Centro de Tecnologia e Urbanismo
Departamento de Engenharia Elétrica

José Carlos Marinello Filho

Desempenho e Otimização de Sistemas de Comunicação Sem-Fio MIMO com Elevado Número de Antenas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Estadual de Londrina para obtenção do Título de Mestre em Engenharia Elétrica.

Londrina, PR
2014

José Carlos Marinello Filho

Desempenho e Otimização de Sistemas de Comunicação Sem-Fio MIMO com Elevado Número de Antenas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Estadual de Londrina para obtenção do Título de Mestre em Engenharia Elétrica.

Área de concentração: Sistemas Eletrônicos
Especialidade: Sistemas de Telecomunicações

Orientador:
Prof. Dr. Taufik Abrão

Londrina, PR
2014

Catalogação elaborada pela Divisão de Processos Técnicos da Biblioteca Central da Universidade Estadual de Londrina.

Dados Internacionais de Catalogação-na-Publicação (CIP)

M338d Marinello Filho, José Carlos.
Desempenho e otimização de sistemas de comunicação sem-fio MIMO com elevado número de antenas / José Carlos Marinello Filho. – Londrina, 2014.
87 f. : il.

Orientador: Taufik Abrão.
Dissertação (Mestrado em Engenharia Elétrica) – Universidade Estadual de Londrina, Centro de Tecnologia e Urbanismo, Programa de Pós-Graduação em Engenharia Elétrica, 2014.
Inclui bibliografia.

1. Sistemas de comunicação sem fio – Teses. 2. Sistemas de telecomunicação – Teses. 3. Sistema de antenas – Teses. 4. Engenharia elétrica – Teses. I. Abrão, Taufik. II. Universidade Estadual de Londrina. Centro de Tecnologia e Urbanismo. Programa de Pós- graduação em Engenharia Elétrica. III. Título.

CDU 621.39

José Carlos Marinello Filho

Desempenho e Otimização de Sistemas de Comunicação Sem-Fio MIMO com Elevado Número de Antenas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Estadual de Londrina para obtenção do Título de Mestre em Engenharia Elétrica.

Área de concentração: Sistemas Eletrônicos
Especialidade: Sistemas de Telecomunicações

Comissão Examinadora

Prof. Dr. Taufik Abrão
Depto. de Engenharia Elétrica
Universidade Estadual de Londrina
Orientador

Prof. Dr. Cristiano Magalhães Panazio
Depto. de Eng. de Telecomunicações e Controle
Universidade de São Paulo

Prof. Dr. Bruno Augusto Angélico
Depto. de Eng. de Telecomunicações e Controle
Universidade de São Paulo

19 de novembro de 2014

“Feliz aquele que transfere o que sabe e aprende o que ensina”

Cora Coralina

Agradecimentos

Agradeço primeiramente a Deus, por iluminar meu caminho e permitir que chegasse até aqui; ao meu orientador Taufik Abrão, por tudo que me ensinou como professor e amigo; a minha família, por todo o apoio e suporte que me deram para que conseguisse alcançar meus objetivos; a minha namorada Luciana, pelo carinho e companheirismo que sempre me oferece; e a todos os professores, amigos e companheiros de curso que muito me ajudaram ao longo desse trabalho.

Resumo

Neste trabalho, diferentes aspectos de sistemas de comunicação sem-fio empregando múltiplas antenas transmissoras e receptoras (*multiple-input-multiple-output*) (MIMO) são abordados, privilegiando a análise dos sistemas com elevada dimensionalidade, de forma a avaliar cenários onde características MIMO desejáveis possam ser exploradas. Assim, na primeira parte deste trabalho, um esquema de detecção aplicado ao *uplink* de tais sistemas é proposto, o qual combina as técnicas de redução treliça (*lattice reduction*) (LR) e otimização por colônia de formigas (*ant colony optimization*) (ACO), sendo o seu desempenho avaliado para elevadas dimensões do sistema (até 20×20, 4-QAM) e canais espacialmente correlacionados, apresentando um bom compromisso desempenho×complexidade. Na segunda parte do trabalho, o foco passa a ser a otimização do *downlink* de sistemas MIMO: uma formulação convexa para o problema da transmissão multiusuário (*multiuser transmission*) (MuT) visando a minimização da taxa de erro de bit (MinBER) é desenvolvida, a partir da qual diversos métodos de otimização determinísticos e heurísticos são aplicados. Ao final da análise desenvolvida, conclui-se que a técnica proposta busca em linha quase-Newton - função penalidade (*line search quasi-Newton - penalty function*) (LSQN-PF) apresentou o melhor compromisso desempenho×complexidade para elevadas dimensões de sistema. Na terceira parte do trabalho, analisa-se sistemas multi-celulares não-cooperativos baseados em duplexagem por divisão de tempo (*time division duplex*) (TDD), em que uma enorme quantidade de antenas é empregada na estação rádio-base (*base station*) (BS), sendo a informação de estado do canal (*channel state information*) (CSI) estimada por meio da transmissão de sequências piloto no *uplink*. Nesse cenário, caracteriza-se o desempenho sob a métrica da taxa de erro de bit (*bit-error-rate*) (BER) para a condição limite do número de antenas na BS tendendo a infinito, derivando-se expressões e comprovando sua validade por meio de simulações computacionais. Então, com base na expressão derivada e em outras presentes na literatura, uma metodologia de otimização do sistema é proposta, a qual consiste em distribuir as sequências de treinamento entre os usuários da célula de forma eficiente.

Palavras-chave: MIMO, detecção, precodificação, otimização, MinBER, MIMO massivo.

Abstract

In this work, different features of a wireless communication system deploying MIMO are adressed, focusing the analysis of increased dimension systems, in order to evaluate scenarios where many MIMO desirable characteristics can be found. Hence, in the first part of this work, a detection scheme applied to the *uplink* of such systems is proposed, that combines the techniques LR and ACO, being its performance evaluated for high system dimensions (up to 20×20, 4-QAM) and spatially correlated channels, presenting a good performance×complexity trade-off. In the second part of the work, our focus turns to be the *downlink* optimization of the MIMO system: a convex formulation for the MuT problem using the MinBER criterion is derived, from which several deterministic and heuristic optimization methods are applied. After the analysis performed, we conclude that the proposed LSQN-PF technique presented the best performance×complexity trade-off for large dimensions of the system. In the third part of the work, it is analysed noncooperative TDD multi-cellular systems, where a massive number of antennas are deployed in the BS, being CSI estimated by means of *uplink* pilot sequences transmission. In this scenario, the performance under the metric of BER is characterized for the limit of the number of BS antennas tending to infinity, deriving expressions and validating them by means of computational simulations. Then, based on this derived expression and others found in literature, a system optimization approach is proposed, which consists of assigning the training sequences among the users of the cell in an efficient manner.

Keywords: MIMO, detection, precoding, optimization, MinBER, massive MIMO.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

ACO otimização por colônia de formigas (*ant colony optimization*)

AWGN ruído aditivo gaussiano branco (*additive white gaussian noise*)

BC canal de difusão (*broadcast channel*)

BER taxa de erro de bit (*bit-error-rate*)

BS estação rádio-base (*base station*)

CDF função de distribuição cumulativa (*cumulative distribution function*)

CDMA múltiplo acesso por divisão de código (*code division multiple access*)

CSI informação de estado do canal (*channel state information*)

DPC codificação de papel sujo (*dirty paper coding*)

FDD duplexagem por divisão de frequências (*frequency division duplex*)

LR redução treliça (*lattice reduction*)

LSQN-EN busca em linha quase-Newton - ruído efetivo (*line search quasi-Newton - effective noise*)

LSQN-PF busca em linha quase-Newton - função penalidade (*line search quasi-Newton - penalty function*)

LTE *Long Term Evolution*

LTE-A *Long Term Evolution - Advanced*

MAC canal de múltiplo acesso (*multiple access channel*)

MACO otimização por colônia de formigas modificada (*modified ant colony optimization*)

MCS simulação Monte-Carlo (*Monte-Carlo simulation*)

MIMO múltiplas antenas transmissoras e receptoras (*multiple-input-multiple-output*)

MinBER minimização da taxa de erro de bit

ML máxima verossimilhança (*maximum likelihood*)

MMSE mínimo erro quadrático médio (*minimum mean squared error*)

MUI interferência multiusuário (*multiuser interference*)

MuT transmissão multiusuário (*multiuser transmission*)

MT terminais móveis (*mobile terminals*)

MU-MIMO MIMO multiusuário

OFDM multiplexagem por divisão de frequências ortogonais (*orthogonal frequency division multiplexing*)

PSO otimização por nuvem de partículas (*particle swarm optimization*)

RF rádio frequência

RZF *zero forcing* regularizado (*regularized zero forcing*)

SD detector esférico (*sphere decoding*)

SDMA múltiplo acesso por divisão espacial (*spatial division multiple access*)

SINR relação sinal-ruído-interferência (*signal-to-interference-plus-noise ratio*)

SIR relação sinal-interferência (*signal-to-interference ratio*)

SM modulação espacial (*spatial modulation*)

SNR relação sinal-ruído (*signal-to-noise ratio*)

SQP programação quadrática sequencial (*sequential quadratic programming*)

TDD duplexagem por divisão de tempo (*time division duplex*)

TSP problema do caixeiro viajante (*travelling salesman problem*)

TMF transmissor por filtro casado (*transmitter matched filtering*)

V-BLAST *Vertical-Bell Laboratories Layered Space-Time*

WiMAX *Worldwide Interoperability for Microwave Access*

ZF *zero forcing*

Convenções e Lista de Símbolos

Na notação das fórmulas, as seguintes convenções foram utilizadas:

- letras maiúsculas em negrito são matrizes, exemplo: $\mathbf{P}$, $\mathbf{H}$;
- letras minúsculas em negrito são vetores, exemplo: $\mathbf{x}$, $\mathbf{z}$;
- letras em itálico são escalares, exemplo: a, β;
- v_i é o i-ésimo elemento do vetor $\mathbf{v}$;
- A_{ij} é o elemento da i-ésima linha e j-ésima coluna da matriz $\mathbf{A}$;
- $\text{diag}(a_1, a_2, \ldots, a_k)$ é uma matriz diagonal com elementos $a_1, a_2, \ldots, a_k$;
- $\mathbf{I}_K$ é uma matriz identidade de ordem K;
- $\{\cdot\}^*$ é o operador conjugado complexo;
- $\{\cdot\}^T$ é o operador matriz transposta;
- $\{\cdot\}^H$ é o operador matriz conjugada transposta;
- $\{\cdot\}^{-1}$ é o operador matriz inversa;
- $\{\cdot\}^\dagger$ é o operador matriz pseudo-inversa;
- $\Re\{\cdot\}$ é o operador parte real;
- $\Im\{\cdot\}$ é o operador parte imaginária;
- $[x]^+$ é o valor máximo entre x e 0;
- $|\cdot|$ é o valor absoluto de um escalar ou vetor;
- $||\cdot||$ é a norma dois de um vetor;
- $\lceil \cdot \rfloor$ é o operador inteiro mais próximo;
- $\text{sgn}(\cdot)$ é o operador sinal;
- $\mathcal{U}(x, y)$ é um processo aleatório com distribuição uniforme entre as variáveis x e y;

- $\mathcal{N}(m, \sigma^2)$ é um processo aleatório com distribuição normal de média m e variância σ^2;
- $\mathbb{E}[\cdot]$ é o operador Esperança Estatística;
- $Var(\cdot)$ é o operador Variância;
- $Q(\cdot)$ é a função Q Gaussiana;
- $\nabla_{\mathbf{s}} P(\mathbf{s}, \alpha, \beta)$ é o operador gradiente de $P(\mathbf{s}, \alpha, \beta)$ em relação a $\mathbf{s}$;
- $\in$ significa pertence ao conjunto;

Palavras em itálico são utilizadas para identificar termos de língua inglesa não traduzidos.

Os seguintes símbolos serão utilizados:

símbolo	descrição
$\mathbf{H}$	Matriz contendo os coeficientes de desvanecimento de pequena escala do canal
N	Número de antenas na BS
M	Ordem da modulação QAM
K	Número de usuários atendidos por determinada BS
ρ	índice de correlação do canal

Visto que os símbolos não citados aqui podem assumir diferentes significados em cada parte do trabalho, sugere-se ao leitor consultar a descrição de notação contida na seção introdutória de cada manuscrito reproduzido no Apêndice A.

Para tornar mais expedita o referenciamento dos trabalhos contidos no Apêndice A, a seguinte convenção será utilizada ao longo deste texto de Dissertação para referenciar equação, seção, figura, tabela e algoritmo, respectivamente:

- A.1.Eq.n refere-se à Equação n do trabalho mostrado no Apêndice A.1;
- A.1.Sec.p refere-se à Seção p do trabalho mostrado no Apêndice A.1;
- A.2.Fig.k refere-se à Figura k do trabalho mostrado no Apêndice A.2;
- A.3.Tab.m refere-se à Tabela m do trabalho mostrado no Apêndice A.3;
- A.1.Alg.j refere-se ao Algoritmo j do trabalho mostrado no Apêndice A.1;

1 Introdução

Avanços tecnológicos atuais, tais como aplicações multimídia em tempo real, vídeo e imagem de alta definição, e uma ampla disponibilidade de serviços *on-line*, têm impulsionado uma demanda por taxas cada vez maiores dos serviços de comunicação. Essa demanda tecnológica porém surge em um cenário em que as disponibilidades de espectro e de energia se tornam cada vez mais limitadas, devido à variedade de serviços que compartilham o canal sem-fio, e as crescentes preocupações em economizar energia e reduzir a emissão de gases (FEHSKE et al., 2011), respectivamente. Nesse contexto, a tecnologia de múltiplas antenas transmissoras e receptoras (*multiple-input-multiple-output*) (MIMO) é um dos principais atributos dos padrões modernos de comunicação sem-fio, como por exemplo *Long Term Evolution* (LTE), *Long Term Evolution - Advanced* (LTE-A), Wi-Fi e *Worldwide Interoperability for Microwave Access* (WiMAX) (HANZO et al., 2010; LI et al., 2010), devido às elevadas eficiências energética e espectral que podem ser alcançadas. Quando implementada em grandes dimensões, na forma conhecida como *Massive* ou *Large* MIMO, a tecnologia MIMO também é vista como uma das potenciais tecnologias a serem aplicadas em padrões futuros, tais como o 5G (BOCCARDI et al., 2014).

Embora a proposta do uso de múltiplas antenas em sistemas de comunicação seja antiga (TELATAR, 1999; FOSCHINI; GANS, 1998), sua implementação em sistemas comerciais pode ser considerada relativamente recente (LI et al., 2010; PAUL; OGUNFUNMI, 2009), de forma que há muito interesse a respeito do tema, tanto por parte da comunidade científica como de grandes indústrias, e resultados importantes são constantemente divulgados.

Quando se fala de comunicação MIMO ponto-a-ponto, existem técnicas concebidas de forma a alcançar ganhos de diversidade e melhorar a confiabilidade do canal, como as relacionadas a códigos espaço temporais (TAROKH; SESHADRI; CALDERBANK, 1998; SUGIURA; CHEN; HANZO, 2012), e seleção de antenas (GHRAYEB, 2006; JALDEN; OTTERSTEN, 2007), e aquelas relacionadas à multiplexação espacial, direcionadas a maximização das taxas de dados e ganhos

de multiplexação (WOLNIANSKY et al., 1998; NISHIMOTO et al., 2007). Existem também técnicas capazes de combinar os ganhos de diversidade e multiplexação, como algumas variantes de seleção de antenas (GHRAYEB, 2006), e a modulação espacial (*spatial modulation*) (SM) (RENZO et al., 2014). Uma interessante revisão das técnicas propostas para MIMO ponto-a-ponto pode ser encontrada em (MIETZNER et al., 2009).

No entanto, o cenário de maior interesse nos últimos anos tem sido o MIMO multiusuário (MU-MIMO) (GESBERT et al., 2007), em que uma estação rádio-base (*base station*) (BS) equipada com N antenas serve um determinado número de terminais móveis (*mobile terminals*) (MT); ademais, sistemas MIMO baseados em múltiplo acesso por divisão espacial (*spatial division multiple access*) (SDMA) têm se tornado cada vez mais comuns. Nesses sistemas, cada MT pode ser um simples e econômico dispositivo de única antena, entre os quais os ganhos de taxa multiusuário são compartilhados. Além disso, sistemas MU-MIMO são mais tolerantes às condições de propagação do canal sem-fio do que sistemas ponto-a-ponto. Para condições de propagação sob linha de visada, por exemplo, os ganhos de multiplexação podem desaparecer para um sistema ponto-a-ponto, enquanto permanecem presentes em um sistema multiusuário desde que a separação angular dos usuários seja superior à resolução Rayleigh do *array* de antenas (MARZETTA, 2010).

No *uplink* dos sistemas MU-MIMO, também conhecido por canal de múltiplo acesso (*multiple access channel*) (MAC), o desenvolvimento de técnicas para detecção dos dados transmitidos combatendo eficientemente a interferência multiusuário (*multiuser interference*) (MUI) pode ser visto como uma simples generalização das técnicas de detecção para os sistemas ponto-a-ponto, com a vantagem da matriz de canal ser geralmente melhor condicionada em termos de correlação entre colunas, uma vez que se referem a terminais com localização distinta. Por outro lado, para o *downlink*, conhecido também como canal de difusão (*broadcast channel*) (BC), a estratégia ótima de transmissão, do ponto de vista da teoria da informação, consiste em uma técnica de pré-cancelamento de interferência denominada codificação de papel sujo (*dirty paper coding*) (DPC) (CAIRE; SHAMAI, 2003), cuja complexidade de implementação se torna proibitiva em sistemas práticos com elevadas taxas de dados e/ou número de usuários.

Tanto em sistemas MIMO ponto-a-ponto quanto em sistemas multiusuários, trabalhar com dimensões elevadas do sistema (particularmente em relação ao número de antenas na BS) pode trazer muitos benefícios, como alcançar elevadas eficiências energética e/ou espectral, bem como elevado número de usuários

atendidos. Quando o número de usuários atendidos K aumenta na mesma proporção de N, i.e., quando a razão $\frac{K}{N}$ permanece constante, exige-se muito mais das técnicas de combate a MUI, sejam elas implementadas na transmissão ou na recepção. Eficientes técnicas para o *uplink* destes cenários foram propostas em (VARDHAN et al., 2008), e o presente trabalho faz também uma análise de diferentes técnicas aplicadas ao *downlink* desses sistemas. Por outro lado, quando a dimensão do sistema aumenta indefinidamente mantendo fixo o número de usuários atendidos, é obtido um cenário em que a interferência dentro de uma mesma célula (intra-celular) desaparece, bem como o ruído aditivo gaussiano branco (*additive white gaussian noise*) (AWGN), de forma que o desempenho do sistema passa a ser limitado apenas pela interferência entre células diferentes (inter-celular) (MARZETTA, 2010). Nesse cenário, pode ser demonstrado que até mesmo as técnicas mais simples de precodificação e detecção alcançam o desempenho ótimo do sistema, sob o ponto de vista da máxima taxa de transmissão de dados alcançável (MARZETTA, 2010; RUSEK et al., 2013).

Esse trabalho tem como foco a análise e a otimização de sistemas MIMO de elevada dimensão, particularmente na condição em que o número de antenas na BS é elevada, sob diferentes configurações e cenários. Dessa forma, são analisadas técnicas capazes de operar de forma eficiente nesse cenário e alcançar os benefícios citados. A seguir são apresentadas as motivações para os três problemas abordados, enquanto uma revisão bibliográfica mais detalhada no contexto de cada assunto pode ser encontrada nas seções introdutórias dos manuscritos reproduzidos no Apêndice A, os quais correspondem aos artigos desenvolvidos no âmbito desta Dissertação de Mestrado.

1.1 Motivação: Detecção em Sistemas MIMO

Motivação para Detecção em MIMO

Sistemas MIMO são amplamente conhecidos por possibilitarem uma considerável melhoria na eficiência espectral (WOLNIANSKY et al., 1998) e/ou energética (ALAMOUTI, 1998) de sistemas de comunicação sem fio. Na configuração de multiplexação espacial (figura 1.1), é permitido um aumento significativo na taxa de transmissão de dados, para uma mesma potência e banda utilizadas, ao se explorar a diversidade de caminhos entre antenas transmissoras e receptoras (TSE; VISWANATH, 2010). Por outro lado, na configuração de diversidade espaço-temporal, esses múltiplos percursos são aproveitados de forma a reduzir a potência necessária para um certo desempenho do sistema, mantendo fixa a banda utilizada.

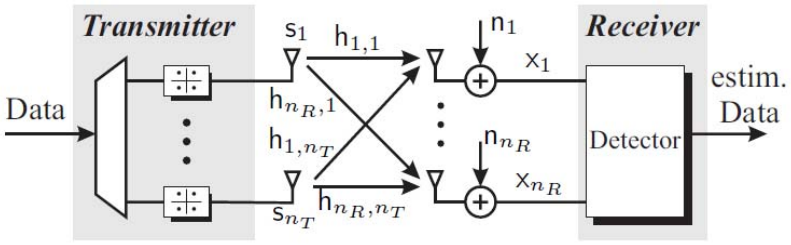

Figura 1.1: Configuração de um sistema MIMO ponto-a-ponto utilizando multiplexação espacial. [Fonte: (WUBBEN et al., 2004)].

Neste trabalho são abordadas técnicas de detecção para sistemas MIMO de multiplexação espacial, que podem estar operando tanto em comunicação ponto-a-ponto, quanto no *uplink* de um sistema multiusuário. Devido aos diferentes fluxos de dados provenientes de cada antena transmissora, os ganhos de diversidade que podem se apresentar em tais sistemas se devem unicamente às múltiplas antenas receptoras. Os detectores lineares, tais como o *zero forcing* (ZF), o de mínimo erro quadrático médio (*minimum mean squared error*) (MMSE) e os que operam por cancelamento de interferência, tais como o *Vertical-Bell Laboratories Layered Space-Time* (V-BLAST), são conhecidos por apresentarem uma baixa complexidade computacional, porém apresentam um desempenho bem inferior ao de máxima verossimilhança (*maximum likelihood*) (ML). Por outro lado, o detector esférico (*sphere decoding*) (SD) tradicional (AGRELL et al., 2002; JALDEN; OTTERSTEN, 2005) é conhecido por apresentar um desempenho ótimo, mas com uma complexidade computacional para regiões de baixa relação sinal-ruído (*signal-to-noise ratio*) (SNR) da mesma ordem daquela necessária ao ML. Além disso, modelos mais realistas de canais MIMO com elevado número de antenas devem levar em conta a correlação entre as colunas da matriz de canal, visto que, devido a limitações físicas, assumir antenas largamente espaçadas se torna uma situação bastante idealizada, principalmente em terminais móveis, mesmo que em frequências de microondas[1]. As principais técnicas de detecção MIMO consideradas nesta primeira parte do trabalho são introduzidas com detalhes em A.1.Sec.3, enquanto que o modelo de sistema adotado nesta parte do trabalho é descrito em A.1.Sec.2.

Como uma forma de melhorar o desempenho dos detectores lineares para canais mal condicionados (fortemente correlacionados), sem aumentar demasiadamente sua complexidade computacional, foram propostos detectores baseados na técnica de redução treliça (*lattice reduction*) (LR) (WUBBEN et al., 2004, 2011).

[1]Por exemplo, a banda S apresenta comprimentos de onda na faixa de 7,5 a 15 cm.

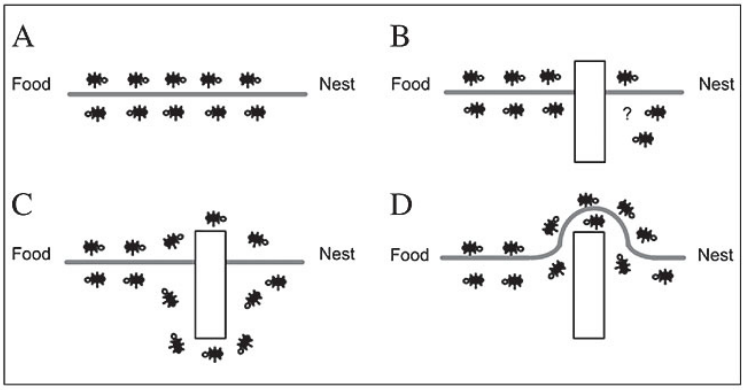

Figura 1.2: Comportamento das formigas na natureza, ilustrando sua capacidade de otimizar percursos. [Fonte: (LOPES; PERRETTO, 2004)].

Essa técnica transforma o canal parcialmente correlacionado em um equivalente, cuja matriz apresenta melhores condições de ortogonalidade. De posse dessa matriz de canal quase ortogonal no receptor, a detecção ocorre utilizando-se um detector linear de baixa complexidade.

A técnica de otimização por colônia de formigas (*ant colony optimization*) (ACO) (DORIGO; BIRATTARI; STUTZLE, 2006) foi proposta originalmente para problemas de natureza combinatória, como o problema do caixeiro viajante (*travelling salesman problem*) (TSP), porém foi estendida para problemas de natureza contínua em (SOCHA; DORIGO, 2008). Ela é baseada no comportamento das formigas em seu estado natural, que, em busca de alimento, possuem a capacidade de explorar o território a sua volta, encontrar os caminhos que levam às melhores fontes de comida, e deixar rastros que irão auxiliar na busca realizada por outras formigas (figura 1.2). Vários trabalhos já foram publicados aplicando essa técnica a diversos problemas em telecomunicações, como detecção em sistemas de múltiplo acesso por divisão de código (*code division multiple access*) (CDMA) (XU; YANG; HANZO, 2007; MARINELLO; SOUZA; ABRÃO, 2012), detecção em sistemas MIMO (KHURSHID; IRTEZA; KHAN, 2010; LAIN; CHEN, 2010; MARINELLO; ABRÃO, 2013b), alocação de recursos em redes sem fio (ZHAO et al., 2010; MARQUES et al., 2012), entre outros.

Um simples detector MIMO baseado em ACO foi proposto em (KHURSHID; IRTEZA; KHAN, 2010), o qual combina a técnica de otimização heurística aos detectores ZF e V-BLAST. Entretanto, os resultados numéricos mostraram um desempenho distante do ML para os detectores baseados em ACO propostos, especialmente em regiões de média e alta SNR. Em (LAIN; CHEN, 2010), um diferente método de detecção baseado em ACO foi proposto; no entanto, uma

vez que seu desempenho mostrou-se não muito satisfatório, foi introduzido um detector baseado na otimização por colônia de formigas modificada (*modified ant colony optimization*) (MACO) como alternativa para melhorar o compromisso desempenho×complexidade em um processo de detecção MIMO quase ótima.

No manuscrito do Apêndice A.1, propusemos um detector que, ao combinar as técnicas LR e ACO, oferece uma melhor alternativa em relação aos detectores MIMO baseado em ACO existentes, mostrando-se portanto eficiente para canais MIMO correlacionados e/ou de elevada dimensão. A escolha pela técnica ACO se deve à sua versatilidade, pela característica de se comportar bem quando aplicada a problemas de otimização de natureza combinatória, e pela relativa escassez de trabalhos científicos investigando sua aplicação ao problema de detecção em MIMO. Como já conhecido que o desempenho de uma técnica de otimização heurística depende da escolha adequada de seus parâmetros de entrada, é realizada uma etapa inicial de ajuste de parâmetros, sendo obtida uma única combinação que assegura um desempenho eficiente do algoritmo sob diferentes configurações do sistema MIMO. Os principais resultados numéricos de simulação Monte-Carlo (*Monte-Carlo simulation*) (MCS) obtidos são discutidos na seção 2.1. Observe-se que o artigo compilado no Apêndice A.1 foi publicado na revista *Wireless Personal Communications* (MARINELLO; ABRÃO, 2013b), e uma versão resumida (MARINELLO; ABRÃO, 2013a) publicada nos anais do evento *IEEE Wireless Communications and Networking Conference* (WCNC) 2013.

Escolher de forma fundamentada determinada técnica de otimização dado o conhecimento da classe de problemas a serem otimizados pode não ser uma tarefa simples. Os teoremas *No Free Lunch* (WOLPERT; MACREADY, 1997) afirmam, de forma generalizada, que a média dos desempenhos alcançados por certo algoritmo avaliados sobre todas as possíveis funções de otimização independe do algoritmo de otimização adotado. Isso implica que não há uma técnica mais adequada a ser adotada com o objetivo de otimizar problemas de qualquer natureza, pois se um algoritmo se mostra melhor que outro sobre determinada classe de problemas, necessariamente o oposto deve ser verdade para as demais classes. Geometricamente na figura 1.3, é demonstrado que, no hiperespaço das funções de otimização, as probabilidades dos algoritmos obterem determinado desempenho podem ser representadas por um cone, centrado no vetor diagonal (cujas coordenadas são simultaneamente unitárias) e com projeção constante sobre este. Tal vetor unitário representa a probabilidade uniforme de qualquer função de otimização ser tomada, ou seja, nenhum conhecimento *a priori* sobre o problema. Por outro lado, algum conhecimento *a priori* do problema pode ser representado

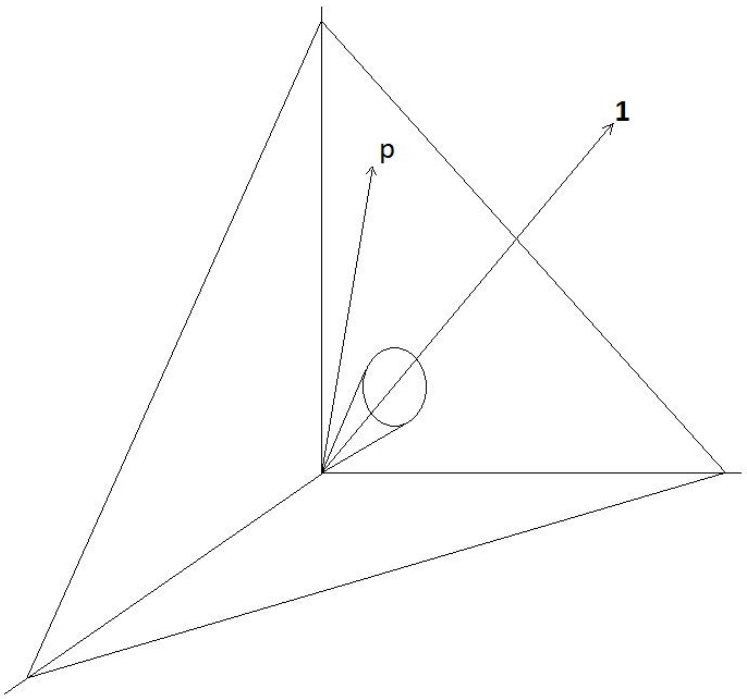

Figura 1.3: Representação geométrica das probabilidades de desempenho das técnicas de otimização (cone) e das probabilidades da função de otimização a ser resolvida (vetor p), no hiperespaço de todas as funções de otimização existentes. [Fonte: Autoria própria].

por um vetor nesse espaço, e assim é demonstrado, pelo teorema de Bayes, que o algoritmo mais eficiente para a classe de problemas em questão é aquele representado pelo vetor, contido no cone, com maior projeção sobre o vetor do problema de otimização.

Para o problema da detecção em sistemas MIMO, podemos imaginar as diferentes funções de otimização a serem encontradas, as quais dependem da matriz de canal e vetor de informação, como um vetor de probabilidades nesse hiperespaço, enquanto as técnicas baseadas em ACO, com diferentes parâmetros de entrada, sendo representadas por um cone centrado no vetor diagonal. Nesse ponto, a etapa de otimização de parâmetros, executada sobre diferentes realizações de canal e vetores de informação, pode ser vista como uma forma de encontrar o vetor contido no cone mais alinhado ao vetor da classe de problemas de otimização, ou, de maneira equivalente, como uma forma de usar o conhecimento *a priori* do problema para tornar a técnica de otimização mais adequada para o problema em questão.

1.2 Motivação: Transmissão Multiusuário em Sistemas MIMO

Motivação para MuT

Em sistemas de comunicação sem fio, é muito comum que as unidades móveis tenham uma única antena cada, pois, dada a frequência de operação e as dimensões do dispositivo, implementá-las com múltiplas antenas se torna uma tarefa desafiadora em termos de *hardware*. Adicionalmente, terminais móveis com alta eficiência energética são convenientes uma vez que resultam em autonomias satisfatórias de bateria para o usuário móvel; por outro lado, isto implica que tais terminais são incapazes de executar técnicas sofisticadas de processamento simultâneo *multistream* e/ou de detecção multiusuário. Com o objetivo de lidar com essas limitações dos MT's, sem prejudicar o desempenho do sistema, as técnicas de transmissão multiusuário (*multiuser transmission*) (MuT) têm sido propostas (VOJCIC; JANG, 1998).

A idéia central desse esquema é precodificar os dados a serem transmitidos, de forma que o receptor, equipado com um dispositivo de recepção convencional de baixa complexidade, possa fazer a detecção simplesmente quantizando o sinal banda-base recebido (figura 1.4). As primeiras técnicas desenvolvidas foram as lineares denominadas transmissor por filtro casado (*transmitter matched filtering*) (TMF) e aquelas baseadas em ZF, as quais apresentam baixa complexidade, mas um desempenho pobre quando comparadas com outras mais sofisticadas. Sabe-se que a técnica TMF é capaz de proporcionar a superposição de forma coerente entre os sinais que chegam no receptor provenientes de diferentes caminhos, e proporcionalmente ao ganho de cada percurso, porém os níveis de interferência resultantes se tornam elevados. Por outro lado o esquema ZF é capaz de eliminar totalmente a MUI, reduzindo, contudo, a SNR no receptor, em um efeito análogo à amplificação do ruído característico do receptor ZF. Já a técnica MuT baseada na abordagem MMSE apresenta um desempenho melhor em relação às anteriores, uma vez que leva em conta a variância do ruído juntamente com a interferência multiusuário. No entanto, todas as técnicas citadas até aqui não levam em conta a informação a ser transmitida na obtenção da matriz de precodificação, e assim minimizam a taxa de erro de bit (*bit-error-rate*) (BER) apenas indiretamente.

A matriz de precodificação pode ser projetada de forma a otimizar diferentes métricas: cancelamento de interferência (GUO; HUANG, 2008), minimização do erro quadrático médio (MMSE) (VOJCIC; JANG, 1998; CHOI; PERREAU, 2004), maximização da eficiência energética (HE et al., 2013) e/ou espectral (NGUYEN et al., 2013), e de forma análoga maximização da capacidade total (WANG; ZHANG,

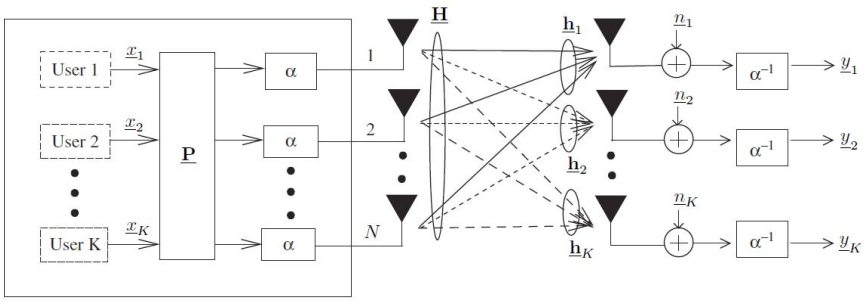

Figura 1.4: Configuração MuT MIMO, em que N antenas na BS se comunicam com K unidades móveis com uma única antena cada. [Fonte: (YAO; CHEN; HANZO, 2012), modificado].

2013), maximização da mínima distância entre os pontos da treliça de sinais recebidos sem ruído (KAPETANOVIC et al., 2013), minimização da probabilidade de erro no receptor (MinBER) (IRMER et al., 2003), entre outras. Neste trabalho de Dissertação, será abordado o critério de otimização da minimização da taxa de erro de bit (MinBER). Para sua implementação, é necessário que o transmissor possua certas informações, tais como a informação de estado do canal (*channel state information*) (CSI) e a SNR no receptor, além de obviamente ter conhecimento dos dados a serem transmitidos. Para sistemas de duplexagem por divisão de tempo (*time division duplex*) (TDD), a BS já possui as informações do canal provenientes do *uplink*, as quais serão as mesmas para o *downlink* uma vez que utilizam a mesma banda de frequências, e é assumido um tempo de coerência do canal suficientemente grande. O mesmo já não ocorre para sistemas com duplexagem por divisão de frequências (*frequency division duplex*) (FDD), e assim essas informações devem ser fornecidas pelo MT à BS por um canal de *feedback*.

A técnica MinBER-MuT foi desenvolvida baseada no critério de minimização da BER no receptor estimada pelo transmissor (IRMER et al., 2003). Devido a potência de transmissão fixa e limitada, o problema é formulado como a otimização de uma função não linear com restrições quadráticas, podendo assim ser utilizado o algoritmo de otimização não linear programação quadrática sequencial (*sequential quadratic programming*) (SQP) (NOCEDAL; WRIGHT, 1999), o qual representa o estado da arte para o problema (YAO et al., 2009; MATHWORKS, 2014). Porém, seu custo computacional pode, em alguns casos, torná-lo inviável para implementação em sistemas MIMO de altas taxas de transmissão de dados. É mostrado em (YAO et al., 2009) que a complexidade da técnica SQP formulada visando otimizar a matriz de precodificação do sistema se torna excessivamente elevada. No entanto, esse algoritmo pode ser formulado visando otimizar diretamente o vetor de sinais transmitidos, resultando em uma considerável redução de

complexidade devido à menor dimensão da variável de otimização, como mostrado no trabalho reproduzido no Apêndice A.2.

Uma aproximação sem restrições para o problema é formulada em (HABENDORF; FETTWEIS, 2007), calculando-se uma medida denominada variância de ruído efetivo, a qual é solucionada pelo algoritmo busca em linha quase-Newton - ruído efetivo (*line search quasi-Newton - effective noise*) (LSQN-EN); porém, nenhum estudo mais detalhado sobre a complexidade computacional foi realizado. Com o objetivo de obter um desempenho quase ótimo, ao custo de uma complexidade computacional aceitável, foi proposto em (YAO et al., 2009) um transmissor em que a busca pela matriz de precodificação MinBER é feita pela técnica de otimização por nuvem de partículas (*particle swarm optimization*) (PSO). Para tornar o problema sem restrições, é utilizado o método da penalidade, que incorpora a restrição na função objetivo de forma a penalizar as soluções que violem as restrições originais do problema.

No Apêndice A.2, é mostrado um trabalho a respeito do problema MinBER-MuT. Primeiramente é desenvolvida, a partir do modelo de sistema em A.2.Sec.2, uma formulação convexa para o problema, a qual torna mais simples a aplicação de técnicas de otimização, sem perda de desempenho. A partir daí, podemos elencar as seguintes contribuições do trabalho: **a**) a implementação de algoritmos de otimização para o problema proposto, tanto determinísticos como heurísticos, é descrita com detalhes em A.2.Sec.3, e as eficiências alcançadas são investigadas e comparadas de forma unificada; **b**) os grandes benefícios da abordagem MinBER-MuT são revelados para elevadas capacidades do sistema, ou seja, elevado número de antenas na BS e usuários atendidos; **c**) uma vez que em tais sistemas de elevadas dimensões a complexidade computacional pode se tornar proibitiva, uma cuidadosa análise de complexidade é realizada. Após um detalhado estudo sobre a complexidade computacional e de desempenho MinBER-MuT para elevadas dimensões do sistema, o trabalho é finalizado indicando os algoritmos com melhor compromisso desempenho×complexidade, conforme a dimensão do sistema MIMO. Os principais resultados são discutidos na seção 2.2. O manuscrito reproduzido no Apêndice A.2 foi submetido à revista *IEEE Transactions on Vehicular Technology*.

1.3 Motivação: Análise de Sistemas MIMO Massivos

Motivação para MIMO Massivo

Sabe-se que o desempenho e/ou confiabilidade de sistemas de comunicação sempre podem ser melhorados explorando formas de diversidade, sejam elas no tempo, frequência ou espaço, pois isso resulta em graus de liberdade adicionais (GOLDSMITH, 2005). Para sistemas multi-celulares MIMO (esquematizado na figura 1.5), uma forma conveniente de explorar graus de liberdade adicionais consiste em incrementar o número de antenas N na BS, pois as diversidades de tempo e frequência impactariam diretamente na capacidade do sistema.

Trabalhos recentes têm investigado o limite da melhoria de desempenho (confiabilidade) dos sistemas MIMO equipados com inúmeras antenas. Foi demonstrado em (MARZETTA, 2010), e posteriormente de forma mais generalizada em (FERNANDES; ASHIKHMIN; MARZETTA, 2013), que para sistemas multi-celulares TDD não-cooperativos, em que a CSI é estimada pela transmissão no *uplink* de sequências piloto (tal modelo de sistema é descrito em A.3.Sec.2), tanto o ruído AWGN quanto a interferência intra-celular desaparecem no limite quando $N \to \infty$. O único efeito que permanece restringindo o desempenho assintótico do sistema é a interferência inter-celular. De fato, o tempo de coerência do canal, inversamente proporcional à mobilidade dos usuários, limita também o tamanho das sequências piloto, as quais possuem tempo e bandas finitas para serem transmitidas. Por sua vez, o tamanho limitado das sequências piloto impactará no número de sequências ortogonais disponíveis, de forma que o reuso de tais sequências entre células vizinhas se torna inevitável. Assumindo, como uma situação de pior caso, que a etapa de treinamento ocorra de forma sincronizada entre todas as células, a estimação da CSI dos usuários servidos por uma determinada BS estará "contaminada" pela CSI dos usuários de células vizinhas, as quais estarão reutilizando as mesmas sequências de treinamento, em um fenômeno denominado "contaminação de pilotos". Este problema é sintetizado na equação A.3.Eq.2.

Sistemas MIMO nos quais N tende a infinito podem levar a reduções consideráveis nos níveis necessários de potência de transmissão. Em (NGO; LARSSON; MARZETTA, 2013) é mostrado, para o *uplink* de sistemas unicelulares, que as potências de transmissão podem ser tomadas inversamente proporcionais a N, para estimativas perfeitas de CSI, ou a $\sqrt{N}$, para estimativas imperfeitas, sem nenhuma perda de desempenho. Por outro lado, para sistemas multi-celulares, é demonstrado em (MARZETTA, 2010) que a relação sinal-ruído-interferência (*signal-*

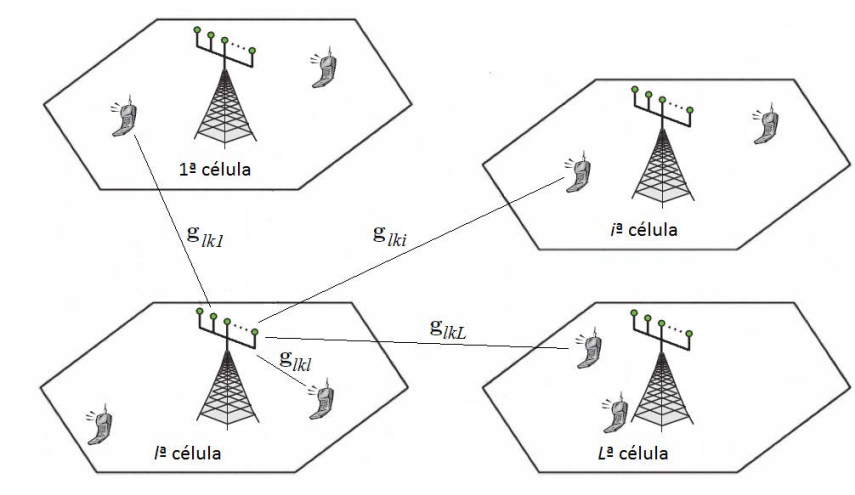

Figura 1.5: Sistema MIMO multi-celular, em que uma BS equipada com N antenas é posicionada ao centro de cada uma de L células hexagonais. [Fonte: autoria própria].

to-interference-plus-noise ratio) (SINR) no limite de $N \to \infty$ depende apenas dos coeficientes de desvanecimento de larga escala, sendo independente da potência de transmissão, a qual pode ser tornada tão baixa quanto se queira, uma vez que o nível de interferência diminui da mesma forma que o nível do sinal de interesse.

