases with increasing EbNo and then undergoes saturation yielding an error floor. The derived BER bounds are also validated using MC simulations rendering a tighter bound for higher EbNo values. GS improves the average BER as compared to the NS scenario but PS and HS maintain their superior performance. Signaling schemes of GS, PS, and HS offer a percentage reduction of 54.55%, 63.64%, and 70.45%, respectively, in the average BER performance at 40 dB EbNo.

In a nutshell, we can conclude that the GS offers significant BER reduction at higher SNR values as opposed to the PS which offers universal gains. Moreover, the perks of HS are also prominent for higher SNR and higher M-ary modulation but depicts PS comparable performance at lower SNR values. Therefore, we recommend to employ HS given high SNR but resort to PS for lower SNR values to save additional computational expense. Additionally, GS is a better choice for slightly distorted systems whereas PS/HS are the optimal choice for moderate to severely distorted systems. Furthermore, we can achieve improved performance by employing higher-order QAM constellations for PS/HS given adequate resources. On the other hand, the throughput gains are eminent at considerably lower SNR values and higher distortion values.

VII. Conclusion

This work proposes probabilistic and hybrid shaping to realize asymmetric signaling in digital wireless communication systems suffering from improper HWD. Instinctively, all forms of asymmetric shaping are capable of decreasing the BER, and this performance gain improves with

TABLE II: First Order Necessary KKT Conditions

Index	KKT Conditions	Satisfied with	Reason
$1 : M_{\mathrm{nu}}$	$\nabla_{\mathbf{p}}\mathcal{L}\left(\mathbf{p}^{*},\lambda^{*}\right)=0,\ \forall 1\leq m\leq M_{\mathrm{nu}}$	$\nabla_{\mathbf{p}}\mathcal{L}\left(\mathbf{p}^{*},\lambda_1^{*},\lambda_2^{*},\lambda_3^{*}\right)=0$	Saddle point of the dual problem
M_{nu}+1	$\lambda_1^{*}\left(\sum\limits_{m=1}^{M_{\mathrm{nu}}}\|x_m\|^2 p_m^{*}-1\right)=0$	$\sum\limits_{m=1}^{M_{\mathrm{nu}}}\|x_m\|^2 p_m^{*}=1,\ \lambda_1^{*}\geq 0$	Maximum power transmission
M_{nu}+2	$\lambda_2^{*}\left(\sum\limits_{m=1}^{M_{\mathrm{nu}}} p_m^{*}-1\right)=0$	$\sum\limits_{m=1}^{M_{\mathrm{nu}}} p_m^{*}=1,\ \lambda_2^{*}\geq 0$	Equality constraint
M_{nu}+3	$\lambda_3^{*}\left(\log_2\left(M_{\mathrm{u}}\right)-\mathrm{H}(\mathbf{p}^{*})\right)=0$	$\mathrm{H}(\mathbf{p}^{*})=\log_2\left(M_{\mathrm{u}}\right),\ \lambda_3^{*}\geq 0$	BER -Rate tradeoff

increasing SNR and/or increasing HWD levels with respect to NS. However, PS outperforms GS and performs equally well as HS. We can achieve more than 50% BER reduction with PS/HS over traditional GS. The perks of PS come at the cost of increased complexity in the design and decoding process. The HS scheme is capable of improving the system performance in terms of the BER as well as throughput. However, for less HWD levels and low EbNo, the benefits of HS over PS are limited while requiring additional complications in optimization, modulation, and detection procedures. Therefore, PS emerges as the best choice in the trade-off between enhanced performance and added complexity.

APPENDIX A
KKT CONDITIONS

The convex non-linear constraint problem **P1a** can be efficiently solved using the first order necessary KKT conditions. We begin by writing the Lagrangian function $\mathcal{L}$ as

$$\mathcal{L}\left(\mathbf{p},\lambda_1,\lambda_2,\lambda_3\right)=\tilde{\mathrm{P}}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p},\mathbf{p}^{(k)}\right)+\lambda_1\left(\sum_{m=1}^{M}\left|x_m\right|^2 p_m-1\right)+\lambda_2\left(\sum_{m=1}^{M}p_m-1\right)+\lambda_3\left(\log_2\left(M_{\mathrm{u}}\right)-\mathrm{H}(\mathbf{p})\right), \tag{44}$$

