Random matrix theory for advanced communication systems.

Jakob Hoydis

N° d'ordre : 2012-04-TH

THÈSE DE DOCTORAT
SPÉCIALITÉ: PHYSIQUE

École doctorale "Sciences et Technologies de l'Information, des Télécommunications et des Systèmes"

Présentée par :

Jakob HOYDIS

Sujet:

Matrices aléatoires pour les futurs systèmes de communication

(Random Matrix Methods for Advanced Communication Systems)

Soutenue le 5 avril 2012 devant les membres du jury:

M. Giuseppe Caire,	University of Southern California	Rapporteur
M. Mérouane Debbah,	Supélec	Examinateur, Encadrant de Thèse
M. Pierre Duhamel,	CNRS/Supélec	Examinateur, Président du Jury
M. Gerhard Fettweis,	Technische Universität Dresden	Examinateur
M. Walid Hachem,	CNRS/Télécom ParisTech	Examinateur
Mme. Mari Kobayashi,	Supélec	Examinatrice, Directrice de Thèse
M. Erik G. Larsson,	Linköping University	Examinateur
M. Philippe Loubaton,	Université de Marne la Vallée	Rapporteur
M. Hikmet Sari,	Supélec	Invité
M. Stephan ten Brink,	Alcatel-Lucent, Bell Labs	Invité

Contents

Acknowledgments

I would like to express my deepest gratitude to both of my supervisors Mari Kobayashi and Mérouane Debbah. Mari guided me especially at the beginning of my PhD with constructive feedback and careful reviews of my first papers. Mérouane provided me with a constant flow of creative research ideas (often more than I could handle) and gave me the freedom to work on the topics I liked. I am particularly thankful for the numerous opportunities both gave me to present my work, for their permanent availability for questions and discussions, for their encouragement and their invaluable support.

I would like to thank all jury members for their participation in my PhD defense and for their kind and motivating comments. Special thanks goes to Giuseppe Caire and Philippe Loubaton for their careful reviews of the manuscript and for pointing out mistakes in some of the proofs, that I hope to have corrected by now.

I am also deeply indebted to many people without whom my PhD would not have been the same. In particular, I would like to thank

- Romain Couillet who showed me the first paper on random matrices and stimulated my interest for this topic. I owe most of my knowledge in this area to him and to his book which I reviewed at different levels of completion. I will always look back with pleasure on the countless hours of intensive work, fruitful discussions, and brainstorming on new research problems we have spent together.
- Jamal Najim for his explanations and discussions on random matrices and the fundamentals of asymptotic statistics which helped me a lot.
- Pablo Piantanida who made me familiar with the world of finite block-length codes and broadened my information-theoretic understanding of various topics.
- Stephan ten Brink for making my visit at Alcatel-Lucent in Stuttgart such a great experience. I admire his critical eye for interesting research problems and his drive to squeeze the last bit of insight out of technical results. I look forward to having him as my future boss.
- All of the members of the Alcatel-Lucent chair, the telecommunications department, and LSS for making my PhD a very pleasant experience; among them in particular Joffrey, Marios, Salam, Subash and my office mates Ahmed, Arash, Ayaz, Daniel, and Najett.
- Alexandre de Baynast for supporting my decision to start a PhD in France.

- My good friend Adi for sharing our studies together and discussing the ups and downs of our lives as PhD students.

My last acknowledgments are for those who are closest to me:

Je voudrais remercier Marie-José, Louis et Christophe pour leur aide, soutien et encouragement. J'ai beaucoup apprécié les nombreux repas passés ensemble ainsi que les après-midis dans votre jardin. Vous êtes devenus comme une deuxième famille pour moi.

Von tiefstem Herzen danke ich meinen Eltern für ihre Liebe, Anteilnahme und bedingungslose Unterstüzung, und meiner Schwester Julia, die mir immer einen Schritt voraus ist und mir oft zeigt, was wirklich wichtig im Leben ist.

Finalement, je voudrais remercier de tout cœur ma meilleure amie et future femme Murielle qui m'a supporté et encouragé pendant ces trois dernières années de travail intense. J'espère que tu me pardonnes mes moments d'absence lors de nombreux week-ends et soirées. Tu apportes beaucoup de stabilité à ma vie et je suis infiniment reconnaissant de t'avoir.

Jakob Hoydis, April 2012

Abstract

Advanced mobile communication systems are characterized by an increasingly dense deployment of different types of wireless access points. Since these systems are primarily limited by interference, multiple-input multiple-output (MIMO) techniques as well as coordinated transmission and detection schemes are necessary to mitigate this limitation. As a consequence, mobile communication systems become more complex which requires that also the mathematical tools for their theoretical analysis must evolve. In particular, these must be able to take the most important system characteristics into account, such as fading, path loss, interference, and imperfect channel state information. The aim of this thesis is to develop such tools based on large random matrix theory and to demonstrate their usefulness with the help of several practical applications. These include the performance analysis of network MIMO and large-scale MIMO systems, the design of low-complexity polynomial expansion detectors, and the study of random beamforming techniques as well as multi-hop relay and double-scattering channels. In summary, the methods developed in this work provide deterministic approximations of the system performance (e.g., in terms of mutual information, achievable rates, or signal-to-interference-plus-noise ratio (SINR)) which become arbitrarily tight in the large system regime with an unlimited number of transmitting and receiving devices. This leads in many cases to surprisingly simple and close approximations of the finite-size system performance and allows one to draw relevant conclusions about the most significant parameters. One can think of these methods as a way to provide a deterministic abstraction of the physical layer which substantially reduces the system complexity. Due to this complexity reduction, it is possible to carry out a system optimization which would otherwise be intractable.

Abstract (français)

Les futurs systèmes de communication mobile sont caractérisés par un déploiement de plus en plus dense de différents types de points d'accès sans fil. Étant donné que ces systèmes sont principalement limitée par les interférences, les techniques aux entrées multiples, sorties multiples, dit “multiple-input multiple-output (MIMO)” ainsi que la cooperation entre les émetteurs et/ou les récepteurs sont nécessaires pour atténuer cette limitation. En conséquence, les systèmes de communication mobiles deviennent plus complexes, qui exige que aussi les outils mathématiques pour leur analyse théorique doit évoluer. En particulier, ces outils doivent être en mesure de prendre en compte les caractéristiques du système les plus importants, tels que l'affaiblissement de propagation, les interférences, et l'information imparfaite d'état du canal. Le but de cette thèse est de développer de tels outils basés sur la théorie des grandes matrices aléatoires et de démontrer leur utilité à l'aide de plusieurs applications pratiques. Il s'agit notamment de l'analyse des performances des systèmes “network MIMO” et des systèmes MIMO à grande échelle, dit “massive MIMO”, la conception de détecteurs de faible complexité à expansion polynomiale, l'étude des techniques de precodage unitaire aléatoire ainsi que l'analyse de canaux à relais multiples et de canaux à double diffusion. En résumé, les méthodes développées dans ce travail fournissent des approximations déterministes de la performance du système (par exemple, en termes d'information mutuelle, des taux réalisables, ou du rapport signal sur bruit plus interférence) qui deviennent arbitrairement serrés dans une régime asymptotique avec un nombre illimité d'émetteurs et de récepteurs. Cette approche conduit souvent à des approximations de la performance du système étonnamment simples et précises et permet de tirer des conclusions importantes sur les paramètres les plus pertinents. On peut penser à cette méthode comme un moyen pour fournir une abstraction déterministe de la couche physique qui réduit significativement la complexité du système. Grâce à cette réduction de la complexité, il est par exemple possible d'optimiser certains paramètres du système, ce qui aurait été impossible autrement.

Acronyms

AF	amplify-and-forward
BER	bit-error rate
BF	eigen beamforming
BS	base station
CDMA	code-division multiple access
CLT	central limit theorem
CoMP	coordinated multi-point
CS	central station
CSI	channel state information
DMT	diversity-multiplexing trade-off
DoF	degrees of freedom
e.s.d.	empirical spectral distribution
HARQ	hybrid automatic repeat request
i.i.d.	independent and identically distributed
ICT	information and communication technology
LDPC	low density parity-check
LHS	left-hand side
LOS	line-of-sight
MAC	multiple access channel
MF	matched filter
MIMO	multiple-input multiple-output
MMSE	minimum-mean-square-error
OFDM	orthogonal frequency-division multiplexing
RHS	right-hand side
RMT	random matrix theory
RZF	regularized zero-forcing
SDMA	space division multiple access
SIC	successive interference cancellation
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
TDD	time-division duplexing
TDMA	time-division multiple access
UT	user terminal

Notations

Linear algebra	
$\mathbf{X}$	matrix
$\mathbf{I}_N$	identity matrix of size $N \times N$
$\mathbf{0}_{N\times K}$	all zero matrix of size $N \times K$
$\mathrm{diag}(x_1, \ldots, x_N)$	diagonal matrix with entries $x_1, \ldots, x_N$
$[\mathbf{X}]_{i,j}$, $(\mathbf{X})_{i,j}$, $\mathbf{X}_{i,j}$	(i, j) entry of matrix $\mathbf{X}$
$\mathbf{X}^{\mathsf{T}}$	transpose of $\mathbf{X}$
$\mathbf{X}^{\mathsf{H}}$	complex conjugate transpose of $\mathbf{X}$
$\mathbf{X}^*$	complex conjugate of $\mathbf{X}$
$\mathrm{tr}\,\mathbf{X}$	trace of $\mathbf{X}$
$\det \mathbf{X}$	determinant of $\mathbf{X}$
$\mathrm{rank}\,(\mathbf{X})$	rank of matrix $\mathbf{X}$
$\mathbf{X} \otimes \mathbf{Y}$	Kronecker or Tensor product of $\mathbf{X}$ and $\mathbf{Y}$
$\\|\mathbf{X}\\|$	spectral norm of $\mathbf{X}$
$\mathbf{x}$	column vector
x_i	ith entry of vector $\mathbf{x}$
$\mathbf{1}_N$, $\mathbf{0}_N$	all one and all zero column vector of size N
$\geq, \leq, >, <$	component-wise inequalities, e.g., $\mathbf{x} \geq \mathbf{y}$ implies that $x_i \geq y_i \, \forall i$
$\succeq, \preceq, \succ, \prec$	matrix inequalities, e.g., $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is nonnegative definite

Analysis	
$\mathbb{C}$, $\mathbb{R}$, $\mathbb{N}$	the complex, real, and natural numbers
$\mathbb{C}^+$	$\{z \in \mathbb{C} : \Im\{z\} > 0\}$
$\mathbb{R}^+, \mathbb{R}^-$	the nonnegative and nonpositive real numbers
$\mathbb{N}^\star$	the positive natural numbers
$\lvert x \rvert$	absolute value
$(x)^+$	$\max(x, 0)$
$\Re\{z\}$	real part of z
$\Im\{z\}$	imaginary part of z
$\mathbf{i}$	$\mathbf{i} = \sqrt{-1}$ with $\Im\{\mathbf{i}\} = 1$
$\mathbb{1}\{A\}$	indicator function, i.e., $\mathbb{1}\{A\} = 1$ iff A is true, $\mathbb{1}\{A\} = 0$ otherwise
$\mathcal{O}(y_n)$	Landau's big-$\mathcal{O}$ notation, i.e., $x_n = \mathcal{O}(y_n)$ implies $\lvert x_n \rvert \leq C \lvert y_n \rvert$, for $C > 0$ and $n > n_0$
$(x_n)_{n\geq 1}$	infinite sequence of numbers (or sets) $x_1, x_2, \ldots$

$f'(x)$	first derivative of $f(x)$
$\log(x)$	natural logarithm
$\limsup_n x_n$	limit superior of $(x_n)_{n\geq 1}$, i.e., for every $\epsilon > 0$, there exists $n_0(\epsilon)$, such that $x_n \leq \limsup_n x_n + \epsilon \ \forall n > n_0(\epsilon)$
$\liminf_n x_n$	limit inferior, i.e., $\liminf_n x_n = -\limsup_n -x_n$

Probability theory

$(\Omega, \mathcal{F}, P)$	probability space Ω with σ-Algebra $\mathcal{F}$ and measure P
F_X	distribution function of X, i.e., $F_X(x) = P(X \leq x)$
$\mathrm{supp}(F)$	support of the distribution function F
$\mathbb{E}[X]$	expectation of X, i.e., $\mathbb{E}[X] = \int_\Omega X(\omega) dP(\omega)$
$\xrightarrow{\text{a.s.}}$	almost sure convergence
$\Rightarrow$	convergence in distribution
$\sim$	distributed as, e.g., $X \sim \mathcal{N}(0,1)$
$x_n \asymp y_n$	asymptotic equivalence, i.e., $x_n - y_n \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$
$\mathcal{CN}(\mathbf{m}, \mathbf{Q})$	complex Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{Q}$
$\mathcal{N}(\mathbf{m}, \mathbf{Q})$	Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{Q}$
$Q(x)$	Q-function, i.e., $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

Résumé

Introduction

La demande croissante en applications multimédia mobiles a engendré une très forte augmentation du flux de données échangées entre les systèmes de communication sans fil [1]. L'usage des "smartphones" a notamment contribué au passage des communications traditionnelles vers des communications multimédia à haute vitesse, partout et à tout moment. Par conséquent, les attentes des utilisateurs évoluent également dans le sens d'une interconnexion continue, aussi bien à domicile qu'en déplacement. Cependant, du fait de cette pénétration croissante des "smartphones", les réseaux atteignent déjà leurs limites de capacité lorsque la demande est forte. En effet, des problèmes de congestion apparaissent à la fois au niveau du lien sans fil et du réseau physique filaire.

L'intérêt porté aux technologies écologiques a récemment été suscité par l'étude "SMART 2020" [2], décrivant les effets potentiels des technologies de l'information et des communications (TIC) sur les émissions mondiales de dioxyde de carbone. Bien que les TIC ne semblent que modérément contribuer aux émissions mondiales, avec 1,25 % en 2002 et une prévision d'environ 2,5 % en 2020, cela représente tout de même une augmentation de 10 % par an. Ceci signifie qu'en dépit de progrès significatifs dans le domaine des technologies économes en énergie, notre capacité à réduire (ou du moins à maintenir) la consommation énergétique et les émissions associées, sera très vite contrariée par l'augmentation des flux de données. Ainsi l'augmentation de la capacité des réseaux d'une part et la réduction de la consommation en énergie d'autre part sont deux besoins futurs, a priori contradictoires. D'où la question : comment les opérateurs mobiles peuvent-ils satisfaire les futures demandes de transport de données d'un point de vue aussi bien économique qu'écologique ? Dans un premier temps, regardons les différentes options afin d'augmenter l'efficacité spectrale des réseaux mobiles.

Plus de spectre

Presque tous les systèmes de communication mobile actuels utilisent des fréquences entre 300 MHz–3 GHz. Ceci est dû aux caractéristiques de propagation des ondes électromagnétiques qui sont favorables dans cette bande. Par conséquent, cette partie du spectre est quasiment complètement allouée et les ressources spectrales libres se trouvent uniquement dans des fréquences plus élevées. Pour cette raison, les systèmes de communication dits "millimeter-wave (mmWave)" exploitant le spectre entre 3-300GHz attirent de plus en plus l'attention [3]. Aujourd'hui, seuls les systèmes de communication non-mobile ou à courte distance utilisent ces bandes. En raison d'un très grand affaiblissement de propagation, la communication avec des signaux de très courte longueur d'onde est essentiellement limitée aux liens à visibilité directe, dits “line-of-sight (LOS)”. Ainsi, assurer une couverture du réseau à l'intérieur à partir de stations de bases (SBs) déployées à l'extérieur semble impossible. Néanmoins, plusieurs projets de recherches actuels étudient la possibilité d'utilisation de mmWaves pour des réseaux mobiles [4]. Cependant, cette recherche en est encore à ses premiers balbutiements.

Meilleurs codages et techniques de modulation

Grâce aux “low-density parity check” (LDPC) [5] et turbo codes [6], l'“orthogonal frequency-division multiplexing” (OFDM), l'adaptation de lien et l'“hybrid automatic request” (HARQ), les systèmes modernes de communication mobile tels que LTE [7] ou WiMAX [8] atteignent déjà des efficacités spectrales proches des limites théoriques. Ainsi, aucune percée scientifique dans les domaines du codage et de la modulation n'est attendue dans les années à venir. Parmi les objectifs de recherche en cours on trouve des techniques de modulation non-orthogonales comme l'“Isotropic Orthogonal Transform Algorithm” (IOTA)-OFDM [9], ou des techniques comme le “Vandermonde frequency-division multiplexing” (VFDM) [10] qui exploitent le préfixe cyclique de l'OFDM pour permettre des émissions à un réseau secondaire sans interférence. En outre, les “Fountain codes” ou les “rateless codes” [11] qui généralisent dans un certain sens la notion de HARQ, ainsi que les codes polaires [12] promettent une augmentation supplémentaire de l'efficacité spectrale. Toutefois, en dépit de ces efforts, il est peu probable que les progrès en matière de codage ou de modulation soient le principal moteur d'une augmentation de la capacité des futurs systèmes de communication mobile.

Plus d'antennes

Depuis les articles fondateurs [13, 14] dans les années 1990, les systèmes à “entrées multiples, sorties multiples” des communications sans fil, dits MIMO (“multiple-input multiple-output”), ont suscité un grand intérêt de la part des chercheurs et de l'industrie. A l'heure actuelle, les bénéfices générés par des émetteurs et/ou récepteurs à plusieurs antennes sont bien compris [15]. Les techniques MIMO peuvent apporter des gains de puissance, améliorer la fiabilité des liaisons et augmenter le débit par un multiplexage spatial de plusieurs flux de données sur la même ressource temps-fréquence. Pour cette raison, les techniques MIMO font déjà partie intégrante des normes modernes de communication sans fil [7, 8], permettant aux SBs et aux terminaux d'utilisateurs (TUs) de supporter jusqu'à huit antennes. Une autre option intéressante, qui fait l'objet de nombreux projets de recherche [16, 17, 18], réside dans les systèmes MIMO de grande échelle, dits “large-scale MIMO” ou “massive MIMO”, où les SBs sont équipées de centaines d'antennes. En théorie, ceci peut permettre une réduction significative de la puissance d'émission, tout en réalisant une très grande efficacité spectrale. Bien que ne ressortissant pas directement des techniques MIMO, la sectorisation de cellule d'ordre supérieur [19] et l'utilisation d'antennes directives et intelligentes, dits “3D beamforming”, [20, 21], apparaissent comme des moyens efficaces pour augmenter l'efficacité spectrale. Ces techniques visent à accroître la répartition spatiale tout en concentrant l'énergie transmise vers les TUs destinataires et en réduisant l'interférence vers les autres TUs.

Plus de cellules

Il est bien connu que la réduction de la taille des cellules est le moyen le plus simple et le plus efficace pour augmenter le débit sans fil [22, 23]. La physique nous apprend que rapprocher un émetteur et un récepteur réduit la puissance d'émission nécessaire pour surmonter l'affaiblissement de propagation et d'autres phénomènes, tels que le "fading" (des évanouissements rapides de la puissance reçue) et le bruit thermique. En outre, le débit par unité de surface augmente en théorie linéairement avec la densité cellulaire. Cependant, le déploiement de SBs traditionnelles, telles que les macro- ou microcellules, nécessite des dépenses d'investissement de capital (CAPEX) et d'exploitation (OPEX) colossales pour l'acquisition de sites, la planification du réseau, la fourniture d'un réseau de transport, dit "backhaul", l'exploitation et la maintenance de ce réseau. Ainsi, on constate une popularité croissante des femto-cellules [24], i.e., des SBs de petite taille installées par l'utilisateur lui-même à son domicile ou sur son lieu de travail, en utilisant la connexion internet existante comme "backhaul". Celles-ci permettent de décharger les réseaux de données mobiles et d'assurer une bonne couverture à l'intérieur à coût minimal pour l'opérateur.

Actuellement, les milieux universitaires et industriels portent un intérêt considérable aux réseaux à petites cellules, dits "small-cell networks" (SCNs). Pour simplifier, les SCNs reposent sur un déploiement extrêmement dense de SBs à faible coût et faible puissance d'émission, qui sont bien plus petites que les équipements des macro-cellules traditionnelles. L'action des SCNs est rendue possible grâce au partage du lien filaire avec les réseaux sans fil et les points d'accès filaires existants (FTTx ("fiber to the x") ou les points d'accès filaires VDSL ("very-high-bitrate digital subscriber line") par exemple), mais aussi grâce à l'installation des SBs sur les équipements urbains (lampadaires et abris bus par exemple). Par ailleurs, le fonctionnement des SCNs repose sur les fonctionnalités d'auto-organisation des petites cellules. Ainsi, les SCNs permettent d'éviter l'acquisition de sites d'installation, la planification détaillée du réseau et la maintenance continue. Les SCNs permettent par conséquent de réduire le CAPEX et l'OPEX tout en assurant de hauts débits délivrés de manière uniforme dans la zone de couverture du réseau [25, 26].

Coopération et coordination

Les réseaux cellulaires sont avant tout limités par les interférences intercellulaires. Dans le cas où les SBs seraient autorisées à coopérer ou à coordonner, ces interférences pourraient être exploitées ou réduites [27]. En général, on fait la distinction entre la coopération (également "network MIMO" ou "multi-cell processing"), i.e. plusieurs SBs sont reliées entre elles par un réseau de transport et traitent conjointement leurs données, et la coordination, i.e. des groupes de SBs, dits "clusters", décident conjointement des stratégies de précodage/décodage mais ne partagent pas les données des utilisateurs. Les deux techniques, communément désignées sous le terme "coordinated multi-point (CoMP)", peuvent améliorer la couverture et le débit, en particulier pour les TUs au bord de la cellule, et ce sans la nécessité de déployer de nouveaux sites ou des antennes supplémentaires. Cependant, de nombreux défis techniques doivent être résolus avant que les techniques CoMP puissent voir le jour [28]. En outre, les gains effectifs générés par ces techniques sont moins prometteurs à partir du moment où les frais généraux relatifs à l'obtention d'information d'état du canal (CSI) sont pris en compte [29]. La recherche actuelle fait également état d'autres techniques intéressantes telles que les schémas d'alignement d'interférences [30], les concepts de relais [31], et les réseaux "cloud" sans fil, dits "wireless network clouds" [32]. Ce dernier concept diffère radicalement de l'architecture classique des réseaux les SBs sont désormais remplacées par des têtes de radio à distance connectées par des fibres optiques, le traitement du signal étant effectué sur des serveurs centralisés.

Radio cognitive

"La plupart des bandes de fréquences est sous-utilisée dans la plupart des endroits, la plupart du temps" [33, 34]. L'idée de radios cognitives [35] émane de cette observation, celles-ci sont autorisées à se servir de certaines parties du spectre sous licence, étant donné qu'elles peuvent détecter de manière fiable si une bande est utilisée ou non. Dans ce contexte, on fait souvent référence aux trous du spectre ("spectrum holes") dans le temps, dans la fréquence, ou dans l'espace, que les radios cognitives cherchent à exploiter. Malgré l'intensité de la recherche sur ce sujet au cours de la dernière décennie, les radios cognitives n'ont en pratique pas encore connu de succès. Ceci est dû à plusieurs raisons, à savoir un manque d'algorithmes de détection fiables, de protocoles, d'équipements, et de contraintes réglementaires. Il semble ainsi peu probable que les radios cognitives joueront un rôle majeur dans la prochaine génération de systèmes de communication mobile. Cependant, le concept de radios intelligentes ou flexibles, capables d'interagir avec l'environnent et de prendre des décisions intelligentes, est une idée prometteuse [36].

D'autres techniques

Il y a bien sûr beaucoup d'autres techniques qui offrent des possibilités intéressantes pour l'amélioration de l'efficacité spectrale. Parmi elles, on trouve les émetteurs-récepteurs full-duplex, qui pourraient théoriquement doubler la capacité des réseaux actuels [37], la polarisation électromagnétique [38, 39], qui pourrait conduire à une triplement de la capacité, et les approches croix-couche, dites "cross-layer", tel que le décodage conjoint de la source et du canal [40], dit "joint source-channel decoding", qui exploite la redondance aux différentes couches de protocole.

Comme mentionné ci-dessus, il devient clair que les futurs systèmes de communication mobile sont susceptibles d'être constitués d'un déploiement dense de différents types de points d'accès sans fil (macro-, micro-, femto-cellules) avec des caractéristiques différentes (puissance d'émission, nombre d'antennes, installé par l'operateur/l'utilisateur, accès fermé/ouvert). De plus, les techniques CoMP ont le bénéfice d'atténuer les interférences croissantes dans ces réseaux denses. Afin d'évaluer lesquelles des techniques présentées ci-dessus sont les plus adaptées pour atteindre un objectif fixé (augmentation de la capacité, réduction de la consommation d'énergie ou réduction des coûts), une analyse théorique de la performance est indispensable. Toutefois, les modèles de canaux et les outils qui ont été développés pour l'analyse de liens point à point ne parviennent que rarement à fournir des indications pertinentes dans le cas de grands réseaux hétérogènes de plus en plus denses. Il est donc nécessaire de développer de nouveaux outils théoriques pour leur analyse, ceci étant l'objectif principal de cette thèse. Dans ce qui suit, nous allons nous concentrer presque exclusivement sur la caractérisation des limites de performance théorique. L'efficacité énergétique, malgré son importance, ne sera pas prise en compte dans ce travail.

Pourqoui aller à l'infini ?

La principale difficulté d'une analyse de la performance théorique significative réside à la fois (i) dans le choix du modèle le plus simple qui reflète suffisamment les caractéristiques principales du système étudié et (ii) dans le choix de l'outil mathématique qui permet de tirer des conclusions générales et, éventuellement, de fournir des lignes techniques directrices. Par exemple, il est difficile de justifier pourquoi les conclusions établies par un modèle de canal sans affaiblissement de propagation et sans interférence devraient être valables pour des scénarios plus complexes combinant ces deux aspects. Les caractéristiques les plus importantes des systèmes de communication sans fil considérées dans cette thèse sont résumées ci-dessous :

- **Fading** : L'évanouissement ou "fading", décrivant les fluctuations aléatoires de la force du signal reçu dans le temps et en fréquence, constitue l'une des caractéristiques principales des canaux sans fil. On distingue généralement l'évanouissement à petite échelle, dit "fast fading", de celui à grande échelle, dit "shadowing". Le premier est causé par des petits mouvements du récepteur, de l'émetteur ou des diffuseurs, tandis que le second est dû au "shadowing" par des obstacles importants, tels que des bâtiments ou des murs.

- **Affaiblissement de propagation** : L'intensité moyenne du signal reçu est principalement déterminée par l'affaiblissement de propagation, dit "path loss", dépendant de la distance entre l'émetteur et le récepteur. La façon dont l'affaiblissement de propagation est calculé dépend de l'environnement (urbain ou rural par exemple) et des conditions de propagation radio (conditions météorologiques par exemple). Les effets du "path loss" sont particulièrement importants lorsque les techniques coopératives sont considérées, car leur avantage dépend de la position des TUs dans une cellule.

- **Interférences** : Les réseaux mobiles sont limités par les interférences créées par les transmissions d'autres stations de base et des TUs. Par conséquent, cet aspect ne peut être ignoré et doit être explicitement pris en compte.

- **Information imparfaite d'état du canal** : L'hypothèse d'une connaissance parfaite de l'état du canal, dit "channel state information (CSI)", est pratique pour l'analyse théorique, mais conduit souvent à des résultats

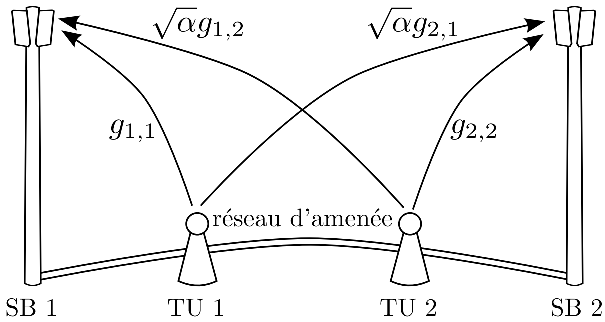

Figure 1: Illustration de modele de système: Deux stations de base reliées par un réseau de transport de capacité infinie décodent conjointement les messages des deux TUs.

trop optimistes. Ainsi, l'acquisition de CSI doit être modélisée en prenant en compte les ressources nécessaires. En particulier, dans les systèmes multicellulaires, la réutilisation des séquences d'apprentissage ou de symboles pilotes dans les cellules adjacentes peut conduire à une source supplémentaire d'interférence, souvent désignée comme " pilot contamination" [41].

- **Réseau de transport ("backhaul") d'une capacité finie** : Si les techniques COMP sont analysées, il est nécessaire d'être précis sur la quantité de données qui doivent être échangées entre les SBs coopérantes. Dans certains cas, la capacité du réseau de transport, dite "backhaul capacity", peut être un facteur limitant et donc non négligeable.

- **Liens à visibilité directe et corrélation d'antennes** : Plus un réseau est dense, plus il est probable que le lien entre un émetteur et un récepteur soit à visibilité directe, dite "line-of-sight (LOS)". Ainsi, le canal sans fil qui les sépare est constitué d'une composante déterministe LOS et d'une composante aléatoire liée aux évanouissements rapides. En outre, l'hypothèse d'un environnement de diffusion riche et d'antennes non corrélées peut s'avérer fausse pour des systèmes MIMO de grande échelle. Il est donc nécessaire de modéliser les liens à visibilité directe et la corrélation d'antennes qui peuvent avoir un impact significatif sur les performances du système.

Afin de se faire une idée de la difficulté liée à l'analyse théorique lorsque seulement l'un de ces aspects est pris en compte, nous considérons l'exemple simple suivant, qui se concentre exclusivement sur les effets d'affaiblissement de propagation. Comme le montre la figure 1, deux stations de base reliées par un réseau de transport de capacité infinie décodent conjointement les messages des deux TUs. Les SBs et les TUs sont tous équipés d'une antenne unique. Le vecteur $\mathbf{y} = [y_1, y_2]^{\mathsf{T}} \in \mathbb{C}^2$ des signaux reçus par les deux stations de base à un instant donné est

$$\begin{aligned}\mathbf{y} &= \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{n} \\ &= \sqrt{\rho}\begin{pmatrix} g_{1,1} & \sqrt{\alpha}\, g_{1,2} \\ \sqrt{\alpha}\, g_{2,1} & g_{2,2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \end{aligned} \tag{1}$$

où $g_{i,j} \sim \mathcal{CN}(0,1)$ est le coefficient du canal entre TU j et SB i, $x_j \sim \mathcal{CN}(0,1)$ est le signal transmis de TU j, $n_i \sim \mathcal{CN}(0,1)$ est un bruit thermique, ρ est la puissance d'émission et $\alpha \in [0,1]$ est un facteur d'affaiblissement de propagation. Ce modèle de canal représente le fait qu'un TU est généralement relié par un canal fort à sa plus proche SB, tandis que le canal qui le relie à l'autre SB est faible. Par simplicité, nous considérons un scénario symétrique par rapport aux deux TUs. Nous supposons, en outre, que la matrice H est inconnue aux TUs, mais parfaitement connue aux SBs. Sous ces hypothèses, le meilleur taux possible est l'information mutuelle ergodique $\mathbb{E}[I(\rho)]$ entre $\mathbf{x}$ et $\mathbf{y}$, qui s'écrit comme [42]

$$\mathbb{E}[I(\rho)] = \mathbb{E}\left[\log\det\left(\mathbf{I}_2 + \rho\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)\right] \tag{2}$$

où l'espérance est calculée par rapport à la matrice $\mathbf{H}$. Même sur ce simple exemple, une expression exacte de l'information mutuelle $\mathbb{E}[I(\rho)]$ n'est connue que dans le cas dégénéré où $\alpha = 1$ (cf. [43]). De toute évidence, α a un effet significatif sur la performance du système et on pourrait s'attendre à ce que $\mathbb{E}[I(\rho)]$ soit une fonction croissante de ce paramètre. Comme aucune expression analytique de l'information mutuelle ergodique n'est disponible, cette affirmation est difficile à vérifier. Pour contourner ce problème, on peut soit s'appuyer sur des simulations, soit considérer des régimes asymptotiques $\rho \to 0$ ou $\rho \to \infty$, pour lesquels $\mathbb{E}[I(\rho)]$ devient facilement calculable. Toutefois, les régimes asymptotiques de puissance d'émission infiniment faible ou infiniment grande n'ont pas d'application pratique pour les systèmes de communication, les simulations ne permettant pas non plus une analyse plus approfondie. Sans aucun autre outil d'évaluation de $\mathbb{E}[I(\rho)]$ disponible, l'histoire s'arrêterait à ce stade.

L'approche adoptée dans cette thèse est de considérer une autre limite asymptotique, à savoir la limite lorsque les dimensions de la matrice $\mathbf{H}$ tendent vers l'infini à la même vitesse. Sous cette hypothèse, quelque peu artificielle, on peut montrer que pour plusieurs modèles de canal, des fonctionnelles de la matrice $\mathbf{H}$ peuvent être approximées par des quantités déterministes qui ne dépendent que des propriétés statistiques de $\mathbf{H}$. Ces quantités sont souvent données sous formes concises, qui permettent alors de comprendre les principaux paramètres du système. Par ailleurs, pour des petites dimensions de $\mathbf{H}$, ces approximations asymptotiques sont souvent très bonnes, ce qui les rend intéressantes d'un point de vue pratique. Le principal outil mathématique conduisant à ce type de résultats est la théorie des grandes matrices aléatoires, qui sera largement discuté dans les sections 2.2, 2.3 et 2.4. Pour en revenir au modèle de canal (1), on peut obtenir une approximation asymptotique de $\mathbb{E}[I(\rho)]$ en utilisant le théorème 14 (*iv*) de la section 2.3, qui s'exprime comme

$$\begin{aligned}\mathbb{E}[I(\rho)] \approx\ & 2\log\left(1 + 2(1+\alpha)\rho + \sqrt{1+4(1+\alpha)\rho}\right) - 2\log(2e) \\ & + \frac{4}{1+\sqrt{1+4(1+\alpha)\rho}}.\end{aligned} \tag{3}$$

Il est naturel de se demander si une telle approximation est justifiée dans le cas de systèmes de petites dimensions comportant seulement deux émetteurs et deux récepteurs, chacun d'eux étant équipé d'une antenne. Étonnamment, la réponse est "oui", comme le montre la figure 2, qui représente $\mathbb{E}[I(\rho)]$ et son approximation asymptotique (3) en fonction de ρ pour $\alpha = 0.5$. Les deux courbes

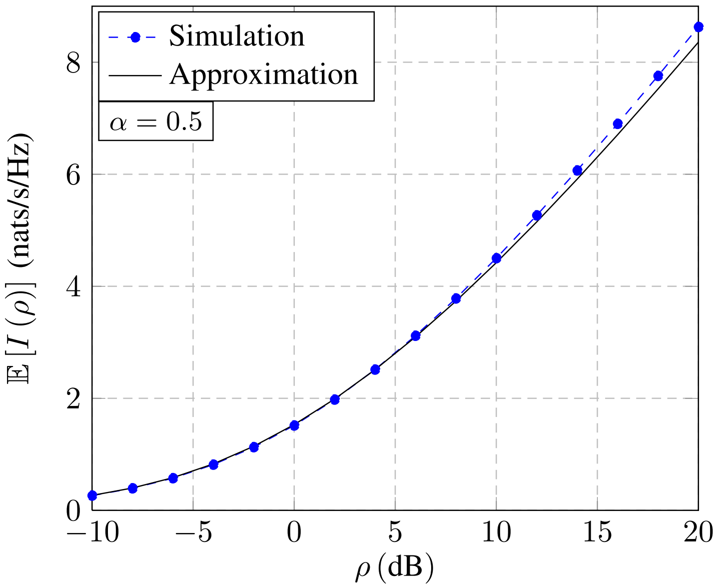

Figure 2: L'information mutuelle ergodique $\mathbb{E}[I(\rho)]$ et son approximation asymptotique (3) en fonction de ρ pour $\alpha = 0.5$.

sont quasiment superposées, même si le rapprochement devient moins précis pour les grandes valeurs de ρ. On peut aussi montrer que (3) est effectivement une fonction croissante de α.

Ceci constitue bien sûr un exemple extrême et, en fonction de ρ et de α, le rapprochement peut être moins précis. Toutefois, cet exemple suggère qu'une analyse asymptotique peut s'avérer utile pour l'étude de systèmes complexes, dont l'analyse exacte est impossible. En outre, ce type de résultats asymptotiques offre un intérêt pratique car il permet une approximation proche de la performance des systèmes de petite taille à des points de fonctionnement réalistes. Dans cette thèse, nous développons et appliquons ces approximations asymptotiques à l'étude de modèles de systèmes complexes qui intègrent (en partie) les caractéristiques décrites précédemment.

Résumé des contributions

Cette thèse est divisée en trois parties distinctes mais interdépendantes: la théorie (chapitre 2), les applications (chapitre 3) et les conclusions et perspectives (chapitre 4). La structure de la thèse et des relations entre les différentes sections est illustré par un schéma en figure 3.

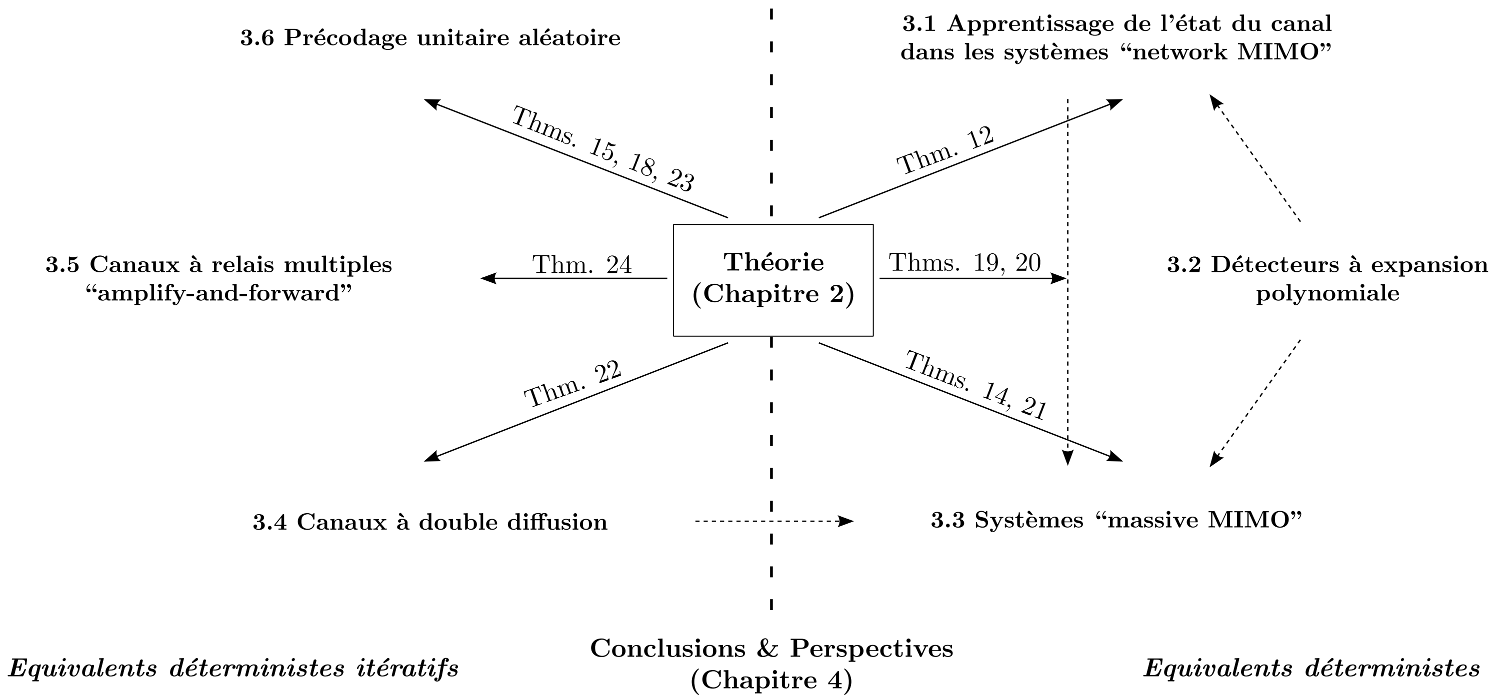

Figure 3: Schéma de la structure de la thèse.

Le chapitre 2 donne un aperçu de plusieurs outils et techniques mathématiques qui seront fréquemment utilisés dans la suite. Parmi eux, nous trouvons quelques notions de base de la théorie des probabilités, de la théorie des matrices aléatoires (RMT) et des outils d'analyse réelle et complexe. Nous allons également rappeler quelques résultats importants asymptotiques de RMT qui seront nécessaires ou approfondis dans les sections suivantes. Ensuite, nous allons introduire la notion d'équivalents déterministes en présentant certains des résultats existants, mais aussi en en établissant de nouveaux. Ces résultats seront appliqués à des problèmes pratiques dans les sections 3.1, 3.2 et 3.3. Nous allons également élaborer une nouvelle approche pour le calcul des équivalents déterministes pour certaines fonctionnelles de matrices unitaires aléatoires. Les applications pratiques de ces résultats sont présentées dans la section 3.6. Enfin, nous allons expliquer le concept d'équivalents déterministes itératifs qui a été développé dans cette thèse, avant de présenter plusieurs résultats basés sur cette notion. Plusieurs applications pratiques seront discutées dans les sections 3.4, 3.5 et 3.6.

Dans le chapitre 3, nous appliquons les résultats théoriques du chapitre 2 aux problèmes des communications sans fil. Tous ces problèmes adressent de manière différente certaines des principales caractéristiques des futurs systèmes de communication avancés, comme décrits ci-dessus.

L'apprentissage optimal de l'état du canal dans les systèmes "network MIMO"

Comme la coopération des SBs est l'une des techniques les plus prometteuses pour accroître l'efficacité spectrale des réseaux cellulaires, nous commençons dans la section 3.1 avec une étude des systèmes "network MIMO", où les SBs sont connectées par un réseau de transport. Plusieurs contraintes pratiques sont considérées, à savoir une connaissance imparfaite de l'état du canal, un "backhaul" de capacité finie et les affaiblissements de propagation arbitraires entre les SBs coopératives et les TUs. Le but de cette section est de calculer la fraction optimale du temps de cohérence du canal qui doit être utilisée pour l'apprentissage du canal afin de maximiser une borne inférieure de l'information mutuelle. Nous utilisons la RMT (Théorème 12) pour fournir une approximation de cette borne, qui est ensuite optimisée par rapport à la longueur des séquences pilotes. Notre principale contribution est la prise en compte simultanée de plusieurs contraintes pratiques des systèmes "network MIMO" tout en considérant un modèle de canal réaliste. L'utilisation de la RMT dans le contexte de ce problème d'optimisation est également une idée originale qui pourrait être appliquée à d'autres scénarios. Par exemple, la longueur optimale des séquences pilotes pour les systèmes MIMO de grande échelle (voir section 3.3) qui souffrent de "pilot contamination" pourrait être obtenue de manière similaire.

Les résultats de cette section ont été publiés dans les articles suivants :

- J. Hoydis, M. Kobayashi, M. Debbah, "Optimal channel training in uplink network MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824–2833, Juin 2011.
- J. Hoydis, M. Kobayashi, M. Debbah, "On optimal channel training for uplink network MIMO systems," *Proc. IEEE International Conference on*

Acoustics, Speech and Signal Processing (ICASSP), Prague, République Tchèque, Mai 2011, pp. 3056–3059.

Détecteurs à expansion polynomiale

Comme le nombre d'observations conjointement traitées augmente (par exemple, dans le système "network MIMO" ou les systèmes MIMO de grande échelle), la mise en œuvre de récepteurs optimaux devient un défi. Ceci fait donc apparaître un besoin de structures de récepteurs moins complexes qui est adressée dans la section 3.2. Ici, nous considérons la conception et l'analyse des récepteurs à expansion polynomiale de faible complexité, adaptés à des canaux sans fil caractérisés par un profil de variance généralisé. Dans ce modèle de canal, le vecteur de canal entre chaque émetteur et le récepteur considéré peut avoir une matrice de covariance différente. Ce type de détecteur est pertinent pour des systèmes MIMO de grande échelle ou des réseaux MIMO distribués, où le récepteur MMSE nécessiterait l'inversion d'une grande matrice, méthode coûteuse en temps de calcul. La principale difficulté de cette section réside dans le calcul des moments asymptotiques d'un type particulier de matrices aléatoires (voir les théorèmes 19 et 20). Les simulations montrent qu'un tel détecteur peut fournir une bonne performance tout en réduisant la complexité numérique.

Les résultats de cette section ont été publiés dans l'article suivant :

- J. Hoydis, M. Debbah, M. Kobayashi, "Asymptotic moments for interference mitigation in correlated fading channels," *Proc. IEEE International Symposium on Information Theory (ISIT)*, Saint Petersbourg, Russie, Août 2011, pp. 2796–2800.

Les systèmes "massive MIMO"

Dans la section 3.3, nous faisons l'analyse de la performance des systèmes MIMO de grande échelle ou "massive MIMO" qui sont caractérisés par un plus grand nombre d'antennes par SB que de TUs par cellule. Bien que "massive MIMO" soit une technique non-coopérative, elle peut théoriquement atteindre de très hautes efficacités spectrales avec un traitement du signal linéaire et simple. Pour un modèle de canal très général qui représente la connaissance imparfaite de l'état du canal, la "pilot contamination", la corrélation d'antennes et l'affaiblissement de propagation, nous dérivons des approximations asymptotiques des taux atteignables en liaisons montante et descendante avec des détecteurs et précodeurs linéaires. Notre analyse, basée sur les théorèmes 14 et 21, fournit des approximations de la performance qui sont faciles à calculer et précises pour les systèmes de dimensions finies, comme le montre les simulations effectuées. En outre, nous démontrons que, dans certains scénarios, l'utilisation du récepteur MMSE ou du précodeur RZF (dit "regularized zero-forcing") peut réduire le nombre d'antennes d'un ordre de grandeur tout en atteignant les mêmes performances qu'un filtre adapté (dit "matched filter") dans la liaison montante ou descendante.

Les résultats de cette section ont été publiés dans les articles suivants :

- J. Hoydis, S. ten Brink, M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, Jan. 2012, soumis.

- J. Hoydis, S. ten Brink, M. Debbah, "Comparison of linear precoding schemes for the massive MIMO downlink," *Proc. IEEE International Conference on Communications (ICC)*, Ottawa, Canada, Juin 2012.
- J. Hoydis, S. ten Brink, M. Debbah, "Massive MIMO: How many antennas do we need?" *Proc. IEEE Allerton Conference on Communication, Control, and Computing (ALLERTON)*, Urbana-Champaign, IL, Etas-Unis, Septembre 2011.

Dans les trois dernières sections, nous présentons des applications pratiques des équivalents déterministes itératifs (cf. section 2.4), à savoir l'analyse de la performance asymptotique (cf. 3.4) de canaux à double diffusion ("double-scattering") à accès multiples ("multiple access channel (MAC)"), le calcul d'un équivalent déterministe de l'information mutuelle d'un canal MIMO à relais multiples "amplify-and-forward" (cf. section 3.5) et l'analyse de la performance du précodage unitaire aléatoire pour des canaux quasi-statiques et pour des canaux avec évanouissement et à antennes corrélées (cf. section 3.6).

Canaux à double diffusion

Le modèle de canal à double diffusion, dit "double-scattering", peut expliquer l'existence de canaux de rang bas, sans corrélation entre les antennes de l'émetteur ou du récepteur. C'est par exemple le cas si l'émetteur et le récepteur sont entourés par des anneaux de diffuseurs locaux, dont les diamètres sont de petite taille par rapport à leur distance. Cet effet peut devenir visible dans les systèmes MIMO de grande échelle, où un grand réseau d'antennes est monté sur un bâtiment haut ou une tour, et où la plupart des TUs est entourée par des obstacles. Notre analyse asymptotique (en termes d'information mutuelle et de taux réalisables avec détection MMSE), reposant principalement sur le théorème 22, montre que l'approche par les équivalents déterministes itératifs peut faciliter l'étude de ce modèle de canal plutôt complexe. En outre, ces résultats pourraient être utilisés pour étendre l'analyse des systèmes "massive MIMO" présentés dans la section 3.3 à des modèles de canaux encore plus réalistes.

Les résultats de cette section ont été publiés dans les articles suivants :

- J. Hoydis, R. Couillet, M. Debbah, "Iterative deterministic equivalents for the capacity analysis of communication systems," *IEEE Trans. Inf. Theory*, Décembre 2011, soumis.
- J. Hoydis, R. Couillet, M. Debbah, "Asymptotic analysis of double-scattering channels," *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, Etas-Unis, Novembre 2011. **Finaliste du concours du meilleur article étudiant.**

Canaux à relais multiples "amplify-and-forward"

Comme mentionné précédemment, les techniques de relais pourraient jouer un rôle important dans les futures architectures de réseau et sont déjà utilisées dans les normes actuelles des systèmes de communication mobile [31]. Ainsi, leur compréhension théorique est aussi pertinente en pratique. La performance asymptotique du canal à relais multiples "amplify-and-forward" a été étudiée dans de nombreux ouvrages (cf. [44, 45]), qui soit ignorent les effets d'amplification du bruit à chaque relais, soit ne conduisent pas à des expressions faciles à manipuler. En s'appuyant sur les équivalents déterministes itératifs, nous sommes en mesure de trouver une expression simple et récursive de l'information mutuelle asymptotique de ce modèle de canal (théorème 24). Des simulations démontrent que les approximations asymptotiques sont valables pour un nombre d'antennes raisonnable à chaque nœud.

Les résultats de cette section ont été publiés dans l'article suivant :

- J. Hoydis, R. Couillet, M. Debbah, "Iterative deterministic equivalents for the capacity analysis of communication systems," *IEEE Trans. Inf. Theory*, Décembre 2011, soumis.

Précodage unitaire aléatoire

Les précodeurs unitaires ont suscité un intérêt majeur grâce à leur capacité à réduire la quantité de données échangées sur le chemin retour, nécessaires à la formation de faisceaux, dite "beamforming". Pour cette raison, ils sont déjà considérés dans des futures normes de réseaux mobiles [46, 47, 48]. Ainsi, l'évaluation de la performance des systèmes qui utilisent les précodeurs unitaires est obligatoire et constitue un domaine de recherche très actif [49]. Une nouvelle approche pour les modèles de matrices aléatoires impliquant des matrices unitaires via les équivalents déterministes (itératifs) nous permet de traiter des canaux quasi-statiques et des canaux corrélés avec évanouissement et d'en déduire ainsi des approximations précises de l'information mutuelle et de la somme des taux avec un détecteur MMSE. Les résultats théoriques apparentés sont résumés dans les théorèmes 15, 18 et 23.

Les résultats de cette section ont été publiés dans les articles suivants :

- R. Couillet, J. Hoydis, M. Debbah, "Random beamforming over quasi-static and fading channels: A deterministic equivalent approach," *IEEE Trans. Inf. Theory*, Novembre 2011, accepté.

- J. Hoydis, R. Couillet, M. Debbah, "Deterministic equivalents for the performance analysis of isometric random precoded systems," *Proc. IEEE International Conference on Communications (ICC)*, Kyoto, Japan, Juin 2011, pp. 1–5.

D'autres contributions

Par ailleurs, les publications suivantes ont résulté de cette thèse :

- J. Hoydis, M. Debbah, *"Random Matrix Methods for Cooperation in Small-cell Networks" dans Femtocell Networks: Deployment, PHY Techniques, and Resource Management.* Cambridge University Press, 2012, édité par : T. Q. S. Quek, G. de la Roche, I. Guvenc, M. Kountouris.

- J. Hoydis, M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mars 2011.

- J. Hoydis, R. Couillet, M. Debbah, "Les réseaux a petites cellules économes en energie," *Techniques de l'Ingénieur*, no. Reference IN 141, Août 2011.

- J. Hoydis, R. Couillet, M. Debbah., “Les réseaux a petites cellules économes en energie,” *Revue de l'Electricité et de l'Electronique*, no. 11, pp. 51–55, Décembre 2010.

- J. Hoydis, A. Müller, R. Couillet, M. Debbah, “Analysis of multicell cooperation with random user locations via deterministic equivalents,” *Proc. Workshop on Spatial Stochastic Models for Wireless Networks (SPASWIN)*, Paderborn, Germany, Mai 2012.

- A. Müller, J. Hoydis, R. Couillet, M. Debbah, “Optimal 3D cell planning: A random matrix theory approach,” *Proc. IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, Etats-Unis, Décembre 2012, soumis.

- J. Hoydis, R. Couillet, P. Piantanida, M. Debbah, “A random matrix approach to the finite blocklength regime of MIMO fading channels,” *Proc. IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, Etats-Unis, Juillet 2012, soumis.

- J. Hoydis, A. Kammoun, J. Najim, M. Debbah, “Outage performance of cooperative small-cell systems under Rician fading channels,” *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, San Francisco, CA, Etats-Unis, Juin 2011, pp. 551–555.

- J. Hoydis, M. Kobayashi, M. Debbah, “Asymptotic performance of linear receivers in network MIMO,” *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, Etats-Unis, Novembre 2010, pp. 942–948.

- J. Hoydis, M. Kobayashi, M. Debbah,“Random matrix theory for small-cell networks,” *25th Meeting of the Wireless World Research Forum*, London, Royaume-Uni, Novembre 2010.

- J. Hoydis, J. Najim, R. Couillet, M. Debbah, “Fluctuations of the mutual information in large distributed antenna systems with colored noise,” *Proc. IEEE Allerton Conference on Communication, Control, and Computing (ALLERTON)*, Urbana-Champaign, IL, Etats-Unis, Septembre 2010, pp. 240–245.

- S. Lakshminaryana, J. Hoydis, M. Debbah, M. Assaad, “Asymptotic analysis of distributed multi-cell beamforming,” *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIM RC)*, Istanbul, Turquie, Septembre 2010, pp. 2105–2110.

Conclusions et perspectives

Afin de faire face à une demande croissante de services mobiles, les systèmes avancés de communication mobile seront caractérisés par un déploiement dense des différents types de points d'accès sans fil. Celui-ci sera probablement un mélange de petites cellules à faible puissance, de femto cellules et de macro SBs équipées de centaines d'antennes intelligentes. L'atténuation des interférences et la réduction de la consommation d'énergie dans de tels réseaux sont d'une importance primordiale. Pour cette raison, les techniques COMP ainsi que les fonctionnalités de l'auto-optimisation sont non seulement souhaitables, mais aussi nécessaires. Comme les réseaux mobiles deviennent plus complexes, les méthodes requises pour l'analyse de leur performance théorique doivent évoluer. Cela implique qu'elles doivent tenir compte des caractéristiques les plus importantes de ces réseaux, à savoir le "fading", l'affaiblissement de propagation, les interférences, l'information imparfaite de l'état du canal, le "pilot contamination", la corrélation d'antennes, les liens à visibilité directe et la coopération avec échange de données limité. Dans cette thèse, nous avons développé de nouvelles méthodes basées sur la théorie des grandes matrices aléatoires qui sont capables de prendre ces caractéristiques en considération. En particulier, le concept des équivalents déterministes basé sur un système de grande taille conduit souvent à des approximations de la performance du système étonnamment simples et précises et permet de tirer des conclusions importantes sur les paramètres les plus pertinents. On peut penser à cette méthode comme un moyen pour fournir une abstraction déterministe de la couche physique qui réduit significativement la complexité du système. Grâce à cette réduction de la complexité, il est par exemple possible d'optimiser certains paramètres du système (la longueur des pilotes, les matrices de précodage, etc.), ce qui aurait été impossible autrement. Cette approche pourrait aussi s'avérer bénéfique pour l'optimisation conjointe de plusieurs couches de la pile des protocoles.

Nous avons démontré l'utilité des équivalents déterministes dans le contexte de plusieurs scénarios d'intérêt pratique, tels que l'analyse de la performance et l'optimisation des systèmes CoMP et "massive MIMO", et l'étude de canal à plusieurs relais. En outre, plusieurs nouvelles contributions au domaine de la théorie des matrices aléatoires proviennent de cette thèse. Les plus importants sont le concept des équivalents déterministes itératifs et le calcul des équivalents déterministes pour une certaine classe de fonctionnelles de matrices unitaires aléatoires.

Quelques mots de prudence sont nécessaires en ce qui concerne la nature asymptotique de nos résultats. Nous avons mis l'accent sur le fait que la plupart des équivalents déterministes fournissent des approximations très précises pour des systèmes de dimensions finies. Il est parfois impossible de faire la distinction entre des résultats asymptotiques et exacts dans le cas d'un système où l'émetteur et le récepteur sont équipés avec seulement deux antennes chacun. Toutefois, rien ne garantit que ce soit le cas pour tous les choix de paramètres possibles. En général, les approximations deviennent moins précises pour les grandes valeurs de SNR (cf. [50] qui explique cet effet pour l'information mutuelle d'un canal point à point MIMO). Parmi les indicateurs importants de la précision des approximations asymptotiques, nous trouvons les taux de convergence de certaines quantités vers leurs limites asymptotiques. Par exemple, tandis que la variance de l'information mutuelle normalisée de certains canaux MIMO $N \times N$ diminue à la vitesse de $1/N^2$ [51], la variance du SINR avec un détecteur MMSE diminue à la vitesse de $1/N$ [52]. Ainsi, la précision des équivalents déterministes ne dépend pas uniquement de la taille du système, mais davantage de la fonctionelle de la variable aléatoire en considération.

Dans ce qui suit, nous allons exposer quelques sujets intéressants pour des travaux futurs :

Topologies aléatoires du réseau

Tout au long du document, nous avons implicitement supposé que les émetteurs et les récepteurs sont situés à des positions fixes et connues. Cependant, cette hypothèse est rarement rencontrée dans la pratique. Les TUs se déplacent naturellement et les positions des points d'accès déployés par l'utilisateur, tels que les femto-cellules, ne sont ni connues, ni vouée à rester statiques. Ainsi, la topologie du réseau est un paramètre aléatoire supplémentaire des systèmes de communication mobile qui doit aussi être pris en considération. Une approche récente pour s'attaquer à ce problème est la géométrie stochastique [53]. Dans ce cadre, les TUs et les points d'accès sont considérés comme des processus ponctuels aléatoires dans l'espace. Ces processus ponctuels sont généralement caractérisés par leur densité et leur tendance à former des "clusters", c'est à dire, des points d'attraction à densité plus élevée. Bien que cette technique ait conduit à de nombreux résultats intéressants sur la performance moyenne de réseaux à plusieurs couches [54, 55], dits "multi-tier networks", ce n'est que très récemment que la coopération entre les points d'accès a été considérée [56, 57]. Cependant, l'examen de la coopération entre les nœuds introduit des interdépendances complexes entre les points du processus de point sous-jacent qui sont difficilement résolubles par les techniques existantes. Une combinaison de la RMT et de la géométrie stochastique pourrait surmonter ce problème. La RMT pourrait supprimer le caractère aléatoire causé par le "fading" dans de tels réseaux, tandis que la géométrie stochastique pourrait moyenner toutes les positions possibles prises par les utilisateurs. Ceci est particulièrement important si l'on veut maximiser le rendement moyen du système par rapport à certains paramètres, tels que les positions optimales des SBs. Les premiers résultats utilisant ces concepts sont établis dans les publications suivantes :

- J. Hoydis, A. Müller, R. Couillet, M. Debbah, "Analysis of multicell cooperation with random user locations via deterministic equivalents," *Proc. Workshop on Spatial Stochastic Models for Wireless Networks (SPASWIN)*, Paderborn, Germany, Mai 2012.
- A. Müller, J. Hoydis, R. Couillet, M. Debbah, "Optimal 3D cell planning: A random matrix theory approach," *Proc. IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, Etats-Unis, Décembre 2012, soumis.

Codes de longueur finie

L'information mutuelle et les taux réalisables considérés dans cette thèse sont tous basés sur l'hypothèse cruciale de mots de code de longueur infinie. Ainsi, bien que le temps de cohérence du canal soit limité, les messages sont supposés être codés et transmis sur un nombre infini de blocs de cohérence. Dans la pratique, cette hypothèse impliquerait un délai infini. Récemment, les publications [58, 59] ont suscité un grand intérêt pour les limites ultimes de la performance de codes de longueur finie. Toutefois, les limites précises et explicites de la probabilité d'erreur pour une longueur de bloc donnée sont en général difficiles à établir pour des canaux MIMO. L'application de la RMT à ce domaine de la théorie de l'information pourrait conduire à de significatives simplifications. Nous avons fourni quelques premiers résultats sur ce sujet dans :

- J. Hoydis, R. Couillet, P. Piantanida, M. Debbah, "A random matrix approach to the finite blocklength regime of MIMO fading channels," *Proc. IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, Etats-Unis, Juillet 2012, soumis.

Utilisation intelligente des antennes "excédentaires"

Dans la section 3.3, nous avons analysé la performance des systèmes MIMO de grande échelle ("massive MIMO") où le SBs sont équipées d'un nombre d'antennes beaucoup plus important que le nombre de TUs par cellule. Toutefois, ce type de réseaux d'antennes de grande taille est plus ou moins une approche de type force brute pour contrer l'interférence intercellulaire et réduire les puissances d'émission. Cependant, il serait intéressant d'étudier d'autres façons d'utiliser plus intelligemment les antennes supplémentaires ou "excédentaires" dans un réseau. Par exemple, une macro SB pourrait sacrifier certaines de ses antennes pour l'annulation d'interférence vers des couches inférieures du réseau, comme les femto-cellules. En supposant un protocole duplex par séparation temporelle inversée, dit "time-division duplex (TDD)", entre des macro- et femto-cellules [60], la connaissance de l'état du canal à la macro-SB pourrait être obtenue "gratuitement". Dans ce contexte, il serait intéressant d'étudier si la perte de performance dans la macro-cellule est suffisamment compensée par les améliorations des taux résultants dans le réseau femto. Il serait également intéressant de se pencher sur la question du placement optimal des antennes: pour une région donnée, avec une distribution des TUs donnée, quelle est la meilleure façon de déployer N antennes pour couvrir cette région ?

Les canaux variant dans le temps

Les modèles de canaux que nous avons utilisés dans cette thèse ne présentent pas de corrélation dans le temps. Autrement dit, les réalisations de canaux à deux instants différents sont des variables aléatoires indépendantes. En exploitant les résultats récents sur les matrices aléatoires variant dans le temps [61, 62], il serait intéressant d'étudier comment la performance du réseau varie au fil du temps pour un processus stochastique donné. Toutefois, la recherche à ce sujet en est encore à ses balbutiements et il faudra encore un certain temps et un certain effort jusqu'à ce que ces méthodes soient suffisamment bien comprises pour envisager des modèles de matrices pertinents pour les systèmes de communication.

Chapter 1

Introduction

1.1 Mobile communications: Future challenges

During the last years, mobile data traffic has skyrocketed, paralleling the development of the wired Internet traffic at the beginning of this millennium. Since 2008, the total global mobile data volume has more than doubled every year and is expected to grow at a similar rate in the future [1]. The main drivers of this massive growth are smartphones, tablet PCs, and laptops, whose prominent usage is about data rather than voice. Soon, there will be as many wireless devices as humans on earth. As a result of this development, current networks reach their capacity limits, especially in highly populated metropolitan areas. Congestion problems arise in the wireless and the backhaul networks alike.

At the same time, there is a growing concern about the possible effects of information and communication technology (ICT) on global carbon emissions [2, 18]. Although ICT's contribution to the global emissions is and will remain a rather small percentage of the global figures (with 1.25 % in 2002 and around 2.5 % in 2020), the general trend of a 10 % yearly increase in ICT-related carbon emissions is alarming. This means that, despite significant progresses in energy-efficient technologies, the growth in data traffic will outpace our ability to reduce or even maintain the overall energy consumption and related emissions. Thus, more network capacity on the one hand and less energy consumption on the other are two seemingly contradictory future requirements on ICT. This begs the question how mobile operators can satisfy the future traffic demands, both economically and ecologically.

Let us now have a look at several different ways how the capacity of wireless networks could be generally increased:

More spectrum

Essentially all commercial mobile communication systems are operated on frequency bands in the range from 300 MHz–3 GHz. This is due to the favorable radio propagation conditions at these frequencies. As a consequence, this part of the spectrum is almost entirely allocated and new spectral resources can only be found at higher frequencies. For this reason, there is a growing interest in millimeter-wave (mmWave) communication systems [3] which exploit the spectrum from 3–300 GHz. Today, mainly short-range indoor and fixed-wireless communication systems are operated in these bands. Owing to high reflection and penetration losses, the communication at very short wavelength is more or less limited to line-of-sight (LOS) conditions. Providing indoor coverage from outdoor base stations (BSs) seems impossible. Nevertheless, there are current research activities which assess the feasibility of mmWaves for mobile communications [4]. However, this research is still in its infancy.

Better coding and modulation schemes

With low density parity-check (LDPC) [5] and Turbo [6] codes, orthogonal frequency-division multiplexing (OFDM), link adaptation, and hybrid automatic repeat request (HARQ), modern mobile communication systems like LTE [7] or WiMAX [8] operate at spectral efficiencies close to the theoretical limits. Thus, revolutionary breakthroughs in coding or modulation are not to be expected in the near future. Current research targets, for example, non-orthogonal modulation schemes which do not require a cyclic prefix, such as Isotropic Orthogonal Transform Algorithm (IOTA)-OFDM [9], or techniques like Vandermonde frequency-division multiplexing (VFDM) [10] which exploit the cyclic prefix of OFDM to allow for interference-free underlay networks. Also rate-less or Fountain codes [11], which generalize in some sense the concept of HARQ, as well as Polar codes [12] promise further improvements in spectral efficiency. However, despite these efforts, advances in coding or modulation are unlikely to be the main capacity driver of future mobile communication systems.

