UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Ricardo Bohaczuk Venturelli

Optimization of Integer-Forcing Precoding for Multi-User MIMO Downlink

Florianópolis
2021

RICARDO BOHACZUK VENTURELLI

OPTIMIZATION OF INTEGER-FORCING PRECODING FOR MULTI-USER MIMO DOWNLINK

Tese submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Engenharia Elétrica
Orientador: Prof. Danilo Silva, Ph.D.

FLORIANÓPOLIS
2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Bohaczuk Venturelli, Ricardo
Optimization of Integer-Forcing Precoding for Multi User MIMO Downlink / Ricardo Bohaczuk Venturelli ; orientador, Danilo Silva , 2021.
72 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2021.

Inclui referências.

1. Engenharia Elétrica. 2. Multi-user MIMO. 3. Downlink channel. 4. Integer-forcing precoding. 5. Adaptive modulation. I. , Danilo Silva. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

Ricardo Bohaczuk Venturelli

OPTIMIZATION OF INTEGER-FORCING PRECODING FOR MULTI-USER MIMO DOWNLINK

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Bartolomeu Ferreira Uchôa Filho, Ph.D.
Universidade Federal De Santa Catarina

Prof. Richard Demo Souza, Dr.
Universidade Federal de Santa Catarina

Prof. Gustavo Fraidenraich, Dr.
Universidade Estadual de Campinas

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Engenharia Elétrica.

ICP-Edu
Documento assinado digitalmente
Telles Brunelli Lazzarin
Data: 10/05/2021 15:54:47-0300
CPF: 023.400.719-25
Verifique as assinaturas em https://v.ufsc.br

Prof. Telles Brunelli Lazzarin, Dr. Eng.
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Santa Catarina

ICP-Edu
Documento assinado digitalmente
Danilo Silva
Data: 07/05/2021 10:37:15-0300
CPF: 031.557.924-26
Verifique as assinaturas em https://v.ufsc.br

Prof. Danilo Silva, Ph.D.
Orientador
Universidade Federal de Santa Catarina

Florianópolis, May 7, 2021.

Agradecimentos

Desejo expressar meu reconhecimento a todos que, de uma maneira ou outra, colaboraram na realização deste trabalho, em especial

aos meus pais Ivo José Venturelli e Marly Bohaczuk Venturelli, por todo o suporte que me deram e ainda me dão ao longo da vida;

à minha irmã Rafaela Bohaczuk Venturelli Knop e meu cunhado André Felipe Knop, por todo apoio ao longo dos anos;

ao professor Danilo Silva, pela dedicação ao trabalho, pela excelente orientação, por todo o incentivo para procurar soluções dos problemas encontrados e pelas críticas, correções e revisões dessa tese;

aos professores Bartolomeu Ferreira Uchôa Filho, Richard Demo Souza e Gustavo Fraidenraich, membros da banca examinadora, pela leitura, correções e sugestões para a tese que, com certeza, ajudaram a melhorá-la;

aos professores Carlos Aurélio Faria da Rocha, Leonardo Silva Resende, Raimes Moraes e demais professores do Laboratório de Comunicação e Sistemas Embarcados por todo apoio moral durante os seminários do laboratório;

aos amigos Bruno, Gustavo, Ana, Raffaela, Renata, Jean, Daniel, companheiros para o café da tarde, para os *happy-hours* e, claro, para discussão da pesquisa, sempre incentivando e propondo novas maneiras de solucionar algum problema;

aos amigos que conheci pela internet, em especial, todos do grupo dos Piratas do Quizz e aos amigos André, Cassiano, Thiago e Giordano;

ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro dado através do edital 153535/2016-4;

as demais pessoas que contribuíram para a tese mas não foram aqui citadas.

RESUMO

A tecnologia com múltiplas antenas (MIMO) vem sido amplamente considerada em canais sem fio, uma vez que a capacidade-soma cresce com o número de antenas. Esta tese foca no canal *downlink* com multiusuários (MU-MIMO), em que a estação rádio base deseja se comunicar com múltiplos usuários. Técnicas cujo desempenho se aproximam da capacidade-soma têm um custo computacional muito elevado, o que é inviável em cenários práticos. Por outro lado, métodos lineares, como forçagem a zero (ZF) e ZF regularizado (RZF), que são de baixa complexidade, têm um desempenho muito aquém da capacidade-soma. Como uma alternativa, técnicas de forçagem a inteiros (IF), que podem ser vistas como uma generalização dos métodos lineares tradicionais, foram propostas. O objetivo da pré-codificação IF é produzir um canal efetivo que é aproximadamente uma matriz inteira, ao invés da matriz identidade. Encontrar os parâmetros ótimos para a pré-codificação IF é uma tarefa difícil já que requer uma otimização inteira. Silva *et al.* propuseram dois métodos para a pré-codificação IF chamados DIF e RDIF. Eles também mostraram como encontrar os parâmetros ótimos de forma analítica para o caso especial $K = 2$ usuários. Nesta tese, é proposto um método sub-ótimo de baixa complexidade para encontrar parâmetros do esquema IF para qualquer número de usuários. O método proposto consiste em solucionar um problema de otimização relaxado e, em seguida, aplicar um algoritmo de redução de base de reticulado. É mostrado que o método proposto tem complexidade de $\mathcal{O}(K^3)$. Resultados de simulação mostram que o método proposto tem um desempenho superior aos métodos tradicionais de pré-codificação linear em todos cenários simulados. Uma segunda contribuição desta tese é combinar o esquema de pré-codificação proposto com modulação adaptativa, em que o transmissor seleciona a taxa e a energia para cada usuário baseado na probabilidade de erro de bit. Devido ao canal efetivo com coeficientes inteiros produzido pela abordagem IF, é necessário que o transmissor seja capaz de operar com modulações de diferentes cardinalidades. Os resultados de simulação mostram que, para valores medianos de SNR, o método proposto tem uma taxa-soma maior que os demais métodos comparados.

Palavras-chave: MIMO multiusuários, canal *downlink*, pré-codificação forçagem a inteiros, modulação adaptativa.

Resumo Expandido

Introdução

Técnicas de pré-codificação são usadas para mitigar a interferência de usuários em canais *downlink* MIMO (do inglês, *multiple-input-multiple-output*) com multiusuários [1]. Métodos de pré-codificação linear, como forçagem a zero (ZF, do inglês *zero-forcing*) e ZF regularizado (RZF, do inglês *regularized ZF*) são amplamente usados devido a sua baixa complexidade, entretanto, seu desempenho fica aquém da capacidade soma [2, 3]. Por outro lado, técnicas não-lineares, como *vector-pertubation* [4], podem alcançar taxa-somas mais altas em troca de um custo computacional potencialmente maior.

A pré-codificação LRA (do inglês, *lattice-reduction aided*) é uma técnica não-linear de baixa complexidade que, diferente dos métodos lineares, consegue atingir total diversidade disponível pelo canal [5, 6]. Na pré-codificação LRA, uma matriz de pré-codificação $\mathbf{T}$ é aplicada antes da transmissão para transformar a matriz do canal $\mathbf{H}$ em uma base mais apropriada (de acordo com alguma heurística), que é obtida através de redução de base em reticulados. Com essa abordagem, a matriz do canal efetivo, após um escalonamento apropriado pelos usuários, se torna uma matriz (unimodular) de inteiros $\mathbf{A}$. Para anular esse interferência com coeficientes inteiros, antes de aplicar a matriz $\mathbf{T}$, os símbolos de modulação são pré-multiplicados pela inversa de $\mathbf{A}$ e, em seguida, é aplicado um operador de modulo para limitar a energia transmitida. Como codificação de canal pode ser aplicada sobre o esquema de pré-codificação LRA, o desempenho desses esquemas são tipicamente medidos baseados na probabilidade de erro de símbolo em um cenário sem códigos.

Uma generalização da pré-codificação LRA é a chamada pré-codificação de forçagem a inteiros (IF, do inglês *integer-forcing*), cuja principal diferença é que a codificação de canal é aplicada imediatamente antes da multiplicação por $\mathbf{T}$ [7, 8, 9]. Essa abordagem tem a vantagem que prover maior confiabilidade com um custo computacional similar. Além disso, ela permite deduzir explicitamente expressões para taxas alcançáveis, ao invés de ser avaliada por simulações numéricas como em pré-codificação LRA, o que a torna mais receptiva para otimização.

Entretanto, pré-codificação IF ótima (assim como pré-codificação linear ótima) é, em geral, NP-hard [9]. Por essa razão, trabalhos nessa área focam em desenvolver algoritmos sub-ótimos de baixa complexidade. Uma das abordagens mais simples é escolher $\mathbf{T}$ de tal modo que $\mathbf{HT} = c\mathbf{A}$ [7] or $\mathbf{HT} \approx c\mathbf{A}$ [8], em que $c > 0$ é uma constante. Essas abordagens são semelhantes ao LRA para escolher $\mathbf{T}$, e requerem uma redução de reticulado para encontrar $\mathbf{A}$ [10]. Uma abordagem mais geral, porém mais complexa, é o algoritmo iterativo baseado em dualidade, que requer uma redução de reticulado em cada iteração. Em [9], Silva *et al.* mostraram que, em alta SNR, a escolha ótima para $\mathbf{T}$ satisfaz $\mathbf{HT} = c\mathbf{DA}$, em que $\mathbf{D}$ é uma matriz diagonal. Para uma SNR qualquer, o desempenho pode ser melhorado escolhendo $\mathbf{HT} \approx c\mathbf{DA}$. Para o caso especial $K = 2$ usuários, e em alta SNR, as escolhas ótimas para $\mathbf{D}$ e $\mathbf{A}$ foram encontradas de forma analítica. Contudo, para $K > 2$ usuários as escolhas ainda estão em aberto.

No cenário MU-MIMO, a modulação adaptativa permite uma melhor alocação de recursos (taxa e energia) uma vez que o transmissor, baseado na SNR e na realização do canal, seleciona os recursos para maximizar a taxa-soma atendendo algum critério (geralmente em termos de probabilidade de error de bit) [11]. Modulação adaptativa já foi proposta no cenário MU-MIMO usando pré-codificação ZF [12], na qual aproximações para as expressões de probabilidade de error de bit foram usadas uma vez que elas são mais simples de se manipular. Até onde sabemos, modulação adaptativa em conjunto com pré-codificação IF não foi proposto.

Objetivos

Nesta tese, nós propomos um método sub-ótimo de baixa complexidade para escolher $\mathbf{D}$ e $\mathbf{A}$ para qualquer número de usuários K. Em seguida, nós combinamos o método proposto com modulação adaptativa pra mostrar seu desempenho em cenários mais práticos.

Metodologia

O método proposto é baseado no esquema RDIF proposto em [9], o qual considera a abordagem ótima em alta SNR para solucionar uma heurística para valores gerais de SNR. Para solucionar o problema de otimização com essa abordagem, consideramos que a $\mathbf{A}$ é uma matriz unitriangular (triangular com 1 na diagonal principal) superior e, em seguida, consideramos uma problema de otimização relaxado, em que os coeficientes de bA podem ser qualquer número complexo (ao invés de inteiro Gaussiano). Com essas considerações encontramos as soluções ótimas para o problema relaxado $\tilde{\mathbf{D}}$ e $\tilde{\mathbf{A}}$. De maneira geral, se $\mathbf{D}$ é conhecido, a escolha de $\mathbf{A}$ que maximiza a taxa-soma corresponde em resolver o problema de vetores independentes de menor norma em reticulados, que pode ser resolvido (de forma sub-ótima) utilizando o algoritmo LLL, cuja complexidade é $\mathcal{O}(K^4 \log K)$. Entretanto, se $\mathbf{D} = \tilde{\mathbf{D}}$, devido ao fato de $\mathbf{A}$ ser unitriangular, o algoritmo LLL requer apenas $\mathcal{O}(K^3)$ operações [13], o que é a mesma complexidade dos métodos de pré-codificação linear.

Uma das dificuldades em se utilizar modulação adaptativa em conjunto com pré-codificação IF é que, devido ao embaralhamento feito ao pré-multiplicar pela inversa de $\mathbf{A}$, precisa-se que todas as possíveis constelações sejam um sub-conjunto de uma constelação-mãe. Para contornar esse problema, consideramos apenas constelações QAM, uma vez que essas constelações são mais simples de relacionar com reticulados. Além disso, expressões para probabilidade de erro de bit foram desenvolvidas para essas constelações.

Resultados e Discussão

Nas simulações, nós comparamos a taxa-soma média do método proposto com outros métodos em dois cenários. O primeiro é considerando teoria de informação, em que taxa-soma é também chamada de taxa alcançável. O segundo considera modulação adaptativa em um cenário sem códigos. Neste caso, a taxa-soma é também a soma das eficiências espectrais dos usuários. Em cada simulação, o número de usuários $K \leq N$ foi escolhido de forma a maximizar a taxa-soma, em que N é o número de antenas transmissoras.

Em termos de taxas alcançáveis, para $N = 16$ antenas transmissoras, nós mostramos que o método proposto tem melhor desempenho que métodos lineares para todos os valores de SNR simulados. Nós mostramos também que mesmo variando o valor de N (para uma SNR fixa) o desempenho do método proposto ainda é melhor que os demais. Entretanto, podemos perceber que a diferença para a capacidade-soma aumenta conforme N aumenta. Para o caso $N = K = 4$, nós comparamos o método proposto com o método de [9], que é baseado em busca exaustiva, e podemos constatar que, de fato, o método proposto é sub-ótimo.

Para modulação adaptativa, as simulações mostraram que o método proposto tem um desempenho melhor que os demais para valores medianos de SNR. Entretanto, para baixa e alta SNR o método LRA, cuja tendência é fazer com que todos os usuários utilizem a mesma constelação, tem um desempenho melhor que o método proposto. Note que, como o método proposto é baseado em uma heurística para alta SNR, espera-se que seu desempenho seja degradado em baixa SNR. Já a degradação em alta SNR acontece devido à limitação das constelações disponíveis, i.e., para alguns usuários, o transmissor não pode escolher uma constelação de maior cardinalidade já que ela não está entre as possíveis escolhas de constelação.

Considerações Finais

Esta tese propôs um projeto para escolhas de parâmetros para o método RDIF para $K \geq 2$ usuários e, em seguida, combinou o esquema proposto com modulação adaptativa.

O esquema proposto considera uma estrutura para a matriz $\mathbf{A}$ e então, soluciona um problema de otimização relaxado. Em seguida, a matriz $\mathbf{A}$ é encontrada utilizando um algoritmo de redução de base de reticulado que, nesse caso especial, tem complexidade de $\mathcal{O}(K^3)$. O esquema tem uma complexidade global de $\mathcal{O}(NK^2)$ que é a mesma de métodos lineares, e um desempenho superior.

Ao considerar constelações QAM, foi possível combinar o método proposto com modulação adaptativa e mostramos que, para valores medianos de SNR, esse esquema tem desempenho superior aos demais.

Palavras-chave: MIMO multiusuários, canal *downlink*, pré-codificação forçagem a inteiros, modulação adaptativa.

ABSTRACT

Multiple-input-multiple-output (MIMO) technology has been vastly considered in wireless channels since the sum capacity grows with the number of antennas. This thesis focuses on the multi-user MIMO (MU-MIMO) downlink scenario, where the base-station wants to communicate with multiple users. Techniques whose performance approach the sum capacity have a high computational cost, which is infeasible in practical scenarios. On the other hand, linear methods, such as zero-forcing (ZF) and regularized ZF (RZF), which are low-complexity, have a performance far below the sum capacity. As an alternative, integer-forcing (IF) techniques have been proposed, which can be seen as a generalization of traditional linear methods. The goal of IF precoding is to produce an effective channel that is approximately an integer matrix, rather than an identity matrix. Finding optimal parameters in IF precoding is a difficult task since it requires integer optimization. Silva *et al.* proposed two methods for IF precoding called DIF and RDIF. They also show how to analytically obtain optimal parameters in the special case $K = 2$ users. In this thesis, a low-complexity suboptimal method is proposed to optimize the parameters of an IF scheme for any number of K users. The proposed method involves solving a relaxation of the problem followed by the application of a lattice reduction algorithm and is shown to have an overall complexity of $\mathcal{O}(K^3)$. Simulation results show that the proposed method achieves a higher sum rate than a heuristic choice of parameters and significantly outperforms conventional linear precoding in all simulated scenarios. A second contribution of this thesis is combining the proposed precoding scheme with adaptive modulation, where the transmitter selects power and rate for each user based on the bit-error probability. Due to the integer effective channel in the IF approach, the transmitter must be able to operate with different modulation sizes. Simulation results show that for a medium range of SNR the proposed method allows a higher sum of spectral efficiency than other methods.

Keywords: Multi-user MIMO, downlink channel, integer-forcing precoding, adaptive modulation.

List of Symbols

Symbol	**Description**
$\mathbb{C}$	Set of complex numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{Z}$	Set of integers numbers
$\mathbb{Z}[j]$	Set of Gaussian integer numbers, i.e., a complex number $z = a + b\jmath$, where $a, b \in \mathbb{Z}$
$\mathbb{Z}_n$	Set of integers numbers modulo n
$\mathcal{A}^n$	Set of n-dimensional vector which entries is in $\mathcal{A}$
$\mathcal{A}^{n\times m}$	Set of matrices with n rows and m columns whose entries are in $\mathcal{A}$
$\mathbf{A}^{-1}$	The inverse of matrix $\mathbf{A}$
$\mathbf{A}^{\mathrm{T}}$	The transpose of matrix $\mathbf{A}$
$\mathbf{A}^{\mathrm{H}}$	The Hermitian (conjugate transpose) of matrix $\mathbf{A}$
$(\mathbf{A})_{ij}$	The (i, j)-th element of matrix $\mathbf{A}$
$\det \mathbf{A}$	The determinant of matrix $\mathbf{A}$
$\operatorname{rank} \mathbf{A}$	The rank of matrix $\mathbf{A}$
$\operatorname{Tr} \mathbf{A}$	The trace of matrix $\mathbf{A}$
$\lvert x\rvert$	The absolute value of number x
$\lvert\lvert\mathbf{x}\rvert\rvert$	The Euclidean norm of vector $\mathbf{x}$
$\lfloor x\rceil$	Rounding to nearest (Gaussian) integer operator
$\Re\{x\}, \Im\{x\}$	The real and imaginary component of x, respectively
$\mathcal{CN}(\mu, \mathbf{\Gamma})$	A circularly symmetric complex Gaussian (normal) distribution with mean μ and covariance $\mathbf{\Gamma}$
$\mathbb{E}[X]$	The expected value of X
$\Pr[A]$	The probability of event A

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Multiple-input-multiple-output (MIMO) technology has been widely considered in wireless communications systems since its performance can grow linearly with the number of antennas [1]. The benefits of MIMO communications include, but are not limited to, spatial diversity gain and spatial multiplexing gain [14].

MIMO technology was first investigated in a single-user scenario and then it was extended to the multi-user (MU) scenario motivated by the use in cellular networks [14, 2]. MU-MIMO allows users to share the same time and frequency resources in order to increase the overall performance (i.e. achievable rate) [15]. There are two types of MU-MIMO channels: the *uplink* channel where multiple users send their message to a base station (BS), and the *downlink* channel where a BS sends messages to multiple users [1].

In this work, we are interested in the downlink scenario, also called *broadcast* (BC) channel, where a BS with N antennas wants to communicate with K users, each one having a single antenna[1].

The channel to the ith user is assumed to be flat-fading, which can be represented by a complex-valued vector. We also assume that the transmitter has total knowledge of the channel gains, which is known as *channel state information* (CSI). There are few ways to obtain the CSI, for example, in time-division duplexing (TDD) systems, the BS can exploit channel reciprocity to infer the CSI, while in the frequency-division duplexing (FDD) systems, the BS can transmit a pilot signal and then each user sends its CSI through a feedback link [15].

In this thesis, we consider a total power constraint, i.e., the power used by all transmit antennas is limited. Some works may also consider a per-antenna constraint, where each transmit antenna has limited power [16].

The user performance is usually limited by the signal-to-interference-and-noise ratio (SINR), which itself depends on the transmitter power available and of the channel realization. Precoding techniques are often used in order to mitigate user interference in multi-user (MU) multiple-input-multiple-output (MIMO) downlink channels [1].

Throughout this work, we are interested in a scheme that maximizes the *sum rate* (also called *throughput*), which measures the number of bits that all users can receive. We investigate some precoding techniques with two different approaches. In the first approach, which is based on information theory, the precoding parameters are chosen to maximize the achievable sum

[1] We assume single-antenna receivers in order to reduce the complexity of the analysis. However, this assumption has some practical benefits, for example, it requires a simpler hardware on user devices and less channel knowledge at the transmitter [16].

rate. We say that a sum rate is achievable if there exist a coded scheme with arbitrarily large code length and an error probability as small as desired. The second approach, which is more practical, consider that the transmitter uses adaptive modulation in order to select modulation size for each user in a uncoded scenario such that the bit-error probability does not exceed a determined value.

1.1 Information Theoretical Approach

The sum-capacity is the supremum of all achievable sum rates. For a MU-MIMO BC channel, the sum-capacity is well known [17, 18, 19, 20]. A technique that can achieve the sum-capacity is the "dirty paper coding" (DPC) [21], where the BS takes the non-causal interference (known by the transmitter) into accounting before selecting the transmitted signal. However, due to its high complexity, DPC is not very used in practical scenarios.