where the Lagrange multipliers are $\lambda_1,\lambda_2,\ \lambda_3\geq 0$. Next. we evaluate the gradient of the (44) with respect to the optimization variables in $\mathbf{p}$

$$\nabla_{\mathbf{p}}\mathcal{L}=\left[\frac{\partial\mathcal{L}}{\partial p_1}\quad\frac{\partial\mathcal{L}}{\partial p_2}\quad\ldots\quad\frac{\partial\mathcal{L}}{\partial p_{M_{\mathrm{nu}}}}\right], \tag{45}$$

where the partial derivative of $\mathcal{L}$ with respect to p_m is given by

$$\frac{\partial\mathcal{L}}{\partial p_m}=\frac{\partial\mathrm{P}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p}^{(k)}\right)}{\partial p_m}+\lambda_1\left|x_m\right|^2+\lambda_2+\lambda_3\left(\frac{1}{\ln(2)}+\log_2\left(p_m\right)\right),\quad\forall\,1\leq m\leq M_{\mathrm{nu}} \tag{46}$$

Suppose that there is a local solution $\mathbf{p}^{*}$ of **P1a** and the objective function $\tilde{\mathrm{P}}_{\mathrm{b}}^{\mathrm{UB}}\left(\mathbf{p},\mathbf{p}^{(k)}\right)$ along with the constraints (24b) and (24c) are continuously differentiable. Then, there exists a Lagrange multiplier vector λ^*, with components λ_i, where $i \in (1, 2, 3)$, such that the necessary first order KKT conditions (as presented in Table II) are satisfied at $(\mathbf{p}^*, \lambda^*)$. Interestingly, the KKT conditions are satisfied with

$$\nabla_{\mathbf{p}} \mathcal{L} \left(\mathbf{p}^*, \lambda_1^*, \lambda_2^*, \lambda_3^*\right) = 0, \; \sum_{m=1}^{M} |x_m|^2 p_m^* = 1, \; \sum_{m=1}^{M} p_m^* = 1, \; \mathrm{H}(\mathbf{p}^*) = \log_2 (M_{\mathrm{u}}) \,. \tag{47}$$

owing to the maximum transmission power preference, equality constraint and BER -Rate trade-off , respectively. Interestingly, the complimentary slackness for both inequality constraints is satisfied with strictly positive Lagrange multipliers yielding a feasible optimal solution. These $M_{\mathrm{nu}} + 3$ solution parameters $\left(p_1^*, p_2^*, \ldots, p_{M_{\mathrm{nu}}}^*, \lambda_1^*, \lambda_2^*, \lambda_3^*\right)$ can be efficiently obtained by simultaneously solving the equations in (47) using Levenberg-Marquardt algorithm [64].

APPENDIX B
TRANSLATION WITHIN POWER BUDGET

In this appendix we present the proof of Remark 1. It is straightforward to prove that the translation $\mathbf{v} = \mathbf{A}\mathbf{w}$ does not change the variance/power but only introduce asymmetry/improperness. Considering the transformation caused by the translation $v = \sqrt{1+\zeta} w_I + i\sqrt{1-\zeta} w_Q$, the power/variance is given by

$$\sigma_v^2 = (1+\zeta)\sigma_{w_I}^2 + (1-\zeta)\sigma_{w_Q}^2. \tag{48}$$

Using the symmetric nature of r.v. w, i.e., $\sigma_{w_I}^2 = \sigma_{w_Q}^2$, it is clear that $\sigma_v^2 = \sigma_w^2$. On the other hand, the pseudo-variance can be calculated as

$$\tilde{\sigma}_v^2 = (1+\zeta)\sigma_{w_I}^2 - (1-\zeta)\sigma_{w_Q}^2 + i2\sqrt{1-\zeta^2} E\left\{w_I w_Q\right\}. \tag{49}$$

Again, the symmetry implies $E\left\{w_I w_Q\right\} = 0$. Thus, the circularity coefficient can be derived from (49), i.e., $|\tilde{\sigma}_v^2|/\sigma_v^2 = \zeta$.