More antennas

Since the seminal papers [13, 14] in the mid–1990s, multiple-input multiple-output (MIMO) wireless communications have attracted enormous interest from researchers and industry alike. By now, the benefits of multiple antennas at the transmitters and/or receivers are well understood [15]. MIMO techniques can provide power gains, improve the link reliability, and increase the throughput by multiplexing several independent data stream on the same time-frequency resource. Simple MIMO schemes are already an integral part of modern mobile communication standards [7, 8] which support today up to eight antennas at the BSs and the user terminals (UTs). Large-scale MIMO systems where BSs are equipped with hundreds of antennas are the subject of numerous ongoing research projects [16, 17, 18]. In theory, large-scale or "massive" MIMO can significantly reduce transmit powers while achieving high spectral efficiencies. Although not directly MIMO techniques, one needs to consider also higher-order cell-sectorization [19] and 3D-beamforming [20, 21] as effective means to increase the spectral efficiency. These techniques aim at increasing the spatial reuse by focusing the transmitted energy towards the intended UTs and by reducing interference to other UTs at the same time.

More cells

It is well known that cell-size shrinking is the simplest and most effective way to increase wireless throughput [22, 23]. Physics tells us that bringing a radio transmitter and receiver closer together reduces the necessary transmit power to overcome path loss and other phenomena, such as fading and noise. Moreover, the area throughput increases theoretically linearly with the cell density.

However, the deployment of traditional BSs, such as macro and micro cells, requires huge capital and operational expenditures for cell-site acquisition, network planning, backhaul provisioning, operation, and maintenance. Hence there is a growing popularity of femto cells [24], i.e., user-deployed home-BSs utilizing the existing Internet connection as backhaul. Femto cells allow to offload traffic from the macro cells and to provide high-capacity indoor coverage at minimal cost for the network operator. Currently, also "small cells", i.e., self-organizing operator-deployed outdoor/indoor femto cells, receive considerable interest from academia and industry as a promising means to provide localized high-capacity coverage at low energy-consumption and cost [25, 26].

Cooperation and coordination

Cellular networks are first and foremost limited by intercell interference. If the BSs were allowed to cooperate or to coordinate, this interference could be either exploited or reduced [27]. One distinguishes generally between cooperation (also network MIMO or multi-cell processing), i.e., multiple BSs are connected together via backhaul links and jointly process their data, and coordination, i.e., clusters of BSs jointly decide on precoding/decoding strategies but generally do not share user data. Both techniques, commonly referred to as coordinated multi-point (CoMP), can improve the coverage and throughput, especially for cell-edge UTs, without the need to deploy new cell sites or additional antennas. However, many technical challenges need to be overcome before CoMP can be successfully introduced in practice [28]. Moreover, the effective gains of CoMP are less promising once the overhead for the acquisition of channel state information (CSI) is taken into account [29]. Other interesting techniques under current research are interference alignment schemes [30], relaying concepts [31], and wireless network clouds [32] where BSs are replaced by fiber-connected remote radio heads and the processing is carried out on centralized server-farms.

Cognitive radio

"Most bands in most places are underused most of the time" [33, 34]. This observation stimulated the idea of cognitive radios [35] which are allowed to use parts of the licensed spectrum, given that they can reliably sense if it is currently used or not. In this context, one often speaks about "spectrum holes" in either time, frequency, or space which cognitive radios try to exploit. Despite the heavy research on this topic during the last decade, cognitive radios have not yet been successful in practice. This is mainly due to a lack of reliable sensing algorithms, protocols, hardware, and regulatory constraints. Thus, it is unlikely that cognitive radio will play a major role in next generation mobile communication systems. However, the concept of more "intelligent" or "flexible" radios, which moves away from the classical centralized network architecture to self-organizing networks with intelligent decision making at the nodes, is a promising idea [36].

Other techniques

There are of course many other techniques which provide interesting opportunities for spectral-efficiency improvements. Among them are for example full-duplex transceivers [37] which could theoretically double the capacity of current networks, a consequent exploitation of electromagnetic polarization [38, 39] which could lead in principle to a threefold capacity increase, and cross-layer approaches like joint source-channel decoding [40] which exploits redundancy and side information at different protocol layers.

From the discussion above, it becomes clear that advanced mobile communications systems are likely to consist of a dense deployment of different types of wireless access points (macro/micro/femto/small cells) with different characteristics (transmit powers, number of antennas, open/closed access, user/operator deployed). Additionally, CoMP techniques will be a desirable feature to mitigate the increasing interference in such networks. In order to assess which of the possible techniques presented above is the most appropriate with respect to a given goal, it is necessary to provide a fundamental theoretical performance analysis. However, channel models and tools which were developed for the analysis of simple point-to-point links often fail to provide meaningful insights for large, increasingly dense, heterogeneous networks. New theoretical tools for their analysis are needed and the development of these tools is the main goal of this thesis. In what follows, we will focus almost exclusively on the characterization of theoretical performance limits. Energy efficiency, although an important parameter, will not be considered in this work.

1.2 Why go to large dimensions?

The main difficulty of a meaningful theoretical performance analysis is (i) to find the simplest model which sufficiently reflects the main characteristics of the system under study and (ii) to choose the right mathematical tool which allows one to draw general conclusions and, possibly, to provide engineering guidelines. For example, it is hard to justify why conclusions which are drawn for a channel model without path loss and interference should hold for more complex scenarios where both aspects are taken into account. Some of the most important characteristics of the wireless communication systems considered in this thesis are:

- **Fading**: Fading describes the random fluctuations of the received signal strength over time and frequency and is a main characteristic of all wireless channels. One generally distinguishes between small-scale and large-scale fading. The former is caused by small movements of the receiver, the transmitter, or the scatterers while the latter is due to shadowing by large obstacles, such as buildings or walls.

- **Path loss**: The average received signal strength is mainly determined by the distance-dependent path loss between a transmitter and a receiver. How this path loss is calculated depends on the environment (e.g., urban or rural) and the radio propagation conditions (e.g., precipitation). The effects of path loss are especially important once CoMP techniques are considered. This is because the benefit of these techniques depends on the positions of the UTs within a cell.

- **Interference**: Mobile networks are limited by interfering transmissions from other BSs and UTs. Consequently, this aspect cannot be ignored

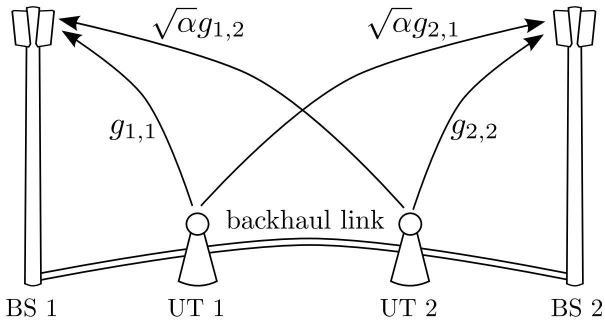

Figure 1.1: Schematic illustration of the uplink system model: Both BSs jointly decode the messages from both UTs.

and must be explicitly taken into account.

- **Imperfect CSI** : The assumption of perfect CSI is convenient for the theoretical analysis but often leads to overoptimistic results. Thus, the acquisition of CSI must be modeled and the necessary resources must be accounted for. In particular, in multi-cell systems, the reuse of pilot sequences in adjacent cells can lead to an additional source of interference, often referred to as pilot contamination [41].

- **Backhaul capacity**: If CoMP techniques are analyzed, it is necessary to be specific about the amount of data which needs to be exchanged between the cooperating BSs. In some cases, the backhaul capacity might be a limiting factor which cannot be overlooked.

- **Line-of-sight channels and antenna correlation**: The denser a network, the more likely it is that a transmitter and a receiver are under LOS conditions. Hence, the wireless channel between them is composed of a deterministic LOS component superimposed with random channel fluctuations due to fading. Moreover, the assumption of a rich scattering environment and uncorrelated antennas might fail to hold once large-scale MIMO systems are considered. Both LOS channels and antenna correlation can have a significant impact on the system performance.

In order to provide an idea why already some of these aspects taken alone are difficult problems for a theoretical analysis, let us consider the following simple example which focuses exclusively on the effects of path loss. A schematic system model is sketched in Fig. 1.1. Two BSs, connected via backhaul links of infinite capacity, cooperate to jointly decode the messages from two UTs. Both the BSs and the UTs are equipped with a single antenna. The vector $\mathbf{y} = [y_1, y_2]^\mathsf{T} \in \mathbb{C}^2$ of the received base-band signals at the two BSs at a given time reads

$$\begin{aligned}\mathbf{y} &= \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{n} \\ &= \sqrt{\rho}\begin{pmatrix} g_{1,1} & \sqrt{\alpha}\, g_{1,2} \\ \sqrt{\alpha}\, g_{2,1} & g_{2,2}\end{pmatrix}\begin{pmatrix} x_1 \\ x_2\end{pmatrix} + \begin{pmatrix} n_1 \\ n_2\end{pmatrix} \end{aligned} \tag{1.1}$$

where $g_{i,j} \sim \mathcal{CN}(0,1)$ is the fast-fading channel coefficient between UT j and BS i, $x_j \sim \mathcal{CN}(0,1)$ is the transmitted signal of UT j, $n_i \sim \mathcal{CN}(0,1)$ is thermal noise at BS i, ρ is the transmit power per UT, and $\alpha \in [0,1]$ is a path loss factor. This channel model accounts for the fact that a UT has generally a strong channel to its closest BS while it has a weaker channel to the other BS. For simplicity, we consider a scenario which is symmetric with respect to both UTs. We further suppose that $\mathbf{H}$ is unknown to the UTs but perfectly known to the BSs. Under these assumptions, the best possible achievable rate is the ergodic mutual information $\mathbb{E}\left[I(\rho)\right]$ between $\mathbf{x}$ and $\mathbf{y}$, given as [42]

$$\mathbb{E}\left[I(\rho)\right] = \mathbb{E}\left[\log\det\left(\mathbf{I}_2 + \rho\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)\right] \tag{1.2}$$

where the expectation is with respect to the matrix $\mathbf{H}$. Even in this simple setting, an exact expression of $\mathbb{E}\left[I(\rho)\right]$ is only known in the degenerate case $\alpha = 1$ (see e.g., [43]). Obviously, α has a significant effect on the system performance and one would expect that $\mathbb{E}\left[I(\rho)\right]$ is an increasing function of α. Since no analytic expression of the ergodic mutual information is available, this claim is difficult to verify. To circumvent this problem, one could either rely on simulations or consider the asymptotic regimes $\rho \to \infty$, $\rho \to 0$, where $\mathbb{E}\left[I(\rho)\right]$ has a tractable expression. However, the asymptotic regimes of infinitely low or high transmit powers are generally not the relevant operating points of communication systems and simulations do not allow for further analysis. The story would normally end here if no other tools to evaluate $\mathbb{E}\left[I(\rho)\right]$ were available.

The approach taken in this thesis is to consider another asymptotic limit, namely the limit when the dimensions of $\mathbf{H}$ grow infinitely large at the same speed. Under this somewhat artificial assumption, one can show for many channel models of interest—much more complicated than the one considered in this example—that functionals of the random matrix $\mathbf{H}$ can be well approximated by deterministic quantities which only depend on the statistical properties of $\mathbf{H}$. These quantities are often given in concise form and allow one to draw insight about the most important system parameters. Moreover, the asymptotic approximations are often very tight for small dimensions of $\mathbf{H}$ and, therefore, of practical value. The main mathematical tool to derive these kinds of results is the theory of large random matrices which will be discussed in detail in Sections 2.2, 2.3, and 2.4. Coming back to the channel model (1.1), one can derive from Theorem 14 (*iv*) in Section 2.3 a large-system approximation of $\mathbb{E}\left[I(\rho)\right]$ which is given as

$$\begin{aligned}\mathbb{E}\left[I(\rho)\right] \approx{}& 2\log\left(1 + 2(1+\alpha)\rho + \sqrt{1+4(1+\alpha)\rho}\right) - 2\log(2e) \\ &+ \frac{4}{1+\sqrt{1+4(1+\alpha)\rho}}.\end{aligned} \tag{1.3}$$

It is natural to ask whether such an approximation is justified for small systems with only two single-antenna transmitters and receivers. Surprisingly, the answer is yes! This can be seen from Fig 1.2 which shows $\mathbb{E}\left[I(\rho)\right]$ and its large-system approximation by (1.3) as a function of ρ for $\alpha = 0.5$. The match between both curves is almost perfect, although the approximation becomes less accurate for large ρ. One can also show that (1.3) is indeed an increasing function of α.

This is of course a very extreme example and, depending on ρ and α, the approximation can be less accurate. However, this example suggests that a large-system analysis is useful for the study of complex systems which could not

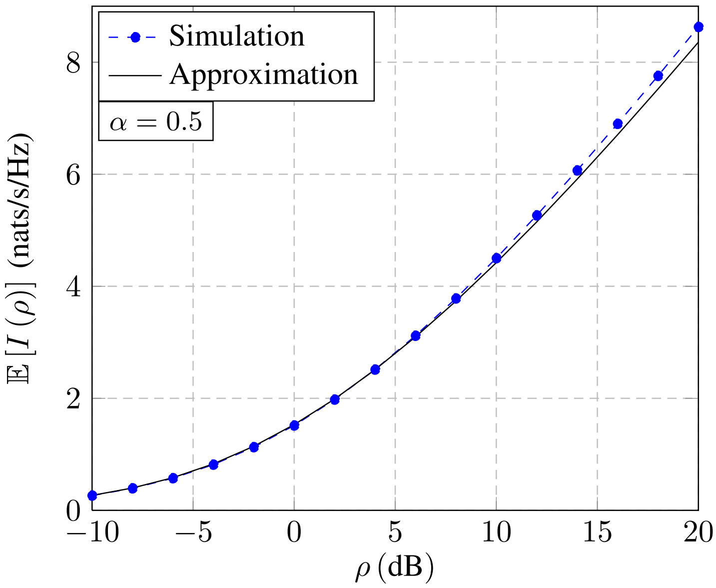

Figure 1.2: Ergodic mutual information $\mathbb{E}[I(\rho)]$ and its large system approximation by (1.3) versus transmit SNR ρ for $\alpha = 0.5$.

have been treated by exact analysis. Moreover, these types of asymptotic results are of practical interest as they provide tight performance approximations for small system dimensions at realistic operating points. This thesis develops and applies these asymptotic approximations to the study of involved system models which integrate (some of) the characteristic features detailed above.

1.3 Outline and contributions

This thesis is divided into three separate but interdependent parts: *Theory* (Chapter 2), *Applications* (Chapter 3), and *Conclusions & Outlook* (Chapter 4). A schematic diagram of the thesis' structure and the relations between different sections is shown in Figure 1.3 (see p. 10).

In Chapter 2, we will first provide an overview of several mathematical tools and techniques which will be of repeated use. These are in particular some basic notions of probability theory, random matrix theory (RMT), and the Stieltjes transform. We will also recall some important asymptotic results of RMT which will be either needed or extended in the following sections. Then, we will introduce so called "deterministic equivalents" and present some existing as well as novel results. These will be applied to practical problems in Sections 3.1, 3.2, and 3.3. We will also elaborate on a novel approach to the derivation of deterministic equivalents for certain functionals of random unitary matrices. Practical applications of these results are presented in Section 3.6. Finally, we will explain the concept of "iterative deterministic equivalents" which has been developed in this thesis and present several related results. Applications will we discussed in Sections 3.4, 3.5, and 3.6.

Chapter 3, is dedicated to applications of the theoretical results of Chapter 2 to problems in wireless communications. All of these problems address in different ways some of the main characteristics of advanced communication systems outlined in Section 1.2:

As BS-cooperation is one of the most promising techniques to increase the spectral efficiency of cellular networks, we begin in Section 3.1 with a study of uplink network MIMO systems under several practical constraints. These are imperfect CSI, limited backhaul capacity, and arbitrary path losses between the cooperative BSs and the UTs. The aim of this section is to derive the optimal fraction of the channel coherence time which should be used for channel training in order to maximize a lower bound on the mutual information. We use RMT (Theorem 12) to provide an approximation of this bound which is then optimized with respect to the length of the pilot sequences. Our main contribution is the simultaneous consideration of several practical constraints of network MIMO systems together with a realistic channel model. The use of RMT in the context of this optimization problem is also an original idea which could be applied to other scenarios. For example, the optimal amount of channel training for large-scale MIMO systems (Section 3.3) which suffer from pilot contamination could be obtained in a similar manner.

As the number of jointly processed observations grows (e.g., in network MIMO or large-scale MIMO systems), the implementation of optimal receivers becomes a challenge. Thus, there is a need for less complex receiver structures which is addressed in Section 3.2. Here, we consider the design and analysis of low-complexity polynomial expansion receivers, suitable for wireless channels with a generalized variance profile structure, i.e., the channel vector between each transmitter and the receiver is allowed to have a different covariance matrix. This type of detector is relevant to large-scale MIMO or distributed antenna systems where the minimum-mean-square-error (MMSE) detector would require the computationally expensive inversion of a large matrix. At the heart of this section is the derivation of the asymptotic moments of a particular type of random matrices (see Theorem 19, 20). Simulations show that such a detector can provide a good performance with reduced implementation complexity.

Section 3.3 is about the performance analysis of large-scale MIMO or "massive MIMO" systems which are characterized by a much larger number of BS-antennas than UTs per cell. Although massive MIMO is a non-cooperative technique, it can theoretically provide very high spectral efficiencies with simple linear signal processing. For a very general channel model which accounts for imperfect CSI, pilot contamination, antenna correlation, and path loss, we derive asymptotic approximations of achievable up- and downlink rates with linear detectors and precoders. Our analysis is fundamentally based on Theorems 14 and 21 and provides easy computable performance approximations which are shown to be accurate for finite system dimensions. Moreover, we demonstrate that, in certain scenarios, the use of MMSE detection or regularized zero-forcing (RZF) can reduce the number of antennas by one order of magnitude to achieve the same performance as a matched filter (MF) in the uplink or eigen beamforming (BF) in the downlink.

In the last three sections, we present practical applications of iterative deterministic equivalents as detailed in Section 2.4. These are the asymptotic performance analysis of double-scattering multiple access channels (MACs) in

Section 3.4, the derivation of a deterministic equivalent of the mutual information of the K-hop amplify-and-forward (AF) MIMO relay channel in Section 3.5, and the performance analysis of random unitary beamforming over correlated fading channels in Section 3.6:

The double-scattering channel model can explain the existence of low-rank channels without receive and transmit antenna correlation. This is for example the case if the transmitter and the receiver are surrounded by rings of local scatterers whose diameters are small compared to their distance. This effect might become visible in large-scale MIMO systems where a large antenna array is mounted on a high building or tower while most of the UTs are located in an irregular clutter environment. Our asymptotic analysis (in terms of mutual information and achievable rates with MMSE detection) is mainly based on Theorem 22 and shows that the iterative deterministic equivalent approach can facilitate the study of this rather involved channel model. These results could be further used to extend the analysis of massive MIMO systems in Section 3.3 to even more realistic channel models.

As mentioned earlier, relaying techniques might play an important role in future network architectures and are already considered in current cellular standards [31]. Thus, their theoretical understanding is also of practical relevance. The asymptotic performance of the AF multi-hop relay channel has been investigated in numerous works (see e.g., [44, 45]). However, these ignore either the effects of noise amplification at each relay stage or do not lead to tractable expressions. Based on the iterative deterministic equivalent approach, we are able to find a simple, recursive expression of the asymptotic mutual information of this channel model (Theorem 24). Simulations demonstrate that the asymptotic results provide valid approximations for reasonable numbers of antennas at each node.

The last practical application we consider are unitary precoders which have gained significant interest in wireless communications as limited feedback beamforming solutions for future wireless standards [46, 47, 48]). Thus, the performance evaluation of unitary precoded systems is compulsory and a field of active research [49]. A novel approach to random matrix models involving unitary matrices via (iterative) deterministic equivalents allows us to treat both quasi-static and correlated fading channels and to derive tight approximations of the mutual information and the MMSE sum rate. The related theoretical results are summarized in Theorems 15, 18, and 23.

The thesis is concluded in Chapter 4 which summarizes some of the main results and provides an outlook to future work. This comprises in particular the application of RMT to the error analysis of MIMO block-fading channels in the finite blocklength regime and the combination of RMT and stochastic geometry for the study of cooperative multicell systems with random user locations.

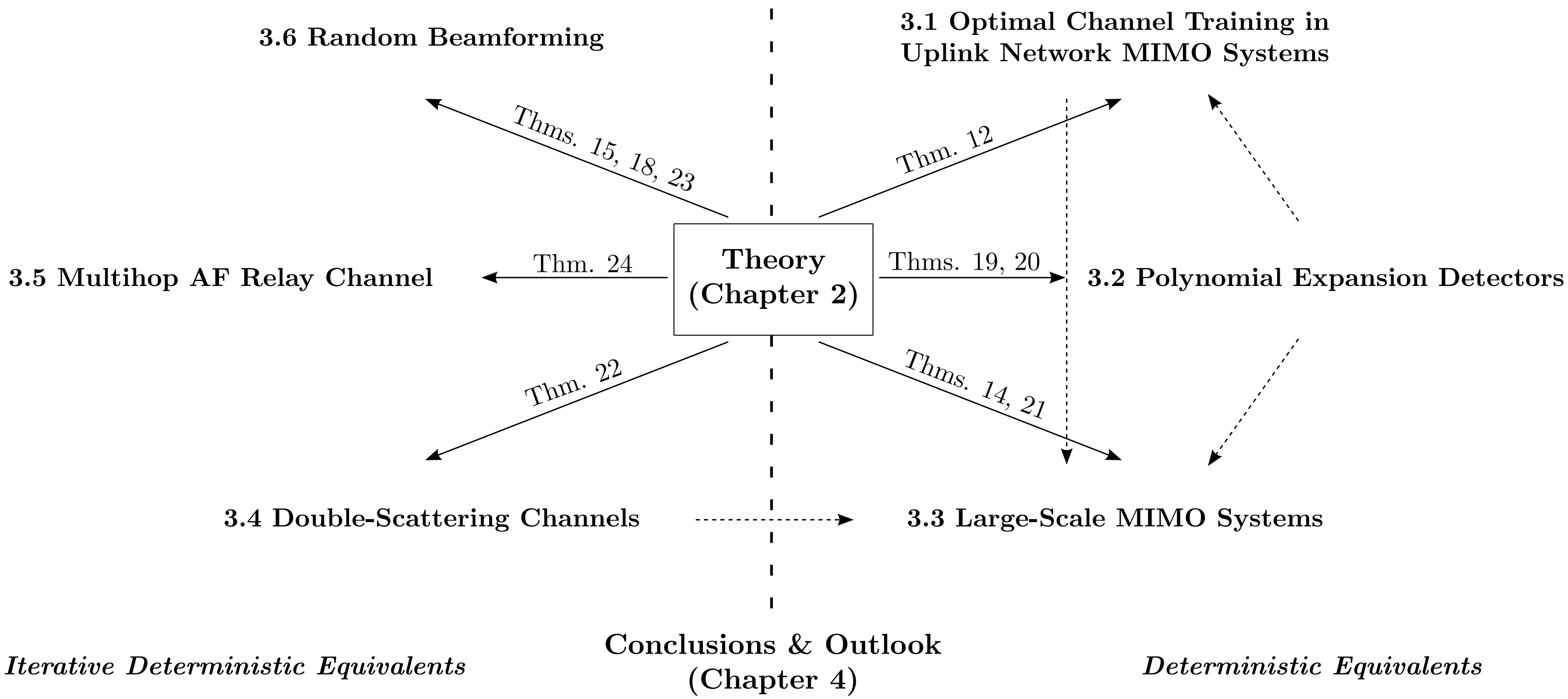

Figure 1.3: Schematic structure of the thesis.

1.4 Publications

The following publications have been produced in the course of this thesis:

Book chapters

[1] J. Hoydis and M. Debbah, *"Random Matrix Methods for Cooperation in Small-cell Networks" in Femtocell Networks: Deployment, PHY Techniques, and Resource Management.* Cambridge University Press, 2012, edited by: T. Q. S. Quek, G. de la Roche, I. Guvenc, and M. Kountouris.

Journal papers

[1] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, Jan. 2012, submitted.

[2] S. Lakshminaryana, J. Hoydis, M. Debbah, and M. Assaad, "Asymptotic analysis of distributed multi-cell beamforming," *IEEE Trans. Signal Process.*, 2012, to be submitted.

[3] J. Hoydis, R. Couillet, and M. Debbah, "Iterative deterministic equivalents for the capacity analysis of communication systems," *IEEE Trans. Inf. Theory*, Dec. 2011, submitted.

[4] R. Couillet, J. Hoydis, and M. Debbah, "Random beamforming over quasi-static and fading channels: A deterministic equivalent approach," *IEEE Trans. Inf. Theory*, Nov. 2011, accepted. [Online]. Available: http://arxiv.org/abs/1011.3717

[5] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal channel training in uplink network MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824–2833, Jun. 2011.

[6] ——, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[7] J. Hoydis, R. Couillet, and M. Debbah, "Les réseaux a petites cellules économes en energie," *Techniques de l'Ingénieur*, no. Reference IN 141, Aug. 2011.

[8] J. Hoydis, R. Couillet, and M. Debbah., "Les réseaux a petites cellules économes en energie," *Revue de l'Electricité et de l'Electronique*, no. 11, pp. 51–55, Dec. 2010.

[9] J. Hoydis and M. Debbah, "Green, cost-effective, flexible, small-cell networks," *IEEE ComSoc MMTC E-Letter Special Issue on "Multimedia Over Femto Cell"*, vol. 5, no. 5, pp. 23–26, Sep. 2010. [Online]. Available: http://committees.comsoc.org/mmc/e-news/E-Letter-September10.pdf

Conference papers

[1] J. Hoydis, T. Wild, S. ten Brink, and M. Debbah, "On coherent combining of distributed observations," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2012, invited paper.

[2] J. Hoydis, R. Couillet, P. Piantanida, and M. Debbah, "A random matrix approach to the finite blocklength regime of MIMO fading channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012, submitted.

[3] J. Hoydis, A. Müller, R. Couillet, and M. Debbah, "Analysis of multi-cell cooperation with random user locations via deterministic equivalents," in *Proc. Workshop on Spatial Stochastic Models for Wireless Networks (SPASWIN)*, Paderborn, Germany, May 2012.

[4] A. Müller, J. Hoydis, R. Couillet, and M. Debbah, "Optimal 3D cell planning: A random matrix approach," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, submitted.

[5] J. Hoydis, S. ten Brink, and M. Debbah, "Comparison of linear precoding schemes for the massive MIMO downlink," in *Proc. IEEE International Conference on Communications (ICC)*, Ottawa, Canada, Jun. 2012.

[6] J. Hoydis, R. Couillet, and M. Debbah, "Asymptotic analysis of double-scattering channels," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2011, **Finalist of the student paper contest**.

[7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?" in *Proc. IEEE Allerton Conference on Communication, Control, and Computing (ALLERTON)*, Urbana-Champaign, IL, USA, Sep. 2011.

[8] J. Hoydis, M. Debbah, and M. Kobayashi, "Asymptotic moments for interference mitigation in correlated fading channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2796–2800.

[9] J. Hoydis, A. Kammoun, J. Najim, and M. Debbah, "Outage performance of cooperative small-cell systems under Rician fading channels," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, San Francisco, CA, USA, Jun. 2011, pp. 551–555.

[10] J. Hoydis, R. Couillet, and M. Debbah, "Deterministic equivalents for the performance analysis of isometric random precoded systems," in *Proc. IEEE International Conference on Communications (ICC)*, Kyoto, Japan, Jun. 2011, pp. 1–5.

[11] J. Hoydis, M. Kobayashi, and M. Debbah, "On optimal channel training for uplink network MIMO systems," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3056–3059.

[12] ——, “Asymptotic performance of linear receivers in network MIMO,” in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2010, pp. 942–948, invited paper.

[13] ——, “Random matrix theory for small-cell networks,” in *25th Meeting of the Wireless World Research Forum*, London, UK, Nov. 2010. [Online]. Available: http://hal-supelec.archives-ouvertes.fr/hal-00554958/

[14] J. Hoydis, J. Najim, R. Couillet, and M. Debbah, “Fluctuations of the mutual information in large distributed antenna systems with colored noise,” in *Proc. IEEE Allerton Conference on Communication, Control, and Computing (ALLERTON)*, Urbana-Champaign, IL, USA, Sep. 2010, pp. 240–245.

[15] S. Lakshminaryana, J. Hoydis, M. Debbah, and M. Assaad, “Asymptotic analysis of distributed multi-cell beamforming,” in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Istanbul, Turkey, Sep. 2010, pp. 2105–2110.

Chapter 2

Theory

In this chapter, we will first provide some necessary background on probability and random matrix theory. This includes several standard inequalities, convergence types, the continuous mapping, Fubini, Vitali, and dominated convergence theorems. We will also discuss some trace lemmas, results on unitary matrices, the Stieltjes transform and its properties. Then, we will introduce and motivate the concept of deterministic equivalents and present some known and novel results. The chapter is concluded with a section on iterative deterministic equivalents which extend deterministic equivalents to a broader class of random matrix models. The theoretical results presented in this chapter find several practical applications which are detailed in Chapter 3.

2.1 Some notions of probability theory

Consider a probability space $(\Omega, \mathcal{F}, P)$ with sample space Ω, a σ-field $\mathcal{F}$ on Ω and probability measure P on $\mathcal{F}$. Denote $\omega \in \Omega$ a sample point of Ω. Let $(\mathcal{R}, \mathcal{G})$ be a measurable space. A random variable $X = X(\omega)$ is a map $X : \Omega \mapsto \mathcal{R}$. As we will deal with random matrices and associated functionals in this thesis, the observation space $\mathcal{R}$ is typically either the set of real $\mathbb{R}^{N\times K}$ or complex numbers $\mathbb{C}^{N\times K}$ and $\mathcal{G}$ are the Borel sets of $\mathcal{R}$. If $X \in \mathbb{R}$ is an $\mathcal{F}$-measurable random variable, i.e., $X^{-1}(A) \in \mathcal{F}\, \forall A \in \mathcal{G}$, we define the probability distribution function $F(x)$ of X as

$$F(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\}). \tag{2.1}$$

The expected value $\mathbb{E}[X]$ of X is the integral of X with respect to P, i.e.,

$$\mathbb{E}[X] = \int_\Omega X(\omega)P(d\omega) = \int_\Omega X(\omega)dP(\omega). \tag{2.2}$$

Several standard inequalities related to the expected value will be needed:

Lemma 1 (Markov's inequality [63, (5.31)]). *Let X be a nonnegative random variable and $\epsilon > 0$. Then,*

$$P(X \geq \epsilon) \leq \frac{1}{\epsilon}\mathbb{E}[X].$$

In particular, for an arbitrary random variable X and some integer k,

$$P\left(|X| \geq \epsilon\right) \leq \frac{1}{\epsilon^k}\mathbb{E}\left[|X|^k\right].$$

Lemma 2 (Hoelder's inequality [63, 5.35]). *For X, Y two arbitrary random variables and $p, q > 1$, satisfying $\frac{1}{q} + \frac{1}{p} = 1$,*

$$\mathbb{E}\left[|XY|\right] \leq \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} \left(\mathbb{E}\left[|Y|^q\right]\right)^{1/q}.$$

In particular, for two sets $\{x_1, \dots x_n\}$ and $\{y_1, \dots y_n\}$ of complex numbers,

$$\sum_i |x_i y_i| \leq \left(\sum_i |x_i|^p\right)^{1/p} \left(\sum_i |y_i|^q\right)^{1/q}.$$

We will often deal with infinite sequences $X_1(\omega), X_2(\omega), \dots$ of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Here, each ω generates the entire sequence $(X_n(\omega))_{n\geq 1}$ and not only a single random variable $X_n(\omega)$. Similarly, we consider sequences of distribution functions $(F_n)_{n\geq 1}$, e.g., when X_n has distribution F_n. One can define several types of convergence behavior related to both types of sequences.

Definition 1 (Weak convergence). *The sequence of distribution functions $(F_n)_{n\geq 1}$ converges weakly to the function F, if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for each $x \in \mathbb{R}$ at which F is continuous. This is denoted by $F_n \Rightarrow F$. If X_n and X have distributions F_n and F, respectively, we also write $X_n \Rightarrow X$ or $X_n \Rightarrow F$.

Definition 2 (Convergence in probability). *The sequence of random variables $(X_n)_{n\geq 1}$ converges in probability to X, if for all $\epsilon > 0$*

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

This is denoted by $X_n \xrightarrow{\mathcal{P}} X$.

The notion of convergence in probability does not play a role in this thesis.

Definition 3 (Almost sure convergence). *The sequence of random variables $(X_n)_{n\geq 1}$ converges almost surely to X, if*

$$P\left(\limsup_{n\to\infty} |X_n - X| = 0\right) = 1.$$

This is denoted by $X_n \xrightarrow{a.s.} X$.

Definition 4 (Convergence in the rth mean). *The sequence of random variables $(X_n)_{n\geq 1}$ converges in the rth mean to X, if for $r \geq 1$*

$$\limsup_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0.$$

Remark 1. *It will be of particular interest in this thesis to prove the almost sure weak convergence of sequences of "random" distribution functions* $(F_n)_{n\geq 1}$ *generated by some probability space* $(\Omega, \mathcal{F}, P)$*. That is, we will show that for all* $\omega \in A$*,* $A \subset \Omega$ *with* $P(A) = 1$*,* $\lim_{n\to\infty} F_n(x, \omega) = F(x)$*, for some distribution function* F*.*

Remark 2. *Note that the almost sure convergence of some sequence of random variables implies their convergence in probability, while convergence in probability implies weak convergence. Convergence in the* r*th mean implies convergence in the* $(r-1)$*th mean and convergence in probability.*

In order to prove the almost sure convergence of $(X_n)_{n\geq 1}$ to some constant X, one often relies on the combination of Markov's inequality (Lemma 1) and the first Borel-Cantelli lemma:

Lemma 3 (First Borel-Cantelli lemma [63, Theorem 4.3])**.** *Let* $(A_n)_{n\geq 1}$ *be a sequence of sets,* $A_n \in \mathcal{F}$ *for some probability space* $(\Omega, \mathcal{F}, P)$*. If* $\sum_n P(A_n) \leq \infty$*, then* $P(\limsup_n A_n) = 0$*.*

One first defines $A_n^\epsilon = \{\omega : |X_n(w) - X| \geq \epsilon\}$, for some $\epsilon > 0$, and shows by Markov's inequality that $P(A_n^\epsilon) \leq \frac{1}{\epsilon^k}\mathbb{E}\left[|X_n - X|^k\right] = f_n$, where $\sum_n f_n < \infty$. In many cases, we have $k = 4$ and $f_n = \mathcal{O}(n^{-2})$. This implies by the Borel-Cantelli lemma that $P(\limsup_n A_n^\epsilon) = 0$. In particular, let $\epsilon_{p,q} = \frac{p}{q}$ for $p, q \in \mathbb{N}^\star$. Thus, $P(\cup_{p,q\in\mathbb{N}^\star} \limsup_n A_n^{\epsilon_{p,q}}) = 0$ since the countable union of sets of probability zero has also zero measure [63]. For each $p, q \in \mathbb{N}^\star$ and ω in the complement of this set (a set of probability one), there exists $n_0(\omega)$, such that, for all $n \geq n_0(\omega)$, $|X_n(w) - X| < \frac{p}{q}$. Thus, $X_n \xrightarrow{\text{a.s.}} X$. Note that all proofs relying on this technique automatically establish the convergence of $(X_n)_{n\geq 1}$ to X in the kth mean.

We are often interested in the behavior of functions f of random variables X_n. The portmanteau lemma provides an important equivalent description if $(X_n)_{n\geq 1}$ converges weakly to X:

Lemma 4 (A Portmanteau lemma [64, Lemma 2.2 (i) and (ii)])**.** *Let* $(X_n)_{n\geq 1}$ *be a sequence of random variables, where* X_n *has distribution* F_n*, and let* X *be a random variable with distribution* F*. The following statements are equivalent:*

(i) $\lim_{n\to\infty} F_n(x) = F(x)$ *for all continuity points* $x \in \mathbb{R}$ *of* F*;*

(ii) $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ *for all bounded continuous functions* f*.*

The continuous mapping theorem is a very useful result if f is continuous:

Theorem 1 (Continuous mapping theorem [64, Theorem 2.3])**.** *Let* $(X_n)_{n\geq 1}$ *be a sequence of real random variables and let* $f : \mathbb{R} \mapsto \mathbb{R}$ *be continuous at every point of a set* A *such that* $P(X \in A) = 1$*, for some random variable* X*.*

(i) If $X_n \Rightarrow X$*, then* $f(X_n) \Rightarrow f(X)$*;*

(ii) If $X_n \xrightarrow{\mathcal{P}} X$*, then* $f(X_n) \xrightarrow{\mathcal{P}} f(X)$*;*

(iii) If $X_n \xrightarrow{a.s.} X$*, then* $f(X_n) \xrightarrow{a.s.} f(X)$*.*

In several cases, one is able to prove that $X_n \xrightarrow{\text{a.s.}} X$, but one would like to show that $(X_n)_{n\geq 1}$ converges also in the mean to X, i.e., $\lim_n \mathbb{E}\left[|X_n - X|\right] = 0$. This can often be done by the dominated convergence theorem:

Theorem 2 (Dominated convergence theorem [63, Theorem 16.4]). *Let $(f_n)_{n\geq 1}$ be a sequence of real measurable functions on some measure space $(\Omega, \mathcal{F}, P)$. Assume that for all $\omega \in A \subset \Omega$ with $P(A) = 1$, $\lim_{n\to\infty} f_n(\omega) \to f(\omega)$ for some measurable function f and $|f_n| \leq g$ for some integrable function g, i.e., $\int_\Omega |g(\omega)|dP(\omega) < \infty$. Then,*

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)dP(\omega) = \int_\Omega f(\omega)dP(\omega).$$

Define the functions $f_n(\omega) = |X_n(\omega) - X|$ for all n. Since $X_n \xrightarrow{\text{a.s.}} X$, it follows that $f_n \xrightarrow{\text{a.s.}} f = 0$. If one can show that $f_n \leq g$ and $\mathbb{E}\left[|g|\right] < \infty$, it follows from the dominated convergence theorem that $\lim_{n\to\infty} \mathbb{E}\left[|X_n(\omega) - X|\right] = 0$.

Sometimes we will deal with functions of two (or more) random variables X and Y, defined on the probability spaces $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$, respectively. Denote by $(\Omega\times\Omega', \mathcal{F}\times\mathcal{F}', Q)$ their product space (see [63, Sec. 18]). The Fubini theorem (also referred to as Tonelli's theorem for nonnegative functions) allows us to evaluate expectations with respect to the product measure Q by two iterated integrals with respect to the measures P and P'. The most important application of the Fubini theorem in this thesis will be described in detail in Section 2.4 where we deal with functional of products of random matrices.

Theorem 3 (Fubini theorem [63, Theorem 18.3]). *Let $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$ be two probability spaces. Denote $(\Omega \times \Omega', \mathcal{F} \times \mathcal{F}', Q)$ their product space. Let $f : \Omega \times \Omega' \to \mathbb{R}$ be $(\mathcal{F} \times \mathcal{F}')$-integrable, i.e., $\int_{\Omega\times\Omega'} |f(\omega, \omega')|dQ(\omega, \omega') < \infty$. Then,*

$$\begin{aligned}\int_{\Omega\times\Omega'} f(\omega, \omega')dQ(\omega, \omega') &= \int_\Omega \left[\int_{\Omega'} f(\omega, \omega')dP'(\omega')\right] dP(\omega)\\ &= \int_{\Omega'} \left[\int_\Omega f(\omega, \omega')dP(\omega)\right] dP'(\omega').\end{aligned}$$

2.2 Background on random matrix theory

Before we present concepts and results related to random matrices, we first recall some standard lemmas and identities which will be of repeated use throughout the thesis.

Lemma 5 (Resolvent identity). *For invertible matrices $\mathbf{A}$ and $\mathbf{B}$, we have the following identity:*

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}.$$

Lemma 6 (Matrix inversion lemma 1 [65, Eq. (2.2)]). *Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be invertible. Then, for any vector $\mathbf{x} \in \mathbb{C}^N$ and any scalar $c \in \mathbb{C}$ such that $\mathbf{A}+c\mathbf{x}\mathbf{x}^\mathsf{H}$ is invertible,*

$$\mathbf{x}^\mathsf{H}\left(\mathbf{A} + c\mathbf{x}\mathbf{x}^\mathsf{H}\right)^{-1} = \frac{\mathbf{x}^\mathsf{H}\mathbf{A}^{-1}}{1 + c\mathbf{x}^\mathsf{H}\mathbf{A}^{-1}\mathbf{x}}.$$

Lemma 7 (Matrix inversion lemma 2)**.** *Under the assumptions of Lemma 6,*

$$\left(\mathbf{A} + c\mathbf{x}\mathbf{x}^{\mathsf{H}}\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}c\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}}{1 + c\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}}.$$

Proof. By Lemmas 5 and 6, we have

$$\begin{aligned}\left(\mathbf{A} + c\mathbf{x}\mathbf{x}^{\mathsf{H}}\right)^{-1} - \mathbf{A}^{-1} &= -\left(\mathbf{A} + c\mathbf{x}\mathbf{x}^{\mathsf{H}}\right)^{-1} c\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1} \\ &= -\frac{\mathbf{A}^{-1}c\mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}}{1 + c\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}}.\end{aligned}$$

Adding $\mathbf{A}^{-1}$ to both sides of the last equation concludes the proof. □

Lemma 8 (Rank-1 perturbation lemma [66, Lemma 2.1])**.** *Let* $z \in \mathbb{C}\setminus\mathbb{R}^+$, $\mathbf{A} \in \mathbb{C}^{N\times N}$ *and* $\mathbf{B} \in \mathbb{C}^{N\times N}$ *with* $\mathbf{B}$ *Hermitian nonnegative definite, and* $\mathbf{x} \in \mathbb{C}^N$*. Then,*

$$\left|\operatorname{tr}\left(\left(\mathbf{B} - z\mathbf{I}_N\right)^{-1} - \left(\mathbf{B} + \mathbf{x}\mathbf{x}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\right)\mathbf{A}\right| \leq \frac{\|\mathbf{A}\|}{\operatorname{dist}(z, \mathbb{R}^+)}$$

where dist *is the Euclidean distance.*

Lemma 9 (A trace inequality)**.** *For any* $\mathbf{A} \in \mathbb{C}^{N\times N}$ *and* $\mathbf{B} \in \mathbb{C}^{N\times N}$*,*

$$|\operatorname{tr}\mathbf{A}\mathbf{B}| \leq \sqrt{\operatorname{tr}\mathbf{A}\mathbf{A}^{\mathsf{H}}\operatorname{tr}\mathbf{B}\mathbf{B}^{\mathsf{H}}} \leq N\|\mathbf{A}\|\|\mathbf{B}\|.$$

Lemma 10 (A matrix norm inequality)**.** *For any* $\mathbf{A} \in \mathbb{C}^{N\times N}$ *and* $\mathbf{B} \in \mathbb{C}^{N\times N}$*,*

$$\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|.$$

Lemma 11 ([67, Lemma B.26])**.** *Let* $\mathbf{A} \in \mathbb{C}^{N\times N}$ *be deterministic and* $\mathbf{x} = [x_1 \dots x_N]^{\mathsf{T}} \in \mathbb{C}^N$ *be a random vector of independent entries. Assume* $\mathbb{E}[x_i] = 0$, $\mathbb{E}\left[|x_i|^2\right] = 1$*, and* $\mathbb{E}\left[|x_i|^\ell\right] \leq \nu_\ell$*. Then, for any* $p \geq 1$*,*

$$\mathbb{E}\left[\left|\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} - \operatorname{tr}\mathbf{A}\right|^p\right] \leq C_p\left(\operatorname{tr}\mathbf{A}\mathbf{A}^{\mathsf{H}}\right)^{\frac{p}{2}}\left(\nu_4^{\frac{p}{2}} + \nu_{2p}\right)$$

where C_p *is a constant which only depends on* p*.*

A sometimes useful result, similar to Lemma 11, can be stated as follows:

Lemma 12 ([68, Lemma 3])**.** *Let* $\mathbf{A} \in \mathbb{C}^{N\times N}$ *be deterministic and* $\mathbf{x} = [x_1 \dots x_N]^{\mathsf{T}} \in \mathbb{C}^N$ *be a random vector of independent entries. Assume* $\mathbb{E}[x_i] = 0$, $\mathbb{E}\left[|x_i|^2\right] = \frac{1}{N}$*, and* $\mathbb{E}\left[|\sqrt{N}x_i|^l\right] \leq \nu_l$*. Then, for any integer* $p, q \geq 1$*,*

$$\mathbb{E}\left[\left|\left(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right)^q - \left(\frac{1}{N}\operatorname{tr}\mathbf{A}\right)^q\right|^p\right] \leq \frac{\|\mathbf{A}\|^{pq}}{N^{\frac{p}{2}}} q^{p-\frac{1}{2}} \sqrt{C_{2p}\left(\nu_4^{p} - \nu_{4p}\right)\sum_{k=0}^{q-1}\nu_{4pk}}$$

where C_{2p} *is a constant which only depends on* p*.*

Proof. The proof is provided in Appendix 2.5.1. □

From now on, we will consider sequences of random matrices $(\mathbf{A}_N)_{N\geq 1}$ and vectors $(\mathbf{x}_N)_{N\geq 1}$ with growing dimensions, e.g., $\mathbf{A}_N \in \mathbb{C}^{N\times N}$ and $\mathbf{x}_N \in \mathbb{C}^N$. Therefore, a sample point $\omega \in \Omega$ of the underlying probability space $(\Omega, \mathcal{F}, P)$ is assumed to create the entire sequences $(\mathbf{A}_N(\omega))_{N\geq 1}$ and $(\mathbf{x}_N(\omega))_{N\geq 1}$ rather than a single realization of $\mathbf{A}_N$ and $\mathbf{x}_N$. One of the most important results, which can be easily obtained from Lemma 11 (originally proved in [69]), is given by the next lemma:

Lemma 13 (A trace lemma [69, Lemma 2.7])**.** *Let $(\mathbf{A}_N)_{N\geq 1}$, $\mathbf{A}_N \in \mathbb{C}^{N\times N}$, be a sequence of matrices and $(\mathbf{x}_N)_{N\geq 1}$, $\mathbf{x}_N = [x_{1,N} \dots x_{N,N}]^\mathsf{T} \in \mathbb{C}^N$, a sequence of random vectors of i.i.d. entries, independent of $(\mathbf{A}_N)_{N\geq 1}$. Assume that $\mathbb{E}\left[x_{i,j}\right] = 0$, $\mathbb{E}\left[|x_{i,j}|^2\right] = 1$, $\mathbb{E}\left[|x_{i,j}|^8\right] < \infty$, and $\limsup_N \|\mathbf{A}_N\| < \infty$. Then,*

$$\frac{1}{N}\mathbf{x}_N^\mathsf{H}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\operatorname{tr}\mathbf{A}_N \xrightarrow{\text{a.s.}} 0.$$

Proof. By Lemma 11, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{N}\mathbf{x}_N^\mathsf{H}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\operatorname{tr}\mathbf{A}_N\right|^4\right] \\
&\leq \frac{C_4}{N^2}\left(\frac{1}{N}\operatorname{tr}\mathbf{A}_N\mathbf{A}_N^\mathsf{H}\right)^2\left(\left(\mathbb{E}\left[|x_{1,N}|^4\right]\right)^2 + \mathbb{E}\left[|x_{1,N}|^8\right]\right) \\
&\leq \frac{C_4}{N^2}\|\mathbf{A}_N\|^4\left(\left(\mathbb{E}\left[|x_{1,N}|^4\right]\right)^2 + \mathbb{E}\left[|x_{1,N}|^8\right]\right) \qquad (2.3)
\end{aligned}$$

where the last inequality follows from Lemma 10. By assumption, for every $\epsilon > 0$, there exists an N_0, such that for all $N > N_0$, $\|\mathbf{A}_N\| < D + \epsilon$ for some finite constant D. Thus, for $N > N_0$, the RHS of (2.3) is a $\mathcal{O}(N^{-2})$ and hence summable. By Lemma 3, it follows that $\frac{1}{N}\mathbf{x}_N^\mathsf{H}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\operatorname{tr}\mathbf{A}_N \xrightarrow{\text{a.s.}} 0$. □

Remark 3. *Several refinements of Lemma 13 are possible. First, the elements of $\mathbf{x}_n$ do not need to be independent and identically distributed (i.i.d.) but only independent. Second, in [70, Lemma 4], the condition $\limsup_N \|\mathbf{A}_N\| < \infty$ is relaxed to hold only almost surely. Third, the lemma also holds if the condition $\limsup_N \|\mathbf{A}_N\| < \infty$ is replaced by $\limsup_N \mathbb{E}\left[\frac{1}{N}\operatorname{tr}\left(\mathbf{A}_N\mathbf{A}_N^\mathsf{H}\right)^2\right] < \infty$ as can be seen by (2.3), since $\frac{1}{N^2}(\operatorname{tr}\mathbf{A}\mathbf{A}^\mathsf{H})^2 \leq \frac{1}{N}\operatorname{tr}\left(\mathbf{A}\mathbf{A}^\mathsf{H}\right)^2$.*

In a similar spirit, the next lemma shows under which conditions terms of the form $\mathbf{x}_N^\mathsf{H}\mathbf{A}_N\mathbf{y}_N$ for independent sequences of random vectors $(\mathbf{x}_N)_{N\geq 1}$ and $(\mathbf{y}_N)_{N\geq 1}$ vanish.

Lemma 14 ([71, Lemma 3.7])**.** *Let $(\mathbf{A}_N)_{N\geq 1}$, $\mathbf{A}_N \in \mathbb{C}^{N\times N}$, be a sequence of matrices and $(\mathbf{x}_N)_{N\geq 1}$, $\mathbf{x}\,[x_{1,N} \dots x_{N,N}]^\mathsf{T} \in \mathbb{C}^N$, and $(\mathbf{y}_N)_{N\geq 1}$ two sequences of random vectors of independent entries, independent of $(\mathbf{A}_N)_{N\geq 1}$. Assume that $\mathbb{E}\left[x_{i,j}\right] = 0$, $\mathbb{E}\left[|x_{i,j}|^2\right] = 1$, $\mathbb{E}\left[|x_{i,j}|^4\right] < \infty$, and $\limsup_N \|\mathbf{A}_N\| < \infty$. Then,*

$$\frac{1}{N}\mathbf{x}_N^\mathsf{H}\mathbf{A}_N\mathbf{y}_N \xrightarrow{\text{a.s.}} 0.$$

A special class of random matrices which will play an important role in this thesis are the so-called Haar matrices, isometric matrices or unitarily invariant unitary matrices:

Definition 5 (Unitary matrix)**.** *A matrix* $\mathbf{U} \in \mathbb{C}^{N\times N}$ *is unitary if it satisfies* $\mathbf{U}\mathbf{U}^{\mathsf{H}} = \mathbf{U}^{\mathsf{H}}\mathbf{U} = \mathbf{I}_N$*. We denote by* $\mathcal{U}(N)$ *the set of* $N \times N$ *unitary matrices.*

Definition 6 (Haar matrix)**.** *A Haar matrix* $\mathbf{U} \in \mathbb{C}^{N\times N}$*, is a matrix-valued random variable which takes its values uniformly from* $\mathcal{U}(N)$*. In particular, the eigenvalues of* $\mathbf{U}$ *are uniformly distributed on the complex unit circle.*

Remark 4. *There are several ways to create Haar matrices. Let* $\mathbf{X} \in \mathbb{C}^{N\times N}$ *be random of i.i.d. entries* $\mathbf{X}_{i,j} \sim \mathcal{CN}(0,1)$*. Then, the matrix* $\mathbf{W} \in \mathbb{C}^{N\times N}$*, defined as* $\mathbf{W} = \mathbf{X}\left(\mathbf{X}^{\mathsf{H}}\mathbf{X}\right)^{-\frac{1}{2}}$ *is a Haar matrix. Alternatively, let* $\mathbf{Y} \in \mathbb{C}^{N\times N}$ *be random of i.i.d. entries* $\mathbf{Y}_{i,j} \sim \mathcal{CN}(0,1)$*. Denote by* $\mathbf{Y} = \mathbf{QR}$ *its QR-decomposition such that* $\mathbf{R}$ *has nonnegative diagonal entries. Then,* $\mathbf{Q}$ *is a Haar matrix.*

Definition 7 (Unitarily invariant matrix)**.** *Let* $\mathbf{W} \in \mathbb{C}^{N\times n}$ *be a random matrix and let* $\mathbf{U} \in \mathbb{C}^{N\times N}$*,* $\mathbf{V} \in \mathbb{C}^{n\times n}$ *be two unitary matrices, independent of* $\mathbf{W}$*. The matrix* $\mathbf{W}$ *is said to be left-unitarily invariant (right-unitarily invariant) if the distribution of* $\mathbf{UW}$ $(\mathbf{WV})$ *equals that of* $\mathbf{W}$*. The matrix* $\mathbf{W}$ *is bi-unitarily invariant if it is both left- and right-unitarily invariant.*

An important trace lemma which can be seen as the counterpart to Lemma 13 for i.i.d. matrices, is given as follows:

Lemma 15 (Trace lemma for Haar matrices [72, 73] (see also [71]))**.** *Let* $(\mathbf{W}_N)_{N\geq 1}$*,* $\mathbf{W}_N \in \mathbb{C}^{N\times n}$*, be a sequence of random matrices, where* $\mathbf{W}_N$ *consists of* $n < N$ *columns of an* $N \times N$ *Haar matrix, and suppose that* $\mathbf{w}_N \in \mathbb{C}^N$ *is a column of* $\mathbf{W}_N$*. Let* $(\mathbf{B}_N)_{N\geq 1}$*,* $\mathbf{B}_N \in \mathbb{C}^{N\times N}$*, be a sequence of random matrices, where* $\mathbf{B}_N$ *depends on all columns of* $\mathbf{W}_N$ *except* $\mathbf{w}_N$*. Let* $(\mathbf{A}_N)_{N\geq 1}$*,* $\mathbf{A}_N \in \mathbb{C}^{N\times N}$*, be a sequence of random matrices independent of* $(\mathbf{W}_N)_{N\geq 1}$*. Assume that* N *and* n *grow infinitely large, such that* $c = \limsup_N n/N < 1$*,* $\limsup_N \|\mathbf{B}_N\| < \infty$*, and* $\limsup_N \|\mathbf{A}_N\| < \infty$*. Then,*

$$(i) \quad \mathbb{E}\left[\left|\mathbf{w}_N^{\mathsf{H}}\mathbf{B}_N\mathbf{w}_N - \frac{1}{N-n}\operatorname{tr}\mathbf{\Pi}_N\mathbf{B}_N\right|^4\right] = \mathcal{O}\left(\frac{1}{N^2}\right)$$

$$(ii) \quad \mathbf{w}_N^{\mathsf{H}}\mathbf{B}_N\mathbf{w}_N - \frac{1}{N-n}\operatorname{tr}\mathbf{\Pi}_N\mathbf{B}_N \xrightarrow{\text{a.s.}} 0$$

$$(iii) \quad \mathbf{w}_N^{\mathsf{H}}\mathbf{B}_N\mathbf{w}_N - \frac{1}{N-n}\operatorname{tr}\left(\mathbf{I}_N - \mathbf{W}_N\mathbf{W}_N^{\mathsf{H}}\right)\mathbf{B}_N \xrightarrow{\text{a.s.}} 0$$

$$(iv) \quad \mathbf{w}_N^{\mathsf{H}}\mathbf{A}_N\mathbf{w}_N - \frac{1}{N}\operatorname{tr}\mathbf{A}_N \xrightarrow{\text{a.s.}} 0$$

where $\mathbf{\Pi}_N = \mathbf{I}_N - \mathbf{W}_N\mathbf{W}_n^{\mathsf{H}} + \mathbf{w}_N\mathbf{w}_N^{\mathsf{H}}$*.*

2.2.1 The Stieltjes transform

Definition 8 (Stieltjes transform)**.** *Let* μ *be a finite nonnegative measure with support* $\operatorname{supp}(\mu) \subset \mathbb{R}$*, i.e.,* $\mu(\mathbb{R}) < \infty$*. The Stieltjes transform* $m(z)$ *of* μ *is defined for* $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$ *as*

$$m(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} d\mu(\lambda).$$

We denote by $\mathcal{S}(\mathbb{R}^+)$ *the class of Stieltjes transforms of probability measures carried by* $\mathbb{R}^+$*, i.e.,* $\mu(\mathbb{R}^+) = 1$*.*

We will now summarize several important properties of the Stieltjes transform. These results can be found for example in [74] or [75].

Property 1. *Let $m(z)$ be the Stieltjes transform of a finite nonnegative measure μ on $\mathbb{R}$. Then,*

(i) $m(z)$ *is analytic over* $\mathbb{C} \setminus \operatorname{supp}(\mu)$,

(ii) $z \in \mathbb{C}^+$ *implies* $m(z) \in \mathbb{C}^+$,

(iii) if $z \in \mathbb{C}^+$, $|m(z)| \leq \frac{1}{\Im\{z\}}$ *and* $\Im\{\frac{1}{m(z)}\} \leq -\Im\{z\}$,

(iv) if $\mu((-\infty, 0)) = 0$, *then* $m(z)$ *is analytic over* $\mathbb{C} \setminus \mathbb{R}^+$. *In addition,* $z \in \mathbb{C}^+$ *implies* $zm(z) \in \mathbb{C}^+$ *and the following inequalities hold:*

$$|m(z)| \leq \begin{cases} \frac{1}{\Im\{z\}} & , \ z \in \mathbb{C} \setminus \mathbb{R} \\ |z|^{-1} & , \ z < 0 \\ \frac{1}{\operatorname{dist}(z, \mathbb{R}^+)} & , \ z \in \mathbb{C} \setminus \mathbb{R}^+ \end{cases}$$

where dist *is the Euclidean distance.*

There is another set of properties which allow one to recover μ when only its Stieltjes transform $m(z)$ is known.

Property 2. *Let $m(z)$ be the Stieltjes transform of a finite measure μ on $\mathbb{R}$. Then,*

(i) $\mu(\mathbb{R}) = \lim_{y\to\infty} -\mathbf{i}ym(\mathbf{i}y)$,

(ii) $\mu([a,b]) = \lim_{y\to 0^+} \int_a^b \Im\{m(x + \mathbf{i}y)\}dx$, *if a, b are continuity points of μ.*

The following property is useful if one wants to prove that a given function is the Stieltjes transform of a finite measure.

Property 3. *Let $m(z)$ be an analytic function over $\mathbb{C}^+$ such that $m(z) \in \mathbb{C}^+$ if $z \in \mathbb{C}^+$. If $\limsup_{y\to\infty} |\mathbf{i}ym(\mathbf{i}y)| < \infty$, then $m(z)$ is the Stieltjes transform of a finite nonnegative measure on $\mathbb{R}$. If additionally, $zm(z) \in \mathbb{C}^+$ for $z \in \mathbb{C}^+$, then $\mu(\mathbb{R}^-) = 0$ and $m(z)$ has an analytic continuation on $\mathbb{C} \setminus \mathbb{R}^+$.*

Finally, the Stieltjes transform of a probability measure can be related to the moments of the underlying distribution.

Theorem 4 ([71, Theorem 3.3])**.** *Let μ be a probability measure on $\mathbb{R}$ and denote by $m(z)$ and F its Stieltjes transform and distribution function, respectively. Assume that* $\operatorname{supp}(\mu) \subset [a,b]$ *for* $0 < a < b < \infty$. *Then, for* $z \in \mathbb{C} \setminus \mathbb{R}$, $|z| > b$,

$$m(z) = -\frac{1}{z} \sum_{k=0}^{\infty} \frac{M_k}{z^k}$$

where M_k are the moments of F, defined as

$$M_k = \int_{\mathbb{R}} \lambda^k d\mu(\lambda) = \int_{\mathbb{R}} \lambda^k dF(\lambda).$$

Remark 5. *The function* $G(z) = \frac{1}{z}m(-1/z) = \sum_{k=0}^{\infty}(-1)^k z^k M_k$ *can be seen as a moment generating function of* F*, since* $M_k = \frac{(-1)^k}{k!}G^{(k)}(0)$*, where* $G^{(k)}(z)$ *is the* k*th derivative of* $G(z)$*.*

We will now discuss the important relation of the Stieltjes transform and the eigenvalue distribution or the empirical spectral distribution (e.s.d.) of Hermitian matrices:

Definition 9 (Empirical spectral distribution)**.** *Let* $\mathbf{A} \in \mathbb{C}^{N\times N}$ *be a Hermitian matrix with eigenvalues* $\lambda_1, \dots, \lambda_N$*. The e.s.d.* $F^{\mathbf{A}}(x)$ *of the eigenvalues of* $\mathbf{A}$ *is defined as*

$$F^{\mathbf{A}}(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{\lambda_i \le x\}.$$

The Stieltjes transform $m_{\mathbf{A}}(z)$ of the of the e.s.d. $F^{\mathbf{A}}$ of some Hermitian matrix $\mathbf{A}$ can be written in several ways:

$$\begin{aligned} m_{\mathbf{A}}(z) &= \int_{\mathbb{R}} \frac{1}{\lambda - z} dF^{\mathbf{A}}(\lambda) \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\lambda_i - z} \\ &= \frac{1}{N}\operatorname{tr}\left(\mathbf{A} - z\mathbf{I}_N\right)^{-1} \end{aligned} \tag{2.4}$$

since the eigenvalues of $\mathbf{A} - z\mathbf{I}_N$ equal $\lambda_1 - z, \dots, \lambda_N - z$ and the trace equals the sum of the eigenvalues of a matrix. For the sake of brevity, we will call $m_{\mathbf{A}}(z)$ the Stieltjes transform of $\mathbf{A}$, rather than the Stieltjes transform of $F^{\mathbf{A}}$. This is also consistent with our notation as we write $m_{\mathbf{A}}$ instead of $m_{F^{\mathbf{A}}}$. Once $m_{\mathbf{A}}(z)$ is known, one can recover the moments M_k of $\mathbf{A}$ by Theorem 4 since

$$M_k = \frac{1}{N}\operatorname{tr}\mathbf{A}^k = \int_{\mathbb{R}} \lambda^k dF^{\mathbf{A}}(\lambda). \tag{2.5}$$

2.3 Deterministic equivalents

Let us begin with a rigorous definition of a deterministic equivalent:

Definition 10 (Deterministic equivalent)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $(f_n)_{n\ge1}$ *a series of measurable complex-valued functions,* $f_n : \Omega \times \mathbb{C} \to \mathbb{C}$*. Let* $(g_n)_{n\ge1}$ *be a series of complex-valued functions,* $g_n : \mathbb{C} \to \mathbb{C}$*. Then,* $(g_n)_{n\ge1}$ *is said to be a deterministic equivalent of* $(f_n)_{n\ge1}$ *on* $D \subset \mathbb{C}$*, if there exists a set* $A \subset \Omega$ *with* $P(A) = 1$*, such that*

$$f_n(\omega, z) - g_n(z) \xrightarrow[n\to\infty]{} 0$$

for all $\omega \in A$ *and for all* $z \in D$*. This will be denoted by* $f_n \asymp g_n$*.*

Otherwise stated, a deterministic equivalent for $(f_n)_{n\ge1}$ is a series $(g_n)_{n\ge1}$ such that $g_n(z)$ approximates $f_n(\omega, z)$ arbitrarily closely as n grows, for every $z \in D$ and almost every ω. In particular, if $(f_n)_{n\ge1}$ converges almost surely to a limiting function f, i.e., for all $(\omega, z) \in A \times D$ with $A \subset \Omega$, $P(A) = 1$ and $D \subset \mathbb{C}$, we have

$$f_n(\omega, z) \xrightarrow[n\to\infty]{} f(z) \tag{2.6}$$

then $(g_n)_{n\geq 1}$ defined by $g_n = f$, for all n, is also a deterministic equivalent of $(f_n)_{n\geq 1}$. In many cases of practical interest, one can also show that $\int_\Omega f_n(\omega, z)dP(\omega) - g_n(z) \to 0$ as $n \to \infty$. Thus g_n is also an approximation of the expected value of f_n.

In order to further illustrate the difference between deterministic equivalents and asymptotic limits, let us consider the infinite sequence $(x_n)_{n\geq 1}$ of random numbers, where $x_n \sim \mathcal{CN}\left((-1)^n, \frac{1}{n}\right)$. Clearly, there is no $\bar{x}$ such that $x_n \to \bar{x}$, almost surely. However, we can define the deterministic series $(\bar{x}_n)_{n\geq 1}$, where $\bar{x}_n = (-1)^n$, which satisfies $x_n - \bar{x}_n \to 0$, almost surely. The last result follows directly from the Markov inequality and the Borel-Cantelli lemma since $\mathbb{E}[|x_n - \bar{x}_n|^4] = 2n^{-2}$. Thus, a deterministic equivalent of $(x_n)_{n\geq 1}$ can be defined although the sequence has no asymptotic limit (in the almost sure sense). Note that in this example, $(\bar{x_n})_{n\geq 1}$ is also a deterministic equivalent of $(\mathbb{E}[x_n])_{n\geq 1}$, since $\bar{x}_n = \mathbb{E}[x_n]$).