Linear precoding is a low-complexity alternative, where the signals are pre-multiplied by a well-chosen matrix before being transmitted. This matrix is called *precoding* or *beamforming matrix* [16]. Among traditional linear precoding techniques, the two most well-known techniques are zero-forcing (ZF) and regularized ZF (RZF) [2, 22, 3]. The former chooses a precoding matrix such that the effective channel is free from interference. While the latter chooses a precoding matrix in order to maximize the signal-to-interference-and-noise ratio (SINR). Even though this method produces some interference to the users, the amount of interference can be controlled by the *regularization factor*.

Lattice-reduction-aided (LRA) precoding [5, 6] is a low-complexity non-linear technique that, differently from linear methods, can achieve full diversity supported by the channel. In LRA precoding, a linear precoding matrix $\mathbf{T}$ is applied before transmission, in order to transform the channel matrix $\mathbf{H}$ to a more suitable basis (according to some heuristic), which is obtained through lattice basis reduction [6]. With this approach, the effective channel matrix, after appropriate scaling by the users, becomes a (unimodular) integer-valued matrix $\mathbf{A}$. In order to cancel this integer interference, prior to the application of $\mathbf{T}$, the modulation symbols are pre-multiplied by the inverse of $\mathbf{A}$, followed by a modulo operator to limit the transmit power. Since channel coding can be applied on top of an LRA precoding scheme, the performance of the latter is typically measured based on uncoded symbol error probability [5, 6].

A generalization of LRA precoding is the so-called integer-forcing (IF) precoding [6, 7, 8, 9], whose main difference is that channel encoding is applied immediately before the multiplication by $\mathbf{T}$. This approach has the advantage of providing higher reliability at a similar computational cost. Moreover, it allows achievable rate expressions to be derived explicitly, rather than evaluated by numerical simulation as in LRA precoding, leading to a scheme much more amenable to optimization.

However, *optimal* IF precoding (as well as optimal linear precoding) is NP-hard in general [9] and for this reason prior work has focused on developing low-complexity suboptimal algorithms. The simplest approach is to choose $\mathbf{T}$ such that $\mathbf{HT} = c\mathbf{A}$ [7] or $\mathbf{HT} \approx c\mathbf{A}$ [6], where $c > 0$ is some constant. This turns out to be equivalent to the LRA approach to choosing $\mathbf{T}$, requiring lattice reduction to find $\mathbf{A}$ [6]. A more general but much more complex approach is the iterative duality-based algorithm in [8], which requires a lattice reduction step at every iteration.[2] In

[2] Another difficulty with the approach of [8] is that it requires a more complicated transmission scheme using multiple shaping lattices, so in effect it cannot be applied to the problem considered in this paper.

[9], Silva *et al.* showed that, for high SNR, the optimal choice of $\mathbf{T}$ satisfies $\mathbf{HT} = c\mathbf{DA}$, where $\mathbf{D}$ is a diagonal matrix, while, for general SNR, the performance can be improved by choosing $\mathbf{HT} \approx c\mathbf{DA}$. For the special case of $K = 2$ users, at high SNR, the optimal choice of $\mathbf{D}$ and $\mathbf{A}$ is found analytically in [9], however, the general case remains open.

1.2 Adaptive Modulation

In a MU-MIMO adaptive modulation (AM) scenario, the transmitter selects the power and rate for each user based on the SNR and the channel realization in order to maximize the sum rate such that some criteria (generally in terms of a target bit-error rate) is satisfied [11]. Differently of non-adaptive scenario, where those parameters are chosen based on the worst-case or on average channel condition, AM allows a more efficiency of resources allocation.

AM techniques were first studied in single-user with single antennas system [23, 24, 25] and then extended to single-user MIMO scenario [11]. In [12] a scheme AM in MU-MIMO scenario was proposed using ZF precoding combining with user scheduling. In most of those works, based on estimated instantaneous SINR at the receiver, the transmitter choose a modulation size, for each user, such that the bit-error probability is not greater than a determined value. Generally, an approximation or a bound of the bit-error probability expression is used, since it is simpler to manipulate.

In this thesis, we combine adaptive modulation with LRA and IF precoding schemes. Note that, due to the pre-multiplication by the inverse of $\mathbf{A}$, it is required that each constellation is a subset of a "mother" constellation. One way to ensure this requirement is using only QAM constellation. Moreover, QAM constellations fits well with using of lattice. One major difficult is that, to the best of our knowledge, there is no closed-form neither approximations of the bit-error probability for lattice modulations.

1.3 Contributions

In this thesis, we first propose a low-complexity sub-optimal method for choosing $\mathbf{D}$ and $\mathbf{A}$ for any number of K users. We show how to find the optimal choice of $\mathbf{D}$ for a certain relaxation of the problem, after which $\mathbf{A}$ can be found with a single lattice reduction step. Remarkably, due to the special structure that we stipulate for $\mathbf{A}$, the latter problem can be solved much more efficiently than the general case, leading to an algorithm with overall complexity $O(K^3)$, the same as linear precoding methods and lower than previous IF precoding methods [7, 6, 8]. Simulation results show that the proposed method achieves a higher sum rate than the heuristic choice $\mathbf{D} = \mathbf{I}$ and significantly outperforms conventional linear precoding in all simulated scenarios.

After that, we show how adaptive modulation can be applied for LRA/IF precoding in an uncoded scenario and with QAM constellations. We also find expressions for the bit-error probability for lattice-modulations. Simulation results show that for medium values of SNR our schemes achieves higher spectral efficiency than LRA or linear precoding.

1.4 Publications

- R. B. Venturelli and D. Silva, “Optimization of Integer-Forcing Precoding for Multi-User MIMO Downlink,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1860–1864, 2020

- R. B. Venturelli and D. Silva, “Um Algoritmo para Escolha de Parâmetros da Pré-codificação de Forçagem a Inteiros para Canais Downlink MU–MIMO,” in *Anais de XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais. Sociedade Brasileira de Telecomunicações*, Sep. 2018.

1.5 Organization

This thesis is organized as follows.

In Chapter 2, we define all parameters of the channel model. Moreover, we review linear precoding methods such as ZF and RZF, as well as non-linear precoding methods such as LRA and IF. In Chapter 3 we mathematically define our optimization problem and then, after considering specific structure for matrix $\mathbf{A}$, we solve a relaxed version of this optimization problem. In Chapter 4, we combine the proposed method with adaptive modulation, where we make some considerations about the constellation with lattice and the bit-error probability. Finally, in Chapter 5, we have our conclusions and suggestion for future works.

Chapter 2

Preliminaries

2.1 System Model

Consider a downlink MIMO channel with an N-antenna transmitter and $K \leq N$ single-antenna users. Let $\mathbf{w}_i \in \mathcal{W}_i$ be the message destined to the ith user, where $\mathcal{W}_i$ is the message space of the ith user with cardinality $|\mathcal{W}_i|$, $i = 1, \ldots, K$, and let $\mathbf{x}'_j \in \mathbb{C}^n$ be the vector sent by the jth transmitter's antenna, $j = 1, \ldots, N$. For $i = 1, \ldots, K$, the signal received by the ith user is given by

$$\mathbf{y}_i = \mathbf{h}_i \mathbf{X}' + \mathbf{z}_i \tag{2.1}$$

where $\mathbf{X}' = \begin{bmatrix} \mathbf{x}_1'^{\mathrm{T}} & \cdots & \mathbf{x}_N'^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N \times n}$, $\mathbf{h}_i \in \mathbb{C}^N$ is the channel coefficients to the ith user, and $\mathbf{z}_i \in \mathbb{C}^n$ is the noise vector, such that $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. We can express (2.1) in matrix form as

$$\mathbf{Y} = \mathbf{H}\mathbf{X}' + \mathbf{Z} \tag{2.2}$$

where $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^{\mathrm{T}} & \cdots & \mathbf{y}_K^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{K \times n}$, $\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^{\mathrm{T}} & \cdots & \mathbf{h}_K^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{K \times N}$ and $\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1^{\mathrm{T}} & \cdots & \mathbf{z}_K^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{K \times n}$.

The transmitted signals must satisfy an average total power constraint, namely

$$\frac{1}{n}\mathbb{E}[\mathrm{Tr}(\mathbf{X}'\mathbf{X}'^{\mathrm{H}})] \leq \mathrm{SNR} \tag{2.3}$$

where $\mathrm{SNR} > 0$.

The ith user will try to infer a message $\hat{\mathbf{w}}_i \in \mathcal{W}_i$ from $\mathbf{y}_i$. An error occurs if $\hat{\mathbf{w}}_i \neq \mathbf{w}_i$ for any i. The sum rate is given by $R_{\mathrm{sum}} = R_1 + \cdots + R_K$, where $R_i = \frac{1}{n}\log_2|\mathcal{W}_i|$. A sum rate R is said to be achievable if, for any $\epsilon > 0$ and a sufficiently large n, there exists a coding scheme with sum rate at least R and error probability less than ϵ.

2.2 The Sum Capacity

The *sum capacity* of the channel is the supremum of all achievable sum rates and is given by [19, 18, 20]

$$C_{\mathrm{sum}} = \sup_{\mathbf{Q}:\mathrm{Tr}(\mathbf{Q})\leq 1} \log_2 \det\left(\mathbf{I} + \mathrm{SNR}\mathbf{H}^{\mathrm{H}}\mathbf{Q}\mathbf{H}\right) \tag{2.4}$$

where $\mathbf{Q} \in \mathbb{R}^{K \times K}$ is a diagonal matrix with nonnegative entries.

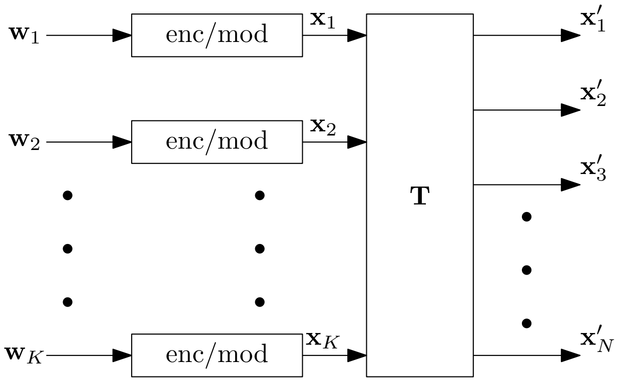


Figure 2.1: A transmitter using linear precoding.

It is known that it is possible to achieve the sum capacity by using *dirty-paper coding* (DPC) [1]. The main idea in DPC comes from the single antenna point-to-point channel with interference: If the transmitter knows the non-causal interference, the capacity of a channel with interference is the same as a non-interference AWGN channel [21]. This idea can be extended for multiuser MIMO downlink channel since the ith user considers the signals sent to other users as interference [1]. For example, after choosing the codeword to the first user, the transmitter can take it into account to select codeword to the second user in such a way that it is no longer treated as interference. This process goes on until the codeword to Kth user, which takes into account all previously codewords.

However, DPC is not used in practical scenarios due to its high complexity implementation [1]. Despite performance far from the sum capacity, an alternative to DPC is a family of low-complexity precoding methods, such as that of linear precoding.

2.3 Linear Precoding

Linear precoding is a low-complexity alternative to the DPC. Here, the messages $\mathbf{w}_1, \dots, \mathbf{w}_K$ are encoded and then modulated into vector $\mathbf{x}_1, \dots, \mathbf{x}_K$, respectively, where $\mathbf{x}_i \in \mathbb{C}^n$ and $\mathbb{E}[\|\mathbf{x}_i\|^2] \leq n\text{SNR}$. Let $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1^{\mathrm{T}} & \cdots & \mathbf{x}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. The signals sent by the transmitter are

$$\mathbf{X}' = \mathbf{T}\mathbf{X} \tag{2.5}$$

where $\mathbf{T} \in \mathbb{C}^{N\times K}$ is called the *precoding matrix* or *beamforming matrix.* A linear precoding transmitter is shown in Fig. 2.1.

Note that, due to (2.3) we have

$$\begin{aligned}\text{SNR} &\geq \frac{1}{n}\mathbb{E}\left[\operatorname{Tr}\left(\mathbf{T}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\right)\right] && (2.6)\\ &\geq \frac{1}{n}\operatorname{Tr}\left(\mathbb{E}\left[\mathbf{X}\mathbf{X}^{\mathrm{H}}\right]\mathbf{T}^{\mathrm{H}}\mathbf{T}\right) && (2.7)\\ &\geq \text{SNR}\operatorname{Tr}(\mathbf{T}^{\mathrm{H}}\mathbf{T}) && (2.8)\end{aligned}$$

and therefore

$$\operatorname{Tr}\left(\mathbf{T}^{\mathrm{H}}\mathbf{T}\right) \leq 1 \tag{2.9}$$

is a constraint to the precoding matrix.

In this scenario, we can rewrite (2.2) as

$$\begin{aligned}\mathbf{Y} &= \mathbf{HTX} + \mathbf{Z} && (2.10)\\ &= \mathbf{H}'\mathbf{X} + \mathbf{Z} && (2.11)\end{aligned}$$

where $\mathbf{H}' \triangleq \mathbf{HT}$. Note that, the signal received by ith user is

$$\begin{aligned}\mathbf{y}_i &= \mathbf{h}_i\mathbf{X}' + \mathbf{z}_i && (2.12)\\ &= \mathbf{h}_i\mathbf{TX} + \mathbf{z}_i && (2.13)\\ &= \mathbf{h}_i'\mathbf{X} + \mathbf{z}_i && (2.14)\\ &= \underbrace{h_{ii}'\mathbf{x}_i}_{\text{desired signal}} + \underbrace{\sum_{j\neq i} h_{ij}'\mathbf{x}_j}_{\text{interference}} + \underbrace{\mathbf{z}_i}_{\text{noise}} && (2.15)\end{aligned}$$

where $\mathbf{h}_i' = \begin{bmatrix} h_{i1}' & \cdots & h_{iK}' \end{bmatrix}$ is the ith row of $\mathbf{H}'$.

Thus, the signal-to-interference-and-noise ratio (SINR) of the ith user is given by

$$\mathrm{SINR}_i = \frac{|h_{ii}'|^2\,\mathrm{SNR}}{\sum_{j\neq i}\left|h_{ij}'\right|^2\mathrm{SNR} + 1}. \qquad (2.16)$$

and the achievable sum-rate of linear precoding schemes is

$$R_{\text{sum}}(\mathbf{H},\mathbf{T},\mathrm{SNR}) = \sum_{i=1}^{K}\log(1+\mathrm{SINR}_i). \qquad (2.17)$$

The two best known linear precoding methods are the zero-forcing (ZF) precoding and the regularized zero-forcing (RZF) precoding. The first one tries to pre-invert the channel matrix in order to produce zero interference, while the second one tries to minimize both the interference and the noise simultaneously.

2.3.1 Zero Forcing

In the zero-forcing precoding, the matrix $\mathbf{T}$ is designed to achieve zero interference, i.e., we want that $h_{ii}' \neq 0$ and $h_{ij}' = 0$ for $j \neq i$. Note that, without loss of generality, we can always assume that h_{ii}' is real and positive. This means that $\mathbf{HT} = \mathrm{diag}(\sqrt{\mathbf{p}})$ where $\mathbf{p} = \begin{bmatrix} p_1 & \cdots & p_K \end{bmatrix} \in \mathbb{R}^K$ is a real vector with non-negative entries. We can mathematically express the choice of $\mathbf{T}$ by an optimization problem

$$\mathbf{T} = \underset{\substack{\mathbf{T}':\mathbf{HT}'=\mathrm{diag}(\sqrt{\mathbf{p}})\\ \mathrm{Tr}(\mathbf{T}'\mathbf{T}'^{\mathrm{H}})\leq 1}}{\arg\max}\; R_{\text{sum}}(\mathbf{H},\mathbf{T}',\mathrm{SNR}) \qquad (2.18)$$

whose optimal solution is given by [2, 22]

$$\mathbf{T} = \mathbf{H}^{\mathrm{H}}\left(\mathbf{HH}^{\mathrm{H}}\right)^{-1}\mathrm{diag}(\sqrt{\mathbf{p}}). \qquad (2.19)$$

In this case, it is clear that $\mathrm{SINR}_i = p_i\mathrm{SNR}$ and therefore $R_{\text{sum}} = \sum_i \log(1 + p_i\mathrm{SNR})$.

It is possible to choose $p_1 = p_2 = \ldots = p_K$, however an optimized power allocation can improve the performance of ZF precoding. More precisely, the following optimization problem

$$\text{maximize} \quad \sum_{i=1}^{K} \log(1 + p_i \mathrm{SNR}) \tag{2.20}$$

$$\text{subject to} \quad \sum_{i=1}^{K} p_i [(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}]_{ii} = 1 \tag{2.21}$$

where the constraint comes from (2.9), can be solved using the well-known water-filling algorithm [22, 26, 27].

One of the main problems of ZF appears when $\mathbf{H}\mathbf{H}^{\mathrm{H}}$ has some small eigenvalue since it limits the achievable sum rate of ZF schemes [3]. To demonstrate this limitation, let $\mathbf{U}\Lambda\mathbf{U}^{-1}$ be the eigendecomposition of $\mathbf{H}\mathbf{H}^{\mathrm{H}}$, i.e., $\mathbf{H}\mathbf{H}^{\mathrm{H}} = \mathbf{U}\Lambda\mathbf{U}^{-1}$, where $\mathbf{U} \in \mathbb{C}^{K \times K}$ is the matrix whose ith column is the ith eigenvector and $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_K) \in \mathbb{C}^{K \times K}$ is a diagonal matrix whose elements are the eigenvalues. Since $\mathbf{H}\mathbf{H}^{\mathrm{H}}$ is Hermitian matrix and nonsingular, $\mathbf{U}$ is also a unitary matrix, i.e., $\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{H}}$, and all eigenvalues are real and (strictly) positive. Consider now the case where we allocate the same power to all users, i.e., $p_1 = p_2 = \cdots = p_K$. Note that (2.19) can be rewritten as

$$\mathbf{T} = \sqrt{p}\mathbf{H}^{\mathrm{H}} \left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1} \tag{2.22}$$

where $p \in \mathbb{R}$ is chosen to satisfy the constraint (2.9) and therefore

$$p = \frac{1}{\mathrm{Tr}\left(\left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1}\right)} = \frac{1}{\sum_i \lambda_i^{-1}}. \tag{2.23}$$

Note that since the right-hand side of (2.23) is a scaled version of the harmonic mean, we have that $p \leq \min(\lambda_1, \lambda_2, \ldots, \lambda_K)$. The sum rate, in this case, is given by $K \log(1 + p\mathrm{SNR})$, thus, it becomes clear that a small eigenvalue limits the performance of ZF schemes.

2.3.2 Regularized Zero Forcing

One way to overcome the limitation of small eigenvalues is to regularize the inverse in (2.19), i.e., adding a multiple of the identity matrix before calculating the inverse [3], which means that

$$\mathbf{T} = \mathbf{H}^{\mathrm{H}} \left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \alpha\mathbf{I}\right)^{-1} \mathrm{diag}(\sqrt{\mathbf{p}}) \tag{2.24}$$

where $\alpha \geq 0 \in \mathbb{R}$ is called the *regularization factor* and determines the amount of interference that we tolerated and $\mathbf{p} = \begin{bmatrix} p_1 & \cdots & p_K \end{bmatrix} \in \mathbb{R}^K$ is the vector that contains the power allocated for each user. The optimal value for α can be found by maximizing the SINR and it is given by $\alpha = K/\mathrm{SNR}$ [3, 2]. Note that, as $\mathrm{SNR} \to \infty$ (and therefore $\alpha \to 0$), (2.24) reduces to ZF precoding [3].

Since (2.24) corresponds to a regularized version of (2.19), this method is called *regularized zero-forcing* (RZF). The exact same matrix $\mathbf{T}$ can be found if we minimize the mean-square-error between the transmitted signal $\mathbf{X}' = \mathbf{T}\mathbf{X}$ and the received signal $\mathbf{Y}$, and therefore, this method is also called *minimum-mean-square-error* (MMSE) [28]. Other names can also be used to refer to this method, such as *signal-to-leakage-and-noise ratio* (SLNR) or *Wiener filter* [2, 16].

As well as in ZF scheme, we can choose $\mathbf{p}$ such that each element is equal to each other or we can use a power allocation algorithm. Note that since **HT** is no longer a diagonal matrix, we cannot apply the water-filling algorithm directly. However, it is possible to use a heuristic approach [2], where we use the objective function (2.20) as in ZF while using a more appropriate constraint [2]. Specifically, we substitute the constraint (2.21) by

$$\sum_{i=1}^{K} p_i \left[\left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \frac{K}{\mathrm{SNR}}\mathbf{I} \right)^{-1} \mathbf{H}\mathbf{H}^{\mathrm{H}} \left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \frac{K}{\mathrm{SNR}}\mathbf{I} \right)^{-1} \right]_{ii} = 1 \tag{2.25}$$

and then we can use the water-filling algorithm.

2.4 Lattices

A lattice $\Lambda \subseteq \mathbb{R}^n$ is a discrete subgroup of $\mathbb{R}^n$ [29]. A lattice is closed under reflection, i.e., if $\lambda \in \Lambda$ then $-\lambda \in \Lambda$, and under addition, i.e., if $\lambda_1, \lambda_2 \in \Lambda$ then $\lambda_1 + \lambda_2 \in \Lambda$. Thus, any integer linear combination of lattices points must be in the lattice.