Em (FERNANDES; ASHIKHMIN; MARZETTA, 2013) é proposto uma forma de distribuir os intervalos de treinamento dentro de um mesmo intervalo de coerência de canal entre usuários de células próximas, de forma a evitar transmissões simultâneas e minimizar a contaminação de pilotos. Embora a técnica proposta alcance consideráveis aumentos da capacidade assintótica, esta ainda permanece saturada. Técnicas tradicionais de combate à interferência multi-celular, como a de reuso de frequências fracional, que consiste em dividir a banda disponível por um determinado fator de reuso, de forma que parcelas de frequências diferentes sejam disponibilizadas entre células de um mesmo *cluster*, podem ser aplicadas, porém resultam em capacidades médias totais de sistema ainda menores ao dividir a banda disponível (MARZETTA, 2010). Todavia, para fatores de reuso maiores, a capacidade do sistema é distribuída de forma mais igualitária entre os usuários, de forma que a capacidade individual e a relação sinal-interferência (*signal-to-interference ratio*) (SIR) de probabilidades maiores que 95% sejam maiores, como mostra a tabela 1.1.

Mesmo com o número de antenas na BS tendendo a infinito, o número de usuários que podem ser atendidos pela mesma BS permanece limitado pelo tempo de coerência do canal. Assumindo-se um sistema que utiliza MIMO em conjunto com multiplexagem por divisão de frequências ortogonais (*orthogonal fre-*

Tabela 1.1: Desempenho para *downlink* do sistema em (MARZETTA, 2010) em função do fator de reuso de frequências.

Fator de Reuso de Freq.	Prob.$> 0,95$ SIR(dB)	Prob.$> 0,95$ Capacidade por MT (Mbits/s)	Cap. Média por MT (Mbits/s)	Cap. Média por Célula (Mbits/s)
1	-29	0,016	44	1800
3	$-5,8$	0,89	28	1200
7	8,9	3,6	17	730

Tabela II de (MARZETTA, 2010)

quency division multiplexing) (OFDM), sabe-se que o coeficiente de canal de um usuário permanece constante dentre $\mathtt{N}_{\mathrm{smooth}}$ subportadoras adjacentes (HANZO et al., 2010), o qual pode ser visto como o número de subportadoras em que é dividido um intervalo de banda de coerência do canal[2]. Assim, dentro de uma mesma célula, a mesma sequência pode ser atribuída a $\mathtt{N}_{\mathrm{smooth}}$ usuários diferentes, os quais podem ser distinguidos pela BS durante a etapa de treinamento por meio de filtragens apropriadas. Logo, o número de usuários que podem ser atendidos resulta $K = \tau \cdot \mathtt{N}_{\mathrm{smooth}}$, em que τ é o comprimento das sequências piloto[3]. Considerando os parâmetros OFDM adotados em (MARZETTA, 2010), idênticos ao do padrão LTE, tem-se um intervalo de coerência do canal de 500μs, que é dividido em 7 símbolos OFDM, dos quais 3 são dedicados ao treinamento no *uplink*, 1 para processamento, e os outros 3 para transmissão de dados, que pode ser em qualquer direção. Sendo o intervalo de suavização em frequência de $\mathtt{N}_{\mathrm{smooth}} = 14$ subportadoras, é possível acomodar até 42 usuários em uma mesma célula, os quais podem estar se movendo em velocidades de até 288km/h.

Em (ASHIKHMIN; MARZETTA, 2012) é proposto uma técnica de precodificação capaz de eliminar o efeito de contaminação de pilotos, resultando em ganhos assintóticos ilimitados. Porém, o esquema proposto exige o compartilhamento de algumas informações entre as diferentes células, e/ou a uma unidade de processamento central. Primeiramente, é necessário que os coeficientes de desvanecimento lento de todos os usuários sejam compartilhados em uma unidade de processamento central, o que não é tão problemático de ser alcançado, uma vez que tais coeficientes são válidos por um tempo relativamente longo, e o número de usuários é limitado na célula. No entanto, o esquema exige que os dados de informação de todos os usuários, para o *downlink*, e os sinais recebidos e disponíveis na saída

[2]Definido como a largura de banda cuja resposta em frequência do canal pode ser considerada constante

[3]Quantidade de símbolos OFDM dentro de um intervalo de coerência do canal dedicados ao envio de tais sequências

de um combinador, para o *uplink*, sejam disponíveis entre todas as células, o que pode resultar em um compartilhamento excessivo de informação entre as mesmas, sobrecarregando os canais de *backhaul*. Para minimizar parcialmente o problema, as células podem ser agrupadas em *clusters*, de forma a limitar o volume de dados enviados a uma mesma unidade de processamento central e compartilhado entre células.

Uma formulação MMSE para sistemas multi-celulares não-cooperativos TDD, considerando o efeito de contaminação de pilotos, é derivada em (JOSE et al., 2011). A função objetivo a ser minimizada é formulada como o valor esperado do erro quadrático entre o vetor de símbolos transmitidos e o sinal recebido pelos usuários estimado pela BS, somado à interferência quadrática vista pelos usuários das outras células. Assim, o valor ótimo para esse problema, encontrado na forma fechada, é capaz de minimizar as interferências intra-celular e inter-celular de forma conjunta, resultando em apreciáveis capacidades para o sistema como um todo, e superando as demais técnicas unicelulares de precodificação consideradas.

Diversos problemas práticos surgem quando se busca implementar uma BS com elevado número de antenas (RUSEK et al., 2013), como limitações de tamanho, correlação e acoplamento mútuo entre antenas, uma vez que em sistemas MIMO massivos a tendência é a redução do espaçamento entre as antenas, consumo de energia dos circuitos de rádio frequência (RF), entre outros. A implementação de uma BS cujas antenas são distribuídas em duas dimensões é proposta em (NAM et al., 2013). Embora essa configuração permita uma maior densidade de antenas na BS, os efeitos de correlação e acoplamento mútuo se tornam mais hostis. Como exemplo ilustrativo, uma implementação bi-dimensional de *large*-MIMO com 32 antenas alcança de 2 a 3,6 vezes a capacidade média de um sistema LTE equipado com 2 antenas na BS.

Uma análise de capacidade quando o número de antenas cresce é feita em (HOYDIS; BRINK; DEBBAH, 2013) para cenários mais realistas, em que N não é extremamente grande quando comparado a K. Os autores derivaram expressões para a quantidade de antenas por MT necessárias para se alcançar uma determinada porcentagem do limite assintótico de capacidade, chegando então à conclusão de que o esquema *zero forcing* regularizado (*regularized zero forcing*) (RZF) pode alcançar o desempenho assintótico da técnica TMF com um número significantemente reduzido de antenas para o *downlink*. Resultados análogos também foram encontrados para o *uplink*.

No Apêndice A.3, é apresentado um trabalho em que a análise do sistema

Massive MIMO multi-celular não-cooperativo TDD é feita sob a métrica de desempenho BER. A escolha por esse critério de qualidade se deve à sua importância em sistemas de comunicação digital, e pela escassez de trabalhos científicos investigando seu comportamento em sistemas MIMO em que $N \to \infty$. Primeiramente, é demonstrado que a técnica de precodificação ZF converge assintoticamente para o mesmo desempenho da TMF quando se leva em conta a restrição de potência de transmissão na BS. Isto difere do que é mostrado em (RUSEK et al., 2013), onde os autores desprezam tal restrição, justificando que os níveis reais de potência não são relevantes na condição assintótica do sistema. Argumenta-se aqui que sistemas de comunicação energeticamente eficientes terão mais e mais importância e significado em sistemas 4G e 5G. Portanto, a restrição de potência máxima disponível na BS deve ser considerada no projeto. Depois, é feita uma análise da probabilidade de erro no *downlink* do sistema, derivando-se expressões fechadas em função dos níveis de potência e dos coeficientes de desvanecimento de larga escala associados aos usuários móveis. Por fim, com base na expressão derivada e em outra demonstrada em (FERNANDES; ASHIKHMIN; MARZETTA, 2013), é proposta uma estratégia de otimização do sistema, chamada de alocação de pilotos, a qual consiste em distribuir entre os usuários da célula as sequências piloto disponíveis de forma eficiente. Na seção 2.3 são discutidos os principais resultados. O artigo mostrado foi submetido para a revista *IEEE Systems Journal*.

2 Discussão dos Resultados

Nesta seção são discutidos os principais resultados obtidos para os três problemas tratados. A discussão de tais resultados seguirá a mesma ordem dos trabalhos apresentados no Apêndice A. Todas as curvas e tabelas apresentadas nesta seção são resultados numéricos de simulação computacional baseada no método Monte-Carlo, e portanto de autoria própria, sendo essa informação omitida das legendas visando não sobrecarregar a notação.

2.1 Detecção em Sistemas MIMO

No trabalho do Apêndice A.1, é proposto um detector para sistemas MIMO que combina as técnicas de LR e ACO, denominado LR-ACO (A.1.Alg.2). Seu desempenho em termos de BER é avaliado e comparado com outras técnicas aplicadas ao problema em A.1.Sec.4, tais como as lineares MMSE e LR-MMSE, o esquema baseado em ACO proposto em (LAIN; CHEN, 2010) (ACO_1), e uma versão modificada que usa a solução LR-MMSE como ponto de partida do algoritmo (ACO_2), sempre visando elevar o número de antenas do sistema e/ou a ordem de modulação da constelação. Como referência de detecção ótima, também é mostrado o desempenho da técnica SD, sempre que as configurações do canal MIMO permitirem sua simulação em tempo computacional factível. Embora a relação E_b/N_0 seja usualmente designada, por simplicidade de notação, como SNR nessa seção, tais grandezas estão relacionadas por $\text{SNR} = log_2(M) \cdot E_b/N_0$.

Inicialmente, o desempenho das técnicas de detecção MIMO para 4×4, 64-QAM, considerando o aumento do índice de correlação de canal ρ, é avaliado na figura 2.1. A mesma análise é feita para 8×8 16-QAM e para 20×20 4-QAM, na figura 2.2 e na figura 2.3, respectivamente. Como o tempo de simulação para a técnica SD se tornou excessivamente alto para 20×20 4-QAM com $\rho = 0,9$, foram mostrados apenas os resultados de canal fracamente e medianamente correlacionados na figura 2.3.

Analisando as figuras 2.1, 2.2 e 2.3, pode-se concluir que o desempenho em

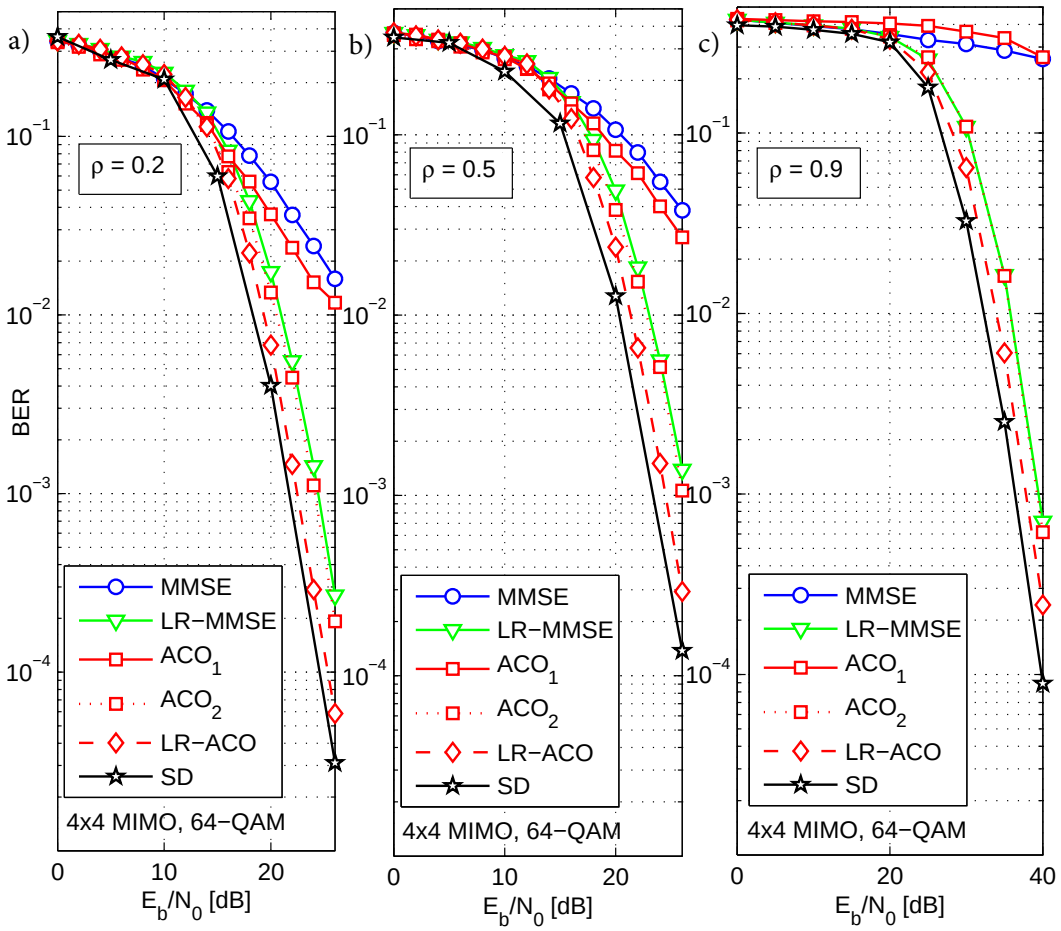

Figura 2.1: Desempenho detectores MIMO em função da SNR, considerando 4×4, 64-QAM, para canais: a) fracamente correlacionados ($\rho = 0,2$); b) medianamente correlacionados ($\rho = 0,5$); c) fortemente correlacionados ($\rho = 0,9$).

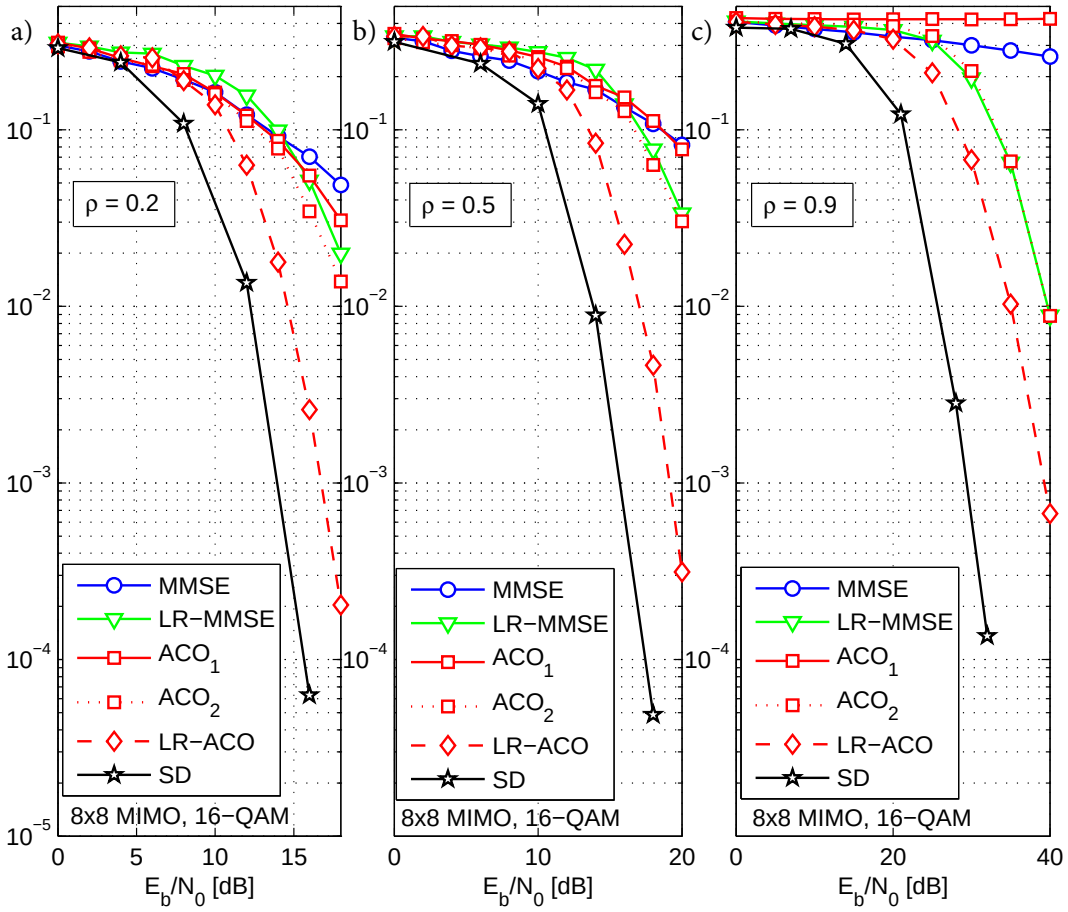

Figura 2.2: Desempenho detectores MIMO em função da SNR, considerando 8×8, 16-QAM, para canais: a) fracamente correlacionados ($\rho = 0,2$); b) medianamente correlacionados ($\rho = 0,5$); c) fortemente correlacionados ($\rho = 0,9$).

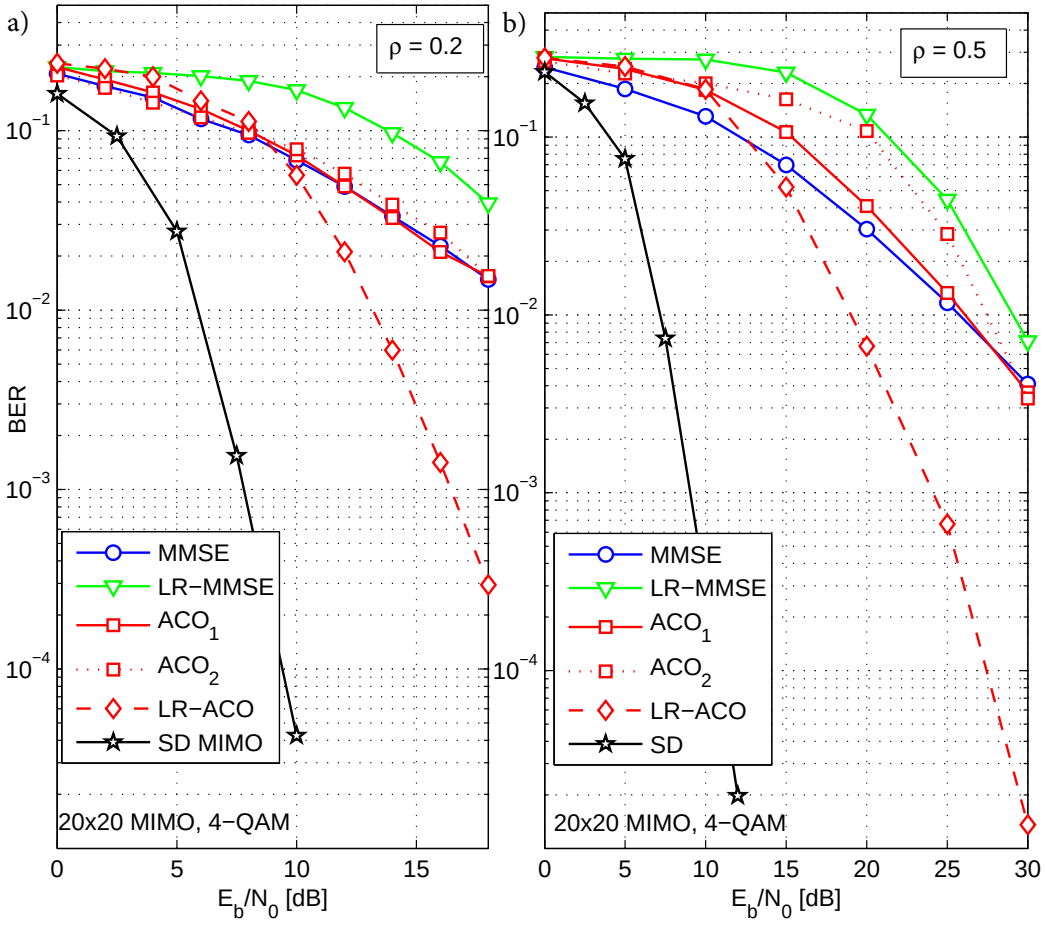

Figura 2.3: Desempenho detectores MIMO em função da SNR, considerando 20×20, 4-QAM, para canais: a) fracamente correlacionados ($\rho = 0,2$); b) medianamente correlacionados ($\rho = 0,5$).

termos de BER da técnica LR-ACO para 4×4 64-QAM permaneceu próximo da referência de desempenho SD, até mesmo para o canal fortemente correlacionado. Para 8×8 16-QAM, o deslocamento de SNR, para as regiões de elevada SNR, se torna aproximadamente o mesmo dos canais descorrelacionados para $\rho = 0,2$ e $\rho = 0,5$, ao passo que se torna um pouco maior para $\rho = 0,9$. Adicionalmente, para 20×20 4-QAM, o deslocamento de SNR, que era $\approx$ 8 dB para canais descorrelacionados, como mostra os resultados em A.1.Sec.4, se torna $\approx$ 10 dB para canais fracamente correlacionado, e $\approx$ 17 dB para canais medianamente correlacionados. Um resultado interessante que pode ser notado a partir das figuras é que o detector LR-ACO alcançou a ordem de diversidade completa, definida como a taxa de inclinação da curva de BER para a técnica ML em regiões de alta SNR, para todas as configurações investigadas.

As figuras 2.4 e 2.5 mostram o comportamento dos detectores MIMO analisados para condição de eficiência espectral crescente em canais medianamente correlacionados ($\rho = 0,5$). Enquanto na figura 2.4 a análise é feita sendo a ordem de modulação fixa (4-QAM), na figura 2.5 a configuração espacial MIMO é mantida constante (8×8 MIMO). A tabela 2.1 mostra as eficiências espectrais alcançadas na figura 2.4, enquanto a tabela 2.2 faz o mesmo para a figura 2.5. A partir das curvas obtidas, pode-se notar novamente a ordem de diversidade do detector LR-ACO muito próxima à do detector SD[1]. De forma geral, nota-se também que a perda de desempenho do esquema LR-ACO em relação ao SD é mais sensível ao aumento do número de antenas do que em relação ao aumento da cardinalidade do sistema. Por exemplo, o deslocamento de SNR passa de $\approx$ 1dB na figura 2.4.a para $\approx$ 7dB na figura 2.4.b, e para $\approx$ 17dB na figura 2.4.c. Por outro lado, esse deslocamento se mantém aproximadamente fixo em $\approx$ 3dB da figura 2.5.a para a figura 2.5.b, passando para $\approx$ 5dB na figura 2.5.c. Ainda assim, a figura 2.6 mostra que, para uma eficiência espectral fixa (24 b/s.Hz) em canais medianamente correlacionados, trabalhar em configurações com maior número de antenas é mais eficiente do ponto de vista da eficiência energética, dado o ganho de diversidade proporcionado por graus de liberdade adicionais. No entanto, outros desafios surgem, tais como complexidade de implementação em termos de *hardware*, limitações dimensionais, e uma maior complexidade computacional dos algoritmos, como será discutido a seguir.

Importantes conclusões podem ser obtidas analisando-se as curvas de complexidades mostradas na figura 2.7, tanto em termos de operações de ponto flutuante

[1]Para eficiências espectrais mais elevadas, a precisão da curva de desempenho SD se torna menor devido ao menor número de erros do método MCS, haja vista a complexidade excessiva desse detector nesses cenários

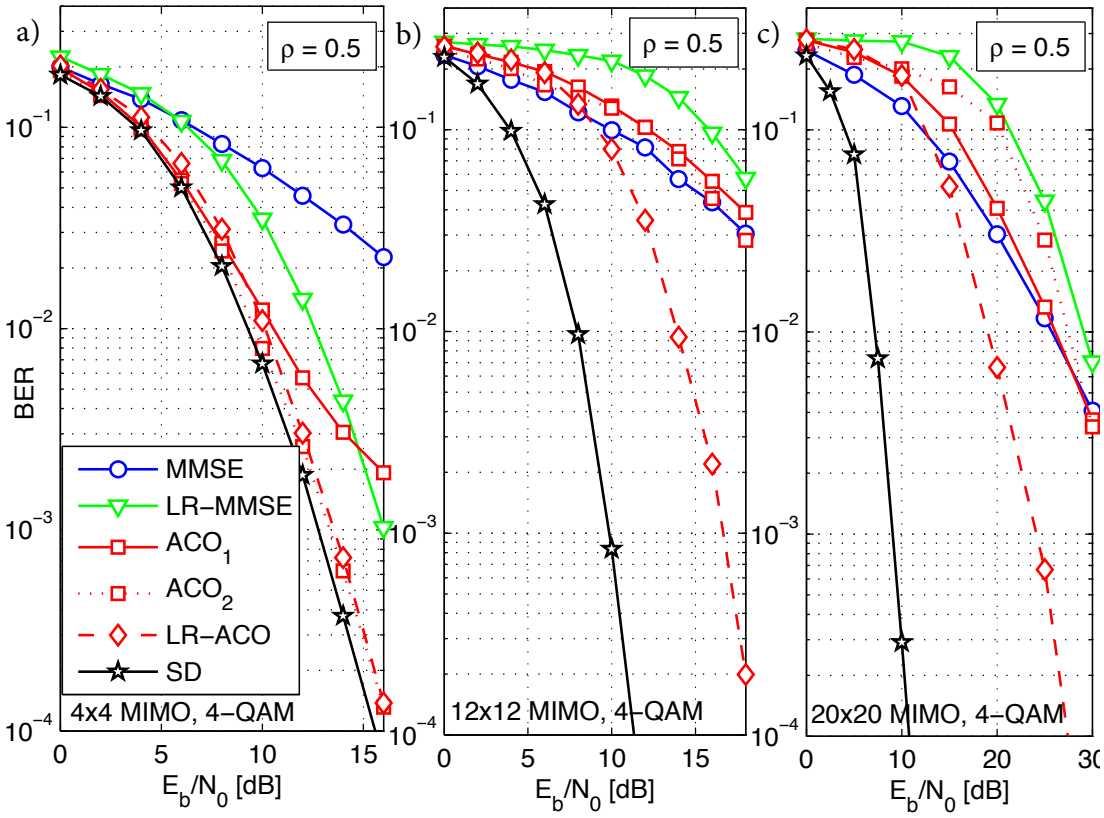

Figura 2.4: Desempenho detectores MIMO em função da SNR, considerando canais medianamente correlacionados ($\rho = 0,5$), para eficiência espectral crescente, fixada a ordem de modulação em 4-QAM: a) 4×4 MIMO; b) 12×12 MIMO; c) 20×20 MIMO.

Tabela 2.1: Eficiência espectral [b/s.Hz] $\times$ configuração espacial MIMO.

MIMO	4×4	12×12	20×20
4-QAM	8	24	40

Tabela 2.2: Eficiência espectral [b/s.Hz] $\times$ ordem de modulação M-QAM.

MIMO	4-QAM	16-QAM	64-QAM
8×8	16	32	48

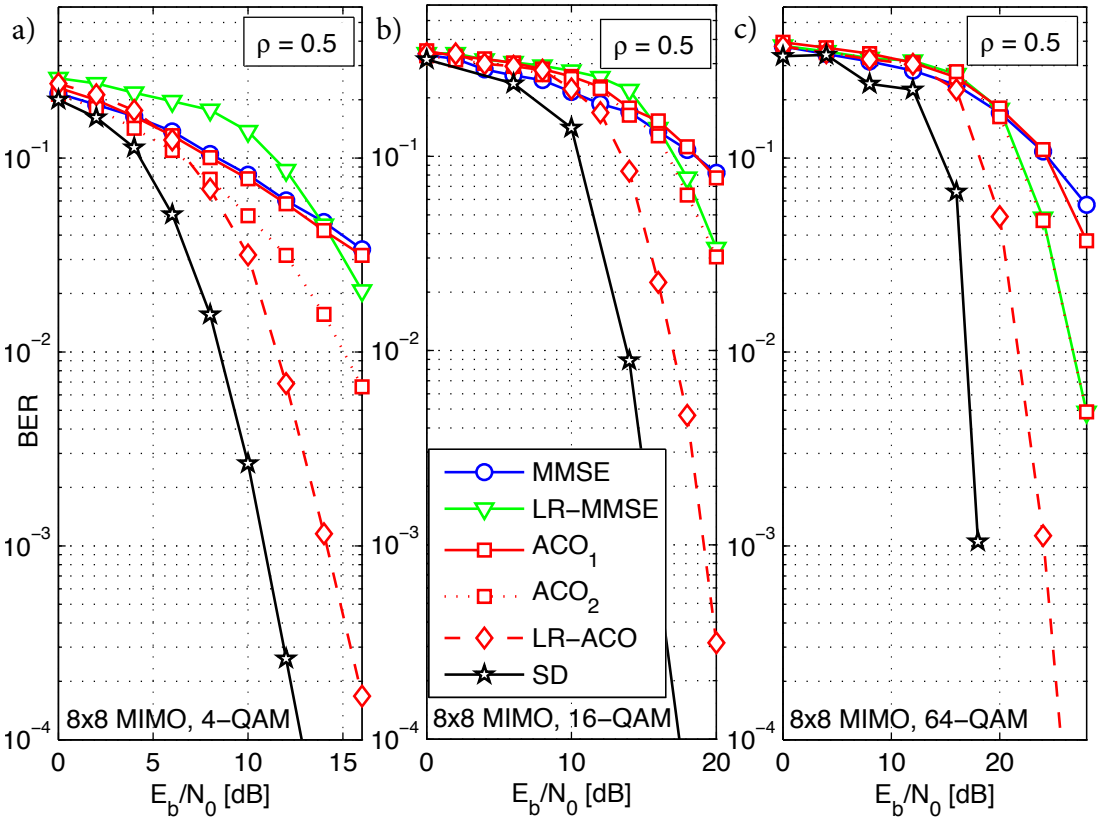

Figura 2.5: Desempenho detectores MIMO em função da SNR, considerando canais medianamente correlacionados ($\rho = 0,5$), para eficiência espectral crescente, fixada a configuração espacial em 8×8 MIMO: a) 4-QAM; b) 16-QAM; c) 64-QAM.

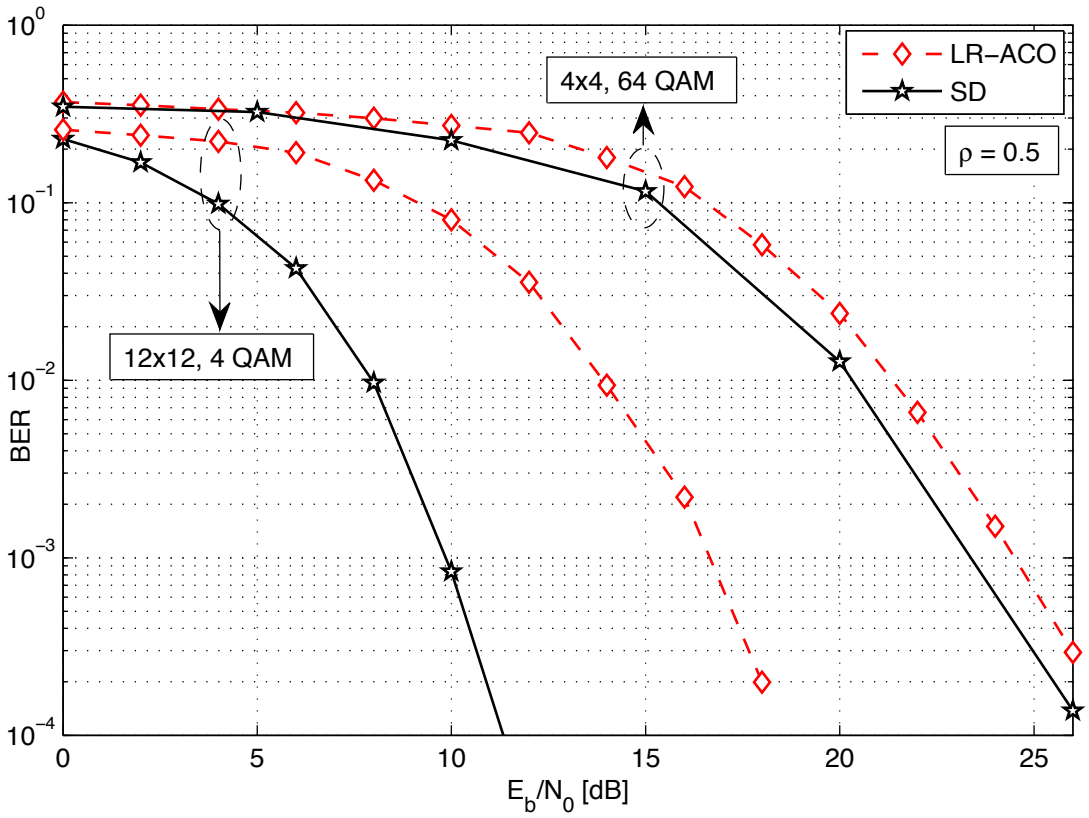

Figura 2.6: Desempenho detectores MIMO em função da SNR, considerando canais medianamente correlacionados ($\rho = 0,5$) e eficiência espectral constante em 24 b/s.Hz: 4×4, 64-QAM *vs* 12×12, 4-QAM.

(*flops*) (GOLUB; LOAN, 1996), quanto tempo computacional[2]. Primeiramente, são comprovados alguns resultados importantes de (JALDEN; OTTERSTEN, 2005), onde foi demonstrado analiticamente que a complexidade do SD cresce de forma exponencial. De fato, isso pode ser visto nas duas análises da figura 2.7, tanto em função do número de antenas como da ordem de modulação M-QAM, uma vez que as curvas de complexidade para o SD rapidamente assumem valores fora das escalas do eixo-y adotadas na figura 2.7 (em termos de *flops* e tempo computacional, respectivamente). Além disso, a figura mostra que a técnica LR-ACO apresenta sempre uma menor complexidade em relação aos outros detectores baseados em ACO considerados, enquanto seu desempenho é substancialmente melhorado, bem como a ordem de diversidade alcançada é completa. As pequenas diferenças de formato entre as curvas de complexidade em termos de *flops* e *time* podem ser justificadas por simplificações adotadas no cálculo da complexidade em termos de *flops*, como descrito com mais detalhes em A.1.Sec.4.4. Tendo em vista novamente a figura 2.6, pode-se notar que, para ambas as técnicas mostradas, a configuração com maior número de antenas resulta em maior complexidade de processamento. Para o detector LR-ACO, tal conclusão é imediata da figura 2.7. Já para o detector SD, embora o número de soluções disponíveis seja o mesmo ($64^4 = 4^{12}$), o fato de operar em uma SNR menor (para uma mesma BER) torna a complexidade muito mais elevada (JALDEN; OTTERSTEN, 2005).

2.2 Transmissão Multiusuário em Sistemas MIMO

Uma análise das técnicas de transmissão multiusuário (MuT) baseadas no critério MinBER é desenvolvida no trabalho apresentado no Apêndice A.2. Em particular, uma formulação convexa para o problema de otimização é apresentada na equação (A.2.Eq.14), disciplinando a aplicação de técnicas de otimização, tanto determinísticas quanto heurísticas. Em seguida, a implementação de um algoritmo SQP adequado à resolução do problema é descrita em detalhes, bem como outras duas técnicas determinísticas, a por projeção de gradiente e a por busca em linha quase-Newton, além de duas técnicas heurísticas, a ACO e a PSO. Tais técnicas são aplicadas sobre o problema sem restrição, a partir do uso dos métodos ruído efetivo e função penalidade, como mostrado na seção A.2.Sec.3. Novamente, o objetivo é examinar configurações MIMO de elevada dimensão do sistema, no caso o número de antenas na BS correspondente ao número de usuários servidos.

[2]Todos os dados de tempo computacional dos algoritmos foram obtidos a partir de uma *Workstation* Dell Precision T7500, com processador Intel Xeon E5620 de 2,40GHz, e 4GB de memória RAM

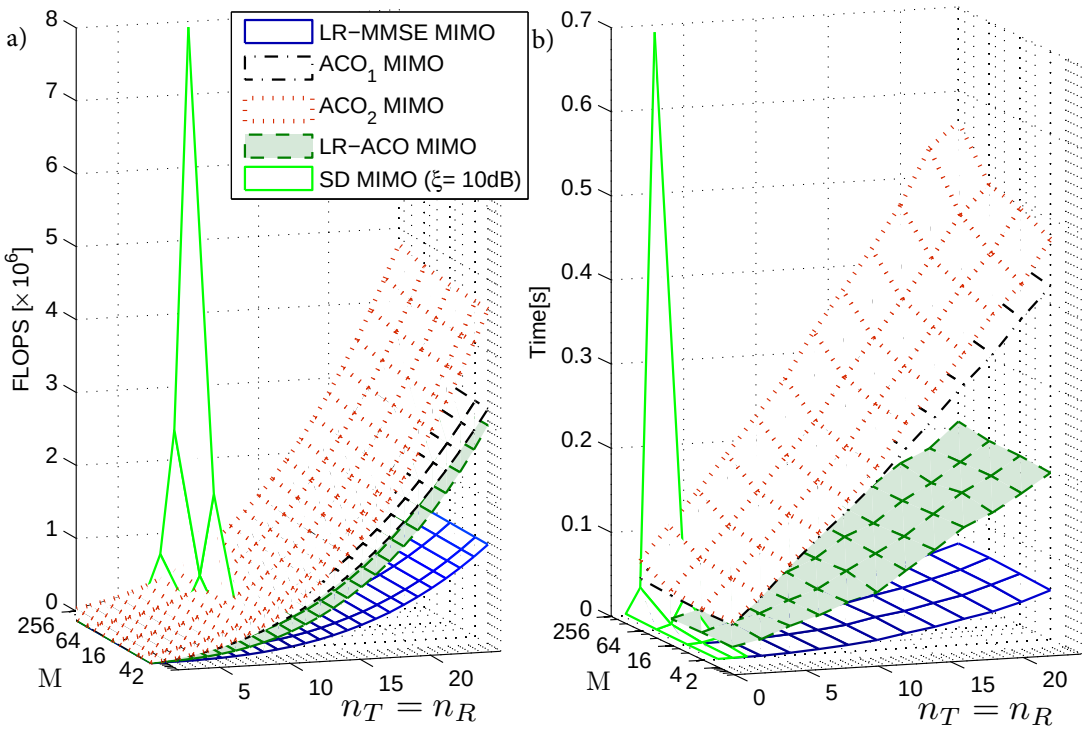

Figura 2.7: Complexidade Detectores MIMO em função da dimensão do sistema, considerando $\rho = 0$ e SNR de 10dB para o SD, e análise por: a) *flops*; b) tempo computacional.

Resultados numéricos de simulação foram obtidos para as seguintes figuras de mérito: curvas de convergência, desempenho BER e complexidade computacional, considerando sistemas 4×4 4-QAM e 12×12 4-QAM. O objetivo foi analisar quais técnicas são capazes de alcançar o desempenho MinBER, sob a condição de elevado número de antenas, bem como registrar o número de iterações necessárias para isso (A.2.Tab.2).

A figura 2.8 mostra como as técnicas MinBER-MuT têm seu desempenho melhorado com o aumento da dimensão do sistema, além de se tornarem mais vantajosas em relação à técnica MuT linear MMSE. Pode-se notar que o ganho de SNR, para uma BER não codificada alvo de 10^{-3} (geralmente suficiente para sistemas aplicando códigos corretores de erro), passa de 3,5dB, na configuração MIMO 4×4, para ≈ 8dB, nas configurações MIMO 12×12 e 20×20. Isso implica que o esquema de precodificação MinBER-MuT além de possibilitar a operação de um sistema com maiores capacidades, reduz a potência necessária para uma mesma confiabilidade, aumentando consideravelmente sua eficiência energética. Adicionalmente, a figura indica as ordens de diversidade alcançadas, e evidencia outra vantagem da abordagem MinBER em relação à MMSE com o aumento da dimensão do sistema: enquanto a ordem de diversidade para a técnica MMSE aumenta de forma lenta para maiores dimensões, para a técnica MinBER esse

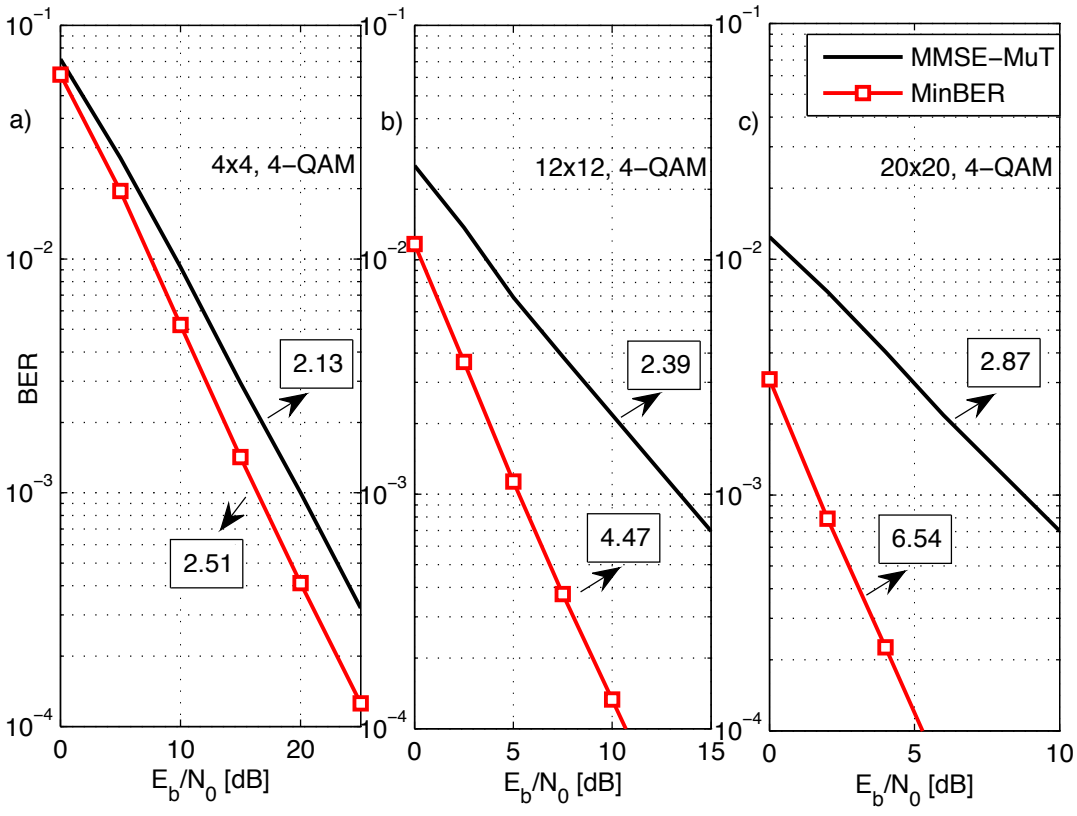

Figura 2.8: Desempenho MuT-MIMO em função da SNR, considerando: a) 4×4, 4-QAM; b) 12×12, 4-QAM; c) 20×20, 4-QAM. Os valores indicados nos retângulos mostram o ganho de diversidade para cada curva.

aumento ocorre de forma bem mais notável, indo de 2,51 em 4×4, 4-QAM, para 4,47 em 12×12, 4-QAM, e finalmente para 6,54 em 20×20, 4-QAM.