We have already seen several deterministic equivalents for functionals of random vectors and matrices in Section 2.2. For example, in Lemma 13, $\frac{1}{N}\operatorname{tr}\mathbf{A}_N$ is a deterministic equivalent of the quadratic form $\mathbf{x}_N^{\mathsf{H}}\mathbf{A}_N\mathbf{x}_N$. In wireless communications, one is often interested in the behavior of functionals $f_n(\mathbf{H}_n, z)$, where $\mathbf{H}_n \in \mathbb{C}^{N\times n}$ is a matrix describing the input-output relation of a wireless channel. In particular, $f_n(\mathbf{H}_n, z) = \frac{1}{N}\log\det(\mathbf{I}_N + z\mathbf{H}_n\mathbf{H}_n^{\mathsf{H}})$, $z \in \mathbb{R}^+$, is the (normalized) mutual information of the MIMO channel $\mathbf{H}_n$ between an n-antenna transmitter and an N-antenna receiver at signal-to-noise ratio (SNR) z. Other quantities of interest are the signal-to-interference-plus-noise ratio (SINR) with linear detectors or precoders and the associated rates. The goal of a large system analysis based on RMT is to provide deterministic approximations of these random quantities, which become arbitrarily tight as the system dimensions grow. Thus, deterministic equivalents provide a *deterministic abstraction of the physical layer*. This is particularly interesting for involved channel models which are intractable by exact analysis.

The next lemma is important when one deals with products or ratios of deterministic equivalents.

Lemma 16. *[76, Lemma 1] Let $(a_n)_{n\geq 1}$, $(\overline{a}_n)_{n\geq 1}$, $(b_n)_{n\geq 1}$, and $(\overline{b}_n)_{n\geq 1}$ four infinite sequences of complex random variables. Assume that $a_n \asymp \overline{a}_n$ and $b_n \asymp \overline{b}_n$.*

(i) If $|a_n|$, $|\overline{b}_n|$ and/or $|\overline{a}_n|$,$|b_n|$ are almost surely bounded, then

$$a_n b_n \asymp \overline{a}_n \overline{b}_n.$$

(ii) If $|a_n|$, $|\overline{b}_n|^{-1}$ and/or $|\overline{a}_n|$,$|b_n|^{-1}$ are almost surely bounded, then

$$a_n / b_n \asymp \overline{a}_n / \overline{b}_n.$$

Another important result shows that the weak convergence of two distributions implies the convergence of their respective Stieltjes transforms, and vice versa.

Theorem 5 ([67, Theorem B.9]). *Let $(F_n)_{n\geq 1}$ be a sequence of bounded real functions, satisfying* $\lim_{x\to-\infty} F_n(x) = 0$. *Then, there exists a sequence $(\bar{F}_n)_{n\geq 1}$ such that* $\lim_{x\to-\infty} \bar{F}_n(x) = 0$ *and* $|F_n(x) - \bar{F}_n(x)| \xrightarrow[n\to\infty]{} 0$ *for all $x \in \mathbb{R}$, if and only if*

$$m_{F_n}(z) - m_{\bar{F}_n}(z) \xrightarrow[n\to\infty]{} 0 \;\forall z \in \mathbb{C}^+$$

where $m_{F_n}(z)$ and $m_{\bar{F}_n}(z)$ are the Stieltjes transforms of F_n and $\bar{F}_n$, respectively.

Often, one is only able to show the almost sure point-wise convergence $m_{F_n}(z) - m_{\bar{F}_n}(z) \to 0$ for a given $z \in A \subset \mathbb{C}^+$, e.g., $|z| > \epsilon$. Since $|m_{F_n}(z)| \leq |z|^{-1}$ for all $z \in \mathbb{C} \setminus \mathrm{supp}(F_n)$, the next theorem can be used to prove that this convergence holds on a larger region, e.g., all regions excluding $\mathrm{supp}(F_n)$.

Theorem 6 (Vitali's convergence theorem [77, Theorem 5.21]). *Let $(f_n)_{n\geq 1}$ be a sequence of functions, analytic on $D \subset \mathbb{C}$, such that $|f_n(z)| \leq M$ $\forall n, \forall z \in D$. Assume that $f_n(z_j)$ converges for a countable set $z_1, z_2, \ldots \in D$, having a limit point in D. Then $f_n(z)$ converges uniformly in any region bounded by a contour interior to D. Moreover, this limit is an analytic function.*

From now on, all matrices and vectors should be understood as sequences of matrices and vectors with growing dimensions. For notational convenience, we drop the index n, e.g., we write $\mathbf{X}$ instead of $(\mathbf{X}_n)_{n\geq 1}$.

2.3.1 Existing results

One of the most famous examples is the Marčenko-Pastur law which provides a deterministic equivalent of the e.s.d. of the random matrix $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, where $\mathbf{X} \in \mathbb{C}^{N\times n}$ has i.i.d. entries with zero mean and variance $1/n$. It was first proven in [78], but then generalized to the following theorem:

Theorem 7 (Marčenko-Pastur law [67, Theorem 3.10]). *Let $\mathbf{X} \in \mathbb{C}^{N\times n}$ be a random matrix of independent entries, satisfying $\mathbb{E}[X_{i,j}] = 0$, $\mathbb{E}\left[|X_{i,j}|^2\right] = \frac{1}{n}$, and $\mathbb{E}\left[|\sqrt{n}X_{i,j}|^{2+\epsilon}\right] < \infty$, for $\epsilon > 0$. Denote by F_N the e.s.d. of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ and let $c = \frac{N}{n}$. Assume that $N, n \to \infty$ such that $0 < \liminf_N c \leq \limsup_N c < \infty$. Then, almost surely,*

$$F_N - F_c \Rightarrow 0$$

where F_c has density

$$dF_c(x) = \left(1 - c^{-1}\right)^+ \mathbb{1}\{x = 0\} + \frac{1}{2\pi c x}\sqrt{(x-a)^+(b-x)^+}$$

where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$. Moreover, the Stieltjes transform $m_c(z)$ of F_c is given by

$$m_c(z) = \frac{1-c}{2cz} - \frac{1}{2c} - \frac{\sqrt{(1-c-z)^2 - 4cz}}{2cz}$$

where the branch of $\frac{\sqrt{(1-c-z)^2-4cz}}{2cz}$ is chosen such that $m_c(z) \in \mathbb{C}^+$ for $z \in \mathbb{C}^+$ and $m_c(z) > 0$ for $z < 0$.

By Lemma 4, $F_N - F_c \Rightarrow 0$, almost surely, implies that

$$\int_{\mathbb{R}^+} f(\lambda) dF_N(\lambda) - \int_{\mathbb{R}^+} f(\lambda) dF_c(\lambda) \xrightarrow{\text{a.s.}} 0 \tag{2.7}$$

for any bounded continuous function $f(x)$. However, even in some cases where $f(x)$ is not bounded, e.g., $f(x) = \log(1+x)$, the convergence can be shown to hold since the support of F_N lies almost surely within a compact interval $[a, b]$.

Theorem 8 (No eigenvalues outside the support [69, Theorem 1.1] (see also [79, Theorem 3]))**.** *Let* $\mathbf{X}$ *be defined as in Theorem 7 and assume that* $\mathbb{E}\left[|\sqrt{n}X_{i,j}|^4\right] < \infty$*. Denote by* λ_{max} *and* λ_{min} *the largest and smallest eigenvalue of* $\mathbf{X}\mathbf{X}^{\mathsf{H}}$*, respectively. Then,*

(i) $\lambda_{max} - (1+\sqrt{c})^2 \xrightarrow{\text{a.s.}} 0$,

(ii) $\lambda_{min} - (1-\sqrt{c})^2 \mathbb{1}\{c \leq 1\} \xrightarrow{\text{a.s.}} 0$.

By Theorem 8, $f(x)$ can then be replaced by the function $f(x)\mathbb{1}\{x \leq M\}$, for some $M > \lambda_{\max}$, for which (2.7) holds. This leads to the following result:

Theorem 9 (see e.g. [80])**.** *Let* $\mathbf{X}$ *be defined as in Theorem 7 and assume that* $\mathbb{E}\left[|\sqrt{n}X_{i,j}|^4\right] < \infty$*. For* $x > 0$*, define* $I_N(x) = \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{X}\mathbf{X}^{\mathsf{H}}\right)$*. Then,*

$$I_N(x) - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0$$

where

$$\bar{I}_N(x) = \frac{1}{c}\log(1 + c m_c(-x)) + \log\left(1 + \frac{1}{x}\frac{1}{1+cm_c(-x)}\right) - \frac{m_c(-x)}{1+cm_c(-x)}.$$

Remark 6. *One can also show that the convergence holds in the first mean, i.e.,* $\mathbb{E}[I_N(x)] - \bar{I}_N(x) \to 0$ *(see for example [74, Theorem 4.1]. Moreover, if the entries of* $\mathbf{X}$ *are complex Gaussian random variables, one can prove the stronger result [81, Theorem 1]:* $\mathbb{E}[I_N(x)] = \bar{I}_N(x) + \mathcal{O}(N^{-2})$*.*

Once a deterministic equivalent of a random quantity is established, it is of interest to study its fluctuations around its deterministic approximation. This is often done in the form of a central limit theorem (CLT) as exemplarily shown in the next theorem:

Theorem 10 ([71, Theorem 3.18])**.** *Let* $I_N(x)$ *be defined as in Theorem 9. Then,*

$$\frac{N}{\Theta_c}\left(I_N(x) - \mathbb{E}[I_N(x)]\right) \Rightarrow \mathcal{N}(0,1)$$

where

$$\Theta_c^2 = -\log\left(1 - \frac{cm_c(-x)^2)}{(1+cm_c(-x))^2}\right) + \kappa\frac{cm_c(-x))}{(1+cm_c(-x))^2}$$

$\kappa = \mathbb{E}\left[|\sqrt{n}X_{1,1}|^4\right] - 2$ *and* $m_c(z)$ *is the Stieltjes transform of the Marčenko-Pastur law as given by Theorem 7.*

Based on Theorem 4, it is also possible to establish deterministic equivalents for the moments of certain random matrix models:

Theorem 11 (Marčenko-Pastur law [82])**.** *Let* $\mathbf{X}$ *be defined as in Theorem 9 and denote by* $M_k = \frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}}\right)^k$ *its* k*th moment. Then,*

$$M_k - \bar{M}_k \xrightarrow{\text{a.s.}} 0$$

where

$$\bar{M}_k = \sum_{i=0}^{k-1} \binom{k}{i} \binom{k}{i+1} \frac{c^i}{k}.$$

The following theorem provides a deterministic equivalent of the ergodic mutual information for channel matrices with a variance profile and non-centered entries. This model is also referred to as the Rician model and can be used to account for LOS components in a channel. The fluctuations of the mutual information for this channel model are described by Claim 1.

Theorem 12. *[74, Theorems 2.4, 2.5, 4.1] Let* $\mathbf{X} \in \mathbb{C}^{N\times n}$ *be random,* $\mathbf{A} \in \mathbb{C}^{N\times n}$ *deterministic, and define* $\mathbf{B}_N = (\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathsf{H}}$*. Assume that* $\mathbf{X}$ *has i.i.d. entries satisfying* $\mathbb{E}[X_{i,j}] = 0$*,* $\mathbb{E}[|X_{i,j}|] = \frac{\sigma_{i,j}^2}{n}$*, and* $\mathbb{E}\left[|\sqrt{n}X_{i,j}|^{4+\epsilon}\right] < \infty$ *for some* $\epsilon > 0$*. Assume that* $\sup_N \max_{i,j} \sigma_{i,j} < \infty$ *and and that the Euclidean norms of the rows and columns of* $\mathbf{A}$ *are bounded. Denote* $\mathbf{D}_j = \operatorname{diag}\left(\sigma_{1,j}^2, \ldots, \sigma_{N,j}^2\right)$ *and* $\widetilde{\mathbf{D}}_i = \operatorname{diag}\left(\sigma_{i,1}^2, \ldots, \sigma_{i,n}^2\right)$ $\forall i, j$*.*

(i) The following set of $N+n$ *deterministic equations,*

$$\psi_i(z) = \frac{-1}{z\left(1 + \frac{1}{n}\operatorname{tr}\widetilde{\mathbf{D}}_i\widetilde{\mathbf{T}}(z)\right)}, \quad 1 \le i \le N$$

$$\widetilde{\psi}_j(z) = \frac{-1}{z\left(1 + \frac{1}{n}\operatorname{tr}\mathbf{D}_j\mathbf{T}(z)\right)}, \quad 1 \le j \le n$$

where

$$\begin{aligned}
\mathbf{\Psi}(z) &= \operatorname{diag}\left(\psi_1(z), \ldots, \psi_N(z)\right) \\
\widetilde{\mathbf{\Psi}}(z) &= \operatorname{diag}\left(\widetilde{\psi}_1(z), \ldots, \widetilde{\psi}_n(z)\right) \\
\mathbf{T}(z) &= \left(\mathbf{\Psi}(z)^{-1} - z\mathbf{A}\widetilde{\mathbf{\Psi}}(z)\mathbf{A}^{\mathsf{H}}\right)^{-1} \\
\widetilde{\mathbf{T}}(z) &= \left(\widetilde{\mathbf{\Psi}}(z)^{-1} - z\mathbf{A}^{\mathsf{H}}\mathbf{\Psi}(z)\mathbf{A}\right)^{-1}
\end{aligned}$$

admits a unique solution $(\psi_1(z), \ldots, \psi_N(z), \widetilde{\psi_1}(z), \ldots, \widetilde{\psi_n}(z)) \in \mathcal{S}(\mathbb{R}^+)^{N+n}$ *for* $z \in \mathbb{C}\setminus\mathbb{R}^+$*.*

(ii) For $x > 0$*, let* $I_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N\right)$ *and consider the quantity:*

$$\begin{aligned}
\bar{I}_N(x) = {} & \frac{1}{N}\log\det\left(\frac{\mathbf{\Psi}(-x)^{-1}}{x} + \mathbf{A}\widetilde{\mathbf{\Psi}}(-x)\mathbf{A}^{\mathsf{H}}\right) + \frac{1}{N}\log\det\left(\frac{\widetilde{\mathbf{\Psi}}(-x)^{-1}}{x}\right) \\
& - \frac{x}{Nn}\sum_{i,j}\sigma_{ij}^2\mathbf{T}_{ii}(-x)\widetilde{\mathbf{T}}_{jj}(-x).
\end{aligned}$$

Then, for $N, n \to \infty$*, such that* $0 < \liminf_N \frac{N}{n} \le \limsup_N \frac{N}{n} < \infty$*,*

$$\mathbb{E}[I_N(x)] - \bar{I}_N(x) \xrightarrow[N\to\infty]{} 0.$$

Deterministic equivalents for another random matrix model with numerous applications to wireless communications are provided by the next theorem. This model is the so-called Kronecker model where random matrices with independent entries are multiplied from the left and right side by deterministic correlation matrices. This result is fundamental for the analysis of the double-scattering channel model in Section 3.4 and the proof of Theorem 22.

Theorem 13 ([83, Corollary 1 and Theorem 2]). *For $k \in \{1, \ldots, K\}$, let $\mathbf{R}_k \in \mathbb{C}^{N\times N}$, $\mathbf{T}_k \in \mathbb{C}^{n_k\times n_k}$, and $\mathbf{D}_N \in \mathbb{C}^{N\times N}$ be Hermitian nonnegative definite, satisfying $\limsup_N \|\mathbf{R}_k\| < \infty$, $\limsup_N \|\mathbf{T}_k\| < \infty$, and $\limsup_N \|\mathbf{D}_N\| < \infty$. Let $\mathbf{X}_k \in \mathbb{C}^{N\times n_k}$ have i.i.d. elements satisfying $\mathbb{E}[X_{i,j}] = 0$, $\mathbb{E}\left[|X_{i,j}|^2\right] = \frac{1}{n_k}$, and $\mathbb{E}\left[|\sqrt{n_k}X_{i,j}|^8\right] < \infty$. Denote $\mathbf{B}_N = \sum_k \mathbf{R}_k^{1/2}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathsf{H}}\mathbf{R}_k^{1/2}$ and, for $x > 0$, define $I_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N\right)$. Let $c_k = \frac{n_k}{N}$ and assume that $n_k, N \to \infty$, such that that $0 < \liminf_N c_k \le \limsup_N c_k < \infty\ \forall k$.*

(i) The following set of K equations $(1 \le k \le K)$,

$$\bar{e}_k = \frac{1}{n_k}\operatorname{tr}\mathbf{T}_k\left(e_k\mathbf{T}_k + \mathbf{I}_{n_k}\right)^{-1}$$
$$e_k = \frac{1}{n_K}\operatorname{tr}\mathbf{R}_k\left(\sum_{i=1}^{K}\bar{e}_i\mathbf{R}_i + x\mathbf{I}_N\right)^{-1}$$

has a unique solution such that $\bar{e}_k, e_k > 0$ for all k.

(ii) $\frac{1}{N}\operatorname{tr}\mathbf{D}_N\left(\mathbf{B}_N + x\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{D}_N\left(\sum_{i=1}^{K}\bar{e}_i\mathbf{R}_{i,N} + x\mathbf{I}_N\right)^{-1} \xrightarrow[N\to\infty]{a.s.} 0$

(iii) $I_N(x) - \bar{I}_N(x) \xrightarrow[N\to\infty]{a.s.} 0$, *where*

$$\bar{I}_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\bar{e}_{k,N}\mathbf{R}_{k,N}\right) + \sum_{k=1}^{K}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + e_{k,N}\mathbf{T}_{k,N}\right) - \frac{1}{N}\sum_{k=1}^{K}n_k e_{k,N}\bar{e}_{k,N}.$$

The next theorem extends Theorem 13 to a more general class of random matrices where each column of $\mathbf{X}$ can have a different covariance matrix. Moreover, a deterministic matrix $\mathbf{S}$ is added. This matrix model will be used extensively in Sections 3.2, 3.3 and 3.6. Additionally, we derive deterministic equivalents of the matrix moments in Theorem 20.

Theorem 14 ([70, Theorem 1],[84, Theorem 2.3]). *Let $\mathbf{B}_N = \mathbf{X}\mathbf{X}^{\mathsf{H}} + \mathbf{S}_N$, where $\mathbf{X} \in \mathbb{C}^{N\times n}$ is random and $\mathbf{S}_N \in \mathbb{C}^{N\times N}$ is Hermitian nonnegative definite. The jth column $\mathbf{x}_j$ of $\mathbf{X}$ is given as $\mathbf{x}_j = \tilde{\mathbf{R}}_j\mathbf{z}_j$, where $\mathbf{z}_j = [z_{j,1}, \ldots, z_{j,N}]^{\mathsf{T}} \in \mathbb{C}^N$ has i.i.d. elements and $\tilde{\mathbf{R}}_j \in \mathbb{C}^{N\times N}$ is deterministic. Denote $\mathbf{R}_j = \tilde{\mathbf{R}}_j\tilde{\mathbf{R}}_j^{\mathsf{H}}$ and assume that $\mathbb{E}[z_{j,i}] = 0$, $\mathbb{E}\left[|z_{j,i}|^2\right] = \frac{1}{N}$, $\mathbb{E}\left[|\sqrt{N}z_{i,j}|^8\right] < \infty$, and $\limsup_N \|\mathbf{R}_j\| < \infty$. Let $\mathbf{D}_N \in \mathbb{C}^{N\times N}$ be a deterministic Hermitian which satisfies $\limsup_N \|\mathbf{D}_N\| < \infty$. Then, as $n, N \to \infty$ such that $0 < \liminf N/n \le \limsup N/n < \infty$, the following holds for any $z \in \mathbb{C}\setminus\mathbb{R}^+$:*

(i) The following set of n equations ($1 \le j \le n$),

$$\delta_j(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\frac{1}{N} \sum_{k=1}^{n} \frac{\mathbf{R}_k}{1+\delta_k(z)} + \mathbf{S}_N - z\mathbf{I}_N \right)^{-1} \tag{2.8}$$

has a unique solution such that $(\delta_1(z), \ldots, \delta_n(z))$ *are Stieltjes transforms of nonnegative finite measures on* $\mathbb{R}^+$. *For* $z < 0$, $\delta_1(z), \ldots, \delta_{N,n}(z)$ *are the unique nonnegative solutions to* (2.8) *and can be obtained by a standard fixed-point algorithm with initial values* $\delta_j^{(0)}(z) > 0$ *for* $j = 1, \ldots, n$.

(ii) $\frac{1}{N} \operatorname{tr} \mathbf{D}_N (\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}_N(z) \xrightarrow{\text{a.s.}} 0$, *where*

$$\mathbf{T}_N(z) = \left(\frac{1}{N} \sum_{j=1}^{n} \frac{\mathbf{R}_j}{1+\delta_j(z)} + \mathbf{S}_N - z\mathbf{I}_N \right)^{-1}.$$

(iii) Let F_N *be the e.s.d. of* $\mathbf{B}_N$ *and denote by* $\bar{F}_N$ *the distribution function with Stieltjes transform* $\frac{1}{N} \operatorname{tr} \mathbf{T}_N(z)$. *Then, almost surely,*

$$F_N - \bar{F}_N \Rightarrow 0.$$

(iv) For $x > 0$, *let* $I_N(x) = \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N\right)$. *Then,*

$$\mathbb{E}\left[I_N(x)\right] - \bar{I}_N(x) \to 0$$

where

$$\bar{I}_N(x) = \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{S}_N + \frac{1}{x}\frac{1}{N} \sum_{j=1}^{n} \frac{\mathbf{R}_j}{1+\delta_j(-x)} \right) + \frac{1}{N} \sum_{j=1}^{n} \log\left(1+\delta_j(-x)\right) - \frac{1}{N} \sum_{j=1}^{n} \frac{\delta_j(-x)}{1+\delta_j(-x)}.$$

2.3.2 New results

In this section, we present several new deterministic equivalents which were derived in the context of different application scenarios. Let us begin with the following theorem which can be seen as an analogous result to Theorem 13 where the matrices $\mathbf{X}_k$ are replaced by Haar matrices. Applications of this result to the performance analysis of random beamforming over quasi-static channels can be found in Section 3.6.

Theorem 15 ([85, Theorem 7])**.** *For* $i \in \{1, \ldots, K\}$, *let* $\mathbf{P}_i \in \mathbb{C}^{n_i \times n_i}$ *be Hermitian nonnegative, satisfying* $\limsup_{n_i} \|\mathbf{P}_i\| < \infty$, *and let* $\mathbf{W}_i \in \mathbb{C}^{N_i \times n_i}$ *be* $n_i < N_i$ *columns of a Haar distributed random matrix. Let* $\mathbf{H}_i \in \mathbb{C}^{N \times N_i}$ *be a random matrix such that* $\mathbf{R}_i \triangleq \mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} \in \mathbb{C}^{N \times N}$ *satisfies* $\limsup_N \|\mathbf{R}_i\| < \infty$, *almost surely. Define* $c_i = \frac{n_i}{N_i}$, $\bar{c}_i = \frac{N_i}{N}$,

$$\mathbf{B}_N = \sum_{i=1}^{K} \mathbf{H}_i \mathbf{W}_i \mathbf{P}_i \mathbf{W}_i^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}}$$

and denote F_N *the e.s.d. of* $\mathbf{B}_N$.

(i) For $z \in D \triangleq \{z = x + \mathbf{i}y : x < 0, |y| \leq |x|\frac{1-c_i}{c_i}\}$*, the following system of* $2K$ *equations* $(1 \leq i \leq K)$

$$\bar{e}_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(e_i(z)\mathbf{P}_i + [\bar{c}_i - e_i(z)\bar{e}_i(z)]\mathbf{I}_{n_i}\right)^{-1}$$
$$e_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{e}_j(z)\mathbf{R}_j - z\mathbf{I}_N\right)^{-1} \tag{2.9}$$

has a unique solution such that $(e_1(z) \ldots, e_K(z))$ *are Stieltjes transforms of finite nonnegative measures over* $\mathbb{R}^+$ *which satisfy for* $z < 0$, $0 \leq e_i(z) < c_i\bar{c}_i/\bar{e}_i(z)$ $\forall i$*, where they are explicitly given by*

$$\bar{e}_i(z) = \lim_{t\to\infty} \bar{e}_i^{(t)}(z)$$
$$e_i(z) = \lim_{t\to\infty} e_i^{(t)}(z)$$
$$\bar{e}_i^{(t)}(z) = \lim_{k\to\infty} \bar{e}_i^{(t,k)}(z)$$

where for $k \geq 1$,

$$e_i^{(t)}(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{e}_j^{(t-1)}(z)\mathbf{R}_j - z\mathbf{I}_N\right)^{-1}$$
$$\bar{e}_i^{(t,k)}(z) = \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(e_i^{(t)}(z)\mathbf{P}_i + [\bar{c}_i - e_i^{(t)}(z)\bar{e}_i^{(t,k-1)}(z)]\mathbf{I}_{n_i}\right)^{-1}$$

with the initial values $\bar{e}_i^{(t,0)}(z) = 0$ *and* $e_i^{(0)}(z) = 0$ $\forall i$.

(ii) Assume that $N \to \infty$*, such that* $0 < \liminf \bar{c}_i \leq \limsup \bar{c}_i < \infty$ *and* $0 \leq \liminf c_i \leq \limsup c_i < 1$ *for all* i*. Then, almost surely,*

$$F_N - \bar{F}_N \Rightarrow 0$$

where $\bar{F}_N$ *is the distribution function whose Stieltjes transform* $\bar{m}_N(z)$ *is defined for* $z \in D$ *as*

$$\bar{m}_N(z) = \frac{1}{N} \operatorname{tr} \left(\sum_{i=1}^{K} \bar{e}_i(z)\mathbf{R}_i - z\mathbf{I}_N\right)^{-1}.$$

Proof. The proof is given in Appendix 2.5.2. □

Remark 7. *An important aspect of Theorem 15 (i) is that we provide an explicit algorithm to compute the solution of the fundamental equations* (2.9) *for* $z < 0$*. This is achieved by proving that the implicit equations in* $e_i(z)$ (2.9) *belong to the class of so-called* standard interference functions*, defined as follows:*

Definition 11 (Standard interference function [86])**.** *A* K*-variate function* $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_K(\mathbf{x})]^\mathsf{T} \in \mathbb{R}^K$ *for* $\mathbf{x} \in \mathbb{C}^K$ *is said to be standard if it fulfills the following conditions:*

1. *Positivity: if* $\mathbf{x} \geq 0$*, then* $\mathbf{h}(\mathbf{x}) > 0$*;*

2. *Monotonicity: if* $\mathbf{x} \geq \mathbf{x}'$*, then* $\mathbf{h}(\mathbf{x}) \geq \mathbf{h}(\mathbf{x}')$*;*

3. *Scalability: if* $\alpha > 1$*, then* $\alpha\mathbf{h}(\mathbf{x}) > \mathbf{h}_j(\alpha\mathbf{x})$*.*

The next theorem then ensures that the solution to (2.9) *can be computed by a standard fixed-point algorithm.*

Theorem 16 (Fixed-point theorem [86, Theorem 2])**.** *If a K-variate function* $\mathbf{h}(\mathbf{x})$ *is standard and there exists* $\mathbf{x}$ *such that* $\mathbf{x} \geq \mathbf{h}(\mathbf{x})$*, then the algorithm that consists in setting*

$$\mathbf{x}^{(t+1)} = \mathbf{h}\left(\mathbf{x}^{(t)}\right), \qquad t \geq 1$$

for any initial value $\mathbf{x}^{(0)} \geq 0$*, converges to the unique fixed point of* $\mathbf{x} = \mathbf{h}(\mathbf{x})$*.*

This result ensures the uniqueness of solutions to (2.9) *for* $z < 0$*. However, since the functions* $e_i(z)$ *have analytic extensions on* $\mathbb{C} \setminus \mathbb{R}^+$ *which satisfy the fundamental equations for* $z \in D$ *and can be shown to be Stieltjes transforms of finite nonnegative measures on* $\mathbb{R}^+$*, it follows by the identity theorem that* (2.9) *has also a unique functional solution:*

Theorem 17 (Identity theorem [87, Theorem 3.2.6])**.** *Let* $f(z)$ *and* $g(z)$ *be two functions, analytic in a common domain* D*. If* $f(z)$ *and* $g(z)$ *coincide on* $D' \subset D$ *and* D' *has a limit point, then* $f(z) = g(z)$ *everywhere in* D*.*

The approach of proving the point-wise and functional uniqueness of fundamental equations via standard interference functions and the identity theorem is an original contribution of this thesis which has been successfully applied to various random matrix models (see also Theorems 22, 23, 24). Moreover, having an explicit algorithm to compute such solutions is of practical importance.

The next result provides a deterministic equivalent of the (ergodic) mutual information associated with the matrix model of Theorem 15.

Theorem 18 ([85, Theorem 4])**.** *For* $\mathbf{B}_N$ *as defined in Theorem 15, denote* $I_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N\right)$ *for* $x > 0$*. Assume that* $N \to \infty$*, such that* $0 < \liminf \bar{c}_i \leq \limsup \bar{c}_i < \infty$ *and* $0 \leq \liminf c_i \leq \limsup c_i < 1$ *for all* i*. Then,*

$$\begin{aligned} (i) \quad & \mathbb{E}\left[I_N(x)\right] - \bar{I}_N(x) \to 0 \\ (ii) \quad & I_N(x) - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0 \end{aligned}$$

where

$$\begin{aligned} \bar{I}_N(x) = & \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\bar{e}_k(-x)\mathbf{R}_k\right) \\ & + \sum_{k=1}^{K}\frac{1}{N}\log\det\left([\bar{c}_k - e_k(-x)\bar{e}_k(-x)]\mathbf{I}_{n_k} + e_k(-x)\mathbf{P}_k\right) \\ & + \sum_{k=1}^{K}(1-c_k)\bar{c}_k\log(\bar{c}_k - e_k(-x)\bar{e}_k(-x)) - \bar{c}_k\log(\bar{c}_k) \end{aligned}$$

with $e_k(-x)$*,* $\bar{e}_k(-x)$ $\forall k$ *as given by Theorem 15 (i).*

Proof. The proof is provided in Appendix 2.5.3. □

Based on the relation of the Stieltjes transform of a distribution function and its moments (see Theorem 4 and Remark 5), one can derive the following result from Theorem 14 (*ii*):

Theorem 19. *Let $\mathbf{B}_N$ be defined as in Theorem 14 for $\mathbf{S}_N = \mathbf{0}$ and let $\bar{F}_N$ be the distribution function as defined in Theorem 14 (iii). Denote by $\bar{M}_k \triangleq \int_0^\infty \lambda^k d\bar{F}_N(\lambda)$ the kth moment of $\bar{F}_N$. Then,*

$$\bar{M}_k = \frac{(-1)^k}{k!}\frac{1}{N}\operatorname{tr}\mathbf{T}_{N,k}, \qquad n \geq 0$$

where $\mathbf{T}_{N,k}, k \geq 0$ is defined recursively by the following set of equations:

$$\begin{aligned}
\mathbf{T}_{N,k+1} =& \sum_{i=0}^{k}\sum_{j=0}^{i}\binom{k}{i}\binom{i}{j}\mathbf{T}_{N,k-i}\mathbf{Q}_{i-j+1}\mathbf{T}_{N,j} \\
\mathbf{Q}_{k+1} =& \frac{k+1}{N}\sum_{j=1}^{n} f_{j,k}\mathbf{R}_j \\
f_{j,k+1} =& \sum_{i=0}^{k}\sum_{l=0}^{i}\binom{k}{i}\binom{i}{l}(k-i+1)f_{j,l}f_{j,i-l}\delta_{j,k-i} \quad , \qquad 1 \leq j \leq n \\
\delta_{j,k+1} =& \frac{1}{N}\operatorname{tr}\mathbf{R}_j\mathbf{T}_{N,k+1} \quad , \qquad 1 \leq j \leq n
\end{aligned}$$

with the initial values $\mathbf{T}_{N,0} = \mathbf{I}_N$, $f_{j,0} = -1$ and $\delta_{j,0} = \frac{1}{N}\operatorname{tr}\mathbf{R}_j$ $\forall j$.

Proof. The proof can be found in Appendix 2.5.4. □

Remark 8. *While Theorem 19 allows us to compute the moments $\bar{M}_k$ of $\bar{F}_N$, it does not imply that the difference between the empirical moments $M_k \triangleq \frac{1}{N}\operatorname{tr}\mathbf{B}_N^k = \int_0^\infty \lambda^k dF_N(\lambda)$ and $\bar{M}_k$ converges almost surely to zero. The next theorem provides some sufficient conditions for which this convergence holds.*

Remark 9. *Although difficult to show analytically, one can verify numerically that Theorem 19 coincides with [88, Theorem 1] for $\mathbf{R}_j = \operatorname{diag}(r_{1j}, \ldots, r_{Nj})$ $\forall j$. Moreover, for $\mathbf{R}_j = \mathbf{I}_N$ our result can be shown to coincide with Theorem 11. Note that we assume a normalization of the variance of the matrix entries by $1/N$ while Theorem 11 assumes a normalization of $1/n$.*

If the matrices $\mathbf{R}_j$ are drawn from a finite set of matrices, we get the following stronger result:

Theorem 20. *For fixed $L > 0$, let $\mathcal{R} = \{\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_L\}$ be a set of complex $N \times N$ matrices and let $\mathbf{D}_N \in \mathbb{C}^{N\times N}$ be nonnegative definite Hermitian. Consider the matrix $\mathbf{B}_N$ as defined in Theorem 14 and assume that $\mathbf{R}_j \in \mathcal{R}$ $\forall j$. Assume that $\limsup_N \|\mathbf{D}_N\| < \infty$, $\limsup_N \max_l \|\tilde{\mathbf{R}}_l\| < \infty$, and that $N, n \to \infty$, such that $0 < \liminf \frac{n}{N} \leq \limsup \frac{n}{N} < \infty$. Then,*

$$\frac{1}{N}\operatorname{tr}\mathbf{D}_N\mathbf{B}_N^k - \frac{(-1)^k}{k!}\frac{1}{N}\operatorname{tr}\mathbf{D}_N\mathbf{T}_{N,k} \xrightarrow{a.s.} 0, \qquad k \geq 0$$

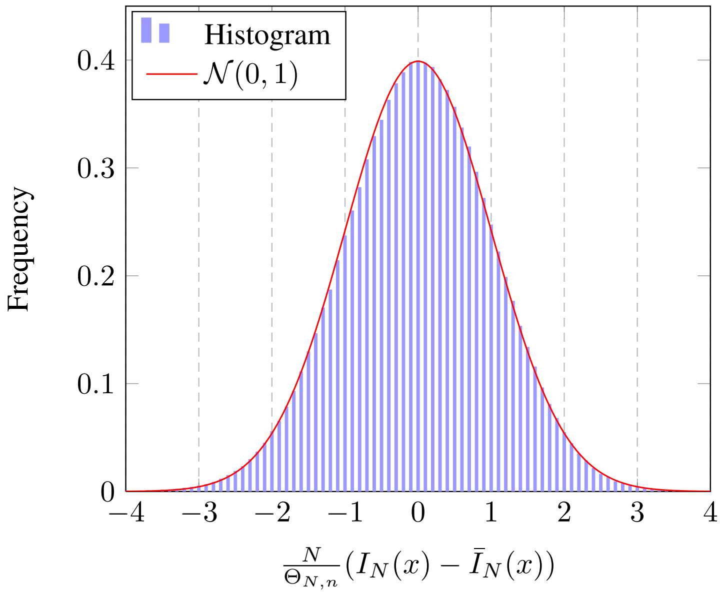

Figure 2.1: Histogram of the random variable $N(I_N(x)-\bar{I}_N(x))\Theta_{N,n}^{-1}$ versus the standard normal distribution $\mathcal{N}(0,1)$ for $x = -5\,\mathrm{dB}$.

where $\mathbf{T}_{N,k}$ is given by Theorem 19. This implies in particular,

$$\frac{1}{N}\operatorname{tr}\mathbf{B}_N^k - \bar{M}_k \xrightarrow{a.s.} 0, \qquad k \geq 0.$$

Proof. The proof is provided in Appendix 2.5.5. □

In the next claim, we provide the fluctuations of the mutual information for the Rician matrix model of Theorem 12 in form of a CLT .

Claim 1 ([89, Theorem 3], [90, Claim 1])**.** *Under the assumptions of Theorem 12 (ii) and the condition that $X_{i,j} \sim \mathcal{CN}\left(\frac{\sigma_{i,j}^2}{n}\right)$, the following holds:*

$$\frac{N}{\Theta_{N,n}}\left(I_N(x) - \bar{I}_N(x)\right) \Rightarrow \mathcal{N}\left(0,1\right)$$

where

$$\Theta_{N,n}^2 = -\log\det(\mathbf{I}_{N+n} - \mathbf{J})$$

and the matrix $\mathbf{J} \in \mathbb{C}^{(N+n)\times(N+n)}$ is defined as

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{J}_2 \\ \mathbf{J}_3 & \mathbf{J}_4 \end{pmatrix}$$

where $\mathbf{J}_1 \in \mathbb{C}^{n\times n}$, $\mathbf{J}_2 \in \mathbb{C}^{n\times N}$, $\mathbf{J}_3 \in \mathbb{C}^{N\times n}$, *and* $\mathbf{J}_4 \in \mathbb{C}^{N\times N}$ *have entries*

$$
\begin{aligned}
[\mathbf{J}_1]_{k,m} &= \frac{1}{n(1+\delta_m)^2}\mathbf{a}_m^{\mathsf{H}}\mathbf{T}(-x)\mathbf{D}_k\mathbf{T}(-x)\mathbf{a}_m\\
[\mathbf{J}_2]_{k,m} &= -\frac{x}{n}\mathbf{t}_m^{\mathsf{H}}\mathbf{D}_k\mathbf{t}_m\\
[\mathbf{J}_3]_{k,m} &= -\frac{x}{n}\widetilde{\mathbf{t}}_m^{\mathsf{H}}\widetilde{\mathbf{D}}_k\widetilde{\mathbf{t}}_m\\
[\mathbf{J}_4]_{k,m} &= \frac{1}{n(1+\widetilde{\delta}_m)^2}\mathbf{b}_m^{\mathsf{H}}\widetilde{\mathbf{T}}(-x)\widetilde{\mathbf{D}}_k\widetilde{\mathbf{T}}(-x)\mathbf{b}_m
\end{aligned}
$$

and $\mathbf{a}_i$, $\mathbf{b}_i$, $\mathbf{t}_i$, *and* $\widetilde{\mathbf{t}}_i$ *denote the columns of* $\mathbf{A}$, $\mathbf{A}^{\mathsf{H}}$, $\mathbf{T}(-x)$, *and* $\widetilde{\mathbf{T}}(-x)$.

Proof. A justification of this claim is provided in Appendix 2.5.6. □

Remark 10. *Consider the matrix model* $\mathbf{B}'_N = (\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathsf{H}} + \mathbf{S}$ *where* $\mathbf{S} \in \mathbb{C}^{N\times N}$ *is a deterministic Hermitian nonnegative matrix. Define* $\mathbf{X}' = (\mathbf{X}\ \mathbf{0}_{N\times N})$ *and* $\mathbf{A}' = \left(\mathbf{A}\ \mathbf{S}^{1/2}\right)$. *Then* $\mathbf{B}'_N = (\mathbf{X}'+\mathbf{A}')(\mathbf{X}'+\mathbf{A}')^{\mathsf{H}}$. *Since the matrices* $\mathbf{X}'+\mathbf{A}'$ *and* $\mathbf{X}+\mathbf{A}$ *are of the same form, both Theorem 12 and Claim 1 also hold for this extended matrix model. This is useful for the analysis of channels with colored noise (see [90]).*

In order to give some further validation of Claim 1, let us provide a short numerical example. Let $x = -5\,\mathrm{dB}$, $N = 8$, $n = 4$, and let $\mathbf{A} \in \mathbb{C}^{N\times n}$ be a random realization of a complex Gaussian matrix where $\mathbf{A}_{i,j} \sim \mathcal{CN}(0, 1/n)$. Further, let $\sigma^2_{i,j}$ be a random realization of a uniformly distributed random variable over the interval $[0,1]$. Fig. 2.1 compares the empirical histogram of $N(I_N(x) - \bar{I}_N(x))\Theta^{-1}_{N,n}$ as obtained by Monte Carlo simulations against the standard normal distribution. The overlap between both results is surprisingly good for the rather small matrix dimensions 8×4. Also note that Claim 1 was used in [89] to calculate an approximation of the outage probability in a cooperative small cell network with Rician fading channels.

The last theorem of this section provides a deterministic equivalent of a slightly more involved type of functionals of random matrices. This result will be needed in Section 3.3 for the performance analysis of linear detectors and precoders in large-scale MIMO systems.

Theorem 21. *Let* $\mathbf{\Theta}_N \in \mathbb{C}^{N\times N}$ *be a Hermitian nonnegative definite matrix satisfying* $\limsup_N \|\mathbf{\Theta}_N\| < \infty$. *Then, under the same conditions as in Theorem 14, the following holds true for* $z < 0$:

$$
\frac{1}{N}\operatorname{tr}\mathbf{D}_N\left(\mathbf{B}_N+\mathbf{S}_N - z\mathbf{I}_N\right)^{-1}\mathbf{\Theta}_N\left(\mathbf{B}_N+\mathbf{S}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{D}_N\mathbf{T}'_N(z) \xrightarrow{a.s.} 0
$$

where

$$
\mathbf{T}'_N(z) = \mathbf{T}_N(z)\mathbf{\Theta}_N\mathbf{T}_N(z) + \mathbf{T}_N(z)\frac{1}{N}\sum_{j=1}^{n}\frac{\mathbf{R}_j\delta'_j(z)}{\left(1+\delta_j(z)\right)^2}\mathbf{T}_N(x)
$$

with $\mathbf{T}_N(z)$, $\delta_j(z)\ \forall j$ *as defined in Theorem 14 (i) and* $\boldsymbol{\delta}'(z) = [\delta'_1(z)\cdots\delta'_K(z)]^{\mathsf{T}}$

given by

$$\begin{aligned}\boldsymbol{\delta}'(z) &= \left(\mathbf{I}_K - \mathbf{J}(z)\right)^{-1}\mathbf{v}(z)\\ [\mathbf{J}(z)]_{kl} &= \frac{\frac{1}{N}\operatorname{tr}\mathbf{R}_k\mathbf{T}_N(z)\mathbf{R}_l\mathbf{T}_N(z)}{N\left(1+\delta_k(z)\right)^2}\\ [\mathbf{v}(z)]_k &= \frac{1}{N}\operatorname{tr}\mathbf{R}_k\mathbf{T}_N(z)\boldsymbol{\Theta}_N\mathbf{T}_N(z)\end{aligned}$$

where $\mathbf{J}(z) \in \mathbb{C}^{K\times K}$ *and* $\mathbf{v}(z) \in \mathbb{C}^K$*.*

Proof. The proof is given in Appendix 2.5.7. □

2.4 Iterative deterministic equivalents

2.4.1 Definition and motivation

Deterministic equivalents are convenient to study the performance of wireless communication systems when a single system parameter can be modeled by a random matrix, e.g., the fading channel or a precoding matrix. In order to tackle the performance analysis of more complex systems which are characterized by functionals of several random matrices, e.g., products, it is necessary to extend the notion of deterministic equivalents. In this section, we develop a systematic approach to generalize deterministic equivalents (see Definition 10) to *iterative deterministic equivalents*.

At the heart of the concept of iterated deterministic equivalents is the Fubini theorem (see Theorem 3 in Section 2.1). Let $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$ be two probability spaces and denote $(\Omega \times \Omega', \mathcal{F} \times \mathcal{F}', Q)$ their product space. Let us consider now a set $A \in \mathcal{F} \times \mathcal{F}'$. Then, we have from Theorem 3 that

$$\begin{aligned}Q(A) &= \int_{\Omega\times\Omega'} \mathbb{1}_A(\omega,\omega')dQ(\omega,\omega')\\ &= \int_{\Omega'}\left[\int_{\Omega}\mathbb{1}_A(\omega,\omega')dP(\omega)\right]dP'(\omega'). \end{aligned}\tag{2.10}$$

Equation (2.10) is the core ingredient for the definition of iterative deterministic equivalents: Let $(\mathbf{H}_n(\omega))_{n\geq 1}$ and $(\mathbf{H}'_n(\omega'))_{n\geq 1}$ be two series of random matrices generated by the spaces $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$, respectively, and denote by Q the product-space measure. Let $f_n((\mathbf{H}_n(\omega), \mathbf{H}'_n(\omega')), z)$ be a functional of the matrices $\mathbf{H}_n(\omega)$ and $\mathbf{H}'_n(\omega')$. Assume that there is a function $\tilde{g}_n(\mathbf{H}_n(\omega), z)$, such that, for each $\omega \in A \subset \Omega$ with $P(A) = 1$, there exists a subset $B(\omega) \subset \Omega'$ with $P'(B(\omega)) = 1$, on which

$$f_n((\mathbf{H}_n(\omega), \mathbf{H}'_n(\omega')), z) - \tilde{g}_n(\mathbf{H}_n(\omega), z) \to 0. \tag{2.11}$$

Although $\tilde{g}_n(\mathbf{H}_n(\omega), z)$ is a random function (as it depends on ω), it is independent of $\mathbf{H}'_n(\omega')$. Thus, we can see $\tilde{g}_n(\mathbf{H}_n(\omega), z)$ as a deterministic equivalent of $f_n((\mathbf{H}_n(\omega), \mathbf{H}'_n(\omega')), z)$ with respect to $(\mathbf{H}'_n(\omega'))_{n\geq 1}$. Now, let us assume that there is a second function $g_n(z)$, such that for $\omega \in C \subset \Omega$ with $P(C) = 1$,

$$\tilde{g}_n(\mathbf{H}_n(\omega), z) - g_n(z) \to 0. \tag{2.12}$$

Call $D = \{(\omega, \omega') : \omega \in A \cap C \ , \ \omega' \in B(\omega)\} \subset \Omega \times \Omega'$, the space on which $f_n((\mathbf{H}_n, \mathbf{H}'_n), z) - g_n(z) \to 0$. Then, from (2.10), this space has probability

$$\begin{aligned} Q(D) &= \int_\Omega \left[\int_{\Omega'} \mathbb{1}_D(\omega, \omega') dP'(\omega') \right] dP(\omega) \\ &\overset{(a)}{\geq} \int_{A \cap C} \left[\int_{B(\omega)} \mathbb{1}_D(\omega, \omega') dP'(\omega') \right] dP(\omega) \\ &\overset{(b)}{=} \int_{A \cap C} dP(\omega) \\ &\overset{(c)}{=} 1 \end{aligned} \tag{2.13}$$

where (a) is due to $A \cap C \subset \Omega$ and $B(\omega) \subset \Omega'$, (b) follows since $P'(B(\omega)) = 1$ for $\omega \in A$ and (c) holds since $P(A \cap C) = P(A) + P(C) - P(A \cup C) = 1$.

To summarize, if a deterministic equivalent g_n exists for a functional f_n of a *random* series $(\mathbf{H}'_n)_{n \geq 1}$ and a *deterministic* series $(\mathbf{H}_n)_{n \geq 1}$ of matrices, and if additionally it can be proved that this deterministic equivalent holds true for almost every such $(\mathbf{H}_n)_{n \geq 1}$ generated by a space Ω, then the latter is also a deterministic equivalent for the random series $((\mathbf{H}_n, \mathbf{H}'_n))_{n \geq 1}$.

This is the mathematical key idea behind our method to derive iterative deterministic equivalents of functionals $f_n((\mathbf{H}_n(\omega), \mathbf{H}'_n(\omega')), z)$ of two (or more) random matrices. First, one considers one of the sequences of random matrices, e.g., $(\mathbf{H}_n(\omega))_{n \geq 1}$, to be deterministic and derives a deterministic equivalent with respect to $(\mathbf{H}'_n(\omega'))_{n \geq 1}$. In the example above, this was the role of the functional $\tilde{g}_n(\mathbf{H}_n(\omega), z)$ which is independent of $\mathbf{H}'_n(\omega')$. In a second step, one assumes the matrices $(\mathbf{H}_n(\omega))_{n \geq 1}$ to be random and derives an iterative deterministic equivalent $g_n(z)$ of $\tilde{g}_n(\mathbf{H}_n(\omega), z)$. Of course, this procedure can be carried out for any finite number of random matrices where in each step the "randomness" related to one of the matrices is removed. From the above construction, we will call $(g_n)_{n \geq 1}$ an *iterative deterministic equivalent*.

As application examples, we will provide deterministic equivalents of the (ergodic) capacity as well as the sum-rate with MMSE detection of double-scattering MACs (Section 3.4), deterministic equivalents of the mutual information of the multi-hop AF MIMO relay channel (Section 3.5), and an asymptotic performance analysis of random beamforming over correlated fading channels (Section 3.6). These applications are based on the novel theoretical results, summarized in the next section. We recall that all matrices should be understood as sequences of matrices with growing dimensions. For notational convenience, we drop the index n, e.g., we write $\mathbf{H}$ instead of $(\mathbf{H}_n)_{n \geq 1}$.

2.4.2 Results

Our first result is related to the double-scattering channel model as introduced in [91]. We provide a set of fundamental equations whose solutions are needed to compute a deterministic equivalent of the mutual information. The analysis is based on the crucial observation that this random matrix model can be seen as the Kronecker model considered in Theorem 13 with random correlation matrices.

Theorem 22 ([92, Theorems 4 and 5])**.** *For $k \in \{1, \dots, K\}$, let $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ be defined as*

$$\mathbf{H}_k = \frac{1}{\sqrt{N_k n_k}} \mathbf{R}_k^{\frac{1}{2}} \mathbf{W}_{1,k} \mathbf{S}_k^{\frac{1}{2}} \mathbf{W}_{2,k} \mathbf{T}_k^{\frac{1}{2}}$$

where $\mathbf{R}_k \in \mathbb{C}^{N \times N}$ and $\mathbf{T}_k \in \mathbb{C}^{n_k \times n_k}$ are Hermitian nonnegative definite, $\mathbf{S}_k = \mathrm{diag}(s_{k,1}, \dots, s_{k,N_k}) \in (\mathbb{R}^+)^{N_k \times N_k}$, and $\mathbf{W}_{1,k} \in \mathbb{C}^{N \times N_k}$, $\mathbf{W}_{2,k} \in \mathbb{C}^{N_k \times n_k}$ are independent standard complex Gaussian matrices. Let $\mathbf{Q}_k \in \mathbb{C}^{n_k \times n_k}$ be Hermitian nonnegative definite, $x > 0$, and assume that $\limsup_N \|\mathbf{R}_k\| < \infty$, $\limsup_N \|\mathbf{S}_k\| < \infty$ and $\limsup_N \|\mathbf{T}_k \mathbf{Q}_k\| < \infty$.

(i) The following system of $3K$ implicit equations ($1 \le k \le K$):

$$\begin{aligned}
\bar{g}_k &= \frac{1}{n_k} \mathrm{tr}\, \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}} \left(g_k \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}} + \mathbf{I}_{n_k} \right)^{-1} \\
g_k &= \frac{1}{n_k} \sum_{j=1}^{N_k} \frac{s_{k,j} \delta_k}{1 + \bar{g}_k s_{k,j} \delta_k} \qquad (2.14) \\
\delta_k &= \frac{1}{N_k} \mathrm{tr}\, \mathbf{R}_k \left(\sum_{i=1}^{K} \frac{n_i}{N_i} \frac{\bar{g}_i g_i}{\delta_i} \mathbf{R}_i + x \mathbf{I}_N \right)^{-1}
\end{aligned}$$

has a unique solution satisfying $\bar{g}_k, g_k, \delta_k > 0\ \forall k$.

(ii) Let $\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^{\mathsf{H}}$ and denote $I_N(x) = \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N\right)$. Assume that N, N_k, and n_k grow infinitely large, satisfying $0 < \liminf \frac{N_k}{N} \le \limsup \frac{N_k}{N} < \infty$ and $0 < \liminf \frac{n_k}{N} \le \limsup \frac{n_k}{N} < \infty\ \forall k$. Then,

$$\begin{aligned}
&(a) \quad I_N(x) - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0 \\
&(b) \quad \mathbb{E}\left[I_N(x)\right] - \bar{I}_N(x) \to 0
\end{aligned}$$

where

$$\begin{aligned}
\bar{I}_N(x) = \frac{1}{N} \log\det &\left(\mathbf{I}_N + \frac{1}{x} \sum_{k=1}^{K} \frac{n_k}{N_k} \frac{\bar{g}_k g_k}{\delta_k} \mathbf{R}_k \right) \\
&+ \frac{1}{N} \sum_{k=1}^{K} \log\det\left(\mathbf{I}_{N_k} + \bar{g}_k \delta_k \mathbf{S}_k\right) \\
&+ \frac{1}{N} \sum_{k=1}^{K} \left[\log\det\left(\mathbf{I}_{n_k} + g_k \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}} \right) - 2 n_k g_k \bar{g}_k \right].
\end{aligned}$$

Proof. The proof is given in Appendix 2.5.8. □

Remark 11. *The values of $\bar{g}_k$, g_k, and δ_k can be computed by a standard fixed-point algorithm which iteratively computes (2.14), starting from some arbitrary initialization $\bar{g}_k^{(0)}, g_k^{(0)}, \delta_k^{(0)} > 0$. This algorithm is proved to converge, generally terminates within a few iterations (depending on the system size and the desired accuracy), and does not pose any computational challenge.*

Our next result is an extension of Theorem 15 to the case where the matrices $\mathbf{W}_k$ *and* $\mathbf{H}_k$ are random. This is an example which demonstrates that iterative deterministic equivalents can be also computed for combinations of random correlated and random unitary matrices. Further practical applications of this result can be found in Section 3.6.

Theorem 23 ([85, Theorem 2 and 4]). *For $k \in \{1, \ldots, K\}$, let $\mathbf{P}_k \in \mathbb{C}^{n_k \times n_k}$ be Hermitian nonnegative, satisfying $\limsup_{n_k} \|\mathbf{P}_k\| < \infty$, and let $\mathbf{W}_k \in \mathbb{C}^{N_k \times n_k}$ be $n_k < N_k$ columns of a Haar distributed random matrix. Let $\mathbf{H}_k \in \mathbb{C}^{N \times N_k}$ be a random matrix whose jth column vector $\mathbf{h}_{kj} \in \mathbb{C}^N$ is modeled as $\mathbf{h}_{kj} = \mathbf{R}_{kj}^{1/2}\mathbf{z}_{kj}$, where $\mathbf{R}_{kj} \in \mathbb{C}^{N \times N}$ are Hermitian nonnegative definite and $\mathbf{z}_{kj} \in \mathbb{C}^N$ have i.i.d. elements with zero mean, variance $1/N$ and finite $(4+\epsilon)$th moment, $\epsilon > 0$. Assume that $\limsup_N \max_{k,j} \|\mathbf{R}_{kj}\| < \infty$ and $\limsup_N \max_i \|\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}\| < \infty$, almost surely. Denote $\mathbf{B}_N = \sum_{k=1}^K \mathbf{H}_k\mathbf{W}_k\mathbf{P}_k\mathbf{W}_k\mathbf{H}_k$, $c_k = \frac{n_k}{N_k}$, $\bar{c}_k = \frac{N_k}{N}$ and assume that $N, N_k, n_k \to \infty$, such that $0 \leq \liminf c_k \leq \limsup c_k < 1$ and $0 < \liminf_N \bar{c}_k \leq \limsup_N \bar{c}_k < \infty \ \forall k$.*

(i) Let $x > 0$. Then, the system of equations $(1 \leq k \leq K,\ 1 \leq j \leq N_k)$:

$$\bar{b}_k = \frac{1}{N} \operatorname{tr} \mathbf{P}_k \Big(b_k \mathbf{P}_k + \left[\bar{c}_k - b_k \bar{b}_k \right] \mathbf{I}_{n_k} \Big)^{-1}$$

$$b_k = \frac{1}{N} \sum_{j=1}^{N_k} \frac{\zeta_{kj}}{1 + \bar{b}_k \zeta_{kj}}$$

$$\zeta_{kj} = \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \left(\frac{1}{N} \sum_{i=1}^{K} \sum_{l=1}^{N_k} \frac{\bar{b}_i \mathbf{R}_{i,l}}{1 + \bar{b}_i \zeta_{il}} + x \mathbf{I}_N \right)^{-1}$$

has a unique solution satisfying $\zeta_{kj}, b_k, \bar{b}_k \geq 0$ and $0 \leq b_k \bar{b}_k < c_k \bar{c}_k \ \forall k, j$.

(ii) For $x > 0$, denote $I_N(x) = \frac{1}{N} \log\det \left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N \right)$. Then,

$$(a) \quad I_N(x) - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0$$
$$(b) \quad \mathbb{E}\left[I_N(x) \right] - \bar{I}_N(x) \to 0$$

where

$$\bar{I}_N(x) = \bar{V}_N(x) + \frac{1}{N} \sum_{k=1}^{K} \log\det \left(\left[\bar{c}_k - b_k \bar{b}_k \right] \mathbf{I}_{n_k} + b_k \mathbf{P}_k \right)$$
$$+ \sum_{k=1}^{K} (1 - c_k) \bar{c}_k \log(\bar{c}_k - b_k \bar{b}_k) - \bar{c}_k \log(\bar{c}_k)$$

$$\bar{V}_N(x) = \frac{1}{N} \log\det \left(\mathbf{I}_N + \frac{1}{x} \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{N_k} \frac{\bar{b}_k \mathbf{R}_{k,j}}{1 + \bar{b}_k \zeta_{kj}} \right)$$
$$- \sum_{k=1}^{K} \bar{b}_k b_k + \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{N_k} \log \left(1 + \bar{b}_k \zeta_{kj} \right). \tag{2.15}$$

Proof. The proof is given in Appendix 2.5.9. □

Remark 12. *The solution to the fundamental equations is given explicitly by the following fixed-point algorithm*

$$b_k = \lim_{t\to\infty} g_k^{(t)}, \qquad \bar{b}_k = \lim_{t\to\infty} \bar{b}_k^{(t)}, \qquad \zeta_{kj} = \lim_{t\to\infty} \zeta_{kj}^{(t)}$$

where

$$\begin{aligned}
\bar{b}_k^{(t)} &= \lim_{l\to\infty} \bar{b}_k^{(t,l)}, \quad \zeta_{kj}^{(t)} = \lim_{l\to\infty} \zeta_{kj}^{(t,l)} \\
b_k^{(t)} &= \frac{1}{N} \sum_{j=1}^{N_k} \frac{\zeta_{kj}^{(t)}}{1 + \bar{b}_k^{(t-1)} \zeta_{kj}^{(t)}} \\
\bar{b}_k^{(t,l)} &= \frac{1}{N} \operatorname{tr} \mathbf{P}_k \left(b_k^{(t-1)} \mathbf{P}_k + \left[\bar{c}_k - b_k^{(t-1)} \bar{b}_k^{(t,l-1)} \right] \mathbf{I}_{n_k} \right)^{-1} \\
\zeta_{kj}^{(t,l)} &= \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \left(\frac{1}{N} \sum_{i=1}^{K} \sum_{l=1}^{N_k} \frac{\bar{b}_i^{(t-1)} \mathbf{R}_{i,l}}{1 + \bar{b}_i^{(t-1)} \zeta_{il}^{(t,l-1)}} + x\mathbf{I}_N \right)^{-1}
\end{aligned}$$

with initial values $\zeta_{kj}^{(t,0)} = 1/x$, $\bar{b}_k^{(t,0)} = 0$ *and* $b_k^{(0)} = 0\ \forall k, j$.

Our last result will be used in Corollary 7 (Section 3.5) to compute a deterministic equivalent of the mutual information of the K-hop AF relay channel. In contrast to the previous results of this section, the concept of iterative deterministic equivalents is applied here multiple times (one time for each hop) and leads to a set of recursive fixed-point equations.