It is possible to define a lattice Λ by its *generator matrix* (also called *basis matrix*) $\mathbf{B} \in \mathbb{R}^{n \times n}$ as

$$\Lambda \triangleq \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{B}\mathbf{u}, \mathbf{u} \in \mathbb{Z}^n\}. \tag{2.26}$$

where the *columns* of $\mathbf{B}$ are independent vectors over $\mathbb{R}^n$. For example, if $\mathbf{B}$ is the identity matrix, we get the integer lattice $\Lambda = \mathbb{Z}^n$. Note that the generator matrix in not unique for a given lattice. More specifically, $\mathbf{B}$ and $\mathbf{B}' = \mathbf{B}\mathbf{A}$ generate the same lattice if and only if $\mathbf{A}$ is unimodular, i.e., $\mathbf{A}$ is an integer matrix with $|\det \mathbf{A}| = 1$.

This concept of lattice can be extended to the complex field. A (complex) lattice Λ is a discrete $\mathbb{Z}[\jmath]$-submodule of $\mathbb{C}^n$, i.e., any Gaussian integer[1] linear combination of lattice points lies in the lattice. The definition in (2.26) still holds, however with $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{B} \in \mathbb{C}^{n \times n}$ and $\mathbf{u} \in \mathbb{Z}[\jmath]^n$. In the remainder of this work, we consider that all lattices are in the complex field.

A *fundamental cell* $\mathcal{P}_\Lambda$ of a lattice Λ is a bounded set, which, when shifted by lattice points generates a partition of $\mathbb{C}^n$. That is:

(*i*) $\mathcal{P}_\Lambda(\lambda) = \lambda + \mathcal{P}_\Lambda$, for all $\lambda \in \Lambda$,

(*ii*) $\mathcal{P}_\Lambda(\lambda) \cap \mathcal{P}_\Lambda(\lambda') = \emptyset$, for all $\lambda \neq \lambda' \in \Lambda$,

(*iii*) $\bigcup_{\lambda \in \Lambda} \mathcal{P}_\Lambda(\lambda) = \mathbb{C}^n$.

Note that, every element $\mathbf{x} \in \mathbb{C}^n$ can be *uniquely* expressed as the sum of a lattice point and a point in the fundamental cell, i.e., $\mathbf{x} = \lambda + \mathbf{x}_e$, where $\lambda \in \Lambda$ and $\mathbf{x}_e \in \mathcal{P}_\Lambda$. Since this expansion is unique, we can define a quantization function $\mathcal{Q}_\Lambda : \mathbb{C}^n \to \Lambda$, such that $\mathcal{Q}_\Lambda(\mathbf{x}) = \lambda$ and a lattice-modulus operator, where $\mathbf{x} \bmod \Lambda \triangleq \mathbf{x} - \mathcal{Q}_\Lambda(\mathbf{x}) \in \mathcal{P}_\Lambda$. A lattice-modulus operator has the following properties:

(*i*) $(\mathbf{x} + \lambda) \bmod \Lambda = \mathbf{x} \bmod \Lambda$, for all $\lambda \in \Lambda$

(*ii*) $(\mathbf{x} \bmod \Lambda + \mathbf{y}) \bmod \Lambda = (\mathbf{x} + \mathbf{y}) \bmod \Lambda$, for $\mathbf{y} \in \mathbb{C}^n$.

Clearly, both the quantization function and the lattice-modulus operator depend on the fundamental cell used.

[1] A Gaussian integer is any number of the form $a + \jmath b$, where $a, b \in \mathbb{Z}$.

One of the most important fundamental cell is the *Voronoi region*, denoted by $\mathcal{V}_\Lambda$. This cell uses, as quantization function, the nearest neighbor quantizer, i.e.,

$$\mathcal{Q}_\Lambda(\mathbf{x}) = \arg\min_{\lambda \in \Lambda} \|\mathbf{x} - \lambda\| \tag{2.27}$$

and therefore, we can define the Voronoi region as

$$\mathcal{V}_\Lambda = \{\mathbf{x} \in \mathbb{C}^n : \mathcal{Q}_\Lambda(\mathbf{x}) = \mathbf{0}\}. \tag{2.28}$$

The second moment (per dimension) of a lattice is defined as

$$P_\Lambda \triangleq \frac{1}{n}\mathbb{E}[\|\mathbf{x}\|^2] \tag{2.29}$$

where $\mathbf{x} \in \mathbb{C}^n$ is a random vector uniformly distributed over $\mathcal{V}_\Lambda$.

A pair of lattices Λ and Λ_s are called *nested* if $\Lambda_s \subseteq \Lambda$. For each $\lambda \in \Lambda$, we say that $\lambda + \Lambda_s$ is a coset of Λ_s relative to Λ. The set of relative cosets, denoted by $\Lambda/\Lambda_s = \{\lambda + \Lambda_s : \lambda \in \Lambda\}$, is called the *quotient group*. The point $\lambda \bmod \Lambda_s$ is the coset leader of $\lambda + \Lambda_s$. Moreover, the set of coset leaders, denoted by $\mathcal{C} \triangleq \Lambda \bmod \Lambda_s = \{\lambda \bmod \Lambda_s : \lambda \in \Lambda\}$, is a *nested lattice codebook*. The lattice codebook can also be written as $\mathcal{C} = \Lambda \cap \mathcal{V}_{\Lambda_s}$. In this case, the Voronoi region of Λ_s is called the *shaping region* and Λ_s is referred as *shaping lattice*.

2.4.1 Lattice Reduction

An algorithm that, for a given $\mathbf{B}$, finds a "good" basis $\mathbf{B}' = \mathbf{BA}$ (in column notation) is called a *lattice reduction algorithm* [29]. Generally, the idea of a "good" basis refers to using vectors that are as short as possible and nearly orthogonal. There are many lattice reductions techniques, each one having different criteria [30]: (a) the Minkowski reduction (b) the Korkine-Zolotareff (KZ) reduction and (c) the Lenstra-Lenstra-Lovász (LLL) reduction [31] (see also [32] for an overview of the LLL algorithm in the complex field). Minkowski and KZ reduction produce better results (i.e., shorter bases)[33], however they require a high computational cost, since they need to solve the shortest vector problem (SVP), which is NP-Hard. Meanwhile, the LLL algorithm, which is one of the most used, can find a suboptimal basis in polynomial time.

LLL Reduction

Let $\mathbf{B} = \begin{bmatrix}\mathbf{B}(1) & \cdots & \mathbf{B}(n)\end{bmatrix}$ be a basis of lattice Λ, where $\mathbf{B}(i) \in \mathbb{C}^n$ is the ith column of $\mathbf{B}$, $i = 1, \ldots, n$. Let $\mathbf{B}' = \begin{bmatrix}\mathbf{B}'(1) & \cdots & \mathbf{B}'(n)\end{bmatrix}$ be the Gram-Schmitd orthogonalization of $\mathbf{B}$, i.e.,

$$\mathbf{B}'(i) = \mathbf{B}(i) - \sum_{j=1}^{i-1} \mu_{ij}\mathbf{B}'(j) \tag{2.30}$$

where

$$\mu_{ij} = \frac{\mathbf{B}(i)^{\mathrm{H}}\mathbf{B}'(j)}{\|\mathbf{B}'(j)\|^2}. \tag{2.31}$$

We say that $\mathbf{B}'$ is a LLL-reduced basis if it satisfied both the *size-reduction condition*, i.e.,

$$|\Re\{\mu_{ij}\}| \leq \frac{1}{2} \qquad \text{and} \qquad |\Im\{\mu_{ij}\}| \leq \frac{1}{2} \tag{2.32}$$

for $1 \leq j < i \leq n$, and the *Lovász condition*, i.e.,

$$\|\mathbf{B}'(i) + \mu_{i,i-1}\mathbf{B}'(i-1)\|^2 \geq \delta \, \|\mathbf{B}'(i-1)\|^2 \tag{2.33}$$

for $i = 2, \dots, n$, where $\delta \in \left(\frac{1}{4}, 1\right]$ (typically $\delta = \frac{3}{4}$ is chosen) [31, 32].

The LLL algorithm consists in, for a given $\mathbf{B}$, find a LLL-reduced basis $\mathbf{B}' = \mathbf{BA}$. Starting with $i = 2$, it computes (2.30) and verifies if (2.32) and (2.33) are satisfied. Whenever (2.32) is not satisfied then $\mathbf{B}(i)$ is updated to $\mathbf{B}(i) - \lfloor \mu_{ij} \rceil \mathbf{B}(j)$, where $\lfloor \cdot \rceil$ is the rounding to the nearest (Gaussian) integer operator, and μ_{ij} and $\mathbf{B}'(i)$ are recomputed. Then, if (2.33) is not satisfied, $\mathbf{B}(i)$ and $\mathbf{B}(i-1)$ are swapped and the iteration index i is replaced by $i-1$ (provided that is still greater than 2). The complete procedure of (complex) LLL algorithm can be found in [32].

The complexity of LLL algorithm is hard to precisely estimate since it mostly depends on the number of times that the Lovász condition is not satisfied [31, 33]. It is estimated that, on average, the LLL algorithm requires $\mathcal{O}(n^2 \log n)$ iterations [32]. Each iteration has a complexity of $\mathcal{O}(n^2)$, which results in a total complexity of $\mathcal{O}(n^4 \log n)$.

Note that, if $\mathbf{B}$ is an upper unitriangular, i.e., triangular with ones in the main diagonal, and $\mathbf{B}'$ and μ_{ij} are defined as (2.30) and (2.31), respectively, then the Lovász condition (2.33) is automatically satisfied [13]. In this case, the LLL algorithm has a fixed number n of iterations, and therefore, a total complexity of $\mathcal{O}(n^3)$.

2.5 Lattice-Reduction Aided (LRA) Precoding

As said before, although low complexity, the performance of linear methods is far from the sum capacity. If one allows a more complex implementation in order to increase the performance, non-linear techniques often are used as an alternative. Some non-linear schemes that appear to be interesting are the lattice-reduction aided (LRA) approach [5, 6] and the integer-forcing (IF) approach [8, 7, 9], since both can potentially achieve higher rates than linear precoding, without a significant increase in complexity. In this section, we explain about LRA precoding. The IF precoding is presented in the next section.

Let $\Lambda \subseteq \mathbb{C}^n$ and Λ_s be a nested lattice of Λ. We assume that the second moment (per dimension) of Λ_s is $P_{\Lambda_s} = \text{SNR}$. We define the constellation symbols as $\mathcal{M} = (\Lambda + \mathbf{s}) \cap \mathcal{V}_{\Lambda_s}$, where $\mathbf{s} \in \mathbb{C}^n$ is a shift chosen to ensure zero-mean constellations. Note that, if $\Lambda = \mathbb{Z}[j]^n$ and $\Lambda_s = \sqrt{M}\Lambda$, where $\sqrt{M} \in \mathbb{N}$, then, with an appropriate choice of $\mathbf{s}$, $\mathcal{M}$ is equivalent to a square M-QAM constellation.

Let $\mathbf{w}_i \in \mathcal{M}$ be the data to be sent to the ith user, $i = 1, \dots, K$ and $\mathbf{W} \triangleq \begin{bmatrix} \mathbf{w}_1^{\mathrm{T}} & \cdots & \mathbf{w}_K^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$. Let $\mathbf{A} \in \mathbb{Z}[j]^{K \times K}$ be a unimodular matrix, i.e., an integer matrix with $|\det \mathbf{A}| = 1$.

In the first step of LRA precoding, the transmitter computes the precoded message $\mathbf{w}'_1, \dots, \mathbf{w}'_K$ as

$$\mathbf{W}' = \mathbf{A}^{-1}\mathbf{W} \tag{2.34}$$

where $\mathbf{W}' = \begin{bmatrix} \mathbf{w}'^{\mathrm{T}}_1 & \cdots & \mathbf{w}'^{\mathrm{T}}_K \end{bmatrix}^{\mathrm{T}}$.

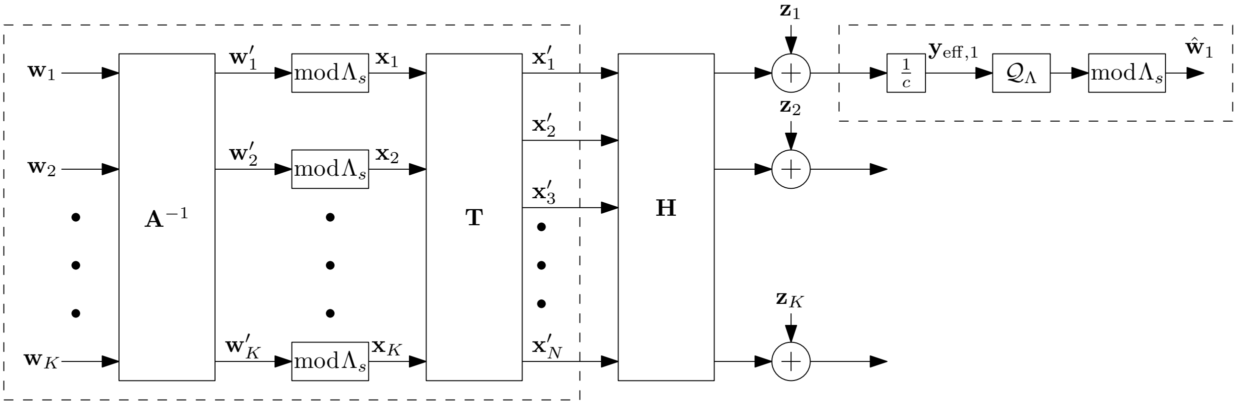


Figure 2.2: A LRA scheme. On the left the transmitter. On the right, only one receiver is shown.

After that, a lattice-modulus operation is applied in each row of $\mathbf{W}'$, producing

$$\mathbf{x}_i = \mathbf{w}'_i \bmod \Lambda_s. \tag{2.35}$$

This step ensures that $\mathbb{E}\left[\|\mathbf{x}_i\|^2\right] \leq n\text{SNR}$, for $i = 1, \dots, K$.

Finally, similar to the linear techniques, a precoding matrix $\mathbf{T} \in \mathbb{C}^{N \times K}$ is applied to the matrix $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}} & \cdots & \mathbf{x}_K^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$. Recall from (2.5) that the transmitted matrix $\mathbf{X}'$ is given by

$$\mathbf{X}' = \mathbf{T}\mathbf{X}. \tag{2.36}$$

Note that the constraint (2.9) must still be satisfied.

At the receiver side, a gain $\alpha_i \in \mathbb{C}$ is applied, producing

$$\begin{aligned} \mathbf{y}_{\text{eff},i} &= \alpha_i \mathbf{y}_i = \alpha_i \left(\mathbf{h}_i \mathbf{X}' + \mathbf{z}_i\right) && (2.37)\\ &= \mathbf{a}_i \mathbf{X} + \mathbf{z}_{\text{eff},i} && (2.38)\\ &= \mathbf{a}_i (\mathbf{A}^{-1}\mathbf{W} \bmod \Lambda_s) + \mathbf{z}_{\text{eff},i} && (2.39)\\ &= \mathbf{w}_i + \mathbf{A}\lambda + \mathbf{z}_{\text{eff},i} && (2.40) \end{aligned}$$

where $\mathbf{z}_{\text{eff},i} = (\alpha_i \mathbf{h}_i \mathbf{T} - \mathbf{a}_i)\mathbf{X} + \alpha_i \mathbf{z}_i$. Recall that $\mathbf{x} \bmod \Lambda_s = \mathbf{x} + \lambda$, and since $\mathbf{A}\lambda \in \Lambda$, then $\mathbf{w}_i + \mathbf{A}\lambda$ correspond to a periodically extension of the data points. The receiver can now estimate its data by

$$\begin{aligned} \hat{\mathbf{w}}_i &= \mathcal{Q}_\Lambda \left(\mathbf{y}_{\text{eff},i}\right) \bmod \Lambda_s && (2.41)\\ &= \mathbf{w}_i + \mathcal{Q}_\Lambda \left(\mathbf{z}_{\text{eff},i}\right) \bmod \Lambda_s && (2.42) \end{aligned}$$

Note that if $\mathcal{Q}_\Lambda(\mathbf{z}_{\text{eff},i}) \in \Lambda_s$ then the receiver can correctly infer $\mathbf{w}_i$, i.e., $\hat{\mathbf{w}}_i = \mathbf{w}_i$.

A LRA scheme tries to find $\mathbf{A}$ and $\mathbf{T}$ based on some heuristic in order to reduce the error probability. Those parameters can be obtained through a factorization of the channel matrix $\mathbf{H}$. By using any lattice reduction algorithm, for example the LLL algorithm, we can decompose $\mathbf{H} = \mathbf{A}\mathbf{H}_r$,[2] where $\mathbf{H}_r$ is the reduced channel matrix with same dimension as $\mathbf{H}$. In the classical

[2] In according of notation in 2.4.1, we should write the decomposition as $\mathbf{H}_r^{\mathrm{T}} = \mathbf{H}^{\mathrm{T}}\mathbf{A}^{-\mathrm{T}}$.

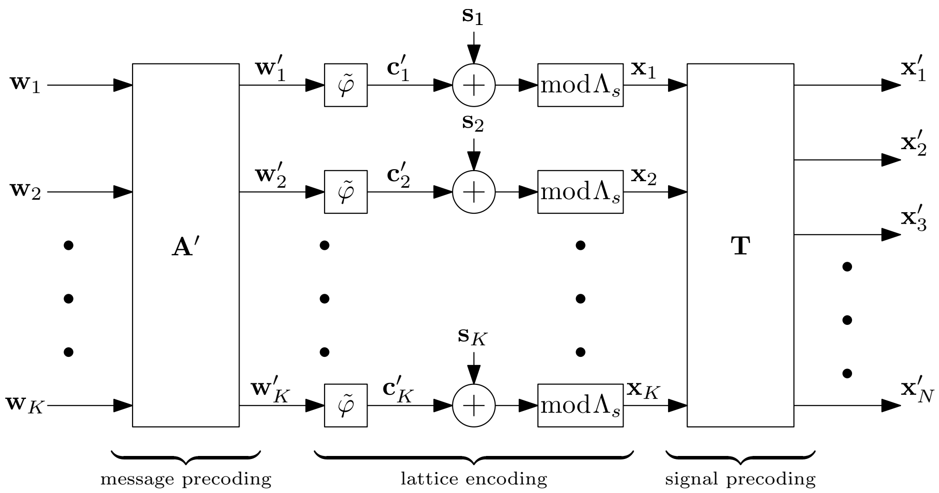


Figure 2.3: A transmitter using an IF precoding.

approach of LRA [5, 6], the precoding matrix $\mathbf{T}$ is chosen as

$$\begin{aligned}\mathbf{T} &= c\mathbf{H}_r^{\mathrm{H}}(\mathbf{H}_r\mathbf{H}_r^{\mathrm{H}})^{-1} \\ &= c\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{A}.\end{aligned} \tag{2.43}$$

where c is a constant to ensure (2.3). If $\mathbf{A} = \mathbf{I}$ then (2.43) corresponds to ZF precoding (2.19) where we allocate the same power to all users.

In [6], advanced strategies are presented for LRA precoding. We would like to emphasize the one called *factorization of the Hermitian of the inverse augmented channel matrix.* Let $\mathcal{H} \triangleq \left[\mathbf{H} \quad \sqrt{\frac{K}{\mathrm{SNR}}}\mathbf{I}\right] \in \mathbb{C}^{K\times N+K}$ be the augmented channel matrix and consider the following factorization

$$\boldsymbol{\mathcal{H}}^{\mathrm{H}}\left(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^{\mathrm{H}}\right)^{-1} = \boldsymbol{\mathcal{T}}\mathbf{A}^{-1}. \tag{2.44}$$

The precoding matrix for LRA $\mathbf{T}$ is obtained from the first N rows of

$$\boldsymbol{\mathcal{T}} = c\boldsymbol{\mathcal{H}}^{\mathrm{H}}\left(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^{\mathrm{H}}\right)^{-1}\mathbf{A} \tag{2.45}$$

i.e.,

$$\mathbf{T} = c\mathbf{H}^{\mathrm{H}}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \frac{K}{\mathrm{SNR}}\mathbf{I}\right)^{-1}\mathbf{A} \tag{2.46}$$

where c is a constant to ensure (2.3). Note that if $\mathbf{A} = \mathbf{I}$ then the precoding matrix of LRA corresponds to the RZF precoding matrix (2.24) where the power allocated for each user is the same.

2.6 Integer-Forcing (IF) Precoding

2.6.1 Transmitter side

At the transmitter side, IF precoding can be divided in three main steps: shows a block diagram for IF precoding with the three main steps.

Let $\Lambda \subseteq \mathbb{C}^n$ be a lattice and $\Lambda_s = p\Lambda$ be the shapping lattice, where $p \in \mathbb{Z}$ is a prime satisfying $p \bmod 4 \equiv 3$ so that $\mathbb{Z}_p[j]$ is a finite field [34]. We assume that the second moment (per dimension) of Λ_s is $P_{\Lambda_s} = \mathrm{SNR}$. Let $\mathcal{C} = \Lambda \cap \mathcal{V}_{\Lambda_s}$ be a nested lattice codebook, which is used by the transmitter.

The transmitter selects a message $\mathbf{w}_i \in \mathcal{W}_i$ to be sent to the ith user, where $\mathcal{W}_i \subseteq \mathcal{W}$ is the *message space* of the ith user and $\mathcal{W} = \mathbb{Z}_p[j]^n$ is the *ambient space*.