A mesma conclusão pode ser tomada analisando-se a figura 2.9, a qual mostra o desempenho do sistema, em termos de BER, para dimensão de sistema crescente ($N = K$), mantendo fixa a SNR do sistema em 6dB. A partir da figura pode-se concluir que os ganhos de desempenho da técnica MinBER-MuT em relação à linear MMSE-MuT se tornam cada vez maiores conforme cresce a dimensão do sistema. A melhoria de desempenho da técnica MMSE pode ser justificada a partir de uma analogia[3] aos resultados de (POOR; VERDU, 1997), onde é demonstrado que a distribuição do termo correspondente à MUI à saída de um detector MMSE converge para uma distribuição gaussiana conforme aumenta o número de usuários em um sistema de espalhamento espectral por sequência direta. Como demonstrado em (POOR; VERDU, 1997), a fidelidade da aproximação gaussiana para a interferência à saída do detector MMSE resulta em uma superioridade em termos de probabilidade de erro de bit em relação a outros detectores lineares. Embora tanto o esquema MMSE quanto o esquema MinBER sejam ótimos

[3] A analogia sugerida consiste em *uplink*/*downlink*, detecção/precodificação, diversidade de código/diversidade de espaço.

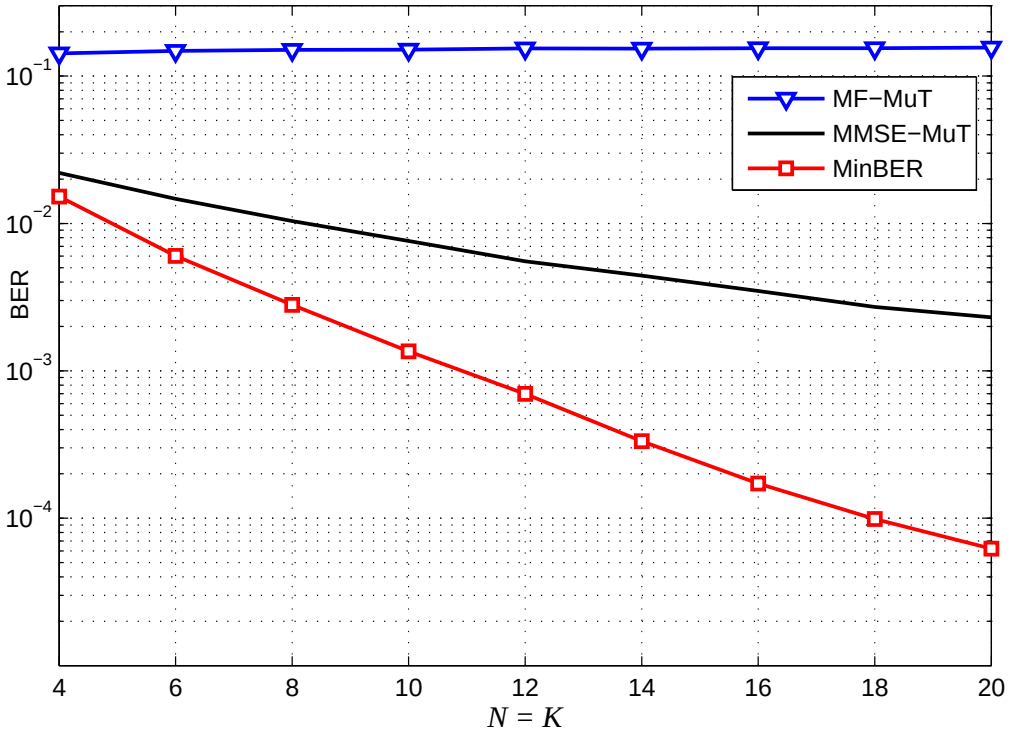

Figura 2.9: Desempenho MuT-MIMO em função da dimensão do sistema, considerando $N = K$, 4-QAM.

quanto a seus respectivos critérios de optimalidade, a superioridade da abordagem MinBER nas figuras 2.8 e 2.9 se deve ao fato do desempenho estar sendo avaliado justamente em termos da BER.

A análise de complexidade computacional das técnicas MinBER-MuT é mostrada na figura 2.10. Pode-se notar que as técnicas de otimização determinísticas apresentaram complexidades bastante reduzidas quando comparadas às heurísticas, mostrando-se mais eficientes para o problema investigado. Focando a análise para as de menor complexidade, notamos que para reduzidas dimensões do sistema ($N < 5$) a técnica SQP se mostra mais eficiente, enquanto que para dimensões moderadas ($5 \leq N < 12$) isso é alcançado pelo esquema LSQN-EN de (HABENDORF; FETTWEIS, 2007). No entanto, para dimensões elevadas do sistema ($N > 12$), a técnica que atingiu o melhor compromisso desempenho×complexidade foi a de busca em linha quase-Newton - função penalidade (*line search quasi-Newton - penalty function*) (LSQN-PF) proposta neste trabalho de Dissertação, uma vez que atinge a referência de desempenho MinBER ao custo de reduzida complexidade computacional.

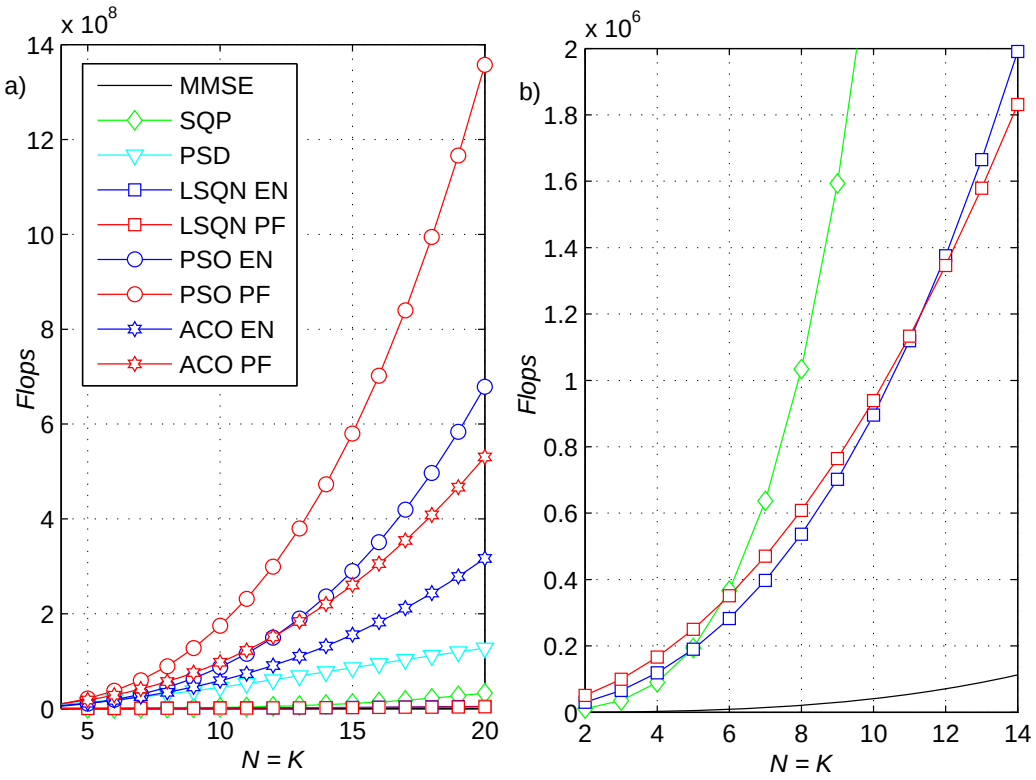

Figura 2.10: Complexidade MuT-MIMO em função da dimensão do sistema, considerando $N = K$: a) Todas as técnicas investigadas; b) Detalhe mostrando apenas as de menores complexidade computacional.

2.3 Análise da BER e Alocação de Pilotos em Sistemas *Massive* MIMO

Sabe-se que o desempenho de sistemas multi-celulares TDD não-cooperativos com ilimitado número de antenas na BS, conforme modelo descrito em A.3.Sec.2, apresenta-se saturado devido ao fenômeno da contaminação de pilotos. Diversos trabalhos abordaram essa limitação de desempenho presente em tais sistemas, geralmente tendo como métrica de desempenho a capacidade total do sistema multi-celular. Em contrapartida, a análise sugerida neste trabalho, conforme descrição do Apêndice A.3, adota como critério de desempenho a BER. Sob este ponto de vista, nossos resultados numéricos mostram que o desempenho de sistemas multicelulares também se torna limitado quando o fenômeno da contaminação de pilotos é levado em conta.

No trabalho em questão, adota-se um cenário multi-celular, em que cada uma das $L = 7$ células hexagonais possui uma BS equipada com N antenas ao seu centro, servindo K usuários móveis portando um MT de única antena cada, uniformemente distribuídos em seu interior. Por simplicidade de análise, porém sem perda de generalidade, assume-se modulação 4-QAM, SNR de 10dB, tanto

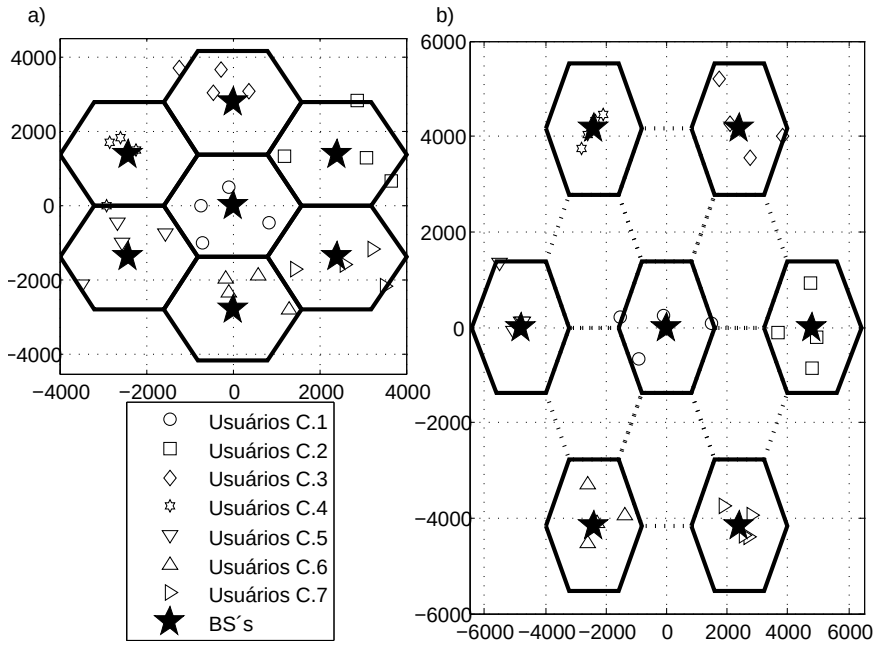

Figura 2.11: Exemplo de realização da configuração espacial multi-celular adotada, para $\tau = 4$. a) Fator de reuso igual a 1; b) Fator de reuso igual a 3.

para *downlink* como para *uplink*; investigou-se os cenários com fator de reuso em frequência (RF) igual a 1 e igual a 3. Uma realização do cenário adotado é mostrada na figura 2.11. Para tornar a simulação mais realista, apenas os parâmetros de desempenho referentes à célula central são considerados, uma vez que seus usuários e sua BS experimentam uma condição de interferência intercelular mais realista.

Uma vez que em (MARZETTA, 2010) e (FERNANDES; ASHIKHMIN; MARZETTA, 2013) o *downlink* do sistema MIMO massivo foi investigado considerando a técnica de precodificação TMF, o presente trabalho analisou o comportamento do sistema considerando a técnica de precodificação ZF. Em (RUSEK et al., 2013, Fig. 11), os autores mostram que a técnica TMF supera a ZF na condição limite de $N \to \infty$. No entanto, a restrição de potência de transmissão limitada na BS é desprezada pelos autores, que justificam que na condição assintótica os valores reais de potência não são relevantes. Em (FERNANDES; ASHIKHMIN; MARZETTA, 2013) essa hipótese é desmentida, e os autores demonstram que o efeito dessa restrição continua presente mesmo no cenário MIMO massivo, e expressões mais precisas para o desempenho do sistema são derivadas. Tendo em mente esse modelo mais realista, no trabalho do Apêndice A.3 é demonstrado que a técnica de precodificação ZF converge na condição de infinitas antenas na BS para o mesmo vetor de precodificação do esquema TMF. Portanto, os mesmos resultados encontrados em (FERNANDES; ASHIKHMIN; MARZETTA, 2013) para a precodificação TMF

também são válidas para o esquema ZF. A figura 2.12 mostra o desempenho das duas técnicas de precodificação com o aumento do número de antenas na BS, e assim comprova este fato.

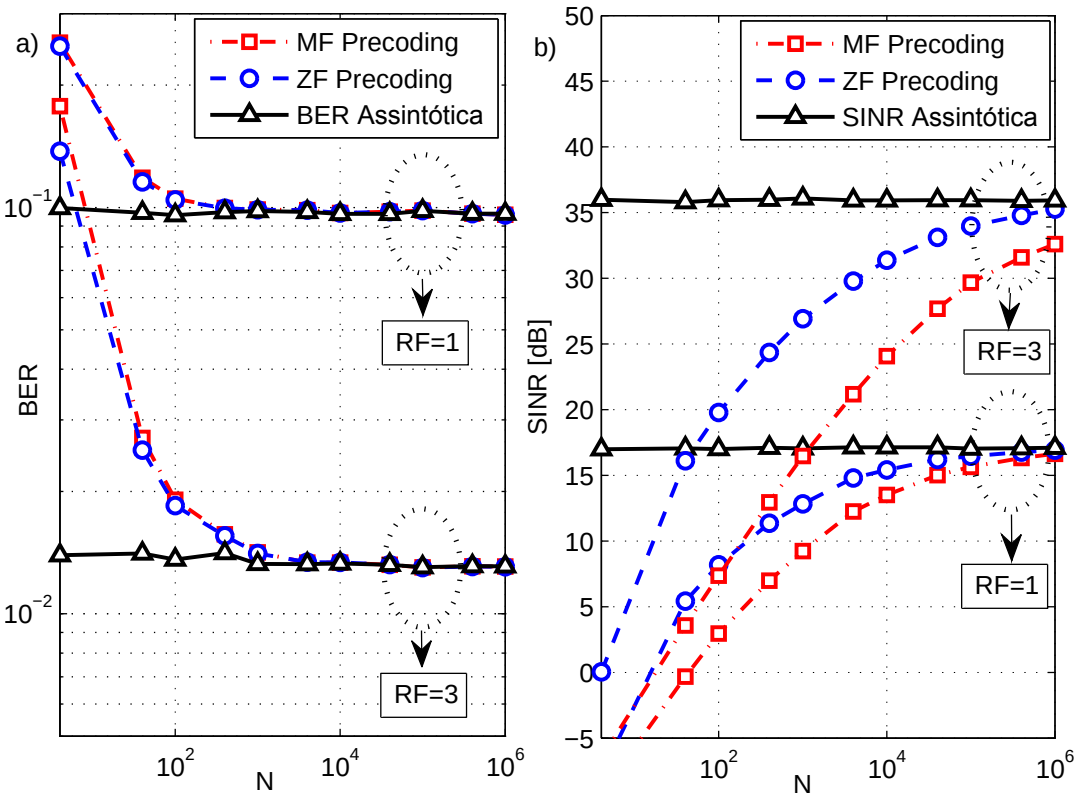

Figura 2.12: Desempenho das técnicas de precodificação MIMO em função do número de antenas na BS, considerando 4-QAM e fatores de reuso em frequência iguais a 1 e 3. a) BER; b) SINR.

Além disso, o desempenho do sistema MIMO massivo é investigado sob a métrica da BER. Analisando o sinal recebido pelos usuários no *downlink*, nota-se que ele é composto predominantemente pelo sinal desejado acrescido de interferências provenientes das células vizinhas. Tais interferências são resultantes do fenômeno da contaminação de pilotos. De fato, como a estimativa do canal de certo usuário possui uma parcela de "contaminação" dos usuários das células vizinhas que reutilizam a mesma sequência de treinamento, parte do sinal que é direcionado no *downlink* para esse usuário também será direcionado aos usuários que compartilham a sequência piloto nas células adjacentes. Esse termo de interferência direcionada pela técnica de precodificação, ou *beamforming*, não desaparece na condição de $N \to \infty$; pelo contrário, se torna mais significativo uma vez que aumenta-se a capacidade de direcionamento da técnica de precodificação. Assim, considerando uma modulação 4-QAM, determina-se no trabalho do Apêndice A.3 a probabilidade de erro para cada usuário, derivando-se expressões da BER nesse cenário. Essas probabilidades de erro são expressas em

função das potências e dos coeficientes de desvanecimento de larga-escala dos usuários daquela sequência de treinamento. A figura 2.12 comprova a validade da expressão derivada, descrita em A.3.Eq.17.

Uma vez que o sinal recebido por determinado usuário no *downlink* do sistema MIMO massivo sofre interferência apenas dos usuários que reutilizam a mesma sequência piloto, este trabalho propõe uma metodologia de otimização do sistema que consiste em distribuir de forma eficiente as sequências de treinamento entre os usuários da célula. É demonstrado que a interferência proveniente das células adjacentes para uma dada sequência piloto depende também do usuário que a utiliza na célula em que se procede a otimização, ou seja, alocando de forma eficiente as sequências piloto é possível melhorar o desempenho do sistema, sob diferentes perspectivas. Um certo grau de coordenação entre as células é necessário para a metodologia de otimização proposta, uma vez que informações como coeficientes de desvanecimento de larga escala e potência dos usuários devem ser compartilhadas. No entanto, esse requisito não é tão problemático de ser atendido, pois tais informações são válidas por um tempo relativamente longo, não aumentam de acordo com o número de antenas na BS, e o número de usuários é limitado pelo tempo de coerência do canal. São propostas diversas métricas de otimização: maximizar a SINR média da célula, minimizar a BER média, maximizar a mínima SINR, e minimizar a máxima BER. Todas essas técnicas são avaliadas sob diferentes perspectivas nas figuras 2.13, 2.14, e 2.15, as quais consideram um número de antenas na BS tendendo a infinito.

A figura 2.13 traz a função de distribuição cumulativa (*cumulative distribution function*) (CDF) da BER entre os usuários da célula, para diferentes fatores de reuso em frequência. A partir do gráfico, pode-se observar um comportamento interessante da BER do *downlink* do sistema MIMO massivo. Enquanto uma parcela dos usuários estabelece sua comunicação livre de erros, i.e., $BER = 0$, outra parcela se comunica com desempenho precário em termos de BER ($BER \approx 10\%$). Essa disparidade se torna mais agravada para fator de reuso unitário, quando a parcela de usuários com elevada BER se torna mais notável. No entanto, a figura 2.13 ressalta uma grande vantagem alcançada pelas técnicas de alocação de pilotos, as quais são capazes de aumentar significantemente a parcela de usuários que se comunicam livre de erros, para ambos fatores de reuso investigados. Consequentemente, elas também são capazes de reduzir a parcela de usuários com desempenho degradado em termos de BER. Conforme desenvolvido no Apêndice A.3, a comunicação livre de erros estabelecida para parte dos usuários se justifica pelo fato da pior condição de interferência que possa vir a

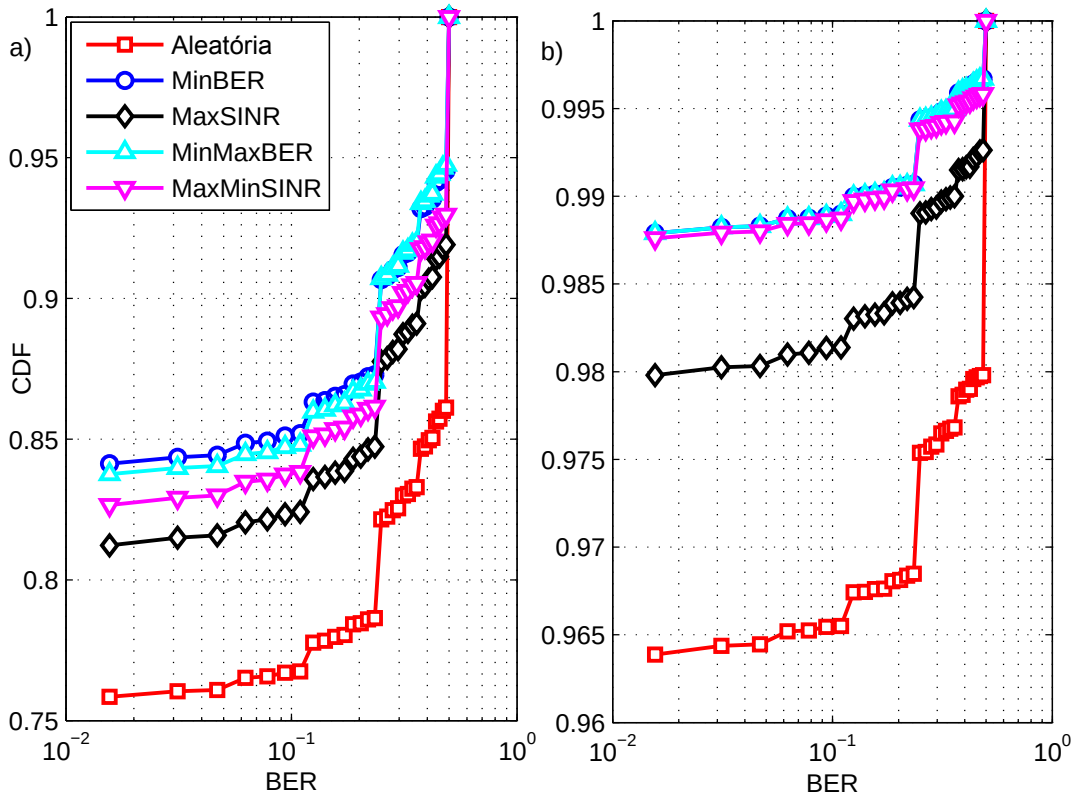

Figura 2.13: Distribuição cumulativa da BER entre os usuários para as diferentes estratégias de alocação de pilotos, considerando 4-QAM. a) Fator de reuso unitário; b) Fator de reuso igual a 3.

afetá-los ainda implicar em potência menor que a do sinal de interesse. Assim, lembrando que o efeito do ruído de fundo desaparece na condição de infinitas antenas na BS, a probabilidade de erro no *downlink* é nula para tais usuários.

A figura 2.14 mostra a porcentagem de usuários acima de determinada SINR, para diferentes técnicas de alocação de pilotos e fatores de reuso em frequência. Nota-se que aumentar o fator de reuso tem o efeito de melhorar significativamente a SINR dos usuários, enquanto os formatos e inclinações das curvas não sofrem alterações notáveis, como se tais curvas fossem apenas deslocadas para a direita em aproximadamente 18dB. A técnica que visa maximizar a SINR média da célula resulta em uma melhoria de desempenho para os usuários melhor posicionados, aumentando em 4dB a SINR dos 20% de usuários com melhor desempenho. Por outro lado, a estratégia de maximizar a mínima SINR da célula aumenta em cerca de 10dB a SINR no patamar de 95%, de forma a favorecer os usuários com localizações desfavoráveis à comunicação.

Finalmente, a figura 2.15 mostra a parcela de usuários acima de determinada taxa de dados, tendo em vista diferentes métricas de alocação de pilotos e fatores de reuso. Nota-se que a inclinação das curvas para fator de reuso igual a 3 é

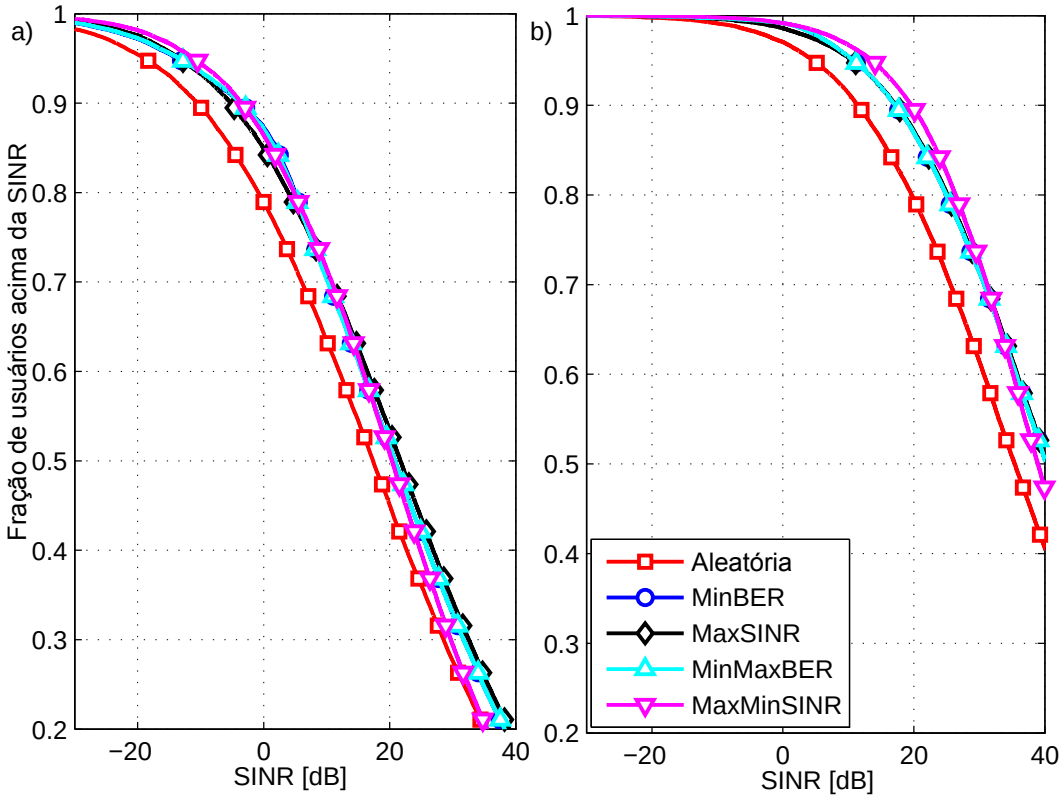

Figura 2.14: Parcela de usuários acima de determinada SINR, para diferentes estratégias de alocação de pilotos. a) Fator de reuso unitário; b) Fator de reuso igual a 3.

muito mais acentuada que as referentes a fator de reuso unitário. Isso significa que as taxas de dados para maior fator de reuso apresentam-se mais uniformemente distribuídas entre os usuários da célula, enquanto para fator de reuso unitário existem maiores disparidades entre as taxas de usuários melhor e pior posicionados. Embora com menor fator de reuso seja alcançada um melhor desempenho médio, tanto em termos de BER como em capacidade, com fator de reuso maior se consegue uma melhor qualidade de serviço para os usuários, a qual tem sido um parâmetro mais visado nos modernos sistemas de comunicação.

Embora todos os usuários se beneficiem do maior fator de reuso em frequência, em termos de SINR, como mostrado na figura 2.14, o mesmo não ocorre em termos de taxa de dados. Este fato se justifica pela relação logarítmica entre capacidade e SINR. Enquanto para os usuários pior posicionados o aumento de SINR resulta em um considerável aumento de capacidade, pois o logaritmo se encontra na região linear, para aqueles bem posicionados o aumento de SINR não compensa a redução de banda devido ao maior fator de reuso.

As Tabelas A.3.Tab.1 e A.3.Tab.2 mostram diversos resultados numéricos alcançados pelas diferentes abordagens de alocação de pilotos, para fator de reuso

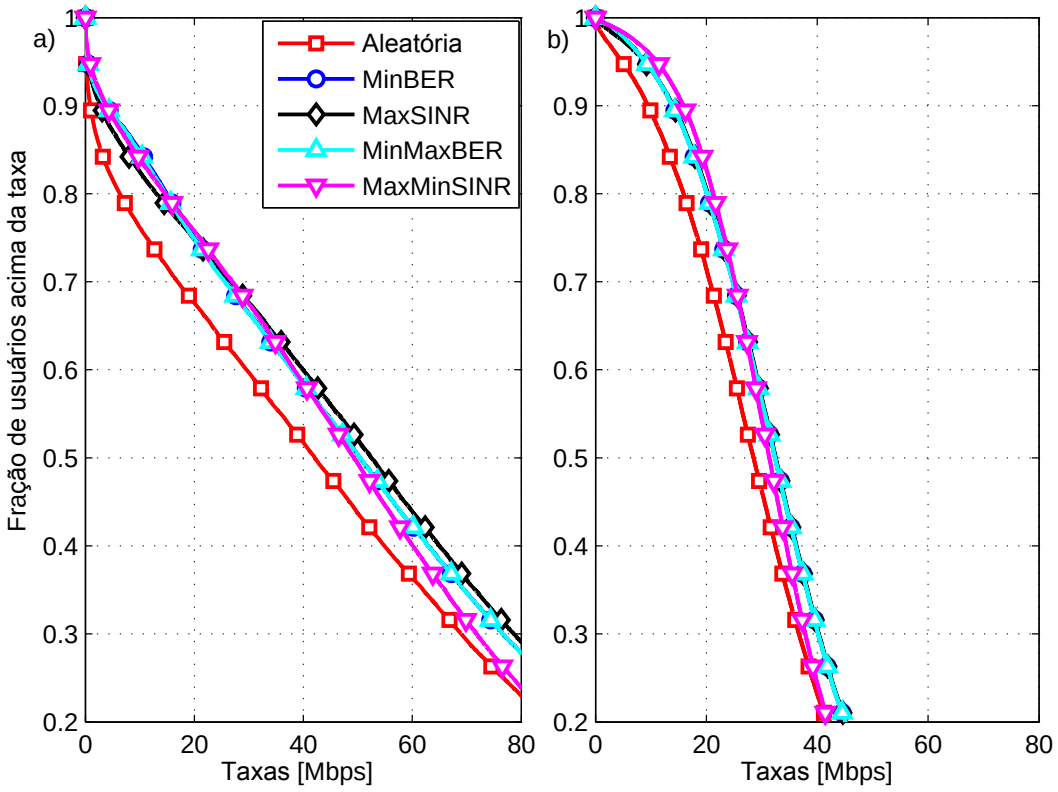

Figura 2.15: Parcela de usuários acima de determinada taxa de dados, para diferentes estratégias de alocação de pilotos. a) Fator de reuso unitário; b) Fator de reuso igual a 3.

um e três, respectivamente. Dentre os resultados alcançados, podemos destacar a taxa de dados alcançada com probabilidade maior que 95% para fator de reuso unitário, que passa de 0,1344Mbps para 0,7937Mbps, enquanto é obtida uma capacidade média de 52,62Mbps. A parcela de usuários com $BER > 0.1$ é reduzida de 23,63% para 14,92% nesse cenário quando estratégias de alocação de pilotos são empregadas. Já para fator de reuso igual a 3, essa parcela de usuários é reduzida de 3,47% para 1,06%, enquanto as capacidades de probabilidade maior que 95% aumentam de 4,79Mbps para 11,15Mbps. Portanto, quando combinadas convenientemente as estratégias de alocação de pilotos e reuso de frequências, pode-se implementar um sistema de comunicação em que 98,78% dos usuários se comunicam com BER consideravelmente reduzidas, tendo asseguradas uma capacidade de 11,15Mbps, e uma capacidade média de 31,68Mbps. Desempenhos ainda melhores podem ser alcançados combinando técnicas de alocação de potência, como aquelas de (FERNANDES; ASHIKHMIN; MARZETTA, 2013) e (RASTI; SHARAFAT, 2011). A aplicação/combinação de tais técnicas permanece como possível continuação deste trabalho.

Quanto à complexidade computacional do esquema de alocação de pilotos, nota-se que, ao realizar uma busca exaustiva entre todas as combinações de

sequências entre os usuários, a complexidade das técnicas cresce de forma fatorial com relação ao número de usuários dentro de determinada portadora na etapa de treinamento[4]. No entanto, o número de usuários dentro de determinada portadora na etapa de treinamento é baixo em sistemas práticos, pois é limitado pelo tempo de coerência do canal, pela mobilidade dos usuários, e pela banda também limitada. Assim a complexidade computacional do esquema proposto não constitui um fator limitante para sua implementação. Vale ressaltar também que após encontrado uma combinação de alocação de pilotos que otimiza certa métrica, esta permanece válida enquanto as potências e os coeficientes de desvanecimento de larga-escala dos usuários forem os mesmos, o que corresponde a um intervalo de tempo relativamente grande em sistemas de comunicação de interesse prático.

[4]Lembrando que vale a relação $K = \tau \cdot \mathtt{N}_{\mathrm{smooth}}$, em que K é o número real de usuários por célula, τ é o comprimento das sequências, que corresponde ao número de usuários em cada subportadora na etapa de treinamento, e $\mathtt{N}_{\mathrm{smooth}}$ é o número de portadoras dentro de uma banda de coerência do canal.

3 Conclusões e Trabalhos Futuros

Podemos concluir que trabalhar com um sistema MIMO de elevada dimensão traz muitos benefícios, mas também pode impor muitos desafios. Para os casos em que o número de usuários cresce na mesma proporção de N, os benefícios incluem maximizar sua capacidade, aumentando sua eficiência energética e espectral, enquanto dentre as diversas dificuldades se destacam os problemas de implementação, a exigência de técnicas de decodificação/precodificação sofisticadas, bem como estimar a CSI de forma confiável e computacionalmente factível. Já quando o número de usuários atendidos permanece fixo, temos como benefícios mitigar os efeitos de interferência e ruído de fundo, maximizando as capacidades, e eficiências espectral e energética, e tornar ótimas as técnicas de decodificação/precodificação mais simples, baseadas no conjugado transposto da matriz de coeficientes de propagação. Como dificuldades para esses cenários temos as dificuldades de implementação, e como combater de forma eficiente o efeito de contaminação de pilotos.

Para o problema de detecção MIMO em canais correlacionados e grande número de antenas no transmissor e receptor, foi proposto um esquema que combina as técnicas ACO e LR, o qual se mostrou mais eficiente em termos do compromisso desempenho×complexidade do que as outras técnicas baseadas em ACO investigadas, para sistemas com elevado número de antenas e canais correlacionados. No caso da Transmissão Multiusuário em MIMO, foi derivada uma formulação convexa para o problema da precodificação visando a minimização da BER, a partir da qual diversos métodos de otimização foram investigados. Importante observar que a técnica proposta LSQN-PF se mostrou mais eficiente do ponto de vista do compromisso desempenho×complexidade, sob a condição de elevada dimensão. Além disso, mostrou-se que a abordagem MinBER opera de forma muito eficiente para dimensões elevadas do sistema MIMO, considerando $N = K$, sendo seu ganho em termos de SNR, para um mesmo desempenho em relação à técnica MMSE, cada vez maior com a crescente dimensão do sistema.

Por fim, analisou-se sistemas multi-celulares TDD não-cooperativos em que o número de antenas na BS cresce indefinidamente (*Massive* MIMO). Para tais sistemas, foi investigado o desempenho do *downlink* sob a métrica da BER, sendo derivada uma expressão para a probabilidade de erro de bit assintótica dos usuários. Com base na expressão derivada e na SINR assintótica de (FERNANDES; ASHIKHMIN; MARZETTA, 2013), foi proposta uma estratégia de otimização de desempenho do sistema, sob diferentes critérios, chamada de alocação de pilotos. Nossos resultados numéricos mostraram que as estratégias propostas podem alcançar ganhos consideráveis para o sistema MIMO massivo em relação ao critério aleatório de alocação de pilotos, não sendo sua complexidade computacional um fator restritivo para sua implementação em sistemas reais.

3.1 Trabalhos Futuros

Muito ainda deve ser investigado quando se fala em sistemas *Massive* MIMO, uma vez que esse é um tema recente com grande potencial de aplicação em sistemas 5G. No entanto, muitas dificuldades devem ainda ser superadas a fim de amadurecer a tecnologia.

Quando falamos em detecção, interessantes resultados foram divulgados em (VARDHAN et al., 2008), (SRINIDHI et al., 2011), e trabalhos relacionados, mostrando que as técnicas propostas baseadas em busca ascendente de verossimilhança e busca *Tabu* podem ser eficientemente aplicadas ao problema. Dessa forma, combiná-las de alguma forma com as técnicas ACO e/ou LR pode gerar resultados promissores.

No contexto da Transmissão Multiusuário, foi demonstrado que a abordagem MinBER pode ser aplicada de forma muito eficiente a sistemas unicelulares. Porém, nenhum estudo foi feito ainda, dentro do conhecimento do autor à ocasião do desenvolvimento dessa Dissertação, sobre sua aplicação em sistemas multi-celulares. É de se esperar que seu desempenho seja fortemente degradado, uma vez que a formulação do problema de otimização correspondente não leva em conta nenhuma interferência inter-celular. Nesse ponto, a formulação de um problema de otimização adequado para esse cenário seria muito conveniente, porém o desafio é fazer isso sem o conhecimento da CSI das células adjacentes, uma vez que a disponibilização dessa informação entre todas as células é inviável para $N \to \infty$. Uma formulação semelhante foi derivada em (JOSE et al., 2011), e que talvez possa ser adaptada para esse problema.

A respeito de sistemas *Massive* MIMO, uma vez que o limitante do desempenho se encontra na CSI estimada imperfeitamente, técnicas eficientes para sua obtenção, na condição de elevado número de antenas na BS, podem gerar bons resultados. Adicionalmente, uma vez que o efeito da contaminação de pilotos é obtido atribuindo-se sequências piloto ortogonais entre usuários da mesma célula, nenhum trabalho investigou ainda a aplicação de diferentes tipos de sequências. Por exemplo, se fossem utilizadas sequências aleatórias, haveria uma maior disponibilidade de sequências, de forma que usuários de células adjacentes não deveriam necessariamente usar as mesmas. Isso poderia aumentar a interferência intra-celular, porém diminuiria a interferência inter-celular, principalmente para baixos fatores de reuso em frequência, que é a responsável pela saturação do desempenho assintótico. Esses diferentes tipos de sequências de treinamento também poderiam ser interessantes na condição em que não se pode garantir um sincronismo tão preciso do sistema, pois nesse caso as sequências ortogonais podem resultar em elevados índices de correlação e, consequentemente, interferência intra-celular.

Outra perspectiva de continuação deste trabalho diz respeito a combinar a técnica proposta de alocação de pilotos com as técnicas de deslocamento temporal (FERNANDES; ASHIKHMIN; MARZETTA, 2013) e alocação de potência (FERNANDES; ASHIKHMIN; MARZETTA, 2013; RASTI; SHARAFAT, 2011). Uma vez que os resultados numéricos mostrados no trabalho do Apêndice A.3 foram obtidos com potência de transmissão no *downlink* constante entre os usuários, certamente estes resultados podem ser ainda melhorados explorando-se técnicas de alocação de potência.

Apêndice A – Trabalhos Desenvolvidos

Nesta seção são apresentados os artigos desenvolvidos durante o presente trabalho de Mestrado, tanto o já publicado quanto aqueles ainda em fase de submissão. Ao todo, foram elaborados três artigos científicos no período de desenvolvimento das atividades desta Dissertação.

A.1 Detecção em Sistemas MIMO

Título:	*Lattice Reduction Aided Detector for MIMO Communication Via Ant Colony Optimisation*;
Autores:	José Carlos Marinello & Taufik Abrão;
Publicação:	Novembro de 2013 (*online*);
	Julho de 2014 (impressa);
Revista:	*Wireless Personal Communication*.

Wireless Pers Commun
DOI 10.1007/s11277-013-1495-z

Lattice Reduction Aided Detector for MIMO Communication Via Ant Colony Optimisation

José Carlos Marinello · Taufik Abrão

Abstract In this work heuristic ant colony optimisation (ACO) procedure is deployed in conjunction with lattice reduction (LR) technique aiming to improve the performance-complexity tradeoff of detection schemes in MIMO communication. A hybrid LR-ACO MIMO detector using the linear minimum mean squared error (MMSE) criterion as initial guess is proposed and compared with two other traditional (non)linear MIMO detectors, as well as with heuristic MIMO detection approaches from the literature, in terms of both performance and complexity metrics. Numerical results show that the proposed LR-ACO outperforms the traditional ACO-based MIMO detectors and the ACO detector with the MMSE solution as initial guess, with a significant complexity reduction while is able to reach full diversity degree in all scenarios considered, including different channel correlation levels, modulation orders, and antennas configuration.

Keywords Large-MIMO systems · Ant colony optimisation · Evolutionary computing · Lattice reduction · Maximum-likelihood estimation · MMSE · Diversity order

1 Introduction

Multiple-input-multiple-output (MIMO) systems are known by providing a significant spectral efficiency improvement on wireless communication systems [5,20]. Therefore, in the last decade many researches in this field have been carried out due to the new wireless system requirements such as higher data rates and system reliability, imposed by the recent telecommunications services and technologies, allied to large interest on saving resources,

Part of this work has been presented in the IEEE-WCNC'13 Conference.

T. Abrão (✉) · J. C. Marinello
Department of Electrical Engineering, State University of Londrina, Londrina, Brazil
e-mail: taufik@uel.br

J. C. Marinello
e-mail: zecarlos.ee@gmail.com

Published online: 12 November 2013

like spectrum and power. Furthermore, exploiting the path diversity between the multiple-antenna transmitter and multiple-antenna receiver, another MIMO configuration allows to improve the link reliability for a given data rate, or maximise the data rate given a certain reliability requirement [22]. It is also known that very high data rates can be achieved when using a large number of antennas and/or modulations orders (Large MIMO), and that on these cases the detection task becomes challenging [18], since there is a considerable enhancement on the inter layer/antenna interference (ILI), and/or a lower noise robustness, requiring thus more sophisticated and efficient MIMO detection techniques.

Among the linear detectors widely known the zero forcing (ZF) and the minimum mean squared error (MMSE) based techniques present low complexity and the ability to operate under ill-conditioned channel matrices; however both linear MIMO detection techniques is clearly inferior in terms of performance to that achieved by the maximum likelihood (ML) detector. Recently, the sphere decoding (SD) approach [25] has becoming an alternative to the ML detector, presenting a near-optimum performance; however SD approach results in a prohibitive complexity of implementation under low or medium signal-noise ratio (SNR) regions on real communication systems; indeed, under low SNRs, the SD complexity becomes exponential with the modulation order and the number of antennas [9], i.e., the same complexity order of the ML detector.

An appealing procedure to improve the MIMO linear detectors performance under correlated channels, and simultaneously resulting in a feasible complexity, consists in deploying the lattice reduction (LR) technique [26,27]. This technique transforms the partial correlated channel into an equivalent one with a better conditioned channel matrix. Having a near-orthogonal near-uncorrelated channel transformation implemented, at the receiver side the detection can be carried out easily deploying a low-complexity MIMO detector scheme.