The same concept can be extended to the symmetric discrete constellations with uniform prior probabilities. Considering the transformation caused by the translation $v_m = \sqrt{1+\zeta} x_{mI} + i\sqrt{1-\zeta} x_{mQ}$, the power of the transformed constellation is given by

$$P = \frac{1}{M}\left((1+\zeta)\sum_{m=1}^{M} x_{mI}^2 + (1-\zeta)\sum_{m=1}^{M} x_{mQ}^2\right). \tag{50}$$

Using the symmetric property of the original discrete constellation $\sum_{m=1}^{M} x_{mI}^2 = \sum_{m=1}^{M} x_{mQ}^2$, it is clear that the power is preserved as $P = \frac{2}{M}\sum_{m=1}^{M} x_{mI}^2$. Moreover, the non-zero pseudo-variance is given by

$$\tilde{P} = \zeta P + \frac{2i}{M}\sqrt{1-\zeta^2}\sum_{m=1}^{M} x_{mI}\, x_{mQ}. \tag{51}$$

Again, the symmetry implies $\sum_{m=1}^{M} x_{mI}x_{mQ} = 0$. Thus, the circularity coefficient can be derived from (51), i.e., $|\tilde{P}|/P = \zeta$.

Appendix C
Gradient for Optimization

The gradient of the upper bound on BER w.r.t GS parameters is given as

$$\nabla_{\mathcal{G}} \mathrm{P}_\mathrm{b}^{\mathrm{UB}} = \left[\frac{\partial \mathrm{P}_\mathrm{b}^{\mathrm{UB}}}{\partial \zeta} \quad \frac{\partial \mathrm{P}_\mathrm{b}^{\mathrm{UB}}}{\partial \theta} \right], \\ = \frac{1}{\log_2 (M)} \sum_{m=1}^{M} \sum_{\substack{n=1 \\ n \neq m}}^{M} \Delta_{mn} \left[\frac{\partial \gamma_{mn}}{\partial \zeta} \quad \frac{\partial \gamma_{mn}}{\partial \theta} \right], \tag{52}$$

where Δ_{mn} is the common part in both partial derivatives.

$$\Delta_{mn} = \frac{p_m \gamma_{mn}^{-3/2}}{2\sqrt{2\pi}} \sqrt{\frac{1-\rho_z^2}{\alpha}} e^{-\frac{\Omega_{mn}^2}{2}} \left(\ln \left(\frac{p_m}{p_n} \right) - \frac{1}{2\beta_{mn}^2} \right). \tag{53}$$

Moreover, the partial derivative of γ_{mn} with respect to the translation parameter ζ is given as

$$\frac{\partial \gamma_{mn}}{\partial \zeta} = \frac{\bar{\xi}_{mnI}^2}{\sigma_\mathrm{I}^2} + \frac{2\rho_z \bar{\xi}_{mnI} \bar{\xi}_{mnQ}}{\sigma_\mathrm{I} \sigma_\mathrm{Q}} \frac{\zeta}{\sqrt{1-\zeta^2}} - \frac{\bar{\xi}_{mnQ}^2}{\sigma_\mathrm{Q}^2}, \tag{54}$$

where, $\bar{\xi}_{mnI} = \xi_{mnI} \cos(\theta) - \xi_{mnQ} \sin(\theta)$ and $\bar{\xi}_{mnQ} = \xi_{mnI} \sin(\theta) + \xi_{mnQ} \cos(\theta)$. Furthermore, the partial derivative of γ_{mn} with respect to the rotation parameter is evaluated as

$$\frac{\partial \gamma_{mn}}{\partial \theta} = 2 \frac{1+\zeta}{\sigma_\mathrm{I}^2} \big(\xi_{mnI} \cos(\theta) - \xi_{mnQ} \sin(\theta) \big) \big(-\xi_{mnI} \sin(\theta) - \xi_{mnQ} \cos(\theta) \big) + \\ + 2 \frac{1-\zeta}{\sigma_\mathrm{Q}^2} \big(\xi_{mnI} \sin(\theta) + \xi_{mnQ} \cos(\theta) \big) \big(\xi_{mnI} \cos(\theta) - \xi_{mnQ} \sin(\theta) \big) + \\ - \frac{2\rho_z}{\sigma_\mathrm{I} \sigma_\mathrm{Q}} \sqrt{1-\zeta^2} \big({\xi_{mnI}}^2 \cos(2\theta) - {\xi_{mnQ}}^2 \cos(2\theta) - 2\xi_{mnI} \xi_{mnQ} \sin(2\theta) \big). \tag{55}$$

References

[1] L. Chettri and R. Bera, "A comprehensive survey on internet of things (IoT) toward 5G wireless systems," *IEEE Internet of Things J.*, vol. 7, no. 1, pp. 16–32, Jan. 2020.