Theorem 24 ([92, Theorem 2 and 3])**.** *For* $k \in \{1, \ldots, K\}$, *let* $\mathbf{H}_k \in \mathbb{C}^{n_k \times n_{k-1}}$ *be a standard complex Gaussian matrix and define*

$$\mathbf{R}_k \left(\boldsymbol{\beta}_{k-1}\right) = \mathbf{I}_{n_k} + \frac{\alpha_k \beta_{k-1}}{n_{k-1}} \mathbf{H}_k \mathbf{R}_{k-1} \left(\boldsymbol{\beta}_{k-2}\right) \mathbf{H}_k^{\mathsf{H}} \tag{2.16}$$

where $\mathbf{R}_0 = \mathbf{I}_{n_0}$, $\alpha_k > 0$ *and* $\boldsymbol{\beta}_k = [\beta_0 \ldots \beta_k] \geq 0$. *Denote* $c_k = \frac{n_k - 1}{n_k}$ *and let* $n_0, \ldots, n_k \to \infty$ *such that* $0 < \liminf c_k \leq \limsup c_k < \infty\ \forall k$. *Assume that there is* $\bar{\boldsymbol{\beta}}_k = [\bar{\beta}_0 \ldots \bar{\beta}_k]$, *satisfying* $\beta_k - \bar{\beta}_k \xrightarrow{\text{a.s.}} 0\ \forall k$. *Consider the quantities:*

$$\begin{aligned}
m_k \left(x, \boldsymbol{\beta}_k\right) &= \frac{1}{n_{k+1}} \operatorname{tr} \left(\frac{\alpha_{k+1} \beta_k}{n_k} \mathbf{H}_{k+1} \mathbf{R}_k \left(\boldsymbol{\beta}_{k-1}\right) \mathbf{H}_{k+1}^{\mathsf{H}} + \frac{1}{x} \mathbf{I}_{n_{k+1}} \right)^{-1} \\
\mathcal{J}_k \left(x, \boldsymbol{\beta}_{k-1}\right) &= \frac{1}{n_k} \log\det \left(\mathbf{I}_{n_k} + x \frac{\alpha_k \beta_{k-1}}{n_{k-1}} \mathbf{H}_k \mathbf{R}_{k-1} \left(\boldsymbol{\beta}_{k-2}\right) \mathbf{H}_k^{\mathsf{H}} \right).
\end{aligned}$$

(i)

$$m_k \left(x, \boldsymbol{\beta}_k\right) - \bar{m}_k \left(x, \bar{\boldsymbol{\beta}}_k\right) \xrightarrow{\text{a.s.}} 0$$

where $\bar{m}_k \left(x, \bar{\boldsymbol{\beta}}_k\right)$ *is recursively defined for* $k \geq 1$ *as*

$$\bar{m}_k \left(x, \bar{\boldsymbol{\beta}}_k\right) = \frac{x c_{k+1}}{c_{k+1} + \bar{e}_k \left(x, \bar{\boldsymbol{\beta}}_k\right)}$$

and $\bar{e}_k \left(x, \bar{\boldsymbol{\beta}}_k\right) \geq 0$ *is the unique solution to the fixed point equation*

$$\begin{aligned}
\bar{e}_k \left(x, \bar{\boldsymbol{\beta}}_k\right) = c_{k+1} \left(c_{k+1} + \bar{e}_k \left(x, \bar{\boldsymbol{\beta}}_k\right)\right) - \frac{c_{k+1} \left(c_{k+1} + \bar{e}_k \left(x, \bar{\boldsymbol{\beta}}_k\right)\right)^2}{x \alpha_{k+1} \bar{\beta}_k} \\
\times \bar{m}_{k-1} \left(\frac{x \alpha_{k+1} \bar{\beta}_k}{c_{k+1} + x \alpha_{k+1} \bar{\beta}_k + \bar{e}_k \left(x, \bar{\boldsymbol{\beta}}_k\right)}, \bar{\boldsymbol{\beta}}_{k-1} \right).
\end{aligned}$$

The initial values $\bar{m}_0(x, \bar{\beta}_0)$ and $\bar{e}_0(x, \bar{\beta}_0)$ are given in closed form:

$$\bar{m}_0(x, \bar{\beta}_0) = \frac{c_1}{\frac{\alpha_1 \bar{\beta}_0}{c_1 + \bar{e}_0(x, \bar{\beta}_0)} + \frac{1}{x}} + (1 - c_1)x$$

$$\bar{e}_0\left(x, \bar{\beta}_0\right) = -\frac{x\alpha_1\bar{\beta}_0(1 - c_1) + c_1}{2} + \frac{\sqrt{\left(x\alpha_1\bar{\beta}_0(1 - c_1) + c_1\right)^2 + 4x\alpha_1\bar{\beta}_0 c_1^2}}{2}.$$

(ii)

$$\mathcal{J}_k\left(x, \boldsymbol{\beta}_{k-1}\right) - \bar{\mathcal{J}}_k\left(x, \bar{\boldsymbol{\beta}}_{k-1}\right) \xrightarrow{\text{a.s.}} 0$$

where $\bar{\mathcal{J}}_k\left(x, \bar{\boldsymbol{\beta}}_{k-1}\right)$ is recursively defined for $k \geq 2$ as

$$\begin{aligned}\bar{\mathcal{J}}_k\left(x, \bar{\boldsymbol{\beta}}_{k-1}\right) &= c_k \bar{\mathcal{J}}_{k-1}\left(\frac{x\alpha_k\bar{\beta}_{k-1}}{c_k + x\alpha_k\bar{\beta}_{k-1} + \bar{e}_{k-1}\left(x, \bar{\boldsymbol{\beta}}_{k-1}\right)}, \bar{\boldsymbol{\beta}}_{k-1}\right) \\ &\quad + c_k \log\left(1 + \frac{x\alpha_k\beta_{k-1}}{c_k + \bar{e}_{k-1}\left(x, \bar{\boldsymbol{\beta}}_{k-1}\right)}\right) \\ &\quad + \log\left(1 + \frac{\bar{e}_{k-1}\left(x, \bar{\boldsymbol{\beta}}_{k-1}\right)}{c_k}\right) - \frac{\bar{e}_{k-1}\left(x, \bar{\boldsymbol{\beta}}_{k-1}\right)}{c_k + \bar{e}_{k-1}\left(x, \bar{\boldsymbol{\beta}}_{k-1}\right)}.\end{aligned}$$

The initial value $\bar{\mathcal{J}}_1\left(x, \bar{\beta}_0\right)$ is given in closed form:

$$\begin{aligned}\bar{\mathcal{J}}_1\left(x, \bar{\beta}_0\right) &= c_1 \log\left(1 + \frac{x\alpha_1\bar{\beta}_0}{c_1 + \bar{e}_0\left(x, \bar{\beta}_0\right)}\right) + \log\left(1 + \frac{\bar{e}_0\left(x, \bar{\beta}_0\right)}{c_1}\right) \\ &\quad - \frac{\bar{e}_0\left(x, \bar{\beta}_0\right)}{c_1 + \bar{e}_0\left(x, \bar{\beta}_0\right)}.\end{aligned}$$

Proof. The proof is provided in Appendix 2.5.10. ☐

Remark 13. *The quantity $m_k\left(x, \boldsymbol{\beta}_k\right)$ can be seen as the Stieltjes transform of the e.s.d. of the matrix $\alpha_{k+1}\beta_k\frac{1}{n_k}\mathbf{H}_{k+1}\mathbf{R}_k(\boldsymbol{\beta}_{k-1})\mathbf{H}_{k+1}^{\mathsf{H}}$ evaluated at $-\frac{1}{x}$. One can further show that Theorem 24 (i) implies the weak convergence of the e.s.d. $\alpha_{k+1}\beta_k\frac{1}{n_k}\mathbf{H}_{k+1}\mathbf{R}_k(\boldsymbol{\beta}_{k-1})\mathbf{H}_{k+1}^{\mathsf{H}}$ to a distribution function, whose Stieltjes transform is given by $\bar{m}_k$, for almost every $\mathbf{H}_1, \ldots, \mathbf{H}_K$.*

Remark 14. *The values of $\bar{\mathcal{J}}_k$ and $\bar{m}_k$ can be very easily numerically computed. However, due to the recursive structure, the computational complexity grows quickly with k. Calculating $\bar{\mathcal{J}}_k$ and $\bar{m}_k$ with high precision for large values of k (> 10) seems therefore impractical.*

2.5 Appendices

2.5.1 Proof of Lemma 12

Let us begin with the following inequality:

$$\begin{aligned}
&\mathbb{E}\left[\left|\left(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right)^{q} - \left(\frac{1}{N}\operatorname{tr}\mathbf{A}\right)^{q}\right|^{p}\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\left|\left(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\operatorname{tr}\mathbf{A}\right)\right|^{p}\left|\sum_{k=0}^{q-1}\left(\frac{1}{N}\operatorname{tr}\mathbf{A}\right)^{q-1-k}\left(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right)^{k}\right|^{p}\right] \\
&\overset{(b)}{\leq} \sqrt{\mathbb{E}\left[\left|\left(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\operatorname{tr}\mathbf{A}\right)\right|^{2p}\right]}\sqrt{\mathbb{E}\left[\left|\sum_{k=0}^{q-1}\left(\frac{1}{N}\operatorname{tr}\mathbf{A}\right)^{q-1-k}\left(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right)^{k}\right|^{2p}\right]}
\end{aligned} \tag{2.17}$$

where (a) follows from $x^n - y^n = (x-y)\sum_{k=0}^{n-1} x^{n-1-k}y^k$ and (b) follows from Hoelder's inequality. For the first factor, we have from Lemmas 11 and 9

$$\mathbb{E}\left[\left|\left(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\operatorname{tr}\mathbf{A}\right)\right|^{2p}\right] \leq \frac{C_{2p}}{N^p}\|\mathbf{A}\|^{2p}\left(\nu_4^p + \nu_{4p}\right). \tag{2.18}$$

For the second factor, we have again by Hoelder's inequality and Lemma 9

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{k=0}^{q-1}\left(\frac{1}{N}\operatorname{tr}\mathbf{A}\right)^{q-1-k}\left(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right)^{k}\right|^{2p}\right] \\
&\leq q^{2p-1}\sum_{k=0}^{q-1}\|\mathbf{A}\|^{2p(q-1-k)}\mathbb{E}\left[\left|\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right|^{2pk}\right].
\end{aligned} \tag{2.19}$$

Let us treat the terms $\mathbb{E}\left[\left|\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right|^{2pk}\right]$ separately. For every even integer $r \geq 2$,

$$\begin{aligned}
\mathbb{E}\left[\left|\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right|^{r}\right] &= \sum_{\substack{i_1,\dots,i_r \\ j_1,\dots,j_r}} \mathbb{E}\left[x_{i_1}^* \cdots x_{i_r}^* x_{j_1} \cdots x_{j_r}\right]\mathbf{A}_{i_1,j_1}\cdots\mathbf{A}_{i_r,j_r}^* \\
&\overset{(a)}{\leq} \nu_{2r}\left|\frac{1}{N}\sum_{i,j}\mathbf{A}_{i,j}\right|^{r} \\
&\overset{(b)}{\leq} \nu_{2r}\|\mathbf{A}\|^{r}
\end{aligned} \tag{2.20}$$

where (a) is due to $\mathbb{E}\left[x_{i_1}^* \cdots x_{i_r}^* x_{j_1} \cdots x_{j_r}\right] \leq \frac{1}{N^r}\mathbb{E}\left[|\sqrt{N}x_i|^{2r}\right] \leq \frac{\nu_{2r}}{N^r}$ and (b) follows from $\left|\frac{1}{N}\sum_{i,j}\mathbf{A}_{i,j}\right| \leq \|\mathbf{A}\|$. Using (2.20), (2.19) and (2.18) in (2.17) concludes the proof.

2.5.2 Proof of Theorem 15

We first provide an outline of the proof for better understanding.

Sketch of the proof

As a first step, we wish to prove that there exists a matrix $\mathbf{F}$ of the form $\mathbf{F} = \sum_{i=1}^{K} \bar{f}_i \mathbf{R}_i$, with $\bar{f}_i \in \mathbb{C}$, such that, for all nonnegative $\mathbf{A}$ with $\|\mathbf{A}\| < \infty$ uniformly on N and $z < 0$,

$$\frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{F} - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0.$$

Taking $\mathbf{A} = \mathbf{R}_i$ and denoting $f_i \triangleq \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$, we will have in particular that

$$f_i - \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{f}_j \mathbf{R}_j - z\mathbf{I}_N \right)^{-1} \xrightarrow{\text{a.s.}} 0.$$

Contrary to classical deterministic equivalent approaches for random matrices with i.i.d. entries, finding the approximation $\frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}$ for $\frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$ is not straightforward. The reason is that, during the derivation, terms such as $\frac{1}{N_i - n_i} \operatorname{tr} \left(\mathbf{I}_{N_i} - \mathbf{W}_i \mathbf{W}_i^{\mathsf{H}}\right) \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i$ with the $\left(\mathbf{I}_{N_i} - \mathbf{W}_i \mathbf{W}_i^{\mathsf{H}}\right)$ prefix will naturally appear and will be required to be controlled. We proceed as follows.

- We first denote, for all i, $\delta_i \triangleq \frac{1}{N_i - n_i} \operatorname{tr} \left(\mathbf{I}_{N_i} - \mathbf{W}_i \mathbf{W}_i^{\mathsf{H}}\right) \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i$ some auxiliary variable. Then we prove

$$f_i - \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0,$$

  with $\mathbf{G} = \sum_{j=1}^{K} \bar{g}_j \mathbf{R}_j$ and

$$\bar{g}_i = \frac{1}{(1 - c_i)\bar{c}_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{1 + p_{il}\delta_i}} \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}}{1 + p_{il}\delta_i},$$

  where p_{il} denotes the lth eigenvalue of $\mathbf{P}_i$, and δ_i is linked to f_i through

$$f_i - \left((1 - c_i)\bar{c}_i \delta_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{\delta_i}{1 + p_{il}\delta_i} \right) \xrightarrow{\text{a.s.}} 0.$$

- This expression of $\bar{g}_i$, which is not convenient under this form, is then shown to satisfy

$$\bar{g}_i - \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i + p_{il} f_i - f_i \bar{g}_i} = \bar{g}_i - \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(f_i \mathbf{P}_i + [\bar{c}_i - f_i \bar{g}_i] \mathbf{I}_{n_i}\right)^{-1} \xrightarrow{\text{a.s.}} 0,$$

  which induces the $2K$-equation system

$$f_i - \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{g}_j \mathbf{R}_j - z\mathbf{I}_N \right)^{-1} \xrightarrow{\text{a.s.}} 0$$

$$\bar{g}_i - \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(\bar{g}_i \mathbf{P}_i + [\bar{c}_i - f_i \bar{g}_i] \mathbf{I}_{n_i}\right)^{-1} \xrightarrow{\text{a.s.}} 0.$$

- These relations are sufficient to infer the deterministic equivalent, but will be made more attractive for further considerations by introducing $\mathbf{F} = \sum_{i=1}^{K} \bar{f}_i \mathbf{R}_i$, and proving that

$$f_i - \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{f}_j \mathbf{R}_j - z\mathbf{I}_N \right)^{-1} \xrightarrow{\text{a.s.}} 0$$
$$\bar{f}_i - \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(\bar{f}_i \mathbf{P}_i + [\bar{c}_i - f_i \bar{f}_i] \mathbf{I}_{n_i} \right)^{-1} = 0,$$

  where, for $z < 0$, $\bar{f}_i$ lies in $[0, c_i\bar{c}_i/f_i)$ and is now uniquely determined by f_i. In order to establish this convergence, it is necessary to define an analytic extension of $\bar{f}_i$ in a neighborhood of $\mathbb{R}^-$. The function f_i can be immediately extended to $\mathbb{C} \setminus \mathbb{R}^+$ where it verifies the properties of a Stieltjes transform of a finite measure supported by $\mathbb{R}^+$.

This is the very technical part of the proof. We then prove in a second step the existence and uniqueness of a solution to the fixed-point equation

$$e_i - \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{e}_j \mathbf{R}_j - z\mathbf{I}_N \right)^{-1} = 0$$
$$\bar{e}_i - \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(\bar{e}_i \mathbf{P}_i + [\bar{c}_i - e_i \bar{e}_i] \mathbf{I}_{n_i} \right)^{-1} = 0,$$

for all finite N, $z < 0$ and for $\bar{e}_i \in [0, c_i\bar{c}_i/e_i)$. This unfolds from a property of so-called *standard functions*. We will show precisely that the vector application $\mathbf{h} = (h_1, \ldots, h_K)$ defined for $z < 0$ by

$$h_i : (x_1, \ldots, x_K) \mapsto \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{x}_j \mathbf{R}_j - z\mathbf{I}_N \right)^{-1}$$

with $\bar{x}_i$ the unique solution to

$$\bar{x}_i = \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(\bar{x}_i \mathbf{P}_i + [\bar{c}_i - x_i \bar{x}_i] \mathbf{I}_{n_i} \right)^{-1}$$

lying in $[0, c_i\bar{c}_i/x_i)$, is a standard function. It will unfold, from [86, Theorem 2], that the fixed-point equation in $(e_1, \ldots, e_K)$ has a unique solution with positive entries and that this solution can be determined as the limiting iteration of a classical fixed point algorithm. We will further establish that the $e_k(z)$ are Stieltjes transforms of finite measures supported by $\mathbb{R}^+$ which satisfy the fundamental equations for $z \in D$.

The last step proves that the unique solution $(e_1, \ldots, e_N)$ is such that

$$e_i - f_i \xrightarrow{\text{a.s.}} 0,$$

which is solved by standard arguments. This will entail immediately by classical complex analysis arguments that $m_N(z) - \bar{m}_N(z) \xrightarrow{\text{a.s.}} 0$ for all $z \in \mathbb{C} \setminus \mathbb{R}^+$, form which the almost sure convergence $F_N - \bar{F}_N \Rightarrow 0$ unfolds.

Complete proof

We remind that, as N grows, the ratios $c_i = \frac{n_i}{N_i}$ for $i = \{1, \ldots, K\}$ satisfy $\limsup_N c_i < 1$. We also assume for the time being that for all i, $\|\mathbf{R}_i\|$ is uniformly bounded. The case where $\|\mathbf{R}_i\|$ is uniformly bounded only in the almost sure sense will be treated subsequently.

Step 1: Convergence

In this section, we take $z < 0$, until further notice. Let us first introduce the following parameters. We will denote $P = \max_i\{\limsup \|\mathbf{P}_i\|\}$, $R = \max_i\{\limsup \|\mathbf{R}_i\|\}$, $c_+ = \max_i\{\limsup c_i\}$, $\bar{c}_- = \min_i\{\liminf \bar{c}_i\}$ and $\bar{c}_+ = \max_i\{\limsup \bar{c}_i\}$.

Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be a Hermitian nonnegative definite matrix with spectral norm uniformly bounded by A. Recall the definition $\mathbf{R}_i = \mathbf{H}_i\mathbf{H}_i^{\mathsf{H}}$. Taking $\mathbf{G} = \sum_{j=1}^{K} \bar{g}_j\mathbf{R}_j$, with $\bar{g}_1, \ldots, \bar{g}_K$ scalars left undefined for the moment, we have

$$
\begin{aligned}
&\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{G} - z\mathbf{I}_N)^{-1} \\
&\overset{(a)}{=} \frac{1}{N}\operatorname{tr}\left[\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\sum_{i=1}^{K}\mathbf{H}_i\left(-\mathbf{W}_i\mathbf{P}_i\mathbf{W}_i^{\mathsf{H}} + \bar{g}_i\mathbf{I}_{N_i}\right)\mathbf{H}_i^{\mathsf{H}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\right] \\
&\overset{(b)}{=} \sum_{i=1}^{K}\bar{g}_i\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_i(\mathbf{G} - z\mathbf{I}_N)^{-1} \\
&\qquad - \frac{1}{N}\sum_{i=1}^{K}\sum_{l=1}^{n_i} p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il} \\
&\overset{(c)}{=} \sum_{i=1}^{K}\bar{g}_i\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_i(\mathbf{G} - z\mathbf{I}_N)^{-1} \\
&\qquad - \frac{1}{N}\sum_{i=1}^{K}\sum_{l=1}^{n_i}\frac{p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}}{1 + p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}}
\end{aligned}
\tag{2.21}
$$

with $\mathbf{w}_{il} \in \mathbb{C}^{N_i}$ the lth column of $\mathbf{W}_i$, $p_{i1}, \ldots, p_{in_i}$ the eigenvalues of $\mathbf{P}_i$ and $\mathbf{B}_{(i,l)} = \mathbf{B}_N - p_{il}\mathbf{H}_i\mathbf{w}_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}$. The equality (a) follows from Lemma 5, (b) follows from the decomposition $\mathbf{W}_i\mathbf{P}_i\mathbf{W}_i^{\mathsf{H}} = \sum_{l=1}^{n_i} p_{il}\mathbf{w}_{il}\mathbf{w}_{il}^{\mathsf{H}}$, while the equality (c) follows from Lemma 6.

The idea now is to infer the values of the $\bar{g}_i$ such that the differences in (2.21) go to zero almost surely as N grows large. We will therefore proceed by studying the quantities $\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}$ and $\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}$ in the denominator and numerator of the second term in (2.21). For every $i \in \{1, \ldots, K\}$, denote

$$
\delta_i \triangleq \frac{1}{N_i - n_i}\operatorname{tr}\left(\mathbf{I}_{N_i} - \mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}\right)\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\ . \tag{2.22}
$$

Introducing the additional term $(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}$ in the argument of the trace in δ_i, we denote

$$
\beta_i \triangleq \frac{1}{N_i - n_i}\operatorname{tr}\left(\mathbf{I}_{N_i} - \mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}\right)\mathbf{H}_i\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\ . \tag{2.23}
$$

With these notations, according to Lemma 15 (*iii*), the term $\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}$ is asymptotically close to δ_i, and, if $\mathbf{G}$ is independent of $\mathbf{w}_{il}$, the term $\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}$ is asymptotically close to β_i.

We also define

$$f_i \triangleq \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \tag{2.24}$$

for $z \in \mathbb{C} \setminus \mathbb{R}^+$. Note that $f_i(z) \geq 0$ for $z < 0$. Remark first, from standard matrix inequalities and the fact that $\mathbf{w}^{\mathsf{H}}\mathbf{A}\mathbf{w} \leq \|\mathbf{A}\|$ for any Hermitian matrix $\mathbf{A}$ and any unitary vector $\mathbf{w}$, that we have the following bounds on δ_i, β_i and f_i,

$$\delta_i \leq \frac{R}{|z|}, \quad \beta_i \leq \frac{RA}{|z|^2}, \quad f_i \leq \frac{R}{|z|}\ . \tag{2.25}$$

From Lemma 6, we have that

$$\begin{aligned}
(1-c_i)\bar{c}_i\delta_i &= f_i - \frac{1}{N}\sum_{l=1}^{n_i}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il} \\
&= f_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}{1 + p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}\ .
\end{aligned} \tag{2.26}$$

Since $z < 0$, $\delta_i \geq 0$, so $\frac{1}{1+p_{il}\delta_i}$ is well defined. By adding the term $\frac{1}{N}\sum_{l=1}^{n_i}\frac{\delta_i}{1+p_{il}\delta_i}$ on both sides, (2.26) can be re-written as

$$\begin{aligned}
&(1-c_i)\bar{c}_i\delta_i - f_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{\delta_i}{1+p_{il}\delta_i} \\
&= \frac{1}{N}\sum_{l=1}^{n_i}\left[\frac{\delta_i}{1+p_{il}\delta_i} - \frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}{1 + p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}\right] \\
&= \frac{1}{N}\sum_{l=1}^{n_i}\left[\frac{\delta_i - \mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}{(1+p_{il}\delta_i)\left(1 + p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}\right)}\right]\ .
\end{aligned} \tag{2.27}$$

We now apply Lemma 15 and Lemma 8, which together with $\delta_i \leq R|z|^{-1}$ ensures that

$$\mathbb{E}\left[\left|(1-c_i)\bar{c}_i\delta_i - f_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{\delta_i}{1+p_{il}\delta_i}\right|^4\right] \leq 8\frac{C}{N^2} \tag{2.28}$$

for some constant $C > 0$. This determines the asymptotic behavior of δ_i and, thus, the asymptotic behavior of the quantity $\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}$ in the denominator of (2.21).

We now proceed similarly with β_i as with δ_i. Assuming first that $\mathbf{G}$ is independent of $\mathbf{w}_{il}$, we first obtain

$$\begin{aligned}
\beta_i &= \frac{1}{N_i - n_i}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i \\
&- \frac{1}{N_i - n_i}\sum_{l=1}^{n_i}\frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}{1 + p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}
\end{aligned} \tag{2.29}$$

from which we have

$$\frac{1}{N_i - n_i} \operatorname{tr} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i - \frac{1}{N_i - n_i} \sum_{l=1}^{n_i} \frac{\beta_i}{1 + p_{il}\delta_i} - \beta_i$$
$$= \frac{1}{N_i - n_i} \sum_{l=1}^{n_i} \left[\frac{\mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il}}{1 + p_{il} \mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il}} - \frac{\beta_i}{1 + p_{il}\delta_i} \right]. \tag{2.30}$$

With the same inequalities as above, and with

$$\mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il} \leq \frac{RA}{|z|^2} \tag{2.31}$$

we have that

$$\mathbb{E}\left[\left| \frac{\mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il}}{1 + p_{il} \mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il}} - \frac{\beta_i}{1 + p_{il}\delta_i} \right|^4 \right]$$
$$= \mathbb{E}\left[\left| \frac{\mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il} - \beta_i}{(1 + p_{il}\delta_i)(1 + p_{il} \mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il})} \right.\right.$$
$$+ \frac{p_{il}\delta_i \left[\mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il} - \beta_i \right]}{(1 + p_{il}\delta_i)(1 + p_{il} \mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il})}$$
$$\left.\left. + \frac{p_{il}\beta_i \left[\delta_i - \mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il} \right]}{(1 + p_{il}\delta_i)(1 + p_{il} \mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_{(i,l)} - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \mathbf{w}_{il})} \right|^4 \right]$$
$$\leq 8 \frac{C'}{N^2} \left(1 + \frac{P^4 R^4}{|z|^4} \left(1 + \frac{A^4}{|z|^4} \right) \right) \tag{2.32}$$

for some $C' > C$. Multiplying (2.30) by $\frac{N_i - n_i}{N}$, we obtain

$$\mathbb{E}\left[\left| \frac{1}{N} \operatorname{tr} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \right.\right.$$
$$\left.\left. - \beta_i \left((1 - c_i)\bar{c}_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{1 + p_{il}\delta_i} \right) \right|^4 \right]$$
$$\leq 8 \frac{C'}{N^2} \left(1 + \frac{P^4 R^4}{|z|^4} \left(1 + \frac{A^4}{|z|^4} \right) \right). \tag{2.33}$$

This now provides us with the asymptotic behavior of β_i or equivalently of the quantity $\mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} (\mathbf{G} - z\mathbf{I}_N)^{-1} \mathbf{A} (\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1} \mathbf{H}_i \mathbf{w}_{il}$ in the numerator of (2.21).

We are now in position to infer the $\bar{g}_i$ such that $\frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{G} - z\mathbf{I}_N)^{-1}$ is asymptotically small. For the previous derivations to hold, the scalars $\bar{g}_k$, $k \in \{1, \ldots, K\}$, were assumed independent of $\mathbf{w}_{il}$. It is however easy to see that these derivations still hold true (up to the choice of larger constants C, C') if $\bar{g}_k = \bar{g}_k^{(il)} + \varepsilon_{k,N}^{(il)}$ with $\bar{g}_k^{(il)}$ independent of $\mathbf{w}_{il}$ and

$|\varepsilon_{k,N}^{(il)}| \leq C''/N$, for C'' constant independent of k, i, j. This follows from the fact that

$$\left\| \sum_{k=1}^{K} \bar{g}_k \mathbf{R}_k - \sum_{k=1}^{K} \bar{g}_k^{(il)} \mathbf{R}_k \right\| = \left\| \sum_{k=1}^{K} \varepsilon_{k,N}^{(il)} \mathbf{R}_k \right\| \leq \frac{KRC''}{N}.$$

We choose

$$\bar{g}_k = \frac{1}{(1-c_k)\bar{c}_k + \frac{1}{N}\sum_{m=1}^{n_k} \frac{1}{1+p_{km}\delta_k}} \frac{1}{N} \sum_{m=1}^{n_k} \frac{p_{km}}{1+p_{km}\delta_k} \tag{2.34}$$

and remark that $\bar{g}_k - \bar{g}_k^{(il)} = \mathcal{O}(1/N)$ with $\bar{g}_k^{(il)}$ defined similar to $\bar{g}_k$ (2.34), with column $\mathbf{w}_{il}$ removed from the expression of $\mathbf{B}_N$. Indeed, when $\mathbf{w}_{il}$ is removed, $p_{im} = 0$ and $\delta_i = 0$ are no longer defined, while the term $\delta_k^{(il)}$, $k \neq i$, defined equivalently as $\bar{g}_k^{(il)}$, satisfies $|\delta_k^{(il)} - \delta_k| \leq \frac{1}{N_k} \frac{1}{(1-c_k)|z|}$ from Lemma 8, from which the result unfolds.

Summing the previous results over i and l, we then have

$$\begin{aligned}
&\frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{G} - z\mathbf{I}_N)^{-1} \\
&= \sum_{i=1}^{K} \frac{1}{(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}\delta_i}} \\
&\quad \times \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}}{1+p_{il}\delta_i} \frac{1}{N} \operatorname{tr} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i \\
&\quad - \frac{1}{N} \sum_{i=1}^{K} \sum_{l=1}^{n_i} \frac{p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}}{1 + p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}} \\
&= \sum_{i=1}^{K} \frac{1}{N} \sum_{l=1}^{n_i} p_{il} \left[\frac{\frac{1}{N} \operatorname{tr} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i}{((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i} \frac{1}{1+p_{il'}\delta_i})(1+p_{il}\delta_i)} \right. \\
&\qquad \left. - \frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}}{1 + p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{H}_i\mathbf{w}_{il}} \right].
\end{aligned} \tag{2.35}$$

Notice now that $1 + p_{il}\delta_i \geq 1$ and

$$(1-c_i)\bar{c}_i < (1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}\delta_i} \leq \bar{c}_i \tag{2.36}$$

which ensure that we can divide the term in the expectation of the left-hand side of (2.33) by $1 + p_{il}\delta_i$ and $(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}\delta_i}$ without taking the risk of the denominator getting close to 0. This leads to

$$\begin{aligned}
&\mathbb{E}\left[\left| \frac{\beta_i}{1+p_{il}\delta_i} - \frac{\frac{1}{N} \operatorname{tr} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{H}_i}{\left((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}\delta_i}\right)(1+p_{il}\delta_i)} \right|^4 \right] \\
&\leq 8 \frac{C'}{N^2(1-c_i)^4\bar{c}_i^4} \left(1 + \frac{P^4R^4}{|z|^4}\left(1 + \frac{A^4}{|z|^4}\right)\right).
\end{aligned} \tag{2.37}$$

From (2.32) and (2.37), we therefore have that

$$\mathbb{E}\left[\left|\frac{\frac{1}{N}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}\mathbf{H}_i}{\left((1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+p_{il}\delta_i}\right)(1+p_{il}\delta_i)} - \frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}{1+p_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{il}}\right|^4\right]$$
$$\leq 128\frac{C'}{N^2(1-c_i)^4\bar{c}_i^4}\left(1+\frac{P^4R^4}{|z|^4}\left(1+\frac{A^4}{|z|^4}\right)\right). \tag{2.38}$$

We finally obtain

$$\mathbb{E}\left[\left|\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}-\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{G}-z\mathbf{I}_N)^{-1}\right|^4\right]$$
$$\leq 128K^4\frac{C'}{N^2(1-c_+)^4\bar{c}_-^4}\left(1+\frac{P^4R^4}{|z|^4}\left(1+\frac{A^4}{|z|^4}\right)\right). \tag{2.39}$$

This provides a first convergence result as a function of the parameters δ_i, from which a deterministic equivalent can be determined. Nonetheless, the expression of $\bar{g}_i$ is rather impractical as it stands and we need to go further.

Observe in particular that $\bar{g}_i$ can be written under the form

$$\bar{g}_i=\frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{((1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{il'}\delta_i})+p_{il}\delta_i((1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{il'}\delta_i})}. \tag{2.40}$$

We will study the denominator of the above expression and show that it can be synthesized into a much more attractive form.

From (2.28), we first have

$$\mathbb{E}\left[\left|f_i-\delta_i\left((1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+p_{il}\delta_i}\right)\right|^4\right]\leq\frac{8C}{N^2}. \tag{2.41}$$

Multiplying (2.34) by $-\delta_i\left((1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+p_{il}\delta_i}\right)$ and adding $\bar{c}_i$ to both sides yields

$$\bar{c}_i-\bar{g}_i\delta_i\left((1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+p_{il}\delta_i}\right)=(1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+p_{il}\delta_i}. \tag{2.42}$$

By definition, $\bar{g}_i\leq\frac{P}{(1-c_i)\bar{c}_i}$, and we therefore also have

$$\mathbb{E}\left[\left|(\bar{c}_i-f_i\bar{g}_i)-\left((1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+p_{il}\delta_i}\right)\right|^4\right]\leq 8\frac{C}{N^2}\frac{P^4}{(1-c_+)^4\bar{c}_-^4}. \tag{2.43}$$

The equations (2.41) and (2.43) can now be used to approximate the denominator of $\bar{g}_i$ as follows

$$\begin{aligned}
&\mathbb{E}\left[\left|\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i}\right|^4\right] \\
&= \mathbb{E}\left[\left|\frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}^2\left[f_i - \delta_i((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{il'}\delta_i})\right]}{\left[(1+p_{il}\delta_i)((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{il'}\delta_i})\right][\bar{c}_i - f_i\bar{g}_i + p_{il}f_i]}\right.\right. \\
&\qquad \left.\left. + \frac{p_{il}\left[\bar{c}_i - f_i\bar{g}_i - ((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{il'}\delta_i})\right]}{\left[(1+p_{il}\delta_i)((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{il'}\delta_i})\right][\bar{c}_i - f_i\bar{g}_i + p_{il}f_i]}\right|^4\right].
\end{aligned} \tag{2.44}$$

Before providing a useful bound, we need to ensure here that the term $\bar{c}_i - f_i\bar{g}_i + p_{il}f_i$ is uniformly away from zero, for all random f_i and for all N. For this, we recall the bounds $0 \le f_i \le \frac{R}{|z|}$ and $0 \le \bar{g}_i \le \frac{P}{(1-c_i)\bar{c}_i}$.

Let us consider $0 < \varepsilon < 1$ and take from now on $z < -\frac{RP}{(1-c_+)\bar{c}_-(\bar{c}_- - \varepsilon)}$, so that $\bar{c}_i - f_i\bar{g}_i > \varepsilon$ for all i. From (2.41), (2.43) and (2.44), we have

$$\mathbb{E}\left[\left|\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i}\right|^4\right] \le 64\frac{C}{N^2}\frac{P^8}{(1-c_i)^4\bar{c}_i^4\varepsilon^4}\left(1 + \frac{1}{(1-c_i)^4\bar{c}_i^4}\right) \tag{2.45}$$

which is of order $\mathcal{O}(1/N^2)$ since we assumed $\limsup_N c_i < 1$.

We are now ready to introduce the matrix $\mathbf{F}$. Consider

$$\mathbf{F} = \sum_{i=1}^{K}\bar{f}_i\mathbf{R}_i, \tag{2.46}$$

with $\bar{f}_i$ defined as the unique solution to the equation in x

$$x = \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{\bar{c}_i - f_i x + f_i p_{il}} \tag{2.47}$$

within the interval $0 \le x < c_i\bar{c}_i/f_i$. To prove the uniqueness of the solution within this interval, note simply that

$$\begin{aligned}
\frac{c_i\bar{c}_i}{f_i} &> \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{\bar{c}_i - f_i(c_i\bar{c}_i/f_i) + f_i p_{il}} \\
0 &\le \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{\bar{c}_i - f_i\cdot 0 + f_i p_{il}}
\end{aligned} \tag{2.48}$$

and that the function $x \mapsto \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{\bar{c}_i - f_i x + f_i p_{il}}$ is convex for $x \in [0, c_i\bar{c}_i/f_i)$. Hence the uniqueness of the solution in $[0, c_i\bar{c}_i/f_i)$. We also show that this solution is an attractor of the fixed-point algorithm, when correctly initialized. Indeed, let $x_0, x_1, \ldots$ be defined by

$$x_{n+1} = \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i x_n + f_i p_{il}} \tag{2.49}$$

with $x_0 \in [0, c_i\bar{c}_i/f_i)$. Then, $x_n \in [0, c_i\bar{c}_i/f_i)$ implies $\bar{c}_i - f_i x_n + f_i p_{il} > (1 - c_i)\bar{c}_i + f_i p_{il} \geq f_i p_{il}$ and therefore $f_i x_{n+1} \leq c_i \bar{c}_i$, so $x_0, x_1, \ldots$ is contained in $[0, c_i\bar{c}_i/f_i)$. Now observe that

$$x_{n+1} - x_n = \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il} f_i (x_n - x_{n-1})}{(\bar{c}_i + p_{il} f_i - f_i x_n)(\bar{c}_i + p_{il} f_i - f_i x_{n-1})} \tag{2.50}$$

with all terms being nonnegative in the sum, so that the differences $x_{n+1} - x_n$ and $x_n - x_{n-1}$ have the same sign (we also have from the above remarks that $x_{n+1} - x_n \leq c_i(x_n - x_{n-1})$). The sequence $x_0, x_1, \ldots$ is therefore monotonic and bounded: it converges. Calling x_∞ this limit, we have that

$$x_\infty = \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i + p_{il} f_i - f_i x_\infty} \tag{2.51}$$

as required.

To be able to finally prove that $\frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{F} - z\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0$, we want now to show that $\bar{g}_i - \bar{f}_i$ tends to zero at a sufficiently fast rate. For this, we write

$$\begin{aligned}
&\mathbb{E}\left[\left|\bar{g}_i - \bar{f}_i\right|^4\right] \\
&\leq 8\left(\mathbb{E}\left[\left|\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i}\right|^4\right]\right. \\
&\qquad \left. + \mathbb{E}\left[\left|\frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i} - \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i\bar{f}_i + p_{il}f_i}\right|^4\right]\right) \\
&= 8\left(\mathbb{E}\left[\left|\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i}\right|^4\right]\right. \\
&\qquad \left. + \mathbb{E}\left[\left|\bar{g}_i - \bar{f}_i\right|^4 \left|\frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}f_i}{(\bar{c}_i - f_i\bar{f}_i + p_{il}f_i)(\bar{c}_i - f_i\bar{g}_i + p_{il}f_i)}\right|^4\right]\right)
\end{aligned} \tag{2.52}$$

where we have simply written $\bar{g}_i - \bar{f}_i = (\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i})$ $+ (\frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i} - \bar{f}_i)$ and used the triangle inequality on the fourth power of each term.

We only need to ensure now that the coefficient multiplying $\left|\bar{g}_i - \bar{f}_i\right|$ in the right-hand side term is uniformly smaller than 1. For this, observe that, as $z \to -\infty$, $|p_{il}f_i| \leq \frac{PR}{|z|} \to 0$ in the numerator. In the denominator, we already know that $\bar{c}_i - f_i\bar{f}_i + p_{il}f_i \geq (1 - c_i)\bar{c}_i$ and we also have that $\bar{c}_i - f_i\bar{g}_i + p_{il}f_i \geq \bar{c}_i - \frac{RP}{(1-c_i)|z|}$, which is greater than some $\eta > 0$ for $|z|$ taken large.

Take $\eta > 0$ and smaller than 1, and choose z to be such that, for all i,

$$\left|\frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}f_i}{(\bar{c}_i - f_i\bar{f}_i + p_{il}f_i)(\bar{c}_i - f_i\bar{g}_i + p_{il}f_i)}\right| \leq \frac{PR}{|z|(1-c_i)\bar{c}_i\eta} < \frac{1-\eta}{8} \quad (2.53)$$

That is, from now on, we take $z < \min\left(-\frac{8PR}{\eta(1-\eta)(1-c_+)\bar{c}_-}, -\frac{RP}{(1-c_+)\bar{c}_-(1-\varepsilon)}\right)$.

From the inequality (2.52), gathering the terms in $\mathbb{E}\left[\left|\bar{g}_i - \bar{f}_i\right|^4\right]$ on the left side, we finally have

$$\mathbb{E}\left[\left|\bar{g}_i - \bar{f}_i\right|^4\right] \leq \frac{512}{\eta^4}\frac{C}{N^2}\frac{P^8}{(1-c_i)^4\bar{c}_i^4\varepsilon^4}\left(1 + \frac{1}{(1-c_i)^4\bar{c}_i^4}\right). \quad (2.54)$$

We can now proceed to prove the deterministic equivalent relations:

$$\begin{aligned}
&\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1} \\
&= \sum_{i=1}^{K}\frac{1}{N}\sum_{l=1}^{n_i}p_{il}\left[\frac{\frac{1}{N}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i}{((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{i,l'}\delta_i})(1+p_{il}\delta_i)}\right. \\
&\qquad\left. - \frac{\frac{1}{N}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i}{\bar{c}_i - f_i\bar{f}_i + p_{il}f_i}\right] \\
&= \sum_{i=1}^{K}\frac{1}{N}\sum_{l=1}^{n_i}p_{il}\left[\left(\frac{\frac{1}{N}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i}{((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{i,l'}\delta_i})(1+p_{il}\delta_i)}\right.\right. \\
&\qquad\left. - \frac{\frac{1}{N}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i}\right) \\
&\quad + \left(\frac{\frac{1}{N}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i}{\bar{c}_i - f_i\bar{g}_i + p_{il}f_i}\right. \\
&\qquad\left.\left. - \frac{\frac{1}{N}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i}{\bar{c}_i - f_i\bar{f}_i + p_{il}f_i}\right)\right] \\
&= \sum_{i=1}^{K}\frac{1}{N}\operatorname{tr}\mathbf{H}_i^{\mathsf{H}}\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i \\
&\quad\times\frac{1}{N}\sum_{l=1}^{n_i}p_{il}\left[\frac{f_i(\bar{g}_i - \bar{f}_i)}{(\bar{c}_i - f_i\bar{f}_i + p_{il}f_i)(\bar{c}_i - f_i\bar{g}_i + p_{il}f_i)}\right. \\
&\quad + \frac{\left((\bar{c}_i - f_i\bar{g}_i) - ((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{i,l'}\delta_i})\right)}{((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{i,l'}\delta_i})(1+p_{il}\delta_i)(\bar{c}_i - f_i\bar{f}_i + p_{il}f_i)} \\
&\quad\left. + \frac{p_{il}\left(f_i - \delta_i((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{i,l'}\delta_i})\right)}{((1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+p_{i,l'}\delta_i})(1+p_{il}\delta_i)(\bar{c}_i - f_i\bar{f}_i + p_{il}f_i)}\right]. \quad (2.55)
\end{aligned}$$

Therefore, from (2.41), (2.43) and (2.54),

$$\begin{aligned}&\mathbb{E}\left[\left|\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}-\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{F}-z\mathbf{I}_N\right)^{-1}\right|^4\right]\\&\leq\frac{64R^4P^4A^4K}{|z|^8(1-c_+)^8\bar{c}_-^8}\frac{C}{N^2}\left(1+\frac{1}{(1-c_+)^4\bar{c}_-^4}\right)^4\left[1+\frac{64R^4P^4}{|z|^4\eta^4\varepsilon^4}\right]\end{aligned}\tag{2.56}$$

which is of order $\mathcal{O}(1/N^2)$. Together with (2.39), applying Markov inequality and the Borel Cantelli lemma, we finally arrive at

$$\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}-\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{F}-z\mathbf{I}_N\right)^{-1}\xrightarrow{\text{a.s.}}0\tag{2.57}$$

as N grows large for realizations of $\{\mathbf{W}_1,\ldots,\mathbf{W}_K\}$ taken from a set $A_z\subset\Omega$ of probability one (we use Ω here to denote the sample space of the probability space generating the sequences of matrices $\{\mathbf{W}_1,\ldots,\mathbf{W}_K\}$ of growing sizes). This therefore holds true for countably many z (smaller than the established bound) with a cluster point in $\mathbb{R}^-$, on a set $A\subset\Omega$ of probability one.

Before we can extend the convergence on the entire negative real axis, we need to define an analytic extension of $\bar{f}_i$ in a neighborhood of $\mathbb{R}^-$. Take $D=\left\{z=x+\mathbf{i}y : x<0, |y|\leq|x|\frac{1-c_i}{c_i}\right\}$. For $z\in D$, the following holds

$$\Re\{f_i\}\geq 0\qquad\text{and}\qquad|\Im\{f_i\}|\leq\Re\{f_i\}\frac{1-c_i}{c_i}.\tag{2.58}$$

To see this, consider $\mathbf{B}_N=\mathbf{U}\mathbf{D}\mathbf{U}^{\mathsf{H}}$ the eigenvalue decomposition of $\mathbf{B}_N$, where $\mathbf{U}=[\mathbf{u}_1\ldots\mathbf{u}_N]\in\mathbb{C}^{N\times N}$ is unitary and $\mathbf{D}=\operatorname{diag}(d_1,\ldots,d_N)$ contains the nonnegative eigenvalues of $\mathbf{B}_N$. Denoting $z=x+\mathbf{i}y$, we have

$$\begin{aligned}f_i&=\frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}\\&=\frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}\left(\mathbf{B}_N-z^*\mathbf{I}_N\right)\left(\mathbf{B}_N-z^*\mathbf{I}_N\right)^{-1}\\&=\frac{1}{N}\sum_{j=1}^{N}\frac{d_j-x}{|d_j-z|^2}\mathbf{u}_j^{\mathsf{H}}\mathbf{R}_i\mathbf{u}_j+\mathbf{i}y\frac{1}{N}\sum_{j=1}^{N}\frac{1}{|d_j-z|^2}\mathbf{u}_j^{\mathsf{H}}\mathbf{R}_i\mathbf{u}_j.\end{aligned}\tag{2.59}$$

From the last equation, it follows that $x<0$ and $|y|\leq|x|\frac{1-c_i}{c_i}$ imply $\Re\{f_i\}\geq 0$ and $|\Im\{f_i\}|\leq\Re\{f_i\}\frac{1-c_i}{c_i}$.

Consider now the sequence $\{q_{i,n}\}_{n\geq 0}$ of complex numbers, recursively defined as

$$q_{i,n}=\frac{f_i}{(1-c_i)\bar{c}_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+p_{il}q_{i,n-1}}},\quad n\geq 1\tag{2.60}$$

and $q_{i,0}=0$. We will now show that $|q_{i,n}|\leq\frac{|f_i|}{(1-c_i)\bar{c}_i}$ for all n and $z\in D$. First, notice that

$$|q_{i,n}|\leq\frac{|f_i|}{(1-c_i)\bar{c}_i}\tag{2.61}$$

whenever $\Re\{q_{i,n-1}\} \geq 0$. After some simple algebra, one arrives at

$$\Re\{q_{i,n}\} = \frac{\Re\{f_i\} \left[(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{|1+p_{il}q_{i,n-1}|^2}\right]}{\left|(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}q_{i,n-1}}\right|^2} + \frac{\frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}(\Re\{f_i\}\Re\{q_{i,n-1}\} - \Im\{f_i\}\Im\{q_{i,n-1}\})}{|1+p_{il}q_{i,n-1}|^2}}{\left|(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}q_{i,n-1}}\right|^2} \tag{2.62}$$

$$\Im\{q_{i,n}\} = \frac{\Im\{f_i\} \left[(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{|1+p_{il}q_{i,n-1}|^2}\right]}{\left|(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}q_{i,n-1}}\right|^2} + \frac{\frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}(\Im\{f_i\}\Re\{q_{i,n-1}\} + \Re\{f_i\}\Im\{q_{i,n-1}\})}{|1+p_{il}q_{i,n-1}|^2}}{\left|(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}q_{i,n-1}}\right|^2}. \tag{2.63}$$

Now, if we assume $\Re\{q_{i,n-1}\} \geq 0$, we have

$$\begin{aligned}\Re\{q_{i,n}\} &\geq \frac{\Re\{f_i\}(1-c_i)\bar{c}_i - |\Im\{f_i\}|\frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}|\Im\{q_{i,n-1}\}|}{|1+p_{il}q_{i,n-1}|^2}}{\left|(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}q_{i,n-1}}\right|^2} \\ &\geq \frac{\Re\{f_i\}(1-c_i)\bar{c}_i - |\Im\{f_i\}|c_i\bar{c}_i}{\left|(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}q_{i,n-1}}\right|^2}.\end{aligned} \tag{2.64}$$

The right-hand side of the last equations is nonnegative whenever

$$|\Im\{f_i\}| \leq \Re\{f_i\}\frac{1-c_i}{ci}.$$

As this condition is always satisfied for $z \in D$ and we have defined $q_{i,0} = 0$, we can conclude that (2.61) and $\Re\{q_{i,n}\} \geq 0$ hold for all n.

Additionally, we have from (2.62) and (2.63) that

$$\begin{aligned}&\Re\{f_i\}\Re\{q_{i,n-1}\} + \Im\{f_i\}\Im\{q_{i,n-1}\} \\ &= \frac{\left(\Re\{f_i\}^2 + \Im\{f_i\}^2\right)\left[(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1+p_{il}\Re\{q_{n-2}\}}{|1+p_{il}q_{n-2}|^2}\right]}{\left|(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+q_{il}x_{n-2}}\right|^2} \geq 0.\end{aligned} \tag{2.65}$$

Until here, we have proved that $\{q_{i,n}\}$ is a sequence of bounded analytic functions on $z \in D$ (the analyticity follows from the fact that f_i is analytic on $\mathbb{C} \setminus \mathbb{R}^+$ and $q_{i,n}$ is a rational function with no pole in D). Let us now focus on the negative real axis, i.e., $z < 0$, which lies in the interior of D. Here, the following holds

$$\begin{aligned}&q_{i,n+1} - q_{i,n} \\ &= (q_{i,n} - q_{i,n-1})\, f_i \\ &\times \frac{\frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{(1+p_{il}q_{i,n})(1+p_{il}q_{i,n-1})}}{\left[(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}q_{i,n}}\right]\left[(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+p_{il}q_{i,n-1}}\right]}.\end{aligned} \tag{2.66}$$

As f_i and all terms in the fraction of the right-hand side of the last equation are nonnegative, the differences $q_{i,n+1} - q_{i,n}$ and $q_{i,n} - q_{i,n-1}$ have the same sign. Thus, $\{q_{i,n}\}$ is either monotonically increasing or decreasing. Since $\{q_{i,n}\}$ is also bounded, it must converge. This implies by Vitali's convergence theorem that $\{q_{i,n}\}$ converges uniformly on all closed subsets of D and that this limit is an analytic function. Call this limit $q_i = \lim_n q_{i,n}$.

We now define $\tilde{f}_{i,n}$ by the quantities f_i and $q_{i,n}$:

$$\tilde{f}_{i,n} = \frac{1}{f_i}\frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}q_{i,n}}{1+p_{il}q_{i,n}}. \tag{2.67}$$

Clearly, $\{\tilde{f}_{i,n}\}$ is a sequence of analytic bounded functions, converging for $z \in D$ to

$$\tilde{f}_i \triangleq \frac{1}{f_i}\frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}q_i}{1+p_{il}q_i}. \tag{2.68}$$

With the above definition, $q_{i,n+1}$ satisfies

$$\begin{aligned} q_{i,n+1} &= \frac{f_i}{(1-c_i)\bar{c}_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+p_{il}q_{i,n}}} \\ &= \frac{f_i}{\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}q_{i,n}}{1+p_{il}q_{i,n}}} \\ &= \frac{f_i}{\bar{c}_i - f_i\tilde{f}_{i,n}}. \end{aligned} \tag{2.69}$$

Thus, we can write, from (2.67),

$$\tilde{f}_{i,n+1} = \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{(\bar{c}_i - f_i\tilde{f}_{i,n})\left(1+\frac{p_{il}f_i}{\bar{c}_i - f_i\tilde{f}_{i,n}}\right)} = \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}}{\bar{c}_i - f_i\tilde{f}_{i,n} + p_{il}f_i}. \tag{2.70}$$

As a consequence, the restriction of $\tilde{f}_i$ to $z < 0$ is identical to $\bar{f}_i$ and the fixed-point algorithm defined by (2.70) with $\tilde{f}_{i,0} = 0$ converges to $\tilde{f}_i$ for $z \in D$. From this point on, we therefore extend the definition of $\bar{f}_i$ to D by $\bar{f}_i(z) = \tilde{f}_i(z)$.

From (2.65) and for $z \in D$, we have

$$\Re\{\bar{f}_i\} = \frac{1}{N}\sum_{l=1}^{n_i}p_{il}\frac{p_{il}|q_i|^2\Re\{f_i\} + \Re\{f_i\}\Re\{q_i\} + \Im\{f_i\}\Im\{q_i\}}{|f_i + p_{il}q_if_i|^2} \geq 0. \tag{2.71}$$

Since $\mathbf{F} = \sum_{k=1^K}\bar{f}_k\mathbf{R}_k$ and the matrices $\mathbf{R}_k$ are Hermitian nonnegative definite, it follows that $\left|\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\right| \leq \frac{||\mathbf{A}||}{|x|}$ for $z \in D$.

From the Vitali convergence theorem, the identity theorem, the analyticity of the functions under study, and the fact that $\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$ and $\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}$ are uniformly bounded on all closed subsets of $z \in D$, we have that the convergence

$$\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}\left(\sum_{i=1}^{K}\bar{f}_i\mathbf{R}_i - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0 \tag{2.72}$$

holds true for all $z \in D$.

Applying the result for $\mathbf{A} = \mathbf{R}_j$, this is in particular

$$f_j - \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{i=1}^{K} \bar{f}_i \mathbf{R}_i - z \mathbf{I}_N \right)^{-1} \xrightarrow{\text{a.s.}} 0 \tag{2.73}$$

where we remind that $\bar{f}_i$ is the unique solution to

$$x = \frac{1}{N} \sum_{i=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i x + p_{il} f_i} \tag{2.74}$$

within the set $[0, c_i \bar{c}_i / f_i)$. For $\mathbf{A} = \mathbf{I}_N$, this implies

$$m_N(z) - \frac{1}{N} \operatorname{tr} \left(\sum_{i=1}^{K} \bar{f}_i \mathbf{R}_i - z \mathbf{I}_N \right)^{-1} \xrightarrow{\text{a.s.}} 0 \tag{2.75}$$

which proves the convergence.

Step 2: Existence and Uniqueness

We will now prove the existence and the uniqueness of positive solutions $e_1(z), \ldots, e_K(z)$ for $z < 0$ and the convergence of the classical fixed point algorithm to these values. In addition, we will show that the $e_i(z)$ have analytic extensions on $\mathbb{C} \setminus \mathbb{R}^+$ which are Stieltjes transforms of finite measures over $\mathbb{R}^+$ and satisfy the fundamental equations for $z \in D$. We first introduce some notations and useful identities. Until stated otherwise, we assume $z < 0$. Note that, similar to the auxiliary variables δ_i and q_i in Step 1, we can define, for any pair of variables x_i and $\bar{x}_i$, with $\bar{x}_i$ defined as the solution y to $y = \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_j - x_j y + x_j p_{il}}$ such that $0 \leq y < c_j \bar{c}_j / x_j$, the auxiliary variables $\Delta_1, \ldots, \Delta_K$, with the properties

$$\begin{aligned} x_i &= \Delta_i \left((1 - c_i)\bar{c}_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{1 + p_{il} \Delta_i} \right) \\ &= \Delta_i \left(\bar{c}_i - \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il} \Delta_i}{1 + p_{il} \Delta_i} \right) \end{aligned} \tag{2.76}$$

and

$$\begin{aligned} \bar{c}_i - x_i \bar{x}_i &= (1 - c_i)\bar{c}_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{1 + p_{il} \Delta_i} \\ &= \bar{c}_i - \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il} \Delta_i}{1 + p_{il} \Delta_i} \ . \end{aligned} \tag{2.77}$$

Indeed, firstly, there exists a unique mapping between x_i and Δ_i. This unfolds from noticing that

$$\begin{aligned} \frac{d\, x_i}{d\Delta_i} &= \frac{d}{d\Delta_i} \left[\Delta_i \left((1 - c_i)\bar{c}_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{1 + p_{il} \Delta_i} \right) \right] \\ &= (1 - c_i)\bar{c}_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{(1 + p_{il} \Delta_i)^2} > 0 \end{aligned} \tag{2.78}$$

and therefore x_i and Δ_i are one-to-one. Additionally, x_i is a strictly growing function of Δ_i with $\Delta_i = 0$ for $x_i = 0$. This ensures that $\Delta_i > 0$ if and only if $x_i > 0$.

Secondly, from the definition of $\bar{x}_i$, we have

$$\begin{aligned}
\bar{c}_i - x_i\bar{x}_i &= \bar{c}_i - x_i \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{(\bar{c}_i - x_i\bar{x}_i) + p_{il}x_i} \\
&= \bar{c}_i - \Delta_i \left(\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}\Delta_i}{1+p_{il}\Delta_i} \right) \\
&\quad \times \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - x_i\bar{x}_i + p_{il}\Delta_i \left(\bar{c}_i - \frac{1}{N}\sum_{l=1'}^{n_i} \frac{p_{il'}\Delta_i}{1+p_{il'}\Delta_i} \right)} \ .
\end{aligned} \tag{2.79}$$

Note in particular that for $x_i\bar{x}_i = \frac{1}{N}\sum_{l=1'}^{n_i} \frac{p_{il'}\Delta_i}{1+p_{il'}\Delta_i}$, the above equation simplifies to

$$\begin{aligned}
&\bar{c}_i - \Delta_i \left(\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}\Delta_i}{1+p_{il}\Delta_i} \right) \\
&\quad \times \frac{1}{N}\sum_{l=1}^{n_1} \frac{p_{il}}{\left(\bar{c}_i - \frac{1}{N}\sum_{l=1'}^{n_i} \frac{p_{il'}\Delta_i}{1+p_{il'}\Delta_i} \right) + p_{il}\Delta_i \left(\bar{c}_i - \frac{1}{N}\sum_{l'=1}^{n_i} \frac{p_{il'}\Delta_i}{1+p_{il'}\Delta_i} \right)} \\
&= \bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}
\end{aligned} \tag{2.80}$$

and therefore $\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}$ is one of the solution of the implicit equation in u,

$$u = \bar{c}_i - x_i \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{u + p_{il}x_i} \ . \tag{2.81}$$

Equivalently, writing $u = \bar{c}_i - x_i y$, it follows that $\frac{1}{x_i}\frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}$ is one of the solutions of the equation in y

$$y = \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - x_i y + p_{il}x_i} \ . \tag{2.82}$$

Since

$$x_i \left(\frac{1}{x_i}\frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}\Delta_i}{1+p_{il}\Delta_i} \right) < c_i\bar{c}_i \tag{2.83}$$

this solution lies in $[0, c_i\bar{c}_i/x_i)$ and is exactly equal to $\bar{x}_i$. This proves that the equations in $(x_i, \bar{x}_i)$ can be written under the form of the equations in $(\Delta_i, \bar{x}_i)$, as presented above.

We take the opportunity of the above definitions to notice that, for $x_i > x_i'$ and $\bar{x}_i'$, Δ_i' defined similarly as $\bar{x}_i$ and Δ_i,

$$x_i\bar{x}_i - x_i'\bar{x}_i' = \frac{1}{N}\sum_{l=1}^{n_i} \frac{p_{il}(\Delta_i - \Delta_i')}{(1+p_{il}\Delta_i)(1+p_{il}\Delta_i')} > 0 \tag{2.84}$$

whenever $\mathbf{P}_i \neq 0$. Therefore $x_i \bar{x}_i$ is a growing function of x_i (or equivalently of Δ_i). This will turn out to be a useful remark later.

We are now in position to prove the step of uniqueness. Define for $i \in \{1, \ldots, K\}$, the functions

$$h_i : (x_1, \ldots, x_K) \mapsto \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{x}_j \mathbf{R}_j - z \mathbf{I}_N \right)^{-1} \tag{2.85}$$

with $\bar{x}_j$ the unique solution of the equation in y

$$y = \frac{1}{N} \sum_{l=1}^{n_j} \frac{p_{jl}}{\bar{c}_j + x_j p_{jl} - x_j y} \tag{2.86}$$

such that $0 \leq \bar{x}_j < c_j \bar{c}_j / x_j$.

We will prove in the following that the multivariate function $\mathbf{h} = (h_1, \ldots, h_K)$ is a standard interference function as introduced in Definition 11. In order to prove that there exist $x_1, \ldots, x_K$ such that $x_j \geq h_j(x_1, \ldots, x_K)$ for all j, it is sufficient to notice that $h_j(x_1, \ldots, x_K) \leq R/|z|$ for all j. Thus, for $x_j \geq R/|z|$ for all j, $x_j \geq h_j(x_1, \ldots, x_K)$ holds for all j. Therefore, by showing that $\mathbf{h} \triangleq (h_1, \ldots, h_K)$ is a standard function, we will prove with the help of Theorem 16 that the classical fixed point algorithm converges to the unique set of positive solutions $e_1, \ldots, e_K$, when $z < 0$.

The positivity condition is straightforward as $\bar{x}_i$ is positive for x_i positive and therefore $h_j(x_1, \ldots, x_K)$ is always positive whenever $x_1, \ldots, x_K$ are nonnegative.

The scalability is also rather direct. Let $\alpha > 1$, then

$$\begin{aligned} &\alpha h_j(x_1, \ldots, x_K) - h_j(\alpha x_1, \ldots, \alpha x_K) \\ &= \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{k=1}^{K} \frac{\bar{x}_k}{\alpha} \mathbf{R}_k - \frac{z}{\alpha} \mathbf{I}_N \right)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{k=1}^{K} \bar{x}_k^{(\alpha)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1} \end{aligned} \tag{2.87}$$

where we denoted $\bar{x}_j^{(\alpha)}$ the unique solution to (2.86) within $[0, c_j \bar{c}_j / (\alpha x_j))$ with x_j replaced by αx_j.

At this point, we will require the following lemma:

Lemma 17. *Let $\mathbf{A}, \mathbf{B}, \mathbf{R} \in \mathbb{C}^{N \times N}$, where $\mathbf{A}$ and $\mathbf{B}$ are nonnegative-definite, satisfying $\mathbf{B} \succ \mathbf{A}$, and $\mathbf{R}$ is nonnegative-definite. Then*

$$\operatorname{tr} \mathbf{R} \left(\mathbf{A}^{-1} - \mathbf{B}^{-1} \right) > 0. \tag{2.88}$$

Proof. Note that $\mathbf{B} \succ \mathbf{A}$ implies by [93, Corollary 7.7.4] $\mathbf{B}^{-1} \prec \mathbf{A}^{-1}$. Thus, for any vector $\mathbf{x} \in \mathbb{C}^N$,

$$\mathbf{x}^{\mathsf{H}} \left(\mathbf{A}^{-1} - \mathbf{B}^{-1} \right) \mathbf{x} > 0. \tag{2.89}$$

Consider now the eigenvalue decomposition of the matrix $\mathbf{R} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{\mathsf{H}}$, where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]$ and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$. Since $\lambda_i \geq 0 \ \forall i$, we have

$$\operatorname{tr} \mathbf{R} \left(\mathbf{A}^{-1} - \mathbf{B}^{-1} \right) = \sum_{i=1}^{N} \lambda_i \mathbf{u}_i^{\mathsf{H}} \left(\mathbf{A}^{-1} - \mathbf{B}^{-1} \right) \mathbf{u}_i > 0. \tag{2.90}$$

$\square$

From Lemma 17, it suffices to show that

$$\sum_{k=1}^{K} \left[\bar{x}_k^{(\alpha)} - \frac{\bar{x}_k}{\alpha}\right] \mathbf{R}_k + \left[z - \frac{z}{\alpha}\right] \mathbf{I}_N \tag{2.91}$$

is positive definite. Since $\alpha x_i > x_i$, we have from the property (2.84) that

$$\alpha x_k \bar{x}_k^{(\alpha)} - x_k \bar{x}_k > 0 \tag{2.92}$$

or equivalently

$$\bar{x}_k^{(\alpha)} - \frac{\bar{x}_k}{\alpha} > 0. \tag{2.93}$$

Along with $1 - 1/\alpha > 0$ and $z < 0$, this ensures that $\alpha h_j(x_1, \ldots, x_K) > h_j(\alpha x_1, \ldots, \alpha x_K)$.

The monotonicity requires some more calculus. This unfolds from considering $\bar{x}_i$ as a function of Δ_i, by verifying that $\frac{d}{d\Delta_i}\bar{x}_i$ is negative.

$$\begin{aligned}
&\frac{d}{d\Delta_i}\bar{x}_i \\
&= \frac{1}{\Delta_i^2}\left(1 - \frac{\bar{c}_i}{\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}}\right) + \frac{\bar{c}_i}{\Delta_i^2}\left(\frac{\frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{(1+p_{il}\Delta_i)^2}}{\left(\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right)^2}\right) \\
&= \frac{1}{\Delta_i^2\left(\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right)^2}\Bigg[-\frac{1}{N}\left(\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right)\left(\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right) \\
&\qquad + \frac{\bar{c}_i}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{(1+p_{il}\Delta_i)^2}\Bigg] \\
&= \frac{1}{\Delta_i^2\left(\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right)^2}\Bigg[\left(\frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right)^2 \\
&\qquad - \frac{\bar{c}_i}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i} + \frac{\bar{c}_i}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{(1+p_{il}\Delta_i)^2}\Bigg] \\
&= \frac{1}{\Delta_i^2\left(\bar{c}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right)^2}\left[\left(\frac{1}{N}\sum_{l=1}^{n_i}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right)^2 - \frac{\bar{c}_i}{N}\sum_{l=1}^{n_i}\frac{(p_{il}\Delta_i)^2}{(1+p_{il}\Delta_i)^2}\right].
\end{aligned} \tag{2.94}$$

From the Cauchy-Schwarz inequality (or Hoelder's inequality for $p = q = 2$), we have

$$\begin{aligned}
\left(\sum_{l=1}^{n_i}\frac{1}{N}\frac{p_{il}\Delta_i}{1+p_{il}\Delta_i}\right)^2 &\le \sum_{k=1}^{n_i}\frac{1}{N^2}\sum_{l=1}^{n_i}\frac{(p_{il}\Delta_i)^2}{(1+p_{il}\Delta_i)^2} \\
&= c_i\bar{c}_i\frac{1}{N}\sum_{l=1}^{n_i}\frac{(p_{il}\Delta_i)^2}{(1+p_{il}\Delta_i)^2} \\
&< \frac{\bar{c}_i}{N}\sum_{l=1}^{n_i}\frac{(p_{il}\Delta_i)^2}{(1+p_{il}\Delta_i)^2}
\end{aligned} \tag{2.95}$$

which is sufficient to conclude that $\frac{d}{d\Delta_i}\bar{x}_i < 0$. Since Δ_i is an increasing function of x_i, we have that $\bar{x}_i$ is a decreasing function of x_i, i.e., $\frac{d}{dx_i}\bar{x}_i < 0$. Therefore, for two sets $x_1, \ldots, x_K$ and $x'_1, \ldots, x'_K$ of positive values such that $x_j > x'_j$, defining $\bar{x}'_j$ equivalently as $\bar{x}_j$ for the terms x'_j, we have $\bar{x}'_k > \bar{x}_k$. Therefore, from Lemma 17, we finally have

$$\begin{aligned}
& h_j(x_1, \ldots, x_K) - h_j(x'_1, \ldots, x'_K) \\
&= \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{k=1}^{K} \bar{x}_k \mathbf{R}_k - z\mathbf{I}_N \right)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{k=1}^{K} \bar{x}'_k \mathbf{R}_k - z\mathbf{I}_N \right)^{-1} \\
&> 0.
\end{aligned} \tag{2.96}$$

This proves the monotonicity condition and, finally, that $\mathbf{h} = (h_1, \ldots, h_K)$ is a standard function.

It follows from Theorem 16 that $(e_1, \ldots, e_K)$ is uniquely defined and that the classical fixed-point algorithm converges to this solution from any initialization point (remember that, at each step of the algorithm, the set $\bar{e}_1, \ldots, \bar{e}_K$ must be evaluated, possibly thanks to a further fixed-point algorithm).

We will now show that $e_i(z)$ has an analytic extension on $z \in \mathbb{C} \setminus \mathbb{R}^+$ which is the Stieltjes transform of a finite measure supported by $\mathbb{R}^+$. For this proof, consider the matrices $\mathbf{P}_{[p],i} \in \mathbb{C}^{n_i p}$ and $\mathbf{H}_{[p],i} \in \mathbb{C}^{Np \times N_i p}$ for all i defined as the Kronecker products $\mathbf{P}_{[p],i} \triangleq \mathbf{P}_i \otimes \mathbf{I}_p$, $\mathbf{H}_{[p],i} \triangleq \mathbf{H}_i \otimes \mathbf{I}_p$, such that $\mathbf{P}_{[p],i}$ and $\mathbf{R}_{[p],i} = \mathbf{H}_{[p],i}\mathbf{H}^{\mathsf{H}}_{[p],i}$ have the same spectral distributions as the matrices $\mathbf{P}_i$ and $\mathbf{R}_i$, respectively. It is easy to see that the solutions of the implicit equations (2.9) for $z \in \mathbb{C} \setminus \mathbb{R}^+$ remain unchanged by substituting the $\mathbf{P}_{[p],i}$ and $\mathbf{R}_{[p],i}$ to the $\mathbf{P}_i$ and $\mathbf{R}_i$, respectively, for any p. Denoting similarly $f_{[p],i}$ the f_i adapted to $\mathbf{P}_{[p],i}$ and $\mathbf{H}_{[p],i}$, from the convergence result of Step 1, we can choose $f_{[1],i}, f_{[2],i}, \ldots$ a sequence of the set of probability one where convergence is ensured as p grows large (N and the n_i are kept fixed). This sequence is uniformly bounded (by $R/|z|$) in $\mathbb{C} \setminus \mathbb{R}^+$, and we can, therefore, extract a converging subsequence $f_{[\phi(p)],i}$ out of it. Call $e'_i(z)$ this limit.