Let $\varphi : \Lambda \to \mathcal{W}$ be a $\mathbb{Z}[j]$-linear mapping with kernel Λ_s and let $\tilde{\varphi} : \mathcal{W} \to \mathcal{C}$ be a bijective encoding function, such that $\varphi(\tilde{\varphi}(\mathbf{w})) = \mathbf{w}$ for all $\mathbf{w} \in \mathcal{W}$. Finally, we define the lattice code $\mathcal{C}_i \subseteq \mathcal{C}$ such that $\mathcal{C}_i = \varphi(\mathcal{W}_i)$, which is used as a decoder for the ith user.

Message Precoding

Let $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^{\mathrm{T}} & \cdots & \mathbf{a}_K^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{Z}[j]^{K \times K}$ be a matrix such that $\det(\mathbf{A}) \neq 0 \bmod p$, i.e., $\mathbf{A}$ is invertible modulo p, and let $\mathbf{A}' \in \mathbb{Z}[j]^{K \times K}$ be a matrix such that

$$\mathbf{A}\mathbf{A}' = \mathbf{I} \bmod p. \tag{2.47}$$

Let $\mathbf{W} = \begin{bmatrix} \mathbf{w}_1^{\mathrm{T}} & \cdots & \mathbf{w}_K^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathcal{W}^K$ be a matrix whose ith row is given by $\mathbf{w}_i \in \mathcal{W}_i$. The transmitter computes the *precoded messages* $\mathbf{w}'_1, \ldots, \mathbf{w}'_K \in \mathcal{W}$ as

$$\mathbf{W}' = \mathbf{A}'\mathbf{W} \bmod p \tag{2.48}$$

where $\mathbf{W}' = \begin{bmatrix} \mathbf{w}_1'^{\mathrm{T}} & \cdots & \mathbf{w}_K'^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ is a matrix whose ith row is $\mathbf{w}'_i$.

Lattice Encoding

Then, for each $\mathbf{w}'_i$, the lattice-code $\mathcal{C}$ is applied generating the codeword

$$\mathbf{c}'_i = \tilde{\varphi}(\mathbf{w}'_i). \tag{2.49}$$

After that, each codeword is shifted by the dither vector $\mathbf{u}_i \in \mathbb{C}^n$ and reduced modulus Λ_s generating

$$\mathbf{x}_i = \mathbf{c}'_i + \mathbf{u}_i \bmod \Lambda_s. \tag{2.50}$$

The dither vector may be chosen by a uniform distribution over $\mathcal{V}_{\Lambda_s}$ [29], however, a fixed dither is also possible [35]. The dither vector together with the lattice-modulus operator ensure that $\mathbb{E}[\|\mathbf{x}_i\|^2] \leq n\mathrm{SNR}$.

Signal Precoding

Finally, similar to the linear techniques, a precoding matrix $\mathbf{T} \in \mathbb{C}^{N \times K}$ is applied to the matrix $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}} & \cdots & \mathbf{x}_K^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$. Recall from (2.5) that the transmitted matrix $\mathbf{X}'$ is given by

$$\mathbf{X}' = \mathbf{T}\mathbf{X}. \tag{2.51}$$

Note that the constraint (2.9) must still be satisfied.

2.6.2 Receiver side

Recall from linear techniques that the signal received by the ith user is given by

$$\mathbf{y}_i = \mathbf{h}'_i\mathbf{X} + \mathbf{z}_i \tag{2.52}$$

where $\mathbf{h}'_i = \mathbf{h}_i\mathbf{T}$. The receiver, then, applies a scalar $\alpha_i \in \mathbb{C}$ and computes

$$\begin{aligned}\mathbf{y}_{\text{eff},i} &= \alpha_i\mathbf{y}_i - \mathbf{a}_i\mathbf{U} \bmod \Lambda_s &(2.53)\\ &= \mathbf{a}_i(\mathbf{X}-\mathbf{U}) + (\alpha_i\mathbf{h}'_i - \mathbf{a}_i)\,\mathbf{X} + \alpha_i\mathbf{z}_i \bmod \Lambda_s &(2.54)\\ &= \mathbf{a}_i\mathbf{C}' + \mathbf{z}_{\text{eff},i} \bmod \Lambda_s &(2.55)\\ &= \mathbf{c}_i + \mathbf{z}_{\text{eff},i} \bmod \Lambda_s &(2.56)\end{aligned}$$

where $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1^{\mathrm{T}} & \cdots & \mathbf{u}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$, $\mathbf{c}_i = \mathbf{a}_i\mathbf{C}' \bmod \Lambda_s$ and

$$\mathbf{z}_{\text{eff},i} = (\alpha_i\mathbf{h}'_i - \mathbf{a}_i)\,\mathbf{X} + \alpha_i\mathbf{z}_i \tag{2.57}$$

is the effective noise.

The receiver can now infer a codeword $\hat{\mathbf{c}}_i$ from $\mathbf{y}_{\text{eff},i}$ by applying a quantization function over the fine lattice, i.e.,

$$\begin{aligned}\hat{\mathbf{c}}_i &= \mathcal{Q}_\Lambda(\mathbf{y}_{\text{eff},i}) \bmod \Lambda_s &(2.58)\\ &= \mathcal{Q}_\Lambda(\mathbf{c}_i + \mathbf{z}_{\text{eff},i} \bmod \Lambda_s) \bmod \Lambda_s &(2.59)\\ &= \mathbf{c}_i + \mathcal{Q}_\Lambda(\mathbf{z}_{\text{eff},i}) \bmod \Lambda_s &(2.60)\end{aligned}$$

Note that if $\mathcal{Q}_\Lambda(\mathbf{z}_{\text{eff},i}) \in \Lambda_s$ then the receiver can correctly infer $\mathbf{c}_i$, i.e., $\hat{\mathbf{c}}_i = \mathbf{c}_i$. Suppose now that $\mathbf{c}_i$ was successfully recovered, the user can now obtain its desired message $\mathbf{w}_i$ by applying the mapping φ. More precisely,

$$\varphi(\mathbf{c}_i) = \varphi(\mathbf{a}_i\mathbf{C}' \bmod \Lambda_s) = \mathbf{a}_i\varphi(\mathbf{C}') = \mathbf{a}_i\mathbf{W}' = \mathbf{a}_i\mathbf{A}'\mathbf{W} = \mathbf{w}_i \tag{2.61}$$

since φ is $\mathbb{Z}[j]$-linear. Note that this means that $\mathbf{c}_i = \tilde{\varphi}(\mathbf{w}_i) \in \mathcal{C}$, in other words, $\mathbf{c}_i$ is the codeword corresponding to the original message $\mathbf{w}_i$.

Theorem 1. *[9, 36, 37, 38] For p and n sufficiently large, there is an IF precoding scheme with achievable sum rate*

$$R_{\text{IF}}(\mathbf{H},\mathbf{A},\mathbf{T}) = \sum_{i=1}^{K} R_{\text{comp}}(\mathbf{h}_i\mathbf{T},\mathbf{a}_i) \tag{2.62}$$

where

$$R_{\text{comp}}(\mathbf{h}'_i,\mathbf{a}_i) = \log_2^+\left(\frac{\text{SNR}}{\sigma^2_{\text{eff},i}}\right) \tag{2.63}$$

is the individual rate for each user (also called computational rate *in the compute-and-forward framework) and*

$$\sigma^2_{\text{eff},i} = \frac{1}{n}\mathbb{E}[\|\mathbf{z}_{\text{eff},i}\|^2] = \|\alpha_i\mathbf{h}'_i - \mathbf{a}_i\|^2\,\text{SNR} + |\alpha_i|^2 \tag{2.64}$$

is the variance of the effective noise.

The optimal value of α_i can be easily calculated by minimizing the variance of the effective noise $\sigma^2_{\mathrm{eff},i}$, and it is given by

$$\alpha_i^{\mathrm{opt}} = \frac{\mathbf{a}_i \mathbf{h}_i'^{\mathrm{H}} \mathrm{SNR}}{\|\mathbf{h}_i'\|^2 \, \mathrm{SNR} + 1}. \tag{2.65}$$

Assuming that (2.65) is used, we can rewrite (2.63) as

$$\begin{aligned} R_{\mathrm{comp}}(\mathbf{h}_i', \mathbf{a}_i) &= \log_2^+ \left(\frac{1}{\mathbf{a}_i \left(\mathbf{I} - \frac{\mathrm{SNR}}{\mathrm{SNR}\|\mathbf{h}_i'\|^2+1} \mathbf{h}_i'^{\mathrm{H}} \mathbf{h}_i' \right) \mathbf{a}_i^{\mathrm{H}}} \right) & (2.66) \\ &= \log_2^+ \left(\frac{1}{\|\mathbf{a}_i\|^2 - \frac{1}{\|\mathbf{h}_i'\|^2 + \mathrm{SNR}^{-1}} |\mathbf{a}_i \mathbf{h}_i'^{\mathrm{H}}|^2} \right). & (2.67) \end{aligned}$$

Even though, for practical implementations, it is necessary that $\mathbf{A}$ be invertible modulo p, for achievable rates, $\mathbf{A}$ only needs to be full rank (equivalently, invertible over $\mathbb{C}^n$) since p can be arbitrarily large. Therefore, the IF precoding tries to find a matrix $\mathbf{A} \in \mathbb{Z}[j]^{K\times K}$ with $\mathrm{rank}(\mathbf{A}) = K$ and a matrix $\mathbf{T} \in \mathbb{C}^{N\times K}$ with $\mathrm{Tr}(\mathbf{T}^{\mathrm{H}}\mathbf{T}) = 1$ that maximizes (2.62).

Note that, if $\mathbf{A} = \mathbf{I}$, then finding an optimal $\mathbf{T}$ corresponds to solving the optimal linear beamforming problem [2]. Moreover, if we choose $\mathbf{A} = \mathbf{I}$ and $\mathbf{T} = \mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\,\mathrm{diag}(\sqrt{\mathbf{p}})$, we have traditional ZF precoding, while choosing $\mathbf{A} = \mathbf{I}$ and $\mathbf{T} = \mathbf{H}^{\mathrm{H}}(K\mathrm{SNR}^{-1}\mathbf{I}+\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\,\mathrm{diag}(\sqrt{\mathbf{p}})$ we recover RZF precoding. As we can see, IF precoding is a generalization of linear precoding [39, 9].

As the authors of [9] showed, if we let the choice of $\mathbf{T}$ be more flexible, the problem can be more structured and potentially easier to solve. They also showed two approaches to choosing $\mathbf{T}$ for IF precoding, called DIF and RDIF, which will be more detailed in the next section.

2.6.3 DIF and RDIF Schemes

The first approach proposed by [9] chooses a precoding structure such that the channel seen by the users is an integer matrix, up to scaling for each user. More precisely, for any full rank integer matrix $\mathbf{A}$, the precoding matrix $\mathbf{T}$ is given

$$\mathbf{T} = c\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{D}\mathbf{A} \tag{2.68}$$

where $\mathbf{D} \in \mathbb{C}^{K\times K}$ is a diagonal matrix with nonzero entries such that $|\det \mathbf{D}| = 1$ and $c > 0$ is chosen to satisfy (2.9).

This method is called *diagonally-scaled exact integer-forcing* (DIF) precoding. It is clear that DIF is a generalization of ZF schemes, which corresponds to $\mathbf{A} = \mathbf{I}$ and $c\mathbf{D} = \mathrm{diag}(\sqrt{\mathbf{p}})$.

The DIF precoding is optimal in the high SNR regime [9], where it can achieve a sum rate given by

$$R_{\mathrm{DIF}}^{\mathrm{HI}}(\mathbf{H}, \mathbf{A}, \mathbf{D}) \triangleq K \log_2 \left(\frac{\mathrm{SNR}}{\mathrm{Tr}\left(\mathbf{A}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1}\mathbf{D}\mathbf{A}\right)} \right) \tag{2.69}$$

which is also shown to be a lower bound on the achievable rate for any SNR.

The second approach proposed in [9], which is called *regularized DIF* (RDIF), attempts to improve the performance of DIF for finite SNR by regularizing the matrix inversion in (2.68).

Specifically, (2.68) is modified as

$$\mathbf{T} = c\mathbf{H}^{\mathrm{H}} \left(\frac{K}{\mathrm{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^{\mathrm{H}} \right)^{-1} \mathbf{D}\mathbf{A}. \tag{2.70}$$

Just as DIF generalizes ZF, RDIF is a generalization of RZF scheme, which is obtained by making $\mathbf{A} = \mathbf{I}$ and $c\mathbf{D} = \mathrm{diag}(\sqrt{\mathbf{p}})$. In particular, RDIF reduces to DIF when SNR $\to \infty$.

Note that the diagonal of $\mathbf{D}$ in DIF/RDIF has a similar role as the vector $\mathbf{p}$ in ZF/RZF, i.e., the square of the ith element in the diagonal of $\mathbf{D}$ can be seen as the "power" gain applied to an integer linear combination $\mathbf{a}_i\mathbf{X}$. However, it is important to emphasize that, unless $\mathbf{A} = \mathbf{I}$, $\mathbf{D}$ is not a true power allocation since, in general, we cannot guarantee that $\mathbb{E}[\|\mathbf{a}_i\mathbf{X}\|^2] = n\mathrm{SNR}$.

Chapter 3

Optimization based on Achievable Sum Rates

3.1 Problem Statement

We are interested in finding matrices $\mathbf{A}$ and $\mathbf{D}$ that maximize the sum rate (2.62) for the RDIF scheme, i.e., with $\mathbf{T}$ chosen as in (2.70). In general, this is a hard problem due not only to the integer constraints on $\mathbf{A}$ but also to the complicated objective function (2.62). The latter difficulty is overcome in [9] by solving a simpler optimization problem, which can be interpreted as the minimization of a regularized version of the denominator in (2.69), namely,

$$\begin{aligned} \underset{\mathbf{A},\mathbf{D}}{\text{minimize}} \quad & f(\mathbf{A},\mathbf{D}) \triangleq \operatorname{Tr}(\mathbf{A}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}\mathbf{M}\mathbf{D}\mathbf{A}) \\ \text{s.t.} \quad & |\det \mathbf{D}| = 1 \\ & \operatorname{rank}(\mathbf{A}) = K \end{aligned} \tag{3.1}$$

where $\mathbf{A} \in \mathbb{Z}[j]^{K\times K}$, $\mathbf{D} \in \mathbb{C}^{K\times K}$ is diagonal, and

$$\mathbf{M} \triangleq \left(\frac{K}{\mathrm{SNR}}\mathbf{I} + \mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1}. \tag{3.2}$$

While generally a suboptimal heuristic, solving the above problem indeed maximizes the sum rate for the special case of asymptotically high SNR (where RDIF reduces to DIF).

The above problem was solved analytically in [9] for the special case $K = 2$. For $K > 2$, the problem is still open.

3.2 Fixed D

If $\mathbf{D}$ is fixed, then finding $\mathbf{A}$ that minimizes (3.1) corresponds to the *shortest independent vector problem* (SIVP) as shown by [9]. The SIVP can be sub-optimally solved using lattice basis reduction algorithms, for example the LLL algorithm [31, 32].

Let $\mathbf{B} \in \mathbb{C}^{K\times K}$ be a lattice generator matrix such that $\mathbf{B}^{\mathrm{H}}\mathbf{B} = \mathbf{D}^{\mathrm{H}}\mathbf{M}\mathbf{D}$ ($\mathbf{B}$ is any square root of $\mathbf{D}^{\mathrm{H}}\mathbf{M}\mathbf{D}$, in particular, $\mathbf{B}$ can be obtained using Cholesky decomposition). In this case, note that (3.1) can be rewritten as

$$\begin{aligned}\mathrm{Tr}(\mathbf{A}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}\mathbf{M}\mathbf{D}\mathbf{A}) &= \mathrm{Tr}(\mathbf{A}^{\mathrm{H}}\mathbf{B}^{\mathrm{H}}\mathbf{B}\mathbf{A}) \\ &= \sum_{j=1}^{K} \|\mathbf{B}\mathbf{A}(j)\|^2 \qquad (3.3)\end{aligned}$$

where $\mathbf{A}(j)$ is the jth column of $\mathbf{A}$.

Thus, we wish to find K linearly independent vectors in the lattice generated by the columns of $\mathbf{B}$ that minimizes (3.1). Those vectors will correspond to the columns of $\mathbf{A}$.

When $\mathbf{D} = \mathbf{I}$, the RDIF scheme becomes equivalent to the LRA precoding proposed [6], except for the fact that LRA precoding assumes symbol-level detection, while IF precoding employs codeword-level decoding, as discussed in Section 2.6 (see also [8, 9]). More precisely, the precoding matrix in LRA precoding has the same form as $\mathbf{T}$ in (2.70) (with $\mathbf{D} = \mathbf{I}$), while the integer matrix $\mathbf{A}$ is obtained in both schemes by solving the same lattice reduction problem.

We can clearly see that (2.70) becomes, up to scaling, the first N rows of (2.45) when $\mathbf{D} = \mathbf{I}$. Moreover, the integer matrix $\mathbf{A}$ found in (2.44) is the same solution obtained for RDIF by applying lattice reduction on the columns of a matrix $\mathbf{B}$ such that $\mathbf{B}^{\mathrm{H}}\mathbf{B} = \mathbf{M}$, as discussed in Section 3.1. This is because $(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^{\mathrm{H}})^{-1} = \mathbf{M}$, thus one possible square root of $\mathbf{M}$ is precisely the left-hand side of (2.44), i.e., $\mathbf{B} = \boldsymbol{\mathcal{H}}^{\mathrm{H}}(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^{\mathrm{H}})^{-1}$.

Although the factorization approaches proposed in [6] are based on heuristics, they correspond to a particular case of RDIF which tries to maximize the achievable sum rate. This means that, if we replace the left-hand side of (2.44) by $\boldsymbol{\mathcal{H}}^{\mathrm{H}}(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^{\mathrm{H}})^{-1}\mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix, before performing the factorization, we can potentially improve the performance of LRA precoding.

3.3 Proposed Method

Here we propose a method to find an approximate solution $(\mathbf{A}, \mathbf{D})$ to problem (3.1) for any K. The summary of the algorithm is described in Section 3.3.5. We start by proposing a convenient choice for the structure of $\mathbf{A}$.

3.3.1 Structure of $\mathbf{A}$

Consider the objective function in (3.1) and note that

$$f(\mathbf{A}, \mathbf{D}) = \sum_{i=1}^{K} M_{ii} \|\mathbf{a}_i\|^2 |d_i|^2 + \sum_{i=1}^{K} \sum_{j=i+1}^{K} 2M_{ji}\mathbf{a}_i\mathbf{a}_j^{\mathrm{H}} d_i d_j^* \qquad (3.4)$$

where $\mathbf{a}_i$ is the ith row of $\mathbf{A}$, d_i is the i-th element in the main diagonal of $\mathbf{D}$ and M_{ij} is the element of row i and column j of $\mathbf{M}$.

The first summation in (3.4) contains only nonnegative values. If we focus exclusively on minimizing $\|\mathbf{a}_i\|$, $i = 1, \ldots, K$, then it is easy to see that the optimal choice is $\mathbf{A} = \mathbf{I}$. However, since the second summation can have positive or negative values, we wish some degree of freedom to be able to minimize or maximize the absolute values of the inner products ($\left|\mathbf{a}_i\mathbf{a}_j^{\mathrm{H}}\right|$). To satisfy these conflicting requirements, we propose that $\mathbf{A}$ be *upper unitriangular* (upper triangular with ones along the main diagonal), i.e.

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1K} \\ 0 & 1 & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{3.5}$$

up to permutation of rows. An advantage of this structure is that the restriction of full rank $\mathbf{A}$ is automatically satisfied. Note that, for $K > 2$, a row permutation of $\mathbf{A}$ may change the achievable rate.

We first consider $\mathbf{A}$ exactly in upper unitriangular form. The generalization to other permutations is discussed in 3.3.4.

3.3.2 Relaxed Problem

Even with the proposed structure for $\mathbf{A}$, we still have an integer optimization problem, which is generally hard to solve. In order to circumvent this difficulty, we consider in this section a relaxation of the problem where the indeterminate entries of $\mathbf{A}$ can be any complex number.

Theorem 2. *Under the relaxed constraint that* $\mathbf{A} \in \mathbb{C}^{K\times K}$ *and the additional constraint that* $\mathbf{A}$ *be upper unitriangular, problem* (3.1) *has a solution given by*

$$\tilde{\mathbf{A}} = \mathbf{D}^{-1}\mathbf{U}^{-1}\mathbf{D} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^{-1}\mathbf{\Lambda}^{-\frac{1}{2}} \tag{3.6}$$

$$\tilde{\mathbf{D}} = (\det \mathbf{\Lambda})^{\frac{1}{2K}}\mathbf{\Lambda}^{-\frac{1}{2}} \tag{3.7}$$

where $\mathbf{U} \in \mathbb{C}^{K\times K}$ *is upper unitriangular and* $\mathbf{\Lambda} \in \mathbb{R}^{K\times K}$ *is diagonal such that* $\mathbf{M} = \mathbf{U}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{U}$. *The solution for* $\mathbf{A}$ *as a function of* $\mathbf{D}$ *is unique and the optimal solution for* $\mathbf{D}$ *(with the corresponding optimal* $\mathbf{A}$*) is unique up to a phase shift for each of the diagonal entries. The optimal value of the problem is* $K(\det \mathbf{M})^{1/K}$.