The ant colony optimisation (ACO) is a technique originally proposed for combinatorial optimisation problems, such as the traveling salesman problem [3]. It is inspired on the behaviour of ants in nature, in which looking for food, and having the ability of exploitation of the territory around, find ways that lead to the best sources of food, and leave trails that will help other ants on their searches. Many technical works have been disseminated applying this technique to several combinatorial (discrete) and continuous optimisation problems that arise in telecommunications, such as detection in code division multiple access systems (CDMA) [14,29], detection in MIMO systems [10,12], resource allocation in wireless networks [8,4], pre-coding for MIMO systems [17], among others. Performance and complexity of several heuristic population-based algorithms, including ACO, have been analysed in [23] and [11]; numerical results have corroborated the best ACO performance-complexity tradeoff. A particular advantage of ACO is the pheromone learning mechanism, in which pheromone is deposited on the best trails as the algorithm evolves. The amount of pheromone in a given path can be seen as the probability of it be the shortest way to the food source, or, in the MIMO detection context, the probability of a certain data vector be the actual transmitted one. This soft-decision guided search enables the ACO algorithm to find near optimal solutions even in very difficult optimisation problems, with high dimension and many local optima. Other algorithms are able to present faster convergence rates, e.g, the artificial bee colony optimisation (BCO) algorithm [19]; however, it is known that the BCO premature convergence feature might be also one of its main drawbacks [24], since limits its exploitation capability.

A simple ant colony optimisation applied to MIMO detection problem has been presented in [10]. The ACO heuristic technique is combined with the zero-forcing and V-BLAST MIMO detectors. However, numerical results show that the proposed ACO-ZF and ACO-BLAST MIMO detectors were not able to achieve ML detector performance

for medium and high SNR regions. In [12], a different detection scheme based on ACO has been proposed; however, since the technique developed in [12] results in a remarkable performance loss compared with ML detector, a modified ACO algorithm is suggested herein to improve the performance-complexity trade-off in a near-optimal MIMO detection approach.

As an alternative to the ACO detector of [10] and [12], in this contribution we propose a near-optimum MIMO detector based on ant colony optimisation heuristic approach coupled with the lattice reduction technique, in which part of the search complexity has been shifted to the initial detection stage, by deploying a low complexity linear detector, which works as a start point in the solution search carried out in a second stage. This way, it will be shown that performing the search on a better conditioned domain provided by the LR technique, it is possible to obtain a significant improvement in the performance $\times$ complexity tradeoff with the LR-ACO MIMO detector even when applied to a large number of antennas, namely dense or large MIMO systems. Besides, to ensure an improvement on the ACO convergence rate, we deploy a careful input parameter optimisation procedure, and propose a diminished search space by means of a low-complexity neighbourhood analysis, as explained in the following sections.

The remainder of this paper is organised as follows: in Sect. 2, the system model and notations are introduced; in Sects. 3 and 3.4, different detection schemes with and without reduction of basis, respectively, are revisited. In Sect. 3.5, the state-of-the-art in ACO-MIMO detection is discussed. The proposed detection schemes are introduced in Sects. 3.6 and 3.7, and the numerical results in 4. Concluding remarks can be found in Sect. 5.

2 MIMO System Model

In the adopted MIMO system it is assumed that there are n_T transmit antennas at the transmitter side, and n_R receive antennas at the receiver. Besides, the transmitted information symbol vector is $\mathbf{x} = [x_1\, x_2\, \ldots\, x_{n_T}]^\top$, where x_i takes a value on the squared quadrature amplitude modulation (M-QAM) alphabet and denotes the transmitted symbol at the ith antenna, and $\{\cdot\}^\top$ is the transpose operator; so, the complex-valued symbol (finite) set is given by $\mathcal{S} = \left\{\mathcal{A} + \sqrt{-1}\cdot\mathcal{A}\right\}$, where the real-valued finite set $\mathcal{A} = \left\{\pm\frac{1}{2}a;\ \pm\frac{3}{2}a;\ \ldots;\ \pm\frac{\sqrt{M}-1}{2}a\right\}$, with $\sqrt{M}$ representing the modulation order (per dimension) of the corresponding real-valued amplitude shift keying (ASK) modulation scheme. The parameter $a = \sqrt{6/(M-1)}$ is used for normalizing the power of the complex valued transmit signals to 1. Furthermore, we have assumed that the system is determined, i.e., $n_R \geq n_T$.

For simplicity of analysis, the received complex-valued signal over a MIMO channel is written using matrix notation as:

$$\mathfrak{r} = \mathcal{H}\mathfrak{x} + \mathfrak{n} \tag{1}$$

where $\mathfrak{x}$ denotes the complex valued $n_T \times 1$ transmit signal vector, the corresponding receive signal vector $\mathfrak{r}$ has dimension $n_R \times 1$; besides, $\mathfrak{n} \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$ represents the additive white Gaussian noise (AWGN) with variance $\sigma_n^2 = N_0$, which is observed at the n_R receive antennas while the average transmit power of each antenna is normalised to one, i.e. $\mathbb{E}[\mathfrak{x}\mathfrak{x}^H] = \mathbf{I}_{n_T}$, and $\mathbb{E}[\mathfrak{n}\mathfrak{n}^H] = N_0\mathbf{I}_{n_R}$, thus the bit energy to noise power spectral density ratio is given by $\xi = \frac{E_b}{N_0} = \frac{n_R}{\sigma_n^2 \cdot \log_2 M}$. Furthermore, the $n_R \times n_T$ complex-valued channel matrix $\mathcal{H}$ was assumed uncorrelated complex Gaussian fading gains with unit variance. Furthermore,

non-selective and slow fading environment was assumed, in which the channel coherence time is much larger than the symbol period, leading to a channel coefficient approximately constant over the symbol period, which changes independently from symbol-by-symbol period;[1] besides, the channel coherence bandwidth much larger than the system bandwidth has been assumed.

In order to facilitate the numerical analysis, real and imaginary part of (1) are treated separately; so, the system model can be rewritten as:

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2}$$

with the real-valued channel matrix

$$\mathbf{H} = \begin{bmatrix} \Re\{\boldsymbol{\mathcal{H}}\} & -\Im\{\boldsymbol{\mathcal{H}}\} \\ \Im\{\boldsymbol{\mathcal{H}}\} & \Re\{\boldsymbol{\mathcal{H}}\} \end{bmatrix} \in \mathbb{R}^{n\times m} \tag{3}$$

and the real-valued vectors

$$\mathbf{r} = \begin{bmatrix} \Re\{\mathfrak{r}\} \\ \Im\{\mathfrak{r}\} \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} \Re\{\mathfrak{x}\} \\ \Im\{\mathfrak{x}\} \end{bmatrix}; \quad \mathbf{n} = \begin{bmatrix} \Re\{\mathfrak{n}\} \\ \Im\{\mathfrak{n}\} \end{bmatrix} \in \mathbb{R}^{n} \tag{4}$$

where $\mathbf{r}, \mathbf{n} \in \mathbb{R}^n$, $\mathbf{x} \in \mathcal{A}^m$, $m = 2n_T$ and $n = 2n_R$. Note that now, the information vector assumes values only over the finite set of real-valued: $\mathbf{x} \in \mathcal{A}^{2n_T}$, where $\mathcal{A}$ is the set of real-valued entries in the signal constellation, e.g., $\mathcal{A} = \frac{a}{2} \cdot \{\pm 1, \pm 3, \pm 5 \pm 7\}$ in a 64-QAM signaling.

2.1 MIMO Correlated Channels

This subsection discusses the MIMO channel correlation among different antennas. A baseband discrete-time representation is considered. Now, considering the shadowing effect defined by the coefficients χ, the MIMO channel response (one channel realisation) for the kth antenna consists of:

$$\underline{\mathfrak{h}}^k[t] = \chi^k \cdot \mathfrak{h}^k \tag{5}$$

where $\mathfrak{h}^k$ represents an element of the uncorrelated complex-valued channel matrix $\boldsymbol{\mathcal{H}}$ complex Gaussian fading gains defined previously; while χ^k represents the log-normal shadowing term, given by:

$$\chi^k = 10^{\frac{\sigma_x}{20} w^k}. \tag{6}$$

The shadowing is associated with the Gaussian random variable (r.v.) w^k, where $\sigma_x = 3$ to 6 dB stands for the log-normal shadowing standard deviation.

A recent modification in IEEE 802.15.3a channel model that includes the correlation factor on shadowing factor χ^k among different antennas in a MIMO UWB system was proposed in [30]. In [21], the shadowing correlation between two transmit antennas was considered for the analysis of a MISO DS-UWB system. Let $\boldsymbol{\chi} = \left[\chi_1 \ \chi_2 \cdots \chi_{n_T}\right]^\top$ be the log-normal shadowing r.v.vector with transmit correlation matrix, given by $\mathbf{R}_{\boldsymbol{\chi}} = \left[\rho_{\chi^k,\chi^j}\right]_{n_T\times n_T}$, where ρ_{χ^k,χ^j} represents the spatial correlation coefficient between the kth and jth transmit antenna elements. The idea in [30] was to find a relation between the correlation coefficient for

[1] Block fading channel assumption.

the log-normal r.v.s χ^k, and their corresponding Gaussian distributed variables w^k. Such a relation is given by

$$\rho_{w^k,w^j} = \frac{1}{\xi^2\sigma_x^2} \ln\left\{\left(e^{\xi^2\sigma_x^2} - 1\right)\rho_{\chi^k,\chi^j} + 1\right\} \tag{7}$$

where $\xi = \ln(10)/20$.

After this transformation, the corresponding Gaussian correlation matrix $\mathbf{R}_w = \left[\rho_{w^k,w^j}\right]_{n_T \times n_T}$ is generated. Hence, the correlated channel r.v. vector, $\widetilde{\mathbf{h}} = \left[\widetilde{h}_1\ \widetilde{h}_2\ \cdots\ \widetilde{h}_{n_T}\right]^\top$, is obtained as

$$\widetilde{\mathbf{h}} = \mathbf{R}_h^{1/2}\mathbf{h}, \tag{8}$$

where $\mathbf{h}$ is a random real-valued i.i.d. vector, for instance, the column of real-values Gaussian distributed channel coefficients $\mathbf{H}$ matrix. Thus, the components of the corresponding log-normal correlated shadowing vector are obtained according to (6). As an example and based on the measurement campaign in [30], the correlation matrix for a three antenna MIMO-UWB system case [1] is obtained as

$$\mathbf{R}_\chi = \begin{bmatrix} 1 & 0.86 & 0.54 \\ 0.86 & 1 & 0.86 \\ 0.54 & 0.86 & 1 \end{bmatrix}. \tag{9}$$

A simple single-parameter correlation model proposed in [28] and adopted in [7, Sec.VI.B] assumes that the correlation between the transmit antennas is independent of receive antennas and vice-versa. The MIMO channel is modelled as follows:

$$\widetilde{\mathcal{H}} = \mathbf{R}_{H,Rx}^{1/2}\mathbf{G}\mathbf{R}_{H,Tx}^{1/2} \tag{10}$$

where $\mathbf{G}$ is a $n_R \times n_T$ matrix with i.i.d. complex Gaussian zero mean unit variance elements, $\mathbf{R}_{H,Tx}$ is $n_T \times n_T$ transmit correlation matrix, and $\mathbf{R}_{H,Rx}^{1/2}$ is $n_R \times n_R$ receive correlation matrix. The transmit and receive correlation matrices are modelled in the same way [7]. Hence, both correlation matrices sharing the same structure:

$$\mathbf{R}_{\mathrm{H,\ TX}} = \begin{bmatrix} 1 & \rho_t & \rho_t^4 & \cdots & \rho_t^{(n_T-1)^2} \\ \rho_t & 1 & \rho_t & \cdots & \rho_t^{(n_T-2)^2} \\ \rho_t^4 & \rho_t & 1 & \cdots & \vdots \\ \vdots & \vdots & & \ddots & \rho_t \\ \rho_t^{(n_T-1)^2} & \rho_t^{(n_T-2)^2} & \cdots & \rho_t & 1 \end{bmatrix} \tag{11}$$

If one consider similar environment for transmitter and receiver antennas, for instance indoor wireless LAN setting, then correlation elements $\rho_t = \rho_r = \rho$ could be assumed. In the numerical results section, we have analysed spatially correlated channel in the following scenarios:

$$\rho = \begin{cases} 0.0 : \text{uncorrelated channels;} \\ 0.2 : \text{weakly correlated channels;} \\ 0.5 : \text{medianly correlated channels;} \\ 0.9 : \text{strongly correlated channels.} \end{cases}$$

3 MIMO Detectors

3.1 ML MIMO Detector

The maximum likelihood detector operates by searching the symbol into the set $\mathcal{A}^{2n_T}$ that maximises the likelihood function, i.e.

$$\widehat{\mathbf{x}}_{\text{ML}} = \arg \max_{\mathbf{x}\in\mathcal{A}^{2n_T}} f(\mathbf{x}|\mathbf{r}) \tag{12}$$

where $f(\mathbf{x}|\mathbf{r})$ denotes the likelihood function of $\mathbf{x}$ for a given $\mathbf{r}$. Since the noise in (2) is assumed to be circularly symmetric complex Gaussian (CSCG), the ML detection becomes:

$$\begin{aligned}\widehat{\mathbf{x}}_{\text{ML}} &= \arg \min_{\mathbf{x}\in\mathcal{A}^{2n_T}} ||\mathbf{r}-\mathbf{H}\mathbf{x}||^2 \\ &= \arg \min_{\mathbf{x}\in\mathcal{A}^{2n_T}} (\mathbf{r}-\mathbf{H}\mathbf{x})^\top \mathbf{R}_n^{-1} (\mathbf{r}-\mathbf{H}\mathbf{x})\end{aligned} \tag{13}$$

where the noise covariance matrix $\mathbf{R}_n$ is assumed i.i.d., given by

$$\mathbf{R}_n = \text{diag}\left(\frac{\sigma_{n,1}^2}{2}, \frac{\sigma_{n,2}^2}{2}, \ldots, \frac{\sigma_{n,2n_R}^2}{2}\right) = \frac{N_0}{2}\mathbf{I}_{2n_R},$$

As shown in (13), the ML detection in MIMO systems is equivalent to an exhaustive search of a combinatorial optimisation problem, which becomes prohibitive when the constellation order and number of antennas increase substantially; for example, if $M = 16$ and $n_T = 4$, the number of candidates to be evaluated becomes astonishingly large $M^{n_T} \approx 6.5 \cdot 10^4$.

3.2 Linear Zero-Forcing MIMO Detector

In a conventional zero-forcing detector, the interference is completely suppressed by multiplying the receive signal vector $\mathbf{r}$ with the Moore-Penrose pseudo-inverse of the channel matrix:

$$\mathbf{H}^\dagger = \left(\mathbf{H}^\top\mathbf{H}\right)^{-1}\mathbf{H}^\top \tag{14}$$

The decision step consists in mapping each element of the filter output vector. Finally, the conventional ZF MIMO detector output is given by:

$$\widehat{\mathbf{x}}_{\text{ZF}} = \mathbf{H}^\dagger\mathbf{r} = \mathbf{H}^\dagger(\mathbf{H}\mathbf{x}+\mathbf{n}) = \mathbf{x}+\mathbf{H}^\dagger\mathbf{n} \tag{15}$$

i.e., the linear transformation matrix for zero-forcing is given by the Moore-Penrose pseudo-inverse matrix: $\mathbf{W}_{\text{ZF}} = \mathbf{H}^\dagger = \left(\mathbf{H}^\top\mathbf{H}\right)^{-1}\mathbf{H}^\top$.

3.3 Linear MMSE MIMO Detector

The conventional MMSE for MIMO system is another linear detector, whose preprocessor output is given by:

$$\widehat{\mathbf{x}}_{\text{MMSE}} = \mathbf{W}_{\text{MMSE}}^\top\mathbf{r} = \mathbf{H}\left(\frac{N_0}{E_s}\mathbf{I}+\mathbf{H}^\top\mathbf{H}\right)^{-1}\mathbf{r} \tag{16}$$

where $E_s = E_b \log_2 M$ is the energy per symbol.

Note that both ZF and MMSE MIMO detectors can also be derived by the following unified approach. Let us consider the *extended real-valued channel matrix* as follows:

$$\mathbf{H}_{\text{EXT}} = \begin{bmatrix} \mathbf{H} \\ \kappa \mathbf{I}_{2n_T} \end{bmatrix} \tag{17}$$

where $\kappa \geq 0$ is a constant. Hence, the pseudo-inverse of $\mathbf{H}_{\text{EXT}}$ is obtained applying (17) in (14):

$$\mathbf{H}_{\text{EXT}}^{\dagger} = \left(\kappa^2 \mathbf{I}_{2n_T} + \mathbf{H}^{\top}\mathbf{H}\right)^{-1} \begin{bmatrix} \mathbf{H}^{\top} & \kappa \mathbf{I}_{2n_T} \end{bmatrix} \tag{18}$$

Let us take

$$\left[\mathbf{H}_{\text{EXT}}^{\dagger}\right] = \left(\kappa^2 \mathbf{I}_{2n_T} + \mathbf{H}^{\top}\mathbf{H}\right)^{-1} \mathbf{H}^{\top}$$

then, we can re-define the linear transformation matrix for the ZF and MMSE MIMO detectors as follows:

$$\begin{aligned} \mathbf{W}^{\top} &= \left[\mathbf{H}_{\text{EXT}}^{\dagger}\right] \\ &= \begin{cases} \mathbf{W}_{\text{ZF}}^{\top} = \mathbf{H}^{\dagger} & \text{if } \kappa = 0 \\ \mathbf{W}_{\text{MMSE}}^{\top} & \text{if } \kappa = \sqrt{\frac{2N_0}{E_s}} \end{cases} \end{aligned} \tag{19}$$

3.4 Lattice Reduction Aided MIMO Detectors

A (point) lattice $\mathcal{L}$ is a periodic arrangement of discrete points. Any lattice can be characterised in terms of a non unique basis $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m)$, that allows any lattice point to be characterised as a superposition of integer multiples of the basis vectors $\mathbf{b}_\ell$, for example, any $\mathbf{p} \in \mathcal{L}$ can be written as [27]:

$$\mathbf{p} = \sum_{\ell=1}^{m} a_\ell \mathbf{b}_\ell, \quad a_\ell \in \mathbb{Z} \tag{20}$$

Since the noiseless received signal $\mathbf{Hx}$ can be seen as a lattice point, the aim of lattice-reduction is to transform the basis $\mathbf{H}$ into a new basis $\widetilde{\mathbf{H}}$ with vectors of shortest length or, equivalently, into a basis consisting of roughly orthogonal basis vectors. Usually, $\widetilde{\mathbf{H}}$ is much better conditioned than $\mathbf{H}$.

Hence, let suppose that the lattices generated by the column vectors of MIMO channel $\mathbf{H}$ and LR-reduced channel $\widetilde{\mathbf{H}}$ matrices are the same. This implies there exists an integer unimodular matrix $\mathbf{T}$ that satisfies:

$$\widetilde{\mathbf{H}} = \mathbf{HT} \tag{21}$$

Then, the received signal in (2) can be rewritten as:

$$\begin{aligned} \mathbf{r} &= \mathbf{Hx} + \mathbf{n} \\ &= \mathbf{HTT}^{-1}\mathbf{x} + \mathbf{n} \\ &= \widetilde{\mathbf{H}}\mathbf{z} + \mathbf{n} \end{aligned} \tag{22}$$

where $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x}$

3.4.1 LR-Based Zero-Forcing MIMO Detection

For the zero forcing detector, received signal vector on the n_r receive antennas $\mathbf{r}$ is multiplied by the reduced channel pseudo-inverse matrix $\widetilde{\mathbf{H}}^{\dagger}$, resulting:

$$\mathbf{z}_{\text{ZF}} = \widetilde{\mathbf{H}}^{\dagger}\mathbf{r} = \widetilde{\mathbf{H}}^{\dagger}\left(\widetilde{\mathbf{H}}\mathbf{z} + \mathbf{n}\right) = \mathbf{z} + \widetilde{\mathbf{H}}^{\dagger}\mathbf{n} \tag{23}$$

Since in the lattice reduction procedure the original symbols of the QAM constellation are shifted and scaled by the matrix $\mathbf{T}$, before re-mapping the vector $\mathbf{z}_{\text{ZF}}$ to the original constellation, this shifting-scaling operations must be reverted [15], resulting in:

$$\widehat{\mathbf{z}}_{\text{ZF}} = 2 \cdot \left\lceil \frac{\mathbf{z}_{\text{LR}} - \beta' \mathbf{T}^{-1}\mathbf{1}}{2} \right\rfloor + \beta' \mathbf{T}^{-1}\mathbf{1} \tag{24}$$

where $\mathbf{1}$ is unitary column vector, and $\lceil \cdot \rfloor$ is the rounding function for the near integer.

As shown in [27], this process is equivalent to the quantisation operation (over the vector $\mathbf{z}_{\text{ZF}}$) for the nearest point of the constellation $\mathbf{T}^{-1}\mathbf{x}$, as described by (24). After quantisation operation, the vector $\widehat{\mathbf{z}}_{\text{ZF}}$ is demapping onto the original base by multiplying to the unimodular matrix:

$$\widehat{\mathbf{x}}_{\text{ZF}} = \mathbf{T}\widehat{\mathbf{z}}_{\text{ZF}} \tag{25}$$

Hence, $\widehat{\mathbf{x}}_{\text{ZF}}$ is detected as the symbol associated to the constellation point with the minimal distance.

3.4.2 LR-Based Minimum Mean Squared Error MIMO Detection

For the minimum mean squared error detector, the extended received signal vector $\mathbf{r}_{\text{EXT}}$, defined as $\mathbf{r}_{\text{EXT}} = \begin{bmatrix} \mathbf{r} \\ \mathbf{0}_{2n_T} \end{bmatrix}$, where $\mathbf{0}_{2n_T}$ is a column vector with zeros and length $2n_T$, is multiplied by the pseudo-inverse of the reduced extended channel matrix $\widetilde{\mathbf{H}}^{\dagger}_{\text{EXT}}$, resulting:

$$\mathbf{z}_{\text{MMSE}} = \widetilde{\mathbf{H}}^{\dagger}_{\text{EXT}}\mathbf{r}_{\text{EXT}} \tag{26}$$

Then, the quantisation operation and the demapping onto the original base is made in the same way as the LR-ZF.

3.5 Heuristic ACO-Based MIMO Detector

From (13), the MIMO detection problem can be seen as a combinatorial optimisation problem. Therefore, the ant colony optimisation is a proper technique to be applied. In the ACO-MIMO detector proposed in [12], N_{ants} ants search iteratively for better solutions accordingly to a cost function. For MIMO communications systems, the optimal solution is given by (13), so the following cost function can be adopted in the ant colony optimisation context:

$$F(\mathbf{s}) = ||\mathbf{r} - \mathbf{H}\mathbf{s}||^2 \tag{27}$$

In a convenient way, we can decompose the $\mathbf{H}$ matrix using the QR decomposition, such that $\mathbf{H} = \mathbf{QR}$, being $\mathbf{Q}$ a orthogonal matrix with dimension $n \times m$ and $\mathbf{R}$ an upper triangular matrix, dimension $m \times m$. Hence the vector of receiving signal is readily re-written as:

$$\widehat{\mathbf{r}} = \mathbf{Q}^H\mathbf{r} = \mathbf{Q}^H(\mathbf{H}\mathbf{x} + \mathbf{n}) = \mathbf{R}\mathbf{x} + \mathbf{Q}^H\mathbf{n}. \tag{28}$$

Besides, the cost function (27) is conveniently re-written as:

$$F(\mathbf{s}) = ||\widehat{\mathbf{r}} - \mathbf{R}\mathbf{s}||^2 \tag{29}$$

Hence, the d_{ij} distance related to the "s_{ij}" path, i.e., assumed that the jth symbol of the set $\mathcal{A}$ has been transmitted at the ith antenna, is calculated into the ACO algorithm deploying the recursive relation:

$$d_{ij} = \left| \widehat{r_i} - \sum_{l=i+1}^{m} R_{il}\zeta_l - R_{ii}s_j \right| \tag{30}$$

where i should progressively decrease from m to 1, $s_j \in \mathcal{A}$, and ζ_l is the hard decision version of the transmitted symbol s_l that have been tentatively made decision [12]. Hence, ζ_i can be iteratively obtained as

$$\zeta_i = \arg \min_{s_{ij} \in \mathcal{A}} d_{ij}, \tag{31}$$

Besides, the distances d_{ij} are then converted into the heuristic values η_{ij} using a log-sigmoid function:

$$\eta_{ij} = \frac{1}{1 + e^{d_{ij}}} \tag{32}$$

Such values have influence on the probability calculation of the paths traced by the active ants across the iterations of the algorithm; these ant paths following probability is computed as:

$$p_{ij} = \frac{[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{j \in M}[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}} \tag{33}$$

where τ_{ij} is the pheromone level over the path s_{ij}, and the constants α and β weight the importance of τ_{ij} and η_{ij}, respectively. These levels are a way to implement evolution mechanism in the algorithm along the iterations, being analogue to the substances that real ants deposit on the best paths in searching food process. The pheromone level is set up with a correspondent initial value, and as soon as the N_{ants} complete their paths at nth iteration $\widetilde{\mathbf{s}}_k^{(n)}$, it is updated at the $n+1$ iteration according to:

$$\tau_{ij}^{(n+1)} = (1 + \varphi)\tau_{ij}^{(n)} + \sum_{k=1}^{N_{\text{ants}}} \Delta\tau_{ij}^{k} \tag{34}$$

where φ is the pheromone evaporation rate (ER) and $\Delta\tau_{ij}^{k}$ is given by:

$$\Delta\tau_{ij}^{k} = \begin{cases} F(\widetilde{\mathbf{s}}_k^{(n)}) & \text{if } (i, j) \in \widetilde{\mathbf{s}}_k^{(n)}, \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

At the end of the first iteration, the cost function of the paths taken for each ant had been calculated, being the lower cost function and the associated path assigned to the variables F_{best} and $\mathbf{s}_{best}$, respectively. These variables are updated along the iterations. At the end, $\mathbf{s}_{best}$ represents the solution given by the ACO search algorithm.

3.6 ACO MIMO Detector with Initial Guess

In the following, some modifications in the ACO-MIMO detector described previously are introduced in such a way that the search can be readily started from an initial point, which in fact would be the solution offered by any low-complexity linear MIMO detector; besides, the pheromone update is done on a closest way of that shown in [3]. Finally, adapting this heuristic search algorithm taking into account the improved channel conditions provided by the LR technique, we can considerably improve the overall MIMO system performance at the cost of an affordable increasing in complexity, as discussed in Sect. 4.

Initially, if in the Eq. (30) we use $\boldsymbol{\zeta}$ as the initial solution given by any low-complexity linear MIMO detector, such as ZF or MMSE, with or without LR aiding, the heuristic decision will be obtained taking advantage of the information provided by this detector. This way, as much better the information provided by the initial detector is in terms of bit-error-rate (BER) performance levels, more reliable the final achieved heuristic information, and less processing (time-consuming) is necessary for the ants to search and find a reliable high-quality solution. Hence, in this work we use the initial guess for the ACO input as $\boldsymbol{\zeta} = \widehat{\mathbf{x}}_{\text{MMSE}}$, deployed in the calculation of (30).

The other modification introduced in this work is related to the pheromone updating. Since the cost function (29) should be minimised, Eq. (34) and (35) may be not completely appropriate, since these calculations possibly introduce an excessive pheromone accumulation on the paths controlled by φ, which can provide erroneously a higher pheromone deposition level over the paths under worse evaluation (i.e., high cost function). In order to correct this effect, (35) is re-written as:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{\varrho}{F(\widetilde{\mathbf{s}}_{k}^{(n)})} & \text{if } (i,j) \in \widetilde{\mathbf{s}}_{k}^{(n)}, \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

where ϱ is a constant to be adjusted experimentally.

In a similar way, as suggested in [3], herein we also implemented a second pheromone deposition rule, which takes into account the best solution found so far by the ACO algorithm:

$$\Delta\tau_{\text{elitist}}^{(n)} = \begin{cases} \frac{\Delta}{F(\mathbf{s}_{\text{best}}^{(n)})} & \text{if } (i,j) \in \mathbf{s}_{\text{best}}^{(n)}, \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

where δ is another constant to be adjusted experimentally. As a consequence, Eq. (34) becomes:

$$\tau_{ij}^{(n+1)} = (1-\varphi)\tau_{ij}^{(n)} + \sum_{k=1}^{N_{\text{ants}}} \Delta\tau_{ij}^{k} + \Delta\tau_{\text{elitist}}^{(n)} \tag{38}$$

where $\Delta\tau_{ij}^{k}$ is given by (36), and $\Delta\tau_{\text{elitist}}^{(n)}$ by (37).

3.7 LR-Based ACO MIMO Detector

One of the main issues to be solved in the ACO MIMO detector adaptation to the reduced domain, with better conditioned channel gain matrix, is the fact that in the LR domain there is no fixed constellation, as discussed in [2]. Since $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x}$, each channel matrix remains on a different reduced transformed constellation. Obviously, to calculate all these constellation points for each channel matrix gain realisation is practically unfeasible.

In order to solve this problem, the ants' search for the optimal solution initially takes place on the neighborhood of an initial solution, namely reduced domain neighborhood (RDN) procedure [2]. In this work, the initial solution is provided by the LR-MMSE MIMO detector.

Since every constellation can be seen as a shifted and scaled version of the integer constellation $\mathbb{Z}^n$ [15], the procedure starts taking the shifted and scaled version of the vector $\mathbf{z}_{\text{MMSE}}$, expressed by:

$$\widetilde{\mathbf{z}}_{\text{MMSE}} = \frac{\mathbf{z}_{\text{MMSE}} - \beta' \mathbf{T}^{-1}\mathbf{1}}{2} \tag{39}$$

Hence, the neighborhood is obtained as the combinations of the adjacent integers on each element of $\widetilde{\mathbf{z}}_{\text{MMSE}}$, after quantisation; in other words, for a given dimension i, the neighborhood is obtained as:

$$\begin{aligned} &\left\lceil \widetilde{z}_{\text{MMSE}_i} \right\rfloor - \frac{N-1}{2}, \ \left\lceil \widetilde{z}_{\text{MMSE}_i} \right\rfloor - \frac{N-1}{2} + 1, \ \ldots \\ &\ldots, \left\lceil \widetilde{z}_{\text{MMSE}_i} \right\rfloor + \frac{N-1}{2} - 1, \ \left\lceil \widetilde{z}_{\text{MMSE}_i} \right\rfloor + \frac{N-1}{2} \end{aligned} \tag{40}$$

being N the number of elements per dimension to be calculated. $N = 5$ was adopted in[2]; however, the problem was treated in a complex-values format. Since in this work we are adopting equivalent real-values format, we have adopted $N = 3$. So, the jth neighbour of $\left\lceil \widetilde{z}_{\text{MMSE}_i} \right\rfloor$, $\mathfrak{z}_{ij}$, can be defined as:

$$\mathfrak{z}_{ij} = \left\lceil \widetilde{z}_{\text{MMSE}_i} \right\rfloor - \frac{N+1}{2} + j \tag{41}$$

In a same way, Eq. (30) also should be adapted to the reduced domain. Hence, it becomes [2]:

$$\widetilde{d}_{ij} = \left\| \widetilde{\mathbf{R}} (\widetilde{\mathbf{z}}_{\text{MMSE}} - \widetilde{\mathbf{z}}) \right\|^2 \tag{42}$$

in which i should decrease from m to 1, $\widetilde{\mathbf{z}}$ is a vector formed by the elements of $\lceil \widetilde{\mathbf{z}}_{\text{MMSE}} \rfloor$ on the positions $i+1 \ldots m$, $\mathfrak{z}_{ij}$ at the ith position, and zeros at the positions $i-1 \ldots 1$; matrix $\widetilde{\mathbf{R}}$ is given by the QR decomposition of the reduced channel matrix $\widetilde{\mathbf{H}}$.

Finally, the cost function deployed in the reduced LR domain description becomes:

$$\widetilde{F}(\mathbf{z}) = ||\widetilde{\mathbf{r}} - \widetilde{\mathbf{R}}\mathbf{z}||^2 \tag{43}$$

where $\widetilde{\mathbf{r}} = \widetilde{\mathbf{Q}}\mathbf{r}$.

The remainder of the algorithm is constructed in the same way as the ACO MIMO detector on the original channel gain matrix domain. Pseudo-codes for both heuristic MIMO algorithms are described in Algorithms 1 and 2.

4 Numerical Results

It is well-known that the ACO algorithm performance depends on the appropriate choice for its internal (input) parameters, ensuring this way a desirable acceleration in the algorithm's convergence [14] while circunvemting possible slow convergence issue, and simultaneously guarantees a reduction in the computational complexity. Hence, as a first step in this section, a procedure aiming to obtain optimised input parameters for the ACO-MIMO detector is carried out. Three different MIMO system configurations have been considered: a) 4×4 antennas and

Algorithm 1 ACO-MIMO

Input: $\widehat{\mathbf{r}}$, $\mathbf{R}$, $\widehat{\mathbf{x}}_{\text{MMSE}}$.
Initialisation: $\tau_{ij} \leftarrow \tau_0$, $\mathbf{s}_{best} \leftarrow \widehat{\mathbf{x}}_{\text{MMSE}}$, $F_{best} \leftarrow \infty$.
for each path s_{ij} **do**
  $d_{ij} = |\widehat{r}_i - \sum_{l=i+1}^{m} R_{il}\zeta_l - R_{ii}s_j|$;
  $\eta_{ij} = \frac{1}{1+e^{d_{ij}}}$;
end for
for $n = 1$ to $\mathcal{I}$ **do**
  for each s_{ij} **do**
    Evaluate p_{ij} according to (33);
  end for
  for each k-th ant **do**
    Generate trails $\widetilde{\mathbf{s}}_k$ according to p_{ij};
    Evaluate cost function $F(\widetilde{\mathbf{s}}_k)$;
  end for
  Update F_{best} and $\mathbf{s}_{best}$;
  Update τ_{ij} according to (36), (37) e (38);
end for
Output: $\mathbf{s}_{best}$.

Algorithm 2 LR-ACO-MIMO

Input: $\widetilde{\mathbf{r}}$, $\widetilde{\mathbf{R}}$, $\mathbf{z}_{\text{MMSE}}$, $\mathbf{T}$.
Initialisation: $\tau_{ij} \leftarrow \tau_0$, $\mathbf{z}_{best} \leftarrow \mathbf{z}_{\text{MMSE}}$, $F_{best} \leftarrow \infty$.
Evaluate $\widetilde{\mathbf{z}}_{\text{MMSE}}$ according to (39);
Generate RDN according to (40);
for each neighbor $\mathfrak{z}_{ij}$ **do**
  $\widetilde{d}_{ij} = \left\|\widetilde{\mathbf{R}}(\widetilde{\mathbf{z}}_{\text{MMSE}} - \widetilde{\mathbf{z}})\right\|^2$;
  $\eta_{ij} = \frac{1}{1+e^{\widetilde{d}_{ij}}}$;
end for
for $n = 1$ to $\mathcal{I}$ **do**
  for each $\mathfrak{z}_{ij}$ **do**
    Evaluate p_{ij} according to (33);
  end for
  for each k-th ant **do**
    Generate trails $\widetilde{\mathbf{z}}_k$ according to p_{ij};
    Evaluate cost function $\widetilde{F}(\widetilde{\mathbf{z}}_k)$;
  end for
  Update F_{best} and $\mathbf{z}_{best}$;
  Update τ_{ij} according to (36), (37) e (38);
end for
Output: $\mathbf{s}_{best} = \mathbf{T}\mathbf{z}_{best}$.

64-QAM; b) 8×8 antennas and 16-QAM; c) 20×20 antennas and 4-QAM modulation. After that, performance and complexity analyses under optimised input parameters are developed.

4.1 Input Parameters Optimisation

As generically analysed in [3] and specifically in [14] considering the multiuser DS-CDMA detection problem, the parameters ϱ, δ and evaporation rate (ER) φ present little influence on the ACO algorithm performance. Hence, these parameters were chosen empirically after non-exhaustive tests, being adopted the following values: $\varrho = 1$, $\delta = 3$ and $\varphi = 0.3$.

Table 1 ACO-MIMO input parameters

Parameter	Value
ϱ	1
Δ	3
φ	0.3
N_{ants}	20
τ_0	0.01
α	[0.0; 1.0]; step: 0.2
β	[0.2; 2.0]; step: 0.2

On the other hand, α and β parameters drastically affect the ACO algorithm convergence speed performance depending on the optimisation problem type. As discussed in [3,14], α is related to the importance given to the pheromone levels in the probability calculations of (33), while β is related to the importance given to the *"a priori"* information in (33).

The goal with the input parameters optimisation procedure is to obtain a unique set of values that ensures an affordable and suitable algorithm performance adjusted to deal with the MIMO detection problem under realistic channel scenarios and a wide range of transmit and receive antennas configurations, as well as modulation orders and antenna correlation. Once optimised the parameters set, the algorithm is able to operate in any MIMO channel and system configurations. The numerical results discussed in the following corroborate the good performance $\times$ complexity for the proposed ACO-MIMO detector under a wide varieties of MIMO system configurations.

For the MIMO heuristic ACO-based detection problem, parameters α and β have been optimised while the other input parameters values were fixed as described in Table 1. Varying the parameters $\alpha \in [0.0;\ 1.0]$ and $\beta \in [0.2;\ 2.0]$, with steps of 0.2, all possible parameters combinations on these intervals were performed for three set of antennas, and modulation order combinations: (a) 4×4 and 64-QAM, Fig. 1; (b) 8×8 and 16-QAM, Fig. 2; (c) 20×20 4-QAM, Fig. 3. These figures show the achieved BER performance as a function of both ACO input parameter variations. Hence, for each system configuration analysed, the (α, β) values associated with the minimum bit-error-rate (BER_{min}) were stored.

In order to obtain the best (α, β)-parameters combination, the ratios between the bit error rate of each combination and the minimum BER obtained over a specific MIMO configuration, $\frac{\text{BER}_{\alpha,\beta}}{\text{BER}_{\text{min}}}$, were calculated. Also, for each parameters combination, the mean among the ratio considering the three configurations presented in Figs. 1, 2 and 3 were obtained. Finally, the optimum (α, β)-parameters combination was obtained as that one with the lower mean. As a result, the best combination obtained was $(\alpha, \beta)_{\text{opt}} = (0.8;\ 0.8)$. Hereafter, these values are adopted for the proposed LR-ACO MIMO detector.

4.2 LR-ACO MIMO Performance Under Optimised Input Parameters

Figure 4 shows two figures of performance for all considered MIMO detectors under 4×4 antennas and 64-QAM modulation. For ACO-MIMO detectors, (a) $\mathcal{I} = 40$ iterations for the BER versus E_b/N_0 performance; and (b) correspondent convergence speed was performed at a signal-to-noise ratio of 26 dB. In a same way, Figs. 5 and 6 show the same behavior of the MIMO detectors for 8×8 antennas, 16-QAM, SNR =18 dB; and for 20×20 antennas, 4-QAM, SNR =16 dB, respectively. For a notation simplicity, the ACO-MIMO proposed

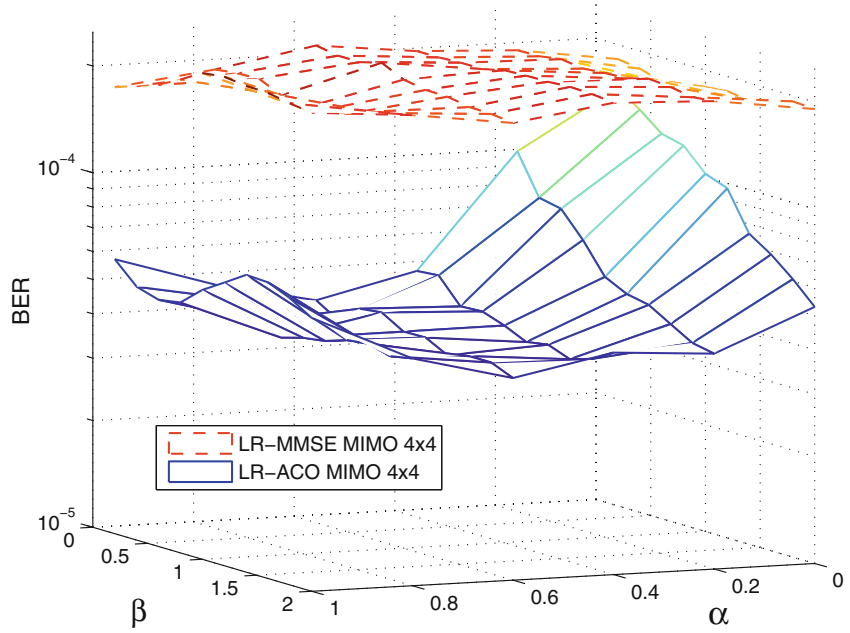

Fig. 1 LR-ACO input α and β parameters optimisation for 4×4, 64-QAM and $\xi = 26\,\text{dB}$

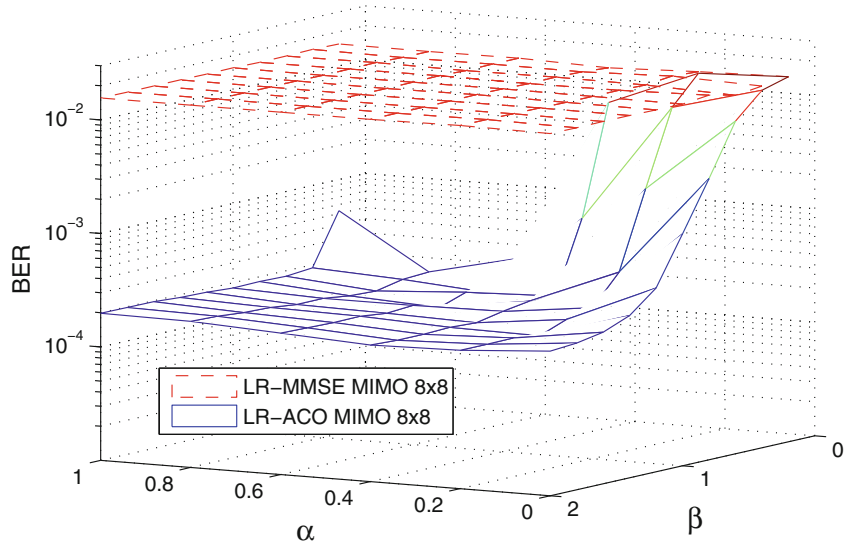

Fig. 2 LR-ACO input α and β parameters optimisation for 8×8, 16-QAM and $\xi = 18\,\text{dB}$

in [12] is identified as "ACO_1", while the ACO-MIMO proposed herein which performs the search starting from the LR-MMSE solution in the original domain named as "ACO_2", and the ACO-MIMO detector proposed herein performing the search in a reduced domain is denominated "LR- ACO". For comparison reference purpose, the sphere decoding MIMO performance of [16], namely "SD MIMO", has been included in the graphs of this section.