We wish to prove that e'_i, seen as a function of z, is the *Stieltjes transform* of a distribution function, whose restriction to $\mathbb{R}^-$ matches e_i. For this, we prove the defining properties of a Stieltjes transform, provided in Property 3. By Vitali's convergence theorem [77], e'_i is analytic on $\mathbb{C}^+$ since e'_i is the limit of a sequence of analytic functions, bounded on every compact of $\mathbb{C} \setminus \mathbb{R}^+$. It is clear that for $z \in \mathbb{C}^+$, $\Im[f_{[p],i}(z)] > 0$, $\Im[zf_{[p],i}(z)] > 0$ and $|yf_{[p],i}(\mathbf{i}y)| \leq R$ for $y > 0$. This implies that for $z \in \mathbb{C}^+$, $\Im[e'_i(z)] \geq 0$, $z\Im[e'_i(z)] \geq 0$ and $\lim_{y\to\infty} -\mathbf{i}ye'_i(\mathbf{i}y) \leq R$. In addition, note that, for $z \in \mathbb{C}^+$,

$$\Im[f_{[p],i}] \geq \frac{1}{N} \frac{r}{(RP + |z|)^2} \Im[z] > 0 \tag{2.97}$$

and

$$\Im[zf_{[p],i}] \geq \frac{1}{N} \frac{Kr^2 t}{(RP + |z|)^2} \Im[z] > 0 \tag{2.98}$$

with r a lower bound on the smallest non-zero eigenvalues of $\mathbf{R}_1, \ldots, \mathbf{R}_K$ (we naturally assume all $\mathbf{R}_k$ non-zero) and t a lower bound on the smallest non-zero eigenvalues of $\mathbf{T}_1, \ldots, \mathbf{T}_K$ (again, none assumed identically zero). Take $z \in \mathbb{C}^+$ and $\varepsilon < \frac{1}{2}\min(\frac{1}{N}\frac{r}{(RP+|z|)^2}\Im[z], \frac{1}{N}\frac{Kr^2 t}{(RP+|z|)^2}\Im[z])$. There now exists p_0 such that

$p \geq p_0$ implies $|\Im[f_{[\phi(p)],i}] - \Im[e'_i]| < \varepsilon/2$ and $|\Im[zf_{[\phi(p)],i}] - \Im[ze'_i]| < \varepsilon/2$, and therefore $\Im[e'_i] > \varepsilon/2$ and $z\Im[e'_i(z)] > \varepsilon/2$ so that $e'_i(z)$ is the Stieltjes transform of a finite measure on $\mathbb{R}^+$. Moreover, since $e'_i(z) = \lim f_{[\phi(p)],i}(z)$ on D, from (2.73), $e'_i(z)$ satisfies the equations (2.9) for all $z \in D$.

Consider now two sets of Stieltjes transforms $(e'_1(z), \ldots, e'_K(z))$ and $(e''_1(z), \ldots, e''_K(z))$, $z \in \mathbb{C}\setminus\mathbb{R}^+$, which are solutions of the fixed-point equation for $z < 0$. Since $e'_i(z) = e''_i(z)$ for all $z < 0$, and $e'_i(z) - e''_i(z)$ is holomorphic on $\mathbb{C} \setminus \mathbb{R}^+$ as the difference of Stieltjes transforms, $e'_i(z) = e''_i(z)$ over $\mathbb{C} \setminus \mathbb{R}^+$ [94] by the identity theorem. This therefore proves, in addition to *point-wise* uniqueness on the negative half-line, the uniqueness of the *Stieltjes transform* solution of the functional implicit equation such that, for $z < 0$, $0 \leq \bar{e}_i < c_i\bar{c}_i/e_i$ for all i. Moreover, this solution satisfies the fundamental equations for $z \in D$.

Step 3: Convergence of $e_i - f_i$

For this step, we follow the same approach as in [74]. Denote

$$\varepsilon_{N,i} \triangleq f_i - \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{k=1}^{K} \bar{f}_k \mathbf{R}_k - z\mathbf{I}_N \right)^{-1} \tag{2.99}$$

and recall the definitions of f_i, e_i, $\bar{f}_i$ and $\bar{e}_i$:

$$f_i = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \tag{2.100}$$

$$e_i = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j-1}^{K} \bar{e}_j \mathbf{R}_j - z\mathbf{I}_N \right)^{-1} \tag{2.101}$$

$$\bar{f}_i = \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - f_i\bar{f}_i + p_{il}f_i}, \qquad \bar{f}_i \in [0, c_i\bar{c}_i/f_i) \tag{2.102}$$

$$\bar{e}_i = \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}}{\bar{c}_i - e_i\bar{e}_i + p_{il}e_i}, \qquad \bar{e}_i \in [0, c_i\bar{c}_i/e_i) \ . \tag{2.103}$$

From the definitions above, we have the following set of inequalities

$$f_i \leq \frac{R}{|z|}, \quad e_i \leq \frac{R}{|z|}, \quad \bar{f}_i \leq \frac{P}{(1-c_i)\bar{c}_i}, \quad \bar{e}_i \leq \frac{P}{(1-c_i)\bar{c}_i} \ . \tag{2.104}$$

We will show in the sequel that

$$e_i - f_i \xrightarrow{\text{a.s.}} 0, \tag{2.105}$$

for all $i \in \{1, \ldots, N\}$. Write the following differences

$$f_i - e_i = \sum_{j=1}^{K} (\bar{e}_j - \bar{f}_j) \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{k=1}^{K} \bar{e}_k \mathbf{R}_k - z\mathbf{I}_N \right)^{-1} \mathbf{R}_j \left(\sum_{k=1}^{K} \bar{f}_k \mathbf{R}_k - z\mathbf{I}_N \right)^{-1} + \varepsilon_{N,i} \tag{2.106}$$

$$\bar{e}_i - \bar{f}_i = \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il}^2(f_i - e_i) - p_{il}\left[f_i\bar{f}_i - e_i\bar{e}_i\right]}{(\bar{c}_i - \bar{e}_i e_i + p_{il}e_i)(\bar{c}_i - \bar{f}_i f_i + p_{il}f_i)} \tag{2.107}$$

$$f_i\bar{f}_i - e_i\bar{e}_i = \bar{f}_i(f_i - e_i) + e_i(\bar{f}_i - \bar{e}_i) \ . \tag{2.108}$$

For notational convenience, we define the following values

$$\alpha \triangleq \sup_i \mathbb{E}\left[|f_i - e_i|^4\right] \tag{2.109}$$

$$\bar{\alpha} \triangleq \sup_i \mathbb{E}\left[|\bar{f}_i - \bar{e}_i|^4\right] \ . \tag{2.110}$$

It is thus sufficient to show that α is summable to prove (2.105). By applying (2.104) to the absolute of the first difference, we obtain

$$|f_i - e_i| \leq \frac{KR^2}{|z|^2} \sup_i |\bar{f}_i - \bar{e}_i| + \sup_i |\varepsilon_{N,i}| \tag{2.111}$$

and hence

$$\alpha \leq \frac{8K^4R^8}{|z|^8}\bar{\alpha} + \frac{8C}{N^2} \tag{2.112}$$

for some $C > 0$ such that $\mathbb{E}[\sup_i |\varepsilon_{N,i}|^4] \leq 8K \sup_i \mathbb{E}[|\varepsilon_{N,i}|^4] \leq C/N^2$. Similarly, we have for the third difference

$$\begin{aligned} |f_i\bar{f}_i - e_i\bar{e}_i| &\leq |\bar{f}_i||f_i - e_i| + |e_i||\bar{f}_i - \bar{e}_i| \\ &\leq \frac{P}{(1-c_+)\bar{c}_-} \sup_i |f_i - e_i| + \frac{R}{|z|} \sup_i |\bar{f}_i - \bar{e}_i| \ . \end{aligned} \tag{2.113}$$

This result can be used to find an upper bound on the second difference term, which writes

$$\begin{aligned} |\bar{f}_i - \bar{e}_i| &\leq \frac{1}{(1-c_+)^2\bar{c}_-^2}\left(P^2 \sup_i |f_i - e_i| + P|f_i\bar{f}_i - e_i\bar{e}_i|\right) \\ &\leq \frac{1}{(1-c_+)^2\bar{c}_-^2}\left(P^2 \sup_i |f_i - e_i|\right. \\ &\qquad \left. + P\left[\frac{P}{(1-c_+)\bar{c}_-} \sup_i |f_i - e_i| + \frac{R}{|z|} \sup_i |\bar{f}_i - \bar{e}_i|\right]\right) \\ &\leq \frac{P^2(\bar{c}_- + 1)}{(1-c_+)^3\bar{c}_-^3} \sup_i |f_i - e_i| + \frac{RP}{|z|(1-c_+)^2\bar{c}_-^2} \sup_i |\bar{f}_i - \bar{e}_i| \ . \end{aligned} \tag{2.114}$$

Hence

$$\bar{\alpha} \leq \frac{8P^8(\bar{c}_- + 1)^4}{(1-c_+)^{12}\bar{c}_-^{12}}\alpha + \frac{8R^4P^4}{|z|^4(1-c_+)^8\bar{c}_-^8}\bar{\alpha} \ . \tag{2.115}$$

For any z satisfying $|z| > \frac{2RP}{(1-c_+)^2}$, we have $\frac{8R^4P^4}{|z|^4(1-c_+)^8} < 1/2$ and thus

$$\bar{\alpha} < \frac{16P^8(\bar{c}_- + 1)^4}{(1-c_+)^{12}\bar{c}_-^{12}}\alpha \ . \tag{2.116}$$

Plugging this result into (2.112) yields

$$\alpha \leq \frac{128K^4R^8P^8(2-c)^4}{|z|^8(1-c_+)^{12}}\alpha + \frac{8C}{N^2} \ . \tag{2.117}$$

Take $0 < \varepsilon < 1$. It is easy to check that for $|z| > \frac{128^{1/8} RP\sqrt{K(\bar{c}_-+1)}}{(1-c_+)^{3/2}\bar{c}_-^{3/2}(1-\varepsilon)^{1/8}}$, $\frac{128K^4R^8P^8(\bar{c}_-+1)^4}{|z|^8(1-c_+)^{12}\bar{c}_-^{12}} < 1-\varepsilon$ and thus

$$\alpha < \frac{8C}{\varepsilon N^2} \ . \tag{2.118}$$

Since C does not depend on N, α is clearly summable which, along with the Markov inequality and the Borel Cantelli lemma, concludes the proof.

Finally, taking the same steps as previously, we also have

$$\mathbb{E}\left[|m_N(z) - \bar{m}_N(z)|^4\right] \leq \frac{8C}{\varepsilon N^2} \tag{2.119}$$

for some $|z|$ large enough. For these z, the same conclusion holds: $m_N(z) - \bar{m}_N(z) \xrightarrow{\text{a.s.}} 0$. From Vitali convergence theorem, since f_i and e_i are uniformly bounded on all closed sets of $\mathbb{C} \setminus \mathbb{R}^+$, we finally have that the convergence is true for all $z \in \mathbb{C} \setminus \mathbb{R}^+$. The almost sure convergence of the Stieltjes transform implies the almost sure weak convergence of $F_N - \bar{F}_N$ to 0, uniformly over every compact set of $\mathbb{R}^+$, which is our final result.

This concludes the proof of Theorem 15 for the matrices $\mathbf{R}_i$ with surely bounded spectral norms.

Almost sure boundedness of $\|\mathbf{R}_i\|$

To extend Theorem 15 to the case where $\|\mathbf{R}_i\|$ is only almost surely bounded, we merely apply the Fubini theorem.

Call $(\Omega_R, \mathcal{F}_R, P_R)$ the probability space that generates the sequences of matrices of growing sizes $\{\mathbf{R}_i, 1 \leq i \leq K, N_i \in \mathbb{N}\}$ and $(\Omega_W, \mathcal{F}_W, P_W)$ the probability space that generates the sequences of matrices of growing sizes $\{\mathbf{W}_i, 1 \leq i \leq K, N_i \in \mathbb{N}\}$ and $(\Omega_R \times \Omega_W, \mathcal{F}_R \times \mathcal{F}_W, Q)$. Denote A the subspace of $\mathcal{F}_R \times \mathcal{F}_W$ for which $F_N - \bar{F}_N \to 0$. Then, from the Fubini theorem,

$$Q(A) = \int_{\Omega_R \times \Omega_W} 1_A(r,w) Q(d(r,w)) = \int_{\Omega_R} \int_{\Omega_W} 1_A(r,w) P_W(dw) P_R(dr). \tag{2.120}$$

Take r such that the $\|\mathbf{R}_i\|$ are all uniformly bounded with growing N. Then, from Theorem 15, for this r, $\int_{\Omega_W} 1_A(r,w)P_W(dw) = 1$. But these $r \in \Omega_R$ belong to a space of probability one, as the intersection of K spaces of probability one, and finally $Q(A) = 1$.

2.5.3 Proof of Theorem 18

It is easy to see (e.g. [71, Definition 3.2]) that, for F a probability distribution function with support in $\mathbb{R}^+$

$$\int_0^\infty \log\left(1 + \frac{t}{x}\right) dF(t) = \int_x^\infty \left(-\frac{1}{t} + m_F(-t)\right) dF(t) \tag{2.121}$$

where $m_F(z)$ is the Stieltjes transform of F (this is sometimes called the Shannon-transform in $1/x$). In particular,

$$I_N(x) = \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N\right) = \int_x^\infty \left(-\frac{1}{t} + m_N(-t)\right) dF_N(t) \tag{2.122}$$

where $m_N(z) = \frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$.

We will first show that the expression $\bar{I}_N(x)$ given in Theorem 18 satisfies the same property with $\bar{F}_N$. For notational simplicity, we will write $e_i = e_i(-x)$ and $\bar{e}_i = \bar{e}_i(-x)$.

First note that the system of equations in Theorem 15 (i) is unchanged if we extend the $\mathbf{P}_i$ matrices into $N_i \times N_i$ diagonal matrices filled with $N_i - n_i$ zero eigenvalues. Therefore, we can assume that all $\mathbf{P}_i$ have size $N_i \times N_i$ although we restrict the $F^{\mathbf{P}_i}$ to have a mass $1 - c_i$ in zero. Since this does not alter the equations in Theorem 15 (i), we have in particular $\bar{e}_i < \bar{c}_i/e_i$ for $x > 0$.

This being said, $\bar{I}_N(x)$ is given by

$$\begin{aligned}\bar{I}_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{i=1}^{K}\bar{e}_i\mathbf{R}_i\right)\\ + \sum_{i=1}^{K}\left[\frac{1}{N}\log\det\left([\bar{c}_i - e_i\bar{e}_i]\mathbf{I}_N + e_i\mathbf{P}_i\right) - \bar{c}_i\log(\bar{c}_i)\right].\end{aligned} \tag{2.123}$$

Calling $\bar{I}$ the function

$$\begin{aligned}&\bar{I} : (x_1, \ldots, x_K, \bar{x}_1, \ldots, \bar{x}_K, x)\\ &\mapsto \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{i=1}^{K}\bar{x}_i\mathbf{R}_i\right)\\ &\quad + \sum_{i=1}^{K}\left[\frac{1}{N}\log\det\left([\bar{c}_i - x_i\bar{x}_i]\mathbf{I}_N + x_i\mathbf{P}_i\right) - \bar{c}_i\log(\bar{c}_i)\right],\end{aligned} \tag{2.124}$$

we have

$$\frac{\partial\bar{I}}{\partial x_i}(e_1, \ldots, e_K, \bar{e}_1, \ldots, \bar{e}_K, x) = \bar{e}_i - \bar{e}_i\frac{1}{N}\sum_{l=1}^{N_i}\frac{1}{\bar{c}_i - e_i\bar{e}_i + e_ip_{il}} \tag{2.125}$$

$$\frac{\partial\bar{I}}{\partial\bar{x}_i}(e_1, \ldots, e_K, \bar{e}_1, \ldots, \bar{e}_K, x) = e_i - e_i\frac{1}{N}\sum_{l=1}^{N_i}\frac{1}{\bar{c}_i - e_i\bar{e}_i + e_ip_{il}}. \tag{2.126}$$

In order to proceed, note that we can write $\bar{c}_i$ in the following way:

$$\begin{aligned}\bar{c}_i &= \frac{1}{N}\sum_{l=1}^{N_i}\frac{\bar{c}_i - e_i\bar{e}_i + e_ip_{il}}{\bar{c}_i - e_i\bar{e}_i + e_ip_{il}}\\ &= (\bar{c}_i - e_i\bar{e}_i)\frac{1}{N}\sum_{l=1}^{N_i}\frac{1}{\bar{c}_i - e_i\bar{e}_i + e_ip_{il}} + \frac{1}{N}\sum_{l=1}^{N_i}\frac{e_ip_{il}}{\bar{c}_i - e_i\bar{e}_i + e_ip_{il}}\\ &= (\bar{c}_i - e_i\bar{e}_i)\frac{1}{N}\sum_{l=1}^{N_i}\frac{1}{\bar{c}_i - e_i\bar{e}_i + e_ip_{il}} + e_i\bar{e}_i\end{aligned} \tag{2.127}$$

from which it follows that

$$(\bar{c}_i - e_i\bar{e}_i)\left(1 - \frac{1}{N}\sum_{l=1}^{N_i}\frac{1}{\bar{c}_i - e_i\bar{e}_i + e_ip_{il}}\right) = 0. \tag{2.128}$$

But we also know that $0 \leq \bar{e}_i < \bar{c}_i/e_i$ and therefore $\bar{c}_i - e_i\bar{e}_i > 0$. This entails

$$\frac{1}{N}\sum_{l=1}^{N_i} \frac{1}{\bar{c}_i - e_i\bar{e}_i + e_i p_{il}} = 1. \tag{2.129}$$

From (2.129), we can conclude that

$$\frac{\partial \bar{I}}{\partial x_i}(e_1, \ldots, e_K, \bar{e}_1, \ldots, \bar{e}_K, x) = 0 \tag{2.130}$$

$$\frac{\partial \bar{I}}{\partial \bar{x}_i}(e_1, \ldots, e_K, \bar{e}_1, \ldots, \bar{e}_K, x) = 0. \tag{2.131}$$

We therefore have, from the chain rule of differentiation,

$$\begin{aligned}
\frac{d}{dx}\bar{I}_N(x) &= \sum_{i=1}^{K}\left[\frac{\partial \bar{I}}{\partial e_i}\frac{\partial e_i}{\partial x} + \frac{\partial \bar{I}}{\partial \bar{e}_i}\frac{\partial \bar{e}_i}{\partial x}\right] + \frac{\partial \bar{I}}{\partial x} \\
&= \frac{\partial \bar{I}}{\partial x} \\
&= -\frac{1}{x^2}\sum_{i=1}^{K}\bar{e}_i \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{I}_N + \frac{1}{x}\sum_{j=1}^{K}\bar{e}_j\mathbf{R}_j\right)^{-1} \\
&= -\frac{1}{x}\frac{1}{N}\operatorname{tr}\left[\left(\sum_{i=1}^{K}\frac{1}{x}\bar{e}_i\mathbf{R}_i + \mathbf{I}_N - \mathbf{I}_N\right)\left(\mathbf{I}_N + \frac{1}{x}\sum_{j=1}^{K}\bar{e}_j\mathbf{R}_j\right)^{-1}\right] \\
&= -\frac{1}{x} + \frac{1}{N}\operatorname{tr}\left(x\mathbf{I}_N + \sum_{j=1}^{K}\bar{e}_j\mathbf{R}_j\right)^{-1}
\end{aligned} \tag{2.132}$$

Recognizing the Stieltjes transform of $\bar{F}_N$, we therefore have, along with the fact that $\bar{I}_N(\infty) = 0$,

$$\bar{I}_N(x) = \int_x^{\infty}\left(\frac{1}{t} - \frac{1}{t^2}\bar{m}_N\left(-\frac{1}{t}\right)\right)dt \tag{2.133}$$

and therefore

$$\bar{I}_N(x) = \int_0^{\infty}\log\left(1 + \frac{t}{x}\right)d\bar{F}_N(t). \tag{2.134}$$

In order to prove the almost sure convergence $I_N(x) - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0$, we simply need to remark that the support of the eigenvalues of $\mathbf{B}_N$ is bounded. Indeed, the non-zero eigenvalues of $\mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}$ have unit modulus and therefore $\|\mathbf{B}_N\| \leq KPR$. Similarly, the support of $\bar{F}_N$ is the support of the eigenvalues of $\sum_{i=1}^{K}\bar{e}_i\mathbf{R}_i$, which are bounded by KPR as well.

As a consequence, for $\mathbf{B}_1, \mathbf{B}_2, \ldots$ a realization for which $F_N - \bar{F}_N \Rightarrow 0$ (these lie in a space of probability one), we have, from Lemma 4

$$\int_0^{\infty}\log\left(1 + \frac{t}{x}\right)d[F_N - \bar{F}_N](t) \xrightarrow{\text{a.s.}} 0 \tag{2.135}$$

Hence the almost sure convergence of the instantaneous mutual information.

Because of sure boundedness of $\|\mathbf{B}_N\|$, an immediate application of the dominated convergence theorem on the probability space Ω that engenders the sequences of matrices $\mathbf{B}_1(\omega), \mathbf{B}_2(\omega), \ldots$, $\omega \in \Omega$, entails convergence in the first mean as well.

2.5.4 Proof of Theorem 19

Let $z > 0$. From Theorem 4, it is easy to see that the moments $\bar{M}_k$ of the distribution function $\bar{F}_N$ can be obtained through successive differentiation of the function $\frac{1}{z}\frac{1}{N}\mathbf{T}_N(-\frac{1}{z})$ (where $\mathbf{T}_N(z)$ is given by Theorem 14 (ii)), i.e.,

$$\begin{aligned}
\bar{M}_k &= \frac{(-1)^k}{k!}\frac{d^k}{dz^k}\left(\frac{1}{z}\frac{1}{N}\operatorname{tr}\mathbf{T}_N(-1/z)\right)\bigg|_{z=0} \\
&= \frac{(-1)^k}{k!}\frac{d^k}{dz^k}\frac{1}{N}\operatorname{tr}\left(\frac{1}{N}\sum_{j=1}^{n}\frac{z\mathbf{R}_j}{1+\delta_j(-1/z)}+\mathbf{I}_N\right)^{-1}\Bigg|_{z=0} \\
&= \frac{(-1)^k}{k!}\frac{1}{N}\operatorname{tr}\mathbf{T}_{N,k}(0) \qquad (2.136)
\end{aligned}$$

where

$$\mathbf{T}_{N,0}(z) \triangleq \left(\frac{1}{N}\sum_{j=1}^{n}\frac{z\mathbf{R}_j}{1+\delta_{j,0}(z)}+\mathbf{I}_N\right)^{-1} \qquad (2.137)$$

and $(\delta_{1,0}(z), \ldots, \delta_{n,0}(z))$ is the unique positive solution to

$$\delta_{j,0}(z) = \frac{1}{n}\operatorname{tr}\mathbf{R}_j^{\mathsf{H}}\mathbf{T}_0(z), \quad j = 1, \ldots, n. \qquad (2.138)$$

Denote $\mathbf{T}_{N,k}(z) = \frac{d^k\mathbf{T}_{N,0}(z)}{dz^k}$. In order to find the derivatives $\mathbf{T}_{N,k}(z)$, we need the following additional definitions. For $j \in \{1, \ldots, n\}$, let

$$\begin{aligned}
g_{j,0}(z) &= z\delta_{j,0}(z) &(2.139)\\
f_{j,0}(z) &= -\frac{1}{1+g_{j,0}(z)} &(2.140)\\
t_{j,0}(z) &= zf_{j,0}(z) &(2.141)
\end{aligned}$$

and denote $\delta_{j,k}(z)$, $g_{j,k}(z)$, $f_{j,k}(z)$, and $t_{j,k}(z)$ their kth derivatives, respectively. Furthermore, let

$$\mathbf{Q}_0(z) = \frac{1}{N}\sum_{j=1}^{n}t_{j,0}(z)\mathbf{R}_j \qquad (2.142)$$

and denote $\mathbf{Q}_k(z) = \frac{d^k\mathbf{Q}_0(z)}{dz^k}$. With theses definitions, we arrive at

$$\mathbf{T}_{N,1}(z) = \mathbf{T}_{N,0}(z)\underbrace{\mathbf{Q}_1(z)\mathbf{T}_{N,0}(z)}_{\triangleq\mathbf{G}_0(z)}. \qquad (2.143)$$

From the Leibniz-rule for the kth derivative of the product of two functions[1]

$$\mathbf{T}_{N,k+1}(z) = \sum_{i=0}^{k} \binom{k}{i} \mathbf{T}_{N,k-i}(z)\mathbf{G}_i(z), \quad n \geq 0 \tag{2.144}$$

$$\mathbf{G}_k(z) = \sum_{i=0}^{k} \binom{k}{i} \mathbf{Q}_{k-i+1}(z)\mathbf{T}_{N,i}(z), \quad n \geq 0 \tag{2.145}$$

where $\mathbf{G}_k(z) = \frac{d^k \mathbf{G}_0(z)}{dz^k}$. Replacing the last equation in the second last yields

$$\mathbf{T}_{N,k+1}(z) = \sum_{i=0}^{k} \sum_{j=0}^{i} \binom{k}{i} \binom{i}{j} \mathbf{T}_{N,k-i}(z)\mathbf{Q}_{i-j+1}(z)\mathbf{T}_{N,j}(z). \tag{2.146}$$

Straight-forward differentiation of $\mathbf{Q}_0(z)$ leads to

$$\mathbf{Q}_k(z) = \frac{1}{N} \sum_{j=1}^{n} t_{j,k}(x)\mathbf{R}_j, \quad k \geq 0. \tag{2.147}$$

The last step is to find explicit expressions of $t_{j,k}(z)$. From the Leibniz-rule

$$t_{j,k}(z) = k f_{j,k-1}(z) + z f_{j,k}(z) \ , \quad k \geq 0. \tag{2.148}$$

Consider now $f_{j,1}(z)$ the first derivative of $f_{j,0}(z)$:

$$f_{j,1}(z) = \frac{g_{j,1}(z)}{(1 + g_{j,0})^2} = \underbrace{f_{j,0}^2(z)}_{\triangleq r_{j,0}(z)} g_{j,1}(z). \tag{2.149}$$

The higher order derivatives are calculated as

$$f_{j,k+1}(z) = \sum_{i=0}^{k} \binom{k}{i} r_{j,i}(z) g_{j,k-i+1}(z) \tag{2.150}$$

where

$$r_{j,k}(z) = \sum_{i=0}^{k} \binom{k}{i} f_{j,i}(zk) f_{j,k-i}(z). \tag{2.151}$$

Combining the last two equations yields

$$f_{j,k+1}(z) = \sum_{i=0}^{k} \sum_{l=0}^{i} \binom{k}{i} \binom{i}{l} f_{j,l}(z) f_{j,i-l}(z) g_{j,k-i+1}(z) \tag{2.152}$$

where $g_{j,k}(z)$ can be easily calculated as

$$g_{j,k}(z) = k\delta_{j,k-1}(z) + z\delta_{j,k}(z) \tag{2.153}$$

[1]For two functions $u(x)$ and $v(x)$, $\frac{d^n(u(x)v(x))}{dx^n} = \sum_{i=0}^{n} \binom{n}{i} \frac{d^{n-i}u(x)}{dx^{n-i}} \frac{d^i v(x)}{dx^i}$.

and $\delta_{j,k}(z)$ is given by

$$\delta_{j,k}(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_j \mathbf{T}_{N,k}(z). \tag{2.154}$$

Since we are only interested in the case $z = 0$, we will drop from now on the dependence on z and write, e.g., $\mathbf{T}_{N,k}$ instead of $\mathbf{T}_{N,k}(0)$. In this case, the expressions of $g_{j,k}(z)$ and $t_{j,k}(z)$ simplify to

$$g_{j,k} = k\delta_{j,k-1} \tag{2.155}$$

$$t_{j,k} = kf_{j,k-1}. \tag{2.156}$$

Replacing these quantities in (2.147) and (2.152), together with (2.146) and (2.154) leads to the desired result. Note that $\mathbf{T}_{N,0} = \mathbf{I}_N$, $f_{j,0} = -1$ and $\delta_{j,0} = \frac{1}{N} \operatorname{tr} \mathbf{R}_j$. Moreover, $\mathbf{T}_{N,k+1}$ depends on $\mathbf{T}_{N,0}, \dots, \mathbf{T}_{N,k}$ and $\mathbf{Q}_1, \dots, \mathbf{Q}_{k+1}$. Since $\mathbf{Q}_{k+1}$ depends only on $f_{j,k}$, $\mathbf{T}_{N,k+1}$ can be recursively calculated.

2.5.5 Proof of Theorem 20

Both $\frac{1}{N} \operatorname{tr} \mathbf{D}_N (\mathbf{B}_N - z\mathbf{I}_N)^{-1}$ and $\frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}_N(z)$ as defined in Theorem 14 are Stieltjes transforms of finite measures which we denote by π and $\overline{\pi}$, respectively. Thus, Theorem 14 and Theorem 5 imply together that, almost surely,

$$\pi - \overline{\pi} \Rightarrow 0.$$

Similar to the proof of Theorem 19 we can express the moments of π and $\overline{\pi}$ as

$$\begin{aligned}\int \lambda^k \pi(d\lambda) &= \frac{(-1)^k}{k!} \frac{d^k}{dz^k} \left(\frac{1}{z} \frac{1}{N} \operatorname{tr} \mathbf{D}_N \left(\mathbf{B}_N + \frac{1}{z} \mathbf{I}_N \right)^{-1} \right) \Bigg|_{z=0} \\ &= \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{B}_N^k\end{aligned}$$

and

$$\begin{aligned}\int \lambda^k \overline{\pi}(d\lambda) &= \frac{(-1)^k}{k!} \frac{d^k}{dz^k} \left(\frac{1}{z} \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}_N(-1/z) \right) \Bigg|_{z=0} \\ &= \frac{(-1)^k}{k!} \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}_{N,k}.\end{aligned}$$

The support of π is almost surely compact as $\mathbf{D}_N$ has bounded spectral norm and the spectral norm of $\mathbf{B}_N$ is almost surely bounded due to the following inequalities:

$$\begin{aligned}\|\mathbf{B}_N\| &\leq \sum_{l=1}^{L} \left\| \tilde{\mathbf{R}}_l \right\| \left\| \frac{1}{N} \mathbf{Z}_l \mathbf{Z}_l^{\mathsf{H}} \right\| \\ &\leq LR \max_l \frac{n_l}{N} \left\| \frac{1}{n_l} \mathbf{Z}_l \mathbf{Z}_l^{\mathsf{H}} \right\| \\ &\xrightarrow{\text{a.s.}} LR \max_l \left(1 + \sqrt{\frac{n_l}{N}} \right)^2 < \infty\end{aligned}$$

for some $R \geq \limsup_N \max_l \|\tilde{\mathbf{R}}_l\|$, $n_l \triangleq \sum_{j=1}^{n} \mathbb{1}\{\mathbf{R}_j = \tilde{\mathbf{R}}_l\}$, and $\mathbf{Z}_l \in \mathbb{C}^{N \times n_l}$ being random matrices of i.i.d. elements with zero mean, unit variance and finite eighth moment. The almost sure convergence of the spectral norm in the last step follows from Theorem 8. The almost sure weak convergence of $\pi - \overline{\pi} \to 0$ implies by Lemma 4, that

$$\int f(\lambda)\pi(d\lambda) - \int f(\lambda)\overline{\pi}(d\lambda) \xrightarrow{\text{a.s.}} 0 \tag{2.157}$$

for any bounded, continuous function. Since the support of π is almost surely bounded and the support of $\overline{\pi}$ can be shown to be bounded following similar steps as in [83, Proof of Theorem 2, Part B], the convergence in (2.157) also holds $f(\lambda) = \lambda^k$. This concludes the proof.

2.5.6 Justification of Claim 1

In the following, $z = -x$. If we define $\delta_j = \frac{1}{n} \operatorname{tr} \mathbf{D}_j \mathbf{T}(-x)$, $\widetilde{\delta}_i = \frac{1}{n} \operatorname{tr} \widetilde{\mathbf{D}}_i \widetilde{\mathbf{T}}(-x)$, and let $\mathbf{\Delta} = \operatorname{diag}(\delta_j, 1 \leq j \leq n)$ and $\widetilde{\mathbf{\Delta}} = \operatorname{diag}\left(\widetilde{\delta}_i, 1 \leq i \leq N\right)$, then, the system of $N+n$ equations in Theorem 12 (i) can be written in an equivalent way as:

$$\delta_j = \Gamma_j(\mathbf{\Delta}, \widetilde{\mathbf{\Delta}}) \ , \quad 1 \leq j \leq n \tag{2.158}$$

$$\widetilde{\delta}_i = \widetilde{\Gamma}_i(\mathbf{\Delta}, \widetilde{\mathbf{\Delta}}) \ , \quad 1 \leq i \leq N \tag{2.159}$$

where

$$\Gamma_j(\mathbf{\Delta}, \widetilde{\mathbf{\Delta}}) \triangleq \frac{1}{n} \operatorname{tr} \mathbf{D}_j \left[x\left(\mathbf{I}_N + \widetilde{\mathbf{\Delta}}\right) + \mathbf{A}\left(\mathbf{I}_n + \mathbf{\Delta}\right)^{-1} \mathbf{A}^{\mathsf{H}} \right]^{-1} \tag{2.160}$$

$$\widetilde{\Gamma}_i(\mathbf{\Delta}, \widetilde{\mathbf{\Delta}}) \triangleq \frac{1}{n} \operatorname{tr} \widetilde{\mathbf{D}}_i \left[x\left(\mathbf{I}_n + \mathbf{\Delta}\right) + \mathbf{A}^{\mathsf{H}} \left(\mathbf{I}_N + \widetilde{\mathbf{\Delta}}\right)^{-1} \mathbf{A} \right]^{-1} . \tag{2.161}$$

It turns out that the functions Γ_j and $\widetilde{\Gamma}_i$ will help in providing a concise expression for the asymptotic variance of the random variable $N(I_N(x) - \bar{I}_N(x))$. Consider the $(N+n)$-variate function $\mathbf{\Gamma}(\mathbf{\Delta}, \widetilde{\mathbf{\Delta}}) = \left(\Gamma_1, \ldots \Gamma_n, \widetilde{\Gamma}_1, \ldots, \widetilde{\Gamma}_N\right)$. It has been shown in several cases, e.g., for a separable variance profile with centered Gaussian entries [95, 81] and for a general variance profile with centered entries [51] (see also Theorem 10 for a degenerate case), that the asymptotic variance $\mathbf{\Theta}_{N,n}^2$ can be written in the following way

$$\mathbf{\Theta}_{N,n}^2 = -\log\det\left(\mathbf{I}_{N+n} - \mathbf{J}\right) \tag{2.162}$$

where $\mathbf{J}$ is the Jacobian matrix of the function $\mathbf{\Gamma}(\mathbf{\Delta}, \widetilde{\mathbf{\Delta}})$. Thus, we claim that also in our case, the last expressions holds.

2.5.7 Proof of Theorem 21

The proof for $\mathbf{\Theta}_N$ invertible is given in [70] (see also [84, Proof of Lemma 3]). Assume now $\mathbf{\Theta}_N$ to be a non-invertible Hermitian nonnegative-definite matrix. Denote by $\mathbf{T}'(z, \mathbf{A})$ the matrix $\mathbf{T}'(z)$ as given by Theorem 21 for $\mathbf{\Theta}_N = \mathbf{A}$ and denote by $m'(z, \mathbf{A})$ the function

$$m'(z, \mathbf{A}) = \frac{1}{N} \operatorname{tr} \mathbf{D}_N \left(\mathbf{B}_N + \mathbf{S} - z\mathbf{I}_N\right)^{-1} \mathbf{A} \left(\mathbf{B}_N + \mathbf{S} - z\mathbf{I}_N\right)^{-1} . \tag{2.163}$$

Then, the following convergence holds

$$m'(z, \mathbf{\Theta}_N + \mathbf{I}_N) - \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}'(z, \mathbf{\Theta}_N + \mathbf{I}_N) \xrightarrow{\text{a.s.}} 0. \tag{2.164}$$

It is straight-forward to show that

$$m'(z, \mathbf{\Theta}_N + \mathbf{I}_N) - m'(z, \mathbf{\Theta}_N) = m'(z, \mathbf{I}_N) \tag{2.165}$$

$$\frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}'(z, \mathbf{\Theta}_N + \mathbf{I}_N) - \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}'(z, \mathbf{\Theta}_N) = \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}'(z, \mathbf{I}_N). \tag{2.166}$$

Hence,

$$\begin{aligned} m'(z, \mathbf{\Theta}_N) - \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}'(z, \mathbf{\Theta}_N) &= m'(z, \mathbf{\Theta}_N + \mathbf{I}_N) - \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}'(z, \mathbf{\Theta}_N + \mathbf{I}_N) \\ &\quad - \left(m'(z, \mathbf{I}_N) - \frac{1}{N} \operatorname{tr} \mathbf{D}_N \mathbf{T}'(z, \mathbf{I}_N) \right) \\ &\xrightarrow{\text{a.s.}} 0. \end{aligned} \tag{2.167}$$

2.5.8 Proof of Theorem 22

Part (i)

The proof follows essentially the same steps as the proof of Theorem 23 (i) and will not be given in full detail here. In order to prove the uniqueness of solutions $(\bar{g}_k, g_k, \delta_k)$, it is sufficient to show by Theorem 16 that the K-variate function $\mathbf{h} : (x_1, \ldots, x_K) \mapsto (h_1, \ldots, h_K)$ as defined below, is a standard interference function (see Definition 11). For $k = 1, \ldots, K$, we define

$$h_k(x_1, \ldots, x_K) \mapsto \frac{1}{n_k} \sum_{j=1}^{N_k} \frac{s_{k,j} \delta_k}{1 + \bar{g}_k s_{k,j} \delta_k} \tag{2.168}$$

where

$$\bar{g}_k = \frac{1}{n_k} \operatorname{tr} \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}} \left(x_k \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}} + \mathbf{I}_{n_k} \right)^{-1} \tag{2.169}$$

and $\delta_k,\ k = 1, \ldots, K$, form the unique jointly positive solution to the following fixed-point equations

$$\delta_k = \frac{1}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{k=1}^{K} \frac{n_k}{N_k} \frac{\bar{g}_k x_k}{\delta_k} \mathbf{R}_k + x \mathbf{I}_N \right)^{-1}. \tag{2.170}$$

The only difference to Theorem 23 (i) is the definition of $\bar{g}_k$. In our case, $\bar{g}_k$ is directly defined as a function of x_k, whereas $\bar{b}_k$ in Theorem 23 (i) is given as the solution of another fixed point equation. However, the behavior of $\bar{b}_k$ and $\bar{g}_k$ as seen as functions of x_k is identical. In particular, let $x_k > x'_k > 0$ and denote by $\bar{g}_k$ and $\bar{g}'_k$ the corresponding values of (2.169), respectively. One can easily verify that the following conditions hold:(i) $\bar{g}_k < \bar{g}'_k$ and (ii) $x_k \bar{g}_k > x'_k \bar{g}'_k$. The remaining steps are identical to the proof of Theorem 23 (i) and will not be repeated here. By showing $\mathbf{h}(x_1, \ldots, x_K)$ to be a standard interference function, we have proved by Theorem 16 that the following fixed-point algorithm, which iteratively computes

$$x_k^{t+1} = h_k(x_1^{(t)}, \ldots, x_K^{(t)}), \qquad k = 1, \ldots, K \tag{2.171}$$

for $t \geq 0$ and some set of initial values $x_1^{(0)}, \ldots, x_K^{(0)}$, converges as $t \to \infty$ to the unique fixed point $(g_1, \ldots, g_K)$.

Part *(ii)*

The key idea is that the random matrix model can be considered as the Kronecker channel model as considered in Theorem 13 with *random* correlation matrices. Assume now a Kronecker model, for which the matrices $\mathbf{H}_k$ are given as

$$\mathbf{H}_k = \frac{1}{\sqrt{n_k}} \mathbf{Z}_k \mathbf{W}_{2,k} \mathbf{T}_k^{\frac{1}{2}} \tag{2.172}$$

where $\mathbf{Z}_k \in \mathbb{C}^{N \times N_k}$ is a *deterministic* matrix and $\mathbf{W}_{2,k}$ and $\mathbf{T}_k$ are defined as in the statement of the theorem. Further assume that $\limsup_N \|\mathbf{Z}_k\| < \infty$ for all k. Thus, we can apply Theorem 13 *(iii)* to obtain the following deterministic equivalent $\bar{V}_N(x)$ of $I_N(x)$:

$$\begin{aligned}\bar{V}_N(x) = \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K} \bar{e}_k \mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}}\right) + \sum_{k=1}^{K} \frac{1}{N} \log\det\left(\mathbf{I}_{n_k} + e_k \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}}\right) \\ - \frac{1}{N}\sum_{k=1}^{K} n_k e_k \bar{e}_k \end{aligned} \tag{2.173}$$

where $\bar{e}_k, e_k, k = 1, \ldots, K$, are given as the unique solutions to the following equations

$$\begin{aligned}\bar{e}_k &= \frac{1}{n_k} \operatorname{tr} \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}} \left(e_k \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}} + \mathbf{I}_{n_k}\right)^{-1} \\ e_k &= \frac{1}{n_k} \operatorname{tr} \mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}} \left(\sum_{i=1}^{K} \bar{e}_i \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}\end{aligned} \tag{2.174}$$

such that $\bar{e}_k, e_k > 0$ for all k. Note that this matrix model is equivalent to the model $\mathbf{H}_k = \frac{1}{\sqrt{n_k}} \mathbf{Z}_k \mathbf{W}_{2,k} \tilde{\mathbf{T}}_k^{\frac{1}{2}}$, where $\tilde{\mathbf{T}}_k^{\frac{1}{2}} = \mathbf{T}^{\frac{1}{2}} \mathbf{Q}_k^{\frac{1}{2}}$.

For our matrix model, the matrices $\mathbf{Z}_k$ are *random* and defined as

$$\mathbf{Z}_k = \frac{1}{\sqrt{N_k}} \mathbf{R}_k^{\frac{1}{2}} \mathbf{W}_{1,k} \mathbf{S}_k^{\frac{1}{2}}. \tag{2.175}$$

Let $(\Omega, \mathcal{F}, P)$ be the probability space generating the random sequences of matrices $(\mathbf{W}_{1,k}(\omega))_{N \geq 1}$. There exists $A \subset \Omega$ with $P(A) = 1$, such that for each $\omega \in A$, we have $\limsup_N \|\mathbf{Z}_k(\omega)\mathbf{Z}_k(\omega)^{\mathsf{H}}\| < \infty$ (see Theorem 8). Thus, for each of these ω, the matrices $\mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}}$ satisfy the criteria of the correlation matrices of Theorem 13. Let $(\Omega', \mathcal{F}', P')$ the probability space generating the matrices $\mathbf{W}_{2,k}$. Thus, for every $\omega \in A$, there exist a $A'(\omega) \subset \Omega'$ with $P'(A') = 1$, such that for all $\omega' \in A'(\omega)$, $\bar{V}_N(x)$ is a deterministic equivalent of $I_N(x)$. Denote by $(\Omega \times \Omega', \mathcal{F} \times \mathcal{F}', Q)$ the product space generating the matrices $\mathbf{W}_{1,k}(\omega)$ and $\mathbf{W}_{1,k}(\omega')$ and denote by $B \subset \Omega \times \Omega'$ the space of all tuples (ω, ω'), such that $\omega \in A$ and $\omega' \in A'(\omega)$. By the Fubini theorem, we have $Q(B) = 1$, which proves that $\bar{V}_N(x) - I_N(x) \to 0$, almost surely. However, $\bar{V}_N(x)$ is a random quantity, which depends on the matrices $\mathbf{Z}_k$. Therefore, we will need to obtain an iterative deterministic equivalent $\bar{I}_N(x)$ of $\bar{V}_N(x)$.

The first step is to replace the fixed-point equations (2.174) that depend on $\mathbf{Z}_k$ by deterministic ones. Let us define the quantities $\bar{e}_{k,i,j}$, $e_{k,i,j}$, for $i \in \{1, \dots, K\}$, $j \in \{1, \dots, N_k\}$, which are given as the unique solutions to the following set of fixed-point equations:

$$\begin{aligned} \bar{e}_{k,i,j} &= \frac{1}{n_k} \operatorname{tr} \tilde{\mathbf{T}}_k \left(e_{k,i,j} \tilde{\mathbf{T}}_k + \mathbf{I}_{n_k} \right)^{-1} \\ e_{k,i,j} &= \frac{1}{n_k} \operatorname{tr} \mathbf{Z}_{k,i,j} \mathbf{Z}_{k,i,j}^{\mathsf{H}} \left(\sum_{\ell=1}^{K} \bar{e}_{\ell,i,j} \mathbf{Z}_{\ell,i,j} \mathbf{Z}_{\ell,i,j}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \end{aligned} \tag{2.176}$$

where

$$\mathbf{Z}_{k,i,j} = \begin{cases} \mathbf{Z}_k & , \quad i \neq k \\ [\mathbf{z}_{k,1} \cdots \mathbf{z}_{k,j-1} \mathbf{z}_{k,j+1} \dots \mathbf{z}_{k,N_k}] & , \quad i = k \end{cases}. \tag{2.177}$$

Obviously, $\bar{e}_{k,i,j}$ and $e_{k,i,j}$ are independent of the vector $\mathbf{z}_{i,j}$. In addition, we define

$$Z = \max_k \limsup_N \|\mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}}\|, \quad T = \max_k \limsup_N \|\tilde{\mathbf{T}}_k\|, \quad n = \min_k n_k, \quad c = \frac{N}{n} \tag{2.178}$$

and

$$\alpha_{i,j} = \max_k |e_{k,i,j} - e_k|, \qquad \bar{\alpha}_{i,j} = \max_k |\bar{e}_{k,i,j} - \bar{e}_k|. \tag{2.179}$$

Thus, we have for N large,

$$\begin{aligned} |\bar{e}_{k,i,j} - \bar{e}_k| &= \left| \frac{1}{n_k} \operatorname{tr} \tilde{\mathbf{T}}_k \left(e_{k,i,j} \tilde{\mathbf{T}}_k + \mathbf{I}_{n_k} \right)^{-1} \left((e_k - e_{k,i,j}) \tilde{\mathbf{T}}_k \right) \left(e_k \tilde{\mathbf{T}}_k + \mathbf{I}_{n_k} \right)^{-1} \right| \\ &\leq \alpha_{i,j} T^2. \end{aligned} \tag{2.180}$$

Since the right-hand side (RHS) of the last inequality is independent of k, we have

$$\bar{\alpha}_{i,j} \leq \alpha_{i,j} T^2. \tag{2.181}$$

On the other hand, for $i \neq k$, we have for N sufficiently large,

$$
\begin{aligned}
&|e_{k,i,j} - e_k| \\
&= \left| \frac{1}{n_k} \operatorname{tr} \mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}} \left(\sum_{\ell=1}^{K} \bar{e}_{\ell,i,j} \mathbf{Z}_{\ell,i,j} \mathbf{Z}_{\ell,i,j}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \right. \\
&\qquad \left. \times \left(\sum_{\ell=1}^{K} (\bar{e}_\ell - \bar{e}_{\ell,i,j}) \mathbf{Z}_\ell \mathbf{Z}_\ell^{\mathsf{H}} + \bar{e}_{i,i,j} \mathbf{z}_{i,j} \mathbf{z}_{i,j}^{\mathsf{H}} \right) \left(\sum_{\ell=1}^{K} \bar{e}_\ell \mathbf{Z}_\ell \mathbf{Z}_\ell^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \right| \\
&\leq \frac{cKZ^2}{x^2} \bar{\alpha}_{i,j} + \frac{\bar{e}_{i,i,j}}{n} \left| \mathbf{z}_{i,j}^{\mathsf{H}} \left(\sum_{\ell=1}^{K} \bar{e}_\ell \mathbf{Z}_\ell \mathbf{Z}_\ell^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}} \right. \\
&\qquad \left. \times \left(\sum_{\ell=1}^{K} \bar{e}_{\ell,i,j} \mathbf{Z}_{\ell,i,j} \mathbf{Z}_{\ell,i,j}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \mathbf{z}_{i,j} \right| \\
&\leq \frac{cKZ^2}{x^2} \bar{\alpha}_{i,j} + \frac{Z^2 T}{x^2 n}
\end{aligned}
\tag{2.182}
$$

where the last inequality is due to $\bar{e}_{k,i,j} \leq T$ and $\left| \mathbf{z}_{i,j}^{\mathsf{H}} \mathbf{A} \mathbf{z}_{i,j} \right| \leq \|\mathbf{z}_{i,j}\|^2 \|\mathbf{A}\| \leq Z \|\mathbf{A}\|$, for any matrix $\mathbf{A}$.

Similarly, one can show that

$$
|e_{k,k,j} - e_k| = \frac{cKZ^2}{x^2} \bar{\alpha}_{k,j} + \frac{Z^2 T}{x^2 n} + \frac{Z}{xn}. \tag{2.183}
$$

It follows from (2.182), (2.183) and (2.181), that

$$
\alpha_{i,j} \leq \frac{cKZ^2}{x^2} \bar{\alpha}_{i,j} + \frac{Z^2 T}{x^2 n} + \frac{Z}{xn} \leq \frac{cKZ^2 T^2}{x^2} \alpha_{i,j} + \frac{Z^2 T}{x^2 n} + \frac{Z}{xn}. \tag{2.184}
$$

Now, for any $x \geq \sqrt{\frac{cKZ^2T^2}{1-\epsilon}}$ and $\epsilon > 0$, we have

$$
\alpha_{i,j} \leq \frac{Z}{\epsilon x n} \left(1 + \frac{ZT}{x} \right) \quad , \quad \bar{\alpha}_{i,j} \leq \frac{ZT^2}{\epsilon x n} \left(1 + \frac{ZT}{x} \right).
$$

Let $\mu = \max\{ \frac{Z}{\epsilon x} \left(1 + \frac{ZT}{x} \right), \frac{ZT^2}{\epsilon x} \left(1 + \frac{ZT}{x} \right) \}$, then we finally have

$$
\alpha_{i,j} \leq \frac{\mu}{n} \quad , \quad \bar{\alpha}_{i,j} \leq \frac{\mu}{n}. \tag{2.185}
$$

The last result establishes that for sufficiently large x, the differences between the solutions $(\bar{e}_{k,i,j}, e_{k,i,j})$ to (2.176) and the solutions $(\bar{e}_k, e_k)$ to (2.174) are uniformly bounded by μ and vanish as $n \to \infty$. Moreover, e_k and $e_{k,i,j}$ have an analytic continuation on $z = -x \in \mathbb{C} \setminus \mathbb{R}^+$ and are uniformly bounded on all closed subsets of $z \in \mathbb{C} \setminus \mathbb{R}^+$. Thus, in particular for $x > 0$ and all k, i, j, we have by the Vitali convergence theorem

$$
e_k - e_{k,i,j} \to 0 \qquad \text{and hence} \qquad \bar{e}_k - \bar{e}_{k,i,j} \to 0. \tag{2.186}
$$

As a consequence of (2.186), we can now write

$$
\begin{aligned}
e_k &= \frac{1}{n_k}\sum_{j=1}^{N_k} \mathbf{z}_{k,j}^{\mathsf{H}} \left(\sum_{i=1}^{K} \bar{e}_i \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \mathbf{z}_{k,j} \\
&\overset{(a)}{\asymp} \frac{1}{n_k}\sum_{j=1}^{N_k} \mathbf{z}_{k,j}^{\mathsf{H}} \left(\sum_{i=1}^{K} \bar{e}_{i,k,j} \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \mathbf{z}_{k,j} \\
&\overset{(b)}{=} \frac{1}{n_k}\sum_{j=1}^{N_k} \frac{\mathbf{z}_{k,j}^{\mathsf{H}} \left(\sum_{i=1}^{K} \bar{e}_{i,k,j} \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} - \bar{e}_{k,k,j}\mathbf{z}_{k,j}\mathbf{z}_{k,j}^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \mathbf{z}_{k,j}}{1 + \bar{e}_{k,k,j}\mathbf{z}_{k,j}^{\mathsf{H}} \left(\sum_{i=1}^{K} \bar{e}_{i,j,k} \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} - \bar{e}_{k,k,j}\mathbf{z}_{k,j}\mathbf{z}_{k,j}^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \mathbf{z}_{k,j}} \\
&\overset{(c)}{\asymp} \frac{1}{n_k}\sum_{j=1}^{N_k} \frac{\frac{s_{k,j}}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \bar{e}_{i,j,k} \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}}{1 + \frac{s_{k,j}\bar{e}_{k,k,j}}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \bar{e}_{i,j,k} \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}} \\
&\overset{(d)}{\asymp} \frac{1}{n_k}\sum_{j=1}^{N_k} \frac{\frac{s_{k,j}}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \bar{e}_{i} \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}}{1 + \frac{s_{k,j}\bar{e}_{k}}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \bar{e}_{i} \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}}
\end{aligned}
\tag{2.187}
$$

where (a) follows from (2.186) since

$$
\begin{aligned}
&\left| \mathbf{z}_{k,j}^{\mathsf{H}} \left(\sum_{i=1}^{K} \bar{e}_i \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \mathbf{z}_{k,j} - \mathbf{z}_{k,j}^{\mathsf{H}} \left(\sum_{i=1}^{K} \bar{e}_{i,k,j} \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \mathbf{z}_{k,j} \right| \\
&\leq \frac{\mu K Z}{n x^2} \xrightarrow[N\to\infty]{} 0,
\end{aligned}
\tag{2.188}
$$

(b) is due to Lemma 6, (c) is a consequence of Lemmas 13 and 8 and (d) is obtained by applying (2.186) a second time. Next, we would like to find deterministic equivalents of the terms $\frac{1}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \bar{e}_i \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}$. We cannot directly apply Theorem 13 at this point since the $\bar{e}_k$ are defined as functions of $\mathbf{Z}_k$. However, based on the relations (2.185) and (2.188), Theorem 13 (see [83, Theorem 1]) can be shown to hold also for the matrix model under study. Thus,

$$
\frac{1}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \bar{e}_i \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \asymp f_k \tag{2.189}
$$

where f_k for $k \in \{1, \ldots, K\}$ are defined as the unique solution to the following fixed-point equations

$$
\bar{f}_k = \frac{1}{N_k}\sum_{j=1}^{N_k} \frac{s_{k,j}\bar{e}_k}{1 + \bar{e}_k s_{k,j} f_k} \tag{2.190}
$$

$$
f_k = \frac{1}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \bar{f}_i \mathbf{R}_i + x\mathbf{I}_N\right)^{-1} \tag{2.191}
$$

such that $f_k > 0$ for all k. Replacing (2.190) in (2.191) leads to

$$f_k = \frac{1}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \frac{\bar{e}_i}{N_i} \mathbf{R}_i \sum_{j=1}^{N_i} \frac{s_{i,j}}{1 + s_{i,j}\bar{e}_i f_i} + x\mathbf{I}_N \right)^{-1}. \tag{2.192}$$

Thus,

$$e_k = \frac{1}{n_k} \sum_{j=1}^{N_k} \frac{s_{k,j} f_k}{1 + s_{k,j}\bar{e}_k f_k} + \epsilon_k \tag{2.193}$$

where ϵ_k is a sequence of random variables, satisfying $\epsilon_k \xrightarrow[N\to\infty]{\text{a.s.}} 0$. Consider now the following system of equations

$$\bar{e}_k = \frac{1}{n_k} \operatorname{tr} \tilde{\mathbf{T}}_k \left(e_k \tilde{\mathbf{T}}_k + \mathbf{I}_{n_k} \right)^{-1} \tag{2.194}$$

$$e_k = \frac{1}{n_k} \sum_{j=1}^{N_k} \frac{s_{k,j} f_k}{1 + s_{k,j}\bar{e}_k f_k} + \epsilon_k \tag{2.195}$$

$$f_k = \frac{1}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \frac{\bar{e}_i}{N_i} \mathbf{R}_i \sum_{j=1}^{N_i} \frac{s_{i,j}}{1 + s_{i,j}\bar{e}_i f_i} + x\mathbf{I}_N \right)^{-1} \tag{2.196}$$

and its deterministic counterpart

$$\bar{g}_k = \frac{1}{n_k} \operatorname{tr} \tilde{\mathbf{T}}_k \left(g_k \tilde{\mathbf{T}}_k + \mathbf{I}_{n_k} \right)^{-1} \tag{2.197}$$

$$g_k = \frac{1}{n_k} \sum_{j=1}^{N_k} \frac{s_{k,j} \delta_k}{1 + s_{k,j}\bar{g}_k \delta_k} \tag{2.198}$$

$$\begin{aligned} \delta_k &= \frac{1}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \frac{\bar{g}_i}{N_i} \mathbf{R}_i \sum_{j=1}^{N_i} \frac{s_{i,j}}{1 + s_{i,j}\bar{g}_i \delta_i} + x\mathbf{I}_N \right)^{-1} \\ &= \frac{1}{N_k} \operatorname{tr} \mathbf{R}_k \left(\sum_{i=1}^{K} \frac{n_i}{N_i} \frac{\bar{g}_i g_i}{\delta_i} \mathbf{R}_i + x\mathbf{I}_N \right)^{-1}. \end{aligned} \tag{2.199}$$

Define the quantities:

$$\gamma_1 = \max_k |e_k - g_k|, \quad \gamma_2 = \max_k |\bar{e}_k - \bar{g}_k|, \quad \gamma_3 = \max_k |f_k - \delta_k|, \quad \epsilon = \max_k |\epsilon_k|.$$

Straight-forward calculations lead to the following bounds:

$$\gamma_1 \le cS\gamma_3 + \frac{cS^2R^2}{x^2}\gamma_2 + \epsilon, \quad \gamma_2 \le \gamma_1 T^2, \quad \gamma_3 \le \frac{KSR^2}{x^2}\gamma_2 + \frac{KS^2R^2T^2}{x^2}\gamma_1. \tag{2.200}$$

Combining these results and using the fact that $\epsilon \xrightarrow{\text{a.s.}} 0$ yields for x sufficiently large,

$$\gamma_1, \gamma_2, \gamma_3 \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{2.201}$$

Since $e_k, g_k, \bar{e}_k, \bar{g}_k, f_k, \delta_k$ have analytic extensions in a neighborhood of $\mathbb{R}^+$ on which they are all (almost surely) bounded, we have by the Vitali convergence theorem that (2.201) holds for any $x > 0$.

Coming now back to $\bar{V}_N(x)$ as given in (2.173), we have from the continuous mapping theorem that

$$\frac{1}{N}\sum_{k=1}^{K}\left[\log\det\left(\mathbf{I}_{n_k} + e_k\tilde{\mathbf{T}}_k\right) - n_k e_k \bar{e}_k\right] - \frac{1}{N}\sum_{k=1}^{K}\left[\log\det\left(\mathbf{I}_{n_k} + g_k\tilde{\mathbf{T}}_k\right) - n_k g_k \bar{g}_k\right] \xrightarrow{\text{a.s.}} 0. \tag{2.202}$$

Moreover, since $\|\sum_{k=1}^{K}\left(\bar{e}_k - \bar{g}_k\right)\mathbf{Z}_k\mathbf{Z}_k^{\mathsf{H}}\| \xrightarrow{\text{a.s.}} 0$, we have

$$\frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\bar{e}_k\mathbf{Z}_k\mathbf{Z}_k^{\mathsf{H}}\right) - \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\bar{g}_k\mathbf{Z}_k\mathbf{Z}_k^{\mathsf{H}}\right) \xrightarrow{\text{a.s.}} 0. \tag{2.203}$$

Applying Theorem 13 to the last term yields

$$\frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\bar{g}_k\mathbf{Z}_k\mathbf{Z}_k^{\mathsf{H}}\right) - \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\frac{n_k}{N_k}\frac{\bar{g}_k g_k}{\delta_k}\mathbf{R}_k\right) - \frac{1}{N}\sum_{k=1}^{K}\log\det\left(\mathbf{I}_{N_k} + \bar{g}_k\delta_k\mathbf{S}_k\right) + n_k\bar{g}_k g_k \xrightarrow{\text{a.s.}} 0. \tag{2.204}$$

Combining (2.202) and (2.204) finally leads to

$$\bar{V}_N(x) - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0 \tag{2.205}$$

where

$$\bar{I}_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\frac{n_k}{N_k}\frac{\bar{g}_k g_k}{\delta_k}\mathbf{R}_k\right) + \frac{1}{N}\sum_{k=1}^{K}\log\det\left(\mathbf{I}_{N_k} + \bar{g}_k\delta_k\mathbf{S}_k\right) + \frac{1}{N}\sum_{k=1}^{K}\left[\log\det\left(\mathbf{I}_{n_k} + g_k\tilde{\mathbf{T}}_k\right) - 2n_k g_k\bar{g}_k\right]. \tag{2.206}$$

This concludes the proof of part (ii) (a).

In order to show the convergence in the first mean (part (ii) (b)), we will pursue the same approach as in [74]. Define the following functions:

$$m_N(z) = \frac{1}{N}\operatorname{tr}\left(\sum_{k=1}^{K}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} \tag{2.207}$$

$$\bar{m}_N(z) = \frac{1}{N}\operatorname{tr}\left(\sum_{k=1}^{K}\frac{n_k}{N_k}\frac{\bar{g}_k g_k}{\delta_k}\mathbf{R}_k - z\mathbf{I}_N\right)^{-1}. \tag{2.208}$$

One can show that

$$\mathbb{E}\left[I_N(x)\right] - \bar{I}_N(x) = \int_x^\infty \left(\left[\frac{1}{\omega} - \mathbb{E}\left[m_N(-\omega)\right] \right] - \left[\frac{1}{\omega} - \bar{m}_N(-\omega) \right] \right) d\omega. \tag{2.209}$$

Since both $m_N(\omega)$ and $\bar{m}_N(\omega)$ are uniformly bounded by $\frac{1}{\omega}$, it follows from dominated convergence arguments, Theorem 13 (ii) and (2.201) that, for all $\omega > 0$,

$$\left[\frac{1}{\omega} - \mathbb{E}\left[m_N(-\omega)\right] \right] - \left[\frac{1}{\omega} - \bar{m}_N(-\omega) \right] \to 0. \tag{2.210}$$

Moreover,

$$\begin{aligned}
&\left| \left[\frac{1}{\omega} - \mathbb{E}\left[m_N(-\omega)\right] \right] - \left[\frac{1}{\omega} - \bar{m}_N(-\omega) \right] \right| \\
&\leq \left| \left[\frac{1}{\omega} - \mathbb{E}\left[m_N(-\omega)\right] \right] \right| + \left| \left[\frac{1}{\omega} - \bar{m}_N(-\omega) \right] \right| \qquad (2.211)\\
&\leq \frac{1}{\omega^2} \left(\frac{1}{N} \operatorname{tr} \mathbb{E}\left[\sum_{k=1}^K \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^{\mathsf{H}} \right] + \frac{1}{N} \operatorname{tr} \left(\sum_{k=1}^K \frac{n_k}{N_k} \frac{\bar{g}_k g_k}{\delta_k} \mathbf{R}_k \right) \right) \\
&\leq \frac{2KRST}{\omega^2} \qquad (2.212)
\end{aligned}$$

where $R = \max_k \limsup \|\mathbf{R}_k\|$, $S = \max_k \limsup \|\mathbf{S}_k\|$, $T = \max_k \limsup \|\mathbf{T}_k \mathbf{Q}_k\|$. Since $\frac{2KRST}{\omega^2}$ is finite and integrable over $[x, \infty)$, it follows from the dominated convergence theorem that

$$\mathbb{E}\left[I_N(x)\right] - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0. \tag{2.213}$$

2.5.9 Proof of Theorem 23

Part (i)

It was shown in (2.86) that, for any fixed $b_k \geq 0$, the equation in $\bar{b}_k$:

$$\bar{b}_k = \frac{1}{N} \operatorname{tr} \mathbf{P}_k \Big(b_k \mathbf{P}_k + \left[\bar{c}_k - b_k \bar{b}_k \right] \mathbf{I}_{n_k} \Big)^{-1}$$

has a unique solution, satisfying $0 \leq \bar{b}_k < c_k \bar{c}_k / b_k$. Thus, $\bar{b}_k$ is uniquely determined by b_k. Consider now the following functions for $k \in \{1, \ldots, K\}$ and $x > 0$:

$$h_k(x_1, \ldots, x_K) \mapsto \frac{1}{N} \sum_{j=1}^{N_k} \frac{\zeta_{kj}}{1 + \bar{b}_k \zeta_{kj}}$$

where $\bar{b}_k \in [0, c_k \bar{c}_k / x_k)$ and $\zeta_{kj} \geq 0$ are the unique solutions to the following fixed-point equations:

$$\bar{b}_k = \frac{1}{N} \operatorname{tr} \mathbf{P}_k \Big(x_k \mathbf{P}_k + \left[\bar{c}_k - x_k \bar{b}_k \right] \mathbf{I}_{n_k} \Big)^{-1} \tag{2.214}$$

$$\zeta_{kj} = \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \left(\frac{1}{N} \sum_{i=1}^K \sum_{l=1}^{N_k} \frac{\bar{b}_i \mathbf{R}_{i,l}}{1 + \bar{b}_i \zeta_{il}} + x \mathbf{I}_N \right)^{-1}. \tag{2.215}$$

Similar to the proof of Theorem 15, it is now sufficient to prove that the K-variate function $\mathbf{h}:(x_1,\dots,x_K)\mapsto(h_1,\dots,h_K)$ is a standard interference function (see Definition 11) and to apply Theorem 16 to conclude on the existence and uniqueness of a solution to $x_k = h_k(x_1,\dots,x_K)$ for all k. The associated fixed-point algorithm follows the recursive equations

$$x_k^{(t+1)} = h_k(x_1^{(t)},\dots,x_K^{(t)}), \qquad k=1,\dots,K$$

for $t\geq 0$ and for any set of initial values $x_1^{(0)},\dots,x_K^{(0)}>0$, which then converge, as $t\to\infty$, to the fixed-point.