[2] S. Javed, O. Amin, B. Shihada, and M.-S. Alouini, "A journey from improper Gaussian signaling to asymmetric signaling," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1539–1591, Apr. 2020.

[3] B. Razavi, "Design considerations for direct-conversion receivers," *IEEE Trans. Circuits Syst. II*, vol. 44, no. 6, pp. 428–435, June 1997.

[4] A. A. Abidi, "Direct-conversion radio transceivers for digital communications," *IEEE J. Solid-State Circuits*, vol. 30, no. 12, pp. 1399–1410, Dec. 1995.

[5] T. Schenk, *RF imperfections in high-rate wireless systems: Impact and digital compensation*. Springer Science & Business Media, 2008.

[6] X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu, “Hardware impairments aware transceiver for full-duplex massive MIMO relaying,” *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6565–6580, Dec. 2015.

[7] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, “Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits,” *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Sept. 2014.

[8] C. Studer, M. Wenk, and A. Burg, “MIMO transmission with residual transmit-RF impairments,” in *Proc. Int. ITG Workshop on Smart Antennas (WSA)*. Bremen, Germany: IEEE, Feb. 2010, pp. 189–196.

[9] S. Javed, O. Amin, S. S. Ikki, and M. S. Alouini, “On the achievable rate of hardware-impaired transceiver systems,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*. Singapore: IEEE, Dec. 2017, pp. 1–6.

[10] M. Soleymani, C. Lameiro, I. Santamaria, and P. J. Schreier, “Improper signaling for SISO two-user interference channels with additive asymmetric hardware distortion,” *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8624–8638, Dec. 2019.

[11] S. Javed, O. Amin, B. Shihada, and M.-S. Alouini, “Improper Gaussian signaling for hardware impaired multihop full-duplex relaying systems,” *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1858–1871, Mar. 2019.

[12] M. M. Alsmadi, N. A. Ali, and S. S. Ikki, “SSK in the presence of improper Gaussian noise: Optimal receiver design and error analysis,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*. Barcelona, Spain: IEEE, Apr. 2018, pp. 1–6.

[13] F. D. Neeser and J. L. Massey, “Proper complex random processes with applications to information theory,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, July 1993.

[14] E. Ollila, “On the circularity of a complex random variable,” *IEEE Signal Process. Lett.*, vol. 15, pp. 841–844, Nov. 2008.

[15] C. Lameiro, I. Santamaría, and P. J. Schreier, “Improper Gaussian signaling for multiple-access channels in underlay cognitive radio,” *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1817–1830, Mar. 2019.

[16] O. Amin, W. Abediseid, and M.-S. Alouini, “Overlay spectrum sharing using improper Gaussian signaling,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 1, pp. 50–62, Jan. 2017.

[17] J.-Y. Lin, R. Y. Chang, C.-H. Lee, H.-W. Tsao, and H.-J. Su, “Multi-agent distributed beamforming with improper Gaussian signaling for MIMO interference broadcast channels,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 136–151, Jan. 2019.

[18] W. Hedhly, O. Amin, and M.-S. Alouini, “Benefits of improper gaussian signaling in interweave cognitive radio with full and partial CSI,” *IEEE Trans. on Cogn. Commun. Netw.*, Feb. 2020.

[19] S. Javed, O. Amin, S. S. Ikki, and M.-S. Alouini, “Asymmetric hardware distortions in receive diversity systems: Outage performance analysis,” *IEEE Access*, vol. 5, pp. 4492–4504, Feb. 2017.

[20] C. Kim, E.-R. Jeong, Y. Sung, and Y. H. Lee, “Asymmetric complex signaling for full-duplex decode-and-forward relay channels,” in *Proc. Int. Conf. on ICT Convergence (ICTC)*. Jeju Island, South Korea: Citeseer, Oct. 2012, pp. 28–29.

[21] S. Javed, O. Amin, S. S. Ikki, and M.-S. Alouini, “Multiple antenna systems with hardware impairments: New performance limits,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1593–1606, Dec. 2018.