Comparing the BER performance results, one can conclude that while the ACO_1-MIMO detector presents a performance near to MMSE MIMO detector for all antenna and modulation order configurations, the ACO_2-MIMO detector achieves a performance near to LR-MMSE MIMO detector when the number of antennas is not so large. It is worth noting the

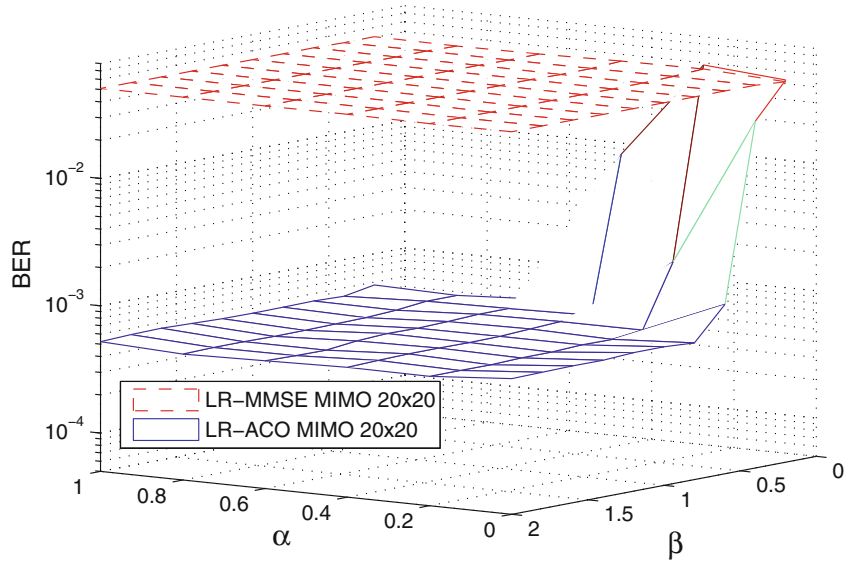

Fig. 3 LR-ACO input α and β parameters variation for 20×20, 4-QAM and $\xi = 16$ dB

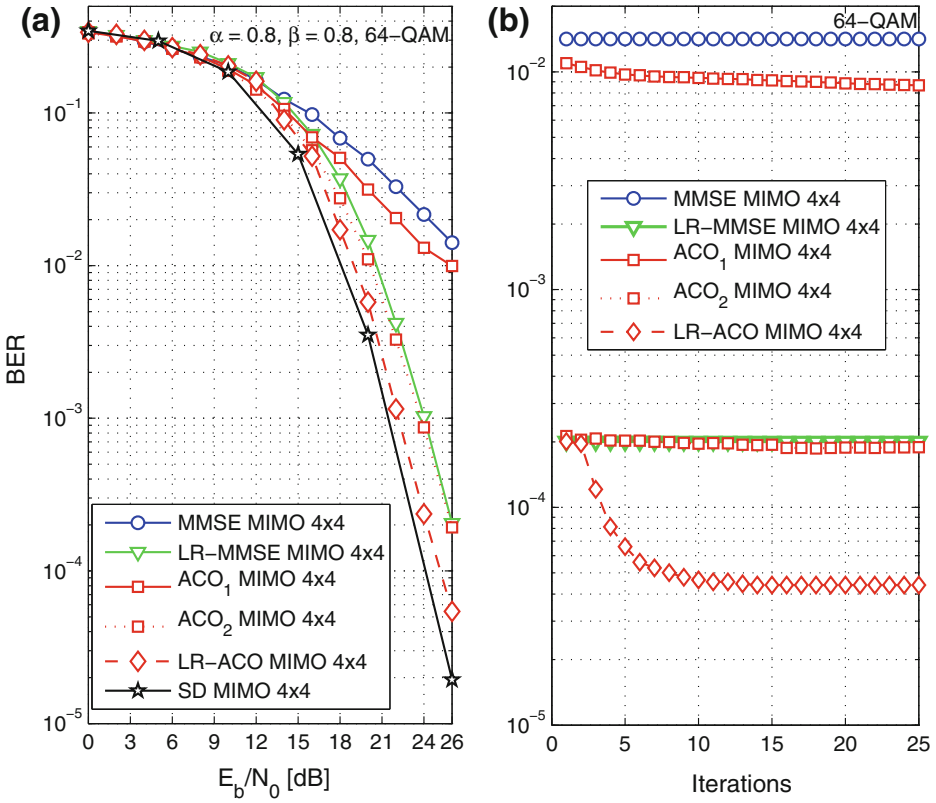

Fig. 4 **a** BER performance after $\mathcal{I} = 40$ iterations for ACO-based detectors; **b** convergence under $\xi = 26$ dB, considering 4×4 antennas, 64-QAM and optimised ACO input parameters

proposed LR-ACO MIMO detector reaches full diversity degree[2] in all system configuration evaluated, leading to a significant performance improvement, although the SNR gap regarding the sphere decoder (SD-MIMO) has grown with the increasing number of antennas.

[2] Defined as the asymptotic slope of the BER curve in each number of antennas scenario.

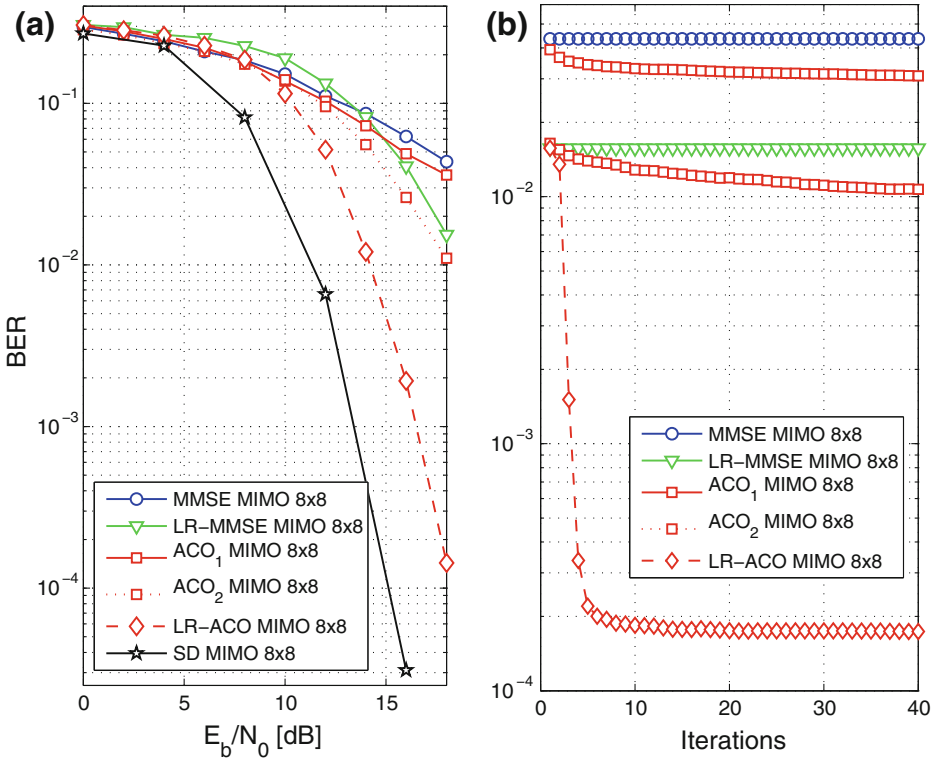

Fig. 5 **a** BER performance after $\mathcal{I} = 40$ iterations for ACO-based detectors; **b** convergence under $\xi = 18$ dB, considering 8×8 antennas, 16-QAM and optimised ACO input parameters

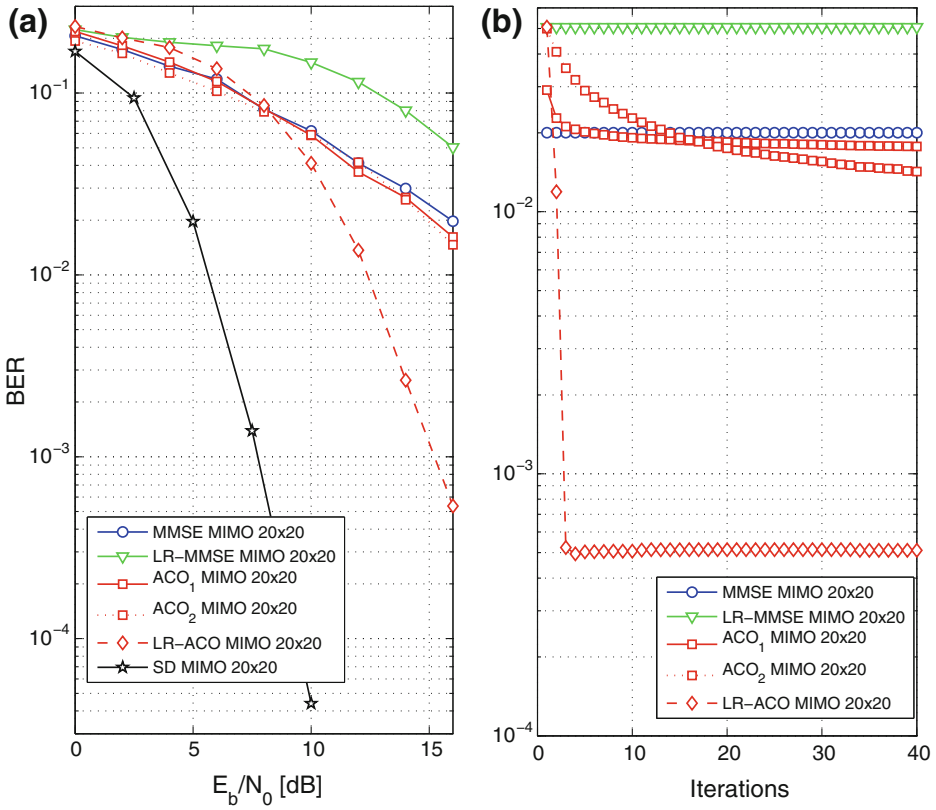

Fig. 6 **a** BER performance after $\mathcal{I} = 40$ iterations for ACO-based detectors; **b** convergence under $\xi = 16$ dB, considering 20×20 antennas, 4-QAM and optimised ACO input parameters

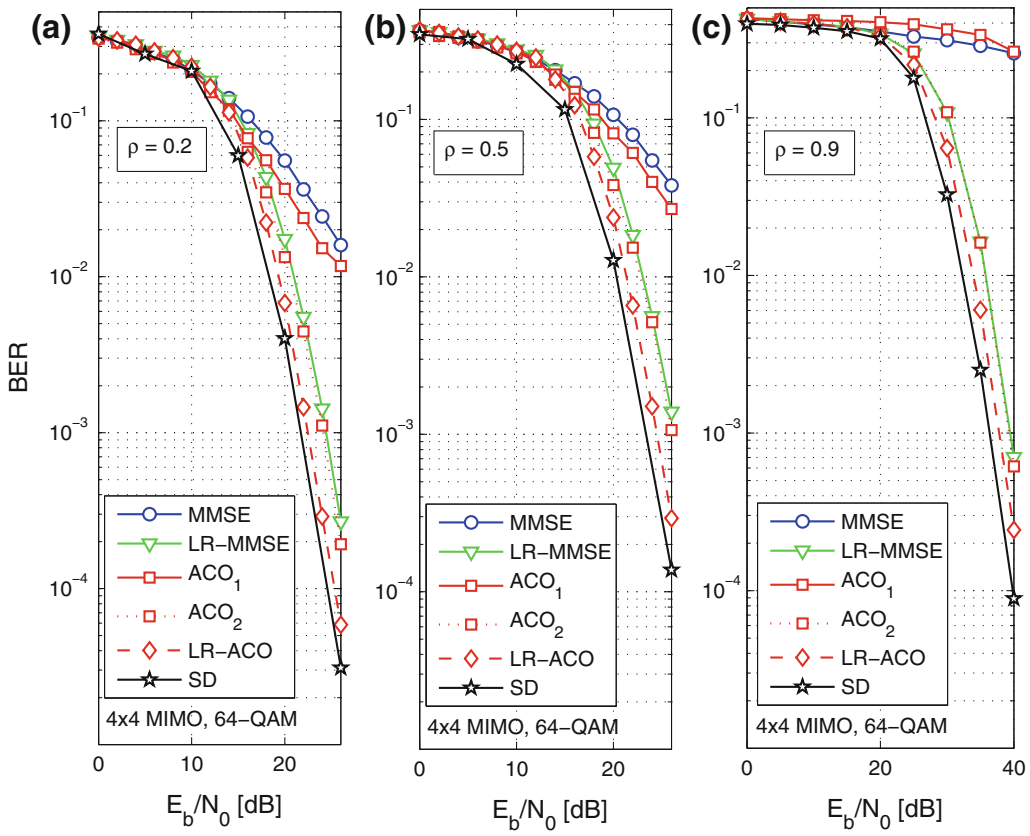

Fig. 7 BER performance for the MIMO detectors considering 4×4 antennas, 64-QAM modulation under correlated channels: **a** weakly correlated ($\rho = 0.2$); **b** medianly correlated ($\rho = 0.5$); **c**) ctrongly correlated ($\rho = 0.9$)

Analysing the convergence speed on Figs. 4b–6b, one can see that the LR-ACO needs at the most $\mathcal{I} = 10$ iterations to achieve total convergence. However, for the other ACO-based detectors either the search remains evolving slowly along the $\mathcal{I} = 40$ iterations, or converge result is very poor (high BER).

4.3 Performance Under Correlated Channels

As explained in Sect. 2, a similar environment for transmitter and receiver antennas is assumed, thus $\rho_t = \rho_r = \rho$. Figure 7 shows the performance of the detectors with the correlation elements increasing, i.e., a) $\rho = 0.2$; b) $\rho = 0.5$; $\rho = 0.9$, for 4×4, 64-QAM. Figure 8 does the same for 8×8, 16-QAM. Finally, for 20×20, 4-QAM, it was not possible to obtain the BER performance for the sphere decoder under strongly correlated channels ($\rho = 0.9$), since the simulation time became excessively high. Hence, Fig. 9 shows the performance of the detectors just for weakly correlated ($\rho = 0.2$) and medianly correlated channels ($\rho = 0.5$) for 20×20, 4-QAM.

From Figs. 7, 8 and 9, one can conclude that the BER performance of the LR-ACO remains near the SD performance, for 4×4, 64-QAM, even for the strongly correlated channel. For 8×8, 16-QAM, the SNR gap for high SNR regions remains approximately the same as the one of uncorrelated channels, for weakly and correlated channels, and became somewhat larger for strongly correlated channels. Hence, under the 20×20, 4-QAM configuration, the SNR gap, that was ≈ 8 dB for uncorrelated channels, became ≈ 10 dB for weakly correlated, and ≈ 17 dB for medianly correlated channels. Again, for all the analysed conditions, the proposed LR-ACO MIMO detector has reached full diversity degree.

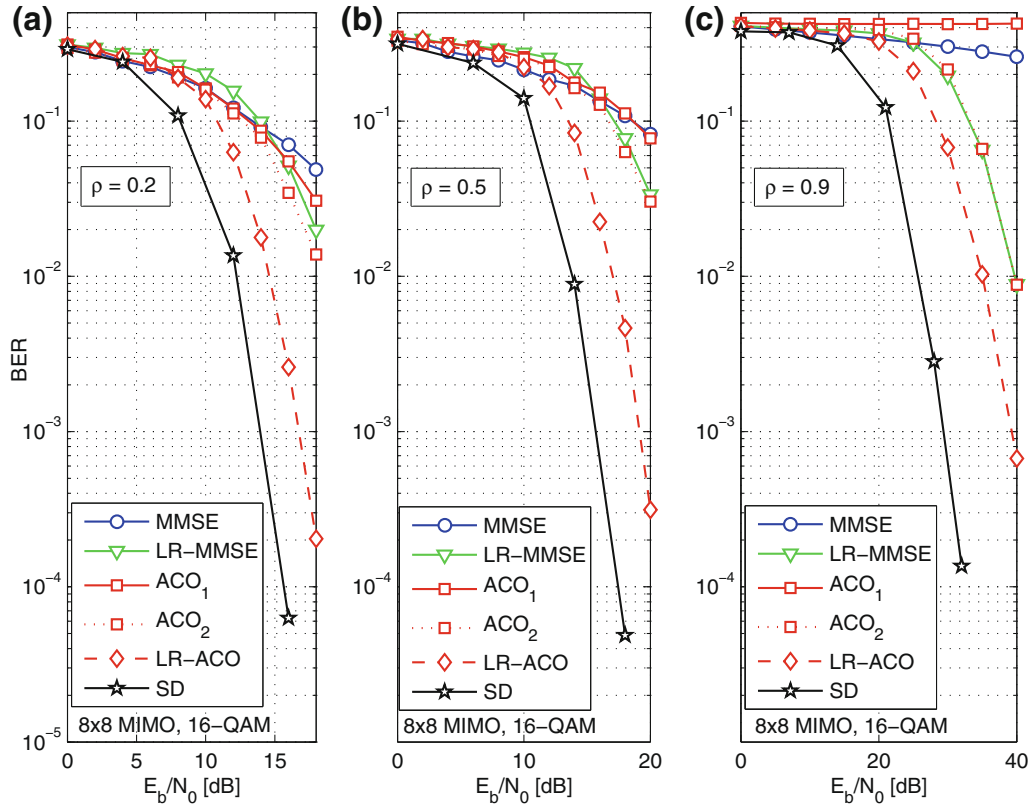

Fig. 8 BER performance for the MIMO detectors considering 8×8 antennas, 16-QAM modulation under correlated channels: **a** weakly correlated ($\rho = 0.2$); **b** medianly correlated ($\rho = 0.5$); **c** strongly correlated ($\rho = 0.9$)

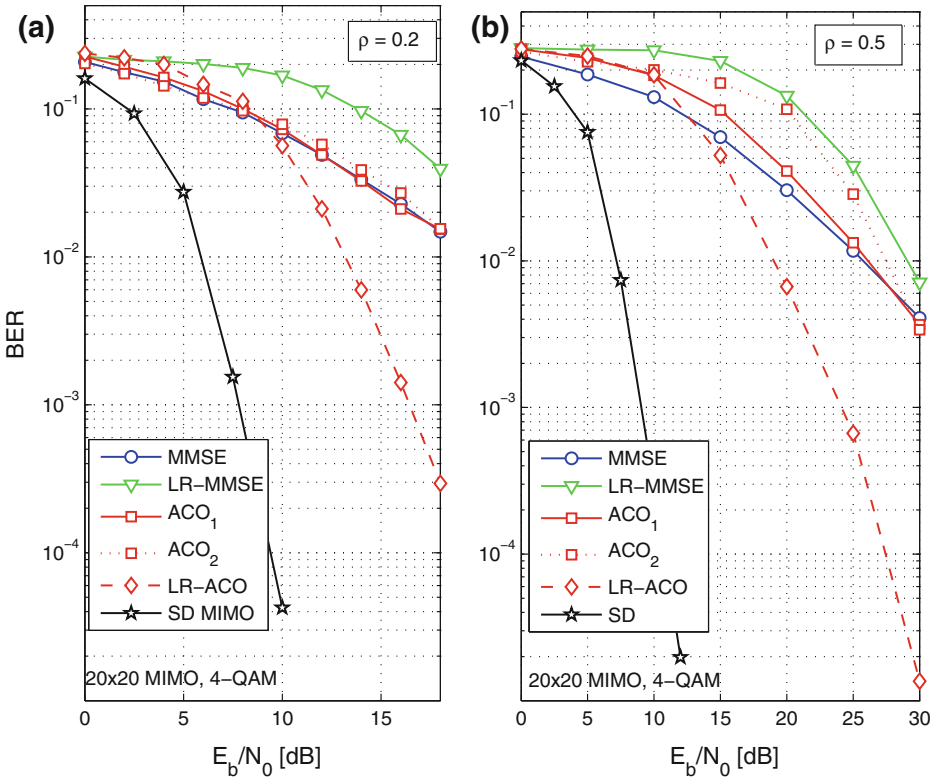

Fig. 9 BER performance for the MIMO detectors considering 20×20 antennas, 4-QAM modulation under correlated channels: **a** weakly correlated ($\rho = 0.2$); **b** medianly correlated ($\rho = 0.5$)

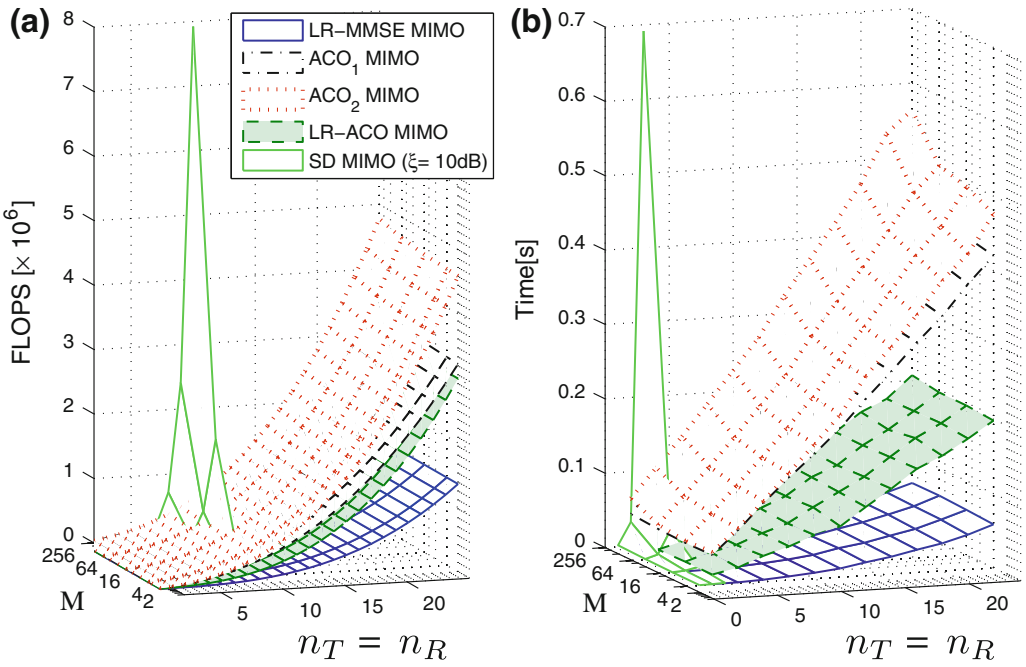

Fig. 10 Detector complexities in terms of: **a** *flops* and **b** *computational time*. For ACO_1 and ACO_2: $N_{ants} = 20$, $\mathcal{I} = 40$; for LR-ACO: $N_{ants} = 20$, $\mathcal{I} = 10$; and for SD: $\xi = 10\,dB$

4.4 Complexity Analysis

The algorithm complexities can be evaluated in terms of the total number of floating-point operations (*flops*)[3] [6]. An addition between two $m \times n$ matrices spends mn *flops*, while a multiplication between a $m \times n$ matrix $\mathbf{A}$ and a $n \times p$ matrix $\mathbf{B}$ spends $(2n - 1)mp$ *flops* [6]. Table 2 shows the number of *flops* for the considered MIMO detectors, assuming equal numbers of antennas ($n_T = n_R = n$) condition and M-QAM modulation order. For simplicity, we considered just the numerical operations, neglecting some basic computational operations, like comparisons, permutations, memory accesses. For instance, the computation of the line 4 in the Algorithm 2 spends $12n^2 - 2n$ *flops*, being $2n_t$ to compute the vector difference, $(4n_t - 1)2n_t$ to compute the multiplication between the matrix and the vector, and $2n_t(2n_t - 1)$ to compute the quadratic norm of the vector. The complexity for all three ACO-based MIMO detectors and the LR-MMSE MIMO detector are of the order $\mathcal{O}(n^3)$.

The complexity of sphere decoding is not trivial to obtain. In fact, it depends on the number of candidates inside the hyper-sphere to be evaluated. It is demonstrated in [9] that the SD complexity always presents an exponential asymptotic behaviour, contrary to previous works that stated its polynomial shaping under certain conditions. This is justified as consequence that to ensure a certain probability of finding a point within the sphere, its radius must grow with the problem size [9]. As the number of points within the sphere is closely related to the search radius, it is also dependent on the problem size.

The MIMO detector complexities in terms of *flops* as a function of $n_T = n_R = n$ antennas and modulation order M are depicted on (Fig. 10a). For SD-MIMO detector, the γ exponent was obtained from the results of [9]. For comparison purpose, the corresponding computational time complexities are also depicted in Fig. 10b. It is clear that, except for SD, the complexity growing for all analysed MIMO detectors shows a polynomial behaviour with the number of antennas n, while the modulation order M affects the complexity only marginally.

[3] A *flop* is defined as an addition, subtraction, multiplication or division between two floating point numbers

Table 2 Number of operations necessary for each MIMO detector

Detector	Number of operations
LR-MMSE	$(258/3)n^3 + 33n^2 - 1$
ACO_1	$(208/3)n^3 + 8n^2 + 4n^2\sqrt{M} + 8n\sqrt{M} + -4n + \mathcal{I}\left[10n\sqrt{M} + (4n^2 + 6n)N_{\text{ants}} + 1\right]$
ACO_2	$(466/3)n^3 + 41n^2 - 4n + 4n^2\sqrt{M} + 8n\sqrt{M} + \mathcal{I}\left[10n\sqrt{M} + (4n^2 + 6n)N_{\text{ants}} + 2n + 1\right]$
LR-ACO	$(466/3)n^3 + 41n^2 - 4n + 4n^2N + 8nN + + \mathcal{I}\left[10nN + (4n^2 + 6n)N_{\text{ants}} + 2n + 1\right]$
SD	$(n^2 + n + 1)2^{\gamma n}$

N_{ants} # ants, $\mathcal{I}$: # iterations, N LR-ACO neighborhood size

On the other hand, the numerical results for SD-MIMO complexity obtained in Fig. 10b corroborates its exponential asymptotic behaviour, as demonstrated analytically by [9], which can be seen both in terms of number of antennas (axis n), as well as in the terms of modulation order (axis M).

It should be noted from Table 2 that the complexities of all ACO detectors do not include the parameters optimisation stage, since it is done just only a single time for adjusting the algorithm, and from then on, the ACO-based algorithms are able to operate under any MIMO modulation order, number of antennas, and channel correlation level, with an improved convergence rate.

In the same way as [2], we have found that another advantage of the LR-ACO consists in its complexity does not depend on the modulation order (M) used, since the neighbourhood size (N) remains independent of M.

Finally, some difference had appeared in the shape of the complexity curves when one compare *flops* and computational time in Fig. 10, which can be justified by the second order simplifications made during the *flops*-based complexity calculations. However, one can see that the LR-ACO detector complexity remains always below the ACO_1 and ACO_2 complexities, whereas it has presented a substantial performance improvement, mainly for large-MIMO systems. Another difference that can be seen is the number of antennas in which the SD complexity achieves very high values. While this condition occurs near $n = 8$ antennas in the *flops* curve, it occurs for $n = 3$ antennas in the *computational time* analysis. It can be justified by some simplifying hypothesis made in [9], since the authors were mainly interested on the asymptotic behaviour of the SD technique, as well as on the strategies deployed to determine the search radius. On the other hand, while herein we have used the pruning strategy of [13], the radius in [9] was assumed to simply satisfy $r \geq \sigma_n\sqrt{n}$.

5 Conclusion

This work proposed a new heuristic-based detector for MIMO communication systems that is able to combine the promising features of both diminished search space propitiated by lattice reduction (LR) technique with the ant colony optimisation; as a result, the proposed LR-ACO MIMO detector is capable to achieve a substantial improvement on BER performance, reaches full diversity degree while expending lower computational burden than others ACO-based MIMO detectors considered, as well as presents certain robustness against the spatially correlated antennas effects. The promising performance-complexity trade-off demonstrated

by the proposed LR-ACO MIMO detector under different number of antennas and modulation order configurations enables its as a good candidate for the next dense-MIMO system implementations.

Acknowledgments This work was supported in part by the National Council for Scientific and Technological Development (CNPq) of Brazil under Grants 202340/2011-2, 303426/2009-8, in part by the Araucaria Foundation of PR-Brazil under Grant 007/2011, in part by CAPES-Brazil (scholarship), and in part by Londrina State University—Paraná State Government (UEL).

References

1. Angélico, B., Burt, P., Jeszensky, P., Hodgkiss, W., & Abrão, T. (2011). Performance analysis of a single-user miso ultra-wideband time reversal system with dfe. *Telecommunication Systems*, *46*, 333–342. doi:10.1007/s11235-010-9295-1.
2. Aubert, S., Nasser, Y., & Nouvel, F. (2012). Lattice reduction-aided minimum mean square error k-best detection for mimo systems. In *2012 International conference on computing, networking and communications (ICNC)* (pp. 1066–1070). doi:10.1109/ICCNC.2012.6167371.
3. Dorigo, M., Birattari, M., & Stutzle, T. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine*, *1*(4), 28–39. doi:10.1109/MCI.2006.329691.
4. de Paula Marques, M., Abrão, T., Adaniya, M. H., Sampaio, L. H. D., & Jeszensky, P. J. E. (2013). Ant colony optimization for resource allocation and anomaly detection in communication networks. In T. Abrão (Ed.), *Search algorithms for engineering optimization*, (chap. 8, pp. 1–34). InTech Open.
5. Foschini, G. J., & Gans, M. J. (1998). On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, *6*, 311–335.
6. Golub, G. H., & Loan, C. F. V. (1996). *Matrix computations*. Maryland: Johns Hopkins University Press.
7. Gowrishankar, R., Demirkol, M., & Yun, Z. (2005). Adaptive modulation for mimo systems and throughput evaluation with realistic channel model. In *2005 International conference on wireless networks, communications and mobile computing* (vol. 2, pp. 851–856). doi:10.1109/WIRLES.2005.1549523.
8. He, Q., Feng, Z., & Zhang, P. (2013). Reconfiguration decision making based on ant colony optimization in cognitive radio network. *Wireless Personal Communications*, *71*(2), 1247–1269. doi:10.1007/s11277-012-0872-3.
9. Jaldén, J., & Ottersten, B. (2005). On the complexity of sphere decoding in digital communications. *IEEE Transactions on Signal Processing*, *53*(4), 1474–1484. doi:10.1109/TSP.2005.843746.
10. Khurshid, K., Irteza, S., & Khan, A. A. (2010). Application of ant colony optimization based algorithm in mimo detection. In *CEC10-IEEE congress on evolutionary computation* (pp. 1–7), doi:10.1109/CEC.2010.5586173.
11. Kotti, M., Benhala, B., Fakhfakh, M., Ahaitouf, A., Benlahbib, B., Loulou, M., et al. (2011) Comparison between pso and aco techniques for analog circuit performance optimization. In *2011 International conference on microelectronics (ICM)* (pp. 1–6). doi:10.1109/ICM.2011.6177367.
12. Lain, J. K., & Chen, J. Y. (2010). Near-mld mimo detection based on a modified ant colony optimization. *IEEE Communications Letters*, *14*(8), 722–724. doi:10.1109/LCOMM.2010.08.100347.
13. Larsson, E. G. (2009). MIMO detection methods: How they work. *IEEE Signal Processing Magazine*, *26*(3), 91–95.
14. Marinello, J. C., de Souza, R. N., & Abrão, T. (2012). Ant colony input parameters optimization for multiuser detection in ds/cdma systems. *Expert Systems with Applications*, *39*(17), 12,876–12,884. doi:10.1016/j.eswa.2012.05.005.
15. Milford, D., & Sandell, M. (2011). Simplified quantisation in a reduced-lattice mimo decoder. *IEEE Communications Letters*, *15*(7), 725–727. doi:10.1109/LCOMM.2011.051011.110485.
16. Mostagi, Y. M., & Abrão, T. (2012). Lattice-reduction-aided over guided search mimo detectors. *International Journal of Satellite Communications Policy and Management*, *1*(2/3), 142–154.
17. Rahhal, J., & Abu-Al-Nadi, D. (2010). Pre-coding for mimo systems in frequency-selective fading channels. *Wireless Personal Communications*, *55*(4), 591–605. doi:10.1007/s11277-009-9821-1.
18. Rusek, F., Persson, D., Lau, B. K., Larsson, E., Marzetta, T., Edfors, O., et al. (2013). Scaling up mimo: Opportunities and challenges with very large arrays. *IEEE Signal Processing Magazine*, *30*(1), 40–60. doi:10.1109/MSP.2011.2178495.
19. Seyman, M. N., & Taspinar, N. (2013). Pilot tones optimization using artificial bee colony algorithm for mimo ofdm systems. *Wireless Personal Communications*, *71*(1), 151–163.

20. Telatar, E. (1999). Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, *10*(6), 585–595. doi:10.1002/ett.4460100604.
21. Torabi, E., Mietzner, J., & Schober, R. (2008). Pre-equalization for pre-rake MISO DS-UWB systems. In *IEEE international conference on communications (ICC '08)* (pp. 4861–4866).
22. Tse, D., & Viswanath, P. (2011). *Fundamentals of wireless communications* (1st ed.). England: Cambridge University Press.
23. Uma, K., Palanisamy, P., & Poornachandran, P. (2011). Comparison of image compression using ga, aco and pso techniques. In *2011 International conference on recent trends in information technology (ICRTIT)* (pp. 815–820). doi:10.1109/ICRTIT.2011.5972298.
24. Wang, B., & Wang, L. (2012). A novel artificial bee colony algorithm for numerical function optimization. In *2012 Fourth international conference on computational and information sciences (ICCIS)* (pp. 172–175). doi:10.1109/ICCIS.2012.32.
25. Wang, P., & Le-Ngoc, T. (2011). A list sphere decoding algorithm with improved radius setting strategies. *Wireless Personal Communications*, *61*(1), 189–200. doi:10.1007/s11277-010-0018-4.
26. Wübben, D., Bohnke, R., Kuhn, V., & Kammeyer, K. D. (2004). Near-maximum-likelihood detection of mimo systems using mmse-based lattice reduction. In *2004 IEEE international conference on communications* (vol. 2, pp. 798–802). doi:10.1109/ICC.2004.1312611.
27. Wübben, D., Seethaler, D., Jaldén, J., & Matz, G. (2011). Lattice reduction. *IEEE Signal Processing Magazine*, *28*(3), 70–91. doi:10.1109/MSP.2010.938758.
28. Zelst, A. V., & Hammerschmidt, J. S. (2002). A single coefficient spatial correlation model for multiple-input multiple-output (mimo) radio channels. In *Proceedings of the URSI XXVIIth general assembly* (pp. 1–4).
29. Zhao, N., Wu, Z., Zhao, Y., & Quan, T. (2012). Population declining ant colony optimization multiuser detection in asynchronous cdma communications. *Wireless Personal Communications*, *62*(4), 783–792. doi:10.1007/s11277-010-0093-6.
30. Zhiwei, L., Xiaoming, P., Png, K. B., & Chin, F. (2007). Kronecker modeling for correlated shadowing in UWB MIMO channels. In *IEEE wireless communications and networking conference (WCNC 2007)* (pp. 1583–1587).

Author Biographies

José Carlos Marinello received his B.S. in Electrical Engineering (Summa Cum Laude) from Londrina State University, PR, Brazil, in December 2012. He is currently working towards his M.S. and Ph.D. at Londrina State University, Brazil. His research interests include physical layer aspects, specially heuristic and convex optimization of 3G and 4G MIMO systems.

A.2 Transmissão Multiusuário em Sistemas MIMO

Título: *BER Minimization in Multiuser Transmission Schemes for MIMO Communication with Increasing Capacity*;

Autores: José Carlos Marinello, Fernando Ciriaco & Taufik Abrão;

Submissão: Setembro de 2014;

Revista: *IEEE Transactions on Vehicular Technology.*

BER Minimization in Multiuser Transmission Schemes for MIMO Communication with Increasing Capacity

José Carlos Marinello, Fernando Ciriaco & Taufik Abrão

***Abstract*—Multiple-input-multiple-output (MIMO) systems substantially improve the spectral and/or energy efficiency of wireless communications. When employed for the sake of supporting multiple users, the space-division multiple access (SDMA) is capable of distinguishing the users supported within the same bandwith and time-slot with the aid of their unique, user-specific channel impulse response (CIR) in the uplink. By the same token, provided that the downlink CIR can be accurately predicted for the ensuing downlink transmission, transmit preprocessing or multiuser transmission (MuT) can be carried out for mitigating the multiuser interference at the base station (BS) for achieving a near-single-user performance at the mobile terminals. These MuT systems can be designed for optimizing different performance metrics, such as the bit-error-rate (BER), which is the focus of this paper. Numerous techniques have been proposed for directly minimizing the BER, which are indeed capable of approaching the single-user performance albeit at a high complexity. In order to mitigate the MuT's complexity, both heuristic and deterministic techniques have been amalgamated for improving the performance of SDMA MIMO MuT systems. Specifically, we investigate their performance *versus* computational complexity upon increasing the number of BS antennas as well as the number of users supported. Our results demonstrate that our Line Search Quasi-Newton algorithm relying on the penalty function approach (LSQN-PF) results in the best performance *versus* complexity trade-off in the context of high capacity multiuser systems.**

***Index Terms*—MIMO systems, SDMA, Multiuser Transmission, Minimum BER, performance-complexity tradeoff, Large systems.**

I. Introduction

MULTIPLE-input-multiple-output (MIMO) systems are able to offer significant advantages in terms of increased throughput and reliability at the price of more complex receiving/transmitting schemes in the wireless communication scenario [1]. In this configuration, several antennas are deployed in both transmitter and receiver side. In contrast to the family of multiplexing and diversity point-to-point schemes, where multiple antennas are activated by a single user to increase throughput and reliability, respectively, in space-division multiple access (SDMA) networks the multiple antennas are employed to support multiple users [2]. Many advantages of SDMA networks have been pointed out in [2], such as range extension, multi-path mitigation, increasing capacity, interference suppression, as well as compatibility with most of the existing modulation schemes, carrier frequencies and other specifications. For a practical reason, an usual configuration consists in equipping mobile terminals (MTs) with a single-receive-antenna, due to space limitations of pocket-sized MTs. Since MTs usually do not have the energy and the processing capability enough to perform multiuser detection (MuD) techniques, a useful strategy to improve the system's performance without harming the energy efficiency, in terms of MTs batteries autonomy, consists in shifting the processing necessary to combat the multiuser interference from the MTs to the base station (BS), leading to the appealing concept of multiuser transmission (MuT) [3], [4].

Multiuser transmission is also known as multiuser beamforming, when the BS is equipped with smart antennas, where the antenna weights (or precoding coefficients) are suitably combined in order to form a beam on the desired users, and a null on the interferers [5], [6], as if each antenna were mechanically pointed to each user. An actual MuT system realizes this directional antennas as the different beams of an antenna array, and the number of users K that can simultaneously communicate in the same frequency with BS is equal to the number of antennas employed in the BS, N. This implies that, when SDMA is employed in a network in conjunction with another multiple access technique, such as frequency (FDMA), time (TDMA) or code (CDMA) division multiple access, its capacity grows linearly with the number of antennas deployed in the BS [5]. This increasing capacity is straightforwardly achieved in TDMA/FDMA networks, since the users can share the same time/frequency slot, while for CDMA this increase occurs based on the enhancement of the desired signal. As the capacity of a CDMA network is determined by the signal-to-interference-plus-noise ratio (SINR), and the desired signal power enhances linearly with N, the capacity of the cell is increased in the same way.

At this point, it would be of great interest to employ a high number of antennas in the BS, since the number of users covered by the cell would increase in the same rate, giving rise to the promising Large-MIMO systems [7], [8]. These efficient communication systems have been in the focus of many technical works recently, and constitutes a new research paradigm, including communication theory, propagation, and electronics areas. Despite of the great benefits of such large systems, some limitations difficult its implementation: physical placement of this great number of antennas in practical BSs, lack of efficient low-complexity algorithms for uplink and downlink communication, and how obtain associated channel

J. C. Marinello, Fernando Ciriaco and T. Abrão are with State University of Londrina, Electrical Engineering Department (DEEL-UEL), Parana, Brazil. E-mails: zecarlos.ee@gmail.com, fciriaco@uel.br, taufik@uel.br

impulse responses (CIR) for these systems. In this paper, we address the second problem, investigating efficient algorithms for the downlink of Large-MIMO systems, while for the uplink similar optimization problem has been studied in [7], [8]. Note that the scenario investigated herein differs from that of [9], [10], where the number of antennas in BS grows keeping fixed the number of users. It was shown in [10] that according the ratio $\frac{N}{K}$ increases, the effects of uncorrelated noise and fast fading become less significant, vanishing at the limit of infinite number of antennas. Thus, the simplest linear techniques can be applied successfully in this scenario. However, as these benefits just arise with a large excess of BS antennas compared with K, typically some hundreds of BS antennas per user [11], the cases in which K grows in the same rate of N are also of great interest.

The MuT system can be designed to be optimum in many different aspects, such as interference supression [6], minimum mean squared error (MMSE) [3], [12], maximizing the energy efficiency [13], and/or the spectral efficiency [14], maximizing the sum rate [15], maximizing the minimum Euclidean distance between vectors of the noiseless received signal lattice [16], minimizing the bit-error-rate (MinBER) [17], [18], [19], among others. Indeed, in the MinBER-MuT approach, the BS is able to predict the error probability of a transmit signal for a given (estimated) CIR, noise power, and transmit data, and hence finds the precoder matrix that results in the MinBER design, which is the focus of this paper.

The idea of minimizing the error probability of a communication system as an optimization criterion was proposed in [20], applied to a CDMA adaptive detector. In the downlink of a multiple input single output system, the MinBER criterion was firstly applied in [17], and extended for higher-order modulation in [21]. Due to the fixed transmit power, the problem is formulated as a quadratically constrained optimization problem with a nonlinear objective function. Thus, it can be typically solved by some nonlinear optimization method, among which the *sequential quadratic programming* (SQP) is the state-of-the-art [22]; however, the SQP optimization method is able to minimize the average BER usually at the expense of an excessive computational burden.

In order to achieve the same performance but with a lower complexity, the problem can be formulated to optimize the transmit signal straightly [4], what simplifies the cost and constraint expressions, and reduces their dimensions. Incorporating the power constraint into the objective function, a new unconstrained MinBER-MuT formulation was proposed in [21], and it was claimed that its computational complexity is significantly reduced compared to the constrained approach; however, a more careful complexity analysis has not been reported. Another way to turn the MinBER optimization problem unconstrained is discussed in [23], in which the constraint is incorporated into the objective function by means of a penalty function, but no insightful description about how to obtain the penalty factors has been provided.

Recently, MinBER-MuT schemes inspired on heuristic methods have been proposed; for example, the *particle swarm optimization* (PSO) based method has been reported in [23]. The PSO was invoked to solve the penalty-based unconstrained MinBER-MuT optimization problem, improving system performance in comparison with the conventional MMSE-MuT scheme, while imposing a significantly reduced complexity compared to the state-of-the-art SQP-based MinBER-MuT design. Another efficient heuristic technique is the ant colony optimization (ACO), that was originally proposed to solve combinatorial optimization problems, such as the traveling salesman problem [24], but was extended for continuous problems in [25], and is inspired on the foraging behavior of ants in nature. Many works have been disseminated successfully applying this technique to several optimization problems that arise in telecommunications, such as multiuser detection in CDMA systems [26] [27], MIMO detection [28] [29], resource allocation on wireless networks [30] [31], among others.