Showing positivity is straightforward: For $x>0$, we have $\zeta_{kj}>0$ by Theorem 14 and $\bar{b}_k\geq 0$ by its definition. Thus, $h_k(x_1,\dots,x_K)>0$ for all $x_1,\dots,x_K>0$.

To prove monotonicity of $h_k(x_1,\dots,x_K)$, we first recall the following result from (2.84). Let $x_k > x_k'$, and consider $\bar{b}_k$ and $\bar{b}_k'$ the corresponding solutions to (2.214). Then,

$$\text{(i)}\;\; \bar{b}_k < \bar{b}_k' \qquad\qquad \text{(ii)}\;\; x_k\bar{b}_k > x_k'\bar{b}_k'. \tag{2.216}$$

We now prove a further result.

Lemma 18. *Let $x>0$ and assume $\bar{b}_k > \bar{b}_k'$. Consider ζ_{kj} and ζ_{kj}' as the unique solutions to* (2.215) *for $\bar{b}_k$ and $\bar{b}_k'$, respectively. Then,*

$$\textit{(i)}\;\; \zeta_{kj}\leq\zeta_{kj}' \qquad\qquad \textit{(ii)}\;\; \bar{b}_k\zeta_{kj} > \bar{b}_k'\zeta_{kj}'. \tag{2.217}$$

Proof. The proof is based on the consideration of an extended version of the random matrix model assumed in Theorem 14. Let us consider the following random matrices $\mathbf{H}_k^L\in\mathbb{C}^{LN\times LN_k}$, given as

$$\mathbf{H}_k^L = \frac{1}{\sqrt{LN}}\left[\left(\mathbf{R}_{k1}^L\right)^{\frac{1}{2}}\mathbf{Z}_{k1}^L,\dots,\left(\mathbf{R}_{kN_k}^L\right)^{\frac{1}{2}}\mathbf{Z}_{kN_k}^L\right] \tag{2.218}$$

where $\mathbf{R}_{kj}^L = \operatorname{diag}(\mathbf{R}_{kj},\dots,\mathbf{R}_{kj})\in\mathbb{C}^{LN\times LN}$ are block-diagonal matrices consisting of L copies of the matrix $\mathbf{R}_{kj}$ and $\mathbf{Z}_{kj}^L\in\mathbb{C}^{LN\times L}$ are random matrices composed of i.i.d. entries with zero mean, unit variance and finite 8th order moment. We define the following matrices which will be of repeated use:

$$\tilde{\mathbf{B}}^L = \sum_{k=1}^K \bar{b}_k\mathbf{H}_k^L\left(\mathbf{H}_k^L\right)^{\mathsf{H}}, \qquad \tilde{\mathbf{B}}'^L = \bar{b}_k'\mathbf{H}_k^L\left(\mathbf{H}_k^L\right)^{\mathsf{H}} + \sum_{l=1,l\neq k}^K \bar{b}_l\mathbf{H}_l^L\left(\mathbf{H}_l^L\right)^{\mathsf{H}}$$
$$\mathbf{Q} = \left(\tilde{\mathbf{B}}^L + x\mathbf{I}_{NL}\right)^{-1}, \qquad \mathbf{Q}' = \left(\tilde{\mathbf{B}}'^L + x\mathbf{I}_{NL}\right)^{-1}.$$

One can verify from Theorem 14 that for any fixed $N,N_1,\dots,N_K$, the following limit holds:

$$\frac{1}{LN}\operatorname{tr}\mathbf{R}_{kj}^L\left(\tilde{\mathbf{B}}^L + x\mathbf{I}_{NK}\right)^{-1} \xrightarrow[L\to\infty]{\text{a.s}} \zeta_{kj}. \tag{2.219}$$

Thus, any properties of the random quantities on the left-hand side of the previous equation also hold for the deterministic quantities ζ_{kj}. We will exploit this fact for the termination of the proof. The matrices $\tilde{\mathbf{B}}^L$ and $\tilde{\mathbf{B}}'^L$ differ only by $\bar{b}_k$. This assumption will be sufficient for the proof since the case $\bar{b}_l > \bar{b}'_l$ for $l \in \{1, \ldots, K\}$ follows by simple iteration of the case $\bar{b}_l = \bar{b}'_l$ for $l \neq k$ and $\bar{b}_k > \bar{b}'_k$.

To prove (i), it is now sufficient to show that, for any L,

$$\frac{1}{N} \operatorname{tr} \mathbf{R}_{k,j}^L \left(\mathbf{Q} - \mathbf{Q}'\right) < 0. \tag{2.220}$$

By Lemma 17, this is equivalent to proving $\left(\mathbf{Q}\right)^{-1} - \left(\mathbf{Q}'\right)^{-1} \succ 0$, which is straightforward since

$$\left(\mathbf{Q}\right)^{-1} - \left(\mathbf{Q}'\right)^{-1} = \tilde{\mathbf{B}}^L - \tilde{\mathbf{B}}'^L = (\bar{b}_k - \bar{b}'_k)\mathbf{H}_k^L \left(\mathbf{H}_k^L\right)^{\mathsf{H}} \succ 0. \tag{2.221}$$

Thus,

$$\frac{1}{NL} \operatorname{tr} \mathbf{R}_{k,j}^L \left(\mathbf{Q} - \mathbf{Q}'\right) \xrightarrow[L\to\infty]{\text{a.s}} \zeta_{kj} - \zeta'_{kj} \leq 0 \tag{2.222}$$

since ζ_{kj} and ζ'_{kj} do not depend on L.

For (ii), we need to show that

$$\bar{b}_k \frac{1}{LN} \operatorname{tr} \mathbf{R}_{kj}^L \mathbf{Q} - \bar{b}'_k \frac{1}{LN} \operatorname{tr} \mathbf{R}_{kj}^L \mathbf{Q}' > 0. \tag{2.223}$$

Similarly to the previous part of the proof, it is sufficient to show that $\left(\bar{b}_k \mathbf{Q}\right)^{-1} - \left(\bar{b}'_k \mathbf{Q}'\right)^{-1} \prec 0$. Hence,

$$\begin{aligned}
\left(\bar{b}_k \mathbf{Q}\right)^{-1} - \left(\bar{b}'_k \mathbf{Q}'\right)^{-1} &= \frac{1}{\bar{b}_k} \left(\tilde{\mathbf{B}}^L + x\mathbf{I}_{NL}\right) - \frac{1}{\bar{b}'_k} \left(\tilde{\mathbf{B}}'^L + x\mathbf{I}_{NL}\right) \\
&= x\left(\frac{1}{\bar{b}_k} - \frac{1}{\bar{b}'_k}\right)\mathbf{I}_{NL} + \left(\frac{1}{\bar{b}_k} - \frac{1}{\bar{b}'_k}\right) \sum_{l=1, l\neq k}^{K} \bar{b}_l \mathbf{H}_l^L \left(\mathbf{H}_l^L\right)^{\mathsf{H}} \\
&\prec 0
\end{aligned} \tag{2.224}$$

since $x > 0$, $\bar{b}_k > \bar{b}'_k$ and $\bar{b}_l \geq 0$ for all l. □

Consider now $(x_1, \ldots, x_K)$ and $(x'_1, \ldots, x'_K)$, such that $x_k > x'_k\ \forall k$, and denote by $(\bar{b}_1, \ldots, \bar{b}_K)$ and $(\bar{b}'_1, \ldots, \bar{b}'_K)$ the corresponding solutions to (2.214). Denote by ζ_{kj} and ζ'_{kj} the unique solutions to (2.215) for $(\bar{b}_1, \ldots, \bar{b}_K)$ and $(\bar{b}'_1, \ldots, \bar{b}'_K)$, respectively. It follows from (2.216), that $\bar{b}_k < \bar{b}'_k\ \forall k$. Equation (2.217) now implies that $\zeta_{kj} \geq \zeta'_{kj}$ and $\bar{b}_k\zeta_{kj} < \bar{b}'_k\zeta'_{kj}$. Combining these results yields

$$h_k(x_1, \ldots, x_K) = \frac{1}{N} \sum_{j=1}^{N_k} \frac{\zeta_{kj}}{1 + \bar{b}_k \zeta_{kj}} > \frac{1}{N} \sum_{j=1}^{N_k} \frac{\zeta'_{kj}}{1 + \bar{b}'_k \zeta'_{kj}} = h_k(x'_1, \ldots, x'_K) \tag{2.225}$$

which proves monotonicity.

To prove scalability, let $\alpha > 1$, and consider the following difference:

$$\begin{aligned}
&\alpha h_k(x_1,\ldots,x_K) - h_k(\alpha x_1,\ldots,\alpha x_K) \\
&= \frac{1}{N}\sum_{j=1}^{N_k} \frac{\alpha\zeta_{kj}}{1+\bar{b}_k\zeta_{kj}} - \frac{\zeta_{kj}^{(\alpha)}}{1+\bar{b}_k^{(\alpha)}\zeta_{kj}^{(\alpha)}} \\
&= \frac{1}{N}\sum_{i=1}^{N_k} \frac{\left[\alpha\zeta_{kj} - \zeta_{kj}^{(\alpha)}\right] + \zeta_{kj}\zeta_{kj}^{(\alpha)}\left[\alpha\bar{b}_k^{(\alpha)} - \bar{b}_k\right]}{\left[1+\bar{b}_k\zeta_{kj}\right]\left[1+\bar{b}_k^{(\alpha)}\zeta_{kj}^{(\alpha)}\right]} \qquad (2.226)
\end{aligned}$$

where we have denoted by $\bar{b}_k^{(\alpha)}$ the solution to (2.214) with x_k replaced by αx_k and by $\zeta_{kj}^{(\alpha)}$ the solution to (2.215) for $\bar{b}_k^{(\alpha)}$. We have from (2.216)-(i) that $\bar{b}_k^{(\alpha)} < \bar{b}_k$ and from (2.216)-(ii) that

$$\alpha x_k \bar{b}_k^{(\alpha)} > x_k\bar{b}_k \Longleftrightarrow \alpha\bar{b}_k^{(\alpha)} - \bar{b}_k > 0. \qquad (2.227)$$

It remains now to show that also $\alpha\zeta_{kj} - \zeta_{kj}^{(\alpha)} > 0$. To this end, consider the following difference:

$$\alpha\zeta_{kj} - \zeta_{kj}^{(\alpha)} = \frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\alpha\mathbf{T} - \mathbf{T}^{(\alpha)}\right) \qquad (2.228)$$

where

$$\mathbf{T} = \left(\frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{N_k}\frac{\bar{b}_k\mathbf{R}_{k,j}}{1+\bar{b}_k\zeta_{kj}} + x\mathbf{I}_N\right)^{-1} \qquad (2.229)$$

$$\mathbf{T}^{(\alpha)} = \left(\frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{N_k}\frac{\bar{b}_k^{(\alpha)}\mathbf{R}_{k,j}}{1+\bar{b}_k^{(\alpha)}\zeta_{kj}^{(\alpha)}} + x\mathbf{I}_N\right)^{-1}. \qquad (2.230)$$

By Lemma 17, it is now sufficient to show that $\left(\mathbf{T}^{(\alpha)}(z)\right)^{-1} \succ (\alpha\mathbf{T}(z))^{-1}$. Write therefore

$$\begin{aligned}
&\left(\mathbf{T}^{(\alpha)}\right)^{-1} - (\alpha\mathbf{T})^{-1} \qquad (2.231)\\
&= x\left(1-\frac{1}{\alpha}\right)\mathbf{I}_N + \frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{N_k}\frac{\left[\alpha\bar{b}_k^{(\alpha)} - \bar{b}_k\right] + \bar{b}_k^{(\alpha)}\bar{b}_k\left[\alpha\zeta_{kj} - \zeta_{kj}^{(\alpha)}\right]}{\alpha\left[1+\bar{b}_k\zeta_{kj}\right]\left[1+\bar{b}_k^{(\alpha)}\zeta_{kj}^{(\alpha)}\right]}\mathbf{R}_{kj}. \qquad (2.232)
\end{aligned}$$

The first summand is positive definite since $x > 0$ and $\alpha > 1$. All other terms are also positive definite since $\alpha\bar{b}_k^{(\alpha)} - \bar{b}_k > 0$ from (2.227) and $\alpha\bar{b}_k^{(\alpha)}\bar{b}_k\zeta_{kj} > \bar{b}_k\bar{b}_k^{(\alpha)}\zeta_{kj}^{(\alpha)}$, since $\alpha\bar{b}_k^{(\alpha)} > \bar{b}_k$ and $\bar{b}_k\zeta_{kj} > \bar{b}_k^{(\alpha)}\zeta_{kj}^{(\alpha)}$ by (2.217)-(ii) and (2.216)-(i). Since the sum of positive definite matrices is also positive definite, we have $\alpha\zeta_{kj} - \zeta_{kj}^{(\alpha)} > 0$. This terminates the proof of scalability.

Thus, we have shown $\mathbf{h} : (x_1,\ldots,x_K) \mapsto (h_1,\ldots,h_K)$ to be an interference standard function. Moreover, from the series convergence in Theorem 15 and Theorem 14

$$\bar{b}_k = \lim_{t\to\infty}\bar{b}_k^{(t)}, \qquad \zeta_{kj} = \lim_{t\to\infty}\zeta_{kj}^{(t)} \qquad (2.233)$$

where

$$\bar{b}_k^{(t)} = \frac{1}{N}\operatorname{tr}\mathbf{P}_k\Big(x_k\mathbf{P}_k + \Big[\bar{c}_k - x_k\bar{b}_k^{(t-1)}\Big]\mathbf{I}_{n_k}\Big)^{-1} \tag{2.234}$$

$$\zeta_{kj}^{(t)} = \frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\frac{1}{N}\sum_{i=1}^{K}\sum_{l=1}^{N_k}\frac{\bar{b}_i\mathbf{R}_{i,l}}{1+\bar{b}_i\zeta_{il}^{(t-1)}} + x\mathbf{I}_N\right)^{-1} \tag{2.235}$$

and $\bar{b}_k^{(0)}$ can take any value in $[0, c_k\bar{c}_k/x_k)$ and $\zeta_{kj}^{(0)} = 1/x$ for all k, j.

Part *(ii)*

We begin by proving the following result:

$$\max_k |\bar{e}_k - \bar{b}_k| \xrightarrow{\text{a.s.}} 0 \tag{2.236}$$

$$\max_k |e_k - b_k| \xrightarrow{\text{a.s.}} 0 \tag{2.237}$$

where $\bar{e}_k$, e_k are defined in Theorem 15 and $\bar{b}_k$, b_k are defined in Theorem 23. From standard lemmas of matrix analysis, we have

$$\begin{aligned}
e_k &= \frac{1}{N}\operatorname{tr}\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}\left(\sum_{i=1}^{K}\bar{e}_i\mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \\
&= \frac{1}{N}\sum_{j=1}^{N_k}\mathbf{h}_{kj}^{\mathsf{H}}\left(\sum_{i=1}^{K}\bar{e}_i\mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}\mathbf{h}_{kj} \\
&= \frac{1}{N}\sum_{j=1}^{N_k}\frac{\mathbf{h}_{kj}^{\mathsf{H}}\left(\sum_{i=1}^{K}\bar{e}_i\mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} - \bar{e}_k\mathbf{h}_{kj}\mathbf{h}_{kj}^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}\mathbf{h}_{kj}}{1+\bar{e}_k\mathbf{h}_{kj}^{\mathsf{H}}\left(\sum_{i=1}^{K}\bar{e}_i\mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} - \bar{e}_k\mathbf{h}_{kj}\mathbf{h}_{kj}^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}\mathbf{h}_{kj}}
\end{aligned}$$

where the last step follows from Lemma 6. If $\bar{e}_i$ were not dependent on $\mathbf{h}_{kj}$, we could now simply proceed by applying Lemma 13 to the individual quadratic forms, i.e.:

$$\begin{aligned}
&\mathbf{h}_{kj}^{\mathsf{H}}\left(\sum_{i=1}^{K}\bar{e}_i\mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} - \bar{e}_k\mathbf{h}_{kj}\mathbf{h}_{kj}^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}\mathbf{h}_{kj} \\
&\asymp \frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\sum_{i=1}^{K}\bar{e}_i\mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} - \bar{e}_k\mathbf{h}_{kj}\mathbf{h}_{kj}^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1}.
\end{aligned} \tag{2.238}$$

However, in order to show that this step is correct, in a similar manner as in the proof of Theorem 22, we need the following intermediate arguments. Define $\bar{e}_{i,kj}$ and $e_{i,kj}$ as the unique solutions to the following fixed-point equations:

$$e_{i,kj} = \frac{1}{N}\operatorname{tr}\mathbf{H}_{i,kj}\mathbf{H}_{i,kj}^{\mathsf{H}}\left(\sum_{l=1}^{K}\bar{e}_{l,kj}\mathbf{H}_{l,kj}\mathbf{H}_{l,kj}^{\mathsf{H}} + x\mathbf{I}_N\right)^{-1} \tag{2.239}$$

$$\bar{e}_{i,kj} = \frac{1}{N}\operatorname{tr}\mathbf{P}_i\left(e_{i,kj}\mathbf{P}_i + [\bar{c}_k - e_{i,kj}\bar{e}_{i,kj}\mathbf{I}_{n_i}]\right)^{-1} \tag{2.240}$$

for $i \in \{1, \ldots, K\}$, where

$$\mathbf{H}_{i,kj} = \begin{cases} \mathbf{H}_i, & k \neq i \\ (\mathbf{h}_{k1} \cdots \mathbf{h}_{kj-1} \mathbf{h}_{kj+1} \cdots \mathbf{h}_{kN_i}), & k = i \end{cases}. \tag{2.241}$$

Thus, $\bar{e}_{i,kj}$ and $e_{i,kj}$ are independent of $\mathbf{h}_{kj}$. Following similar steps as in the proof of Theorem 22 (starting from (2.176)), one can show that for $i \in \{1, \ldots, K\}$ and all k, j,

$$e_{i,kj} - e_i \to 0, \quad \bar{e}_{i,kj} - \bar{e}_i \to 0. \tag{2.242}$$

Thus, we have

$$\begin{aligned}
&\frac{1}{N} \sum_{j=1}^{N_k} \frac{\mathbf{h}_{kj}^{\mathsf{H}} \left(\sum_{i=1}^K \bar{e}_i \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} - \bar{e}_k \mathbf{h}_{kj} \mathbf{h}_{kj}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \mathbf{h}_{kj}}{1 + \bar{e}_k \mathbf{h}_{kj}^{\mathsf{H}} \left(\sum_{i=1}^K \bar{e}_i \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} - \bar{e}_k \mathbf{h}_{kj} \mathbf{h}_{kj}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \mathbf{h}_{kj}} \\
&\underset{\text{(a)}}{\asymp} \frac{1}{N} \sum_{j=1}^{N_k} \frac{\mathbf{h}_{kj}^{\mathsf{H}} \left(\sum_{i=1}^K \bar{e}_{i,kj} \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} - \bar{e}_{k,kj} \mathbf{h}_{kj} \mathbf{h}_{kj}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \mathbf{h}_{kj}}{1 + \bar{e}_k \mathbf{h}_{kj}^{\mathsf{H}} \left(\sum_{i=1}^K \bar{e}_{i,kj} \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} - \bar{e}_{k,kj} \mathbf{h}_{kj} \mathbf{h}_{kj}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} \mathbf{h}_{kj}} \\
&\underset{\text{(b)}}{\asymp} \frac{1}{N} \sum_{j=1}^{N_k} \frac{\frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \left(\sum_{i=1}^K \bar{e}_{i,kj} \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} - \bar{e}_{k,kj} \mathbf{h}_{kj} \mathbf{h}_{kj}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1}}{1 + \bar{e}_k \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \left(\sum_{i=1}^K \bar{e}_{i,kj} \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} - \bar{e}_{k,kj} \mathbf{h}_{kj} \mathbf{h}_{kj}^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1}} \\
&\underset{\text{(c)}}{\asymp} \frac{1}{N} \sum_{j=1}^{N_k} \frac{\frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \left(\sum_{i=1}^K \bar{e}_i \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1}}{1 + \bar{e}_k \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \left(\sum_{i=1}^K \bar{e}_i \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1}} \\
&\underset{\text{(d)}}{\asymp} \frac{1}{N} \sum_{j=1}^{N_k} \frac{\frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \bar{\mathbf{T}}}{1 + \bar{e}_k \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \bar{\mathbf{T}}}
\end{aligned} \tag{2.243}$$

where (a) follows from (2.242), (b) follows from Lemma 13 and Lemma 16, (c) is again due to (2.242) and Lemma 8, and (d) follows from an application of Theorem 14, where we have defined

$$\bar{\mathbf{T}} = \left(\frac{1}{N} \sum_{k=1}^K \sum_{j=1}^{N_k} \frac{\bar{e}_k \mathbf{R}_{kj}}{1 + \bar{e}_k \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \bar{\mathbf{T}}} + x \mathbf{I}_N \right)^{-1}. \tag{2.244}$$

Note again that Theorem 14 cannot be directly applied here since the quantities $\bar{e}_i$ depend on the matrices $\mathbf{H}_i$. However, it is immediate to show that the result extends to this case, by replacing $\bar{e}_i$ by $\bar{e}_{i,kj}$ at each necessary step of the proof.

Hence, we can write

$$e_k = \frac{1}{N} \operatorname{tr} \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}} \left(\sum_{i=1}^K \bar{e}_i \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} + x \mathbf{I}_N \right)^{-1} = \frac{1}{N} \sum_{j=1}^{N_k} \frac{\frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \bar{\mathbf{T}}}{1 + \bar{e}_k \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \bar{\mathbf{T}}} + \epsilon_{N,k} \tag{2.245}$$

for some sequences of reals $\epsilon_{N,k}$, satisfying $\epsilon_{N,k} \to 0$.

Recall now the following definitions for $k = 1, \dots, K$:

$$e_k = \frac{1}{N} \sum_{j=1}^{N_k} \frac{\frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \bar{\mathbf{T}}}{1 + \bar{e}_k \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \bar{\mathbf{T}}} + \epsilon_{N,k} \tag{2.246}$$

$$b_k = \frac{1}{N} \sum_{j=1}^{N_k} \frac{\frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \mathbf{T}}{1 + \bar{b}_k \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \mathbf{T}} \tag{2.247}$$

$$\bar{e}_k = \frac{1}{N} \sum_{j=1}^{n_k} \frac{p_{kj}}{\bar{c}_k - e_k \bar{e}_k + e_k p_{kj}}, \qquad 0 \le \bar{e}_k < c_k \bar{c}_k / e_k \tag{2.248}$$

$$\bar{b}_k = \frac{1}{N} \sum_{j=1}^{n_k} \frac{p_{kj}}{\bar{c}_k - b_k \bar{b}_k + b_k p_{kj}}, \qquad 0 \le \bar{b}_k < c_k \bar{c}_k / b_k \tag{2.249}$$

where

$$\bar{\mathbf{T}} = \left(\frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{N_k} \frac{\bar{e}_k \mathbf{R}_{kj}}{1 + \bar{f}_{N,k} \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \bar{\mathbf{T}}} + x \mathbf{I}_N \right)^{-1} \tag{2.250}$$

$$\mathbf{T} = \left(\frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{N_k} \frac{\bar{b}_k \mathbf{R}_{kj}}{1 + \bar{b}_k \frac{1}{N} \operatorname{tr} \mathbf{R}_{kj} \mathbf{T}} + x \mathbf{I}_N \right)^{-1}. \tag{2.251}$$

Denote $P = \max_k\{\limsup \|\mathbf{P}_k\|\}$, $R = \max_m\{\limsup \|\tilde{\mathbf{R}}_m\|\}$, $c_+ = \max_k\{\limsup c_k\}$ and $\bar{c}_- = \min_k\{\liminf \bar{c}_k\}$, $\bar{c}_+ = \max_k\{\limsup \bar{c}_k\}$. Since we are interested in the asymptotic limit $N \to \infty$, we assume from the beginning that N is sufficiently large, so that the following inequalities hold for all k:

$$c_k \le c_+, \quad \bar{c}_- \le \bar{c}_k \le \bar{c}_+, \quad \|\mathbf{P}_k\| \le P, \quad \|\mathbf{R}_{kj}\| \le R. \tag{2.252}$$

We then have the following properties:

$$\bar{e}_k \le \frac{P}{(1 - c_+)\bar{c}_-}, \quad \bar{b}_k \le \frac{P}{(1 - c_+)\bar{c}_-}, \quad b_k \bar{b}_k < c_+ \bar{c}_+, \quad e_k \bar{e}_k < c_+ \bar{c}_+. \tag{2.253}$$

For notational simplicity, we define the following quantities:

$$\xi = \max_k |e_k - b_k|, \qquad \bar{\xi} = \max_k |\bar{e}_k - \bar{b}_k|. \tag{2.254}$$

We will show in the sequel that $\xi \xrightarrow{\text{a.s.}} 0$ and $\bar{\xi} \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$.

Consider first the following difference:

$$\begin{aligned}
&\sup_{k,j}\left|\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\mathbf{T}-\bar{\mathbf{T}}\right)\right| \\
=&\sup_{k,j}\left|\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\mathbf{T}\left(\frac{1}{N}\sum_{l=1}^{K}\sum_{m=1}^{N_l}\frac{\bar{e}_l\mathbf{R}_{lm}}{1+\bar{e}_l\frac{1}{N}\operatorname{tr}\mathbf{R}_{lm}\bar{\mathbf{T}}}-\frac{\bar{b}_l\mathbf{R}_{lm}}{1+\bar{b}_l\frac{1}{N}\operatorname{tr}\mathbf{R}_{lm}\bar{\mathbf{T}}}\right)\bar{\mathbf{T}}\right| \\
=&\sup_{k,j}\left|\frac{1}{N}\sum_{l=1}^{K}\sum_{m=1}^{N_l}\frac{\bar{e}_l-\bar{b}_l+\bar{e}_l\bar{b}_l\left(\frac{1}{N}\operatorname{tr}\mathbf{R}_{lm}\mathbf{T}-\frac{1}{N}\operatorname{tr}\mathbf{R}_{lm}\bar{\mathbf{T}}\right)}{\left(1+\bar{e}_l\frac{1}{N}\operatorname{tr}\mathbf{R}_{lm}\bar{\mathbf{T}}\right)\left(1+\bar{b}_l\frac{1}{N}\operatorname{tr}\mathbf{R}_{lm}\bar{\mathbf{T}}\right)}\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\bar{\mathbf{T}}\mathbf{R}_{lm}\mathbf{T}\right| \\
\le&\frac{R^2}{x^2}K\max_k\bar{c}_k\left[\max_k|\bar{e}_k-\bar{b}_k|+\max_k|\bar{e}_k\bar{b}_k|\sup_{k,j}\left|\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\mathbf{T}-\bar{\mathbf{T}}\right)\right|\right] \\
\le&\frac{R^2}{x^2}K\bar{c}_+\left[\bar{\xi}+\frac{P^2}{(1-c_+)^2\bar{c}_-^2}\sup_{k,j}\left|\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\mathbf{T}-\bar{\mathbf{T}}\right)\right|\right] \qquad (2.255)
\end{aligned}$$

where the first equality follows from Lemma 5. Rearranging the terms yields:

$$\sup_{k,j}\left|\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\mathbf{T}-\bar{\mathbf{T}}\right)\right|\le\frac{P^2K\bar{c}_+}{x^2-\frac{R^2P^2}{(1-c_+)^2\bar{c}_-^2}}\;\bar{\xi} \qquad (2.256)$$

for $x>\frac{RP}{(1-c_+)\bar{c}_-}$.

Consider now the term $\xi=\max_k|e_k-b_k|$:

$$\begin{aligned}
\xi=&\max_k\left|\frac{1}{N}\sum_{j=1}^{N_k}\frac{\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\bar{\mathbf{T}}-\mathbf{T}\right)+(\bar{b}_k-\bar{e}_k)\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\bar{\mathbf{T}}}{\left(1+\bar{e}_k\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\bar{\mathbf{T}}\right)\left(1+\bar{b}_k\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\mathbf{T}\right)}+\epsilon_{N,k}\right| \\
\le&\bar{c}_+\sup_{kj}\left|\frac{1}{N}\operatorname{tr}\mathbf{R}_{kj}\left(\mathbf{T}-\bar{\mathbf{T}}\right)\right|+\bar{c}_+\frac{R^2}{x^2}\max_k|\bar{e}_k-\bar{b}_k|+\max_k|\epsilon_{N,k}| \\
\le&\frac{P^2K\bar{c}_+^2}{x^2-\frac{R^2P^2}{(1-c_+)^2\bar{c}_-^2}}\bar{\xi}+\frac{\bar{c}_+R^2}{x^2}\bar{\xi}+\max_k|\epsilon_{N,k}| \\
=&\left[\frac{P^2K\bar{c}_+^2}{x^2-\frac{R^2P^2}{(1-c_+)^2\bar{c}_-^2}}+\frac{\bar{c}_+R^2}{x^2}\right]\bar{\xi}+\max_k|\epsilon_{N,k}| \qquad (2.257)
\end{aligned}$$

where the last inequality follows from (2.256). Similarly, we have for $\bar{\xi}=\max_k|\bar{e}_k-\bar{b}_k|$:

$$\begin{aligned}
\bar{\xi}=&\max_k\left|\frac{1}{N}\sum_{j=1}^{n_k}p_{kj}\frac{e_k\bar{e}_k-b_k\bar{b}_k+p_{kj}(b_k-e_k)}{(\bar{c}_k-e_k\bar{e}_k+e_kp_{kj})(\bar{c}_k-b_k\bar{b}_k+b_kp_{kj})}\right| \\
\le&\frac{1}{N}\sum_{j=1}^{n_k}\frac{p_{kj}^2\max_k|e_k-b_k|}{(1-c_+)^2\bar{c}_-^2}+p_{kj}\frac{\max_k\left[\bar{e}_k|e_k-b_k|\right]|+\max_k\left[b_k|\bar{e}_k-\bar{b}_k|\right]}{(1-c_+)^2\bar{c}_-^2} \\
\le&\frac{P^2}{(1-c_+)^2\bar{c}_-^2}\left(1+\frac{1}{(1-c_+)\bar{c}_-}\right)\xi+\frac{PR\bar{c}_+}{x(1-c_+)^2\bar{c}_-^2}\bar{\xi}. \qquad (2.258)
\end{aligned}$$

Thus, for $x \geq \max\left\{\frac{2PR\bar{c}_+}{(1-c_+)^2\bar{c}_-^2}, \frac{RP}{(1-c_+)\bar{c}_-}\right\}$, we have

$$\bar{\xi} \leq \frac{2P^2}{(1-c_+)^2\bar{c}_-^2}\left(1 + \frac{1}{(1-c_+)\bar{c}_-}\right)\xi. \tag{2.259}$$

Replacing (2.259) in (2.257) leads to

$$\xi \leq \left[\frac{P^2 K\bar{c}_+^2}{x^2 - \frac{R^2P^2}{(1-c_+)^2\bar{c}_-^2}} + \frac{\bar{c}_+R^2}{x^2}\right]\frac{2P^2}{(1-c_+)^2\bar{c}_-^2}\left(1 + \frac{1}{(1-c_+)\bar{c}_-}\right)\xi + \max_k |\epsilon_{N,k}|. \tag{2.260}$$

For x sufficiently large, we therefore have

$$0 \leq \xi \leq C\epsilon_{N,k} \to 0$$

for some $C > 0$. This implies that $\xi \xrightarrow{\text{a.s.}} 0$ and, by (2.259), that $\bar{\xi} \xrightarrow{\text{a.s.}} 0$. Since $b_k, \bar{e}_k, \bar{b}_k$ have analytic extensions in a neighborhood of $\mathbb{R}^+$ on which they are (almost surely) bounded, we have from the Vitali convergence theorem that the almost sure convergence holds true for all $x > 0$.

Convergence of the mutual information

Consider now the first term of $\bar{I}_N(x)$ in Theorem 18. Due to the convergence of $\bar{e}_k - \bar{b}_k \xrightarrow{\text{a.s.}} 0$ and the almost sure boundedness of the $\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}$ matrices, $\|\sum_{k=1}^K(\bar{e}_k - \bar{b}_k)\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}\| \xrightarrow{\text{a.s.}} 0$, and we have immediately that

$$\frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^K \bar{e}_k\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}\right) - \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^K \bar{b}_k\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}\right) \xrightarrow{\text{a.s.}} 0. \tag{2.261}$$

Applying Theorem 14 (iv) to the second term yields

$$\frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^K \bar{b}_k\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}\right) - \bar{V}_N(x) \xrightarrow{\text{a.s.}} 0. \tag{2.262}$$

Consider now $\bar{I}_N^{(a)}(x) = \bar{I}_N(x)$ as defined in Theorem 18 and $\bar{I}_N(x)$ as defined in Theorem 23 (ii). It follows from (2.236), (2.237) and (2.262), that

$$\bar{I}_N^{(a)}(x) - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0. \tag{2.263}$$

This implies also that

$$I_N(x) - \bar{I}_N(x) \xrightarrow{\text{a.s.}} 0. \tag{2.264}$$

To prove convergence in the mean, we will use the same arguments as in [74]. Denote

$$m_N(z) = \frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{2.265}$$

$$\bar{m}_N(z) = \frac{1}{N}\operatorname{tr}\left(\frac{1}{N}\sum_{k=1}^K\sum_{j=1}^{N_k}\frac{\bar{b}_k(-z)\mathbf{R}_{k,j}}{1 + b_k(-z)\zeta_{kj}(-z)} - z\mathbf{I}_N\right)^{-1} \tag{2.266}$$

where $m_N(z)$ is the Stieltjes transform of $\mathbf{B}_N$. It is easy to see that

$$\mathbb{E}\left[I_N(x)\right] - \bar{I}_N(x) = \int_x^\infty \left(\left[\frac{1}{\omega} - \mathbb{E}\left[m_N(-\omega)\right]\right] - \left[\frac{1}{\omega} - \bar{m}_N(-\omega)\right]\right) d\omega. \tag{2.267}$$

We now apply the argument from [74, pp. 923] which shows that

$$\begin{aligned}&\left|\int_x^\infty \left(\left[\frac{1}{\omega} - \mathbb{E}\left[m_N(-\omega)\right]\right] - \left[\frac{1}{\omega} - \bar{m}_N(-\omega)\right]\right) d\omega\right| \\ &\leq \int_x^\infty \frac{1}{\omega^2}\left(\left|\mathbb{E}\left[\int_0^\infty t dF_N(t)\right]\right| + \left|\frac{1}{N}\operatorname{tr}\left(\frac{1}{N}\sum_{k=1}^K\sum_{j=1}^{N_k}\frac{\bar{b}_k(\omega)\mathbf{R}_{k,j}}{1+b_k(\omega)\zeta_{kj}(\omega)}\right)\right|\right) d\omega\end{aligned} \tag{2.268}$$

the right-hand side of which exists for all N and is uniformly bounded by $\frac{2}{x}(KPR)$. Since $m_N(-\omega) - \bar{m}_N(-\omega) \xrightarrow{\text{a.s.}} 0$ (as a consequence of the convergence $\bar{e}_k - \bar{b}_k \xrightarrow{\text{a.s.}} 0$), the boundedness of $m_N(-\omega)$ then ensures (by dominated convergence) that $\mathbb{E}\left[m_N(-\omega)\right] - \bar{m}_N(-\omega) \to 0$. Since the integrand tends to zero and is summable independently of N, the dominated convergence theorem now ensures that

$$\mathbb{E}\left[I_N(x)\right] - \bar{I}_N(x) \to 0. \tag{2.269}$$

2.5.10 Proof of Theorem 24

We need several auxiliary results in the course of the proof.

Corollary 1 (Special case of Theorem 13, see also [65]). *Let $\mathbf{R}_N \in \mathbb{C}^{n\times n}$ be Hermitian nonnegative definite, satisfying $\limsup_N \|\mathbf{R}_N\| < \infty$, and let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have i.i.d. complex Gaussian entries with zero mean and variance $1/n$. For $x > 0$, define the following functions $m_N(x) = \frac{1}{N}\operatorname{tr}\left(\mathbf{X}_N\mathbf{R}_N\mathbf{X}_N^{\mathsf{H}} + \frac{1}{x}\mathbf{I}_N\right)^{-1}$ and $J_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + x\mathbf{X}_N\mathbf{R}_N\mathbf{X}_N^{\mathsf{H}}\right)$. Denote $c = \frac{n}{N}$ and assume that $N, n \to \infty$ while $0 < \liminf_N c \leq \limsup_N c < \infty$. Then,*

$$(i)\quad m_N(x) - \bar{m}_N(x) \xrightarrow{\text{a.s.}} 0, \qquad (ii)\quad J_N(x) - \bar{J}_N(x) \xrightarrow{\text{a.s.}} 0$$

where

$$\begin{aligned}\bar{m}_N(x) &= \frac{1}{N}\operatorname{tr}\left(\frac{\mathbf{R}_N}{c+\bar{e}_N} + \frac{1}{x}\mathbf{I}_n\right)^{-1} + (1-c)x \\ \bar{J}_N(x) &= \frac{1}{N}\log\det\left(\left[c+\bar{e}_N\right]\mathbf{I}_n + x\mathbf{R}_N\right) + (1-c)\log\left(c+\bar{e}_N\right) - \frac{\bar{e}_N}{c+\bar{e}_N} - \log(c)\end{aligned}$$

and $\bar{e}_N$ is defined as the unique positive solution to the implicit equation

$$\bar{e}_N = \frac{1}{N}\operatorname{tr}\mathbf{R}_N\left(\frac{\mathbf{R}_N}{c+\bar{e}_N} + \frac{1}{x}\mathbf{I}_n\right)^{-1}. \tag{2.270}$$

Lemma 19. *Let the matrices $\mathbf{R}_k(\boldsymbol{\beta}_{k-1})$, be defined as in (2.16). Then, almost surely:*

$$\limsup_n \|\mathbf{R}_k(\boldsymbol{\beta}_{k-1})\| < \infty, \qquad k = 0, \ldots, K.$$

Proof. For $k \in \{1, \dots, K\}$, denote by $(\Omega_k, \mathcal{F}_k, P_k)$ the probability space generating the sequences of random matrices $\mathbf{H}_k$. By Theorem 8, we have on a space $B_k \subset \Omega_k$ with $P_k(B_k) = 1$,

$$\frac{1}{n_{k-1}} \left\| \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}} \right\| - \left(1 + \frac{1}{\sqrt{c_k}} \right)^2 \to 0. \tag{2.271}$$

Obviously, we have $\|\mathbf{R}_0\| = \|\mathbf{I}_n\| = 1$. Thus, almost surely,

$$\begin{aligned} \limsup_n \|\mathbf{R}_1(\beta_0)\| &\leq 1 + \limsup_n \frac{\alpha_1 \beta_0}{n} \left\| \mathbf{H}_1 \mathbf{H}_1^{\mathsf{H}} \right\| \\ &= 1 + \alpha_1 \beta_0 \limsup_n \left(1 + \frac{1}{\sqrt{c_1}} \right)^2 < \infty. \end{aligned} \tag{2.272}$$

Consider now the product space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, Q_2)$. By the Fubini theorem, we have on a subspace $C_2 \subset \Omega_1 \times \Omega_2$ with $Q_2(C_2) = 1$,

$$\begin{aligned} \limsup_n \|\mathbf{R}_2(\boldsymbol{\beta}_1)\| &\leq 1 + \limsup_n \frac{\alpha_2 \beta_1}{n_1} \left\| \mathbf{H}_2 \mathbf{R}_1(\beta_0) \mathbf{H}_2^{\mathsf{H}} \right\| \\ &\leq 1 + \limsup_n \alpha_2 \beta_1 \|\mathbf{R}_1(\beta_0)\| \frac{1}{n_1} \left\| \mathbf{H}_2 \mathbf{H}_2^{\mathsf{H}} \right\| \\ &= 1 + \alpha_2 \beta_1 \limsup_n \left(1 + \alpha_1 \beta_0 \left(1 + \frac{1}{\sqrt{c_1}} \right)^2 \right) \left(1 + \frac{1}{\sqrt{c_2}} \right)^2 \\ &< \infty. \end{aligned} \tag{2.273}$$

Repeating the last step $k-2$ times concludes the proof. □

Lemma 20. *Let $\mathbf{R} \in \mathbb{C}^{N \times N}$ be Hermitian with smallest eigenvalue $\lambda_{min} \geq 1$ and $a, b, c, d > 0$. Then*

$$\frac{1}{N} \operatorname{tr} \mathbf{R} \left(a\mathbf{R} + b\mathbf{I}_N \right)^{-1} \mathbf{R} \left(c\mathbf{R} + d\mathbf{I}_N \right)^{-1} \geq \frac{1}{(a+b)(c+d)}.$$

Proof. Let $\mathbf{R} = \mathbf{U} \boldsymbol{\Delta} \mathbf{U}^{\mathsf{H}}$, where the matrix $\mathbf{U} \in \mathbb{C}^{N \times N}$ is unitary and $\boldsymbol{\Delta} = \operatorname{diag}(\delta_1, \dots, \delta_N) \geq 1$. Thus,

$$\begin{aligned} &\frac{1}{N} \operatorname{tr} \mathbf{R} \left(a\mathbf{R} + b\mathbf{I}_N \right)^{-1} \mathbf{R} \left(c\mathbf{R} + d\mathbf{I}_N \right)^{-1} \\ &= \frac{1}{N} \operatorname{tr} \boldsymbol{\Delta}^2 \left(a\boldsymbol{\Delta} + b\mathbf{I}_N \right)^{-1} \left(c\boldsymbol{\Delta} + d\mathbf{I}_N \right)^{-1} \\ &= \frac{1}{(a+b)(c+d)} \frac{1}{N} \sum_{i=1}^{N} \frac{\delta_i^2 (a+b)(c+d)}{(a\delta_i + b)(c\delta_i + d)} \\ &\geq \frac{1}{(a+b)(c+d)}. \end{aligned} \tag{2.274}$$

□

Part (i)

From standard matrix inequalities and (2.296), it follows that

$$\begin{aligned}&\left|m_k\left(x,\boldsymbol{\beta}_k\right)-m_k\left(x,\bar{\boldsymbol{\beta}}_k\right)\right|\\&\leq x^2\alpha_{k+1}\left\|\frac{\mathbf{H}_{k+1}\mathbf{H}_{k+1}^{\mathsf{H}}}{n_k}\right\|\left\|\beta_k\mathbf{R}_k\left(\boldsymbol{\beta}_{k-1}\right)-\bar{\beta}_k\mathbf{R}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)\right\|\\&\xrightarrow{\text{a.s.}}0.\end{aligned}\tag{2.275}$$

Thus, we can replace from now on β_k by $\bar{\beta}_k$, for almost every $(\mathbf{H}_1,\ldots,\mathbf{H}_K)$.

From Corollary 1, Lemma 19 and the Fubini theorem, it follows that

$$m_k\left(x,\bar{\boldsymbol{\beta}}_k\right)-\tilde{m}_k\left(x,\bar{\boldsymbol{\beta}}_k\right)\xrightarrow{\text{a.s.}}0\tag{2.276}$$

where

$$\tilde{m}_k(x,\bar{\boldsymbol{\beta}}_k)=\frac{1}{n_{k+1}}\operatorname{tr}\left(\frac{\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)}{c_{k+1}+e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)}+\frac{1}{x}\mathbf{I}_n\right)^{-1}+(1-c_{k+1})x\tag{2.277}$$

and $e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)$ is given as the unique positive solution to

$$e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)=\frac{1}{n_{k+1}}\operatorname{tr}\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)\left(\frac{\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)}{c_{k+1}+e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)}+\frac{1}{x}\mathbf{I}_{n_k}\right)^{-1}.\tag{2.278}$$

In particular, we have $\tilde{m}_0(x,\bar{\boldsymbol{\beta}}_k)=\bar{m}_0(x,\bar{\boldsymbol{\beta}}_k)$, where

$$\bar{m}_0(x,\bar{\beta}_0)=\frac{c_1}{\frac{\alpha_1\bar{\beta}_0}{c_1+\bar{e}_0\left(x,\bar{\beta}_0\right)}+\frac{1}{x}}+(1-c_1)x\tag{2.279}$$

$$\bar{e}_0\left(x,\bar{\beta}_0\right)=-\frac{x\alpha_1\bar{\beta}_0(1-c_1)+c_1}{2}+\frac{\sqrt{\left(x\alpha_1\bar{\beta}_0(1-c_1)+c_1\right)^2+4x\alpha_1\bar{\beta}_0c_1^2}}{2}.\tag{2.280}$$

Replacing $\mathbf{R}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)$ in (2.278) by its recursive definition $\mathbf{R}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)=\mathbf{I}_{n_k}+\frac{\alpha_k\bar{\beta}_{k-1}}{n_{k-1}}\mathbf{H}_k\mathbf{R}_{k-1}\left(\bar{\boldsymbol{\beta}}_{k-2}\right)\mathbf{H}_k^{\mathsf{H}}$ (2.16) yields after straightforward calculus

$$\begin{aligned}&e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)=c_{k+1}\left(c_{k+1}+e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)\right)\\&-\frac{c_{k+1}\left(c_{k+1}+e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)\right)^2}{x\alpha_{k+1}\bar{\beta}_k}m_{k-1}\left(\frac{x\alpha_{k+1}\bar{\beta}_k}{c_{k+1}+x\alpha_{k+1}\bar{\beta}_k+e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)},\bar{\boldsymbol{\beta}}_{k-1}\right).\end{aligned}\tag{2.281}$$

Similarly, one obtains

$$\begin{aligned}&\tilde{m}_k(x,\bar{\boldsymbol{\beta}}_k)\\&=\frac{c_{k+1}\left(c_{k+1}+e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)\right)}{\alpha_{k+1}\bar{\beta}_k}m_{k-1}\left(\frac{x\alpha_{k+1}\bar{\beta}_k}{c_{k+1}+x\alpha_{k+1}\bar{\beta}_k+e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)},\bar{\boldsymbol{\beta}}_{k-1}\right)\\&\quad+(1-c_{k+1})x.\end{aligned}\tag{2.282}$$

Combining the last two equations leads to

$$\tilde{m}_k(x,\bar{\boldsymbol{\beta}}_k) = \frac{xc_{k+1}}{c_{k+1}+e_k\left(x,\bar{\boldsymbol{\beta}}_k\right)}. \tag{2.283}$$

Consider now $\bar{e}_k\left(x,\bar{\boldsymbol{\beta}}_k\right)$, $k \geq 1$, defined as a positive solution to

$$\begin{aligned}\bar{e}_k\left(x,\bar{\boldsymbol{\beta}}_k\right) &= c_{k+1}\left(c_{k+1}+\bar{e}_k\left(x,\bar{\boldsymbol{\beta}}_k\right)\right)\\ &\quad - \frac{c_{k+1}\left(c_{k+1}+\bar{e}_k\left(x,\bar{\boldsymbol{\beta}}_k\right)\right)^2}{x\alpha_{k+1}\bar{\beta}_k}\bar{m}_{k-1}\left(\frac{x\alpha_{k+1}\bar{\beta}_k}{c_{k+1}+x\alpha_{k+1}\bar{\beta}_k+\bar{e}_k\left(x,\bar{\boldsymbol{\beta}}_k\right)},\bar{\boldsymbol{\beta}}_{k-1}\right)\end{aligned} \tag{2.284}$$

where $\bar{m}_k\left(x,\bar{\boldsymbol{\beta}}_k\right)$ is recursively defined for $k \geq 1$ as

$$\bar{m}_k\left(x,\bar{\boldsymbol{\beta}}_k\right) = \frac{xc_{k+1}}{c_{k+1}+\bar{e}_k\left(x,\bar{\boldsymbol{\beta}}_k\right)}. \tag{2.285}$$

It remains to show that a unique solution to (2.284) exists and that $e_k\left(x,\bar{\boldsymbol{\beta}}_k\right) - \bar{e}_k\left(x,\bar{\boldsymbol{\beta}}_k\right) \xrightarrow{\text{a.s.}} 0$. Let us first define the following functions for $k \geq 1$:

$$f_k(z) = c_{k+1}\left(c_{k+1}+z\right) - \frac{c_{k+1}\left(c_{k+1}+z\right)^2}{x\alpha_{k+1}\bar{\beta}_k}m_{k-1}\left(\frac{x\alpha_{k+1}\bar{\beta}_k}{c_{k+1}+x\alpha_{k+1}\bar{\beta}_k+z},\bar{\boldsymbol{\beta}}_{k-1}\right) \tag{2.286}$$

$$\bar{f}_k(z) = c_{k+1}\left(c_{k+1}+z\right) - \frac{c_{k+1}\left(c_{k+1}+z\right)^2}{x\alpha_{k+1}\bar{\beta}_k}\bar{m}_{k-1}\left(\frac{x\alpha_{k+1}\bar{\beta}_k}{c_{k+1}+x\alpha_{k+1}\bar{\beta}_k+z},\bar{\boldsymbol{\beta}}_{k-1}\right). \tag{2.287}$$

From (2.278) and with the help of Lemma 20 (note that the smallest eigenvalue of $\mathbf{R}_k$ is greater or equal to 1 for all k), one can easily verify that $f_k(z)$ satisfies the following properties for $z \geq 0$:

(i)

$$f_k(z) \geq c_{k+1}(c_{k+1}+z)\left[1-\frac{c_{k+1}+z}{c_{k+1}+z+x\alpha_{k+1}\bar{\beta}_k}\right] > 0 \tag{2.288}$$

(ii) for $z > z' \geq 0$,

$$\begin{aligned}&f_k(z)-f_k(z')\\ &\geq \frac{(z-z')c_{k+1}\alpha_{k+1}^2\bar{\beta}_k^2}{(c_{k+1}+z')(c_{k+1}+z)}\\ &\qquad\times\frac{1}{n_k}\operatorname{tr}\mathbf{R}_k\left(\frac{\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k}{c_{k+1}+z}+\frac{1}{x}\mathbf{I}_{n_k}\right)^{-1}\mathbf{R}_k\left(\frac{\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k}{c_{k+1}+z'}+\frac{1}{x}\mathbf{I}_{n_k}\right)^{-1}\\ &\geq \frac{(z-z')c_{k+1}\alpha_{k+1}^2\bar{\beta}_k^2}{\left(\alpha_{k+1}\bar{\beta}_k+\frac{c_{k+1}+z'}{x}\right)\left(\alpha_{k+1}\bar{\beta}_k+\frac{c_{k+1}+z}{x}\right)}\\ &> 0\end{aligned} \tag{2.289}$$

(iii) for $\alpha > 1$,

$$
\begin{aligned}
&\alpha f_k(z) - f_k(\alpha z) \\
\geq\ & \frac{(\alpha-1)c_{k+1}^2\alpha_{k+1}^2\bar{\beta}_k^2}{(c_{k+1}+\alpha z)(\alpha c_{k+1}+\alpha z)} \\
&\quad \times \frac{1}{n_k}\operatorname{tr}\mathbf{R}_k\left(\frac{\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k}{\alpha c_{k+1}+\alpha z}+\frac{1}{\alpha x}\mathbf{I}_{n_k}\right)^{-1}\mathbf{R}_k\left(\frac{\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k}{c_{k+1}+\alpha z}+\frac{1}{x}\mathbf{I}_{n_k}\right)^{-1} \\
&+ \frac{(\alpha-1)c_{k+1}\alpha_{k+1}\bar{\beta}_k}{\alpha x} \\
&\quad \times \frac{1}{n_k}\operatorname{tr}\mathbf{R}_k\left(\frac{\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k}{\alpha c_{k+1}+\alpha z}+\frac{1}{\alpha x}\mathbf{I}_{n_k}\right)^{-1}\left(\frac{\alpha_{k+1}\bar{\beta}_k\mathbf{R}_k}{c_{k+1}+\alpha z}+\frac{1}{x}\mathbf{I}_{n_k}\right)^{-1} \\
\geq\ & \frac{(\alpha-1)c_{k+1}^2\alpha_{k+1}^2\bar{\beta}_k^2}{\left(\alpha_{k+1}\bar{\beta}_k+\frac{c_{k+1}+\alpha z}{x}\right)\left(\alpha_{k+1}\bar{\beta}_k+\frac{\alpha c_{k+1}+\alpha z}{x}\right)} \\
&+ \frac{(\alpha-1)c_{k+1}\alpha_{k+1}\bar{\beta}_k}{\alpha x\left(\frac{\alpha_{k+1}\bar{\beta}_k}{\alpha c_{k+1}+\alpha z}+\frac{1}{\alpha x}\right)\left(\frac{\alpha_{k+1}\bar{\beta}_k}{c_{k+1}\alpha z}+\frac{1}{x}\right)} \\
>\ & 0
\end{aligned}
\tag{2.290}
$$

where $\mathbf{R}_k = \mathbf{R}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)$. All properties are independent of $\mathbf{R}_k$ and therefore hold for $n \to \infty$.

Assume now $k = 1$. For any sequence of bounded non-negative real numbers z_n, we have by (2.276) and the continuous mapping theorem,

$$
f_1(z_n) - \bar{f}_1(z_n) \xrightarrow{\text{a.s.}} 0. \tag{2.291}
$$

Thus, properties $(i) - (iii)$ of $f_1(z)$ also hold for $\bar{f}_1(z)$. By Definition 11 and Theorem 16, these properties imply the uniqueness of positive solutions to the fixed point equations $z = f_1(z)$ and $y = \bar{f}_1(y)$, and hence the uniqueness of solutions to (2.284) for $k = 1$. Moreover, note that

$$
|f_k(a) - f_k(b)| \leq \frac{\alpha_{k+1}^2\bar{\beta}_k^2x^2}{c_{k+1}}\|\mathbf{R}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)\|^2|a-b|. \tag{2.292}
$$

Hence,

$$
\begin{aligned}
&\left|\bar{e}_1\left(x,\bar{\boldsymbol{\beta}}_1\right) - e_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right| \\
=\ & \left|\bar{f}_1\left(\bar{e}_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right) - f_1\left(e_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right)\right| \\
\leq\ & \left|\bar{f}_1\left(\bar{e}_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right) - f_1\left(\bar{e}_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right)\right| + \left|f_1\left(\bar{e}_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right) - f_1\left(e_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right)\right| \\
\leq\ & \epsilon_n + \frac{\alpha_2^2\bar{\beta}_1^2x^2}{c_2}\|\mathbf{R}_1\left(\bar{\beta}_0\right)\|^2\left|\bar{e}_1\left(x,\bar{\boldsymbol{\beta}}_1\right) - e_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right|
\end{aligned}
\tag{2.293}
$$

for some sequence of real numbers ϵ_n, satisfying $\epsilon_n \xrightarrow{\text{a.s.}} 0$. By Lemma 19, $\|\mathbf{R}_1\left(\bar{\beta}_0\right)\| < M$, almost surely, for some $M > 0$. Thus, for $x \leq \sqrt{\frac{c_2(1-\delta)}{\alpha_2^2\bar{\beta}_1^2M^2}}$ and some $\delta > 0$, we have

$$
\left|\bar{e}_1\left(x,\bar{\boldsymbol{\beta}}_1\right) - e_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\right| \leq \frac{\epsilon_n}{\delta} \xrightarrow{\text{a.s.}} 0. \tag{2.294}
$$

Since $\bar{e}_1\left(x,\bar{\boldsymbol{\beta}}_1\right)$ and $e_1\left(x,\bar{\boldsymbol{\beta}}_1\right)$ are (almost surely) bounded on any closed subset of $\mathbb{R}^+\setminus\{0\}$ and have analytic continuations for $x\in\mathbb{C}\setminus\mathbb{R}^-$, we have by Vitali's convergence theorem that the convergence holds for any $x\in\mathbb{R}^+\setminus\{0\}$.

The last convergence implies by the continuous mapping theorem that

$$m_1(x,\bar{\boldsymbol{\beta}}_1)-\bar{m}_1\left(x,\bar{\boldsymbol{\beta}}_1\right)\xrightarrow{\text{a.s.}}0. \tag{2.295}$$

We now assume $k=2$. The last convergence implies $f_2(z)\to\bar{f}_2(z)$, almost surely. The same steps can therefore be applied to show that $m_2(x,\bar{\boldsymbol{\beta}}_1)-\bar{m}_2\left(x,\bar{\boldsymbol{\beta}}_1\right)\xrightarrow{\text{a.s.}}0$. This terminates the proof as this process can be iterated k times.

Part *(ii)*

First, notice that

$$\begin{aligned}\eta_k &\triangleq \left\|\mathbf{R}_k(\boldsymbol{\beta}_{k-1})-\mathbf{R}_k(\bar{\boldsymbol{\beta}}_{k-1})\right\|\\
&=\left\|\frac{\alpha_k\beta_{k-1}}{n_{k-1}}\mathbf{H}_k\mathbf{R}_{k-1}(\boldsymbol{\beta}_{k-2})\mathbf{H}_k^{\mathsf{H}}-\frac{\alpha_k\bar{\beta}_{k-1}}{n_{k-1}}\mathbf{H}_k\mathbf{R}_{k-1}(\bar{\boldsymbol{\beta}}_{k-2})\mathbf{H}_k^{\mathsf{H}}\right\|\\
&\le\alpha_k\left\|\frac{\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}}{n_{k-1}}\right\|\left\|\beta_{k-1}\mathbf{R}_{k-1}(\boldsymbol{\beta}_{k-2})-\bar{\beta}_{k-1}\mathbf{R}_{k-1}(\bar{\boldsymbol{\beta}}_{k-2})\right\|\\
&\le\alpha_k\left\|\frac{\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}}{n_{k-1}}\right\|\left[\left|\beta_{k-1}-\bar{\beta}_{k-1}\right|\left\|\mathbf{R}_{k-1}(\boldsymbol{\beta}_{k-2})\right\|\right.\\
&\qquad\left.+\bar{\beta}_{k-1}\left\|\mathbf{R}_{k-1}(\boldsymbol{\beta}_{k-2})-\mathbf{R}_{k-1}(\bar{\boldsymbol{\beta}}_{k-2})\right\|\right].\end{aligned} \tag{2.296}$$

Since, almost surely, $\limsup\left\|\mathbf{R}_1(\boldsymbol{\beta}_0)-\mathbf{R}_1(\bar{\boldsymbol{\beta}}_0)\right\|\le\limsup\alpha_1|\beta_0-\bar{\beta}_0|\left\|\frac{\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}}}{n}\right\|=0$ and $\limsup\left\|\mathbf{R}_{k-1}(\boldsymbol{\beta}_{k-2})\right\|<\infty$ (see proof of Lemma 19), one can iteratively show that $\eta_k\to0$, almost surely. Thus,

$$\left|\mathcal{J}_k\left(x,\boldsymbol{\beta}_{k-1}\right)-\mathcal{J}_k\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)\right|\xrightarrow[n\to\infty]{\text{a.s.}}0. \tag{2.297}$$

This means that we can from now on replace $\boldsymbol{\beta}_k$ by $\bar{\boldsymbol{\beta}}_k$ and focus on $\mathcal{J}_k\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)$.

As a consequence of Corollary 1, Lemma 19, and the Fubini theorem, we obtain the following relation

$$\mathcal{J}_k(x,\bar{\boldsymbol{\beta}}_{k-1})-\tilde{\mathcal{J}}_k(x,\bar{\boldsymbol{\beta}}_{k-1})\xrightarrow{\text{a.s.}}0,\qquad k\ge1 \tag{2.298}$$

where

$$\begin{aligned}\tilde{\mathcal{J}}_k\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)=&\frac{1}{n_k}\log\det\left(\left[c_k+e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)\right]\mathbf{I}_{n_{k-1}}+x\alpha_k\bar{\beta}_{k-1}\mathbf{R}_{k-1}\left(\bar{\boldsymbol{\beta}}_{k-2}\right)\right)\\
&+(1-c_k)\log\left(c_k+e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)\right)\\
&-\frac{e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)}{c_k+e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)}-\log\left(c_k\right)\end{aligned} \tag{2.299}$$

and $e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)$ is given as the unique positive solution to

$$\begin{aligned}&e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)\\
&=\frac{1}{n_k}\operatorname{tr}\alpha_k\bar{\beta}_{k-1}\mathbf{R}_{k-1}\left(\bar{\boldsymbol{\beta}}_{k-2}\right)\left(\frac{\alpha_k\bar{\beta}_{k-1}\mathbf{R}_{k-1}\left(\bar{\boldsymbol{\beta}}_{k-2}\right)}{c_k+e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)}+\frac{1}{x}\mathbf{I}_{n_{k-1}}\right)^{-1}.\end{aligned} \tag{2.300}$$

In particular, for $k=1$, we have

$$\tilde{\mathcal{J}}_1\left(x,\bar{\beta}_0\right)=\bar{\mathcal{J}}_1\left(x,\bar{\beta}_0\right) \tag{2.301}$$

where

$$\begin{aligned}\bar{\mathcal{J}}_1\left(x,\bar{\beta}_0\right)&=c_1\log\left(1+\frac{x\alpha_1\bar{\beta}_0}{c_1+\bar{e}_0\left(x,\bar{\beta}_0\right)}\right)+\log\left(1+\frac{\bar{e}_0\left(x,\bar{\beta}_0\right)}{c_1}\right)\\&\quad-\frac{\bar{e}_0\left(x,\bar{\beta}_0\right)}{c_1+\bar{e}_0\left(x,\bar{\beta}_0\right)} \end{aligned}\tag{2.302}$$

$$\bar{e}_0\left(x,\bar{\beta}_0\right)=-\frac{x\alpha_1\bar{\beta}_0(1-c_1)+c_1}{2}+\frac{\sqrt{\left(x\alpha_1\bar{\beta}_0(1-c_1)+c_1\right)^2+4x\alpha_1\bar{\beta}_0c_1^2}}{2} \tag{2.303}$$

according to Corollary 1. Note that (2.270) permits a closed-form solution (2.303) in this case.

Using the recursion $\mathbf{R}_{k-1}\left(\bar{\boldsymbol{\beta}}_{k-2}\right)=\mathbf{I}_{n_{k-1}}+\frac{\alpha_{k-1}\bar{\beta}_{k-2}}{n_{k-2}}\mathbf{H}_{k-1}\mathbf{R}_{k-2}\left(\bar{\boldsymbol{\beta}}_{k-3}\right)\mathbf{H}_{k-1}^{\mathsf{H}}$ (2.16) in (2.299), we obtain

$$\begin{aligned}\tilde{\mathcal{J}}_k(x,\bar{\boldsymbol{\beta}}_{k-1})&=c_k\mathcal{J}_{k-1}\left(\frac{x\alpha_k\bar{\beta}_{k-1}}{c_k+x\alpha_k\bar{\beta}_{k-1}+e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)},\bar{\boldsymbol{\beta}}_{k-2}\right)\\&\quad+c_k\log\left(1+\frac{x\alpha_k\bar{\beta}_{k-1}}{c_k+e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)}\right)\\&\quad+\log\left(1+\frac{e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)}{c_k}\right)-\frac{e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)}{c_k+e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)}.\end{aligned}\tag{2.304}$$

In the proof of Part (i), it is shown that

$$e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)-\bar{e}_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)\xrightarrow{\text{a.s.}}0. \tag{2.305}$$

By the continuous mapping theorem, we therefore have

$$\begin{aligned}&\mathcal{J}_{k-1}\left(\frac{x\alpha_k\bar{\beta}_{k-1}}{1+x\alpha_k\bar{\beta}_{k-1}+e_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)},\bar{\boldsymbol{\beta}}_{k-2}\right)\\&\quad-\mathcal{J}_{k-1}\left(\frac{x\alpha_k\bar{\beta}_{k-1}}{1+x\alpha_k\bar{\beta}_{k-1}+\bar{e}_{k-1}\left(x,\bar{\boldsymbol{\beta}}_{k-1}\right)},\bar{\boldsymbol{\beta}}_{k-2}\right)\xrightarrow{\text{a.s.}}0.\end{aligned}\tag{2.306}$$

Applying the last result together with Corollary 1, Lemma 19, the continuous mapping theorem and the Fubini theorem to (2.304) concludes the proof for $k=2$ since $\tilde{\mathcal{J}}_1\left(x,\bar{\beta}_0\right)=\bar{\mathcal{J}}_1\left(x,\bar{\beta}_0\right)$ by (2.301). The convergence for $k>2$ is shown by successive iterations of the last steps.