[22] S. Javed, O. Amin, and M.-S. Alouini, “Full-duplex relaying under I/Q imbalance using improper Gaussian signaling,” in *Proc. Int. Conf. Acoust. , Speech, and Signal Process. (ICASSP 2017)*. New Orleans: IEEE, Mar. 2017, pp. 6538–6542.

[23] I. Santamaria, P. M. Crespo, C. Lameiro, and P. J. Schreier, “Information-theoretic analysis of a family of improper discrete constellations,” *Entropy*, vol. 20, no. 1, p. 45, Jan. 2018.

[24] J. A. López-Fernández, R. G. Ayestarán, I. Santamaria, and C. Lameiro, “Design of asymptotically optimal improper constellations with hexagonal packing,” *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5445–5457, Aug. 2019.

[25] S. Javed, O. Amin, S. S. Ikki, and M.-S. Alouini, “Asymmetric modulation for hardware impaired systems - Error probability analysis and receiver design,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1723–1738, Feb. 2019.

[26] G. Foschini, R. Gitlin, and S. Weinstein, “Optimization of two-dimensional signal constellations in the presence of Gaussian noise,” *IEEE Trans. Commun.*, vol. 22, no. 1, pp. 28–38, Jan. 1974.

[27] N. S. Loghin, J. Zöllner, B. Mouhouche, D. Ansorregui, J. Kim, and S.-I. Park, “Non-uniform constellations for ATSC 3.0,” *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 197–203, Feb. 2016.

[28] J. Zöllner and N. Loghin, "Optimization of high-order non-uniform QAM constellations," in *Proc. IEEE Int. Symp. on Broadband Multimedia Syst. and Broadcast. (BMSB)*. London, UK: IEEE, June 2013, pp. 1–6.

[29] T. Thaiupathump, C. D. Murphy, and S. A. Kassam, "Asymmetric signaling constellations for phase estimation," in *Proc. 10th IEEE Workshop on Statist. Signal and Array Process.* Pocono Manor, PA, USA: IEEE, Aug. 2000, pp. 161–165.

[30] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[31] F. Willems and J. Wuijts, "A pragmatic approach to shaped coded modulation," in *IEEE 1st Symp. on Commun. and Veh. Technol. in the Benelux*. Citeseer, 1993.

[32] R. Laroia, N. Farvardin, and S. A. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1044–1056, 1994.

[33] P. Schulte, "Zero error fixed length distribution matching," Ph.D. dissertation, M.S. thesis, Inst. Commun. Eng., Technische Universität München, Munich, Germany, May 2015.

[34] A. Cirino, "Design and analysis of a Trellis-based syndrome distribution matching algorithm," Master's thesis, University of Bologna, 2018.

[35] M. Dia, V. Aref, and L. Schmalen, "A compressed sensing approach for distribution matching," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*. Vail, CO, USA: IEEE, June 2018, pp. 1266–1270.

[36] A. Elzanaty, A. Giorgetti, and M. Chiani, "Lossy compression of noisy sparse sources based on syndrome encoding," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7073–7087, Oct. 2019.

[37] A. Elzanaty, A. Giorgetti, and M. Chiani, "Limits on sparse data acquisition: RIC analysis of finite Gaussian matrices," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1578–1588, Mar. 2019.

[38] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

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

[40] M. P. Yankov, D. Zibar, K. J. Larsen, L. P. Christensen, and S. Forchhammer, "Constellation shaping for fiber-optic channels with QAM and high spectral efficiency," *IEEE Photon. Technol. Lett.*, vol. 26, no. 23, pp. 2407–2410, Dec. 2014.

[41] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[42] J. Cho, S. Chandrasekhar, R. Dar, and P. J. Winzer, "Low-complexity shaping for enhanced nonlinearity tolerance," in *Proc. 42nd European Conf on Optical Commun. (ECOC 2016)*. Dusseldorf, Germany: VDE, Sept. 2016.