Proof. Let $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{D}}$ be a solution to (3.1) with $\mathbf{A} \in \mathbb{C}^{K\times K}$. We first find $\tilde{\mathbf{A}}$ as a function of $\mathbf{D}$ and then find $\tilde{\mathbf{D}}$.

Let $\boldsymbol{\nabla}_{\mathbf{A}} f$ be a matrix whose (i,j)th element is the partial derivative of (3.1) with respect to a_{ij} if $i < j$ and zero otherwise. Note that $(\boldsymbol{\nabla}_{\mathbf{A}} f)_{ij} = (2\mathbf{D}^{\mathrm{H}}\mathbf{M}\mathbf{D}\mathbf{A})_{ij}$ if $i < j$. The critical points of f with respect to $\mathbf{A}$ are those which satisfy, for all j and all $i < j$,

$$0 = (\boldsymbol{\nabla}_{\mathbf{A}} f)_{ij} = (2\mathbf{D}^{\mathrm{H}}\mathbf{M}\mathbf{D}\mathbf{A})_{ij}. \tag{3.8}$$

Multiplying by $(2d_i^*)^{-1}$ and d_j^{-1} on both sides, this is equivalent to requiring that, for all j and all $i < j$,

$$0 = (\mathbf{M}\mathbf{D}\mathbf{A}\mathbf{D}^{-1})_{ij} = (\mathbf{M}\mathbf{A}')_{ij} \tag{3.9}$$

where $\mathbf{A}' = \mathbf{D}\mathbf{A}\mathbf{D}^{-1} \in \mathbb{C}^{K\times K}$ is also upper unitriangular.

Note that (3.9) implies that a critical point is any matrix $\mathbf{A} = \mathbf{D}^{-1}\mathbf{A}'\mathbf{D}$ such that $\mathbf{M}\mathbf{A}' = \mathbf{L}$ is a lower triangular matrix. Thus, any solution, if it exists, can be found by computing an LU decomposition of $\mathbf{M} = \mathbf{L}\mathbf{A}'^{-1}$. Moreover, since we require that the diagonal of $\mathbf{A}'$ consists of ones, such a decomposition is unique whenever it exists.

Since $\mathbf{M}$ is a symmetric positive definite matrix, such an LU decomposition always exists. Specifically, it admits an LDL decomposition $\mathbf{M} = \mathbf{U}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{U}$, where $\mathbf{U}$ is an upper unitriangular matrix and $\mathbf{\Lambda}$ is a diagonal matrix with real and positive diagonal entries. Thus, $\mathbf{A}' = \mathbf{U}^{-1}$ is the unique solution to (3.9), which gives

$$\tilde{\mathbf{A}} = \mathbf{D}^{-1}\mathbf{U}^{-1}\mathbf{D}. \tag{3.10}$$

Now, substituting $\tilde{\mathbf{A}}$ in (3.1), we have that

$$f(\tilde{\mathbf{A}}, \mathbf{D}) = \mathrm{Tr}(\mathbf{D}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{D}) = \sum_{i=1}^{K} |d_i|^2 \lambda_i \tag{3.11}$$

where $\lambda_i > 0$ and d_i are the ith diagonal element of $\mathbf{\Lambda}$ and $\mathbf{D}$, respectively. Due to the inequality of arithmetic and geometric means, we have that

$$\frac{1}{K} f(\tilde{\mathbf{A}}, \mathbf{D}) = \frac{1}{K}\sum_{i=1}^{K} |d_i|^2 \lambda_i \geq \left(\prod_{i=1}^{K} |d_i|^2 \lambda_i\right)^{\frac{1}{K}} \tag{3.12}$$

with equality if and only if $|d_1|^2 \lambda_1 = \cdots = |d_K|^2 \lambda_K$.

Applying the constraint $|\det \mathbf{D}| = 1$, we have

$$\left(\prod_{i=1}^{K} |d_i|^2 \lambda_i\right)^{\frac{1}{K}} = \left(\prod_{i=1}^{K} \lambda_i\right)^{\frac{1}{K}} = (\det \mathbf{\Lambda})^{\frac{1}{K}}. \tag{3.13}$$

Thus, the bound in (3.12) is achievable by setting each term $|d_i|^2 \lambda_i$ equal to the right hand side of (3.13), i.e.,

$$\mathbf{D}^{\mathrm{H}}\mathbf{D} = (\det \mathbf{\Lambda})^{\frac{1}{K}} \mathbf{\Lambda}^{-1}. \tag{3.14}$$

By choosing each d_i to be real and positive, one solution is given by (3.7), which applied in (3.10) gives (3.6).

Finally, since $\det \mathbf{\Lambda} = \det \mathbf{M}$, we have $f(\tilde{\mathbf{A}}, \tilde{\mathbf{D}}) = K (\det \mathbf{\Lambda})^{\frac{1}{K}} = K (\det \mathbf{M})^{\frac{1}{K}}$, completing the proof. □

Remark 1. *If we let* $\mathrm{SNR} \to \infty$ *and replace the optimal value of the relaxed problem in* (2.69), *we obtain an upper bound on the rate achievable by DIF in this regime. This bound happens to coincide with the high SNR expression for the sum capacity,*

$$C_{\mathrm{sum}}^{\mathrm{HI}} = K \log_2\left(\frac{\mathrm{SNR}}{K}\right) + \log_2(\det \mathbf{H}\mathbf{H}^{\mathrm{H}}). \tag{3.15}$$

Of course, this bound is rarely achievable, since $\mathbf{A}$ *is constrained to be an integer matrix. More precisely, the bound is achievable if and only if each row of* $\mathbf{H}$ *is a multiple of an integer vector.*

3.3.3 Optimization of A

We now show how to find an approximate solution $(\mathbf{A}, \mathbf{D})$ to problem (3.1) satisfying $\mathbf{A} \in \mathbb{Z}[j]^{K \times K}$, starting from a solution $(\tilde{\mathbf{A}}, \tilde{\mathbf{D}})$ to the relaxed problem. First take $\mathbf{D} = \tilde{\mathbf{D}}$, and note that

$$
\begin{aligned}
f(\mathbf{A},\mathbf{D}) &= (\det \mathbf{\Lambda})^{\frac{1}{K}} \operatorname{Tr}(\mathbf{A}^{\mathrm{H}}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{M}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{A}) && (3.16)\\
&= (\det \mathbf{\Lambda})^{\frac{1}{K}} \operatorname{Tr}(\mathbf{A}^{\mathrm{H}}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{A}) && (3.17)\\
&= (\det \mathbf{\Lambda})^{-\frac{1}{K}} \operatorname{Tr}(\mathbf{A}^{\mathrm{H}}\tilde{\mathbf{A}}^{-\mathrm{H}}\tilde{\mathbf{A}}^{-1}\mathbf{A}) && (3.18)\\
&= (\det \mathbf{\Lambda})^{-\frac{1}{K}} \sum_{i=1}^{K} \|\mathbf{B}\mathbf{A}(i)\|^2 && (3.19)
\end{aligned}
$$

where $\mathbf{B} \triangleq \tilde{\mathbf{A}}^{-1} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}}$ and $\mathbf{A}(i)$ is the ith column of $\mathbf{A}$.

Note that, finding a Gaussian integer matrix $\mathbf{A}$ that minimizes (3.19) is the same problem described in Section 3.1 (see also [9]) with $\mathbf{D}$ fixed. That means that the LLL algorithm can be used to find K shortest linearly independent vectors from the lattice generated by the columns of $\mathbf{B}$.

3.3.4 Permutations

Let $\bar{\mathbf{A}}$ be an Gaussian integer matrix with a structure given by (3.5) and suppose we want to solve (3.1) under the constraint that $\mathbf{A} = \mathbf{P}\bar{\mathbf{A}}$ where $\mathbf{P}$ is a permutation matrix.

First, note that

$$
\begin{aligned}
\operatorname{Tr}(\mathbf{A}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}\mathbf{M}\mathbf{D}\mathbf{A}) &= \operatorname{Tr}(\bar{\mathbf{A}}^{\mathrm{H}}\mathbf{P}^{\mathrm{T}}\mathbf{D}^{\mathrm{H}}\mathbf{M}\mathbf{D}\mathbf{P}\bar{\mathbf{A}})\\
&= \operatorname{Tr}(\bar{\mathbf{A}}^{\mathrm{H}}\bar{\mathbf{D}}^{\mathrm{H}}\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P}\bar{\mathbf{D}}\bar{\mathbf{A}})
\end{aligned}
$$

where $\bar{\mathbf{D}} = \mathbf{P}^{\mathrm{T}}\mathbf{D}\mathbf{P}$. Thus, we can use the solution of Theorem 2 with $\mathbf{M}$ replaced by $\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P}$ or, equivalently, $\mathbf{M}^{-1}$ replaced by $\mathbf{P}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{P}$ to obtain

$$
\begin{aligned}
\mathbf{D} &= \mathbf{P}\tilde{\mathbf{D}}\mathbf{P}^{\mathrm{T}} && (3.20)\\
\mathbf{A} &= \mathbf{P}\bar{\mathbf{A}} && (3.21)
\end{aligned}
$$

where $\bar{\mathbf{A}}$ is the output of LLL algorithm.

3.3.5 Summary of the Method

The steps described above allow us to find a choice of $\mathbf{A}$ and $\mathbf{D}$ (and thus $\mathbf{T}$) for any given permutation $\mathbf{P}$ specifying the structure of $\mathbf{A}$. A summary of the proposed method is given in Algorithm 1.

Complexity Analysis

The complexity of an IF scheme is hard to precisely estimate. Generally, the lattice reduction algorithm is the bottleneck on the complexity. It is estimated that the LLL algorithm, one of the most used lattice reduction algorithms, requires $\mathcal{O}(K^4 \log K)$. However, in our case, since $\mathbf{B}$ in step 6 of Alg. 1 is an upper unitriangular matrix, the LLL algorithm can be computed with $\mathcal{O}(K^3)$ [13].

Other operations, such as, the computation of matrix $\mathbf{M}$ in step 1 or the computation of $\mathbf{T}$ in steps 8-10 require $\mathcal{O}(NK^2)$ operations each (recall that we assume $N \geq K$). The LDL

Algorithm 1 Proposed RDIF Design

```
Require: $\mathbf{H}$ and SNR
 1: Compute $\mathbf{M} = \left(K/\mathrm{SNR}\mathbf{I} + \mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1}$
 2: Generate a permutation matrix $\mathbf{P}$
 3: Compute the LDL decomposition $\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P} = \mathbf{U}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{U}$
 4: Compute $\tilde{\mathbf{D}} = (\det \mathbf{\Lambda})^{\frac{1}{2K}}\,\mathbf{\Lambda}^{-\frac{1}{2}}$
 5: Compute $\mathbf{B} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}}$
 6: Use the LLL algorithm using $\mathbf{B}$ as input to find $\bar{\mathbf{A}}$
 7: Set $\mathbf{D} = \mathbf{P}\tilde{\mathbf{D}}\mathbf{P}^{\mathrm{T}}$ and $\mathbf{A} = \mathbf{P}\bar{\mathbf{A}}$
 8: Compute $\mathbf{T}_0 \triangleq \mathbf{H}^{\mathrm{H}}\mathbf{M}\mathbf{D}\mathbf{A}$
 9: Compute $c = \mathrm{Tr}(\mathbf{T}_0^{\mathrm{H}}\mathbf{T}_0)^{-\frac{1}{2}}$
10: Compute $\mathbf{T} = c\mathbf{T}_0$
11: return $\mathbf{A}$ and $\mathbf{T}$
```

Table 3.1: Number of users K, for each method and for each value of SNR, used in Fig. 3.1.

Method	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	30 dB
$\mathbf{M}\downarrow$	16	16	13	15	16	16	16
Random	16	16	13	15	16	16	16
$\mathbf{D} = \mathbf{I}$	16	15	12	13	13	14	15
RZF	16	14	12	12	13	13	14
ZF	7	8	10	11	12	13	14

decomposition in step 3 requires $\mathcal{O}(K^3)$ operations. The remaining operations involves only (upper) triangular and diagonal matrices. Therefore, the total complexity is $\mathcal{O}(NK^2)$, which is the same asymptotic complexity of conventional linear precoding methods.

3.4 Simulation Results for Achievable Rate

In this section we show the average sum-rate performance of the proposed method. In our simulations, the sum rates were obtained through 10000 channel realizations. In each realization, the channel coefficients were randomly obtained considering a circularly symmetric complex Gaussian distribution with zero mean and unit variance.

In each simulation, we compare our proposed RDIF design to sum capacity [1] and to the conventional linear precoding methods, namely, ZF and RZF. We also compare to the RDIF approach mentioned in Section 3.2, where we fix $\mathbf{D} = \mathbf{I}$ and apply the LLL algorithm to find $\mathbf{A}$. This method is denoted by "$\mathbf{D} = \mathbf{I}$".

For our proposed method, we compare two heuristics. Specifically, we compare the heuristic where a random permutation is chosen, which is denoted by "Random", with a heuristic inspired by [40], where the permutation sorts the diagonal elements of $\mathbf{M}$ in descending order, which is denoted by "$\mathbf{M}\downarrow$".

In Figs. 3.1 and 3.2, we show the performance for $N = 16$ and $N = 64$ transmit antennas, respectively. For each method and for each value of SNR, we choose, through exhaustive search, the value of $K \leq N$ that achieves the highest sum rate. The number of users in Fig. 3.1 is shown in Table 3.1, while the number of users in Fig. 3.2 is shown in Table 3.2. As expected, the proposed method outperforms linear techniques as well as the previous RDIF approach

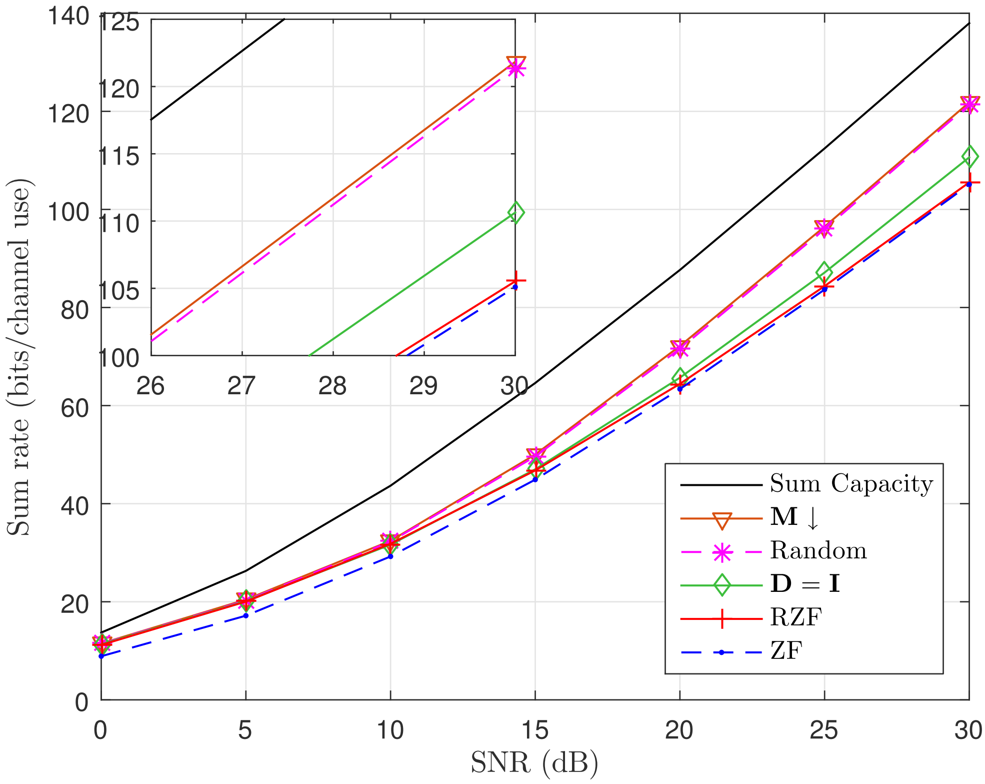


Figure 3.1: Sum rate for $N = 16$ transmit antennas. For each method and each value of SNR, the number of users $K \leq N$ was chosen to maximize the sum rate. On the box, a close up on SNR range of 26 to 30 dB.

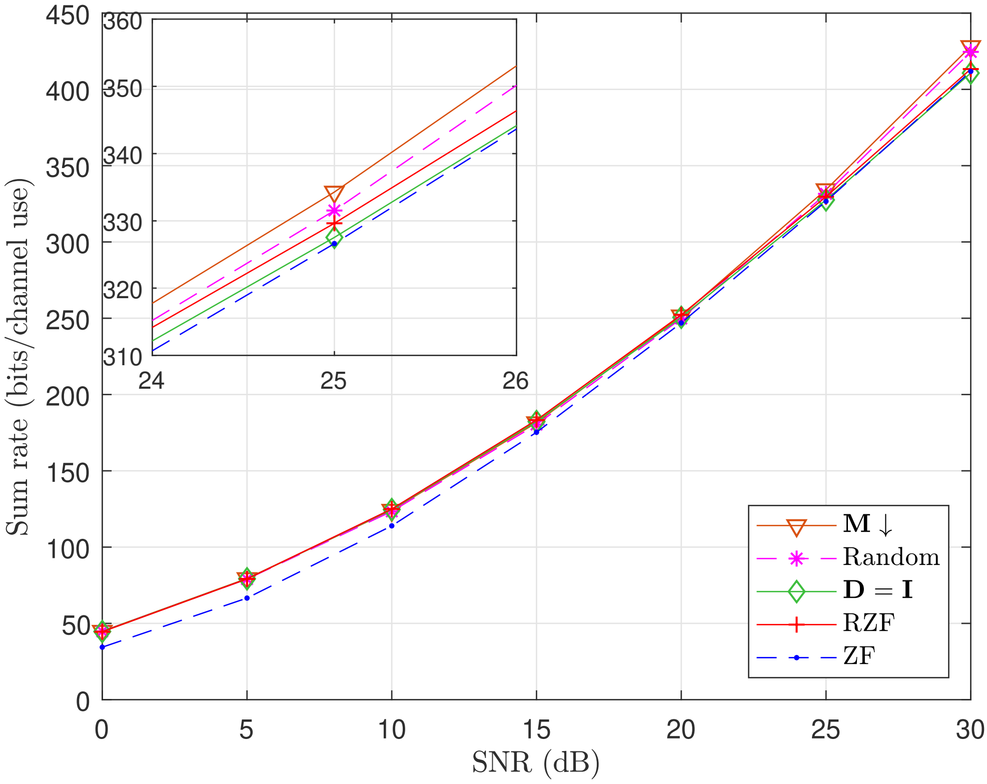


Figure 3.2: Sum rate for $N = 64$ transmit antennas. For each method and each value of SNR, the number of users $K \leq N$ was chosen to maximize the sum rate. On the box, a close up on SNR range of 24 to 26 dB.

Table 3.2: Number of users K, for each method and for each value of SNR, used in Fig. 3.2.

Method	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	30 dB
$\mathbf{M}\downarrow$	64	62	50	48	51	56	60
Random	64	62	50	48	50	55	62
$\mathbf{D} = \mathbf{I}$	64	62	50	48	50	51	53
RZF	64	62	50	48	50	51	53
ZF	24	31	37	44	47	51	53

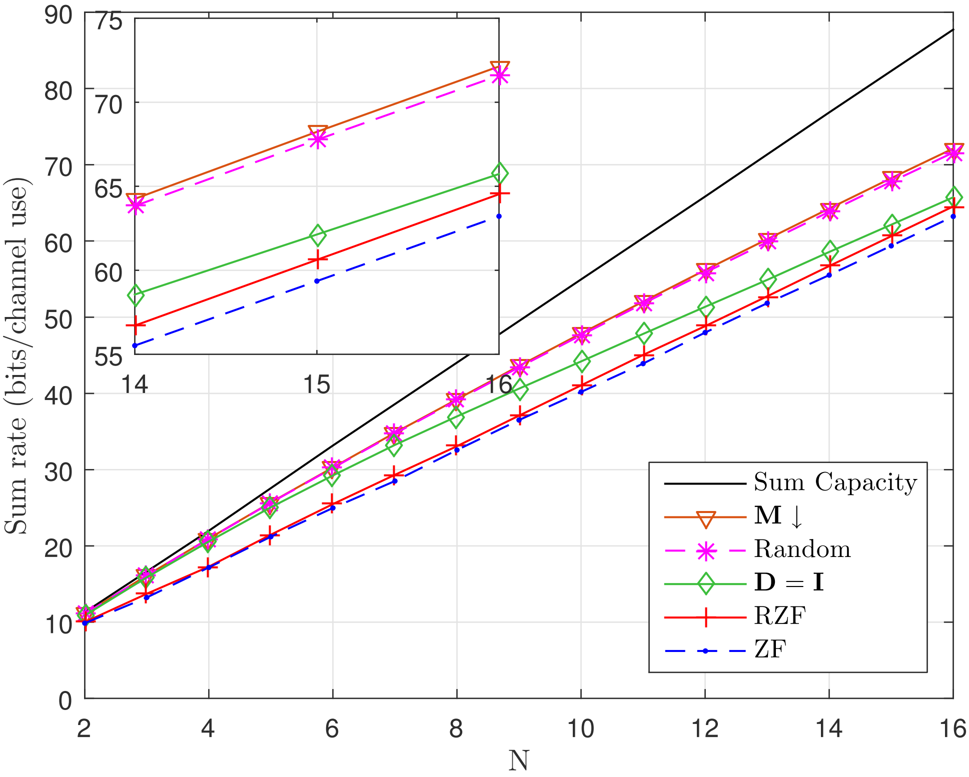


Figure 3.3: Sum rate for SNR = 20 dB. For each method and each value of N, K was chosen to maximize the sum rate. On the box, a close up on the range of N from 14 to 16.