The contributions of this work are fourfold: **i**) Different than [18] and [23][1], a new formulation for the MinBER-MuT optimization problem is proposed, which optimizes the transmit signal straightly, and its convexity is demonstrated. **ii**) Relying on this novel formulation, some deterministic and heuristic optimization algorithms are proposed and investigated in a unified way, being its implementation thoroughly detailed. **iii**) The surprising benefits of the MinBER-MuT approach are unveiled for the increasing capacity of the system, i.e. large number of BS antennas and covered users, in contrast to [19], [21], and [23], that investigated the MinBER-MuT system only under small dimensions. **iv**) Since on these high dimension systems the computational complexity of the algorithms may become prohibitive, a careful convergence analysis is carried out. This analysis allowed a complexity study that pointed out our Line Search Quasi-Newton algorithm relying on the penalty function approach (LSQN-PF) as the most efficient in these scenarios, in terms of the performance$\times$complexity trade-off.

This paper is organized as follows. Section II describes the MuT system model. Section III provides a detailed description of the MinBER-MuT techniques, designed from our formulation of the problem. Section IV shows numerical results that highlight the improved performance of the MinBER-MuT approach, as well as a complexity analysis of the investigated optimization techniques. Section V presents the conclusions, and some mathematical derivations are provided in the Appendices.

II. Multiuser Transmission

In the SDMA downlink transmission shown in Fig. 1, a base station equipped with N transmit antennas communicates with K non-cooperative energy-efficient mobile terminals, each equipped with a single receive antenna. Hence, considering *energy-efficient* downlink (DL) communication design, simple MTs are unable to perform sophisticated multiuser detection in order to mitigate the multiuser interference. As a solution

[1]In [18], both transmitter and receiver filters have been deployed, which may not be desirable when looking for high energy efficient receivers, while a conventional MuT system with a maximum of two antennas has been considered. In [23], the authors primarily investigated a heuristic PSO-based scheme, and the problem formulation was based on the dense precoding matrix, which leads to an increased complexity, as shown herein, both for PSO and SQP, while the number of antennas was limited to 4. No convexity analysis was made about the problem formulations in both papers.

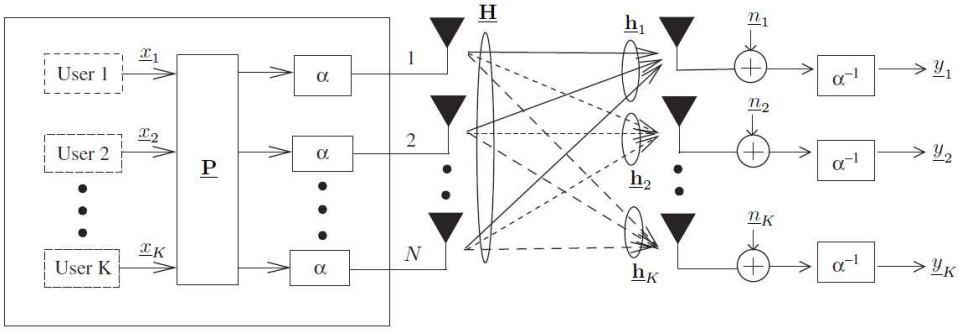

Figure 1. MuT-aided MIMO system with linear precoding. N Tx antennas at BS communicate with K mobile devices equipped with single-receiver-antenna.

adopted herein, the BS performs multiuser transmission, assuming that reliable estimates for the channel state information are available at the transmitter side. A MuT scheme similar to that proposed in [23], [32] is depicted in Fig. 1.

The transmit symbol vector is $\underline{\mathbf{x}} = [\underline{x}_1\, \underline{x}_2\, \ldots\, \underline{x}_K]^T$, where $\underline{x}_k$ denotes the transmitted symbol to the kth MT, with $\underline{x}_k$ assuming a value of the quadrature phase-shift keying (QPSK) alphabet, and $\{\cdot\}^T$ is the transpose operator. The $N \times K$ precoder matrix is

$$\underline{\mathbf{P}} = [\underline{\mathbf{p}}_1\, \underline{\mathbf{p}}_2\, \cdots\, \underline{\mathbf{p}}_K], \tag{1}$$

where $\underline{\mathbf{p}}_k$ is the precoder's coefficient vector for preprocessing the kth user's data stream. MuT schemes require the knowledge of the DL CIR at the transmitter side, since it works as the spatial signature of each user. It can be achieved by reciprocity in time division duplex (TDD) systems, assuming the channel coherence time sufficiently large, or by feedback channels in frequency division duplex (FDD) systems.

Hence, given a total fixed transmit power $\mathcal{P}_{\text{T}}$ at the BS, an appropriate scaling factor α should be used to fulfill this transmit power constraint such that:

$$\alpha = \sqrt{\frac{\mathcal{P}_{\text{T}}}{||\underline{\mathbf{P}}\mathbf{x}||^2}} = \sqrt{\frac{E_{\text{T}}}{T_{\text{s}}||\underline{\mathbf{P}}\mathbf{x}||^2}}, \tag{2}$$

where T_{s} is the symbol period and E_{T} is the total energy available at BS for each symbol period. At the receiver side, the received signal should be then multiplied by α^{-1}, ensuring a unity-gain transmission. This fixed transmit power constraint $\mathcal{P}_{\text{T}}$ allows a fair comparison between different pre-equalization schemes.

Besides, the instantaneous MIMO channel matrix

$$\underline{\mathbf{H}} = [\underline{\mathbf{h}}_1\, \underline{\mathbf{h}}_2\, \ldots\, \underline{\mathbf{h}}_K]^T \tag{3}$$

is assumed constant at each T_{s} time basis.

Finally, the equivalent baseband received signal vector can be written as:

$$\underline{\mathbf{y}} = \underline{\mathbf{H}}\,\underline{\mathbf{P}}\mathbf{x} + \alpha^{-1}\underline{\mathbf{n}}, \tag{4}$$

where $\underline{\mathbf{n}} \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$ represents the additive white Gaussian noise (AWGN) with variance $\sigma_n^2 = \frac{N_0}{2}$ per dimension, evidencing that the scaling factor α directly affects the amplitude of the background noise. Note that the received signal vector $\underline{\mathbf{y}}$ is ready to be detected, since the DL data was preprocessed aiming to mitigate the predicted interference, and hence the data detection at MT's is quite simplified in this scheme.

Since most of the optimization techniques deal with real-valued variables, it is interesting to construct a real equivalent model of the system. Hence, we define the equivalent real-valued channel matrix:

$$\mathbf{H} = \begin{bmatrix} \Re\{\underline{\mathbf{H}}\} & -\Im\{\underline{\mathbf{H}}\} \\ \Im\{\underline{\mathbf{H}}\} & \Re\{\underline{\mathbf{H}}\} \end{bmatrix} \in \mathbb{R}^{2K\times 2N}, \tag{5}$$

and the real-valued vectors

$$\mathbf{y} = \begin{bmatrix} \Re\{\underline{\mathbf{y}}\} \\ \Im\{\underline{\mathbf{y}}\} \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} \Re\{\underline{\mathbf{x}}\} \\ \Im\{\underline{\mathbf{x}}\} \end{bmatrix}; \quad \mathbf{n} = \begin{bmatrix} \Re\{\underline{\mathbf{n}}\} \\ \Im\{\underline{\mathbf{n}}\} \end{bmatrix} \in \mathbb{R}^{2K\times 1}, \tag{6}$$

where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are the real and imaginary operators, respectively.

Thus, (4) holds for the real equivalent system with a real $2N \times 2K$ precoder matrix $\mathbf{P}$. Conventionally, this matrix can be obtained by some linear approach, such as the *matched filtering* (MF) technique [33], [34], where $\mathbf{P}_{\text{MF}} = \alpha\,\mathbf{H}^T$; or the MMSE precoding scheme [34], with $\mathbf{P}_{\text{MMSE}} = \alpha\,\mathbf{H}^T\left(\mathbf{H}\mathbf{H}^T + \sigma_n\mathbf{I}_{2K}\right)^{-1}$, for $N \geq K$. Besides, the precoding matrix $\mathbf{P}$ can also be obtained by solving some optimization criterion.

A. The MinBER MuT Optimization Design

Given a real equivalent binary phase-shift keying (BPSK) symbol vector $\mathbf{x}$, the symbol-specific BER[2] of $\mathbf{y}$ at the receiver as a function of the $2N \times 2K$ precoder matrix predicted at the transmitter side is expressed as [23]:

$$P_{e,\mathbf{x}}(\mathbf{P}) = \frac{1}{2K}\sum_{k=1}^{2K} Q\left(\frac{\text{sgn}(x_k)\mathbf{h}_{k,:}\mathbf{P}\mathbf{x}}{\sigma_n}\right), \tag{7}$$

where $\mathbf{h}_{k,:}$ is the k-th row of $\mathbf{H}$, $Q(\cdot)$ is the standard Gaussian error function, and $\text{sgn}(\cdot)$ is the sign function. Hence, the MinBER-MuT design can be formulated by the following constrained optimization problem [17], [23]:

$$\begin{aligned} \mathbf{P}_{\text{MBER},\mathbf{x}} \;&=\; \arg\min_{\mathbf{P}} \quad P_{e,\mathbf{x}}(\mathbf{P}), \\ \text{s.t.}: \quad & h(\mathbf{P}) = ||\mathbf{P}\mathbf{x}||^2 - \mathcal{P}_{\text{T}} \leq 0. \end{aligned} \tag{8}$$

where the constraint $h(\mathbf{P}) \leq 0$ represents the maximum power available at the BS. Note that, in this scheme, the scaling factor α is not necessary, as opposed to the linear techniques,

[2]Different BER formulations can be adopted, such as the symbol-specific BER, which estimates the BER for the specific vector $\mathbf{x}$, and the average BER formulation, which estimates the BER for all possible vectors $\mathbf{x}$ [23]. Besides, different modulation orders $\mathcal{M}$ result in different BER expressions [21].

since the solution of the optimization problem must satisfy the power constraint. Hence, under this optimization design, the bit error rate predicted at the transmitter side is directly minimized based on the transmit signal $\mathbf{x}$, precoder matrix $\mathbf{P}$, as well as the current channel impulse response $\mathbf{H}$. The only parameter treated statistically is the AWGN noise of the receiver, which its variance σ_n^2 should be estimated at the transmitter side.

By defining the transmit signal $\mathbf{s} = \mathbf{P}\mathbf{x}$, and the estimated reliability vector $\mathbf{r} = \text{diag}(\text{sgn}(\mathbf{x}))\mathbf{H}\mathbf{s}$, the symbol-specific MinBER optimization problem can be formulated as:

$$\begin{aligned} \mathbf{s}_{\text{MBER},\mathbf{x}} &= \arg\min_{\mathbf{s}} \quad P_{e,\mathbf{x}}(\mathbf{s}), \\ \text{s.t.}: & \quad g(\mathbf{s}) = ||\mathbf{s}||^2 - \mathcal{P}_{\text{T}} \leq 0, \\ & \quad r_i \geq 0, \quad 1 \leq i \leq 2K, \end{aligned} \tag{9}$$

in which

$$P_{e,\mathbf{x}}(\mathbf{s}) = \frac{1}{2K}\sum_{k=1}^{2K} Q\left(\frac{\text{sgn}(x_k)\mathbf{h}_{k,:}\mathbf{s}}{\sigma_n}\right). \tag{10}$$

This formulation, which is based on the transmit signal, simplifies the optimization procedure, since the optimization variable changes from $\mathbf{P}$, with dimension $2N \times 2K$, to $\mathbf{s}$, with dimension $2N \times 1$, while the objective and the constraint functions are quite simplified. In addition, the $2K$ new constraints $r_i \geq 0$, for $1 \leq i \leq 2K$ are included just with the purpose of making the problem convex, as explained as follows.

Lemma 1: The MinBER-MuT optimization problem defined in (9) and relying on the transmit signal $\mathbf{s}$ is convex.

Proof: From (9), it can be seen that its feasible set is convex, since it corresponds to a closed region delimited by the intersection of a hypersphere of radius $\sqrt{\mathcal{P}_{\text{T}}}$ and $2K$ hyperplanes that pass through the origin. If each term in the summation of the objective function is convex, then its sum will be convex as well. As the function $Q(x)$ is convex for $x \geq 0$, and it is known from [35] that a composition of a convex function with an affine mapping preserves convexity, we have just to ensure that each Q function is evaluated in its convex region. Indeed, this condition is assured by the $2K$ additional constraints: $r_i \geq 0$, for $1 \leq i \leq 2K$. ■

It can also be proved that if $P_{e,\mathbf{x}}(\mathbf{s}) \leq \frac{1}{4K}$, the $2K$ additional constraints are satisfied. Suppose that not all of them are satisfied, what implies that at least one of them is violated. As $Q(x) > 0.5$ if $x < 0$, the minimum value of $P_{e,\mathbf{x}}(\mathbf{s})$ for this case is $\frac{1}{4K}$. Since practical values of $P_{e,\mathbf{x}}(\mathbf{s}_{\text{MBER}})$ are some magnitudes lower than this bound, the influence of these additional constraints can be considered weak, given that the minimization of the objective function automatically meets them. Thus, from an engineering perspective, several optimization techniques for this problem can be conceived assuming its relaxation, because the algorithm's complexity usually are highly dependent on the number of constraints, specially inequalities [22]. The resultant optimization problem can be stated as:

$$\begin{aligned} \mathbf{s}_{\text{MBER},\mathbf{x}} &= \arg\min_{\mathbf{s}} \quad P_{e,\mathbf{x}}(\mathbf{s}), \\ \text{s.t.}: & \quad g(\mathbf{s}) = ||\mathbf{s}||^2 - \mathcal{P}_{\text{T}} \leq 0. \end{aligned} \tag{11}$$

Although we have considered BPSK and QPSK modulations in our formulation for notational simplicity, extension to higher order modulations are straightforward [21], [36], as shown in Appendix A.

The MinBER MuT is a constrained nonlinear optimization problem, and the sequential quadratic programming (SQP) technique corresponds to the state-of-the-art for these cases [4], [22], [37]. An improved SQP algorithm relying on our novel formulation of the MinBER optimization problem is described in the next section, while its complexity is analysed in Section IV-B. Note that the SQP complexity can be excessive in some cases [4], [23], turning its application infeasible for practical high-rate information systems. As an alternative, there are some ways to turn this optimization problem unconstrained, and hence enabling the application of several unconstrained optimization techniques as done in [21] and [23].

B. Unconstrained MinBER MuT Formulations

1) Penalty Function (PF) Approach: In the same way as [23], the original constrained optimization problem in (11) is converted into an unconstrained one by the introduction of a penalty function. In doing so, the unconstrained problem automatically meets the transmit power constraint. Hence, the cost function is defined by:

$$F_{\mathbf{x}}(\mathbf{s}) = P_{e,\mathbf{x}}(\mathbf{s}) + \lambda\left[g(\mathbf{s})\right]^+, \tag{12}$$

where the operator $[x]^+ = \max(0, x)$, and the penalty factor $\lambda > 0$ should be chosen appropriately. Hence, the original constrained MinBER MuT optimization design (11) becomes the following unconstrained optimization problem:

$$\mathbf{s}_{\text{MBER},\mathbf{x}} = \arg\min_{\mathbf{s}} F_{\mathbf{x}}(\mathbf{s}). \tag{13}$$

Note that a suitable penalty function must be able to penalize the constraints violations in the exact amount, otherwise the objective function can become very difficult to optimize, if excessively high, or have a minimum in an unfeasible point, if somewhat low. As no further information is given in [23] about how the penalty factor λ can be properly obtained, we propose herein an expression to find it:

$$\lambda^* = \frac{1}{2\sqrt{\mathcal{P}_{\text{T}}}}\left|\left|\nabla_{\mathbf{s}} P_{e,\mathbf{x}}(\mathbf{s}_{\text{MBER}})\right|\right|, \tag{14}$$

where

$$\begin{aligned} \nabla_{\mathbf{s}} P_{e,\mathbf{x}}(\mathbf{s}) = & \frac{-1}{2K\sigma_n\sqrt{2\pi}} \times \\ & \sum_{k=1}^{2K} \text{sgn}(x_k)\, \exp\left(\frac{-(\mathbf{h}_{k,:}\mathbf{s})^2}{2\sigma_n^2}\right)(\mathbf{h}_{k,:})^T. \end{aligned} \tag{15}$$

The derivation can be found in Appendix B, and it is quite similar to the Lagrange multipliers obtained in [18]. Equation (14) says that, if $\mathbf{s}_{\text{MBER}}$ were known, we could obtain the exact penalty factor λ^*. However, it is obvious that the *a priori* knowledge of $\mathbf{s}_{\text{MBER}}$ does not make sense. Intuitively, as close as the current solution is to $\mathbf{s}_{\text{MBER}}$, more the penalty factor λ approaches λ^*. Thus, starting from an

initial solution, for instance the MMSE-MuT solution, an initial penalty factor can be obtained and a first iteration of the problem is solved. With this first solution, the penalty factor can be updated and, starting from the previous solution, a new unconstrained problem solved. Deploying some iterations of this optimization procedure, the solution obtained can be as close to the optimum as we wish. Moreover, the convergence for the penalty optimization procedure to $\mathbf{s}_{\text{MBER}}$ is presented in Appendix C.

2) Effective Noise (EN) Approach: Another way to transform the problem in (11) into an unconstrained one is proposed in [21]. Since scaling the received signal by α^{-1} directly affects the background noise amplitude, the equivalent noise measure $\sigma_{\text{eq}} = \sigma_n \sqrt{\frac{||\mathbf{s}||^2}{\mathcal{P}_{\text{t}}}}$ has been suggested, and hence, an unconstrained optimization function can be written as:

$$\mathbf{s}_{\text{MBER},\mathbf{x}} = \arg \min_{\mathbf{s}} \frac{1}{2K} \sum_{k=1}^{2K} Q\left(\frac{\text{sgn}(x_k)\mathbf{h}_{k,:}\mathbf{s}}{\sigma_{\text{eq}}}\right). \quad (16)$$

In [21], the problem is solved deploying a line search quasi-Newton method, which is also explained in the next section. Given these unconstrained formulations of the MinBER MuT optimization problem, we also investigate the application of heuristic techniques to solve them, such as PSO and ACO.

III. MinBER-MuT Techniques

In this Section, the application of some optimization techniques to the MinBER-MuT problem is described in details. We deal both with constrained techniques, based on the formulation in (11), and some unconstrained ones, based on the unconstrained formulations of Section II-B. The described techniques have been carefully adapted to the MinBER-MuT optimization problem considering a large number of antennas, aiming to improve substantially its performance with reduced complexity.

A. Constrained Optimization Techniques

1) Sequential Quadratic Programming (SQP): When solving nonlinear constrained optimization problems, the SQP is considered one of the most efficient methods. A comprehensive overview of this technique can be found in [22], while a short description with focus on implementation issues is presented in [37]. To simplify the analysis herein, the inequality constraint in (11) is treated as an equality one, with no loss of optimality since the optimum transmitted signal takes advantage of the maximum power available for transmission. At each iteration, the objective function is quadratically approximated in the vicinity of the current point $\mathbf{s}_k$ and the constraints are linearized. More precisely, a tangent plane is obtained from the linearization of the active constraints, and the direction in this plane that results in the maximum decrease of the objective function second-order approximation is obtained. Since in our formulation there is only an equality constraint, the tangent plane is the plane orthogonal to its gradient vector $\nabla_{\mathbf{s}} g(\mathbf{s}_k)$, and a $2N \times 2N-1$ basis $\mathbf{Z}_k$ aiming to describe the points in this plane can be obtained as $\mathbf{Z}_k = \mathbf{Q}(:, 2:2N)$, taking the QR decomposition of $\nabla_{\mathbf{s}} g(\mathbf{s}_k)$, such that $\nabla_{\mathbf{s}} g(\mathbf{s}_k) = \mathbf{QR}$. Hence, every $2N \times 1$ vector $\mathbf{d}_{tp}$ in the tangent plane can be denoted by $\mathbf{d}_{tp} = \mathbf{Z}_k \mathbf{d}$, in which $\mathbf{d} \in \mathbb{R}^{2N-1 \times 1}$. The problem of obtaining the maximum decrease direction $\mathbf{d}_{tp}$ of the objective function second-order approximation can be stated as [4], [37]

$$\underset{\mathbf{d}_{tp} \in \mathbb{R}^{2N \times 1}}{\text{minimize}} \ \frac{1}{2} \mathbf{d}_{tp}^T \nabla_{\mathbf{s}}^2 \mathcal{L}(\mathbf{s}_k, \lambda_k)\, \mathbf{d}_{tp} + \mathbf{d}_{tp}^T \nabla_{\mathbf{s}} P_{e,\mathbf{x}}(\mathbf{s}_k), \quad (17)$$

where $\mathcal{L}(\mathbf{s}, \lambda) = P_{e,\mathbf{x}}(\mathbf{s}) + \lambda\, g(\mathbf{s})$ is the Lagrangian, and $\nabla_{\mathbf{s}}^2 \mathcal{L}(\mathbf{s}, \lambda)$ is its Hessian or its BFGS approximation[3]. Incorporating the basis $\mathbf{Z}_k$ of $\mathbf{d}_{tp}$ in (17), leads to

$$\underset{\mathbf{d} \in \mathbb{R}^{2N-1 \times 1}}{\text{minimize}} \ \frac{1}{2} \mathbf{d}^T \mathbf{Z}_k^T \nabla_{\mathbf{s}}^2 \mathcal{L}(\mathbf{s}_k, \lambda_k)\, \mathbf{Z}_k \mathbf{d} + \mathbf{d}^T \mathbf{Z}_k^T \nabla_{\mathbf{s}} P_{e,\mathbf{x}}(\mathbf{s}_k), \quad (18)$$

which can be solved taking its derivative with respect to $\mathbf{d}$, and setting the result to equal a null vector. Finally, the search direction $\mathbf{d}_k$ is obtained as

$$\mathbf{d}_k = -\left(\mathbf{Z}_k^T \nabla_{\mathbf{s}}^2 \mathcal{L}(\mathbf{s}_k, \lambda_k)\, \mathbf{Z}_k\right)^{-1} \mathbf{Z}_k^T \nabla_{\mathbf{s}} P_{e,\mathbf{x}}(\mathbf{s}_k). \quad (19)$$

As discussed in [22], $\nabla_{\mathbf{s}}^2 \mathcal{L}(\mathbf{s}_k, \lambda_k)$ can be computed by the exact Hessian or by its BFGS approximation. The exact Hessian leads to a faster rate of local convergence, typically quadratic, but can be somewhat expensive to obtain, and needs to be positive definite to guarantee the existence of its inverse. On the other hand, although the BFGS computation leads to a superlinear convergence rate, it is more simple to compute, and there are some techniques to keep the matrix definite positive, whenever its Hessian is not, such as the Damped BFGS approximation [22]. So herein we have adopted the BFGS approach.

Obtained the search direction $\mathbf{d}_k$, the next step consists of a line search along it. Since moving in this direction makes the algorithm distances itself from the feasible region, this line search is designed to optimize a merit function [37], that trades off objective function minimization and constraints violation:

$$\Psi(\mathbf{s}_k + \beta \mathbf{d}_k) = P_{e,\mathbf{x}}(\mathbf{s}_k + \beta \mathbf{d}_k) + \mu\, h(\mathbf{s}_k + \beta \mathbf{d}_k), \quad (20)$$

where β is the step length, and μ is a penalty parameter, with efficient methods to obtain suitable values for μ found in literature, for instance [37]. Any one-dimensional optimization technique [38] can be used at this point, such as Fibonacci Search, Golden Section, and others. Herein, for simplicity, we have used the Wolfe conditions in a similar way as [21], in conjunction with an interpolation method. It is important to note that, in most cases, a unity step length is the solution of the one-dimensional search, since $\mathbf{d}_k$ is the exact solution of the quadratic subproblem. At the end of each iteration, as the moviment along $\mathbf{d}_k$ takes the current solution out of the feasible region, it is necessary to bring it back to the feasible set. In this problem, it can be easily done by scaling $\mathbf{s}_k$ by α, forcing $\mathbf{s}_k$ to stay on the hypersphere surface. Note that this simplification is quite simpler than traditional ways of moving $\mathbf{s}_k$ back to the feasible set [22].

2) Projected Steepest Descent: In many cases, the application of SQP to the MinBER MuT problem may be difficult or expensive, due to its excessive computational burden. Thus, we

[3]The BFGS method is the most popular quasi-Newton algorithm, named for its discoverers Broyden, Fletcher, Goldfarb, and Shanno.

adapt herein a quite simple method to solve the MinBER-MuT constrained problem, namely the projected steepest descent (PSD) [38]. It is well known that the classical steepest descent method presents low complexity per iteration, but suffers from slow convergence. Thus its efficiency should be characterized for each specific problem, depending on its difficulty. The algorithm proposed herein simply takes the classical steepest descent direction, given by the negative of the gradient in (15), and computes its projection into the tangent plane

$$\mathbf{d}_k = -\mathbf{Z}_k \, \mathbf{Z}_k^T \, \nabla_{\mathbf{s}} P_{e,\mathbf{x}}(\mathbf{s}_k). \tag{21}$$

Having found the search direction $\mathbf{d}_k$, a one-dimensional search is performed to compute the step length, as in SQP. However, since the initial solution is not as accurate as in SQP, we have deployed the Golden Section Search algorithm [38] for solving the one-dimensional search.

B. Unconstrained Optimization Techniques

In this paper, we have considered two ways to turn the MinBER MuT optimization problem unconstrained as described in Section II-B: the penalty function of [23], and the effective noise approach of [21]. Then we have investigated the application of some deterministic and heuristic unconstrained optimization techniques.

1) Line Search Quasi-Newton (LSQN) Method: Given the effective noise approach to turn the MinBER-MuT problem unconstrained, in [21] is deployed a line-search quasi-Newton method to solve it. This iterative unconstrained optimization technique finds the search direction that leads to the maximum decrease in the second order approximation of the objective function, like the Newton's Method [22], [38]. To prevent the complexity of computing the Hessian matrix of the objective function and its inverse, as well as the positive definiteness requirement of the Hessian matrix, it is used the well-known BFGS strategy [22], [38] to approximate the Hessian inverse, necessary to obtain the Newton direction.

Found the search direction, a one-dimensional problem should be solved in order to compute the step length. However, find a global minimizer can be a computationally prohibitive task even for this problem. For this sake, widely used conditions for efficient inexact line search strategies are invoked, namely the Wolfe conditions [21], [22]. The first one ensures that the adopted step propitiates sufficient decrease in the objective function; while the second, also known as curvature condition, forces the absolute value of the one-dimensional curve slope evaluated in the new point to be smaller than the previous one, being a warning that we cannot expect much more decrease in this direction. If a step simultaneously satisfy both Wolfe conditions, it provides an appreciable decrease in the objective function, at the same time it is sufficiently near the curve valley, so it makes sense to terminate the computation of the step length and evaluate a new search direction.

Herein, we extend the application of this technique for the penalty function approach of Section II-B1, with the proposed penalty parameter in (14). However, to make it possible, it is necessary to perform slight modifications in the penalty function formulation. More precisely, it is clear from (12) that $F_{\mathbf{x}}(\mathbf{s})$ is not continuously differentiable, which is a requirement for the application of gradient-based techniques. So, we can circumvent this issue by applying a slight modification in the penalty formulation

$$F_{\mathbf{x}}(\mathbf{s}) = P_{e,\mathbf{x}}(\mathbf{s}) + \lambda \left([g(\mathbf{s})]^+ \right)^{1.1}, \tag{22}$$

in such a way that its gradient becomes continuously differentiable:

$$\nabla_{\mathbf{s}} F_{\mathbf{x}}(\mathbf{s}) = \nabla_{\mathbf{s}} P_{e,\mathbf{x}}(\mathbf{s}) + 1.1\lambda \left([g(\mathbf{s})]^+ \right)^{0.1} \nabla_{\mathbf{s}} g(\mathbf{s}). \tag{23}$$

2) PSO Assisted MinBER-MuT: In [23], the PSO heuristic is deployed to solve the MinBER-MuT problem, turned unconstrained by the penalty function approach. This population-based stochastic optimization technique is inspired by the social behavior of bird flocks or fish shoals. In the search procedure, a swarm of individuals, namely the particles, adjusts its positions making decisions based on its best own location (cognitive information), and based on the best location found by the whole swarm (social information). Using the time variant acceleration coefficients (TVAC), more importance is given to the cognitive information in the initial steps, improving the exploitation capability of the algorithm, while the social information is preferred in the final iterations, leading to a more accurate convergence of the technique.

Herein, the PSO assisted MinBER-MuT of [23] is also extended to the effective noise approach described in Section II-B2.

3) ACO for Continuous Domains: As shown in [25], in order to adapt the original ACO algorithm [24] to continuous domain optimization problems, it is necessary to represent the pheromone information, which for discrete problems was represented as a table, in the context of a problem with continuous variables. It can be made by using an archive T, that stores the T_{sz} best solutions in ascending order of their respective cost function values found so far by the algorithm.

The transition probabilities, which were previously defined as a discrete probability density function (PDF), also shall be adjusted, becoming a continuous PDF, represented herein by a weighted sum of several Gaussian PDF functions, namely Gaussian kernel, for each problem dimension. As a practical simplification, the sampling of this Gaussian kernel is equivalently performed in two phases: first, a discrete sampling, in which it is chosen what Gaussian function will be sampled (related to the ℓ_g-th solution of T), and then a continuous one, related to it. The probability of choice for each Gaussian function is given by:

$$\Pr(\ell) = \frac{w_\ell}{\sum_{r=1}^{T_{\text{sz}}} w_r}, \tag{24}$$

where w_ℓ is the weight of the ℓ-th solution of T, defined as:

$$w_\ell = \frac{1}{qT_{\text{sz}}\sqrt{2\pi}} \exp\left\{ -\frac{(\ell-1)^2}{2q^2 T_{\text{sz}}^2} \right\}, \tag{25}$$

where q is a parameter of the algorithm, that influences its search diversity since q is related to the probability of being chosen the best solutions of the archive.

At the beginning of each iteration, a new Gaussian function is selected, according to the probabilities given by (24). Then, on that iteration, each one of the N_{ants} available ants follows a new path, which is done sampling the chosen Gaussian at each dimension of the problem; in this case, each element of the transmit signal vector $\mathbf{s} \in \mathbb{R}^{2N\times 1}$. For each dimension, the mean of the Gaussian PDF is given by the value of the element of the ℓ_g-th solution on T: the element $s^i_{\ell_g}$, and the respective standard deviation given by:

$$\sigma^i_\ell = \xi \sum_{e=1}^{T_{\text{sz}}} \frac{|s^i_e - s^i_{\ell_g}|}{T_{\text{sz}} - 1}, \tag{26}$$

where ξ is a parameter related with the algorithm's convergence speed. Note that, as lowest the ξ parameter is, more close the further generated solutions will be to $\mathbf{s}_{\ell_g}$, decreasing the variability of the algorithm, and leading to a enforced convergence. On the other hand, if the ξ value tends to 1, the further generated solutions will present the same deviation than the ones in the current archive, and hence the algorithm's convergence will be very slow.

Finally, at the end of each ACO algorithm iteration, having each ant chosen its path, the archive T is updated. The generated paths are included to T, which is sorted again according to the cost functions of each element. Then, the last N_{ants} elements are discarded, in order to keep always the same size of the solution archive T_{sz}. The algorithm's search finishes when all the solutions of T are equal or when the maximum number of iterations is reached, returning the first solution stored on the top of archive.

In Section IV, we investigate the application of the ACO algorithm for the MinBER-MuT problem, considering both unconstrained approaches described in Sections II-B1 and II-B2.

IV. Numerical Results

Figures of merit, generated by Monte Carlo simulations, for the investigated optimization techniques have been carefully analysed in this Section, including convergence, computational complexity and BER performance. Besides, we have investigated their convergence when the dimension of the problem increases. Having found how the convergence varies with the increasing dimension of the problem while assuming $N = K$, the computational complexity for each MinBER precoding optimization technique is investigated; hence, the most efficient of them, in terms of the performance×complexity trade-off, is highlighted.

A. Performance and Convergence Analysis

For the eight considered techniques, we have investigated its convergence for two MIMO configurations: $N = K = 4$, and $N = K = 12$, both deploying 4-QAM modulation. The number of iterations necessary for convergence is carefully analysed for each technique, in each MIMO configuration, being used in the complexity calculation in the sequel. The MinBER benchmark reference was obtained solving the constrained problem, with the aid of the Matlab Optimization Toolbox [37]. For the PSO-based MuT techniques, the set of parameters used was obtained from [23], that is null inertia weight, the time variant acceleration coefficients, and a swarm size $\mathtt{s} = 20$ particles. On the other hand, for the ACO-MuT techniques, the parameters set was obtained from [25], being adopted $\xi = 0.85$, $q = 0.0001$, and archive size $T_{\text{sz}} = 50$. The only exception is the number of ants $\mathtt{a}$, since the algorithm deploying $\mathtt{a} = 4$ ants has shown a better convergence for this problem than $\mathtt{a} = 2$. Furthermore, for the heuristic techniques based on the penalty approach, we have fixed the outer loops in 4 iterations, varying only the inner iteration number.

Analysing the convergence of the four investigated deterministic optimization techniques depicted in Fig. 2 and 3, one can see that these techniques have reached the MinBER reference in a reasonable number of iterations. The quasi-Newton techniques (SQP-MuT and LSQN-MuT with effective noise and penalty function approaches) have needed tens of iterations to convergence, since the quadratic formulation of the problem results in a more accurate search direction. On the other hand, the projected steepest descent (PSD-MuT) needed some hundreds of iterations to convergence, since the search direction is not so accurate; however, as its complexity per iteration is reduced due to the simple search direction computation, a more elaborated complexity analysis should be invoked to point out the most efficient among them.

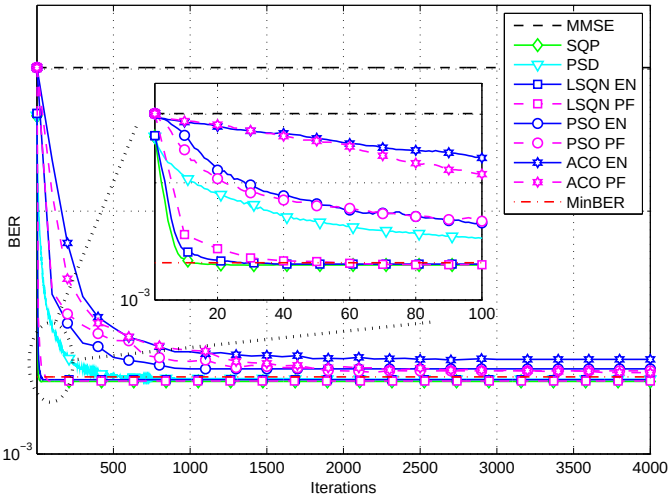

Figure 2. Convergence of the (un)constrained MinBER-MuT optimization techniques under $N = K = 4$, 4-QAM, $SNR = 15$ dB.

For the heuristic techniques, convergence figures have shown that the iteration number to convergence, that is already high for low system dimension, increases to excessively high values for higher dimensions of the MuT optimization problem. Furthermore, in some cases the algorithm did not converge to the MinBER bound, being trapped in local minima[4], as the ACO-EN algorithm in Fig. 2, or the iteration number to convergence becomes so high that it was difficult to be computed even in the simulations, as PSO-PF and PSO-EN in Fig. 3.

[4]Note that the Effective Noise Approach does not assure Convexity, and thus some local minima may appear, as discussed in [21].

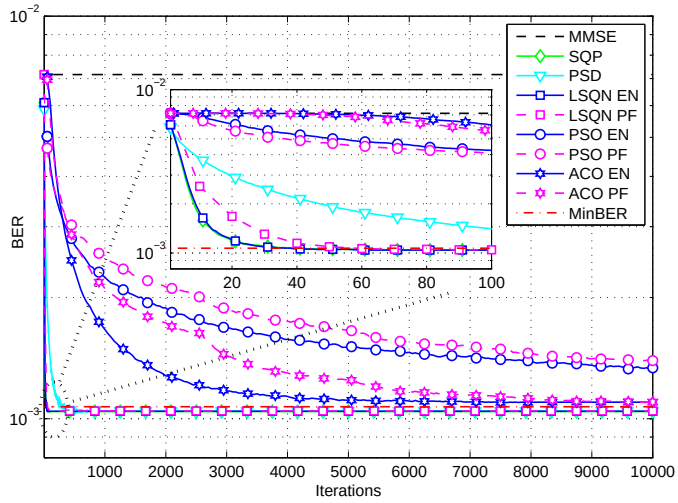

Figure 3. Convergence of the (un)constrained MinBER-MuT optimization techniques under $N = K = 12$, 4-QAM, $SNR = 5$ dB.

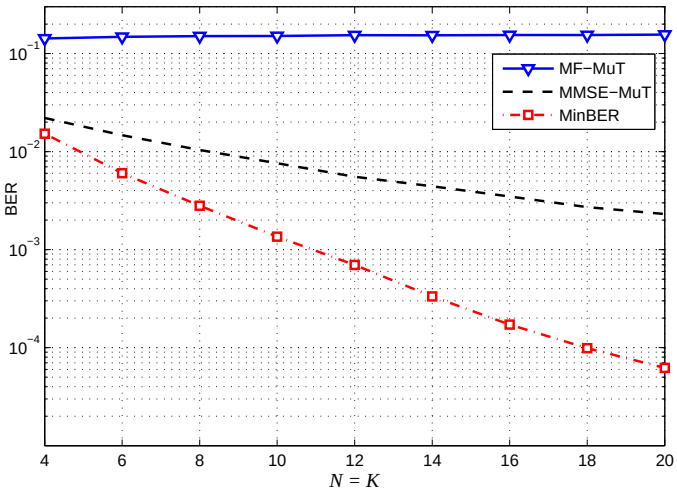

Figure 5. Performance of the MinBER-MuT subject to the increasing dimension of the system ($N = K$), for $SNR = 6$ dB and 4-QAM.

BER *versus* SNR performance for the MinBER MuT optimization techniques, in comparison with the MMSE-MuT, for three MIMO configurations have been considered in Fig. 4, and the achieved diversity orders have been indicated. It can be seen that the performance gain achieved by the MinBER-MuT techniques becomes more noticeable for higher dimensions, i.e., when the number of antennas N and/or users K increases. For a target uncoded BER of 10^{-3}, usually sufficient for systems applying forward error correction coding, the SNR gain passes from ≈ 3.5 dB, in the 4×4 system, to ≈ 7.5 dB, in the 12×12 and 20×20 systems. Furthermore, while the diversity order increases slowly for the MMSE-MuT scheme, this occurs in a quite strongly way for the MinBER-MuT approach, evidencing that very low error rates can be achieved at the expense of low transmit powers. Figure 5 corroborates this result, showing the BER performance with the increasing dimension of the system, for a fixed SNR of 6 dB.

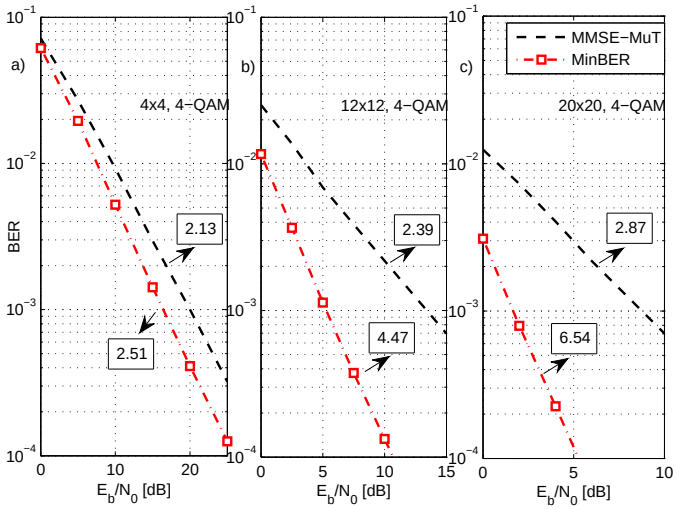

Figure 4. BER Performance of the MinBER-MuT with increasing SNR of the system, for 4-QAM and: a) $N = K = 4$; b) $N = K = 12$; c) $N = K = 20$.

B. Complexity Analysis

The complexity of an algorithm can be evaluated in terms of the total number of floating-point operations (*flops*[5]) [39]. For instance, an addition between two $m \times n$ matrices spends mn *flops*, while a multiplication between a $m \times n$ matrix $\mathbf{A}$ and a $n \times p$ matrix $\mathbf{B}$ spends $(2n-1)mp$ *flops* [39]. Table I shows the number of operations (*flops*) needed for each linear precoding transmitter technique analyzed herein. n represents the system dimension ($N = K$), $\mathtt{s}$ is the number of particles for the PSO algorithm, T_{sz} and $\mathtt{a}$ represent the archive size and the number of ants, respectively, for the ACO algorithm. ϕ represents the number of iterations needed for each Golden section search, such that $\phi = 1 + \frac{\ln(\delta) - \ln(\alpha_{\max})}{\ln(0.618)}$, being δ the uncertainty of the solution, where $\delta = 10^{-10}$ has been adopted herein, and $\alpha_{\max}$ the maximum step length, that we used $\alpha_{\max} = N^2$. For the Penalty based techniques, that present inner and outer loops, It corresponds to the number of inner iterations and Ot to the outer iterations. Besides, It corresponds to the number of iterations for the conventional iterative algorithms.

Fig. 6 compares the computational complexity of the techniques with the increasing number of transmit antennas and MTs, assuming $N = K$ and 4-QAM modulation. Since the iteration number to convergence changes when the system dimension increases, as shown in the convergence graphs of Fig. 2 and 3, and for sake of simplicity, a linear interpolation for the number of iterations to convergence has been adopted. Hence, for each optimization technique, we have registered the iteration number to convergence in the 4×4 and 12×12 MIMO systems, as shown in Table II, and then the intermediate values are obtained from a linear approximation between them, in order to obtain a fairer comparison. As a result, it can be seen from Fig. 6 that the heuristic techniques presented computational complexities much higher than the deterministic ones; therefore, the PSO and ACO heuristic optimization techniques are not so efficient in this problem.

[5]A *flop* is defined as an addition, subtraction, multiplication or division between two floating point numbers.

Table I
NUMBER OF OPERATIONS (FLOPS) FOR EACH MUT TECHNIQUE.