[43] J.-X. Cai, H. G. Batshon, M. V. Mazurczyk, O. V. Sinkin, D. Wang, M. Paskov, W. Patterson, C. R. Davidson, P. Corbett, G. Wolter, T. Hammon, M. Bolshtyansky, D. Foursa, and A. Pilipetskii, "70.4 Tb/s capacity over 7,600 km in C+L band using coded modulation with hybrid constellation shaping and nonlinearity compensation," in *Proc. Opt. Fiber Comm. Conf. and Exhibit. (OFC)*. Los Angeles, CA, USA: Optical Society of America, Mar. 2017, pp. Th5B–2.

[44] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[45] B. Dunbridge, "Asymmetric signal design for the coherent Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 13, no. 3, pp. 422–431, July 1967.

[46] C. Pan and F. R. Kschischang, "Probabilistic 16-QAM shaping in WDM systems," *J. Lightw. Technol.*, vol. 34, no. 18, pp. 4285–4292, Sept. 2016.

[47] A. Elzanaty and M. Alouini, "Adaptive coded modulation for IM/DD free-space optical backhauling: A probabilistic shaping approach," *IEEE Trans. Commun.*, pp. 1–1, July 2020.

[48] Y. C. Gültekin, W. J. van Houtum, S. Şerbetli, and F. M. Willems, “Constellation shaping for IEEE 802.11,” in *Proc. IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. Montreal, QC, Canada: IEEE, Oct. 2017, pp. 1–7.

[49] O. İşcan and W. Xu, “Polar codes with integrated probabilistic shaping for 5G new radio,” in *IEEE 88th Vehicular Tech Conf (VTC-Fall)*. Chicago, IL, USA: IEEE, Aug. 2018, pp. 1–5.

[50] H. Iimori, R.-A. Stoica, and G. T. F. de Abreu, “Constellation shaping for rate maximization in AWGN channels with non-linear distortion,” in *Proc. IEEE 7th Int. Workshop on Comput. Advances in Multi-Sensor Adaptive Process. (CAMSAP)*. Curacao, Netherlands Antilles: IEEE, Dec. 2017, pp. 1–5.

[51] Y. Yao, K. Xiao, B. Xia, and Q. Gu, “Design and analysis of rotated-QAM based probabilistic shaping scheme for rayleigh fading channels,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3047–3063, May 2020.

[52] F. A. Aoudia and J. Hoydis, “Joint learning of probabilistic and geometric shaping for coded modulation systems,” *arXiv preprint arXiv:2004.05062*, 2020.

[53] E. Björnson, M. Matthaiou, and M. Debbah, “A new look at dual-hop relaying: Performance limits with hardware impairments,” *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4512–4525, Nov. 2013.

[54] M. Wenk, *MIMO-OFDM-testbed: challenges, implementations, and measurement results*. ETH Zurich, 2010.

[55] P. Zetterberg, “Experimental investigation of TDD reciprocity-based zero-forcing transmit precoding,” *EURASIP J. Adv. Signal Process.*, vol. 2011, p. 5, Dec. 2011.

[56] A.-A. A. Boulogeorgos, N. D. Chatzidiamantis, and G. K. Karagiannidis, “Energy detection spectrum sensing under RF imperfections,” *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2754–2766, July 2016.

[57] H. Suzuki, T. V. A. Tran, I. B. Collings, G. Daniels, and M. Hedley, “Transmitter noise effect on the performance of a MIMO-OFDM hardware implementation achieving improved coverage,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 867–876, Aug. 2008.

[58] T. T. Duy, T. Q. Duong, D. B. da Costa, V. N. Q. Bao, and M. Elkashlan, “Proactive relay selection with joint impact of hardware impairment and co-channel interference,” *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1594–1606, May 2015.

[59] S. Javed, O. Amin, S. S. Ikki, and M.-S. Alouini, “Impact of improper Gaussian signaling on hardware impaired systems,” in *Proc. 2017 IEEE Intern. Conf. Commun. (ICC)*. Paris, France: IEEE, May 2017, pp. 1–6.

[60] J. G. Proakis and M. Salehi, *Digital communications*. McGraw-Hill, 2008.

[61] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[62] J. J. Moré and D. C. Sorensen, “Computing a trust region step,” *J. Sci. Stat. Comput.*, vol. 4, no. 3, pp. 553–572, Sept. 1983.

[63] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, “Approximate solution of the trust region problem by minimization over two-dimensional subspaces,” *Math. Program.*, vol. 40, no. 1-3, pp. 247–263, Jan. 1988.

[64] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, June 1963.