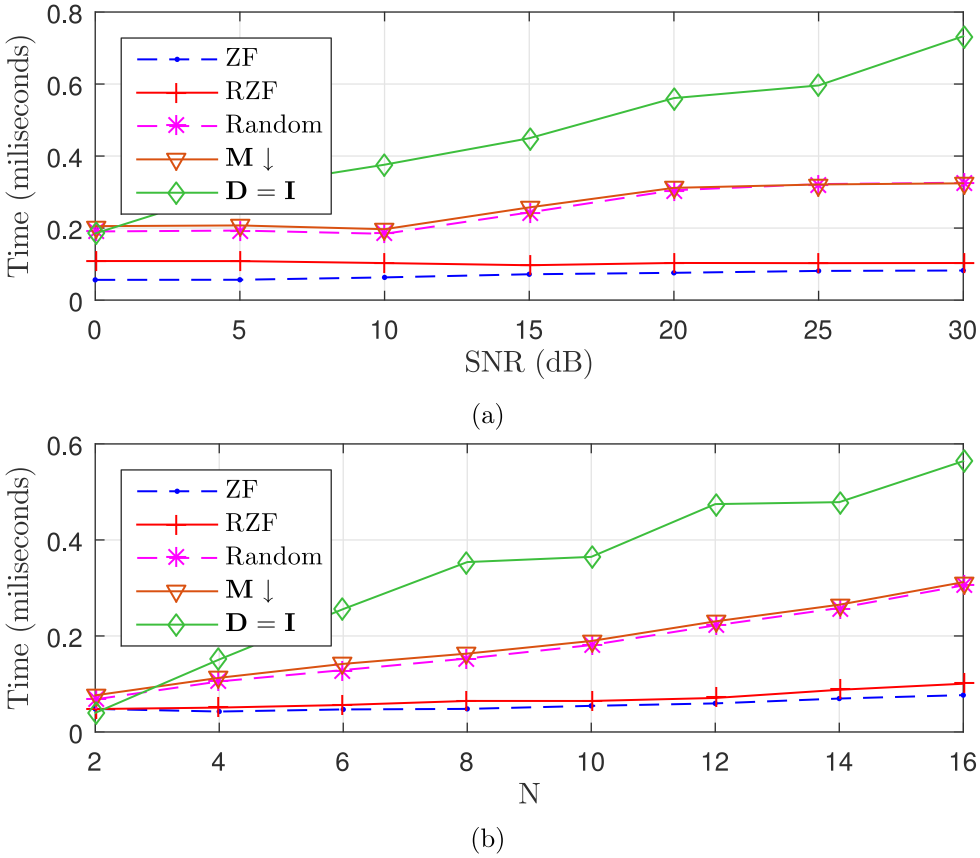
($\mathbf{D} = \mathbf{I}$) for all values of SNR. In particular, for $N = 16$ and for a sum rate of 105 bits/channel use, it outperforms the latter by about 2.1 dB and the former by about 3.2 dB. However, for $N = 64$ transmit antennas and for a sum rate of 330 bits/channel use, the difference is only about 0.4 dB.

Fig. 3.3 shows the performance for a fixed SNR = 20 dB while varying the number of transmit antennas N. The number of users is again chosen in order to maximize the sum rate and it is shown in Table 3.3. Note that, although the gap to capacity increases with K, the difference in performance between our proposed method and the other methods considered also increases.

Fig 3.4 shows the average time for the simulations of Fig. 3.1 and Fig. 3.3. In both situations, we can see that the proposed method is 2 to 3 times slower than conventional linear methods. We can also see that the average time of IF methods (the proposed one and $\mathbf{D} = \mathbf{I}$) increases with SNR (and N) due to the LLL algorithm. However, since the LLL algorithm is less complex for our proposed method, its simulation time is much smaller than that of $\mathbf{D} = \mathbf{I}$ in these scenarios.

Table 3.3: Number of users K, for each method and for each value of N, used in Fig. 3.3.

Method	$N = 2$	$N = 4$	$N = 6$	$N = 8$	$N = 10$	$N = 12$	$N = 14$	$N = 16$
$\mathbf{M}\downarrow$	2	4	6	8	10	12	14	16
Random	2	4	6	8	10	12	14	16
$\mathbf{D} = \mathbf{I}$	2	4	6	8	10	11	13	14
RZF	2	4	5	7	8	10	12	13
ZF	2	4	5	7	8	10	11	13



(a)

(b)

Figure 3.4: Average simulation time for each method. a Parameters as in Fig. 3.1. b Parameters as in Fig. 3.3.

Table 3.4: Sum rate for $N = K = 4$ in bits/channel use.

	SNR (dB)			
Method	0	10	20	30
Sum capacity	3.585	10.992	22.071	34.796
Exhaustive search [9]	3.108	9.970	21.556	34.380
Proposed method ($\mathbf{M}\downarrow$)	3.083	9.884	20.880	33.566
Gap from $\mathbf{M}\downarrow$ to [9]	0.025	0.086	0.676	0.814

Finally, it is worth mentioning that the proposed method for RDIF optimization is indeed suboptimal. As can be seen in Table 3.4, for $N = K = 4$, a small but non-negligible gap exists between the performance of our method and that of the exhaustive search carried out in [9] (which has exponential complexity). Whether this gap can be closed under low complexity is a challenging problem for future work.

Chapter 4

Adaptive Modulation

In adaptive modulation (AM) scenarios, the transmitters selects, for each user, a constellation with the highest spectral efficiency such that the bit-error probability is less than or equal to a pre-determined value.

Based on works of [24, 12] we first show how AM can be applied with linear precoding. Note that, we choose to calculate the exact bit-error probability instead of using an approximation as [12]. While this choice increases the overall complexity of the schemes, we can make better comparison with non-linear schemes. After that, we explain how AM can be used with LRA and IF precoding in a uncoded scenario.

In this chapter, since there are no codes, we consider $n = 1$.

4.1 Adaptive Modulation for Linear Precoding

Recall the notation in Section 2.3, where $\mathbf{w}_i$ is the message desired by the ith user which is modulated into $\mathbf{x}_i \in \mathbb{C}$. Independently of the modulation chosen, we always assume that it has an average symbol energy of SNR, i.e., $\mathbb{E}[|\mathbf{x}_i|^2] \leq \text{SNR}$, for all i. From (2.15), we have that the signal received by the ith user is

$$\begin{aligned}\mathbf{y}_{\text{eff},i} &= \frac{1}{h'_{ii}}\mathbf{y}_i \\ &= \mathbf{x}_i + \mathbf{z}_{\text{eff},i} \end{aligned} \tag{4.1}$$

where $\mathbf{z}_{\text{eff},i} = \frac{1}{h'_{ii}}\left(\sum_{j\neq i} h'_{ij}\mathbf{x}_j + \mathbf{z}_i\right)$ is the effective noise.

Note that (4.1) is an additive noise channel. We assume that the effective noise $\mathbf{z}_{\text{eff},i}$ follows a Gaussian distribution with variance

$$\sigma^2_{\text{eff},i} = \frac{1}{|h'_{ii}|^2}\left(\sum_{j\neq i} |h'_{ij}|^2\,\text{SNR} + 1\right). \tag{4.2}$$

However, this is only true if the ZF precoding matrix is used.

Let $\mathfrak{M}$ be the set of possible cardinalities of M-QAM constellations, where $M \in \mathfrak{M}$ is a power of 2. For a M-QAM constellation, $M \in \mathfrak{M}$, we denote by d_M the minimum distance between symbols and by $P_b^{M\text{-QAM}}(d_M, \sigma^2)$ the bit-error probability, where σ^2 is the variance of the additive noise. In an AM scheme, the transmitter selects a precoding matrix $\mathbf{T} \in \mathbb{C}^{K\times N}$

Table 4.1: Exemple of AM using ZF precoding in a 2 × 2 channel.

	$\sigma^2_{\text{eff},i}$	$P_b^{\text{4-QAM}}$	$P_b^{\text{16-QAM}}$	$P_b^{\text{64-QAM}}$	$P_b^{\text{256-QAM}}$	$P_b^{\text{1024-QAM}}$
User 1	5.0318	$< 10^{-6}$	$1.4 \cdot 10^{-4}$	$2.4 \cdot 10^{-2}$	—	—
User 2	0.9937	$< 10^{-6}$	$< 10^{-6}$	$2.9 \cdot 10^{-5}$	$1.2 \cdot 10^{-2}$	—

and cardinalities $M_1, \dots, M_K \in \mathfrak{M}$, such that

$$\{\mathbf{T}, M_1, \dots, M_K\} = \underset{\substack{\mathbf{T}\in\mathbb{C}^{K\times N}:\mathrm{Tr}(\mathbf{T}\mathbf{T}^{\mathrm{H}})=1 \\ M_1',\dots,M_K'\in\mathfrak{M} \\ P_b^{M_i'\text{-QAM}}(d_{M_i'},\sigma^2_{\text{eff},i})\leq \mathrm{BER}_{\text{target}} \\ i=1,\dots,K}}{\arg\max} \sum_{i=1}^{K} \log_2 M_i' \tag{4.3}$$

where M_i is the cardinality chosen for the ith user, $i = 1, \dots, K$ and $\mathrm{BER}_{\text{target}} > 0$. Note that the noise variance depends on $\mathbf{T}$.

In order to simplify the optimization problem (4.3), we can use the results of achievable rates to select matrix $\mathbf{T}$. More precisely, in a ZF-precoding-AM scheme, $\mathbf{T}$ is chosen as (2.19), and in a RZF-precoding-AM scheme, $\mathbf{T}$ is chosen as (2.24). With this simplification, $\sigma^2_{\text{eff},i}$ is now a given value and (4.3) can solved individually for each user.

Lastly, we need to define the values of d_M for each possible constellation. Since we require that the average energy per symbol is equal to SNR, the minimum distance can be easily obtained as

$$d_M = \begin{cases} 2\sqrt{\mathrm{SNR}} & \text{if } M = 2 \\ \sqrt{\frac{6}{M-1}\mathrm{SNR}} & \text{if } M > 2 \text{ and } \log_2 M \text{ is even} \\ \sqrt{\frac{12}{I^2+J^2-2}\mathrm{SNR}} & \text{if } M > 2 \text{ and } \log_2 M \text{ is odd} \end{cases} \tag{4.4}$$

where I and J are the number of symbols in-phase and quadrature of M-QAM constellation, respectively. The closed-form expressions of $P_b^{M\text{-QAM}}(d_M, \sigma^2)$ are found in A.1.

Example. Consider a 2 × 2 channel, with SNR = 25 dB, $\mathrm{BER}_{\text{target}} = 10^{-3}$ and

$$\mathbf{H} = \begin{bmatrix} 1 & \jmath 1 \\ -\jmath 3 & 1 \end{bmatrix}.$$

Assume that the transmitter uses a ZF precoding schemes. From (2.19), the precoding matrix is given by

$$\mathbf{T} = \begin{bmatrix} -\frac{1}{2} & \jmath\frac{1}{2} \\ -\jmath\frac{3}{2} & -\frac{1}{2} \end{bmatrix} \operatorname{diag}\left(\sqrt{\mathbf{p}}\right)$$

where $\mathbf{p} = \begin{bmatrix} 0.1987 & 1.0063 \end{bmatrix}^{\mathrm{T}}$ is obtained via water-filling.

Suppose that $\mathfrak{M} = \{4, 16, 64, 256, 1024\}$. Table 4.1 contains the values of $\sigma^2_{\text{eff},i}$, $i = 1, 2$ and the bit-error probability for M-QAM constellation, $M \in \mathfrak{M}$. We can clearly see that the transmitter selects a 16-QAM constellation for user 1 and a 64-QAM constellation for user 2, which results in a sum rate of 10 bits/channel use. □

4.2 Adaptive Modulation IF/LRA Schemes

Recall definitions from Section 2.6. Let $\Lambda = d\mathbb{Z}[j]$ and $\Lambda_s = q\Lambda$ be lattices, where q is a power of 2 and d is the minimum distance between symbols of a constellation $\mathcal{S} = \Lambda \cap [0, dq)^2$. Note that $[0, dq)^2$ is a fundamental region of Λ_s. We assume that $\mathcal{S}$ has an average energy equals to SNR. Note that the constellation $\mathcal{S}$ is similar to a q^2-QAM constellation, up to a shift by $\mathbf{u}$ to ensure zero mean.

Let $\mathcal{W} = \mathbb{Z}_q[j]$ be the ambient space and $\mathcal{W}_i \subseteq \mathcal{W}$ be the message space to user i, $i = 1, \ldots, K$. Let $\varphi : \Lambda \to \mathcal{W}$ be a $\mathbb{Z}[j]$-linear map and let $\tilde{\varphi} : \mathcal{W} \to \mathcal{S}$ be a bijective mapping function such that $\varphi(\tilde{\varphi}(\mathbf{w})) = \mathbf{w}$ for all $\mathbf{w} \in \mathcal{W}$.

The transmitter selects a message $\mathbf{w}_i \in \mathcal{W}_i$ to be transmitted to the ith user. Let $\mathbf{W} = \begin{bmatrix}\mathbf{w}_1^{\mathrm{T}} & \cdots & \mathbf{w}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. After an integer matrix $\mathbf{A} \in \mathbb{Z}[j]^{K\times K}$ is chosen, the transmitter computes

$$\mathbf{W}' = \mathbf{A}'\mathbf{W} \bmod q \tag{4.5}$$

where $\mathbf{A}' \in \mathbb{Z}[j]^{K\times K}$ is an integer matrix such that $\mathbf{A}\mathbf{A}' = \mathbf{I}$. Note that $\mathbf{W}' \in \mathbb{Z}_q[j]^{K\times n}$.

After that, the mapping $\tilde{\varphi}$ is applied in each row of $\mathbf{W}'$, generating signal $\mathbf{S}' = \tilde{\varphi}(\mathbf{W}') \in \mathcal{S}$. The transmitter computes

$$\mathbf{X} = \mathbf{S}' + \mathbf{U} \bmod \Lambda_s \tag{4.6}$$

where $\mathbf{U} = \begin{bmatrix}\mathbf{u}^{\mathrm{T}} & \cdots & \mathbf{u}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. Note that, $\mathbf{x}_i$ has zero mean and energy equals to SNR, where $\mathbf{x}_i$ is the ith row of $\mathbf{X}$, $i = 1, \ldots, K$. Then, the signal $\mathbf{X}$ is multiplied by the precoding matrix $\mathbf{T}$ producing the transmitted signal $\mathbf{X}'$.

Recall from (2.52) that the ith user receives signal

$$\mathbf{y}_i = \mathbf{h}_i'\mathbf{X} + \mathbf{z}_i \tag{4.7}$$

where $\mathbf{z}_i$ is the Gaussian noise with unit variance and $\mathbf{h}_i' \triangleq \mathbf{h}_i\mathbf{T}$. After a multiplication by a scalar $\alpha_i \in \mathbb{C}$, the ith user computes

$$\begin{aligned}\mathbf{y}_{\text{eff},i} &= \alpha_i\mathbf{y}_i - \mathbf{a}_i\mathbf{U} \bmod \Lambda_s && (4.8)\\ &= \mathbf{a}_i(\mathbf{X} - \mathbf{U}) + \mathbf{z}_{\text{eff},i} \bmod \Lambda_s && (4.9)\\ &= \mathbf{s}_i + \mathbf{z}_{\text{eff},i} \bmod \Lambda_s && (4.10)\end{aligned}$$

where $\mathbf{s}_i = \mathbf{a}_i\mathbf{S}' \bmod \Lambda_s$ and $\mathbf{z}_{\text{eff},i} = (\alpha_i\mathbf{h}_i' - \mathbf{a}_i)\mathbf{X} + \alpha_i\mathbf{z}_i$. Although the distribution of $\mathbf{z}_{\text{eff},i}$ is hard to estimate, we will assume that it follows a Gaussian distribution with zero mean and variance $\sigma^2_{\text{eff},i}$. Recall from (2.64) that $\sigma^2_{\text{eff},i}$ is given by (using optimal α)

$$\sigma^2_{\text{eff},i} = \text{SNR}\left(\|\mathbf{a}_i\|^2 - \frac{1}{\|\mathbf{h}_i'\|^2 + \text{SNR}^{-1}}\left|\mathbf{a}_i\mathbf{h}_i'^{\mathrm{H}}\right|^2\right). \tag{4.11}$$

Since there is a one-to-one mapping between $\mathcal{W}$ and $\mathcal{S}$ we have that $\varphi(\mathbf{s}_i) = \varphi(\mathbf{a}_i\mathbf{A}'\mathbf{W} \bmod \Lambda_s) = \mathbf{w}_i$. This mean that, from the point of view for the receiver, it can pretend that symbol $\mathbf{s}_i = \tilde{\varphi}(\mathbf{w}_i) \in \mathcal{S}_i$ was sent over a modulo-lattice additive noise (MLAN) channel given by (4.10),

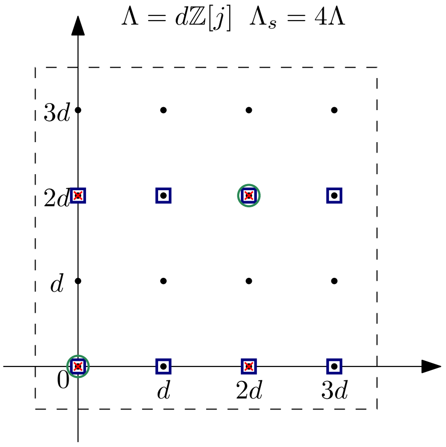


Figure 4.1: Example of possible QAM sub-constellation when $\Lambda = d\mathbb{Z}[j]$ and $\Lambda_s = 4\Lambda$.

where $\mathcal{S}_i \subseteq \mathcal{S}$ is a sub-constellation such that $\mathcal{S}_i = \tilde{\varphi}(\mathcal{W}_i)$.

Let $\mathcal{S}' \subseteq \mathcal{S}$ be a sub-constellation. We define its cardinality by $|\mathcal{S}'|$ and its bit-error probability by $P_b^{\mathcal{S}'}(\sigma^2, \mathcal{S})$, where σ^2 is the variance of additive noise. The transmitter wants to select sub-constellations $\mathcal{S}_1, \ldots, \mathcal{S}_K \subseteq \mathcal{S}$, a precoding matrix $\mathbf{T} \in \mathbb{C}^{K\times N}$, a Gaussian integer matrix $\mathbf{A} \in \mathbb{Z}^{K\times K}$ and q, which is a power of 2, such that

$$\{\mathbf{T}, \mathbf{A}, q, \mathcal{S}_1, \ldots, \mathcal{S}_K\} = \underset{\substack{\mathbf{T}\in\mathbb{C}^{K\times N}:\mathrm{Tr}(\mathbf{T}\mathbf{T}^{\mathrm{H}})=1 \\ \mathbf{A}\in\mathbb{Z}[j]^{K\times K}:\mathrm{rank}(\mathbf{A})=K \\ q\in\mathbb{N}:\log_2 q\in\mathbb{N} \\ \mathcal{S}'_1,\ldots,\mathcal{S}'_K\subseteq\mathcal{S} \\ P_b^{\mathcal{S}'_i}(\sigma^2_{\mathrm{eff},i},\mathcal{S})\leq\mathrm{BER}_{\mathrm{target}} \\ i=1,\ldots,K}}{\arg\max} \quad R_{\mathrm{sum}} = \sum_{i=1}^{K} \log_2 |\mathcal{S}'_i| \tag{4.12}$$

where $\mathrm{BER}_{\mathrm{target}}$ is the maximum BER allowed by the system. Note that, the effective noise variance depends on $\mathbf{T}$ and $\mathbf{A}$, and the constellation $\mathcal{S}$ depends on q. Also note that the minimum distance d can be computed as (4.4) by replacing M with q^2.

Problem (4.12) can be very hard to solve. In order to simplify this problem, we can use the results of achievable rate to determine matrices $\mathbf{T}$ and $\mathbf{A}$. For example, we can use our proposed RDIF design to select those matrices. Moreover, we consider that q is given.

Nevertheless, this optimization problem can still be very hard to solve since, for a general choice $\mathcal{S}' \subseteq \mathcal{S}$ there are no closed-form expression neither approximations for $P_b^{\mathcal{S}'}(\sigma^2_{\mathrm{eff},i}, \mathcal{S})$. In order to circumvent this problem, we limit the set of possible choices of $\mathcal{S}'$.