Chapter 3

Applications

In this chapter, we will present numerous practical applications of the theoretical results developed in Chapter 2. In Section 3.1, we study the impact of channel training on the performance of uplink network MIMO system with finite capacity backhaul links. Assuming a block-fading channel model with finite coherence time, there is an interesting trade-off between the time used for training and data transmissions. We build upon Theorem 12 to determine the optimal amount of channel training which maximizes a deterministic equivalent of the net ergodic achievable rate. In Section 3.2, we use the asymptotic moment results of Theorem 19 and 20 for the implementation and analysis of a polynomial expansion receiver. Simulations suggest that this receiver type could potentially reduce the computational complexity of the MMSE detector while achieving a large fraction of its performance. Section 3.3 deals with the analysis of the up- and downlink performance of large-scale MIMO systems. We derive tight approximations of achievable rates with different linear precoders and detectors for a very general channel model which accounts for imperfect CSI, pilot contamination, path loss, and arbitrary antenna correlation. The analysis relies to a large extent on Theorems 14 and 21. We further critically discuss several assumptions in existing works on large-scale MIMO and investigate if additional antennas can compensate for sub-optimal signal processing. Sections 3.4, 3.5, and 3.6 are mainly based on the concept of iterative deterministic equivalents. In Section 3.4, we apply Theorem 22 to the analysis of double-scattering MACs. We also establish a deterministic equivalent of the sum-rate with MMSE-detection and find the asymptotically optimal precoding matrices. For the special case of Rayleigh-product channels all results can be given in closed form. In Section 3.5, we consider the AF multihop relay channel with noise at each stage. Theorem 24 allows us to provide a deterministic equivalent of the mutual information after each hop. Although this problem has been received considerable research interest, it has not been solved in the literature before. In the last section of this chapter, Section 3.6, we present an asymptotic analysis of random beamforming over quasi-static and fading channels. The analysis relies to a large extent on Theorems 15, 18, and 23. For both scenarios, we derive deterministic equivalents of the mutual information, the SINR and associated rates with MMSE detection. Especially the derivation of deterministic equivalents for random matrix models combining random unitary and i.i.d. matrices is new.

3.1 Optimal channel training in uplink network MIMO systems

Abstract: *We consider a multi-cell frequency-selective fading uplink channel from K single-antenna UTs to B cooperative BSs with M antennas each. The BSs, assumed to be oblivious of the applied codebooks, forward compressed versions of their observations to a central station (CS) via capacity limited backhaul links. The CS jointly decodes the messages from all UTs. Since the BSs and the CS are assumed to have no prior CSI, the channel needs to be estimated during its coherence time. Based on a lower bound of the ergodic mutual information, we determine the optimal fraction of the coherence time used for channel training, taking different path losses between the UTs and the BSs into account. We then study how the optimal training length is impacted by the backhaul capacity. Although our analytical results based on random matrix theory are proved to be tight in the large system limit, we show by simulations that they provide very accurate approximations for even small system dimensions.*

3.1.1 Introduction

Network MIMO has become the synonym for cooperative communications in the cellular context and is regarded as an important concept to boost the interference limited performance of today's cellular networks. It is often also referred to as multi-cell processing or distributed antenna systems and corresponds to a communication system where multiple BSs, connected via high speed backhaul links to a CS, *jointly* process data either received over the uplink or transmitted over the downlink. If the BSs could cooperate without any restrictions with regards to the backhaul capacity, processing delay, computing complexity and the availability of CSI , the multi-cell interference channel would be transformed into a MAC (uplink) or broadcast channel (downlink) without multi-cell interference. This argument motivated the concept of network MIMO and it has been shown in many works, e.g., [96, 97], that BS-cooperation has the potential to realize significant gains in throughput and reliability.

So far, the treatment of multi-cell cooperation in the literature has been either information-theoretic but limited to simple models [98, 99] or based on simulations to account for more realistic and complex network structures [100, 101, 102]. The most common and analytically tractable network models are the Wyner model [103, 104] and the soft-hand-off model [105, 106] which consider cooperation between either two or three adjacent BSs on an infinite linear or circular cellular array. Variants of both models have been studied under various assumptions on the transmission schemes and the fading characteristics.

In practical systems, perfect BS-cooperation or global processing is very difficult, if not impossible, to achieve. The main limitations are threefold: (i) limited backhaul capacity, (ii) local connectivity and (iii) imperfect CSI at the CS and the BSs.[1] Therefore, most of recent research targets the problem of constrained cooperation. For a detailed overview of this topic we refer to the surveys [27, 107, 108]. Information-theoretic implications of limited backhaul capacity have been studied separately for the uplink and downlink in [109] and [110]. Recently, the optimal amount of user data sharing between the BSs for

[1] Also the synchronization of the BSs as well as processing complexity and delay are limiting factors from an implementation perspective but are so far more or less neglected in the literature.

the downlink with linear beamforming and backhaul constraints was studied in [111]. The difficulties related to connecting a large number of BSs to a single CS have motivated the study of systems with only locally connected BSs [110, 112, 113]. Several distributed algorithms for the uplink [114] and downlink [115, 116] have been proposed and it was shown that even with local BS-connection near-optimal performance can be achieved with a reasonable amount of message passing and computational complexity.

One of the most critical limitations of a practical network MIMO system, somehow overlooked compared to (i) and (ii), arises from the substantial overhead related to the acquisition of CSI (iii), indispensable to achieve the full diversity or multiplexing gains. This overhead becomes paramount, in particular for fast fading channels, when the number of antennas, sub-carriers, UTs or BSs grows [16, 101, 102, 29]. Usually, CSI for the uplink is acquired through pilot signals sent by the UTs . This implies that a part of the coherence time of the channel needs to be sacrificed to obtain CSI with a sufficiently high quality. The inherent trade-off between the resources dedicated to channel estimation and data transmission has been studied for the point-to-point MIMO channel [117, 118] and the multi-user downlink [119, 120]. Recently, this problem was also addressed in the context of network MIMO systems, although with a different focus. In [29, 101, 102], the authors compare several multi-cellular system architectures and conclude that the downlink performance of network MIMO systems is mainly limited by the inevitable acquisition of CSI (rather than by limited backhaul capacity). They also demonstrate that a conventional cellular system might outperform a network MIMO system under some circumstances assuming that the number of *coordinated* antennas and the used training overhead for both systems are the same. This means in essence that simply installing more antennas per BS can lead to higher performance improvements than installing costly backhaul infrastructure.

The imperfections detailed above call for robust strategies adapted to restricted BS-cooperation. Some schemes [121, 122] rely on local CSI at the BSs and statistical CSI at the CS, whereas others [123, 100] consider to serve only certain subsets of UTs with multiple BSs. Several BS-cooperation schemes have been studied in [124, 125] for the combination of limited backhaul capacity and imperfect CSI .

In this section, we also consider limited BS-cooperation by focusing especially on the effects of imperfect CSI (iii). More precisely, we study the performance of the multi-cell uplink with partially restricted cooperation assuming that:

- The BSs act as oblivious relays which forward compressed versions of their received signals to the CS via orthogonal error- and delay-free backhaul links, each of fixed capacity C bits/channel use.
- The CS estimates the channel based on pilot tones sent by the UTs .
- The CS jointly processes the received signals from all BSs.

We consider a lower bound of the normalized ergodic mutual information of the network MIMO uplink channel with imperfect CSI and limited backhaul capacity, called the net ergodic achievable rate $R_{\rm net}(\tau)$. For a given channel coherence time T, we attempt to find the optimal length τ^* of the pilot sequences for channel training which maximizes $R_{\rm net}(\tau)$. As this optimization problem is in general intractable, we study a deterministic approximation $\overline{R}_{\text{net}}(\tau)$ of $R_{\text{net}}(\tau)$, based on large random matrix theory.

Our main contribution is to show that optimizing $\overline{R}_{\text{net}}(\tau)$ instead of $R_{\text{net}}(\tau)$ is optimal in the large system limit. To this end, we provide a closed-form expression of the derivative of $\overline{R}_{\text{net}}(\tau)$ (Theorem 26), prove the concavity of $\overline{R}_{\text{net}}(\tau)$ for channel matrices with a doubly regular variance profile (Theorem 27), and show that $\overline{\tau}^*$ which maximizes $\overline{R}_{\text{net}}(\tau)$ converges to τ^* in the large system limit (Theorem 28). We further demonstrate by simulations that our asymptotic results yield tight approximations for systems of small dimensions with as little as three BSs and UTs. In addition, we study the effects of limited backhaul capacity on the optimal channel training length. Since we assume that the CS estimates all channels based on the compressed observations from the BSs, the channel estimates are impaired by thermal noise *and* quantization errors. Thus, increasing the backhaul capacity leads to improved channel estimates and, hence, smaller values of τ^*.

The determination of the optimal training length τ^* in an uplink network MIMO setting with arbitrary path loss between the UTs and BSs and limited backhaul capacity appears to be a novel result, although we limit our investigation to a simple setting where B cooperative BSs do not suffer from interference outside the network. The extension of this work to more realistic networks, such as clustered systems, is left to future investigations. Although the use of random matrix theory in the context of network MIMO is not new, see e.g., [126, 127], we present a novel application to an optimization problem in wireless communications.

3.1.2 System Model

Channel Model

We consider a multi-cell frequency-selective fading uplink channel from K single-antenna UTs to B BSs with M antennas each.[2] A schematic diagram of the channel model for $M = 2$ is given in Fig. 3.1. Communication takes place simultaneously from all UTs to all BSs on L parallel sub-carriers assuming an OFDMtransmission scheme. The stacked receive vector of all BSs on the ℓth sub-carrier $\mathbf{y}(\ell) = [y_1(\ell), \dots, y_{BM}(\ell)]^{\mathsf{T}} \in \mathbb{C}^{BM}$ at a given time reads

$$\mathbf{y}(\ell) = \mathbf{H}(\ell)\mathbf{x}(\ell) + \mathbf{n}(\ell), \qquad \ell = 1, \dots, L \tag{3.1}$$

where $\mathbf{x}(\ell) = [x_1(\ell), \dots, x_K(\ell)]^{\mathsf{T}} \in \mathbb{C}^K$ is the vector of the transmitted signals of all UTs on sub-carrier ℓ, $\mathbf{n}(\ell) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{BM})$ is a vector of additive noise and $\mathbf{H}(\ell) \in \mathbb{C}^{BM \times K}$ is the aggregated channel matrix from all UTs to all BSs on the ℓth sub-carrier.

We consider a discrete-time block-fading channel model where the channel remains constant for a coherence block of T channel uses and then changes randomly from one block to the other. We let $T = T_c W_c$, where W_c is the bandwidth per sub-carrier in Hz and T_c the channel coherence time in seconds. Presuming that the bandwidth of each sub-carrier W_c is on the order of the channel coherence bandwidth, that the antenna spacing at the BSs is sufficiently

[2]Our results can be easily extended to the case where each BS has a different number of antennas.

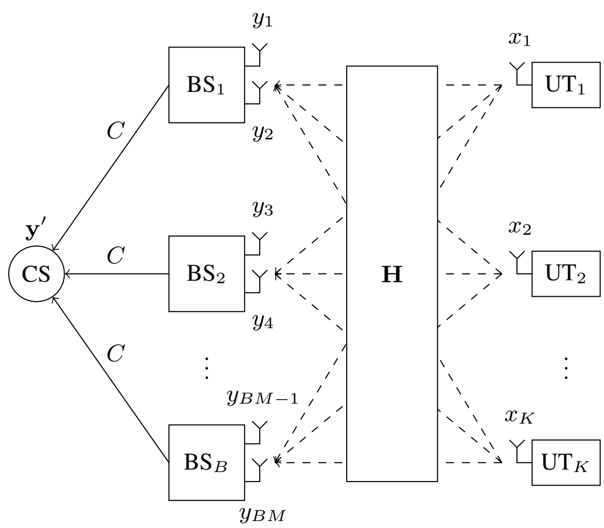

Figure 3.1: Schematic system model for $M = 2$ antennas per BS. The BSs compress and forward their received signals to the CS via orthogonal backhaul links of capacity C bits/channel use. The CS jointly processes the received data from all BSs.

large and that the channels from the UTs to the BSs are uncorrelated, the channel matrices $\mathbf{H}_b(\ell)$, $b = 1, \ldots, B$, from the UTs to the BSs can be modeled as

$$\mathbf{H}_b(\ell) = \mathbf{W}_b(\ell)\,\mathrm{diag}\left(\sqrt{a_{b1}}, \ldots, \sqrt{a_{bK}}\right), \qquad \ell = 1, \ldots, L \tag{3.2}$$

where $\mathbf{W}_b(l) \in \mathbb{C}^{M\times K}$ is a standard complex Gaussian matrix and a_{bk} denotes the inverse path loss between UT k and BS b.[3] For later use, we define the matrix $\mathbf{V} \in (\mathbb{R}^+)^{BM\times K}$ in the following way:

$$\mathbf{V} = \mathbf{A} \otimes \mathbf{1}_M \tag{3.3}$$

where $\mathbf{A} \in (\mathbb{R}^+)^{B\times K}$ is the inverse path loss matrix with elements $\{a_{bk}\}$ and $\mathbf{1}_M$ is a M-dimensional column vector with all entries equal to one, such that the elements $\{v_{ij}\}$ of $\mathbf{V}$ satisfy $v_{ij} = a_{\lceil \frac{M}{i} \rceil j}$. Under these assumptions, the elements $\{h_{ij}(\ell)\}$ of the matrix $\mathbf{H}(\ell)$ are independent circular symmetric complex Gaussian random variables with zero mean and variance v_{ij}, i.e., $h_{ij}(\ell) \sim \mathcal{CN}(0, v_{ij})$. We refer to $\mathbf{V}$ as the *variance profile* of the channel matrix $\mathbf{H}(l)$ and assume in the sequel that $\mathbf{V}$ is perfectly known at the CS while each BS b only knows the distribution of its local channels $\mathbf{H}_b(\ell)$, $\ell = 1, \ldots, L$. In a practical system, the channel coherence bandwidth might be significantly larger than the bandwidth of a sub-carrier so that $\{h_{ij}(\ell)\}$ would exhibit some correlation with respect to ℓ. From a channel estimation perspective, the assumption of i.i.d. channel coefficients represents a worst case since sub-carrier correlation cannot be exploited in the estimation process.

[3]Note that the path loss is independent of the sub-carrier index ℓ. This might not be the case for extremely large bandwidth but it is a reasonable assumption for most practical scenarios.

For simplicity, we assume Gaussian signaling with uniform power allocation, i.e., $x_k(\ell) \sim \mathcal{CN}(0, P/L)$, i.i.d. over ℓ and k, which is not necessarily optimal in the presence of channel estimation errors [128, 117]. Although optimal power allocation over the sub-carriers would provide significant gains, it would require perfect channel knowledge at the UTs or some sort of feedback from the BSs/CS. Since we assume neither feedback nor CSI at the UTs and since the channel statistics are the same for all sub-carriers, uniform power allocation seems to be a reasonable choice.

Compression at the BSs

The BSs are assumed to be oblivious to the applied codebooks of the UTs and forward compressed versions $y_i'(\ell)$ of their received signal sequences $y_i(\ell)$ to the CS via orthogonal backhaul links, each of capacity C bits per channel use.[4] We also assume that the BSs and the CS have no prior knowledge of the instantaneous channel realizations. Under this setting, we consider a simple, sub-optimal compression scheme which neither exploits correlations between the received signals at different antennas nor adapts the employed quantization codebook to the actual channel realization. Thus, a single quantization codebook for the compression of each sequence $y_i(\ell)$ is used. This is in contrast to existing works, e.g., [129], which rely on the assumption of full CSI at the BSs and the CS to apply optimized and channel dependent compression schemes. For a more detailed discussion of different (distributed) compression schemes, we refer to [129, 130, 131, 125] and references therein.

The rate-distortion function for the source $y_i(\ell)$ with squared error distortion is given as [42, Theorem 10.2.1]

$$R_D\left(\sigma_i^2(\ell)\right) = \min_{f_{y_i'(\ell)|y_i(\ell)}:\mathbb{E}\left[|y_i'(\ell)-y_i(\ell)|^2\right]\leq\sigma_i^2(\ell)} I\left(y_i'(\ell); y_i(\ell)\right) \tag{3.4}$$

where the minimization is over all conditional probability density functions $f_{y_i'(\ell)|y_i(\ell)}$ satisfying the expected distortion constraint $\sigma_i^2(\ell)$. Similar to the so-called "elementary compression scheme" in [129], our compression scheme is based on an underlying complex Gaussian "test channel" defined by

$$y_i'(\ell) = y_i(\ell) + q_i(\ell) \tag{3.5}$$

where $q_i(\ell) \sim \mathcal{CN}(0, \sigma_i^2(\ell))$. Note that the test channel (3.5) used for the generation of the quantization codebooks is not optimal since the distribution of $y_i(\ell) = \sum_{j=1}^{K} h_{ij}(\ell)x_j(\ell) + n_i(\ell)$ is not Gaussian. However, one can argue that in a large system with many UTs, the random variable $y_i(\ell)$ is almost Gaussian distributed and the performance degradation due to the sub-optimal choice of $f_{y_i'(\ell)|y_i(\ell)}$ is small. A simple upper bound of the rate distortion function is given

[4]By *orthogonal* backhaul links we mean here that there is no inter-backhaul interference. This is for example the case for a wired backhaul network with a dedicated link between the CS and each BS.

by

$$
\begin{aligned}
I(y_i'(\ell); y_i(\ell)) &= h(y_i'(\ell)) - h(y_i'(\ell)|y_i(\ell)) \\
&\le \log\left(\pi e \left(\mathbb{E}\left[|y_i(\ell)|^2\right] + \sigma_i^2(\ell)\right)\right) - \log\left(\pi e \sigma_i^2(\ell)\right) \\
&= \log\left(1 + \frac{1 + \frac{P}{L}\sum_{j=1}^{K} v_{ij}}{\sigma_i^2(\ell)}\right)
\end{aligned} \tag{3.6}
$$

where the inequality is obtained by upper-bounding the entropy of $y_i'(\ell)$ by the entropy of a complex Gaussian random variable with the same variance. We assume further that each BS uses $C/(ML)$ bits for the compression of each received complex symbol per antenna per sub-carrier. Replacing the left-hand side (LHS) of (3.6) by $C/(ML)$, we can consequently overestimate the quantization noise variance $\sigma_i^2(\ell)$ by choosing

$$
\sigma_i^2 = \sigma_i^2(\ell) = \frac{1 + \frac{P}{L}\sum_{j=1}^{K} v_{ij}}{2^{\frac{C}{ML}} - 1}\ . \tag{3.7}
$$

Since the statistical distribution of $y_i(\ell)$ is the same for all sub-carriers, the quantization noise power σ_i^2 is also independent of ℓ. One can easily verify that the quantization noise vanishes for infinite backhaul capacity, i.e., $\sigma_i^2 \to 0$ for $C \to \infty$, and grows without bounds when the backhaul has zero capacity, i.e., $\sigma_i^2 \to \infty$ for $C \to 0$.

We would like to point out that the field of distributed compression with imperfect CSI is to the best of our knowledge a largely unexplored area. It is for example not clear if each BS should estimate its local channels and forward compressed versions of its estimates to the CS or if the CS should estimate all channels based on compressed signals from the BSs, as assumed here.

Channel Training

Similar to [117], each channel coherence block of length T is split into a phase for channel training and a phase for data transmission. During the training phase of length τ, all K UTs broadcast orthogonal sequences of known pilot symbols of equal power P/L on all sub-carriers. The orthogonality of the training sequences imposes $\tau \ge K$. We assume that the CS estimates the channels $h_{ij}(\ell)$ from all UTs to all BSs based on the observations

$$
r_{ij}(\ell) = \sqrt{\frac{\tau P}{L}} h_{ij}(\ell) + s_{ij}(\ell) \tag{3.8}
$$

where $s_{ij}(\ell) \sim \mathcal{CN}(0, 1 + \sigma_i^2)$ captures the effects of the thermal noise at the BS-antennas and the quantization error on the backhaul links. For details on how the scalar estimation channel (3.8) is obtained, we refer the reader to [117]. It becomes clear from the last equation that the quantization noise degrades the channel estimate. Thus, the backhaul capacity C has a significant influence on the optimal training length τ^*. This point will be further discussed in Section 3.1.4. Computing the MMSE estimate of $h_{ij}(\ell)$ given the observation $r_{ij}(\ell)$, we can decompose $h_{ij}(\ell)$ into the estimate $\hat{h}_{ij}(\ell)$ and the independent estimation error $\tilde{h}_{ij}(\ell)$, such that

$$
h_{ij}(\ell) = \hat{h}_{ij}(\ell) + \tilde{h}_{ij}(\ell). \tag{3.9}
$$

The variance of the estimated channel $\hat{v}_{ij}(\tau)$ and the variance of the estimation error $\tilde{v}_{ij}(\tau)$ are respectively given as

$$\hat{v}_{ij}(\tau) \triangleq \mathbb{E}\left[|\hat{h}_{ij}(\ell)|^2\right] = \frac{\tau\frac{P}{L}v_{ij}^2}{\tau\frac{P}{L}v_{ij}+1+\sigma_i^2} \qquad \forall \ell \tag{3.10}$$

$$\tilde{v}_{ij}(\tau) \triangleq \mathbb{E}\left[|\tilde{h}_{ij}(\ell)|^2\right] = \frac{v_{ij}(1+\sigma_i^2)}{\tau\frac{P}{L}v_{ij}+1+\sigma_i^2} \qquad \forall \ell\ . \tag{3.11}$$

Denote $\hat{\mathbf{V}}(\tau)$ and $\tilde{\mathbf{V}}(\tau)$ the variance profiles of the estimated channel $\hat{\mathbf{H}}(\ell)$ and the estimation error $\tilde{\mathbf{H}}(\ell)$, respectively. One can easily verify that the total energy of the channel is conserved since

$$\mathbf{V} = \hat{\mathbf{V}}(\tau) + \tilde{\mathbf{V}}(\tau). \tag{3.12}$$

Data Transmission

In each channel coherence block, the UTs broadcast their data simultaneously during $T-\tau$ channel uses. The CS jointly decodes the messages from all UTs, leveraging the previously computed channel estimate $\hat{\mathbf{H}}(\ell)$. With the knowledge of $\hat{\mathbf{H}}(\ell)$, the CS "sees" in its received signal $\mathbf{y}'(\ell) = [y_1'(\ell), \ldots, y_{BM}'(\ell)]^{\mathsf{T}}$ the useful term $\hat{\mathbf{H}}(\ell)\mathbf{x}(l)$ and the overall noise term $\mathbf{z}(\ell) = \tilde{\mathbf{H}}(\ell)\mathbf{x}(\ell) + \mathbf{n}(\ell) + \mathbf{q}(\ell)$, i.e.,

$$\mathbf{y}'(\ell) = \hat{\mathbf{H}}(\ell)\mathbf{x}(\ell) + \mathbf{z}(\ell) \tag{3.13}$$

where the quantization noise vector $\mathbf{q} = [q_1(\ell), \ldots, q_{BM}(\ell)]^{\mathsf{T}}$ is defined by (3.5). Since the statistical distributions of all sub-carriers, signals and noise are i.i.d. with respect to the index ℓ, we will hereafter omit the dependence on ℓ and consider a single isolated sub-carrier.

3.1.3 Net Ergodic Achievable Rate

The capacity of the channel (3.13) is not explicitly known. We consider therefore a lower bound of the normalized ergodic mutual information $\frac{1}{BM}I(\mathbf{y}';\mathbf{x}|\hat{\mathbf{H}})$, referred to hereafter as the *ergodic achievable rate* $R(\tau)$. This lower bound is in essence obtained by overestimating the detrimental effect of the estimation error, treating the total noise term $\mathbf{z}$ as independent complex Gaussian noise with covariance matrix $\mathbf{K}_z(\tau) \in (\mathbb{R}^+)^{BM\times BM}$, given as

$$\mathbf{K}_z(\tau) = \mathbb{E}\left[\mathbf{z}\mathbf{z}^{\mathsf{H}}\right] = \mathrm{diag}\left(1+\sigma_i^2 + \frac{P}{L}\sum_{j=1}^{K}\tilde{v}_{ij}(\tau),\ i=1,\ldots,BM\right). \tag{3.14}$$

Thus, the ergodic achievable rate can be written as [128, 117]

$$R(\tau) = \frac{1}{BM}\mathbb{E}_{\hat{\mathbf{H}}}\left[\log\left|\mathbf{I}_{BM} + \frac{P}{L}\overline{\mathbf{H}}(\tau)\overline{\mathbf{H}}(\tau)^{\mathsf{H}}\right|\right] \qquad \text{[nats/channel use]} \tag{3.15}$$

where we have defined the effective channel $\overline{\mathbf{H}}(\tau)$ as

$$\overline{\mathbf{H}}(\tau) = \mathbf{K}_z^{-\frac{1}{2}}(\tau)\hat{\mathbf{H}}. \tag{3.16}$$

Note that the ergodic achievable rate does not account for the fact that only a fraction $(1-\tau/T)$ of the total coherence block length can be used for data transmission. Our goal is thus to find the optimal training length τ^*, maximizing the *net ergodic achievable rate*

$$R_{\rm net}(\tau) \triangleq \left(1-\frac{\tau}{T}\right) R(\tau). \tag{3.17}$$

Here, the difficulty consists in computing the ergodic achievable rate $R(\tau)$ explicitly. Since a closed-form expression of $R(\tau)$ for finite dimensions of the channel matrix $\mathbf{H}$ seems intractable, we resort to an approximation based on the theory of large random matrices. We will demonstrate shortly that this approximation, although only asymptotically tight, yields very close approximations for even small values of B, M, K and L.

Deterministic Equivalent

In this section, we present a deterministic equivalent approximation $\overline{R}(\tau)$ of $R(\tau)$ in the large system limit, i.e., for $K, BM, L \to \infty$ at the same speed. Denote $N = BM$ the product of the number of BSs and the number of antennas per BS. The notation $K \to \infty$ will refer in the sequel to the following two conditions on K, N and L:

$$0 < \liminf_{K\to\infty} \frac{N}{K} \leq \limsup_{K\to\infty} \frac{N}{K} < \infty\ , \quad 0 < \liminf_{K\to\infty} \frac{L}{K} \leq \limsup_{K\to\infty} \frac{L}{K} < \infty. \tag{3.18}$$

Define $\overline{\mathbf{V}}(\tau) = \mathbf{K}_z^{-1}(\tau)\hat{\mathbf{V}}(\tau)$ the variance profile of the effective channel $\overline{\mathbf{H}}(\tau)$ with elements

$$\overline{v}_{ij}(\tau) = \frac{\hat{v}_{ij}(\tau)}{1+\sigma_i^2+\frac{P}{L}\sum_{\ell=1}^K \tilde{v}_{i\ell}(\tau)} \tag{3.19}$$

and consider the following $N \times N$ matrices

$$\mathbf{D}_j(\tau) = \operatorname{diag}\left(\overline{v}_{1j}(\tau), \dots, \overline{v}_{Nj}(\tau)\right), \qquad j = 1, \dots, K. \tag{3.20}$$

We are now in position to state the deterministic approximation $\overline{R}(\tau)$ of $R(\tau)$ based on a direct application of Theorem 12 (*ii*) to our channel model (see also Theorem 14 (*iv*) and [51, Theorem 2.3]).

Theorem 25 (Deterministic equivalent of the ergodic achievable rate)**.** *Let $\tau > 0$. Assume that K, N and L satisfy* (3.18) *and $0 \leq \overline{v}_{ij}(\tau) < v_{max} < \infty\, \forall i, j$. Then:*

(i) The following implicit equation:

$$\mathbf{T}(z) = \left(\frac{1}{K}\sum_{j=1}^K \frac{\mathbf{D}_j(\tau)}{1+\frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}(z)} - z\mathbf{I}_N\right)^{-1} \tag{3.21}$$

admits a unique solution $\mathbf{T}(z) = \operatorname{diag}\left(t_1(z), \dots, t_N(z)\right)$ such that $(t_1(z), \dots, t_N(z)) \in \mathcal{S}(\mathbb{R}^+)^N$.

(ii) Let $P > 0$. Denote $\mathbf{T}_P = \mathbf{T}(-\frac{L}{KP})$ and consider the quantity:

$$\overline{R}(\tau) = \frac{1}{N}\sum_{j=1}^{K}\log\left(1+\frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P\right) - \frac{1}{N}\log\det\left(\frac{L}{KP}\mathbf{T}_P\right) - \frac{1}{N}\sum_{j=1}^{K}\frac{\frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P}{1+\frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P}. \tag{3.22}$$

Then, the following holds true:

$$R(\tau) - \overline{R}(\tau) \xrightarrow[K\to\infty]{} 0.$$

Optimization of the training length τ

In this section, we consider the optimization of the training length τ with the goal of maximizing the net ergodic achievable rate $R_{\text{net}}(\tau)$. In order to find the optimal training length τ^* for a given coherence block length T, we wish to solve the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & R_{\text{net}}(\tau) \triangleq \left(1-\frac{\tau}{T}\right)R(\tau) \\ \text{subject to} \quad & K \le \tau \le T. \end{aligned} \tag{3.23}$$

As this optimization problem is intractable for finite dimensions, we pursue the following approach:

1. We find $\overline{\tau}^*$ maximizing $\overline{R}_{\text{net}}(\tau) = \left(1-\frac{\tau}{T}\right)\overline{R}(\tau)$.
2. We show that $R_{\text{net}}(\tau^*) - \overline{R}_{\text{net}}(\overline{\tau}^*) \xrightarrow[K\to\infty]{} 0$ and $\tau^* - \overline{\tau}^* \xrightarrow[K\to\infty]{} 0$.
3. We verify by simulations that $\overline{\tau}^*$ is very close to τ^* for even small values of K, N and L.

We start by establishing the concavity of $\overline{R}_{\text{net}}(\tau)$, our new objective function. Denote[5]

$$\overline{v}'_{ij}(\tau) = \frac{\hat{v}'_{ij}(\tau)\left[1+\sigma_i^2+\frac{P}{L}\sum_{j=1}^{K}\tilde{v}_{ij}(\tau)\right] - \hat{v}_{ij}(\tau)\frac{P}{L}\sum_{j=1}^{K}\tilde{v}'_{ij}(\tau)}{\left[1+\sigma_i^2+\frac{P}{L}\sum_{j=1}^{K}\tilde{v}_{ij}(\tau)\right]^2} \tag{3.24}$$

where

$$\hat{v}'_{ij}(\tau) = -\tilde{v}'_{ij}(\tau) = \frac{\frac{P}{L}v_{ij}^2\left(1+\sigma_i^2\right)}{\left(1+\sigma_i^2+\tau\frac{P}{L}v_{ij}\right)^2} \tag{3.25}$$

and define the matrices

$$\mathbf{D}'_j(\tau) = \operatorname{diag}\left(\overline{v}'_{1j}(\tau),\ldots,\overline{v}'_{Nj}(\tau)\right), \quad j = 1,\ldots,K\ . \tag{3.26}$$

[5]We use $f'(x)$ to denote the first derivative of the function $f(x)$, i.e., $f'(x) = \frac{d\,f(x)}{d\,x}$.

A simple composition rule [132, Exercise 3.32 (b)] states that the product of a positive decreasing linear function and a positive increasing concave function is also concave. In order to prove the concavity of $\overline{R}_{\text{net}}(\tau) = (1 - \frac{\tau}{T})\overline{R}(\tau)$, it is thus sufficient to show that $\overline{R}(\tau)$ is an increasing concave function in τ. A sufficient condition for concavity is $\overline{R}''(\tau) \leq 0$. We begin by considering the first derivative $\overline{R}'(\tau)$, which allows for a simple concise closed-from expression as provided by the next theorem:

Theorem 26 (Derivative)**.** *Under the same conditions as for Theorem 25, the first derivative of $\overline{R}(\tau)$ permits the explicit expression*

$$\overline{R}'(\tau) = \frac{1}{N}\sum_{j=1}^{K} \frac{\frac{1}{K}\operatorname{tr}\mathbf{D}_j'(\tau)\mathbf{T}_P}{1 + \frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P}$$

where $\mathbf{T}_P = \mathbf{T}(-\frac{L}{KP})$ is given by Theorem 25 (i). Moreover, for any $P, \tau > 0$, $\overline{R}(\tau)$ is an increasing function, i.e.,

$$\overline{R}'(\tau) > 0.$$

Proof. See Appendix 3.7.1. □

Despite the simplicity of the expression of $\overline{R}'(\tau)$ in Theorem 26, it seems intractable to show that $\overline{R}''_{\text{net}}(\tau) \leq 0$ for channel matrices with a general variance profile. This is due to the fact that not only $\mathbf{D}_j(\tau)$ depends on τ, but also $\mathbf{T}_P$. The matrix $\mathbf{T}_P$ is in general given as the solution of an implicit equation which can only be determined numerically, e.g., by a fixed-point algorithm. It is thus difficult to infer the behavior of $\mathbf{T}_P$ with respect to τ. However, one can show for the particular case of a doubly regular variance profile that $\overline{R}(\tau)$ is indeed concave.

Theorem 27 (Concavity)**.** *Let $P, \tau > 0$. Assume that $N = K$ and that $\overline{\mathbf{V}}(\tau)$ is a doubly regular matrix which satisfies the following regularity condition:*

$$\mathcal{K}(\tau) = \frac{1}{N}\sum_{i=1}^{N} \overline{v}_{ik}(\tau) = \frac{1}{N}\sum_{j=1}^{N} \overline{v}_{\ell j}(\tau) \quad \forall k, \ell\ . \tag{3.27}$$

Then, $\overline{R}(\tau)$ is a strictly concave function.

Proof. See Appendix 3.7.2. □

Remark 15. *Based on our simulation results, we conjecture that Theorem 27 also holds for non doubly regular variance profiles $\overline{\mathbf{V}}(\tau)$. Intuitively, $\overline{R}(\tau)$ being a concave function means nothing else than that channel training shows diminishing returns. That is, the marginal benefit of each training symbol decreases until the channel estimation becomes nearly perfect. The previous argument can be made clear considering the two extreme cases $\tau = 0$ and $\tau \to \infty$. One can easily verify that $\mathbf{D}_j(0) = \mathbf{0}$ while $\mathbf{D}_j'(0) > \mathbf{0}$. This implies $\overline{R}'(0) > 0$, i.e., channel training increases the ergodic achievable rate. On the other hand, for $\tau \to \infty$, $\mathbf{D}_j'(\tau) \to \mathbf{0}$, so that also $\overline{R}'(\tau) \to 0$, i.e., the marginal benefit of channel training vanishes. It is thus justified to conjecture that $\overline{R}'(\tau)$ is a decreasing function of τ and hence $\overline{R}(\tau)$ a concave function.*

As a consequence of Theorem 27 and Remark 15, we assume that $\overline{R}_{\rm net}(\tau)$ takes its global maximum in $(0, T]$ and the optimal training length $\overline{\tau}^*$ can be determined as the solution of

$$\overline{R}'_{\rm net}(\tau) = \left(1 - \frac{\tau}{T}\right)\overline{R}'(\tau) - \frac{1}{T}\overline{R}(\tau) = 0 \ . \tag{3.28}$$

The value $\overline{\tau}^*$ can now be easily found, e.g., via the bisection method. It remains to show that the optimal training length $\overline{\tau}^*$ which maximizes $\overline{R}_{\rm net}(\tau)$ is asymptotically optimal for the original objective function $R_{\rm net}(\tau)$. This is done in the next Theorem.

Theorem 28. *Let* $\tau^* = \arg\max_{\tau\in[0,T]} R_{net}(\tau)$ *and* $\overline{\tau}^* = \arg\max_{\tau\in[0,T]} \overline{R}_{net}(\tau)$. *Then, under the same conditions as for Theorem 25, the following holds true:*

(i)

$$R_{net}(\tau^*) - \overline{R}_{net}(\overline{\tau}^*) \xrightarrow[K\to\infty]{} 0.$$

(ii) Further assume that $\overline{\mathbf{V}}(\tau)$ *is a doubly regular matrix which satisfies the conditions of Theorem 27. Then,*

$$\tau^* - \overline{\tau}^* \xrightarrow[K\to\infty]{} 0$$

where $\overline{\tau}^*$ *is given as the solution to*

$$\overline{R}'_{net}(\tau) = \left(1 - \frac{\tau}{T}\right)\overline{R}'(\tau) - \frac{1}{T}\overline{R}(\tau) = 0$$

with $\overline{R}(\tau)$ *and* $\overline{R}'(\tau)$ *given by Theorem 25 (ii) and Theorem 26, respectively.*

Proof. See Appendix 3.7.3. □

Theorem 28 (i) merely states that the maximum point of $R_{\rm net}(\tau)$ can be arbitrarily closely approximated by the maximum point of $\overline{R}_{\rm net}(\tau)$. This result is independent of the structure of the variance profile $\overline{\mathbf{V}}(\tau)$. Theorem 28 (ii) provides a simple way to compute $\overline{\tau}^*$ and states that this value is also asymptotically optimal for $\overline{R}_{\rm net}(\tau)$. However, this result requires $\overline{\mathbf{V}}(\tau)$ to be a doubly regular matrix. Both results together imply that optimizing $\overline{R}_{\rm net}(\tau)$ is asymptotically identical to optimizing $\overline{R}_{\rm net}(\tau)$. We show in the next section via simulations that Theorem 27 and Theorem 28 also hold for non doubly regular variance profiles.

3.1.4 Numerical Results

In order to show the validity of our analysis in the preceding sections, we consider a simple cellular system consisting of $B = 3$ BSs with $M = 2$ antennas and $K = 3$ UTs, as shown in Fig. 3.2. The locations of the UTs are randomly chosen according to a uniform distribution. The inverse path loss factor a_{bk} between UT k and BS b is given as $a_{bk} = d_{bk}^{-3.6}$, where d_{bk} is the distance between UT k and BS b, normalized to the maximum distance within a cell.

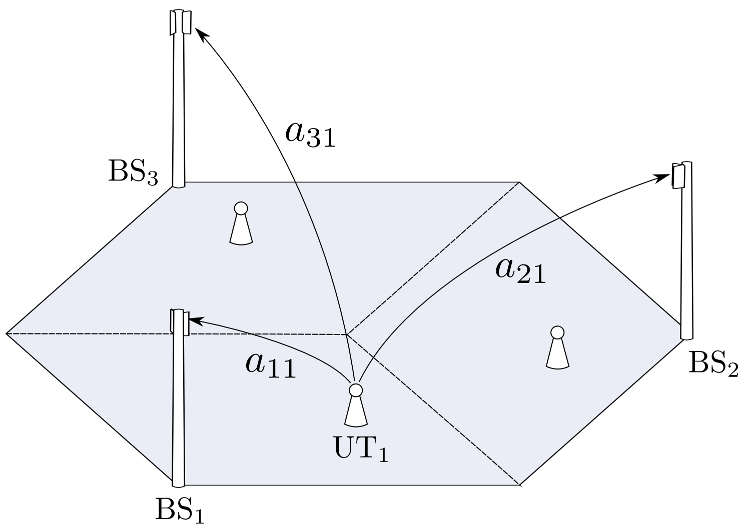

Figure 3.2: Cellular example with $B = 3$ BSs and $K = 3$ UTs.

We consider one random snapshot of user distributions, resulting in the inverse path loss matrix

$$\mathbf{A} = \begin{pmatrix} 2.9775 & 0.0385 & 1.6055 \\ 0.2512 & 2.7826 & 0.1759 \\ 0.0615 & 0.0492 & 1.6376 \end{pmatrix}. \tag{3.29}$$

In the sequel, we assume $\mathbf{A}$ fixed while we average over many independent realizations of the channel matrix $\mathbf{H}$. The cell edge signal-to-noise-ratio is defined as $\text{SNR} = \mathbb{E}\left[|x_i(\ell)|^2\right]/\mathbb{E}\left[|n_i(\ell)|^2\right] = P/L$. Unless otherwise stated, we assume $T = 1000$ and $L = 1$.

Fig. 3.3 depicts the net ergodic achievable rate $R_{\text{net}}(\tau)$ and its deterministic equivalent approximation $\overline{R}_{\text{net}}(\tau)$ by Theorem 25 (ii) as a function of the SNR for a fixed training length of $\tau = 40$ and different values of the backhaul capacity $C = \{1, 5, 10\}$ bits/channel use. Clearly, $\overline{R}_{\text{net}}(\tau)$ gives a very tight approximation of $R_{\text{net}}(\tau)$ over the full range of SNR. The effect of limited backhaul is particularly visible at high SNR where all curves saturate.

For the same set of parameters and $\text{SNR} = 0\,\text{dB}$, we show in Fig. 3.4 $R_{\text{net}}(\tau)$ and $\overline{R}_{\text{net}}(\tau)$ as a function of the training length τ. This plot validates Theorem 27 and the corresponding remark as $\overline{R}_{\text{net}}(\tau)$ is obviously a concave function. Moreover, since the curves of $\overline{R}_{\text{net}}(\tau)$ and $R_{\text{net}}(\tau)$ match very closely, it is reasonable to assume that both take a similar maximum value at a similar value of τ. The validity of Theorem 28 is demonstrated in Fig. 3.5 which shows the optimal training length τ^*, found by an exhaustive search based on Monte Carlo simulations, and the training length $\overline{\tau}^*$ which maximizes $\overline{R}_{\text{net}}(\tau)$ as a function of the SNR for $C = 1$ bits/channel use and $T = 100$. The differences between both values, although very small, are mainly due to the exhaustive search over a necessarily discrete set of values of τ.

Fig. 3.6 shows the dependence of the optimal training length $\overline{\tau}^*$ on the backhaul capacity C for a fixed $\text{SNR} = 10\,\text{dB}$. One can see that $\overline{\tau}^*$ is a decreasing

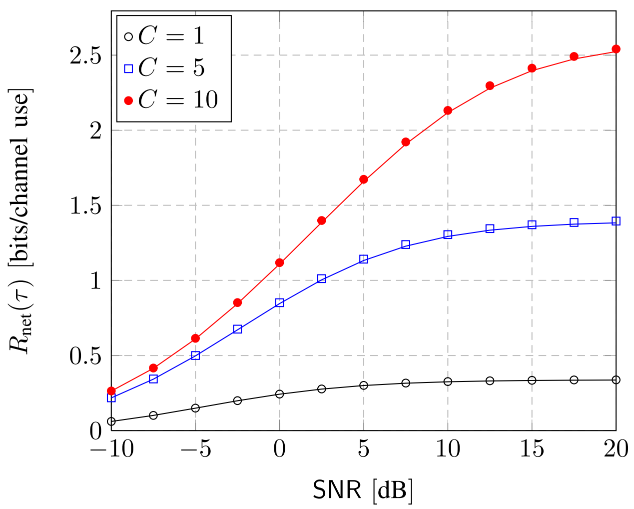

Figure 3.3: Net ergodic achievable rate $R_{\text{net}}(\tau)$ vs SNR for $\tau = 40$ and $T = 1000$. The markers are obtained by simulations, the solid lines correspond to the deterministic equivalent $\overline{R}_{\text{net}}(\tau)$.

function of C which converges quickly to particular value corresponding to infinite capacity backhaul links. The reason for this is the following. The CS estimates the channel coefficients based on the quantized training signals received by the BSs. The channel estimate is hence impaired by thermal noise and quantization errors. Therefore, increasing C results in better channel estimates and reduces the necessary training length. For infinite backhaul capacity, the optimal training length is only dependent on the SNR. In a similar flavor, Fig. 3.7 depicts $R_{\text{net}}(\overline{\tau}^*)$ as a function of the backhaul capacity C. We notice the inefficient utilization of the backhaul links due to sub-optimal compression since the net ergodic achievable rate per BS, i.e., $M \times R_{\text{net}}(\overline{\tau}^*)$, is much lower than the necessary backhaul capacity. For example, it takes $C = 20$ bits/channel use of backhaul capacity to achieve a rate per BS of $2 \times R_{\text{net}}(\overline{\tau}^*) \approx 5.2$ bits/channel use.

3.1.5 Conclusions

We have considered a frequency-selective fading network MIMO uplink channel with arbitrary path losses between the UTs and BSs and finite capacity backhaul links. Using a close approximation of the net ergodic achievable rate based on random matrix theory, we have studied the optimal trade-off between the resources used for channel training and data transmission. Although the asymptotic results are proved to be tight only in the large system limit, our numerical examples show that they provide close approximations even for small system dimensions. Our results also show that limited backhaul capacity has a significant impact on the optimal training length. We wish to conclude this section by pointing out some shortcomings of our system model which remain as future investigations:

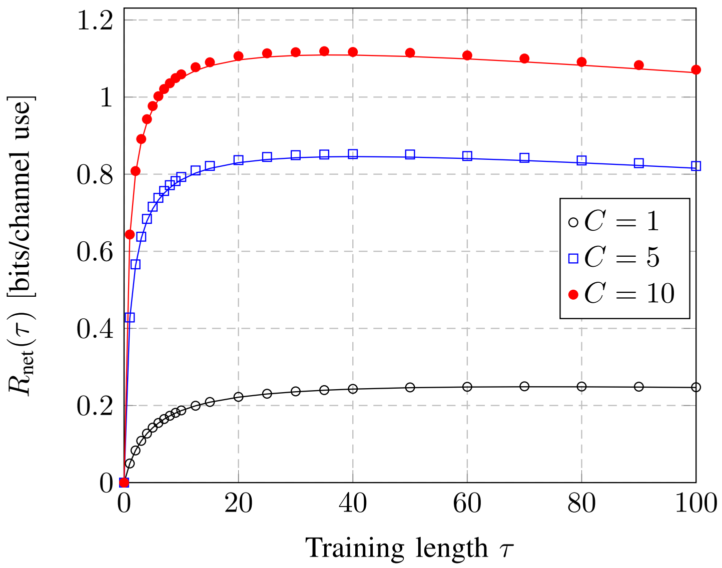

Figure 3.4: Net ergodic achievable rate $R_{\text{net}}(\tau)$ vs training length τ for SNR = 0 dB and $T = 1000$. The markers are obtained by simulations, the solid lines correspond to the deterministic equivalent $\overline{R}_{\text{net}}(\tau)$.

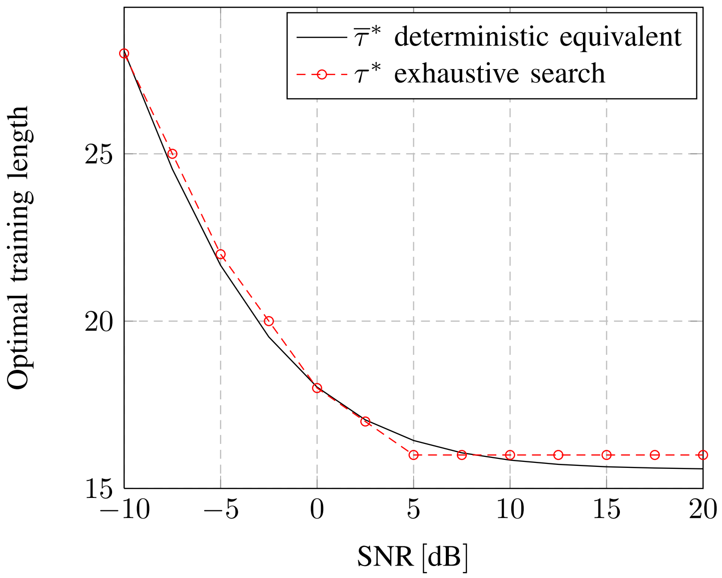

Figure 3.5: Optimal training length τ^* and $\overline{\tau}^*$ vs SNR for $C = 1$ bits/channel use and $T = 100$. The solid line corresponds to $\overline{\tau}^*$ maximizing $\overline{R}_{\text{net}}(\tau)$, the dashed line corresponds to τ^* maximizing $R_{\text{net}}(\tau)$ and is obtained by an exhaustive search based on Monte Carlo simulations.

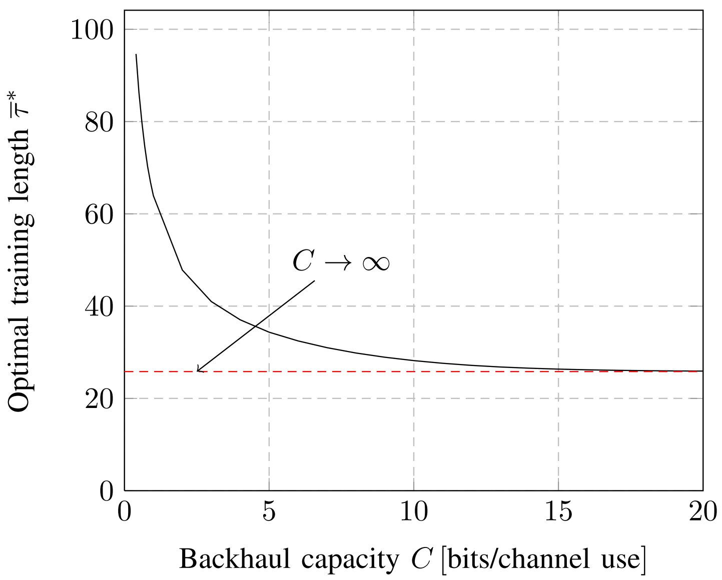

Figure 3.6: Optimal training length $\overline{\tau}^*$ vs backhaul capacity C for SNR $= 10\,\mathrm{dB}$ and $T = 1000$.

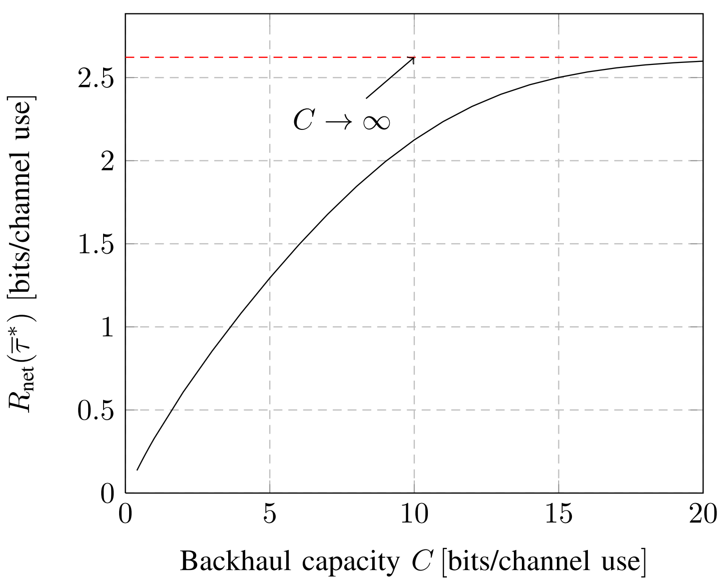

Figure 3.7: Net ergodic achievable rate $R_{\mathrm{net}}(\overline{\tau}^*)$ with optimal channel training $\overline{\tau}^*$ vs backhaul capacity C for SNR $= 10\,\mathrm{dB}$ and $T = 1000$.

Backhaul links and cooperation

A relevant question is how a BS should decide whether to cooperate by forwarding its received data to some central processor or to process its received signals alone. In our model, the net throughput vanishes with a decreasing backhaul capacity although each BSs could theoretically decode a part of the received messages alone. Future work, also motivated by the recent results in [125, 133], comprises the investigation of flexible schemes which adapt the degree of cooperation according to some statistical side-information about the channels, backhaul limitations, quality of CSI, etc.

Inter-cluster interference

We have considered a multi-cell network composed of B cooperative cells without inter-cell interference. In a real system, also the effects of non-orthogonal training sequences leading to "pilot contamination" [41, 16] constitute an important issue for practical system design. Very recently, both aspects have been addressed in [134].

3.2 Polynomial expansion detectors

Abstract: *We consider a certain class of large random matrices, composed of independent column vectors with zero mean and different covariance matrices. This random matrix model arises in several wireless communication systems of recent interest, such as distributed antenna or large-scale MIMO systems. Computing the linear MMSE detector in such systems requires the inversion of a large covariance matrix which becomes prohibitively complex as the number of antennas and users grows. We apply the asymptotic moment result of Theorems 19 and 20 to the design of a low-complexity polynomial expansion detector which approximates the matrix inverse by a matrix polynomial and study its asymptotic performance. Simulation results corroborate the analysis and evaluate the performance for finite system dimensions.*

3.2.1 Introduction

Distributed antenna systems and large antenna arrays have recently attained significant research interest [27, 16]. Both are considered as promising solutions to counter intercell interference and to increase the spectral efficiency of current cellular networks. Since these techniques rely in essence on a significant increase of the number of *coordinated antennas*, the computational complexity of the joint precoding/detection of the transmitted/received signals grows. This calls for low-complexity solutions. In this section, we address this need by assessing the performance of a polynomial expansion detector [135] adapted to the following general channel model.

Consider a discrete-time $N \times K$ MIMO channel with output vector $\mathbf{y} \in \mathbb{C}^N$:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3.30}$$

where $\mathbf{x} = [x_1, \ldots, x_K]^\mathsf{T}$ is the complex channel input vector satisfying $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\mathsf{H}\right] = \mathbf{I}_K$, $\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_K] \in \mathbb{C}^{N \times K}$ is the random channel matrix and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_N)$

is a vector of additive noise. The jth column $\mathbf{h}_j \in \mathbb{C}^N$ of $\mathbf{H}$ is modeled as

$$\mathbf{h}_j = \frac{1}{\sqrt{N}}\mathbf{R}_j\mathbf{g}_j, \quad j = 1,\ldots,K \tag{3.31}$$

where $\mathbf{R}_j \in \mathbb{C}^{N\times N}$ is a deterministic matrix and the elements of $\mathbf{g}_j \in \mathbb{C}^N$ are i.i.d. random variables with zero mean, unit variance and finite eighth order moment. This channel model captures different types of wireless communication systems and generalizes several well-known channel models as discussed below:

Distributed Antenna Systems: Let $\mathbf{R}_j = \operatorname{diag}(r_{1j},\ldots,r_{Nj})$ with elements $r_{ij} = \sqrt{p_j}/d_{ij}^{\beta/2}$, where d_{ij} is the (normalized) distance between transmitter j and receive antenna i, β is the path loss exponent and p_j is the transmit power of transmitter j. This model is suitable for distributed antenna systems [27] where each transmitter sees a different path loss to each of the receive antennas since $d_{1j},\ldots,d_{Nj}$ are different.

Large-scale MIMO: Assume a receiver equipped with a very large antenna array ($N \gg 1$) as in [16]. Unless the antenna spacing is sufficiently large, it is likely that the received signals at different receive antennas are correlated. Our model allows to assign a different correlation matrix $\mathbf{R}_j$ to each transmitter.

MIMO Multiple Access Channel (MAC): Consider a MIMO MAC from M transmitters equipped with K_m, $m = 1,\ldots,M$, antennas to a receiver with N antennas. Each point-to-point link has a different transmit and receive correlation matrix [83]:

$$\mathbf{y} = \sum_{m=1}^{M} \mathbf{\Phi}_{\mathrm{R},m}^{\frac{1}{2}}\mathbf{G}_m\mathbf{\Phi}_{\mathrm{T},m}^{\frac{1}{2}}\mathbf{x}_m + \mathbf{n}$$

where $\mathbf{\Phi}_{\mathrm{R},1},\ldots,\mathbf{\Phi}_{\mathrm{R},M} \in \mathbb{C}^{N\times N}$ are deterministic correlation matrices, $\mathbf{\Phi}_{\mathrm{T},1} \in \mathbb{C}^{K_1\times K_1},\ldots,\mathbf{\Phi}_{\mathrm{T},M} \in \mathbb{C}^{K_M\times K_M}$ are nonnegative definite diagonal matrices, $\mathbf{G}_1 \in \mathbb{C}^{N\times K_1},\ldots,\mathbf{G}_M \in \mathbb{C}^{N\times K_M}$ are random channel matrices with i.i.d. entries with zero mean and variance $1/N$, and $\mathbf{x}_1 \in \mathbb{C}^{K_1},\ldots,\mathbf{x}_M \in \mathbb{C}^{K_M}$ are the transmit vectors. Let $\sum_{m=1}^{M} K_m = K$. Setting $\mathbf{R}_j = \mathbf{\Phi}_{\mathrm{R},m}^{1/2}[\mathbf{\Phi}_{\mathrm{T},m}^{1/2}]_{ii}$ for $j \in \{1+\sum_{l=1}^{m-1} K_l,\ldots,\sum_{l=1}^{m} K_l\}$ and $i = j - \sum_{l=1}^{m-1} K_l$, we fall back to the model in (3.31).

In the sequel, we will study the asymptotic behavior of the moments M_k of the matrix $\mathbf{B}_N \triangleq \mathbf{H}\mathbf{H}^{\mathsf{H}}$, defined as

$$M_k \triangleq \frac{1}{N}\operatorname{tr}\mathbf{B}_N^k, \qquad k = 0,1,2,\ldots \tag{3.32}$$

under the assumption that N and K grow infinitely large at the same speed. In particular, we will derive deterministic approximations $\bar{M}_k$ of M_k , such that $M_k - \bar{M}_k \to 0$ almost surely, for $N,K \to \infty$. This result can be used, for example, to compute low-complexity approximations of the matrix inverse $(\mathbf{B}_N + \sigma^2\mathbf{I}_N)^{-1}$. The computation of this matrix arises in many practical applications, such as for linear multiuser detectors and beamforming strategies. We will focus exemplary on the linear MMSE (LMMSE) detector.

The LMMSE estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$, assuming perfect knowledge of $\mathbf{H}$ at the receiver, is given as [136]

$$\hat{\mathbf{x}} = \mathbf{H}^{\mathsf{H}}(\mathbf{B}_N + \sigma^2\mathbf{I}_N)^{-1}\mathbf{y}. \tag{3.33}$$

The computational complexity of this estimate is of order $\mathcal{O}(r^2)$ [137], where $r = \min(N, K)$. A reduced complexity estimate can be obtained by approximating the matrix inverse in (3.33) by the following matrix polynomial [135]

$$(\mathbf{B}_N + \sigma^2 \mathbf{I}_N)^{-1} \approx \sum_{l=0}^{L-1} w_l \mathbf{B}_N^l \tag{3.34}$$

for some coefficients w_l, where the filter rank $L \leq r$ is chosen according to the allowable complexity. For a given transmitter k, the above *polynomial expansion detector* can be seen as a projection of $\mathbf{y}$ on the Lth Krylov subspace associated to the pair $(\mathbf{B}_N, \mathbf{h}_k)$, i.e., the subspace of $\mathbb{C}^N$ spanned by the vectors $\{\mathbf{h}_k, \mathbf{B}_N \mathbf{h}_k, \ldots, \mathbf{B}_N^{L-1} \mathbf{h}_k\}$, and a weighting of the joint projections by the coefficients w_l. Depending on L, the polynomial expansion detector achieves a performance between the matched filter ($L = 1$) and the LMMSE detector ($L = r$) [135] and allows, consequently, to trade-off performance for complexity. Moreover, (3.34) allows for an efficient multistage implementation [135, 138, 137], where each stage l consists of a matched filter $\mathbf{H}^\mathsf{H}$ and subsequent "re-spreading" by the matrix $\mathbf{H}$. In [139], it was shown that the SINR at the filter output converges in certain cases exponentially in the filter rank L to the SINR output of the LMMSE detector. Thus, L does not need to scale with the system size to achieve close to optimal performance [140].

The optimal weight vector $\mathbf{w} = [w_0 \cdots w_{L-1}]^\mathsf{T}$ can be chosen to minimize the mean square error of the estimated vector $\hat{\mathbf{x}}$, i.e.,

$$\mathbf{w} = \arg \min_{\mathbf{u}=[u_0,\ldots,u_{L-1}]^\mathsf{T}} \mathbb{E}\left[\left\| \mathbf{x} - \mathbf{H}^\mathsf{H} \sum_{l=0}^{L-1} u_l \mathbf{B}_N^l \mathbf{y} \right\|_2^2\right]. \tag{3.35}$$

The solution to this optimization problem is given as [135]

$$\mathbf{w} = \mathbf{\Phi}^{-1} \boldsymbol{\varphi} \tag{3.36}$$

where $\mathbf{\Phi} \in (\mathbb{R}^+)^{L \times L}$ and $\boldsymbol{\varphi} \in (\mathbb{R}^+)^L$ are defined as

$$\begin{aligned} [\mathbf{\Phi}]_{ij} &= M_{i+j} + \sigma^2 M_{i+j-1} \\ [\boldsymbol{\varphi}]_i &= M_i. \end{aligned} \tag{3.37}$$

The computation of the weight vector $\mathbf{w}$ requires the calculation of the moments $M_1, \ldots, M_{2L}$ which is still computational expensive for large L. However, under the assumption that the dimensions of $\mathbf{H}$ grow infinitely large, it was shown for several random matrix models (e.g., [138, 140, 88]) that the moments M_k can be closely approximated by their asymptotic counterparts $\bar{M}_k$. These are independent of a particular realization of $\mathbf{H}$ and can be calculated based on the statistical properties of the channel matrix. If these properties change on a much slower timescale than the fast-fading channel fluctuations, the weight vector $\mathbf{w}$ can be precomputed using $\bar{M}_k$ instead of M_k. Thus, the detector complexity depends only on the complexity of the projection on the Krylov subspace which is of order $\mathcal{O}(r)$ [137].

Multistage or reduced-rank multiuser detectors were mainly considered in the context of code-division multiple access (CDMA) systems as low-complexity solutions to the joint detection of a large number of user terminals with long spreading sequences [135]. The asymptotic (universal) weight design was first studied in [138] for the equal transmit power case and then extended to more involved models, such as different transmit powers [140], multi-path fading [88] and random unitary spreading sequences [141]. These results were then put on a common ground in [137] which compares different types of linear multistage detectors in terms of their complexity and asymptotic performance. Recently, multistage detectors for asynchronous CDMA systems were considered in [142].

The asymptotic results in the above works are based on the almost sure convergence of the e.s.d. of the matrix $\mathbf{B}_N$ to a compactly supported limit distribution. This limit distribution is in general given implicitly by its Stieltjes transform which can be computed based on the statistical properties of the underlying random matrix model. The asymptotic moments are then obtained by writing the Stieltjes transform as a moment generating function (see Theorem 4) and relying on combinatorial arguments [88] or free probability theory [141].

Our technique is different in two aspects. First, we do not require the existence of a limiting eigenvalue distribution of the matrix $\mathbf{B}_N$. Instead, we provide for each pair (N, K) a deterministic approximation $\bar{M}_k$ of the moments M_k which becomes arbitrarily tight as $N, K \to \infty$. Second, the moments are derived through iterated differentiation of the Stieltjes transform and can be computed by simple recursive equations. This is in contrast to [88] which requires an exhaustive search over complicated sets of indices. Hence, our results are more practical from an implementation perspective. Moreover, the asymptotic moments of the random matrix model (3.31) have not been considered in the literature before.

Our main technical results are Theorems 19 and 20 which can be found in Section 2.3.2. Loosely speaking, Theorem 20 states that, for large matrix dimensions, the e.s.d. $F^{\mathbf{B}_N}$ of the matrix $\mathbf{B}_N$ can be closely approximated by a deterministic distribution function F_N. Thus, the optimal weighting vector $\mathbf{w}$ can be approximated by replacing the moments M_k of $F^{\mathbf{B}_N}$ in (3.37) by the moments $\bar{M}_k$ of F_N. Using the result of Theorem 19, we can compute an approximate weight vector $\bar{\mathbf{w}} = [\bar{w}_0 \dots \bar{w}_{L-1}]$ as

$$\bar{\mathbf{w}} = \bar{\mathbf{\Phi}}^{-1} \bar{\boldsymbol{\varphi}} \tag{3.38}$$

where $\bar{\mathbf{\Phi}} \in (\mathbb{R}^+)^{L \times L}$ and $\bar{\boldsymbol{\varphi}} \in (\mathbb{R}^+)^{L}$ are defined by

$$\begin{aligned} \left[\bar{\mathbf{\Phi}}\right]_{ij} &= \bar{M}_{i+j} + \sigma^2 \bar{M}_{i+j-1} \\ \left[\bar{\boldsymbol{\varphi}}\right]_i &= \bar{M}_i. \end{aligned} \tag{3.39}$$

3.2.2 Asymptotic Performance Analysis

We consider now the asymptotic performance of the polynomial expansion receiver in terms of the received SINR γ_m for a given transmitter m. With weight vector $\mathbf{w}$, the mth element $\hat{x}_m$ of the estimated vector $\hat{\mathbf{x}}$ reads

$$\hat{x}_m = \mathbf{h}_m^{\mathsf{H}} \sum_{l=0}^{L-1} w_l \mathbf{B}_N^l \left(\mathbf{H}\mathbf{x} + \mathbf{n}\right). \tag{3.40}$$

One can easily show that the associated SINR γ_m can be expressed as [137, Eq. (18)]

$$\gamma_m = \frac{\mathbf{w}^\mathsf{T} \boldsymbol{\varphi}_m \boldsymbol{\varphi}_m^\mathsf{T} \mathbf{w}}{\mathbf{w}^\mathsf{T} \left(\boldsymbol{\Phi}_m - \boldsymbol{\varphi}_m \boldsymbol{\varphi}_m^\mathsf{T}\right) \mathbf{w}} \tag{3.41}$$

where $\boldsymbol{\Phi}_m \in (\mathbb{R}^+)^{L \times L}$ and $\boldsymbol{\varphi}_m \in (\mathbb{R}^+)^L$ are given as

$$\begin{aligned} [\boldsymbol{\Phi}_m]_{ij} &= \left[\mathbf{B}_N^{i+j}\right]_{mm} + \sigma^2 \left[\mathbf{B}_N^{i+j-1}\right]_{mm} \\ [\boldsymbol{\varphi}_m]_i &= \left[\mathbf{B}_N^{i}\right]_{mm}. \end{aligned} \tag{3.42}$$

The next theorem provides a tight deterministic approximation of the terms $\left[\mathbf{B}_N^k\right]_{mm} = \mathbf{h}_m^\mathsf{H} \mathbf{B}_N^{k-1} \mathbf{h}_m$ in the asymptotic limit.