MuT	Number of Operations
SQP	$(56n^3 + 28n^2 + 86n + 16)It_{\text{SQP}}$
PSD	$(40n^2 + 20n + 10 + \phi(16n^2 + 12n + 12))It_{\text{PSD}}$
LSQN-EN	$(220n^2 + 72n + 63)It_{\text{QNEN}}$
LSQN-PF	$(16n^2 + 10n + 8 + It_{\text{QNPF}}(136n^2 + 114n + 61))Ot_{\text{QNPF}}$
PSO-EN	$(8n^2 + 18n + 6)\mathtt{s} + ((8n^2 + 28n + 9)\mathtt{s} + 9)It_{\text{PSOEN}}$
ACO-EN	$(8n^2 + 18n + 5 + \log(T_{\text{sz}}))T_{\text{sz}} + ((8n^2 + 12n + 8 + T_{\text{sz}}(4n + 1))\mathtt{a} + (T_{\text{sz}} + \mathtt{a})\log(T_{\text{sz}} + \mathtt{a}))It_{\text{ACOEN}}$
PSO-PF	$(16n^2 + 10n + 8 + (8n^2 + 18n + 5)\mathtt{s} + ((8n^2 + 28n + 9)\mathtt{s} + 9)It_{\text{PSOPF}})Ot_{\text{PSOPF}}$
ACO-PF	$(16n^2 + 10n + 8 + (8n^2 + 18n + 5 + \log(T_{\text{sz}}))T_{\text{sz}} + \ldots$ $((8n^2 + 12n + 8 + T_{\text{sz}}(4n + 1))\mathtt{a} + (T_{\text{sz}} + \mathtt{a})\log(T_{\text{sz}} + \mathtt{a}))It_{\text{ACOPF}})Ot_{\text{ACOPF}}$

The projected steepest descent (PSD) technique, although less complex than the heuristic techniques, also presented an excessive complexity, due to the very high number of iterations needed to convergence induced by the simple and inaccurate search direction. On the other hand, the algorithms that deploy second order formulations of the problem, taking advantage of quasi-Newton techniques, have demonstrated the lowest complexities, while achieving the MinBER performance bound for all the MIMO configurations evaluated. As seen in Fig. 6.b, the most efficient MuT optimization technique, in terms of the performance×complexity trade-off, depends on the dimension of the system. While the SQP implemented herein, that directly optimizes the transmit signal vector $\mathbf{s}$, has shown itself as the less complex optimization technique for lower dimensions ($N \leq 4$), and the LSQN technique based on the Effective Noise formulation (LSQN-EN) of [21] the most efficient for moderate number of antennas ($4 < N \leq 11$), the LSQN technique proposed in Section III-B1, that relies on the Penalty function approach (LSQN-PF), has been the most efficient one for higher dimensions ($N \geq 12$).

Finally, although both LSQN-EN and LSQN-PF algorithms are quite similar, the penalty function formulation leads to more simplified expressions, turning the computation of cost functions, gradients, and BFGS updates less complex for higher dimensions.

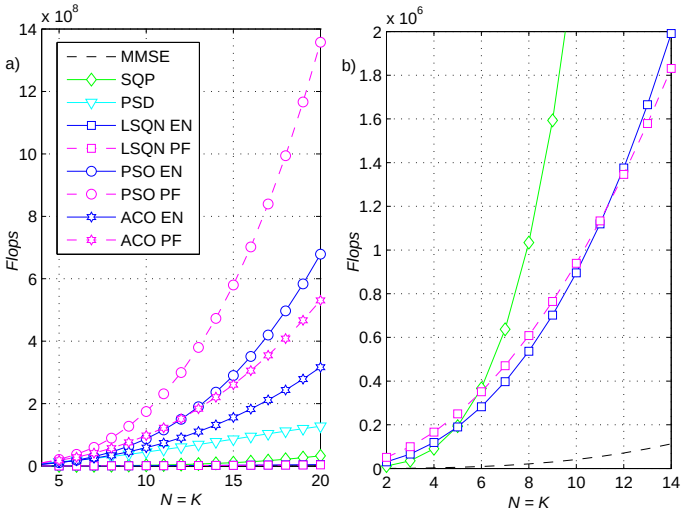

Figure 6. Number of operations × number of antennas: a) All considered techniques; b) A zoom on the less complex MinBER-MuT algorithms.

Table II
NUMBER OF ITERATIONS TO CONVERGENCE FOR EACH MUT OPTIMIZATION TECHNIQUE, AND MIMO CONFIGURATION.

MuT	4×4	12×12
SQP	20	45
PSD	500	400
LSQN-EN	30	40
LSQN-PF[1]	6×10	6×10
PSO-EN	1000	10000
ACO-EN	1500	6000
PSO-PF[1]	4×500	4×2500
ACO-PF[1]	4×600	4×2250

[1]Outer Iterations × Inner iterations

V. CONCLUSION

In this paper, the MIMO SDMA multiuser transmission system under the optimization metric of BER minimization has been investigated. A convex formulation of the problem has been proposed, and quite distinct optimization techniques, among deterministic and heuristic ones, operating on its constrained form, or in unconstrained approaches, have been investigated. A convergence analysis of those eight MinBER-MuT optimization techniques, regarding the increasing system dimension, has allowed a more elaborated computational complexity study. Then the efficiency of those techniques has been characterized, in terms of the performance×complexity trade-off. We have found that the quasi-Newton techniques are the most efficient ones, presenting the lowest complexities with the increasing system dimension, while always achieving the MinBER performance reference. Among them, the LSQN-PF algorithm proposed herein has presented the lowest computational complexity for higher dimensions ($N \geq 12$). It was demonstrates that the advantages of the MinBER-MuT scheme become more noticeable in these scenarios, since the high number of BS antennas and users covered in the cell leads to a system with very high multiuser capacity, operating in a range of low SNR values, with a quite appreciable energy efficiency. Indeed, a more complete analysis for the Large-MIMO system may be taking into account a multicell environment, where the pilot contamination constitutes a much more adverse effect than background noise and small scale fading, which corresponds to the continuity of this work.

APPENDIX A
PROBABILITY OF ERROR FOR 16-QAM MODULATION

When considering 16-QAM modulation, Eq. (10) should be replaced by [21]:

$$P_{e,\mathbf{x}}(\mathbf{s}) = \frac{1}{2K}\sum_{k=1}^{2K} Q\left(\frac{\mathrm{sgn}(x_k)\mathbf{h}_{k,:}\mathbf{s}}{\sigma_n}\right) + \mathrm{sgn}(|x_k|-T)\cdot\left[Q\left(\frac{-T-\mathbf{h}_{k,:}\mathbf{s}}{\sigma_n}\right) - Q\left(\frac{T-\mathbf{h}_{k,:}\mathbf{s}}{\sigma_n}\right)\right],$$

where T is the minimum distance between adjacent symbols of the 16-QAM modulation, and it is associated with the average power of the symbols ($\overline{p}_{avg}$) by $T = 2\sqrt{\frac{\overline{p}_{avg}}{21}}$.

APPENDIX B
PENALTY FACTOR DERIVATION

In order to derive an expression for the penalty factor λ, we may invoke the Lagrange formulation of the problem: $\mathcal{L}(\mathbf{s},\lambda) = P_{e,\mathbf{x}}(\mathbf{s}) + \lambda_k\, g(\mathbf{s})$. The first-order necessary KKT conditions for this case are

$$\nabla_{\mathbf{s}}\mathcal{L}(\mathbf{s}^*,\lambda^*) = 0, \tag{27}$$

$$\nabla_{\lambda}\mathcal{L}(\mathbf{s}^*,\lambda^*) = 0. \tag{28}$$

being $\mathbf{s}^*$ an stationary point for the problem (11), and λ^* the optimal Lagrange multiplier. Note that equation (28) reduces itself to the equality constraint $g(\mathbf{s}) = ||\mathbf{s}||^2 - \mathcal{P}_{\mathrm{T}} = 0$, which tells us that $||\mathbf{s}|| = \sqrt{\mathcal{P}_{\mathrm{T}}}$. Equation (27), by its turn, says that

$$\nabla_{\mathbf{s}}\mathcal{L}(\mathbf{s}^*,\lambda^*) = \nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}^*) + \lambda^*\,\nabla_{\mathbf{s}}g(\mathbf{s}^*) = 0.$$

Since $\nabla_{\mathbf{s}}g(\mathbf{s}) = 2\,\mathbf{s}$, and $||\nabla_{\mathbf{s}}g(\mathbf{s})|| = 2\sqrt{\mathcal{P}_{\mathrm{T}}}$ from (28), the above equation can be rearranged as

$$||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}^*)|| = 2\,\lambda^*\,\sqrt{\mathcal{P}_{\mathrm{T}}}.$$

Isolating λ^*, and since we call $\mathbf{s}^*$ for this problem as $\mathbf{s}_{\mathrm{MBER}}$, we have

$$\lambda^* = \frac{1}{2\sqrt{\mathcal{P}_{\mathrm{T}}}}\,||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_{\mathrm{MBER}})||\,. \tag{29}$$

APPENDIX C
CONVERGENCE OF THE PENALTY APPROACH

Herein we offer some insight on the convergence of the iterative penalty method, with the proposed penalty parameters given by (14). Using as starting point the MMSE solution, i.e. $\mathbf{s}_0 = \mathbf{s}_{\mathrm{MMSE}}$, we have

$$\lambda_0 = \frac{1}{2\sqrt{\mathcal{P}_{\mathrm{T}}}}\,||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_{\mathrm{MMSE}})||\,. \tag{30}$$

Minimizing $\mathcal{L}(\mathbf{s},\lambda_0)$, we arrive at the point $\mathbf{s}_1$ such that

$$\nabla_{\mathbf{s}}\mathcal{L}(\mathbf{s}_1,\lambda_0) = \nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_1) + \lambda_0\,\nabla_{\mathbf{s}}g(\mathbf{s}_1) = 0. \tag{31}$$

Equation (31) says that $\mathbf{s}_1$ satisfies the first-order necessary KKT condition (27) for λ_0. However, we cannot assure the optimality of this solution since (28) is not necessarily satisfied.

Substituting the expression for $\nabla_{\mathbf{s}}g(\mathbf{s}_1)$ and λ_0, we can isolate $\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_1)$ as

$$\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_1) = \frac{-1}{\sqrt{\mathcal{P}_{\mathrm{T}}}}\,||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_0)||\,\mathbf{s}_1. \tag{32}$$

Computing λ_1 based on $\mathbf{s}_1$, we have that

$$\lambda_1 = \frac{1}{2\left(\sqrt{\mathcal{P}_{\mathrm{T}}}\right)^2}\,||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_0)||\,||\mathbf{s}_1||\,, \tag{33}$$

and minimizing $\mathcal{L}(\mathbf{s},\lambda_1)$, we arrive at the point $\mathbf{s}_2$ such that

$$\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_2) = \frac{-1}{\left(\sqrt{\mathcal{P}_{\mathrm{T}}}\right)^2}\,||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_0)||\,||\mathbf{s}_1||\,\mathbf{s}_2. \tag{34}$$

Extending the result to the n-th iteration, we arrive at a point $\mathbf{s}_n$ such that

$$\mathbf{s}_n = \frac{-\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s})|_{\mathbf{s}=\mathbf{s}_n}\left(\sqrt{\mathcal{P}_{\mathrm{T}}}\right)^n}{||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s})|_{\mathbf{s}=\mathbf{s}_0}||\prod_{i=1}^{n-1}||\mathbf{s}_i||}, \tag{35}$$

which can be written in terms of $\mathbf{s}_{n-1}$ as

$$\mathbf{s}_n = \sqrt{\mathcal{P}_{\mathrm{T}}}\,\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_n)\left(\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_{n-1})\right)^{\dagger}\frac{\mathbf{s}_{n-1}}{||\mathbf{s}_{n-1}||}, \tag{36}$$

where $(\cdot)^{\dagger}$ is the pseudo-inverse operator. Taking the norm of above expression leads to

$$||\mathbf{s}_n|| = \sqrt{\mathcal{P}_{\mathrm{T}}}\,\frac{||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_n)||}{||\nabla_{\mathbf{s}}P_{e,\mathbf{x}}(\mathbf{s}_{n-1})||}. \tag{37}$$

From (37), one can see that if the iterative algorithm converges, i.e. if $\mathbf{s}_n = \mathbf{s}_{n-1}$ for some finite n, $||\mathbf{s}_n|| = \sqrt{\mathcal{P}_{\mathrm{T}}}$, satisfying (28). Since (27) is also satisfied by $\mathbf{s}_n$ for λ_{n-1}, the first-order necessary KKT conditions are simultaneously satisfied. Indeed, the convergence of an algorithm that generates a sequence of points that strictly decreases a descent function is assured by the Global Convergence Theorem [38], and thus the convergence assumption is valid.

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. England: Cambridge University Press, 2010.

[2] L. Hanzo, Y. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus Non-coherent and Cooperative Turbo Transceivers*. United Kingdom: IEEE Press and John Wiley & Sons, 2010.

[3] B. Vojcic and W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Transactions on Communications*, vol. 46, no. 10, pp. 1346 –1355, oct 1998.

[4] R. Irmer, "Multiuser transmission in code division multiple access mobile communications systems," Ph.D. dissertation, Technique University of Dresden, Dresden, Germany, April 2005.

[5] A. Molisch, *Wireless Communications*. United Kingdom: John Wiley & Sons, 2010.

[6] L. Guo and Y.-F. Huang, "Interference suppression for multiuser downlink transmission in frequency-selective fading channels," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4386–4397, 2008.

[7] N. Srinidhi, T. Datta, A. Chockalingam, and B. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *Communications, IEEE Transactions on*, vol. 59, no. 11, pp. 2955–2963, November 2011.

[8] K. Vardhan, S. Mohammed, A. Chockalingam, and B. Rajan, "A low-complexity detector for large MIMO systems and multicarrier CDMA systems," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 3, pp. 473–485, April 2008.

[9] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[10] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[11] J. Hoydis, S. ten Brink, and M. Debbah, "Massive mimo in the UL/DL of cellular networks: How many antennas do we need?" *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 160–171, February 2013.

[12] J. Choi and S. Perreau, "MMSE multiuser downlink multiple antenna transmission for CDMA systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 1564–1573, 2004.

[13] S. He, Y. Huang, S. Jin, F. Yu, and L. Yang, "Max-Min energy efficient beamforming for multicell multiuser joint transmission systems," *IEEE Communications Letters*, vol. 17, no. 10, pp. 1956–1959, 2013.

[14] D. Nguyen, L.-N. Tran, P. Pirinen, and M. Latva-aho, "Precoding for full duplex multiuser MIMO systems: Spectral and energy efficiency maximization," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4038–4050, 2013.

[15] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate maximization in multi-antenna broadcast channels with linear preprocessing," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 9, pp. 2338–2342, September 2006.

[16] X. Xu and Z. Chen, "Recursive construction of minimum euclidean distance-based precoder for arbitrary-dimensional MIMO systems," *Communications, IEEE Transactions on*, vol. 62, no. 4, pp. 1258–1271, April 2014.

[17] R. Irmer, R. Habendorf, W. Rave, and G. Fettweis, "Nonlinear multiuser transmission using multiple antennas for TDD-CDMA," in *Proc. Int. Symp. on Wireless Personal Multimedia Communications (WPMC'03)*, Yokosuka, Japan, Oct. 2003, pp. 251–255.

[18] A. Hjorungnes, P. S. R. Diniz, and M. L. R. De Campos, "Jointly minimum BER transmitter and receiver FIR MIMO filters for binary signal vectors," *Signal Processing, IEEE Transactions on*, vol. 52, no. 4, pp. 1021–1036, April 2004.

[19] N. Wang and S. Blostein, "Approximate minimum BER power allocation for MIMO spatial multiplexing systems," *Communications, IEEE Transactions on*, vol. 55, no. 1, pp. 180–187, Jan 2007.

[20] N. B. Mandayam and B. Aazhang, "Gradient estimation for sensitivity analysis and adaptive multiuser interference rejection in code-division multiple-access systems," *Communications, IEEE Transactions on*, vol. 45, no. 7, pp. 848–858, Jul 1997.

[21] R. Habendorf and G. Fettweis, "Nonlinear optimization for the multiuser downlink," in *The 13th European Wireless Conference*, Paris, France, April 2007.

[22] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer Series in Operations Research, 1999.

[23] W. Yao, S. Chen, S. Tan, and L. Hanzo, "Minimum bit error rate multiuser transmission designs using particle swarm optimisation," *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, pp. 5012–5017, 2009.

[24] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28 –39, nov. 2006.

[25] K. Socha and M. Dorigo, "Ant colony optimization for continuous domains," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1155 – 1173, 2008.

[26] C. Xu, B. Hu, L.-L. Yang, and L. Hanzo, "Ant-colony-based multiuser detection for multifunctional-antenna-array-assisted MC DS-CDMA systems," *Vehicular Technology, IEEE Transactions on*, vol. 57, no. 1, pp. 658–663, Jan 2008.

[27] J. C. Marinello, R. N. de Souza, and T. Abrão, "Ant colony input parameters optimization for multiuser detection in DS/CDMA systems," *Expert Systems with Applications*, vol. 39, no. 17, pp. 12 876 – 12 884, 2012.

[28] J. C. Marinello and T. Abrão, "Lattice reduction aided detector for MIMO communication via ant colony optimisation," *Wireless Personal Communications*, pp. 1–23, 2013.

[29] J.-K. Lain and J.-Y. Chen, "Near-MLD MIMO detection based on a modified ant colony optimization," *IEEE Communications Letters*, vol. 14, no. 8, pp. 722–724, 2010.

[30] Y. Liu, M. Tao, B. Li, and H. Shen, "Optimization framework and graph-based approach for relay-assisted bidirectional OFDMA cellular networks," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3490–3500, November 2010.

[31] M. de Paula Marques, T. Abrão, M. H. Adaniya, L. H. D. Sampaio, and P. J. E. Jeszensky, "Ant colony optimization for resource allocation and anomaly detection in communication networks," in *Search Algorithms*, T. Abrão, Ed. InTech Open, December 2012, ch. 8, pp. 1–34.

[32] L.-L. Yang, "A zero-forcing multiuser transmitter preprocessing scheme for downlink communications," *Communications, IEEE Transactions on*, vol. 56, no. 6, pp. 862–865, June 2008.

[33] T. K. Y. Lo, "Maximum ratio transmission," *Communications, IEEE Transactions on*, vol. 47, no. 10, pp. 1458–1461, Oct 1999.

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

[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[36] S. Chen, A. Livingstone, H. Q. Du, and L. Hanzo, "Adaptive minimum symbol error rate beamforming assisted detection for quadrature amplitude modulation," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 4, pp. 1140–1145, April 2008.

[37] The Mathworks, "Optimization toolbox for use with Matlab, users guide, v.2. 2002." [Online]. Available: http://www.mathworks.com/help/releases/R13sp2/pdf_doc/optim/optim_tb.pdf

[38] D. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. New York: Springer, 2008.

[39] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Maryland, USA: Johns Hopkins University Press, 1996.

A.3 Análise da BER e Alocação de Pilotos em Sistemas *Massive* MIMO

Título: *BER Analysis and Pilot Distribution Optimization of Multi-Cellular Very Large MIMO Systems*;

Autores: José Carlos Marinello & Taufik Abrão;

Submissão: Outubro de 2014;

Revista: *IEEE Systems Journal*, edição especial sobre *"5G Wireless Systems with Massive MIMO"*.

BER Analysis and Pilot Distribution Optimization of Multi-Cellular Very Large MIMO Systems

José Carlos Marinello, Taufik Abrão

***Abstract*—In this work, salient characteristics of a wireless communication system deploying a great number of antennas in the base station (BS), namely a Massive MIMO system, are investigated. The asymptotic performance of the linar zero forcing precoding scheme is found, both in terms of signal to interference plus noise ratio (SINR) and bit error rate (BER), and shown to be equivalent to the matched filter beamforming performance. Furthermore, analysis of the Massive MIMO system downlink is carried out from the viewpoint of BER performance, including some realistic adverse effects, such as interference from neighboring cells, channel estimation errors due to background thermal noise, and pilot contamination, which was recently shown to be the only impairment that remains in the MIMO multicell system with infinite number of BS antennas. For this scenario, we derive expressions for the BER of such systems in the limit of infinite number of antennas in BS. Then, a quite simple and efficient method for optimizing the Massive MIMO system performance, under different optimization metrics, is proposed, which consists of simply distributing the pilot sequences among the users of the cell in an efficient manner.**

***Index Terms*—Massive MIMO systems, BER, Multiuser MIMO, Precoding, Pilot sequences.**

I. Introduction

MULTIPLE-input-multiple-output (MIMO) techniques constitutes one of the key features in most of recent telecommunications standards, such as WiFi, WiMAX, LTE [1], [2], due to the large gains in spectral/energy efficiency they can offer. In particular, multiuser MIMO (MU-MIMO) systems have attracted substantial interest since they can achieve spatial multiplexing gains even when serving single antenna mobile terminals (MT's) [3]. Advantages of MU-MIMO also includes a larger robustness to most of propagation issues present in single-user MIMO, such as antenna correlation or channel rank loss. Even when the channel state information (CSI) of some users are highly correlated, multiuser diversity can be extracted by efficient techniques of scheduling the time/frequency resources, yielding to a better exploitation of the additional degrees of freedom (DoF) propitiated by the antenna array at the base station (BS). From an information theoretic point of view, gains of these systems have been demonstrated in [4].

In order to fully exploit the advantages of MIMO systems, a certain number of recent technical works have been concerned with the possibility of increasing the number of BS antennas N to infinity. While early papers have focused on the asymptotic limits of such systems for pure academic interest, practical issues of implementation are each time more present on the latest papers, maturing the technology and turning it gradually ready to be considered in next telecommunications standards, such as 5G [5]. It was shown in [6] that in a time division duplex (TDD) noncooperative multi-cell MIMO system, that employs uplink training pilots for CSI acquisition, with infinite number of BS antennas, the effects of uncorrelated thermal noise and fast fading are averaged out. The only factor that remains limiting performance in the Large-MIMO scenario is inter-cell interference, that when associated with the finite time available to send pilot sequences makes the estimated CSI at one BS "contaminated" by the CSI of users in adjacent cells, in the so-called pilot contamination effect. This phenomenon results from unavoidable reuse of reverse-link pilot sequences by terminals in different cells. As a consequence of increasing the number of BS antennas to infinity, the transmit power can be designed arbitrarily small, since interference decreases in the same rate of the desired signal power, i.e., signal-to-interference-plus-noise ratio (SINR) is independent of transmit power [6].

Alternative strategies to achieve better CSI estimates exist, such as **a)** frequency division duplex (FDD) [7], in which pilots for CSI acquisition are transmitted in downlink, and estimates are fed back to BS in a feedback channel; and **b)** network MIMO [8], where CSI and information data of different coordinated cells are shared among them in a backhaul link, creating a distributed antenna array that serves the users altogether. However, both schemes becomes unfeasible when $N \to \infty$ [6], since lengths of forward pilot sequences and capacity of backhaul links increase substantially with N, respectively.

Operating with a large excess of BS antennas compared with the number of terminals K is a challenging but desirable condition, since some results from random matrix theory become noticeable [9], [10]. It is known, for instance, that very tall/wide matrices tend to be very well conditioned, since their singular values distribution appears to be deterministic, showing a stable behavior (low variances) and a relatively narrow spread [11]. This phenomenon is quite appreciable to enhance the achievable rates of such systems. Besides, the most simple uplink/downlink techniques, i.e., maximum ratio combining (MRC) and matched filtering (MF) precoding, respectively, becomes optimal [11]. The savings in energy consumption are also remarkable. In the uplink of a single-cell system, it is shown in [12] that the power radiated by the terminals can be made inversely proportional to N, with perfect CSI, or to $\sqrt{N}$, for imperfect CSI, with no reduction in performance, resulting in very energy-efficient communication systems.

An interesting investigation about precoding techniques of single-cell MU-MIMO systems downlink is carried out in [13].

J. C. Marinello is a Master of Science candidate at the Electrical Engineering Department, State University of Londrina,. E-mail: zecarlos.ee@gmail.com

T. Abrão is an Associate Professor at the Electrical Engineering Department, State University of Londrina, PR, Brazil. E-mail: taufik@uel.br

Specifically, authors compared MF precoding, also known as conjugate beamforming, and zero forcing (ZF) beamforming, with respect to net spectral-efficiency and radiated energy-efficiency in a simplified single-cell scenario. It is found that, for high spectral-efficiency and low energy-efficiency, ZF outperforms MF, while at low spectral-efficiency and high energy-efficiency the opposite holds. A similar result is found for the uplink in [12], where for low signal-to-noise ratio (SNR), the simple MRC receiver outperforms the ZF receiver. It can be explained since, for reduced power levels, the cross-talk interference introduced by the inferior maximum-ratio combining receiver eventually falls below the noise enhancement induced by ZF and this simple receiver becomes a better alternative. The analog occurs for downlink. On the other hand, when considering the multi-cell environment, it is found in [11] that the asymptotic SINR of MF outperforms ZF, although MF requires much more antennas to approach the asymptotic condition.

A more rigorous expression for the achievable SINR of MF precoding in Massive MIMO systems, with respect to that of [6], is derived in [14]. Authors showed that, for downlink, the effect of the transmit power constraint at BS still accounts in the massive MIMO regime, as opposed to what was assumed in [6]. Besides, authors discuss an efficient technique for temporally distribute the uplink transmissions of pilot sequences, avoiding simultaneous transmissions from adjacent cells and reducing interference as well, in conjunction with power allocation strategy. On the other hand, a precoding technique that eliminates pilot contamination and leads to unlimited gains with $N \to \infty$ is proposed in [15]. However, these gains come at the expense of sharing the information data between base stations, what can overload the backhaul signaling channel for high rate systems, or high number of users per cell.

In this paper, we derive the expression for the downlink SINR of ZF precoding, considering the effect of power constraint at BS, in the same way of [14]. We show that it corresponds to the same value achieved by MF, as opposed to what is found in [11] neglecting the effects of the power constraint. Then, we investigate the bit error rate of the massive MIMO system downlink, and an exact expression for the BER of each user is derived, depending on the transmit power of users and on the large scale fading coefficients. Based on this derived expression, and on the asymptotic SINR expression of [14], we propose a very simplified method of optimizing the massive MIMO downlink, under different metrics. This method consists of simply assigning the available training sequences among the users within a cell in an efficient manner, by knowing only the power and the long-term fading coefficients of users in adjacent cells that reuse the sequences. Depending on the optimization metric adopted, the smart allocation of pilot sequences can lead to appreciable performance gains, both in terms of rate and BER.

This work is organized as follows. Besides this introductory Section, the system model is described in Section II. Some asymptotic limits of the massive MIMO system are revisited and extended in Section III. Our proposed methods of assigning the pilots among the users within the cell in an efficient manner, namely the Pilot Allocation (PA) schemes, are presented in Section IV. Some numerical results are shown in Section V, and, finally, Section VI concludes the paper.

II. System Model

In the adopted MIMO system, it is assumed that each one of L base stations equipped with N transmit antennas communicate with K users equipped with a single-antenna MT. Note that the L base stations share the same spectrum and the same set of K pilot signals. We denote the $1 \times N$ channel vector between the ℓth BS and the kth user of jth cell by $\mathbf{g}_{\ell kj} = \sqrt{\beta_{\ell kj}}\mathbf{h}_{\ell kj}$, in which $\beta_{\ell kj}$ is the long-term fading power coefficient, that comprises path loss and log-normal shadowing, and $\mathbf{h}_{\ell kj}$ is the short-term fading channel vector, that follows $\mathbf{h}_{\ell kj} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. Flat fading environment was assumed, where the channel matrix $\mathbf{H}$ is admitted constant over the entire frame and changes independently from frame to frame (block fading channel assumption). Note that $\beta_{\ell kj}$ is assumed constant for all N BS antennas. Since time division duplex is assumed, reciprocity holds, and thus channel state information is acquired by means of uplink training sequences. During a channel coherence interval, the symbol periods are divided to uplink pilot transmissions, processing, downlink and uplink data transmissions [14]. Using orthogonal pilot sequences, the number of sequences available is equal to its length, and thus, due to mobility of the users, which reduces the coherence time of the channel, the number of terminals served by each BS is limited. The same set of K orthogonal pilots of length K is used in all cells; hence, for the kth user of each cell is assigned the sequence $\boldsymbol{\psi}_k = [\psi_{1,k}\psi_{2,k}\ldots\psi_{K,k}]$, such that $|\psi_{i,k}| = 1$ and $|\boldsymbol{\psi}_{k'}^H\boldsymbol{\psi}_k| = K\delta_{kk'}$ since the sequences are orthogonal, where $\{\cdot\}^H$ is the conjugate transpose operator, and $\delta_{kk'} = 1$ if $k = k'$ and 0 otherwise.

In the training phase, assuming synchronization in the uplink pilot transmissions, that is the worst case[1] [6], we have that the $N \times K$ received signal at the ℓth BS is:

$$\mathbf{Y}_\ell = \sum_{j=1}^{L} \mathbf{G}_{\ell j}^T \sqrt{\boldsymbol{\rho}_j}\boldsymbol{\Psi} + \mathbf{N}, \tag{1}$$

where $\boldsymbol{\rho}_j = \mathrm{diag}(\rho_{1j}\,\rho_{2j}\,\ldots\,\rho_{Kj})$, being ρ_{kj} the uplink transmit power of the kth user of jth cell, $\{\cdot\}^T$ is the transpose operator, $\mathbf{G}_{\ell j} = [\mathbf{g}_{\ell 1j}^T\,\mathbf{g}_{\ell 1j}^T\,\ldots\,\mathbf{g}_{\ell Kj}^T]^T$, such that $\mathbf{G}_{\ell j} = \sqrt{\boldsymbol{\beta}_{\ell j}}\mathbf{H}_{\ell j}$, $\boldsymbol{\beta}_{\ell j} = \mathrm{diag}(\beta_{\ell 1j}\,\beta_{\ell 2j}\,\ldots\,\beta_{\ell Kj})$, $\mathbf{H}_{\ell j} = [\mathbf{h}_{\ell 1j}^T\,\mathbf{h}_{\ell 1j}^T\,\ldots\,\mathbf{h}_{\ell Kj}^T]^T$, $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1\,\boldsymbol{\psi}_2\,\ldots\,\boldsymbol{\psi}_K]$, and $\mathbf{N}$ is a $N \times K$ additive white gaussian noise (AWGN) matrix with zero mean and unitary variance.

In order to generate the estimated CSI matrix $\widehat{\mathbf{G}}_\ell$ of their served users, the ℓth BS correlates its received signal matrix

[1]Unavoidably, the same orthogonal pilot sequences are re-used among the cells; hence, in the course of estimating the channel states to its own MTs, a BS inadvertently learns the channel to MTs in other cells who share the same pilot sequence, or whose pilot sequences are merely correlated with the pilot sequences of its own terminals [6]. In the first case, i.e., same pilot sequence, the worst correlation effect occurs when there is perfect synchronism among pilots.

with the known pilot sequences:

$$\widehat{\mathbf{G}}_\ell^T = \frac{1}{K}\mathbf{Y}_\ell \mathbf{\Psi}^H = \sum_{j=1}^{L} \mathbf{G}_{\ell j}^T \sqrt{\boldsymbol{\rho}_j} + \mathbf{N}'^T, \tag{2}$$

where $\mathbf{N}'$ is an equivalent AWGN matrix with zero mean and variance $\frac{1}{K}$. Note that the channel estimated by the ℓth BS is contaminated by the channel of users that use the same pilot sequence in all other cells.

Besides, information transmit symbol vector of the ℓth cell is denoted by $\mathbf{x}_\ell = [x_{1\ell}\, x_{2\ell}\, \ldots\, x_{K\ell}]^T$, where $x_{k\ell}$ is the transmit symbol to the kth user of the ℓth cell, and takes a value from the squared quadrature amplitude modulation (M-QAM) alphabet; so, the complex-valued symbol (finite) set is given by $\mathcal{S} = \left\{\mathcal{A} + \sqrt{-1}\cdot\mathcal{A}\right\}$, where the real-valued finite set $\mathcal{A} = \left\{\pm\frac{1}{2}a;\, \pm\frac{3}{2}a;\, \ldots;\, \pm\frac{\sqrt{M}-1}{2}a\right\}$, with $\sqrt{M}$ representing the modulation order (per dimension) of the corresponding real-valued ASK scheme. The parameter $a = \sqrt{6/(M-1)}$ is used in order to normalize the power of the complex-valued transmit signals to 1. For analysis simplicity, using matrix notation, the $K \times 1$ complex-valued received signal by the users of the ℓth cell is written as:

$$\mathbf{r}_\ell = \sum_{j=1}^{L} \mathbf{G}_{j\ell}\mathbf{P}_j \sqrt{\boldsymbol{\zeta}_j}\mathbf{x}_j + \mathbf{n}_\ell, \tag{3}$$

where $\boldsymbol{\zeta}_j = \mathrm{diag}(\zeta_{1j}\, \zeta_{2j}\, \ldots\, \zeta_{Kj})$, being ζ_{kj} the downlink transmit power devoted by the jth BS to its kth user, $\mathbf{P}_j$ denotes the complex valued $N \times K$ precoding matrix of the jth BS, being each column $\mathbf{p}_{jk}$ the $N \times 1$ precoding vector of the kth user. Finally, $\mathbf{n}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ represents the AWGN vector with variance $\sigma_n^2 = \frac{1}{2}$ per dimension, which is observed at the K mobile terminals of the ℓth cell.

Under the matched filter beamforming technique, the vector $\mathbf{p}_{jk}$ is computed as [14]:

$$\mathbf{p}_{jk}^{\text{MF}} = \frac{\widehat{\mathbf{g}}_{jk}}{||\widehat{\mathbf{g}}_{jk}||} = \frac{\widehat{\mathbf{g}}_{jk}}{\alpha_{kj}\sqrt{N}}, \tag{4}$$

in which $\alpha_{kj} = \frac{||\widehat{\mathbf{g}}_{jk}||}{\sqrt{N}}$, and $\widehat{\mathbf{g}}_{jk}$ is the kth row of the matrix $\widehat{\mathbf{G}}_j$. Note that the normalization in (4) is necessary to satisfy the maximum transmit power available at the BS.

In the same way, in the zero forcing beamforming technique, the vector $\mathbf{p}_{jk}$ is computed as:

$$\mathbf{p}_{jk}^{\text{ZF}} = \frac{\mathbf{w}_{jk}}{||\mathbf{w}_{jk}||}, \tag{5}$$

in which the vector $\mathbf{w}_{jk} = \widehat{\mathbf{G}}_j^H \mathbf{a}_{jk}$, and $\mathbf{a}_{jk}$ is the kth column of $\mathbf{A}_j = \left[\widehat{\mathbf{G}}_j \widehat{\mathbf{G}}_j^H\right]^{-1}$.

III. Asymptotic Limits of Massive MIMO

Most of the asymptotic limits of Massive MIMO systems are built upon the following well known lemma:

Lemma 1. *Let $\mathbf{s}_1$,$\mathbf{s}_2 \in \mathbb{C}^{N\times 1}$ be two independent complex-valued vectors following a normal distribution, with zero mean and variance σ^2. Then*

$$\lim_{N\to\infty} \frac{\mathbf{s}_1^H \mathbf{s}_2}{N} \overset{a.s.}{=} 0 \quad \textit{and} \quad \lim_{N\to\infty} \frac{\mathbf{s}_1^H \mathbf{s}_1}{N} \overset{a.s.}{=} \sigma^2. \tag{6}$$

Since the channel vectors of different users can be seen as independent random vectors, the above lemma is widely used for deriving limits in the massive MIMO scenarios. It can be justified since as the vector's length grows, the inner products between independent vectors grow at lesser rates than the inner products of vectors with themselves.

A. Asymptotic Limits of MF Beamforming

From (2), it is proved in [14] that $\alpha_{kj}^2 \overset{a.s.}{=} \sum_{l=1}^{L} \rho_{kl}\beta_{jkl} + \frac{1}{K}$. Then authors show that $r_{k\ell}$, i.e., the received signal by the kth user of ℓth cell, can be written as [14, Eq. (5)]:

$$r_{k\ell} = \sum_{l=1}^{L}\sum_{j=1}^{K} \sqrt{\zeta_{jl}\beta_{lk\ell}}\mathbf{h}_{lk\ell}^H \mathbf{p}_{jl}^{\text{MF}} x_{jl} + n_{k\ell}. \tag{7}$$

Based on (4) and Lemma 1, (7) can be simplified when $N \to \infty$ as:

$$\begin{aligned} r_{k\ell} &= \sum_{l=1}^{L} \sqrt{\zeta_{kl}\beta_{lk\ell}}\mathbf{h}_{lk\ell}^H \mathbf{p}_{kl}^{\text{MF}} x_{kl} + n_{k\ell}, \\ &= \sum_{l=1}^{L} \frac{1}{\alpha_{kl}} \sqrt{N\zeta_{kl}\rho_{k\ell}}\beta_{lk\ell} x_{kl} + n_{k\ell}, \\ &= \sqrt{N\rho_{k\ell}} \sum_{l=1}^{L} \frac{\sqrt{\zeta_{kl}}\beta_{lk\ell} x_{kl}}{\alpha_{kl}} + n_{k\ell}. \end{aligned} \tag{8}$$

From (8), it is straightforward to see the asymptotic downlink SINR of the system as:

$$\begin{aligned} \text{SINR}_{k\ell}^{\text{DL}} &= \lim_{N\to\infty} \frac{N\rho_{k\ell}\, \zeta_{k\ell}\beta_{\ell k\ell}^2/\alpha_{k\ell}^2}{N\rho_{k\ell}\left(\sum_{\substack{j=1 \\ j\neq\ell}}^{L} \zeta_{kj}\beta_{jk\ell}^2/\alpha_{kj}^2\right) + 1}, \\ &= \frac{\zeta_{k\ell}\beta_{\ell k\ell}^2/\alpha_{k\ell}^2}{\sum_{\substack{j=1 \\ j\neq\ell}}^{L} \zeta_{kj}\beta_{jk\ell}^2/\alpha_{kj}^2}. \end{aligned} \tag{9}$$

Note that this limit depends mainly on the large scale fading coefficients β_{jki}, which are related to the spatial distribution of the users on the different cells.

B. Asymptotic Limits of ZF Beamforming

For the ZF beamforming, Eq. (7) becomes

$$r_{k\ell} = \sum_{l=1}^{L}\sum_{j=1}^{K} \sqrt{\zeta_{jl}\beta_{lk\ell}}\mathbf{h}_{lk\ell}^H \mathbf{p}_{jl}^{\text{ZF}} x_{jl} + n_{k\ell}. \tag{10}$$

In order to find the asymptotic limits when employing the ZF scheme, we begin analysing the matrix $\mathbf{A}_\ell = \left[\widehat{\mathbf{G}}_\ell \widehat{\mathbf{G}}_\ell^H\right]^{-1}$. From (2), we have that

$$\mathbf{A}_\ell = \left[\left(\sum_{j=1}^{L} \sqrt{\boldsymbol{\rho}_j\boldsymbol{\beta}_{\ell j}}\mathbf{H}_{\ell j} + \mathbf{N}'\right)\left(\sum_{l=1}^{L} \mathbf{H}_{\ell l}^H \sqrt{\boldsymbol{\rho}_l\boldsymbol{\beta}_{\ell l}} + \mathbf{N}'^H\right)\right]^{-1}. \tag{11}$$

Note that from Lemma 1, we can neglect all the terms corresponding to products of independent matrices, since in the limit of $N \to \infty$, they will not account. So, (11) simplifies to:

$$\mathbf{A}_\ell = \left[\sum_{j=1}^{L} \sqrt{\boldsymbol{\rho}_j\boldsymbol{\beta}_{\ell j}}\mathbf{H}_{\ell j} \sum_{l=1}^{L} \mathbf{H}_{\ell l}^H \sqrt{\boldsymbol{\rho}_l\boldsymbol{\beta}_{\ell l}} + \mathbf{N}'\mathbf{N}'^H\right]^{-1}$$

$$= \left[\sum_{j=1}^{L}\sqrt{\boldsymbol{\rho}_j\boldsymbol{\beta}_{\ell j}}\mathbf{H}_{\ell j}\mathbf{H}_{\ell j}^{H}\sqrt{\boldsymbol{\rho}_j\boldsymbol{\beta}_{\ell j}}+\mathbf{N}'\mathbf{N}'^{H}\right]^{-1}$$
$$= \left[N\left(\sum_{j=1}^{L}\boldsymbol{\rho}_j\boldsymbol{\beta}_{\ell j}+\frac{1}{K}\mathbf{I}_K\right)\right]^{-1}$$
$$= [\mathbf{D}_\ell]^{-1}. \quad (12)$$

It can be seen that the matrix $\mathbf{D}_\ell$ is a diagonal matrix, in which the kth term can be written as $[\mathbf{D}_\ell]_{kk} = N\left(\sum_{j=1}^{L}\rho_{kj}\beta_{\ell kj}+\frac{1}{K}\right)$. Thus, the matrix $\mathbf{A}_\ell$ will also be a diagonal matrix, and the kth term can be written as $[\mathbf{A}_\ell]_{kk} = 1/[\mathbf{D}_\ell]_{kk}$. Hence, the vector $\mathbf{w}_{jk} = \widehat{\mathbf{G}}_j^H\mathbf{a}_{jk}$ is simplified to $\mathbf{w}_{jk} = \widehat{\mathbf{g}}_{jk}^H\frac{1}{N\left(\sum_{j=1}^{L}\rho_{kj}\beta_{\ell kj}+\frac{1}{K}\right)}$, and Eq. (5) can be rewritten as

$$\mathbf{p}_{jk}^{\text{ZF}} = \frac{\widehat{\mathbf{g}}_{jk}^H\frac{1}{N\left(\sum_{j=1}^{L}\rho_{kj}\beta_{\ell kj}+\frac{1}{K}\right)}}{\left\|\widehat{\mathbf{g}}_{jk}^H\frac{1}{N\left(\sum_{j=1}^{L}\rho_{kj}\beta_{\ell kj}+\frac{1}{K}\right)}\right\|} = \frac{\widehat{\mathbf{g}}_{jk}^H}{\left\|\widehat{\mathbf{g}}_{jk}^H\right\|} = \mathbf{p}_{jk}^{\text{MF}}. \quad (13)$$

It is important to note that the convergence of the MF precoding to the ZF scheme in the limit of $N \to \infty$ is just achieved when considering the constraint of maximum transmit power available at BS, i.e., normalizing the precoding vector. Otherwise, the asymptotic performances of such techniques will differ, as shown in [11, Fig. 11]. Furthermore, this equality holds only for N very large. For intermediate values, it is seen that the ZF scheme approaches the asymptotic limit faster than the MF beamforming, as will be demonstrated numerically in Section V-A. However, the MF technique is quite less complex, and can be implemented in a decentralized way since the precoding vector of each user is not dependent on the estimated channels of other users, as opposed to ZF.