We denote by M-QAM$(\mathcal{S}) \in \mathcal{S}$ the sub-constellation with a cardinality M such that the constellation points resemble a QAM constellation. More precisely, when $\log_2 M$ is an even number, the constellation points are the traditional square M-QAM constellation, and when $\log_2 M$ is an odd number, the constellation points is a rectangular M-QAM constellation. Let $\mathfrak{M}$ be the set of possible cardinalities of M-QAM$(\mathcal{S})$, i.e., if $M \in \mathfrak{M}$ then M-QAM$(\mathcal{S}) \in \mathcal{S}$. For example, suppose that $\Lambda = d\mathbb{Z}[j]$ and $\Lambda_S = 4\Lambda$. It is easy to note that $\mathfrak{M} = \{2, 4, 8, 16\}$. In Fig. 4.1 , we show the constellation points for each sub-constellation M-QAM$(\mathcal{S})$. The black dots represent a 16-QAM constellation, the blue squares represents a 8-QAM constellation, the red crosses a 4-QAM and finally the green circles represents a 2-QAM (BPSK) constellation.

Table 4.2: Exemple of AM using LRA precoding.

	$\sigma^2_{\text{eff},i}$	$P_b^{4\text{-QAM}(\mathcal{S})}$	$P_b^{16\text{-QAM}(\mathcal{S})}$	$P_b^{64\text{-QAM}(\mathcal{S})}$	$P_b^{256\text{-QAM}(\mathcal{S})}$	$P_b^{1024\text{-QAM}(\mathcal{S})}$
User 1	0.9968	$< 10^{-6}$	$< 10^{-6}$	$3.8 \cdot 10^{-5}$	$1.3 \cdot 10^{-2}$	—
User 2	0.9846	$< 10^{-6}$	$< 10^{-6}$	$3.4 \cdot 10^{-5}$	$1.3 \cdot 10^{-2}$	—

Note that we can rewrite (4.12) as

$$M_i = \underset{\substack{M' \in \mathfrak{M} \\ P_b^{M\text{-QAM}(\mathcal{S})}(\sigma^2_{\text{eff},i}, \mathcal{S}) \leq \text{BER}_{\text{target}}}}{\arg\max} \log_2 M' \tag{4.13}$$

where $M_i \in \mathfrak{M}$ is the cardinality of the constellation for ith user, $i = 1, \ldots, K$, $P_b^{M\text{-QAM}(\mathcal{S})}(\sigma^2, \mathcal{S})$ is the bit-error probability of the sub-constellation M-QAM$(\mathcal{S})$ and σ^2 is the variance of the additive noise.

While this approach may limit the performance of AM for IF/LRA precoding, it allows us to find closed-form expression for the bit-error probability since it becomes similar to the expressions of the bit-error probability for traditional QAM constellations.

Since a M-QAM constellation can be interpreted as a I-PAM constellation in in-phase and a J-PAM constellation in quadrature, we can compute the bit-error probability as

$$P_b^{M\text{-QAM}(\mathcal{S})}(\sigma^2, \mathcal{S}) = \frac{1}{\log_2 M}\left(\log_2 I P_b^{I\text{-PAM}}(d^I, \sigma^2) + \log_2 J P_b^{J\text{-PAM}}(d^Q, \sigma^2)\right) \tag{4.14}$$

where $P_b^{I\text{-PAM}}$ is the bit-error probability of a lattice I-PAM constellation and d^I and d^Q are the minimum distance between symbols of the in-phase and quadrature components, respectively. The expression of $P_b^{I\text{-PAM}}$ are found in Appendix A.2.

Note that, d^I and d^Q depend on the sub-constellation chosen. For example, in Fig. 4.1, the 16-QAM constellation has $d^I = d^Q = d$. The 8-QAM constellation has $d^I = d$ and $d^Q = 2d$, the 4-QAM constellation has $d^I = d^Q = 2d$. And finally the 2-QAM constellation, which is a special case, has $d^I = 2\sqrt{2}d$.

In general, if M is an even power of 2 then let $m = \sqrt{M}$, the minimum distance of a M-QAM$(\mathcal{S})$ constellation is $d^I = d^Q = \frac{q}{m}d$. If M is an odd power of 2 then let $m^{\text{I}} = \sqrt{2M}$ and $m^{\text{Q}} = \sqrt{\frac{M}{2}}$, which are the number of symbols of in-phase and number of symbols in quadrature, respectively. In this case, a M-QAM$(\mathcal{S})$ constellation has minimum distance of $d^{\text{I}} = \frac{q}{m^{\text{I}}}d$ and $d^{\text{Q}} = \frac{q}{m^{\text{Q}}}d$.

Example. Consider again the same channel as before, with SNR = 25 dB, $\text{BER}_{\text{target}} = 10^{-3}$. Suppose that we choose $\Lambda = d\mathbb{Z}[j]$ and $\Lambda_s = q\Lambda$, where $q = 32$ and $d = \sqrt{\frac{6}{q^2-1}\text{SNR}} = 1.362$. Finally, let $\mathfrak{M} = \{4, 16, 64, 256, 1024\}$ be the set of possible cardinalities of M-QAM constellation, where $M \in \mathfrak{M}$.

First, suppose that LRA precoding is used, where $\mathbf{T} = c\mathbf{H}^{\text{H}}\left(\mathbf{H}\mathbf{H}^{\text{H}} + \frac{K}{\text{SNR}}\mathbf{I}\right)^{-1}\mathbf{A}$, c is a constant and $\mathbf{A}$ is obtained via the LLL-reduction. In Table 4.2 we show the bit error probability of each possible constellation for all users. Note that, both users select a 64-QAM constellation, which give a sum rate of 12 bits/channel use.

Now, suppose that RDIF scheme is used, where $\mathbf{T} = c\mathbf{H}^{\text{H}}\left(\mathbf{H}\mathbf{H}^{\text{H}} + \frac{K}{\text{SNR}}\mathbf{I}\right)^{-1}\mathbf{DA}$, c is a constant and $\mathbf{D}$ and $\mathbf{A}$ are obtained using our proposed method. In Table 4.3 we show the bit

Table 4.3: Exemple of AM using IF/proposed scheme.

	$\sigma^2_{\text{eff},i}$	$P_b^{\text{4-QAM}(\mathcal{S})}$	$P_b^{\text{16-QAM}(\mathcal{S})}$	$P_b^{\text{64-QAM}(\mathcal{S})}$	$P_b^{\text{256-QAM}(\mathcal{S})}$	$P_b^{\text{1024-QAM}(\mathcal{S})}$
User 1	4.9414	$< 10^{-6}$	$2.6 \cdot 10^{-4}$	$2.7 \cdot 10^{-2}$	—	—
User 2	0.1987	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$3.8 \cdot 10^{-6}$	$6.1 \cdot 10^{-3}$

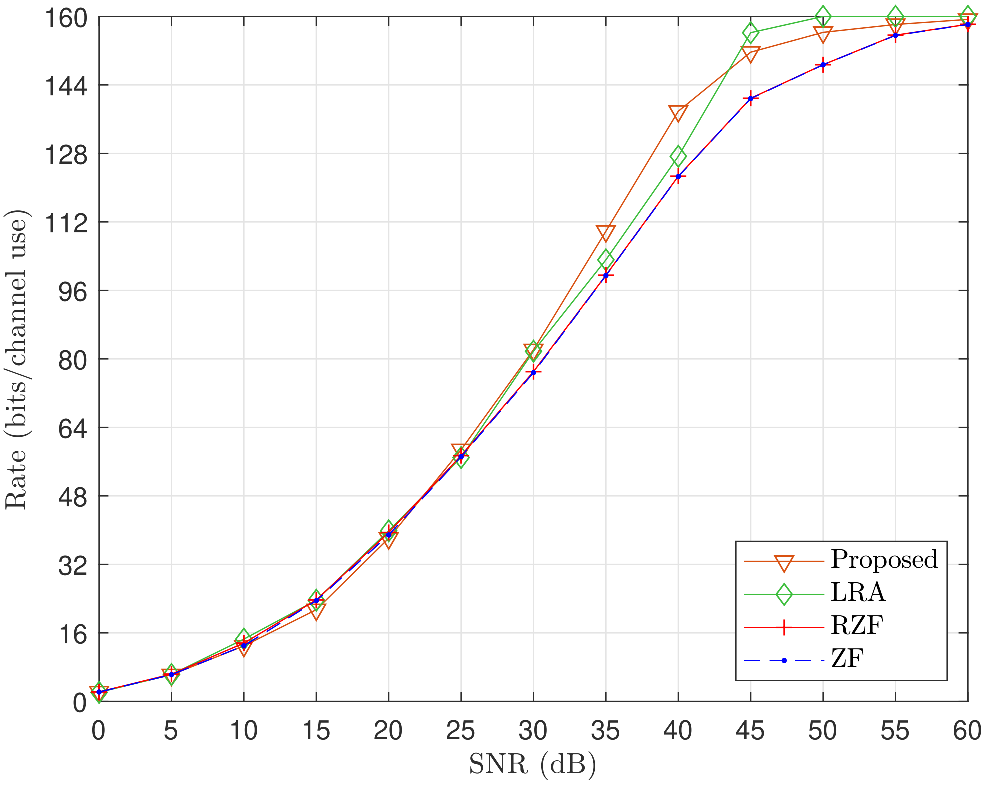


Figure 4.2: Sum of spectral efficiency using adaptive modulation with $N = 16$.

error probability for each possible constellations for all users. Note that, is this case, the user 1 selects a 16-QAM constellation and the user 2 selects a 256-QAM constellation, which give the same sum rate of 12 bits/channel use as the LRA precoding. □

4.3 Numerical Results

In this section we show the average sum of the spectral efficiency of the proposed method. In our simulations, the sum of spectral efficiency were obtained through 10000 channel realizations. In each realization, the channel coefficients were randomly obtained considering a circularly symmetric complex Gaussian distribution with zero mean and unit variance. In all simulation we consider a $\text{BER}_{\text{target}} = 10^{-3}$.

The set of possible modulations contains M-QAM constellation, where $M \in \{2, 4, 8, \dots, 1024\}$. For IF and LRA precoding, $q \in \{2, 4, 8, 16, 32\}$ was chosen such that the sum rate is maximized.

In each simulation, we compare our proposed RDIF design to the conventional linear precoding methods, namely, ZF and RZF, as well as the LRA design, which is obtained by chosen $\mathbf{D} = \mathbf{I}$. For our proposed method, we use only the heuristic where the permutation sorts the diagonal elements of $\mathbf{M}$ in descending order.

Fig. 4.2 shows the performance for $N = 16$ transmit antennas. For each method and for

Table 4.4: Number of users K, for each method and for some values of SNR, used in Fig. 4.2.

Method	0 dB	10 dB	20 dB	30 dB	40 dB	50 dB	60 dB
Proposed	2	8	12	16	16	16	16
LRA	2	8	11	14	16	16	16
RZF	2	8	11	13	14	15	16
ZF	2	7	11	13	14	15	16

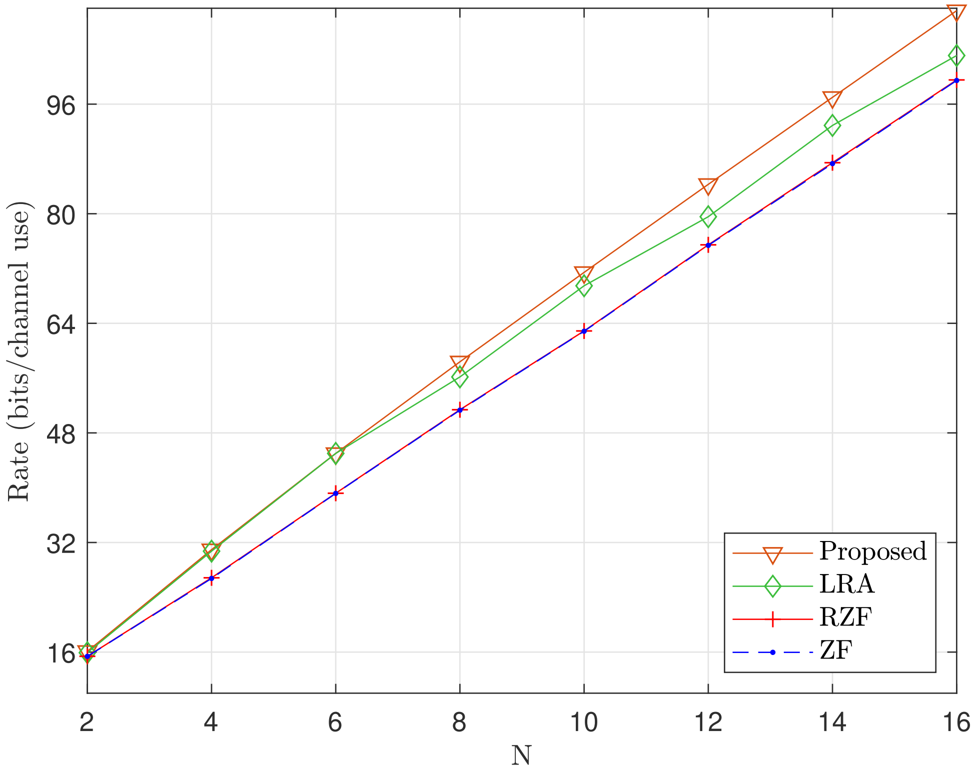


Figure 4.3: Sum of spectral efficiency using adaptive modulation varying the number of transmitter antennas and SNR = 35 dB.

each value of SNR, we choose, through exhaustive search, the value of $K \leq N$ that achieves the highest sum rate and it is shown in Table **??**. Note that for medium values of SNR, our proposed method can achieve higher spectral efficiency than other methods. For example, for a sum of spectral efficiency of 112 bits/channel use, the gap between our method and LRA is about 1.5 dB, and between our method in linear schemes is about 2 dB. In low SNR the performance of our method is degraded since in our design we use a high SNR approximation. Finally, in high SNR, we can note that the LRA schemes outperforms our method. This happens due to the limited set of constellation. For LRA schemes, we expected that most of users selects the same constellation, differently of our schemes where some users selects constellations with higher spectral efficiency while others select lower spectral efficiency constellations. This means that, in high SNR some users can selects constellation with spectral efficiency greater than 10, which is the highest in our set.

Fig. 4.3 shows the performance for a fixed SNR = 35 dB while varying the number of transmit antennas N. Again, we choose $K \leq N$ such that the sum rate is maximized. The number of users K is shown in Table 4.5. Since we consider a medium value of SNR, we can see that the proposed method outperforms both linear and LRA schemes. We can also note

Table 4.5: Number of users K, for each method and for each value of N, used in Fig. 4.3.

Method	$N = 2$	$N = 4$	$N = 6$	$N = 8$	$N = 10$	$N = 12$	$N = 14$	$N = 16$
Proposed	2	4	6	8	10	12	14	16
LRA	2	4	6	8	9	11	12	13
RZF	2	4	5	7	9	10	12	13
ZF	2	4	5	7	8	10	12	13

that gap between method is constant when N varies for the simulated scenarios.

Chapter 5

Conclusion

In this thesis, we first proposed a desing for RDIF method for $K \geq 2$ users and then we combine the proposed scheme with adaptive modulation. In order to obtain a precoding matrix $\mathbf{T}$ and a Gaussian integer matrix $\mathbf{A}$ for RDIF method, we establish a structure for matrix $\mathbf{A}$ and then we solve a relaxed optimization problem, where $\mathbf{A}$ could be a complex matrix. This optimization problem can be solved with $\mathcal{O}(K^3)$ operations. Moreover, the solution for this problem allow us to use a lattice basis reduction with complexity of $\mathcal{O}(K^3)$ since the basis are in unitriangular form. This approach leads to an overall complexity of $\mathcal{O}(NK^2)$, which is the same as linear precoding methods. Simulation results show that our approach not only significantly outperforms conventional linear precoding, but also improves on previous low-complexity IF precoding both in performance and complexity.

To combine the proposed scheme with adaptive modulation, we first needed to ensure that each possible modulation can be used with the chosen shaping lattice, which leads to limit the set of constellation as QAM. By using matrices $\mathbf{T}$ and $\mathbf{A}$ found in the information theoretical approach as well as limiting the set of constellations, the transmitter can select a constellation for each user such that the bit-error probability is less than a BER target. We also have found expressions for the bit-error probability for lattice modulations. Simulation results show that for a medium range for SNR the proposed scheme outperforms both LRA and linear precoding methods.

5.1 Suggestions for Future Works

Some suggestions for future works are:

- Consider the adaptive modulations schemes with codes. One difficulty with this approach is to estimate the bit-error probability in this scenario.
- Consider imperfect channel estimation. Throughout this thesis we consider that CSI in available for the transmitter, however, in cases where the channel changes quickly, it can be hard to have a perfect estimation of the channel.

Appendix A

Bit-Error Probability of M-QAM Constellations

A.1 AWGN Channels

Consider an AWGN channel $y = x + z$, where z is the Gaussian noise with variance of $N_0/2$ per dimension, and x is the transmitted symbol of a M-QAM constellation, where $M = 2^k$ is the cardinality, $k \geq 1$, with a minimum distance between symbols equals to d.

Note that if $M = 2$ then we have the traditional BPSK constellation and if k is even we have a traditional square M-QAM constellation, which can be interpreted as $\sqrt{M}$-PAM constellation in-phase and quadrature. If $k > 1$ is odd, let $M = I \cdot J$, where $I = 2^{\frac{k+1}{2}}$ and $J = 2^{\frac{k-1}{2}}$, then a M-QAM constellation consists of a I-PAM constellation in-phase and a J-PAM constellation in quadrature. We assume that the modulation uses a Gray code mapping. The bit-error probability P_b of a M-QAM is given by [41]

$$P_b^{\text{BPSK}}(d, N_0) = \mathrm{Q}\left(\frac{d/2}{\sqrt{N_0/2}}\right) \tag{A.1}$$

for 2-QAM (BPSK) constellation,

$$\begin{aligned} P_b^{M\text{-QAM}}(d, N_0) = \frac{4}{\sqrt{M}\log_2 M} \sum_{\ell=1}^{\log_2 \sqrt{M}} \sum_{i=0}^{(1-2^{-\ell})\sqrt{M}-1} (-1)^{\left\lfloor \frac{i2^{\ell-1}}{\sqrt{M}} \right\rfloor} \cdot \\ \left(2^{\ell-1} - \left\lfloor \frac{i2^{\ell-1}}{\sqrt{M}} + \frac{1}{2} \right\rfloor\right) \mathrm{Q}\left((2i+1)\frac{d/2}{\sqrt{N_0/2}}\right) \end{aligned} \tag{A.2}$$

for M-QAM constellation if $k = \log_2 M$ is even, and

$$P_b^{M\text{-QAM}}(d, N_0) = \frac{2}{\log_2 M}\left(\frac{1}{I}\sum_{\ell=1}^{\log_2 I} P_I(\ell) + \frac{1}{J}\sum_{\ell=1}^{\log_2 J} P_J(\ell)\right) \tag{A.3}$$

where

$$P_I(\ell) = \sum_{i=0}^{(1-2^{-\ell})I-1} (-1)^{\left\lfloor \frac{i2^{\ell-1}}{I} \right\rfloor} \left(2^{\ell-1} - \left\lfloor \frac{i2^{\ell-1}}{I} + \frac{1}{2} \right\rfloor \right) \mathrm{Q}\left((2i+1)\frac{d/2}{\sqrt{N_0/2}} \right)$$
$$P_J(\ell) = \sum_{j=0}^{(1-2^{-\ell})J-1} (-1)^{\left\lfloor \frac{j2^{\ell-1}}{J} \right\rfloor} \left(2^{\ell-1} - \left\lfloor \frac{j2^{\ell-1}}{J} + \frac{1}{2} \right\rfloor \right) \mathrm{Q}\left((2j+1)\frac{d/2}{\sqrt{N_0/2}} \right)$$

for M-QAM constellation if $k = \log_2 M$ is odd, where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{u^2}{2}\right) \, \mathrm{d}u$ is the Q-function.

A.2 MLAN channel

Let $\Lambda = d\mathbb{Z}$ be a lattice and $\Lambda_s = q\Lambda$, where q is a power of 2 and d is the minimum distance between symbol of Λ. Consider a MLAN channel $y = x + z \bmod \Lambda_s$, where z is a Gaussian noise with zero mean and variance $N_0/2$ per dimension and $x \in \mathcal{M}$, $\mathcal{M} = \Lambda \cap \texttt{Cube}(\Lambda_s)$ is the constellation points with cardinality M. Note that $\mathcal{M}$ resembles a M-PAM constellation, except for a shift to ensure zero mean. We assume that a Gray code mapping is used. Recall that, due to modulo-lattice operator, the receiver sees as the constellation $\mathcal{M}$ is periodically extended over the real field.

Let b_k be the kth bit of a symbol $x_\ell \in \mathcal{M}$, $k = 1, \ldots, \log_2 M$ and $\ell = 1, \ldots, M$. Let $\hat{b}_k$ be the estimated kth bit of $\hat{x}$. We say that an error e_k occurs if $b_k \neq \hat{b}_k$ for any k.

The bit-error probability can be calculated as

$$P_b^{M\text{-PAM}}(d, N_0) = \frac{1}{\log_2 M} \sum_{k=1}^{\log_2 M} P_b(e_k) \tag{A.4}$$

where $P_b(e_k)$ is the probability of an error occurs in the kth bit[1], $k = 1, \ldots, \log_2 M$. Since we assume that the symbols are equiprobable, we have that

$$P_b(e_k) = \frac{1}{M} \sum_{\ell=1}^{M} \Pr\left[e_k | x_\ell\right]. \tag{A.5}$$

where $\Pr\left[e_k | x_\ell\right]$ is the probability of an error occurs in the kth bit if symbols x_ℓ is transmitted.

We define the *error regions* of b_k for x_ℓ as the values of z such that if x_ℓ is transmitted, we have that $\hat{b}_k \neq b_k$. Note that, for $k = 1, \ldots, \log_2 M$, if we know all the error regions of b_k for x_ℓ, $\ell = 1, \ldots, M$, then we can compute $P_b(e_k)$.

For simplicity, let

$$q(x) \triangleq Q\left(x \frac{d/2}{\sqrt{N_0/2}} \right) \tag{A.6}$$

where $Q(\cdot)$ is the Q-function.

A.2.1 2-PAM/BPSK

First, let us consider that $\mathcal{M}$ resembles a 2-PAM/BPSK constellation. In Fig. A.1, we show the

[1] We omit the dependence of d and N_0 in the expression of $P_b(e_k)$ for simplicity.

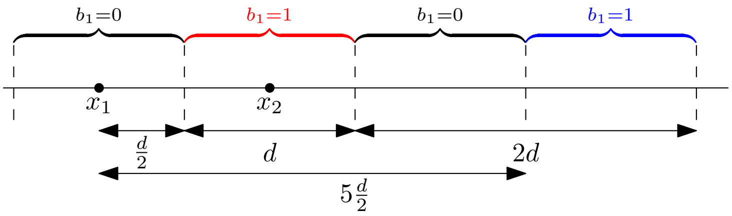


Figure A.1: A 2-PAM constellation in a MLAN channel. The dashed lines are the decision regions.

constellations points as seen by the receiver, i.e., an periodically extended 2-PAM constellation, and its decision regions.

The error regions of b_1 for x_1 can be expressed as $d/2+2id = (1+4i)d/2 < z \leq 3d/2+2id = (3+4i)d/2$, where $i \in \mathbb{Z}$. For example, if $i = 0$ we have the red region in Fig. A.1, and if $i = 1$ we have the blue region in Fig. A.1. Note that, by symmetry, the error regions of b_1 for x_2 are the same. Therefore, we have that

$$\begin{aligned}
P_b^{\text{2-PAM}}(d, N_0) &= \sum_{i=-\infty}^{\infty} \Pr\left[(1+4i)d/2 < z < (3+4i)d/2\right] \\
&= \sum_{i=-\infty}^{\infty} \Pr\left[z > (1+4i)d/2\right] - \Pr\left[z > (3+4i)d/2\right] \\
&= \sum_{i=-\infty}^{\infty} q(1+4i) - q(3+4i) \\
&= \sum_{i=0}^{\infty} q(1+4i) - q(3+4i) + \sum_{j=-\infty}^{-1} q(1+4j) - q(3+4j)
\end{aligned}$$

Now, let $i = n$ and $j = -n-1$ and note that $q(-x) = 1 - q(x)$

$$\begin{aligned}
P_b^{\text{2-PAM}}(d, N_0) &= \sum_{n=0}^{\infty} q(1+4n) - q(3+4n) + \sum_{n=\infty}^{0} q(1+4(-n-1)) - q(3+4(-n-1)) \\
&= \sum_{n=0}^{\infty} q(1+4n) - q(3+4n) + \sum_{n=0}^{\infty} q(-(3+4n)) - q(-(1+4n)) \\
&= \sum_{n=0}^{\infty} q(1+4n) - q(3+4n) + \sum_{n=0}^{\infty} (1-q(3+4n)) - (1-q(1+4n)) \\
&= \sum_{n=0}^{\infty} 2q(1+4n) - 2q(3+4n)
\end{aligned}$$

Moreover, since the argument of $q(\cdot)$ is always an odd number, we can rewrite the expression as

$$P_b^{\text{2-PAM}}(d, N_0) = \sum_{n=0}^{\infty} (-1)^n 2q(2n+1). \tag{A.7}$$

A.2.2 General M-PAM

Now we consider that $\mathcal{M}$ can resemble any general M-PAM constellation. In this case, we are going to find each $P_b(e_k)$, $k = 1, \ldots, \log_2 M$, separately.

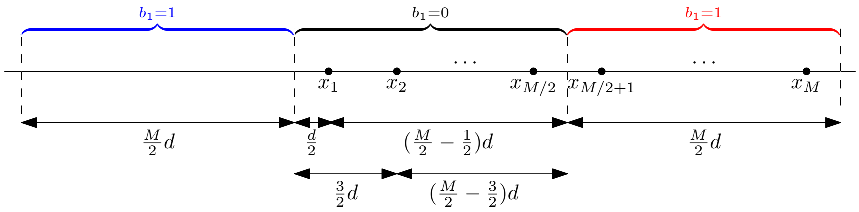


Figure A.2: Regions for the first bit b_1 of a M-PAM constellation using gray mapping.

We start by calculating the bit-error probability of the first bit b_1, i.e., $P_b(e_1)$. Fig. A.2 shows the regions where, based on the received signal, the decoder will infer $b_1 = 0$ or $b_1 = 1$. Note that, symbols $x_1, \ldots, x_{M/2}$ have $b_1 = 0$, while symbols $x_{M/2+1}, \ldots, x_M$ have $b_1 = 1$.

We can see that an error region of b_1 for x_1 can be found in $\left(\frac{M}{2}-\frac{1}{2}\right)d = (M-1)\frac{d}{2} < z \leq \left(\frac{M}{2}-\frac{1}{2}\right)d + \frac{M}{2}d = (2M-1)\frac{d}{2}$ (it corresponds to the red region in Fig. A.2). Note that, if we translate this region by iMd, where $i \in \mathbb{Z}$, other error region can be found. For example, if $i = -1$, we have $(M-1)\frac{d}{2} - Md = -(M+1)\frac{d}{2} < z < (2M-1)\frac{d}{2} - Md = -\frac{d}{2}$, which corresponds to the blue region in Fig. A.2. In general, the error regions of b_1 for x_1 can be expressed as $-(M+1)\frac{d}{2} - iMd = -(2iM+M+1)\frac{d}{2} < z < -\frac{d}{2} - iMd = -(2iM+1)\frac{d}{2}$ for any $i \in \mathbb{Z}$, and therefore

$$\begin{aligned}
\Pr[e_1|x_1] &= \sum_{i=-\infty}^{\infty} \Pr\left[-(2iM+M+1)\frac{d}{2} < z < -(2iM+1)\frac{d}{2}\right] \\
&= \sum_{i=-\infty}^{\infty} \Pr\left[z > -(2iM+M+1)\frac{d}{2}\right] - \Pr\left[z > -(2iM+1)\frac{d}{2}\right] \\
&= \sum_{i=-\infty}^{\infty} q(-(2iM+M+1)) - q(-(2iM+1)) \\
&= \sum_{i=-\infty}^{\infty} q(2iM+1) - q(2iM+M+1) \qquad \text{(A.8)}
\end{aligned}$$

since $q(-x) = 1 - q(x)$.

Note that, the error regions of b_1 for x_2 are the same as the error regions of b_1 for x_1 but translated by d. For example, an error region is found in $-(M+1)\frac{d}{2} - d = -(M+3)\frac{d}{2} < z \leq -\frac{d}{2} - d = 3\frac{d}{2}$ (which corresponds to the blue region in Fig. A.2). Again, if we translate this region by iMd, where $i \in \mathbb{Z}$, other error regions are found. In general, the error regions of b_1 for x_2 can be expressed as $-(2iM+M+3)\frac{d}{2} < z \leq -(2iM+3)\frac{d}{2}$ and therefore

$$\begin{aligned}
\Pr[e_1|x_2] &= \sum_{i=-\infty}^{\infty} q(-(2iM+M+3)) - q(-(2iM+3)) \\
&= \sum_{i=-\infty}^{\infty} q(2iM+3) - q(2iM+M+3). \qquad \text{(A.9)}
\end{aligned}$$

The error regions of b_1 for x_ℓ, $\ell = 2, \ldots, M/2$, are the same error regions of b_1 for $x_{\ell-1}$ but translate by d. For example, the error regions of b_1 for $x_{M/2}$ can be expressed as $-(2iM+M+1)\frac{d}{2} - \left(\frac{M}{2}-1\right)d = -(2iM+2M-1)\frac{d}{2} < z < -(2iM+1)\frac{d}{2} - \left(\frac{M}{2}-1\right)d = -(2iM+M-1)\frac{d}{2}$, where $i \in \mathbb{Z}$ (in particular, if $i = 0$ then the region corresponds to the blue region in Fig. A.2). Moreover, due to symmetry, symbols x_1 and $x_{M/2+1}$ share the same error regions, as well as

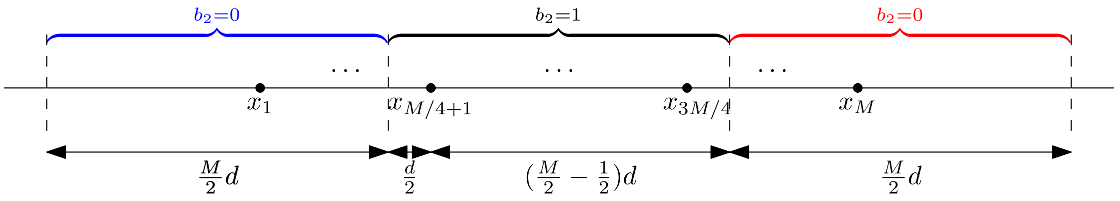


Figure A.3: Regions of the second bit b_2 of M-PAM constellation using Gray mapping.

symbols x_2 and $x_{M/2+2}$, and so on, up to $x_{M/2}$ and x_M. We can show that, for appropriate values of ℓ, $\Pr[e_1|x_\ell] = \Pr[e_1|x_{\ell \pm M/2}]$.

We can now use (A.5), and show that

$$\begin{aligned} P_b(e_1) = \frac{1}{M} \sum_{i=-\infty}^{\infty} & \underbrace{2q(2iM+1) + 2q(2iM+3) + \cdots + 2q(2iM+M-1)}_{M/2 \text{ terms}} \\ & - 2q(2iM+M+1) - \cdots - 2q(2iM+2M-1) \end{aligned}$$

By using a similar trick as we used for 2-PAM (i.e., we separate the summation for i from 0 to ∞ and j from $-\infty$ to -1, and set $i = n$ and $j = -n-1$), it is possible to show that

$$\begin{aligned} P_b(e_1) &= \frac{1}{M} \sum_{n=0}^{\infty} 4q(2nM+1) + \cdots + 4q(2nM+M-1) \\ &\qquad - 4q(2nM+M+1) - \cdots - 4q(2nM+2M-1) \\ &= \frac{1}{M} \sum_{n=0}^{\infty} \sum_{m=0}^{M-1} (-1)^{\lfloor \frac{2m}{M} \rfloor} 4q(2nM+2m+1) \end{aligned}$$

Finally, let $i = m + nM$. Note that $(-1)^{\lfloor \frac{2m}{M} \rfloor} = (-1)^{\lfloor \frac{2i-2nM}{M} \rfloor} = (-1)^{\lfloor \frac{2i}{M} - 2n \rfloor} = (-1)^{\lfloor \frac{2i}{M} \rfloor - 2n} = (-1)^{\lfloor \frac{2i}{M} \rfloor}$. Moreover, we can combine the two summations into one and rewrite the above equation in a more closed form

$$P_b(e_1) = \frac{1}{M} \sum_{i=0}^{\infty} (-1)^{\lfloor \frac{2i}{M} \rfloor} 4q(2i+1). \tag{A.10}$$

Note that, if $M = 2$, this is the same expression as bit-error probability of BPSK constellation (A.7).

We can use the same idea as $P_b(e_1)$ to find $P_b(e_2)$. Fig. A.3 shows the region where $b_2 = 0$ and $b_2 = 1$. For example, symbols $x_{M/4+1}, \ldots, x_{3M/4}$ have $b_2 = 1$, while the other symbols have $b_2 = 0$. Note that the error regions correspond to a translated version of b_1. For example, the error regions of b_2 for $x_{M/4+1}$ can be expressed as $-(2iM+M+1)\frac{d}{2} < z < -(2iM+1)\frac{d}{2}$, where $i \in \mathbb{Z}$ (in Fig. A.3, if $i = 0$, the error region corresponds to the blue region and if $i = -1$, it correspond to the red region). We can conclude that $P_b(e_2) = P_b(e_1)$, i.e.,

$$P_b(e_2) = \frac{1}{M} \sum_{n=0}^{\infty} (-1)^{\lfloor \frac{2n}{M} \rfloor} 4q(2n+1). \tag{A.11}$$

Fig. A.4 shows the regions of $b_3 = 1$ and $b_3 = 0$. An error region of b_3 for $x_{M/8+1}$ can be expressed as $\left(\frac{M}{4} - \frac{1}{2}\right) d = \left(\frac{M}{2} - 1\right) \frac{d}{2} < z < \left(\frac{M}{4} - \frac{1}{2}\right) d + \frac{M}{4} d = (M-1)\frac{d}{2}$ (which corresponds to

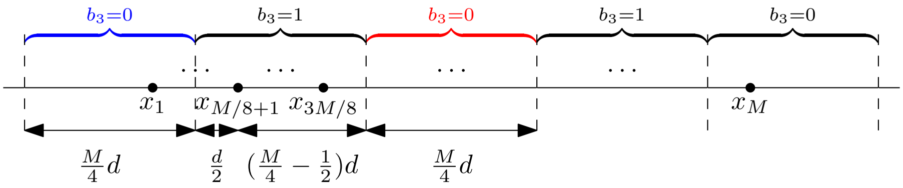


Figure A.4: Regions of the second bit b_3 of M-PAM constellation using Gray mapping.

the red region in Fig. A.4). Note that we can find other error regions by translate it by $i\frac{M}{2}d$ for $i \in \mathbb{Z}$. For example $\left(\frac{M}{2}-1\right)\frac{d}{2} - \frac{M}{2}d = -\left(\frac{M}{2}+1\right)\frac{d}{2} < z < (M-1)\frac{d}{2} - \frac{M}{2}d = -\frac{d}{2}$ corresponds to the blue region in Fig. A.4. By following the same argument as b_1 we can show that

$$\begin{aligned}\Pr[e_3|x_{M/8+1}] &= \sum_{i=-\infty}^{\infty} q\left(-\frac{M}{2}-1-iM\right) - q\left(-1-iM\right)\\ &= \sum_{i=-\infty}^{\infty} q(iM+1) - q\left(iM+\frac{M}{2}+1\right). \qquad \text{(A.12)}\end{aligned}$$

It is possible to show that if symbol x_ℓ is transmitted, for $\ell = \frac{M}{8}+2,\ldots,\frac{3}{8}M$, the error regions are the same as $x_{\ell-1}$ translate by d. Moreover, for appropriate values of ℓ, $\Pr[e_3|x_\ell] = \Pr[e_3|x_{\ell\pm\frac{M}{4}}]$. We have that

$$\begin{aligned}P_b(e_3) = \frac{1}{M}\sum_{i=-\infty}^{\infty} &\underbrace{4q(iM+1)+4q(iM+3)+\cdots+4q\left(iM+\frac{M}{2}-1\right)}_{M/4 \text{ terms}}\\ &- 4q\left(iM+\frac{M}{2}+1\right) - \cdots - 4q(iM+M-1) \qquad \text{(A.13)}\end{aligned}$$

Again, since the summation for negative values of i is equal to the summation for nonnegative values of i, we can express the above equation in a more closed form

$$P_b(b_3) = \frac{1}{M}\sum_{n=0}^{\infty}(-1)^{\lfloor\frac{4n}{M}\rfloor}8q(2n+1). \qquad \text{(A.14)}$$

By using this idea, it is possible to show that, for $k \geq 2$

$$\begin{aligned}P_b(e_k) = \frac{1}{M}\sum_{i=-\infty}^{\infty} &\underbrace{2^{k-1}q\left(i\frac{4M}{2^{k-1}}+1\right)+\cdots+2^{k-1}q\left(i\frac{4M}{2^{k-1}}+\frac{2M}{2^{k-1}}-1\right)}_{M/2^{k-1} \text{ terms}}\\ &- 2^{k-1}q\left(i\frac{4M}{2^{k-1}}+\frac{2M}{2^{k-1}}+1\right) - \cdots - 2^{k-1}q\left(i\frac{4M}{2^{k-1}}+\frac{4M}{2^{k-1}}-1\right)\\ = \frac{1}{M}\sum_{n=0}^{\infty}&(-1)^{\left\lfloor\frac{2^{k-1}n}{M}\right\rfloor}2^k q(2n+1) \qquad \text{(A.15)}\end{aligned}$$

We can now calculate (A.4), where $P_b(e_k)$ is given by (A.10) if $k=1$ and by (A.15) if $k \geq 2$.

Bibliography

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge, UK ; New York: Cambridge University Press, Jul. 2005.

[2] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[3] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[4] B. Hochwald, C. Peel, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: Perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[5] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Transactions on Communications*, vol. 52, no. 12, pp. 2057–2060, Dec. 2004.

[6] S. Stern and R. F. H. Fischer, "Advanced factorization strategies for lattice-reduction-aided preequalization," in *2016 IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 1471–1475.

[7] S.-N. Hong and G. Caire, "Reverse compute and forward: A low-complexity architecture for downlink distributed antenna systems," in *2012 IEEE International Symposium on Information Theory Proceedings*, Jul. 2012, pp. 1147–1151.

[8] W. He, B. Nazer, and S. Shamai (Shitz), "Uplink-Downlink Duality for Integer-Forcing," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1992–2011, Mar. 2018.

[9] D. Silva, G. Pivaro, G. Fraidenraich, and B. Aazhang, "On integer-forcing precoding for the Gaussian MIMO broadcast channel," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4476–4488, Jul. 2017.

[10] W. He, B. Nazer, and S. Shamai (Shitz), "Dirty-paper integer-forcing," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2015, pp. 1053–1060.

[11] Z. Zhou, B. Vucetic, M. Dohler, and Y. Li, “MIMO systems with adaptive modulation,” *IEEE Transactions on Vehicular Technology*, vol. 54, no. 5, pp. 1828–1842, Sep. 2005.

[12] D. Lee, “Performance analysis of zero-forcing-precoded scheduling system with adaptive modulation for multiuser-multiple input multiple output transmission,” *IET Communications*, vol. 9, no. 16, pp. 2007–2012, 2015.

[13] F. T. Luk and D. M. Tracy, “An improved LLL algorithm,” *Linear Algebra and its Applications*, vol. 428, no. 2-3, pp. 441–452, Jan. 2008.

[14] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*, 1st ed. Cambridge: Cambridge University Press, Feb. 2010.

[15] M. M. da Silva and F. A. Monteiro, Eds., *MIMO Processing for 4G and Beyond: Fundamentals and Evolution.* Boca Raton: CRC Press, Jun. 2014.

[16] E. Bjornson and E. Jorswieck, *Optimal Resource Allocation in Coordinated Multi-Cell Systems.* Boston, Mass.: Now Publishers Inc, Jan. 2013.

[17] G. Caire and S. Shamai, “On the achievable throughput of a multiantenna Gaussian broadcast channel,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[18] S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[19] P. Viswanath and D. N. C. Tse, “Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality,” *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[20] W. Yu and J. M. Cioffi, “Sum capacity of Gaussian vector broadcast channels,” *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[21] M. Costa, “Writing on dirty paper (Corresp.),” *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[22] A. Wiesel, Y. C. Eldar, and S. Shamai, “Zero-forcing precoding and generalized inverses,” *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[23] A. Goldsmith and Soon-Ghee Chua, “Variable-rate variable-power MQAM for fading channels,” *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[24] S. T. Chung and A. Goldsmith, “Degrees of freedom in adaptive modulation: A unified view,” *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[25] A. Svensson, “An Introduction to Adaptive QAM Modulation Schemes for Known and Predicted Channels,” *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2322–2336, Dec. 2007.

[26] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[27] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3857–3868, Oct. 2005.

[28] M. Joham, W. Utschick, and J. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[29] R. Zamir, *Lattice Coding for Signals and Networks.* Cambridge: Cambridge University Press, 2014.

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

[31] A. K. Lenstra, H. W. Lenstra, and L. Lovasz, "Factoring polynomials with rational coefficients," *Mathematische Annalen*, vol. 261, no. 4, pp. 515–534, Dec. 1982.

[32] Y. H. Gan, C. Ling, and W. H. Mow, "Complex lattice Reduction Algorithm for Low-Complexity Full-Diversity MIMO detection," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2701–2710, Jul. 2009.

[33] D. Wubben, D. Seethaler, J. Jalden, and G. Matz, "Lattice Reduction," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 70–91, May 2011.

[34] D. S. Dummit and R. M. Foote, *Abstract Algebra, 3rd Edition*, 3rd ed. Hoboken, NJ: Wiley, Jul. 2003.

[35] B. Nazer, V. R. Cadambe, V. Ntranos, and G. Caire, "Expanding the Compute-and-Forward Framework: Unequal Powers, Signal Levels, and Multiple Linear Combinations," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 4879–4909, Sep. 2016.

[36] C. Feng, D. Silva, and F. R. Kschischang, "An algebraic approach to physical-layer network coding," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7576–7596, Nov. 2013.

[37] U. Erez and R. Zamir, "Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[38] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 438–460, Mar. 2011.

[39] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-Forcing Linear Receivers," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7661–7685, Dec. 2014.

[40] P. Xu, "Parallel Cholesky-based reduction for the weighted integer least squares problem," *J Geod*, vol. 86, no. 1, pp. 35–52, Jan. 2012.

[41] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Transactions on Communications*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.