C. Asymptotic BER in Downlink

We demontrated that the MF and ZF schemes converge to the same precoding vector when the number of BS antennas grows to infinity. So, the results obtained in [14] for MF are also valid for ZF, specially Equations (8) and (9). Analysing the received signal of the kth user of the ℓth cell (8), we can also obtain some information about the bit error probability:

$$r_{k\ell} = \sqrt{N\rho_{k\ell}}\left(\frac{\sqrt{\zeta_{k\ell}}\beta_{\ell k\ell}x_{k\ell}}{\alpha_{k\ell}}+\sum_{\substack{l=1\\l\neq\ell}}^{L}\frac{\sqrt{\zeta_{kl}}\beta_{lk\ell}x_{kl}}{\alpha_{kl}}\right)+n_{k\ell}. \quad (14)$$

Indeed, the effect of AWGN is averaged out when $N \to \infty$. For notation simplicity, but with no loss of generality, we consider a QPSK modulation. Thus, the probability of error for this user can be written as:

$$\text{Pe}_{k\ell} = \tfrac{1}{2}\Pr\left(\Re\left\{\frac{\sqrt{\zeta_{k\ell}}\beta_{\ell k\ell}x_{k\ell}}{\alpha_{k\ell}}\right\} < \Re\left\{\sum_{\substack{l=1\\l\neq\ell}}^{L}\frac{\sqrt{\zeta_{kl}}\beta_{lk\ell}x_{kl}}{\alpha_{kl}}\right\}\right) + \tfrac{1}{2}\Pr\left(\Im\left\{\frac{\sqrt{\zeta_{k\ell}}\beta_{\ell k\ell}x_{k\ell}}{\alpha_{k\ell}}\right\} < \Im\left\{\sum_{\substack{l=1\\l\neq\ell}}^{L}\frac{\sqrt{\zeta_{kl}}\beta_{lk\ell}x_{kl}}{\alpha_{kl}}\right\}\right), \quad (15)$$

where $\Pr(\cdot)$ is the probability of an event. This expression can be simplified to

$$\text{Pe}_{k\ell} = \tfrac{1}{2}\Pr\left(\frac{\sqrt{\zeta_{k\ell}}\beta_{\ell k\ell}\Re\{x_{k\ell}\}}{\alpha_{k\ell}} < \sum_{\substack{l=1\\l\neq\ell}}^{L}\frac{\sqrt{\zeta_{kl}}\beta_{lk\ell}\Re\{x_{kl}\}}{\alpha_{kl}}\right) + \tfrac{1}{2}\Pr\left(\frac{\sqrt{\zeta_{k\ell}}\beta_{\ell k\ell}\Im\{x_{k\ell}\}}{\alpha_{k\ell}} < \sum_{\substack{l=1\\l\neq\ell}}^{L}\frac{\sqrt{\zeta_{kl}}\beta_{lk\ell}\Im\{x_{kl}\}}{\alpha_{kl}}\right),$$
$$= \Pr\left(\frac{\sqrt{\zeta_{k\ell}}\beta_{\ell k\ell}\Re\{x_{k\ell}\}}{\alpha_{k\ell}} < \sum_{\substack{l=1\\l\neq\ell}}^{L}\frac{\sqrt{\zeta_{kl}}\beta_{lk\ell}\Re\{x_{kl}\}}{\alpha_{kl}}\right), \quad (16)$$

since both terms in the sum have the same behavior. The errors will occur whenever the interfering signal that reaches the user is greater than its intended signal. To determine the exact value of the probability in (16), we must analyse every possible combination of interfering signals. Thus, the result can be written as

$$\text{Pe}_{k\ell} = \frac{1}{2^{L-1}}\sum_{j=1}^{2^{L-1}}\text{u}\left[\left(\sum_{\substack{l=1\\l\neq\ell}}^{L}\frac{\sqrt{\zeta_{kl}}\beta_{lk\ell}b_{jl}}{\alpha_{kl}}\right)-\frac{\sqrt{\zeta_{k\ell}}\beta_{\ell k\ell}}{\alpha_{k\ell}}\right], \quad (17)$$

where $\text{u}[x]$ is the Heaviside step function ($\text{u}[x] = 1$ if $x \geq 0$, $\text{u}[x] = 0$ otherwise), and b_{jl} is the j,l-th element of the $2^{L-1}\times L$ matrix $\mathbf{B} = [\mathbf{B}'_{1:\ell-1}, \mathbf{1}_{2^{L-1}}, \mathbf{B}'_{\ell:L-1}]$, in which $\mathbf{B}'$ contains every possible combination of $\{\pm 1\}^{L-1}$, and $\mathbf{1}_K$ is an unitary vector of length K.

The expression in (17) gives the exact bit error rate of the kth user of the ℓth cell, as a function of the powers and the long-term fading coefficients of users in adjacent cells sharing the same training sequence. Thus it can be adopted as a performance optimization metric, in the same way as defined in eq. (9). One possible approach is invoking a power allocation algorithm, as done in [14] with respect to (9). In this paper, we prefer a more simple strategy, that consists of simply *optimizing the assignment of pilot sequences to users*, by knowing the interference experienced by each sequence. We call this procedure of Pilot Allocation scheme.

IV. Pilot Allocation Schemes

Eq. (14) shows that the received signal for a given user in the downlink of a massive MIMO system presents interference from another users in adjacent cells that share the same pilot sequence. Besides, from this received signal in the limit of $N \to \infty$, the asymptotic expressions for SINR, eq. (9), and for BER, eq. (17), have been derived. At first glance, it may appear that the interference term in (14) does not depend on which user in ℓth cell is assigned the kth pilot sequence. However, reminding that $\alpha_{kl}^2 \overset{a.s.}{=} \sum_{j=1}^{L}\rho_{kj}\beta_{lkj}+\frac{1}{K}$, one can see it is not true.

Thus, varying to which user is assigned the kth pilot sequence according its long-term fading coefficient can enhance the SINR, eq. (9), and/or[2] decrease the probability of error, eq. (17). This fact allows us the formulation of alternative optimization criteria, as described in the sequel.

Initially, we define the matrix $\mathbf{C}$, of size $K!\times K$, containing every possible combination of pilot sequences to the users, i.e., c_{ij} says that, in the ith combination, the jth pilot sequence is allocated to the c_{ij}th user. In the first pilot allocation scheme, we search the best pilot distribution in the sense that minimizes the mean BER among users of the ℓth cell, leading to the MinBER Pilot Allocation scheme:

$$i_{\text{MB}} = \arg\min_i \frac{1}{K}\sum_{k=1}^{K}\text{Pe}_{c_{ik}\ell}, \quad (18)$$

[2]Maximizing the SINR not necessarily minimizes the BER in the limit of $N \to \infty$, as can be seen from expressions (9) and (17), and discussed in Sec. V-B.

in which

$$\mathrm{Pe}_{c_{ik}\ell} = \frac{1}{2^{L-1}} \sum_{j=1}^{2^{L-1}} \mathrm{u}\left[\left(\sum_{\substack{l=1 \\ l\neq\ell}}^{L} \frac{\sqrt{\zeta_{kl}}\beta_{lk\ell}b_{jl}}{\alpha_{kl}^{(i)}}\right) - \frac{\sqrt{\zeta_{c_{ik}\ell}}\beta_{\ell c_{ik}\ell}}{\alpha_{c_{ik}\ell}^{(i)}}\right] \tag{19}$$

corresponds to the BER of the c_{ik}th user of ℓth cell when the kth pilot sequence is assigned to him. Note that the superscript in $\alpha_{kl}^{(i)}$ and $\alpha_{c_{ik}\ell}^{(i)}$ evidences that these terms depend on the ith pilot distribution, since $(\alpha_{kl}^{(i)})^2 = \rho_{c_{ik}\ell}\beta_{lc_{ik}\ell} + \sum_{\substack{j=1 \\ j\neq\ell}}^{L} \rho_{kj}\beta_{lkj} + \frac{1}{K}$, and $(\alpha_{c_{ik}\ell}^{(i)})^2 = \rho_{c_{ik}\ell}\beta_{\ell c_{ik}\ell} + \sum_{\substack{j=1 \\ j\neq\ell}}^{L} \rho_{kj}\beta_{\ell kj} + \frac{1}{K}$.

In the same way, in the second pilot allocation scheme, we define the pilot distribution that maximizes the mean SINR in the downlink of the ℓth cell, namely MaxSINR Pilot Allocation scheme:

$$i_{\text{MS}} = \arg\max_{i} \frac{1}{K} \sum_{k=1}^{K} \text{SINR}_{c_{ik}\ell}^{\text{DL}}, \tag{20}$$

in which

$$\text{SINR}_{c_{ik}\ell}^{\text{DL}} = \frac{\zeta_{c_{ik}\ell}\beta_{\ell c_{ik}\ell}^2/(\alpha_{c_{ik}\ell}^{(i)})^2}{\sum_{\substack{j=1 \\ j\neq\ell}}^{L} \zeta_{kj}\beta_{jk\ell}^2/(\alpha_{kj}^{(i)})^2} \tag{21}$$

is the downlink SINR of the c_{ik}th user of ℓth cell when the kth pilot sequence is assigned to him.

Both schemes find the pilot distribution by optimizing the mean value of some performance criterion. However, the "average" approach may be not completely adequate in modern communications systems, since it may lead to a great improvement in performance for a few users, while providing low quality of service (QoS) to those users poorly located, typically in the edge of the cell. Hence, we also look for pilot allocation schemes that ensure improvement in QoS for every user within the ℓth cell. The MinimaxBER Pilot Allocation scheme is defined as:

$$i_{\text{MMB}} = \arg\min_{i} \max_{k} \mathrm{Pe}_{c_{ik}\ell}, \tag{22}$$

which minimizes the worst BER within the cell.

On the other hand, the MaxminSINR Pilot Allocation scheme can be defined as:

$$i_{\text{MMS}} = \arg\max_{i} \min_{k} \text{SINR}_{c_{ik}\ell}^{\text{DL}}, \tag{23}$$

which maximizes the lowest SINR among the users of the cell.

Algorithm 1 describes the Pilot Allocation procedure, defining its inputs, outputs, and main steps. After its computation, the ℓth cell assigns the kth pilot sequence to the $c_{i_{opt}k}$th user. Note that each cell finds the optimal pilot combination among its covered users, in a decentralized way, reducing the overall computational complexity.

Algorithm 1 Pilot Allocation Procedure

Input:
$\boldsymbol{\beta}_{jl}, \boldsymbol{\zeta}_j, \boldsymbol{\rho}_j, \forall\, j, l = 1, 2, \ldots L.$

1: Generate matrix $\mathbf{C}$, of size $K! \times K$;
2: **for** each combination $i = 1, 2, \ldots, K!$ **do**
3: **for** each pilot sequence $k = 1, 2, \ldots, K$ **do**
4: Evaluate $\alpha_{c_{ik}\ell}^{(i)} = \sqrt{\rho_{c_{ik}\ell}\beta_{\ell c_{ik}\ell} + \sum_{\substack{j=1 \\ j\neq\ell}}^{L} \rho_{kj}\beta_{\ell kj} + \frac{1}{K}}$;
5: **for** each cell $l = 1, 2, \ldots, L$, $l \neq \ell$ **do**
6: Evaluate $\alpha_{kl}^{(i)} = \sqrt{\rho_{c_{ik}\ell}\beta_{lc_{ik}\ell} + \sum_{\substack{j=1 \\ j\neq\ell}}^{L} \rho_{kj}\beta_{lkj} + \frac{1}{K}}$;
7: **end for**
8: **end for**
9: **end for**
10: Find $i_{opt} \in i = 1, 2, \ldots, K!$, corresponding to the optimal combination in $\mathbf{C}$ according some metric: (18), (20), (22), (23);

Output: i_{opt}.

V. Numerical Results

We adopt in our simulations a multi-cell scenario with hexagonal cells, with radius 1600m, where $K = 4$ users are uniformly distributed in its interior, except in a circle of 100m radius around the cell centered BS. Besides, only the first ring of interfering cells has been considered, both for frequency reuse factor (RF) of one and three. We assume a similar TDD protocol of that in [14], in which the coherence interval is composed of 11 symbol periods: 4 for sending uplink training sequences, 1 for processing, 4 and 2 for downlink and uplink data transmission, respectively. The system uses a carrier frequency of 1.9 GHz and a frequency band of 20 MHz. The log normal shadowing has been modelled with a standard deviation of 8dB, and the path loss decay exponent equal to 3.8. Furthermore, we have considered 4-QAM modulation and SNR's of 10 dB, both in downlink and uplink. It is important to note that the numerical results in this section were obtained via Monte-Carlo simulation method.

Figure 1 depicts a single realization of the multi-cell scenario adopted in our numerical simulations, for frequency reuse factors of one and three. Notice that for clarity purpose, only users sharing the same frequency band, i.e., interfering with each other, have been represented. Indeed, one can see that interfering users are much closer with smaller reuse factors. In our numerical results presented in the sequel, only the performance metrics of users positioned inside the central cell were computed, since these users experience a more realistic condition of interference.

A. Performance Convergence of Precoding Techniques

Considering both MF and ZF precoding techniques, Fig. 2 depicts the asymptotic convergence (as the number of BS antennas increases) for both BER and SINR performance metrics to the bounds defined in (17) and (9), respectively. We have considered frequency reuse factors of one and three. The curves present mean values of each performance metric, taken among the users of the cell. One can note that the SINR of the ZF precoding scheme indeed converges to the same bound of eq. (9), which was derived in [14] as the asymptotic SINR of MF beamforming. This occurs since we have considered the constraint of maximum transmit power available at BS, as opposed to [11]. These numerical results also show that MF needs at least one order of magnitude more BS antennas than ZF to reach that bound. Furthermore, the performance of both schemes are also analysed from the perspective of

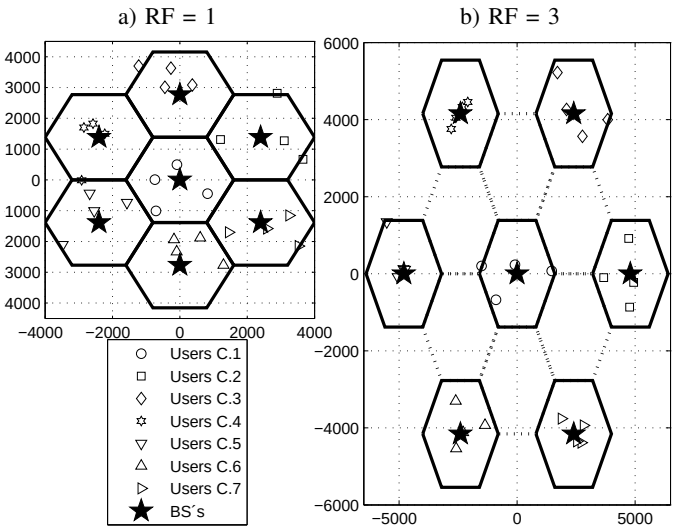

Fig. 1. Single spatial realization for both investigated multi-cell scenarios, with $K = 4$ mobile terminals.

BER, validating eq. (17) as the asymptotic BER that such techniques are able to achieve when $N \to \infty$. Indeed, in terms of BER, the performances of both techniques rapidly approach the asymptotic limit, being necessary $\approx 10^4$ BS antennas for both precoding techniques reach the bound.

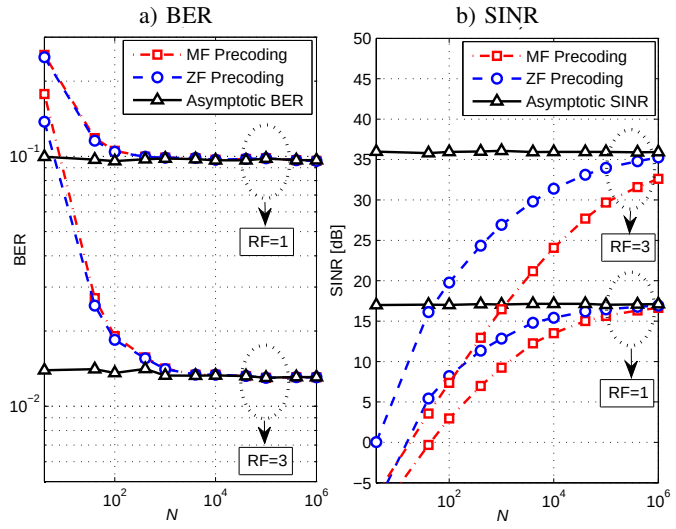

Fig. 2. Asymptotic convergences of MF and ZF precoding techniques to the performance bounds, under reuse factors of one and three, with increasing N.

B. Performance of Pilot Allocation Schemes

In this subsection, we investigate the performance of the four Pilot Allocation schemes proposed in Sec. IV, in terms of mean values, as well as in terms of distribution among users. The simulation results presented here were averaged over 100,000 spatial realizations.

Figure 3 shows the cumulative distribution function (CDF) as a function of the BER of the users, regarding different Pilot Allocation techniques. For reference of comparison, it is also depicted the very large MIMO performance with no optimization in the distribution of pilot sequences, i.e., with random allocation strategy. An interesting behavior on the BER distribution among users in the massive MIMO system can be observed from these numerical results. One can note that a significant portion of users communicates to BS with no errors, i.e., BER = 0. This occurs because, for these users, even the strongest interference that can reach them is lower than their intended signal, and thus the probability of error is null. On the other hand, the other small portion of users, that are not free of errors, presents excessive values of BER. This disparity becomes more noticeable for unitary frequency reuse factor, in which the portion of users that presents excessive error rates is $\approx 10\%$, while for reuse factor of three it is about 1% for a BER ≥ 0.2.

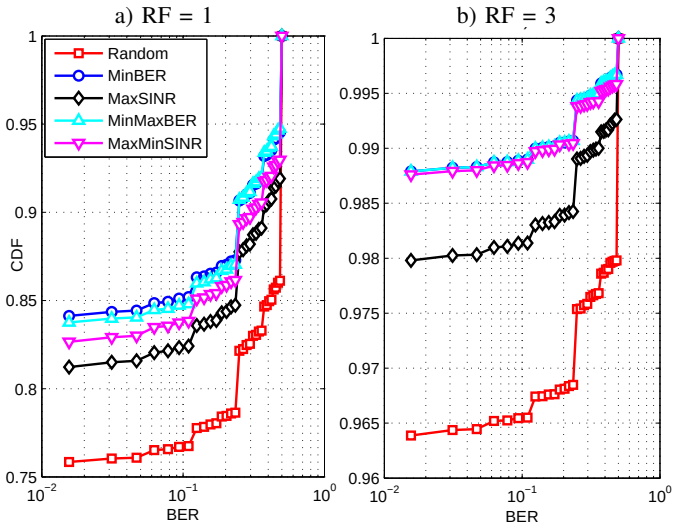

Fig. 3. Cumulative distribution function for the BER of the users, for different pilot allocation schemes.

Furthermore, it shows that the Pillot Allocation schemes are able to significantly decrease the fraction of users with excessive BER's. As shown in Tables I and II, the fraction of users with BER ≥ 0.1 reduces from 23.63% to 14.92%, for frequency reuse factor of one, and from 3.47% to 1.06%, for frequency reuse factor of three, when deploying the MinBER approach.

TABLE I
PERFORMANCE OF PILOT ALLOCATION SCHEMES FOR FREQUENCY-REUSE FACTOR OF ONE.

PA Scheme	Mean BER (%)	Users BER=0 (%)	Users BER≥0.1 (%)	Mean user Rate (Mbps)	95%-likely user Rate (Mbps)
Random	9.84	75.41	23.63	48.50	0.1344
MinBER	5.48	83.98	14.92	55.54	0.4471
MaxSINR	6.85	80.85	17.78	56.59	0.4611
MinimaxBER	5.52	83.68	15.21	55.48	0.4609
MaxminSINR	6.17	82.45	16.28	52.62	0.7937

Figure 4 shows the fraction of users above a given SINR, for frequency reuse factors of one and three. It can be seen that increasing the frequency reuse factor has the effect of significantly improving the SINR of the users, as if the curve was shifted right $\approx$ 18dB, without noticeable changes on its

TABLE II
PERFORMANCE OF PILOT ALLOCATION SCHEMES FOR FREQUENCY-REUSE FACTOR OF THREE.

PA Scheme	Mean BER (%)	Users BER=0 (%)	Users BER≥0.1 (%)	Mean user Rate (Mbps)	95%-likely user Rate (Mbps)
Random	1.41	96.33	3.47	29.07	4.79
MinBER	0.36	98.83	1.06	32.84	8.82
MaxSINR	0.66	98.01	1.86	32.91	8.81
MinimaxBER	0.37	98.82	1.06	32.84	8.82
MaxminSINR	0.39	98.78	1.09	31.68	11.15

format and slope. One can see that the MaxSINR technique has the ability of improving the SINR of the best located users, increasing $\approx$ 4dB of SINR for the best 20% of the users. On the other hand, the MaxminSINR scheme increases $\approx$ 10dB of SINR for the 95% level, i.e., it benefits the less favorably located users.

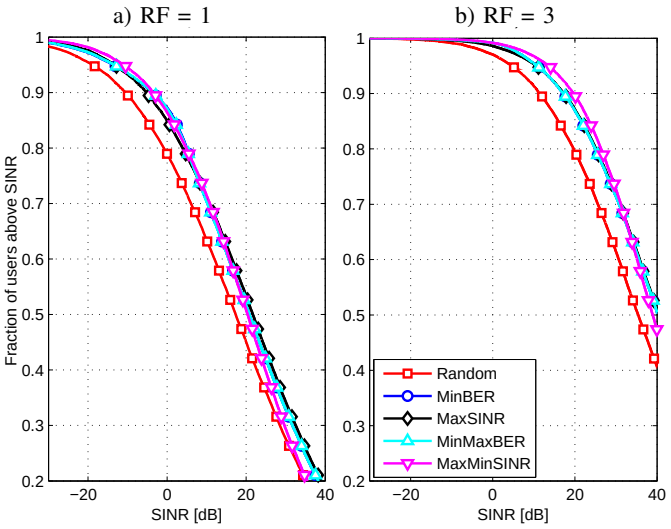

Fig. 4. Fraction of users above a given SINR, for different pilot allocation schemes.

Finally, Figure 5 depicts the fraction of users above a given data rate, for frequency reuse factors of one and three, regarding the different proposed Pilot Allocation schemes. Notice that the downlink data rate $\mathcal{R}_{k\ell}$ of the kth user in the ℓth cell can be defined as:

$$\mathcal{R}_{k\ell} = \left(\frac{\mathrm{BW}}{\mathrm{RF}}\right)\left(\frac{\mathcal{D}}{\mathcal{T}}\right)\log_2\left(1+\mathrm{SINR}_{k\ell}^{\mathrm{DL}}\right), \qquad (24)$$

where BW is the system total bandwidth, $\mathcal{D}$ is the number of symbol periods spent sending downlink data, and $\mathcal{T}$ is the total number of symbol periods within a channel coherence time.

Examining the curves, one can conclude that the slope of curves for reuse factor three is greater than the slope of unitary reuse factor curves. This fact means that the distribution for unitary reuse factor is much more irregular, unequal, in the sense that some users have very high rates while others have low QoS. On the other hand, for reuse factor of three this distribution is much more uniform, guaranteeing simultaneously an improved QoS for much more users.

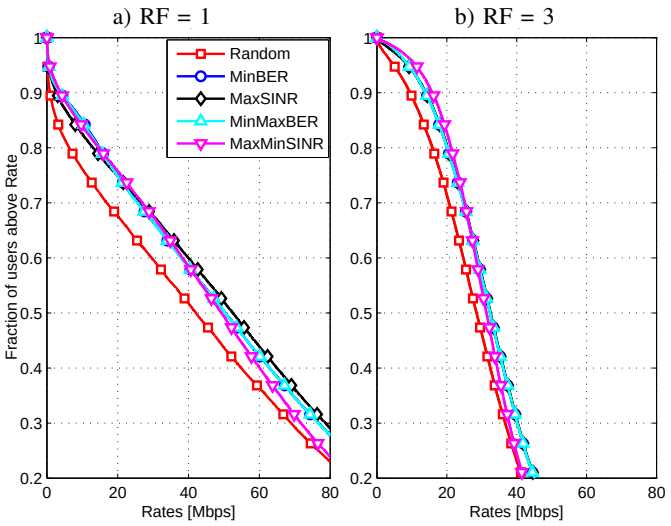

Fig. 5. Fraction of users above a given Rate, for different pilot allocation schemes.

As shown in Table I, 95% of users communicates with rates greater than 0.1344Mbps with random pilot distribution, while when employing the MaxminSINR PA scheme the 95%-likely rate per user increases to 0.7937Mbps. This means that a gain of 6 times can be achieved for unitary frequency reuse factor, while providing a mean rate of 52.62Mbps per user. Furthermore, if the minimum assured performance per terminal is a more important concern, then the MaxminSINR scheme can be employed in conjuntion with a frequency reuse factor of three. As described in Table II, the 95%-likely rate passes from 4.79 Mbps to 11.15 Mbps, a gain greater than 6 Mbps. Note that the mean rate, however, decreases, since the gain in SINR for the best located users does not offset the loss due to reduction in bandwidth, given the logarithmic increase of rate according SINR gains. Larger reuse factors are more beneficial for poor located users, since the logarithm is in its linear region, as discussed in [6].

Comparing both Tables, we note that the assured QoS, in terms of 95%-likely rate, can pass from 0.1344 Mbps, with unitary reuse factor and no optimization in pilot allocation strategy, to 11.15 Mbps, for the MaxminSINR scheme and reuse factor of three. Thus, gains of $\approx$ 85 times can be achieved combining both RF and PA techniques, with appropriate conditions. Besides, the mean BER reduces from 9.84% to 0.39%, and the portion of users communicating in the absense of errors increases from 75.41% to 98.78%. These benefits are achieved by simply assigning the pilot sequences to the users within the cell in a more efficient way, in conjunction with a large frequency reuse factor, and remain valid whenever the long-term fading coefficients stay unchanged.

VI. CONCLUSION

In this work we have characterized the asymptotic performance of the Massive MIMO system downlink under the point of view of the bit-error-rate performance. We showed that both MF and ZF precoding schemes result in the same signals as

$N \to \infty$, and thus the results of [14] are also valid for the ZF beamforming. By numerical simulations, it was shown that the ZF precoding approaches the asymptotic bounds with fewer antennas, about one order of magnitude lesser, although the decentralized implementation of MF can be more preferable in some applications. Then, we derived the exact asymptotic expression of the BER of a given user, based on the long-term fading coefficients and the power levels of other users. In the same way as the asymptotic SINR expression found in [14], the BER expression derived also depends only on the users in neighboring cells that reuse the same pilot sequence.

Furthermore, we have proposed efficient forms of assigning these pilots to the users within the cell, by optimizing several performance metrics, including the asymptotic BER found herein. The significant gains achieved by these pilot allocation techniques were demonstrated numerically. For instance, we have showed that a gain of 6 times regarding the random strategy was achieved for the downlink rate with unitary reuse factor, while it increases from 4.79 Mbps to 11.15 Mbps for reuse factor of three. When combining the MaxminSINR technique with an appropriated reuse factor, we showed that the Massive MIMO system are able to operate with a 95%-likely downlink rate of 11.15 Mbps, providing a communication free of errors for 98.78% of the users.

All of these benefits are achieved in a quite simple way, by only knowing the powers and the long-term fading coefficients of users in adjacent cells, for each pilot sequence. Since these informations do not scale with the number of BS antennas, and remains constant within a long time and frequency interval, the implementation of the proposed method in real systems is surely feasible. Besides, even greater gains might be achieved by combining the proposed schemes with power allocation and time-shifting techniques [14], [16], which is the continuity of this work.

REFERENCES

[1] L. Hanzo, Y. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus Non-coherent and Cooperative Turbo Transceivers*. IEEE Press and John Wiley & Sons, November 2010.

[2] Q. Li, G. Li, W. Lee, M. il Lee, D. Mazzarese, B. Clerckx, and Z. Li, "MIMO techniques in WiMAX and LTE: a feature overview," *Communications Magazine, IEEE*, vol. 48, no. 5, pp. 86–92, May 2010.

[3] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *Signal Processing Magazine, IEEE*, vol. 24, no. 5, pp. 36–46, Sept 2007.

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

[5] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 74–80, February 2014.

[6] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[7] J. Choi, Z. Chance, D. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *Communications, IEEE Transactions on*, vol. 61, no. 12, pp. 5016–5029, December 2013.

[8] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *Wireless Communications, IEEE*, vol. 13, no. 4, pp. 56–61, Aug 2006.

[9] R. Couillet and M. Debbah, "Signal processing in large systems: A new paradigm," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 24–39, Jan 2013.

[10] M. Matthaiou, M. McKay, P. Smith, and J. Nossek, "On the condition number distribution of complex Wishart matrices," *Communications, IEEE Transactions on*, vol. 58, no. 6, pp. 1705–1717, June 2010.

[11] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[12] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *Communications, IEEE Transactions on*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[13] H. Yang and T. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 172–179, February 2013.

[14] F. Fernandes, A. Ashikhmin, and T. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 192–201, February 2013.

[15] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 1137–1141.

[16] M. Rasti and A.-R. Sharafat, "Distributed uplink power control with soft removal for wireless networks," *Communications, IEEE Transactions on*, vol. 59, no. 3, pp. 833–843, March 2011.

Referências

AGRELL, E.; ERIKSSON, T.; VARDY, A.; ZEGER, K. Closest point search in lattices. *Information Theory, IEEE Transactions on*, v. 48, n. 8, p. 2201–2214, Aug 2002. ISSN 0018-9448.

ALAMOUTI, S. A simple transmit diversity technique for wireless communications. *Selected Areas in Communications, IEEE Journal on*, v. 16, n. 8, p. 1451–1458, Oct 1998. ISSN 0733-8716.

ASHIKHMIN, A.; MARZETTA, T. Pilot contamination precoding in multi-cell large scale antenna systems. In: *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. [S.l.: s.n.], 2012. p. 1137–1141. ISSN 2157-8095.

BOCCARDI, F.; HEATH, R.; LOZANO, A.; MARZETTA, T.; POPOVSKI, P. Five disruptive technology directions for 5g. *Communications Magazine, IEEE*, v. 52, n. 2, p. 74–80, February 2014. ISSN 0163-6804.

CAIRE, G.; SHAMAI, S. On the achievable throughput of a multiantenna gaussian broadcast channel. *Information Theory, IEEE Transactions on*, v. 49, n. 7, p. 1691–1706, July 2003. ISSN 0018-9448.

CHOI, J.; PERREAU, S. Mmse multiuser downlink multiple antenna transmission for cdma systems. *IEEE Transactions on Signal Processing*, v. 52, n. 6, p. 1564–1573, 2004. ISSN 1053-587X.

DORIGO, M.; BIRATTARI, M.; STUTZLE, T. Ant colony optimization. *IEEE Computational Intelligence Magazine*, v. 1, n. 4, p. 28 –39, nov. 2006. ISSN 1556-603X.

FEHSKE, A.; FETTWEIS, G.; MALMODIN, J.; BICZOK, G. The global footprint of mobile communications: The ecological and economic perspective. *Communications Magazine, IEEE*, v. 49, n. 8, p. 55–62, August 2011. ISSN 0163-6804.

FERNANDES, F.; ASHIKHMIN, A.; MARZETTA, T. Inter-cell interference in noncooperative tdd large scale antenna systems. *Selected Areas in Communications, IEEE Journal on*, v. 31, n. 2, p. 192–201, February 2013. ISSN 0733-8716.

FOSCHINI, G. J.; GANS, M. J. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, v. 6, p. 311–335, 1998.

GESBERT, D.; KOUNTOURIS, M.; HEATH, R.; CHAE, C.-B.; SALZER, T. Shifting the mimo paradigm. *Signal Processing Magazine, IEEE*, v. 24, n. 5, p. 36–46, Sept 2007. ISSN 1053-5888.

GHRAYEB, A. A survey on antenna selection for mimo communication systems. In: *Information and Communication Technologies, 2006. ICTTA '06. 2nd.* [S.l.: s.n.], 2006. v. 2, p. 2104–2109.

GOLDSMITH, A. *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005. ISBN 0521837162.

GOLUB, G. H.; LOAN, C. F. V. *Matrix Computations*. Maryland, USA: Johns Hopkins University Press, 1996.

GUO, L.; HUANG, Y.-F. Interference suppression for multiuser downlink transmission in frequency-selective fading channels. *IEEE Transactions on Signal Processing*, v. 56, n. 9, p. 4386–4397, 2008. ISSN 1053-587X.

HABENDORF, R.; FETTWEIS, G. Nonlinear optimization for the multiuser downlink. In: *The 13th European Wireless Conference*. Paris, France: [s.n.], 2007.

HANZO, L.; AKHTMAN, Y.; WANG, L.; JIANG, M. *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus Non-coherent and Cooperative Turbo Transceivers*. [S.l.]: IEEE Press and John Wiley & Sons, 2010. ISBN 978-0-470-68669-0.

HE, S.; HUANG, Y.; JIN, S.; YU, F.; YANG, L. Max-min energy efficient beamforming for multicell multiuser joint transmission systems. *IEEE Communications Letters*, v. 17, n. 10, p. 1956–1959, 2013. ISSN 1089-7798.

HOYDIS, J.; BRINK, S. ten; DEBBAH, M. Massive mimo in the ul/dl of cellular networks: How many antennas do we need? *Selected Areas in Communications, IEEE Journal on*, v. 31, n. 2, p. 160–171, February 2013. ISSN 0733-8716.

IRMER, R.; HABENDORF, R.; RAVE, W.; FETTWEIS, G. Nonlinear multiuser transmission using multiple antennas for tdd-cdma. In: *Proc. Int. Symp. on Wireless Personal Multimedia Communications (WPMC'03)*. Yokosuka, Japan: [s.n.], 2003. p. 251–255.

JALDEN, J.; OTTERSTEN, B. On the complexity of sphere decoding in digital communications. *Signal Processing, IEEE Transactions on*, v. 53, n. 4, p. 1474–1484, April 2005. ISSN 1053-587X.

JALDEN, J.; OTTERSTEN, B. On the maximal diversity order of spatial multiplexing with transmit antenna selection. *Information Theory, IEEE Transactions on*, v. 53, n. 11, p. 4273–4276, Nov 2007. ISSN 0018-9448.

JOSE, J.; ASHIKHMIN, A.; MARZETTA, T.; VISHWANATH, S. Pilot contamination and precoding in multi-cell tdd systems. *Wireless Communications, IEEE Transactions on*, v. 10, n. 8, p. 2640–2651, August 2011. ISSN 1536-1276.

KAPETANOVIC, D.; CHENG, H.; MOW, W. H.; RUSEK, F. Optimal two-dimensional lattices for precoding of linear channels. *IEEE Transactions on Wireless Communications*, v. 12, n. 5, p. 2104–2113, 2013. ISSN 1536-1276.

KHURSHID, K.; IRTEZA, S.; KHAN, A. A. Application of ant colony optimization based algorithm in mimo detection. In: *IEEE Congress on Evolutionary Computation (CEC10)*. [S.l.: s.n.], 2010. (DOI:10.1109/CEC.2010.5586173), p. 1–7.

LAIN, J.-K.; CHEN, J.-Y. Near-mld mimo detection based on a modified ant colony optimization. *IEEE Communications Letters*, v. 14, n. 8, p. 722–724, 2010. ISSN 1089-7798.

LI, Q.; LI, G.; LEE, W.; LEE, M. il; MAZZARESE, D.; CLERCKX, B.; LI, Z. Mimo techniques in wimax and lte: a feature overview. *Communications Magazine, IEEE*, v. 48, n. 5, p. 86–92, May 2010. ISSN 0163-6804.

LOPES, H. S.; PERRETTO, M. Reconstruction of phylogenetic trees using the ant colony optimization paradigm. In: *III Brazilian Workshop on Bioinformatics (WOB 04)*. [S.l.: s.n.], 2004. p. 49–56.

MARINELLO, J. C.; ABRÃO, T. Lattice reduction aided detector for dense mimo via ant colony optimization. In: *IEEE Wireless Communications and Networking Conference (WCNC'13)*. [S.l.: s.n.], 2013. p. 2839–2844. ISSN 1525-3511.

MARINELLO, J. C.; ABRÃO, T. Lattice reduction aided detector for mimo communication via ant colony optimisation. *Wireless Personal Communications*, Springer US, p. 1–23, 2013. ISSN 0929-6212.

MARINELLO, J. C.; SOUZA, R. N. de; ABRÃO, T. Ant colony input parameters optimization for multiuser detection in ds/cdma systems. *Expert Systems with Applications*, v. 39, n. 17, p. 12876 – 12884, 2012. ISSN 0957-4174.

MARQUES, M. de P.; ABRÃO, T.; ADANIYA, M. H.; SAMPAIO, L. H. D.; JESZENSKY, P. J. E. Ant colony optimization for resource allocation and anomaly detection in communication networks. In: ABRÃO, T. (Ed.). *Search Algorithms*. [S.l.]: InTech Open, 2012. cap. 8, p. 1–34.

MARZETTA, T. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Transactions on Wireless Communications*, v. 9, n. 11, p. 3590–3600, 2010. ISSN 1536-1276.

MATHWORKS, T. *Optimization Toolbox User's Guide, v.7.0 Release 2014a*. March 2014. Disponível em: <`http://www.mathworks.co.uk/help/pdf_doc/optim/optim_tb.pdf`>.

MIETZNER, J.; SCHOBER, R.; LAMPE, L.; GERSTACKER, W.; HOEHER, P. Multiple-antenna techniques for wireless communications - a comprehensive literature survey. *Communications Surveys Tutorials, IEEE*, v. 11, n. 2, p. 87–105, Second 2009. ISSN 1553-877X.

NAM, Y.-H.; NG, B. L.; SAYANA, K.; LI, Y.; ZHANG, J.; KIM, Y.; LEE, J. Full-dimension mimo (fd-mimo) for next generation cellular technology. *Communications Magazine, IEEE*, v. 51, n. 6, p. 172–179, June 2013. ISSN 0163-6804.

NGO, H. Q.; LARSSON, E.; MARZETTA, T. Energy and spectral efficiency of very large multiuser mimo systems. *Communications, IEEE Transactions on*, v. 61, n. 4, p. 1436–1449, April 2013. ISSN 0090-6778.

NGUYEN, D.; TRAN, L.-N.; PIRINEN, P.; LATVA-AHO, M. Precoding for full duplex multiuser mimo systems: Spectral and energy efficiency maximization. *IEEE Transactions on Signal Processing*, v. 61, n. 16, p. 4038–4050, 2013. ISSN 1053-587X.

NISHIMOTO, H.; OGAWA, Y.; NISHIMURA, T.; OHGANE, T. Measurement-based performance evaluation of mimo spatial multiplexing in a multipath-rich indoor environment. *Antennas and Propagation, IEEE Transactions on*, v. 55, n. 12, p. 3677–3689, Dec 2007. ISSN 0018-926X.

NOCEDAL, J.; WRIGHT, S. J. *Numerical Optimization*. [S.l.]: New York: Springer Series in Operations Research, 1999.

PAUL, T.; OGUNFUNMI, T. Evolution, insights and challenges of the phy layer for the emerging ieee 802.11n amendment. *Communications Surveys Tutorials, IEEE*, v. 11, n. 4, p. 131–150, Fourth 2009. ISSN 1553-877X.

POOR, H.; VERDU, S. Probability of error in mmse multiuser detection. *Information Theory, IEEE Transactions on*, v. 43, n. 3, p. 858–871, May 1997. ISSN 0018-9448.

RASTI, M.; SHARAFAT, A.-R. Distributed uplink power control with soft removal for wireless networks. *Communications, IEEE Transactions on*, v. 59, n. 3, p. 833–843, March 2011. ISSN 0090-6778.

RENZO, M. D.; HAAS, H.; GHRAYEB, A.; SUGIURA, S.; HANZO, L. Spatial modulation for generalized mimo: Challenges, opportunities, and implementation. *Proceedings of the IEEE*, v. 102, n. 1, p. 56–103, Jan 2014. ISSN 0018-9219.

RUSEK, F.; PERSSON, D.; LAU, B. K.; LARSSON, E.; MARZETTA, T.; EDFORS, O.; TUFVESSON, F. Scaling up mimo: Opportunities and challenges with very large arrays. *IEEE Signal Processing Magazine*, v. 30, n. 1, p. 40–60, 2013. ISSN 1053-5888.

SOCHA, K.; DORIGO, M. Ant colony optimization for continuous domains. *European Journal of Operational Research*, v. 185, n. 3, p. 1155 – 1173, 2008. ISSN 0377-2217. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0377221706006333>.

SRINIDHI, N.; DATTA, T.; CHOCKALINGAM, A.; RAJAN, B. Layered tabu search algorithm for large-mimo detection and a lower bound on ml performance. *Communications, IEEE Transactions on*, v. 59, n. 11, p. 2955–2963, November 2011. ISSN 0090-6778.

SUGIURA, S.; CHEN, S.; HANZO, L. A universal space-time architecture for multiple-antenna aided systems. *Communications Surveys Tutorials, IEEE*, v. 14, n. 2, p. 401–420, Second 2012. ISSN 1553-877X.

TAROKH, V.; SESHADRI, N.; CALDERBANK, A. Space-time codes for high data rate wireless communication: performance criterion and code construction. *Information Theory, IEEE Transactions on*, v. 44, n. 2, p. 744–765, Mar 1998. ISSN 0018-9448.

TELATAR, E. Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, Wiley Subscription Services, Inc., A Wiley Company, v. 10, n. 6, p. 585–595, 1999. ISSN 1541-8251.

TSE, D.; VISWANATH, P. *Fundamentals of Wireless Communications*. Inglaterra: Cambridge University Press, 2010.

VARDHAN, K.; MOHAMMED, S.; CHOCKALINGAM, A.; RAJAN, B. A low-complexity detector for large mimo systems and multicarrier cdma systems. *Selected Areas in Communications, IEEE Journal on*, v. 26, n. 3, p. 473–485, April 2008. ISSN 0733-8716.

VOJCIC, B.; JANG, W. M. Transmitter precoding in synchronous multiuser communications. *IEEE Transactions on Communications*, v. 46, n. 10, p. 1346 –1355, oct 1998. ISSN 0090-6778.

WANG, K.; ZHANG, X. Penalty function-based precoding for downlink multiuser mimo systems. *AEU - International Journal of Electronics and Communications*, v. 67, n. 2, p. 167 – 173, 2013. ISSN 1434-8411.

WOLNIANSKY, P.; FOSCHINI, G.; GOLDEN, G.; VALENZUELA, R. V-blast: an architecture for realizing very high data rates over the rich-scattering wireless channel. In: *Signals, Systems, and Electronics, 1998. ISSSE 98. 1998 URSI International Symposium on*. [S.l.: s.n.], 1998. p. 295 –300.

WOLPERT, D.; MACREADY, W. No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on*, v. 1, n. 1, p. 67–82, Apr 1997. ISSN 1089-778X.

WUBBEN, D.; BOHNKE, R.; KUHN, V.; KAMMEYER, K.-D. Near-maximum-likelihood detection of mimo systems using mmse-based lattice reduction. In: *IEEE International Conference on Communications*. [S.l.: s.n.], 2004. v. 2, p. 798 – 802 Vol.2.

WUBBEN, D.; SEETHALER, D.; JALDEN, J.; MATZ, G. Lattice reduction. *Signal Processing Magazine, IEEE*, v. 28, n. 3, p. 70–91, May 2011. ISSN 1053-5888.

XU, C.; YANG, L.-L.; HANZO, L. Ant-colony-based multiuser detection for mc ds-cdma systems. In: *IEEE 66th Vehicular Technology Conference. VTC-2007 Fall.* [S.l.: s.n.], 2007. p. 960 –964. ISSN 1090-3038.

YAO, W.; CHEN, S.; HANZO, L. Particle swarm optimisation aided mimo multiuser transmission designs. *Journal of Computational and Theoretical Nanoscience*, v. 2, n. 9, p. 266–275, 2012.

YAO, W.; CHEN, S.; TAN, S.; HANZO, L. Minimum bit error rate multiuser transmission designs using particle swarm optimisation. *IEEE Transactions on Wireless Communications*, v. 8, n. 10, p. 5012–5017, 2009. ISSN 1536-1276.

ZHAO, Y.; XU, X.; HAO, Z.; TAO, X.; ZHANG, P. Resource allocation in multiuser ofdm system based on ant colony optimization. In: *IEEE Wireless Communications and Networking Conference (WCNC'10)*. [S.l.: s.n.], 2010. p. 1 –6. ISSN 1525-3511.