Theorem 29. *Under the assumptions of Theorem 20, the following convergence holds:*

$$\left[\mathbf{B}_N^k\right]_{mm} - \bar{M}_k^m \xrightarrow{a.s.} 0$$

where

$$\bar{M}_k^m = \sum_{i=0}^{k-1} \bar{M}_{k-i-1}^m \frac{(-1)^i}{i!} \frac{1}{N} \operatorname{tr} \mathbf{R}_m \mathbf{R}_m^\mathsf{H} \mathbf{T}_{N,i}, \qquad k \geq 1$$

and $\mathbf{T}_{N,k}$ is given by Theorem 19. The initial values of the recursion are $\bar{M}_0^m = 1$ and $\mathbf{T}_{N,0} = \mathbf{I}_N$.

Proof of Theorem 29. The proof follows the same steps as [137, Theorem 1] and will not be given here. □

Replacing $\left[\mathbf{B}_N^k\right]_{mm}$ in (3.42) by $\bar{M}_k^m$ and $\mathbf{w}$ in (3.41) by $\bar{\mathbf{w}}$, we can obtain a deterministic approximation of the SINR γ_m at the output of the polynomial expansion receiver.

3.2.3 Numerical Results

Consider a MAC from $K = 40$ single-antenna transmitters to a receiver with $N = 100$ antennas. We use an extended version of Jake's model [83] for the generation of the matrices $\mathbf{R}_j$. Let $\mathbf{R}_j = \boldsymbol{\Theta}_j^{1/2}$ and $\boldsymbol{\Theta}_j \in \mathbb{C}^{N \times N}$ be defined as

$$[\boldsymbol{\Theta}_j]_{kl} = \frac{1}{\phi_{\max}^j - \phi_{\min}^j} \int_{\phi_{\min}^j}^{\phi_{\max}^j} \exp\left(\frac{2\pi \mathbf{i}}{\lambda} d_{kl} \cos(x)\right) dx$$

where $d_{kl} = 2\lambda(k-l)$ and $\phi_{\min}^j, \phi_{\max}^j$ are drawn independently from the intervals $[-\pi, 0]$ and $[0, \pi]$, respectively. The interval $[\phi_{\min}^j, \phi_{\max}^j]$ can be seen as the angular spread of the signal from transmitter j, λ is the wave length, and d_{kl} is the spacing between the receive antennas k and l. We assume Rayleigh fading channels, i.e., $\mathbf{g}_j$ in (3.31) are independent standard complex Gaussian vectors. The covariance matrices $\boldsymbol{\Theta}_j$ are chosen at random at the beginning and then kept fixed while we average over many realizations of the channel matrix $\mathbf{H}$. We denote by $\text{SNR} = 1/\sigma^2$ the transmit SNR .

Fig. 3.8 shows the average received SINR $\mathbb{E}[\gamma_m]$ of a randomly chosen transmitter as a function of the SNR for the matched filter, the LMMSE detector and the polynomial expansion detector with approximate weights for $L = \{2, 3, 6\}$. Markers correspond to simulation results and solid lines to the deterministic SINR approximations. The error bars indicate one standard deviation of γ_k in each direction. Similar to [68], the asymptotic SINR of transmitter m for the LMMSE detector can be easily shown to satisfy

$$\overline{\gamma}_m^{\text{LMMSE}} = \frac{1}{N} \operatorname{tr} \mathbf{R}_m \mathbf{R}_m^{\mathsf{H}} \mathbf{T}_N(-1/\text{SNR})$$

where $\mathbf{T}_N(z)$ is given by Theorem 14 (*ii*). We observe a good fit between the deterministic approximations and the simulation results for the average SINR . However, the standard deviation of the SINR increases with L. This is because the higher order moments converge slower to their deterministic approximations and exhibit therefore stronger fluctuations. Nevertheless, the average SINR performance of the polynomial expansion detector with $L = 6$ is already close to the performance of the LMMSE detector.

Fig. 3.9 depicts the theoretical average bit-error rate (BER) over SNR for the different detectors. Assuming binary phase-shift keying (BPSK) modulation and Gaussian interference, the BER is given as $\mathbb{E}[Q(\sqrt{\gamma_k})]$ where $Q(x)$ is the Gaussian tail function. We can clearly see a performance increase of the polynomial expansion detector with L, although the BER saturates at high SNR. Although not explicitly shown here, one can even observe a performance decrease for large values of L. As mentioned before, this is due to the low accuracy of the approximate weights caused by a slow convergence of the higher-order moments to their deterministic approximations.

3.2.4 Conclusions

We have derived asymptotically tight deterministic approximations of the moments of a certain class of large random matrices, useful for the study of distributed antenna systems and large antenna arrays. We have applied these moment results to the design of a polynomial expansion detector which significantly reduces the computational complexity of multiuser detection compared to the LMMSE detector. Moreover, we have derived an explicit expression of the asymptotic SINR at the output of this detector and verified its accuracy and performance for finite system dimensions by simulations.

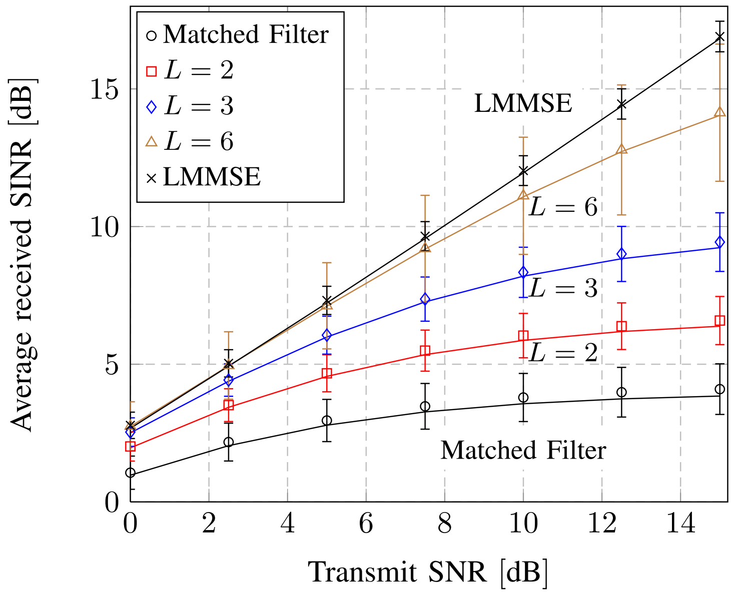

Figure 3.8: Average received SINR versus SNR at the output of the matched filter, LMMSE detector and the polynomial expansion detector with approximate weights for different values of L. Markers correspond to simulation results, solid lines to the deterministic SINR approximations. Error bars indicate one standard deviation of the simulation results in each direction.

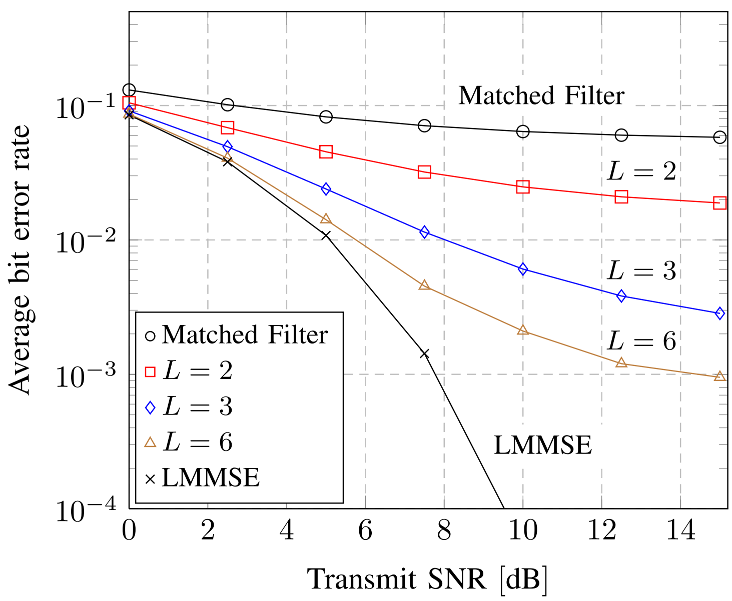

Figure 3.9: Average theoretical bit error rate versus SNR for the matched filter, LMMSE detector and the polynomial expansion detector with approximate weights for different values of L.

3.3 Large-scale MIMO systems

Abstract: *We provide a unified performance analysis of the up- and downlink of noncooperative multi-cellular time-division duplexing (TDD) systems where both the number of antennas per BS N and the number of UTs per cell K are assumed to be large. Our system model accounts for channel estimation, pilot contamination, as well as an arbitrary path loss and antenna correlation for each link. We derive deterministic approximations of achievable rates with several linear precoders and detectors which are asymptotically tight, but accurate for realistic system dimensions, as shown by simulations. It is known from previous work that as $N \to \infty$ while $K/N \to 0$, the system performance is limited by pilot contamination, the simplest precoders/detectors, i.e., BF and MF, are optimal and the transmit power can be made arbitrarily small. We analyze to which extent these conclusions hold if N is large but finite. In particular, we derive how many antennas per UT are needed to achieve η% of the ultimate performance limit and how many antennas more are needed with MF and BF to achieve the performance of MMSE detection and RZF, respectively. Simulations suggest that the use of RZF/MMSE can reduce the number of antennas by one order of magnitude in certain scenarios.*

3.3.1 Introduction

Very large MIMO or "massive MIMO" TDD systems [16, 17] are currently investigated as a novel cellular network architecture with several attractive features: First, the capacity can be increased by simply installing additional antennas to existing cell sites. Thus, massive MIMO provides an alternative to cell-size shrinking, the traditional way of increasing the network capacity [24]. Second, large antenna arrays can potentially reduce uplink and downlink transmit powers through coherent combining and an increased antenna aperture [143]. This aspect is not only relevant from a business point of view but also addresses environmental as well as health concerns related to mobile communications [144, 145]. Third, if channel reciprocity is exploited, the overhead related to channel training scales linearly with the number of UTs per cell K and is independent of the number of antennas per BS N. Consequently, additional antennas do not increase the feedback overhead and therefore "always help" [146]. Fourth, if $N \gg K$, the simplest precoder/detectors are optimal, noise, interference and channel estimation errors vanish, and the only performance limitation is pilot contamination [16], i.e., residual interference caused by the reuse of pilot sequences in adjacent cells.

The features described above are based on several crucial but optimistic assumptions about the propagation conditions, hardware implementations, and the number of antennas which can be deployed in practice. Therefore, recent papers study massive MIMO under more realistic assumptions, e.g., a physical channel model with a finite number of degrees of freedom (DoF) [147] or constant-envelope transmissions with per-antenna power constraints [148].

In this section, we provide a unified performance analysis of the up- and downlink of multicell TDD systems. We consider a realistic system model which accounts for channel estimation, pilot contamination, antenna correlation, and path loss. Assuming that N and K are large, we derive asymptotically tight approximations of achievable rates with several linear precoders/detectors, i.e., BF and RZF in the downlink, MF and MMSE detector in the uplink. These approximations are easy to compute and shown to be accurate for realistic system

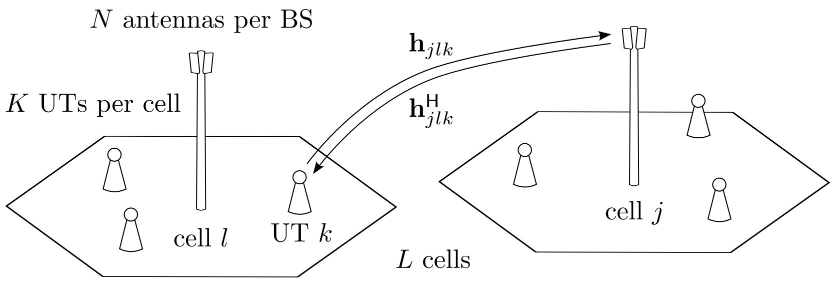

Figure 3.10: System model: In each of the L cells is one BS and K UTs. The BSs are equipped with N antennas. We assume channel reciprocity, i.e., the downlink channel $\mathbf{h}_{jlk}^{\mathsf{H}}$ is the complex conjugate transpose of the uplink channel $\mathbf{h}_{jlk} \in \mathbb{C}^N$.

dimensions. We then distinguish massive MIMO from "classical" MIMO as a particular operating condition of cellular networks where multiuser interference, channel estimation errors and noise are small compared to pilot contamination. Whether this condition is satisfied or not depends on several system parameters, such as the number of UTs per DoF the channel offers (we denote DoF the rank of the antenna correlation matrix which might be smaller than N), the number of antennas per BS, the SNR and the path loss. We then study how many antennas per UTs are needed to achieve η% of the ultimate performance limit and how many antennas more are needed with BF/MF to achieve RZF/MMSE performance. Our simulations suggest that RZF/MMSE can perform as well as BF/MF with one order of magnitude fewer antennas.

3.3.2 System model

Consider a multi-cellular system consisting of $L > 1$ cells with one BS and K UTs in each cell, as schematically shown in Fig. 3.10. The BSs are equipped with N antennas, the UTs have a single antenna. We assume that all BSs and UTs are perfectly synchronized and operate a TDD protocol with universal frequency reuse. We consider transmissions over flat-fading channels on a single frequency band or sub-carrier. Extensions to multiple sub-carriers are straightforward.

Uplink

The received base-band signal vector $\mathbf{y}_j^{\text{ul}} \in \mathbb{C}^N$ at BS j at a given time reads

$$\mathbf{y}_j^{\text{ul}} = \sqrt{\rho_{\text{ul}}} \sum_{l=1}^{L} \mathbf{H}_{jl} \mathbf{x}_l^{\text{ul}} + \mathbf{n}_j^{\text{ul}} \tag{3.43}$$

where $\mathbf{H}_{jl} = [\mathbf{h}_{jl1} \cdots \mathbf{h}_{jlK}] \in \mathbb{C}^{N \times K}$ is the channel matrix from the UTs in cell l to BS j, $\mathbf{x}_l^{\text{ul}} = \left[x_{l1}^{\text{ul}} \cdots x_{lK}^{\text{ul}}\right]^{\mathsf{T}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ is the transmit vector from the UTs in cell l, $\mathbf{n}_j^{\text{ul}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is a noise vector and $\rho_{\text{ul}} > 0$ denotes the uplink SNR. We model $\mathbf{h}_{jlk} \in \mathbb{C}^N$ as

$$\mathbf{h}_{jlk} = \tilde{\mathbf{R}}_{jlk} \mathbf{w}_{jlk} \tag{3.44}$$

where $\mathbf{R}_{jlk} \triangleq \tilde{\mathbf{R}}_{jlk}\tilde{\mathbf{R}}_{jlk}^{\mathsf{H}} \in \mathbb{C}^{N\times N}$ are deterministic and $\mathbf{w}_{jlk} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_N\right)$ are fast fading channel vectors. Our channel model is very versatile as it allows us to assign a different antenna correlation to each channel vector. This is especially important for large antenna arrays with a significant amount of antenna correlation due to either insufficient antenna spacing or a lack of scattering. The channel model is also valid for distributed antenna systems since we can assign a different path loss to each antenna. Moreover, (3.44) can represent a physical channel model with a fixed number of dimensions or angular bins P as in [147], by letting $\tilde{\mathbf{R}}_{jlk} = \sqrt{\ell_{jlk}}\left[\mathbf{A}\ \mathbf{0}_{N\times N-P}\right]$, where $\mathbf{A} \in \mathbb{C}^{N\times P}$ and ℓ_{jlk} denotes the inverse path loss from UT k in cell l to BS j.

Downlink

The received signal $y_{jm}^{\mathrm{dl}} \in \mathbb{C}$ of the mth UT in the jth cell is given as

$$y_{jm}^{\mathrm{dl}} = \sqrt{\rho_{\mathrm{dl}}}\sum_{l=1}^{L}\mathbf{h}_{ljm}^{\mathsf{H}}\mathbf{s}_l + n_{jm}^{\mathrm{dl}} \tag{3.45}$$

where $\mathbf{s}_l \in \mathbb{C}^N$ is the transmit vector of BS l, $n_{jm}^{\mathrm{dl}} \sim \mathcal{CN}(0,1)$ is receiver noise and $\rho_{\mathrm{dl}} > 0$ denotes the downlink SNR. We assume channel reciprocity, i.e., the downlink channels are the complex conjugate transpose of the uplink channel $\mathbf{h}_{ljm}$. The transmit vector $\mathbf{s}_l$ is given as

$$\mathbf{s}_l = \sqrt{\lambda_l}\sum_{k=1}^{K}\mathbf{w}_{lk}x_{lk}^{\mathrm{dl}} = \sqrt{\lambda_l}\mathbf{W}_l\mathbf{x}_l^{\mathrm{dl}} \tag{3.46}$$

where $\mathbf{W}_l = \left[\mathbf{w}_{l1}\cdots\mathbf{w}_{lK}\right] \in \mathbb{C}^{N\times K}$ is a precoding matrix and $\mathbf{x}_l = \left[x_{l1}^{\mathrm{dl}}\cdots x_{lK}^{\mathrm{dl}}\right]^{\mathsf{T}} \in \mathbb{C}^K \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ contains the data symbols for the K UTs in cell l. The parameter λ_l normalizes the average transmit power of BS l per UT to $\mathbb{E}\left[\frac{\rho_{\mathrm{dl}}}{K}\mathbf{s}_l^{\mathsf{H}}\mathbf{s}_l\right] = \rho_{\mathrm{dl}}$, i.e.,

$$\lambda_l = \frac{1}{\mathbb{E}\left[\frac{1}{K}\operatorname{tr}\mathbf{W}_l\mathbf{W}_l^{\mathsf{H}}\right]}. \tag{3.47}$$

Channel estimation

During a dedicated uplink training phase, the UTs in each cell transmit orthogonal pilot sequences which allow the BSs to compute estimates $\hat{\mathbf{H}}_{jj}$ of their local channels $\mathbf{H}_{jj}$. The same set of orthogonal pilot sequences is reused in every cell so that the channel estimate is corrupted by pilot contamination from adjacent cells [16]. Under these assumptions, BS j estimates the channel vector $\mathbf{h}_{jjk}$ based on the observation $\mathbf{y}_{jk}^{\tau} \in \mathbb{C}^N$, given as

$$\mathbf{y}_{jk}^{\tau} = \mathbf{h}_{jjk} + \sum_{l\neq j}\mathbf{h}_{jlk} + \frac{1}{\sqrt{\rho_\tau}}\mathbf{n}_{jk}^{\tau} \tag{3.48}$$

where $\mathbf{n}_{jk}^{\tau} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_N\right)$ and $\rho_\tau > 0$ is the effective training SNR. In general, ρ_τ depends on the pilot transmit power and the length of the pilot sequences. Here, we assume ρ_τ to be given. Assuming MMSE estimation, we can decompose $\mathbf{h}_{jjk}$ as $\mathbf{h}_{jjk} = \hat{\mathbf{h}}_{jjk} + \tilde{\mathbf{h}}_{jjk}$, where $\hat{\mathbf{h}}_{jjk} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Phi}_{jjk}\right)$ is the channel estimate and

$\tilde{\mathbf{h}}_{jjk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{jjk} - \mathbf{\Phi}_{jjk})$ is the independent estimation error. The matrices $\mathbf{\Phi}_{jlk} \in \mathbb{C}^{N \times N}$ are defined as

$$\mathbf{\Phi}_{jlk} = \mathbf{R}_{jjk}\mathbf{Q}_{jk}\mathbf{R}_{jlk}, \qquad \mathbf{Q}_{jk} = \left(\frac{1}{\rho_\tau}\mathbf{I}_N + \sum_{l=1}^{L}\mathbf{R}_{jlk}\right)^{-1}. \tag{3.49}$$

For the analysis later on, it will be useful to write $\hat{\mathbf{h}}_{jjk}$ as

$$\hat{\mathbf{h}}_{jjk} = \mathbf{R}_{jjk}\mathbf{Q}_{jk}\mathbf{y}_{jk}^{\tau} = \mathbf{R}_{jjk}\mathbf{Q}_{jk}\left(\sum_{l=1}^{L}\mathbf{h}_{jlk} + \frac{1}{\sqrt{\rho_\tau}}\mathbf{n}_{jk}^{\tau}\right). \tag{3.50}$$

Achievable uplink rates with linear detection

We consider linear single-user detection, where the jth BS estimates the symbol x_{jm}^{ul} of UT m in its cell by computing the inner product between the received vector $\mathbf{y}_j$ and the linear filter $\mathbf{r}_{jm} \in \mathbb{C}^N$. Two particular filters are of practical interest, namely the matched filter $\mathbf{r}_{jm}^{\mathrm{MF}}$ and the MMSE detector $\mathbf{r}_{jm}^{\mathrm{MMSE}}$, which we define as

$$\mathbf{r}_{jm}^{\mathrm{MF}} = \hat{\mathbf{h}}_{jjm}, \qquad \mathbf{r}_{jm}^{\mathrm{MMSE}} = \left(\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^{\mathsf{H}} + \mathbf{Z}_j^{\mathrm{ul}} + N\varphi_j^{\mathrm{ul}}\mathbf{I}_N\right)^{-1}\hat{\mathbf{h}}_{jjm} \tag{3.51}$$

where $\varphi_j^{\mathrm{ul}} > 0$ and $\mathbf{Z}_j^{\mathrm{ul}} \in \mathbb{C}^{N \times N}$ is and arbitrary Hermitian nonnegative definite matrix. Natural choices are $\varphi_j^{\mathrm{ul}} = \frac{1}{\rho_{\mathrm{ul}}N}$ and

$$\mathbf{Z}_j^{\mathrm{ul}} = \mathbb{E}\left[\tilde{\mathbf{H}}_{jj}\tilde{\mathbf{H}}_{jj}^{\mathsf{H}} + \sum_{l \neq j}\mathbf{H}_{jl}\mathbf{H}_{jl}\right] = \sum_k(\mathbf{R}_{jjk} - \mathbf{\Phi}_{jjk}) + \sum_{l \neq j}\sum_k \mathbf{R}_{jlk}. \tag{3.52}$$

Note that BS j could theoretically estimate all channel matrices $\mathbf{H}_{jl}$ from the observations (3.48) to improve the performance. Nevertheless, high path loss to neighboring cells is likely to render these channel estimates unreliable and the potential performance gains small. Our formulation of $\mathbf{r}_{jm}^{\mathrm{MMSE}}$ allows us to treat φ_j^{ul} and $\mathbf{Z}_j^{\mathrm{ul}}$ as design parameters which could be optimized.

Using a standard bound based on the worst-case uncorrelated additive noise yields the ergodic achievable uplink rate R_{jm}^{ul} of UT m in cell j [117]:

$$R_{jm}^{\mathrm{ul}} = \mathbb{E}\left[\log_2\left(1 + \gamma_{jm}^{\mathrm{ul}}\right)\right] \tag{3.53}$$

with the associated signal-to-interference-plus-noise ratio (SINR) $\gamma_{jm}^{\mathrm{ul}}$, given by

$$\gamma_{jm}^{\mathrm{ul}} = \frac{\left|\mathbf{r}_{jm}^{\mathsf{H}}\hat{\mathbf{h}}_{jjm}\right|^2}{\mathbb{E}\left[\mathbf{r}_{jm}^{\mathsf{H}}\left(\frac{1}{\rho_{\mathrm{ul}}}\mathbf{I}_N + \tilde{\mathbf{h}}_{jjm}\tilde{\mathbf{h}}_{jjm}^{\mathsf{H}} - \mathbf{h}_{jjm}\mathbf{h}_{jjm}^{\mathsf{H}} + \sum_l \mathbf{H}_{jl}\mathbf{H}_{jl}^{\mathsf{H}}\right)\mathbf{r}_{jm}\middle|\hat{\mathbf{H}}_{jj}\right]}. \tag{3.54}$$

We will denote by $\gamma_{jm}^{\mathrm{MF}}$ and $\gamma_{jm}^{\mathrm{MMSE}}$ the SINR with MF and MMSE detection, respectively.

Achievable downlink rates with linear precoding

Since the UTs do not have any channel estimates, we provide an ergodic achievable rate based on the techniques as developed in [149]. To this end, we decompose the received signal y_{jm}^{dl} as

$$\begin{aligned} y_{jm}^{\mathrm{dl}} = \sqrt{\rho_{\mathrm{dl}}\lambda_j}\mathbb{E}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\mathbf{w}_{jm}\right] x_{jm}^{\mathrm{dl}} + \sqrt{\rho_{\mathrm{dl}}\lambda_j}\left(\mathbf{h}_{jjm}^{\mathsf{H}}\mathbf{w}_{jm} - \mathbb{E}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\mathbf{w}_{jm}\right]\right) x_{jm}^{\mathrm{dl}} \\ + \sum_{(l,k)\neq(j,m)} \sqrt{\rho_{\mathrm{dl}}\lambda_l}\mathbf{h}_{ljm}^{\mathsf{H}}\mathbf{w}_{kl}x_{kl}^{\mathrm{dl}} + n_{jm}^{\mathrm{dl}} \end{aligned} \tag{3.55}$$

and assume that the average effective channels $\sqrt{\lambda_j}\mathbb{E}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\mathbf{w}_{jm}\right]$ can be perfectly learned at the UTs. Thus, an ergodic achievable rate R_{jm}^{dl} of UT m in cell j is given as [149, Theorem 1]

$$R_{jm}^{\mathrm{dl}} = \log_2\left(1 + \gamma_{jm}^{\mathrm{dl}}\right) \tag{3.56}$$

with the associated SINR $\gamma_{jm}^{\mathrm{dl}}$:

$$\gamma_{jm}^{\mathrm{dl}} = \frac{\lambda_j \left|\mathbb{E}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\mathbf{w}_{jm}\right]\right|^2}{\frac{1}{\rho_{\mathrm{dl}}} + \lambda_j \mathrm{var}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\mathbf{w}_{jm}\right] + \sum_{(l,k)\neq(j,m)} \lambda_l \mathbb{E}\left[\left|\mathbf{h}_{ljm}^{\mathsf{H}}\mathbf{w}_{lk}\right|^2\right]} \tag{3.57}$$

where $\mathrm{var}\left[x\right] \triangleq \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^{\mathsf{H}}]$ for some random variable x. We consider two different linear precoding strategies $\mathbf{W}_j$ of practical interest, namely eigenbeamforming (BF) $\mathbf{W}_j^{\mathrm{BF}}$ and regularized zero-forcing (RZF) $\mathbf{W}_j^{\mathrm{RZF}}$, which we define as

$$\mathbf{W}_j^{\mathrm{BF}} = \hat{\mathbf{H}}_{jj}, \qquad \mathbf{W}_j^{\mathrm{RZF}} = \left(\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^{\mathsf{H}} + \mathbf{Z}_j^{\mathrm{dl}} + N\varphi_j^{\mathrm{dl}}\mathbf{I}_N\right)^{-1}\hat{\mathbf{H}}_{jj} \tag{3.58}$$

where $\varphi_j^{\mathrm{dl}} > 0$ is a regularization parameter and $\mathbf{Z}_j^{\mathrm{dl}} \in \mathbb{C}^{N\times N}$ is an arbitrary Hermitian nonnegative definite matrix. As the choice of $\mathbf{Z}_j^{\mathrm{dl}}$ and φ_j^{dl} is arbitrary, they could be further optimized (see e.g., [149, Theorem 6]). This aspect is left to future work.

We will denote by $\gamma_{jm}^{\mathrm{BF}}$ and $\gamma_{jm}^{\mathrm{RZF}}$ the SINR with BF and RZF, respectively.

Remark 16. *Under a block-fading channel model with coherence time T, one can account for the rate loss due to channel training by considering the net ergodic achievable rates $\kappa(1-\tau/T)R_{jm}^{ul}$ and $(1-\kappa)(1-\tau/T)R_{jm}^{dl}$ for a given training length $\tau \in [K,T]$ and some $\kappa \in [0,1]$ which determines the fraction of the coherence time used for uplink transmissions.*

3.3.3 Asymptotic analysis

In this section, we present our main technical results. As the ergodic achievable rates R_{jm}^{ul} and R_{jm}^{dl} are difficult to compute for finite system dimensions, we consider the large system limit, where N and K grow infinitely large while keeping a finite ratio K/N. This is in contrast to [16] which assumes that the number of UTs K remains fixed while the number of antennas grows without bound. We will retrieve the results of [16] as a special case. The large system limit implicitly assumes that the coherence time of the channel scales linearly with K (to allow for orthogonal pilot sequences of the UTs in a cell). However, as we use the asymptotic analysis only as a tool to provide approximations for finite N, K, this does not pose any problem. In what follows, we will derive deterministic equivalents $\bar{\gamma}_{jm}^{\mathrm{ul}}$ ($\bar{\gamma}_{jm}^{\mathrm{dl}}$) of the SINR $\gamma_{jm}^{\mathrm{ul}}$ ($\gamma_{jm}^{\mathrm{dl}}$) for MF and MMSE detector (BF and RZF precoder), respectively, such that

$$\gamma_{jm}^{\mathrm{ul}} - \bar{\gamma}_{jm}^{\mathrm{ul}} \xrightarrow{\text{a.s.}} 0, \qquad \gamma_{jm}^{\mathrm{dl}} - \bar{\gamma}_{jm}^{\mathrm{dl}} \to 0. \tag{3.59}$$

One can then show by the dominated convergence theorem and the continuous mapping theorem, respectively, that (3.59) implies that

$$R_{jm}^{\mathrm{ul}} - \log_2\left(1 + \bar{\gamma}_{jm}^{\mathrm{ul}}\right) \to 0, \qquad R_{jm}^{\mathrm{dl}} - \log_2\left(1 + \bar{\gamma}_{jm}^{\mathrm{dl}}\right) \to 0. \tag{3.60}$$

These results must be understood in the way that, for each given set of system parameters (N, K), we provide deterministic approximations of the SINR and the associate rate which become increasingly tight as N and K grow. We will show later by simulations that these approximations are already very tight for realistic system dimensions.

In the sequel, the notation "$N \to \infty$" will refer to $K, N \to \infty$ such that $0 \leq \liminf K/N \leq \limsup K/N < \infty$. Moreover, we assume that the following conditions hold:

A 1. $\limsup_N \|\mathbf{R}_{jlk}\| < \infty$ *for all* j, l, k.

A 2. $\liminf_N \frac{1}{N} \operatorname{tr} \mathbf{R}_{jlk} > 0$ *for all* j, l, k.

A 3. $\limsup_N \|\mathbf{Z}_j^{ul}\| < \infty$, $\limsup_N \|\mathbf{Z}_j^{dl}\| < \infty$ *for all* j.

The next theorems provide SINR approximations in the sense of (3.59) for MF and MMSE detection in the uplink and for BF and RZF precoding in the downlink. The results for MMSE/RZF require Theorems 14 and 21 which are provided in Sections 2.3.1 and 2.3.2, respectively. Due to the similarity of the SINR expressions for the up- and downlink, we only provide the proofs for BF and RZF in the appendix. The proofs for MF and MMSE are very similar and will be omitted.

Theorem 30 (Matched filter)**.**

$$\bar{\gamma}_{jm}^{MF} = \frac{\left(\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjm}\right)^2}{\frac{1}{\rho_{ul}N}\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjm} + \frac{1}{N}\sum_{l,k}\frac{1}{N}\operatorname{tr}\mathbf{R}_{jlk}\mathbf{\Phi}_{jjm} + \sum_{l\neq j}\left|\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jlm}\right|^2}.$$

Theorem 31 (Eigenbeamforming)**.**

$$\bar{\gamma}_{jm}^{BF} = \frac{\bar{\lambda}_j\left(\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjm}\right)^2}{\frac{1}{\rho_{dl}N} + \frac{1}{N}\sum_{l,k}\bar{\lambda}_l\frac{1}{N}\operatorname{tr}\mathbf{R}_{ljm}\mathbf{\Phi}_{llk} + \sum_{l\neq j}\bar{\lambda}_l\left|\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{ljm}\right|^2}$$

where $\bar{\lambda}_j = \left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjk}\right)^{-1}$.

Proof. The proof is provided in Section 3.7.4. □

Theorem 32 (MMSE detector)**.**

$$\bar{\gamma}_{jm}^{MMSE} = \frac{\delta_{jm}^2}{\frac{1}{\rho_{ul}N}\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjm}\bar{\mathbf{T}}'_j + \frac{1}{N}\sum_{l,k}\mu_{jlkm} + \sum_{l\neq j}|\vartheta_{jlm}|^2}$$

where

$$\mu_{jlkm} = \frac{1}{N}\operatorname{tr}\mathbf{R}_{jlk}\mathbf{T}'_{jm} - \frac{2\Re\left(\vartheta_{jlk}^*\vartheta'_{jlkm}\right)(1+\delta_{jk}) - |\vartheta_{jlk}|^2\delta'_{jkm}}{(1+\delta_{jk})^2}$$

$\vartheta_{jlk} = \frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jlk}\mathbf{T}_j$, $\vartheta'_{jlkm} = \frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jlk}\mathbf{T}'_{jm}$, *and where*

(i) $\mathbf{T}_j = \mathbf{T}_N(\varphi_j^{ul})$ *and* $\boldsymbol{\delta}_j = [\delta_{j1}\cdots\delta_{jK}]^\mathsf{T} = \boldsymbol{\delta}(\varphi_j^{ul})$ *are given by Theorem 14 for* $\mathbf{S}_N = \mathbf{Z}_j^{ul}/N$, $\mathbf{D}_N = \mathbf{I}_N$, $\mathbf{R}_k = \mathbf{\Phi}_{jjk}\,\forall k$,

(ii) $\bar{\mathbf{T}}'_j = \mathbf{T}'_N(\varphi_j^{ul})$ *is given by Theorem 21 for* $\mathbf{S}_N = \mathbf{Z}_j^{u;}/N$, $\mathbf{\Theta}_N = \mathbf{I}_N$, $\mathbf{D}_N = \mathbf{I}_N$, *and* $\mathbf{R}_k = \mathbf{\Phi}_{jjk}\,\forall k$,

(iii) $\mathbf{T}'_{jm} = \mathbf{T}'_N(\varphi_j^{ul})$ *and* $\boldsymbol{\delta}'_{jm} = \left[\delta'_{j1m}\cdots\delta'_{jKm}\right]^\mathsf{T} = \boldsymbol{\delta}'(\varphi_j^{ul})$ *are given by Theorem 21 for* $\mathbf{S}_N = \mathbf{Z}_j^{ul}/N$, $\mathbf{\Theta}_N = \mathbf{\Phi}_{jjm}$, $\mathbf{D}_N = \mathbf{I}_N$, $\mathbf{R}_k = \mathbf{\Phi}_{jjk}\,\forall k$.

Theorem 33 (Regularized Zero-Forcing)**.**

$$\bar{\gamma}_{jm}^{RZF} = \frac{\bar{\lambda}_j\delta_{jm}^2}{\frac{(1+\delta_{jm})^2}{\rho_{dl}N} + \frac{1}{N}\sum_{l,k}\bar{\lambda}_l\left(\frac{1+\delta_{jm}}{1+\delta_{lk}}\right)^2\mu_{ljmk} + \sum_{l\neq j}\bar{\lambda}_l\left(\frac{1+\delta_{jm}}{1+\delta_{lm}}\right)^2|\vartheta_{ljm}|^2}$$

where

$$\mu_{ljmk} = \frac{1}{N}\operatorname{tr}\mathbf{R}_{ljm}\mathbf{T}'_{lk} - \frac{2\Re\left(\vartheta_{ljm}^*\vartheta'_{ljmk}\right)(1+\delta_{lm}) - |\vartheta_{ljm}|^2\delta'_{lmk}}{(1+\delta_{lm})^2}$$

$$\bar{\lambda}_l = \frac{K}{N}\left(\frac{1}{N}\operatorname{tr}\mathbf{T}_l - \frac{1}{N}\operatorname{tr}\left(\frac{\mathbf{Z}_j^{dl}}{N} + \varphi_l^{dl}\mathbf{I}_N\right)\bar{\mathbf{T}}'_l\right)^{-1}$$

$\vartheta_{ljm} = \frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{ljm}\mathbf{T}_l$ *and* $\vartheta'_{ljmk} = \frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{ljm}\mathbf{T}'_{lk}$, *and where*

(i) $\mathbf{T}_l = \mathbf{T}_N(\varphi_l^{dl})$ *and* $\boldsymbol{\delta}_l = [\delta_{l1}\cdots\delta_{lK}]^\mathsf{T} = \boldsymbol{\delta}(\varphi_l^{dl})$ *are given by Theorem 14 for* $\mathbf{S}_N = \mathbf{Z}_l^{dl}/N$, $\mathbf{D}_N = \mathbf{I}_N$ *and* $\mathbf{R}_k = \mathbf{\Phi}_{llk}\,\forall k$,

(ii) $\bar{\mathbf{T}}'_l = \mathbf{T}'_N(\varphi_l^{dl})$ *is given by Theorem 21 for* $\mathbf{S}_N = \mathbf{Z}_l^{dl}/N$, $\mathbf{\Theta}_N = \mathbf{I}_N$, $\mathbf{D}_N = \mathbf{I}_N$ *and* $\mathbf{R}_k = \mathbf{\Phi}_{llk}\,\forall k$,

(iii) $\mathbf{T}'_{lk} = \mathbf{T}'_N(\varphi_l^{dl})$ *and* $\boldsymbol{\delta}'_{lk} = [\delta'_{l1k}\cdots\delta'_{lKk}]^\mathsf{T} = \boldsymbol{\delta}'(\varphi_l^{dl})$ *are given by Theorem 21 for* $\mathbf{S}_N = \mathbf{Z}_l^{dl}/N$, $\mathbf{\Theta}_N = \mathbf{\Phi}_{llk}$, $\mathbf{D}_N = \mathbf{I}_N$ *and* $\mathbf{R}_k = \mathbf{\Phi}_{llk}\,\forall k$.

Proof. The proof is provided in Section 3.7.5. □

Remark 17. *Observe the similarity between the results for MF/BF and MMS/RZF, respectively: In the downlink, all transmit powers are multiplied by the power normalization factors* $\bar{\lambda}_j$. *Moreover, the indices* j,l *and* k,m *are swapped for the interference terms.*

Remark 18. *The expressions of $\bar{\gamma}_{jm}^{MMSE}$ and $\bar{\gamma}_{jm}^{RZF}$ can be greatly simplified under a less general channel model, e.g., no antenna correlation, Wyner-type models with a simple path loss factor. We will provide later on a special case in which $\bar{\gamma}_{jm}^{MMSE}$ and $\bar{\gamma}_{jm}^{RZF}$ can be given in closed form.*

Next, we consider the case of an infinite number of antennas per UT, i.e., $K/N \to 0$.

Corollary 2. *Let $N \to \infty$, such that $K/N \to 0$. Denote $\beta_{jlk} = \lim_N \frac{1}{N} \operatorname{tr} \mathbf{\Phi}_{jlk}$, whenever the limit exists, and define $\bar{\lambda}_j^{\infty,BF} = \left(\frac{1}{K}\sum_{k=1}^K \beta_{jjk}\right)^{-1}$ and $\bar{\lambda}_j^{\infty,RZF} = \left(\frac{1}{K}\sum_{k=1}^K \frac{\beta_{jjk}}{\left(\varphi_j^{dl}+\beta_{jjk}\right)^2}\right)^{-1}$. Then,*

$$
\begin{aligned}
\bar{\gamma}_{jm}^{MF} &\to \frac{\beta_{jjm}^2}{\sum_{l\neq j} |\beta_{jlm}|^2}\\
\bar{\gamma}_{jm}^{BF} &\to \frac{\bar{\lambda}_j^{\infty,BF}\beta_{jjm}^2}{\sum_{l\neq j} \bar{\lambda}_l^{\infty,BF}|\beta_{ljm}|^2}\\
\bar{\gamma}_{jm}^{MMSE} &\to \frac{\beta_{jjm}^2}{\sum_{l\neq j} \left(\frac{\varphi_j^{ul}}{\varphi_l^{ul}}\right)^2 |\beta_{jlm}|^2}\\
\bar{\gamma}_{jm}^{RZF} &\to \frac{\bar{\lambda}_j^{\infty,RZF}\beta_{jjm}^2}{\sum_{l\neq j} \left(\frac{\varphi_l^{dl}\varphi_j^{dl}+\varphi_l^{dl}\beta_{jjm}}{\varphi_l^{dl}\varphi_j^{dl}+\varphi_j^{dl}\beta_{llm}}\right)^2 \bar{\lambda}_l^{\infty,RZF}|\beta_{ljm}|^2}.
\end{aligned}
$$

Proof. Note that the first and the second term in the denominator of all SINR expressions vanish as $N \to \infty$ while $K/N \to 0$. For the remaining terms, note that $\mathbf{T}_j(\varphi) \to \varphi^{-1}$ and $\bar{\mathbf{T}}_j'(\varphi) \to \varphi^{-2}$. Lastly, for RZF, we can write $\bar{\lambda}_j$ equivalently as $\bar{\lambda}_j = \left(\frac{1}{K}\sum_k \frac{\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjk}\mathbf{T}_j'}{(1+\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjk}\mathbf{T}_j)^2}\right)^{-1}$. □

Remark 19. *As already observed in [16, Eq. (13)], the performance of MF and MMSE detector coincide with an infinite number of BS-antennas per UT if $\varphi_j^{ul} = \varphi_l^{ul}\ \forall l$. However, even for $\bar{\lambda}_j^{\infty} = \bar{\lambda}_l^{\infty}$ and $\varphi_j^{dl} = \varphi_l^{dl}\ \forall l$, the SINR under RZF and BF are not necessarily identical. This is because the received interference power depends on the correlation matrices $\mathbf{\Phi}_{llm}$.*

3.3.4 On the massive MIMO effect

Let us now consider the simplified channel model

$$
\mathbf{H}_{jj} = \sqrt{\frac{N}{P}}\mathbf{A}\mathbf{W}_{jj}, \quad \mathbf{H}_{jl} = \sqrt{\alpha\frac{N}{P}}\mathbf{A}\mathbf{W}_{jl}, \quad l \neq j \tag{3.61}
$$

where $\mathbf{A} \in \mathbb{C}^{N\times P}$ is composed of $P \leq N$ columns of an arbitrary unitary $N \times N$ matrix, $\mathbf{W}_{jl} \in \mathbb{C}^{P\times K}$ are standard complex Gaussian matrices and $\alpha \in (0,1]$ is an intercell interference factor. Note that this is a special case of (3.44). Under this model, the total energy of the channel grows linearly with the number of antennas N and UTs K, since $\mathbb{E}\left[\operatorname{tr}\mathbf{H}_{jj}\mathbf{H}_{jj}^{\mathsf{H}}\right] = \frac{KN}{P}\operatorname{tr}\mathbf{A}\mathbf{A}^{\mathsf{H}} = KN$. The motivation behind this channel model is twofold. First, we assume that the antenna aperture increases with each additional antenna element. Thus, the captured energy increases linearly with N. This is in contrast to existing works which assume that more and more antenna elements are packed into a fixed volume, see e.g., [150]. An insufficiency of this channel model is that the captured energy grows without bounds as $N \to \infty$. However, we believe that linear energy gains can be achieved up to very large numbers of antennas if the size of the antenna array is scaled accordingly. Second, the number of DoF P offered by the channel does not need to be equal to N [147]. One could either assume P to be large but constant or to scale with N, e.g., $P = cN$, where $c \in (0, 1]$. In general, P depends on the amount of scattering in the channel and, therefore, on the radio environment. A saturation of P for large N was recently confirmed by channel measurements in [151]. Let us further assume that the transmit powers per UT in up- and downlink are equal, i.e., $\rho_{\mathrm{ul}} = \rho_{\mathrm{dl}} = \rho$, and that the matrices $\mathbf{Z}_j^{\mathrm{dl}}$ and $\mathbf{Z}_j^{\mathrm{ul}}$ used for precoding and detection are equal and given by (3.52). Under these assumptions, the performance of MF and BF (MMSE and RZF) coincides and Theorems 30–33 can be given in closed form:

Corollary 3. *For the channel model* (3.61) *and* $\rho_{ul} = \rho_{dl} = \rho$, $\bar{\gamma}_{jm}^{MF}$ *and* $\bar{\gamma}_{jm}^{BF}$, $\forall k, m$, *are given as*

$$\bar{\gamma}^{MF} = \bar{\gamma}^{BF} = \frac{1}{\frac{1}{\eta\rho N} + \frac{K}{P}\frac{\bar{L}}{\eta} + \alpha(\bar{L}-1)}$$
$$= \frac{1}{\underbrace{\frac{\bar{L}}{\rho N}}_{noise} + \underbrace{\frac{1}{\rho_\tau}\left(\frac{P/N}{\rho N} + \frac{K}{N}\bar{L}\right)}_{imperfect\ CSI} + \underbrace{\frac{K}{P}\bar{L}^2}_{interference} + \underbrace{\alpha\left(\bar{L}-1\right)}_{pilot\ contamination}} \tag{3.62}$$

where $\bar{L} = 1 + \alpha(L-1)$ *and* $\eta = \frac{\rho_\tau \frac{N}{P}}{1+\rho_\tau \frac{N}{P}\bar{L}}$.

Corollary 4. *For the channel model* (3.61), $\rho_{ul} = \rho_{dl}$, $\varphi_j^{ul} = \varphi_j^{dl} = \varphi$, *and* $\mathbf{Z}_j^{ul} = \mathbf{Z}_j^{dl} = \mathbf{Z}_j = \sum_{l,k}\mathbf{R}_{jlk} - \sum_k \mathbf{\Phi}_{jjk}$, $\forall j$, $\bar{\gamma}_{jm}^{MMSE}$ *and* $\bar{\gamma}_{jm}^{RZF}$, $\forall k, m$, *are given as*

$$\bar{\gamma}^{MMSE} = \bar{\gamma}^{RZF} = \frac{1}{\frac{1}{\eta\rho N}X + \frac{K}{P}\frac{\bar{L}}{\eta}Y + \alpha(\bar{L}-1)} \tag{3.63}$$

where $\bar{L} = 1 + \alpha(L-1)$, $\eta = \frac{\rho_\tau \frac{N}{P}}{1+\rho_\tau \frac{N}{P}\bar{L}}$, $X = \frac{Z^2}{Z^2 - \frac{K}{P}}$, $Z = \frac{1+\delta}{\delta}$, $S = \frac{\varphi}{\eta} + \frac{K\bar{L}}{P\eta}$,

$$Y = X + \frac{\eta(1+\alpha^2(L-1))(1-2Z)}{\bar{L}(Z^2 - K/P)}, \qquad \delta = \frac{1 - S + \sqrt{(1+S)^2 - 4K/P}}{2(S - K/P)}.$$

Sketch. First, notice that the matrices $\mathbf{\Phi}_{jlk}$ and $\mathbf{Z}_j$ can be simplified to $\mathbf{\Phi}_{jlk} = \max(\mathbb{1}_{l=j}, \alpha)\frac{N}{P}\mathbf{A}\mathbf{A}^{\mathsf{H}}$ and $\mathbf{Z}_j = K(\bar{L} - \eta)\frac{N}{P}\mathbf{A}\mathbf{A}^{\mathsf{H}}$ $\forall j, l, k$. For these values, one can show after some straight-forward but tedious calculus that Theorems 14 and 21 can be given in closed form. □

One can make several observations from (3.62) and (3.63). Obviously, the effective SNR ρN increases linearly with N. Thus, if the number of antennas is doubled, the transmit power can be reduced by a factor of two to achieve the same SNR. However, if the transmit *and* the training power are reduced as N grows, this conclusions fails to hold. As can been seen from the term $\eta\rho N$, the product of transmit and training SNR must satisfy $\rho\rho_\tau \geq \mathcal{O}(1/N)$. Otherwise the SINR converges to zero as $N \to \infty$. As already observed in [143], if $\rho_\tau = \rho$, the transmit power can be made only inversely proportional to $\sqrt{N}$. Surprisingly, the interference depends mainly on the ratio P/K (number of DoF per UT) and not directly on the number of antennas. Thus, interference can only be mitigated by additional antennas if the environment provides sufficient scattering. Moreover, noise, channel estimation errors and interference vanish for $N, P \to \infty$ while pilot contamination is the only performance-limiting factor:

$$\bar{\gamma}^{\mathrm{MF}}, \bar{\gamma}^{\mathrm{BF}}, \bar{\gamma}^{\mathrm{MMSE}}, \bar{\gamma}^{\mathrm{RZF}} \xrightarrow[N,P\to\infty,\ K/N\to 0]{} \gamma_\infty = \frac{1}{\alpha(\bar{L}-1)}. \tag{3.64}$$

We denote by R_∞ the ultimately achievable rate, defined as

$$R_\infty = \log_2(1+\gamma_\infty) = \log_2\left(1 + \frac{1}{\alpha(\bar{L}-1)}\right). \tag{3.65}$$

It is interesting that even with more sophisticated linear single-user detection/precoding, such as MMSE and RZF, the ultimate performance limit γ_∞ cannot be exceeded. Note that without pilot contamination, i.e., for $L = 1$ or $\alpha = 0$, the SINR grows without bounds as $P, N \to \infty$. If P is fixed but large, the SINR saturates at a smaller value than γ_∞. In this case, adding additional antennas only improves the SNR but does not reduce the multiuser interference. Thus, with a finite number of DoF, MMSE/RZF remain also for $N \to \infty$ superior to MF/BF.

Before we proceed, let us verify the accuracy of the approximations $\bar{R}^{\mathrm{MF}}$ and $\bar{R}^{\mathrm{MMSE}}$ for finite N, K. In Fig. 3.11, we depict the ergodic achievable rate R_{jm} of an arbitrary UT with MF and MMSE detection as a function of the number of antennas N for $K = 10$ UTs, $L = 4$ cells, $\rho = 0\,\mathrm{dB}$, $\varphi_l^{\mathrm{ul}} = 1/(\rho N)$ and intercell interference factor $\alpha = 0.1$. We compare two different cases: $P = N$ and $P = N/3$. As expected, the performance in the latter scenario is inferior due to stronger multiuser interference. Most importantly, our closed-form approximations are almost indistinguishable from the simulation results over the entire range of N. The results for the downlink with MF and RZF look similar.

Based on our previous observations, we believe that it is justified to speak about a *massive MIMO effect* whenever the SINR γ_{jm} (in UL or DL) is close to γ_∞, or in other words, whenever noise, channel estimation errors and interference are small compared to the pilot contamination. It becomes evident from (3.62) and (3.63) that the number of antennas needed to achieve this effect depends strongly on the system parameters P, K, L, α, ρ_τ and ρ. In particular, there is no massive MIMO effect without pilot contamination since $\gamma_\infty \to \infty$. Thus, massive MIMO can be seen as a particular operating condition in multi-cellular systems where the performance is ultimately limited by pilot contamination and MF/BF achieve a performance close to this ultimate limit. To make this definition more precise, we say that we operate under massive MIMO conditions if, for some desired "massive MIMO efficiency" $\eta \in (0, 1)$,

$$R = \log(1+\gamma) \ \geq \ \eta R_\infty \tag{3.66}$$

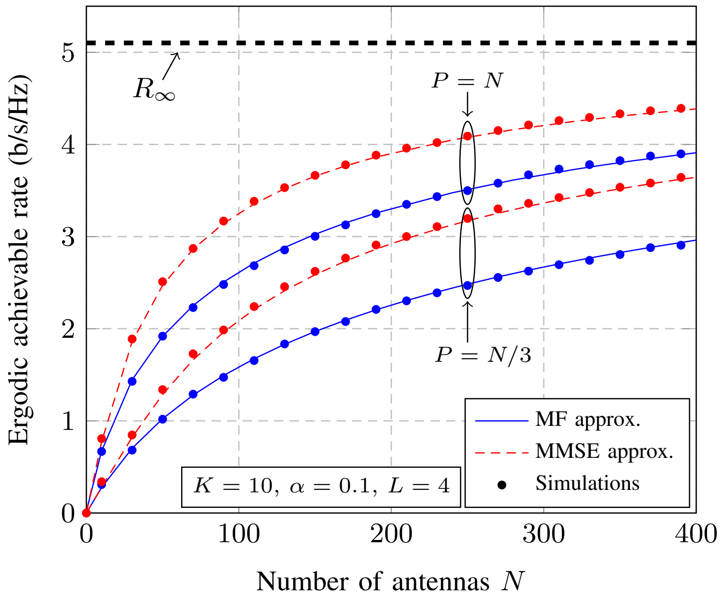

Figure 3.11: Ergodic achievable rate with MF and MMSE detection versus number of antennas N for $P \in \{N, N/3\}$ for $\rho = 0\,\mathrm{dB}$.

where γ is the SINR in the UL/DL with any detection/precoding scheme. This condition implies that we achieve at least the fraction η of the ultimate performance limit. If we assume that $\rho_\tau \gg 1$, i.e., $\eta \approx \bar{L}^{-1}$, the expressions of $\bar{\gamma}^{\mathrm{MF}}$, $\bar{\gamma}^{\mathrm{BF}}$, $\bar{\gamma}^{\mathrm{MMSE}}$, and $\bar{\gamma}^{\mathrm{RZF}}$ in Corollaries 3 and 4 depend on P, K, ρ, and N only through the ratio $\frac{P}{K}$ and the effective SNR ρN. Thus, for a given set of parameters $(\rho, N, \alpha, L, \varphi)$, we can easily find the fraction $\frac{P}{K}$ necessary to satisfy (3.66).

Figs. 3.12 and 3.13 show the necessary DoF per UT $\frac{P}{K}$ for a given effective SNR ρN to achieve a spectral efficiency of ηR_∞ with either MF/BF (solid lines) or MMSE/RZF (dashed lines). We consider $L = 4$ cells, $\varphi = 1/(\rho N)$ and an intercell interference factor $\alpha = 0.3$ and $\alpha = 0.1$, respectively. The plots must be understood in the following way: Each curve corresponds to a particular value of η. In the region above each curve, the condition (3.66) is satisfied.

Let us first focus on Fig. 3.12 with $\alpha = 0.3$. For an effective SNR $\rho N = 20\,\mathrm{dB}$ (e.g., $\rho = 0\,\mathrm{dB}$ and $N = 100 = 20\,\mathrm{dB}$), we need about $P/K = 90$ DoF per UT with MF/BF to achieve $90\,\%$ of the ultimate performance R_∞, i.e., $0.9 \times 2.2 \approx 2\,\mathrm{b/s/Hz}$. If $P \approx N$, only a single UT could be served (Note that this is a simplifying example. Our analysis assumes $K \gg 1$.). However, if we had $N = 1000 = 30\,\mathrm{dB}$ antennas, the transmit power ρ could be decreased by $10\,\mathrm{dB}$ and 10 UTs could be served with the same performance. At the same operating point, the MMSE/RZF requires only ~ 60 DoF per UT to achieve $90\,\%$ of the ultimate performance. Thus, the use of MMSE/RZF would allow us to increase the number of simultaneously served UTs by a factor $\frac{90}{60} = 1.5$. This example also demonstrates the importance of the relation between N and P. In particular, if P saturates for some N, adding more antennas increases the effective SNR but does not reduce the multiuser interference. Thus, the number

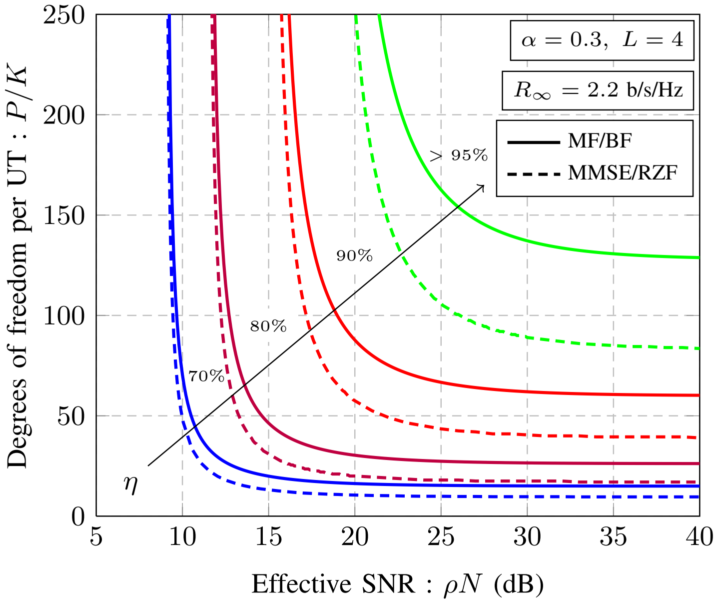

Figure 3.12: Degrees of freedom per UT P/K necessary to achieve ηR_∞ versus effective SNR ρN for $L = 4$ and $\alpha = 0.3$.

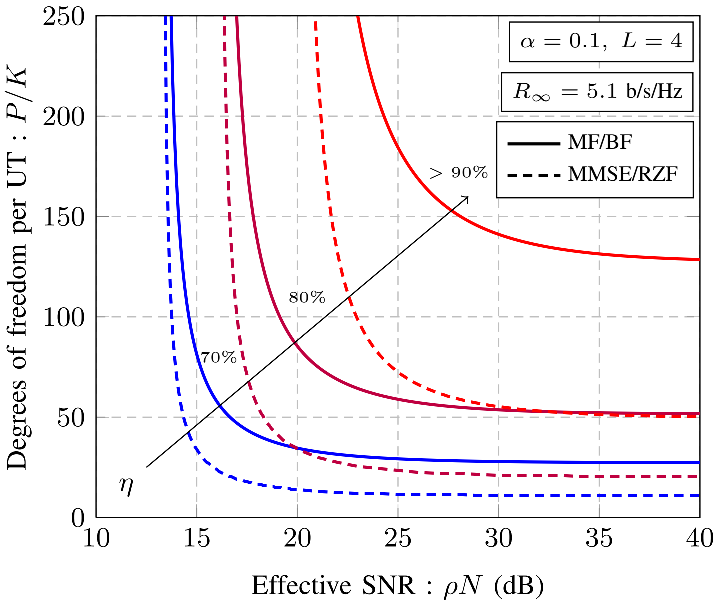

Figure 3.13: Degrees of freedom per UT P/K necessary to achieve ηR_∞ versus effective SNR ρN for $L = 4$ and $\alpha = 0.1$.

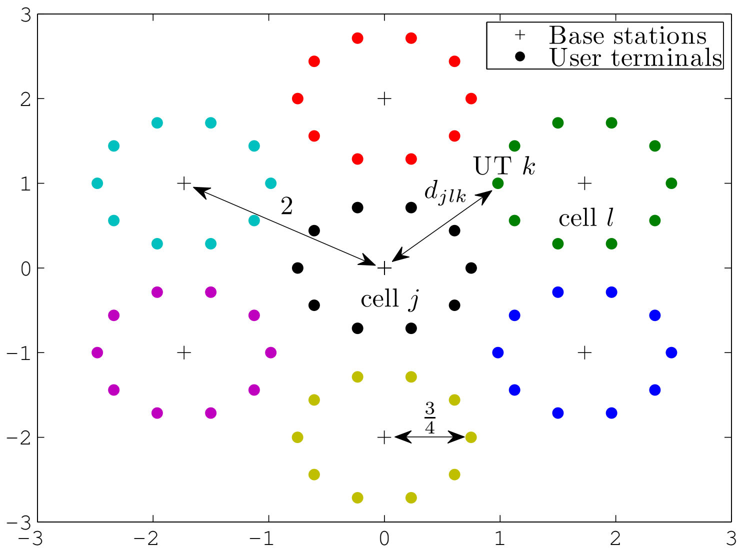

Figure 3.14: 7–cell hexagonal system layout. The distance between two adjacent cells is normalized to 2. There are $K = 10$ UTs uniformly distributed on a circle of radius 3/4 around each BS.

of UTs which can be simultaneously supported depends significantly on the radio environment. We can further see that adding antennas shows diminishing returns. This is because the distances between the curves for different values of η grow exponentially fast. Remember that for $\eta = 1$, a ratio of $P/K = \infty$ would be needed. A last observation we can make is that the absolute difference between MF/BF and MMSE/RZF is marginal for small values of η but gets quickly pronounced as $\eta \to 1$.

Turning to Fig. 3.13 for $\alpha = 0.1$, we can see that for the same effective SNR $\rho N = 20\,\text{dB}$ and the same number of DoF per UT $P/K = 90$ as in the previous example, only 80 % of the ultimate performance are achieved by MF/BF. However, since the intercell interference is significantly smaller compared to the previous example, this corresponds to $0.9 \times 5.1 \approx 4.6\,\text{b/s/Hz}$. Thus, although we operate further away from the ultimate performance limit, the resulting spectral efficiency is still higher. With MMSE/RZF, only 35 DoF per UT are necessary to achieve the same performance and, consequently, $90/35 \approx 2.5$ times more UTs could be simultaneously served. With decreasing intercell interference (and hence decreasing pilot contamination) the advantages of MMSE/RZF become more and more important.

3.3.5 Numerical results

Let us now validate the accuracy of Theorems 31 and 33 for finite N, K in a more realistic downlink scenario. Simulations for the uplink, i.e., Theorems 30 and 32, are omitted but provide very similar results (see Remark 17). We consider a hexagonal system with $L = 7$ cells as shown in Fig. 3.14. The inner cell radius is normalized to one and we assume a distance-based path loss model with path loss exponent $\beta = 3.7$. To allow for reproducibility of our results, we distribute

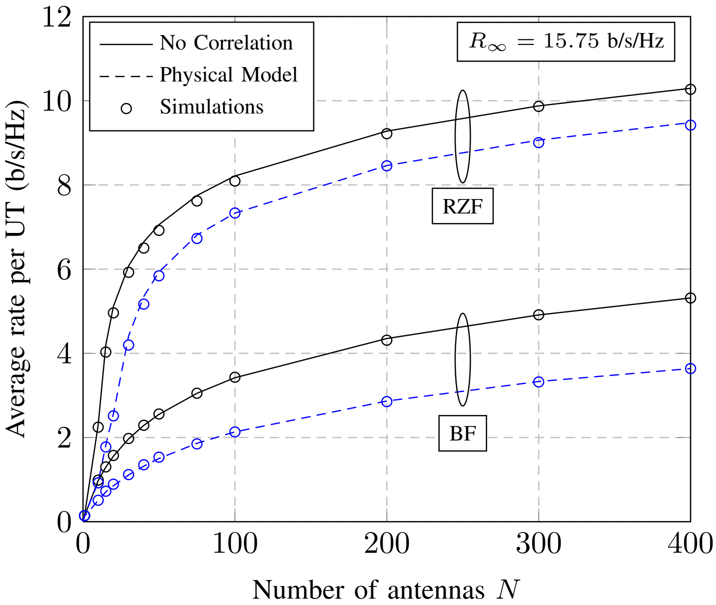

Figure 3.15: Average per-user rate versus number of antennas N for the RZF and BF precoders. Solid and dashed lines depict the asymptotic approximations, markers the simulation results.

$K = 10$ UTs uniformly on a circle of radius $3/4$ around each BS and do not consider shadowing. We further assume a training SNR $\rho_\tau = 6\,\text{dB}$ and transmit SNR $\rho_{\text{dl}} = 10\,\text{dB}$. For RZF, we use a regularization factor $\varphi_j^{\text{dl}} = 1/\rho_{\text{dl}}$ and $\mathbf{Z}_j^{\text{dl}} = 0$. Average rates are then calculated for the UTs in the center cell.

First, we consider a simple channel model without antenna correlation, i.e., $\tilde{\mathbf{R}}_{jlk} = d_{jlk}^{-\beta/2}\mathbf{I}_N$, where d_{jlk} is the distance between BS j and the kth UTs in cell l (cf. (3.44)). Both precoding schemes lead to the ultimate rate $R_\infty = 15.75\,\text{b/s/Hz}$ for an unlimited number of antennas per UT. In Fig. 3.15, we show the achievable rates under both precoding techniques and their approximations by Theorems 31 and 33 as a function of the number of antennas N. Both results match very well, even for small N. We can observe that RZF achieves significant performance gains over BF as it reduces multiuser interference. As a rule of thumb, RZF allows us to reduce the number of antennas by one order of magnitude to achieve BF-performance. Nevertheless, even for $N = 400$ both precoders are far away from the ultimate performance limit.

Second, we consider a physical channel model with a fixed number of dimensions P as in [147]. For a uniform linear array, the matrices $\tilde{\mathbf{R}}_{jlk}$ are given as $\tilde{\mathbf{R}}_{jlk} = d_{jlk}^{-\beta/2}\,[\mathbf{A}\ \mathbf{0}_{N\times N-P}]$, where $\mathbf{A} = [\mathbf{a}(\phi_1)\cdots\mathbf{a}(\phi_P)] \in \mathbb{C}^{N\times P}$ is composed of the "steering vectors" $\mathbf{a}(\phi_p) \in \mathbb{C}^N$ defined as

$$\mathbf{a}(\phi_p) = \frac{1}{\sqrt{P}}\left[1, e^{-\mathbf{i}2\pi\omega\sin(\phi)}, \ldots, e^{-\mathbf{i}2\pi\omega(N-1)\sin(\phi)}\right]^{\mathsf{T}} \tag{3.67}$$

where ω is the antenna spacing in multiples of the wavelength and $\phi_p = -\pi/2 + (p-1)\pi/P$, $p = 1, \ldots, P$, are the uniformly distributed angles of transmission.

We assume that the physical dimensions P scale with the number of antennas as $P = N/2$ and let $\omega = 0.3$. Since $\frac{1}{N}\operatorname{tr}\mathbf{A}\mathbf{A}^{\mathsf{H}} = 1$, the ultimately achievable rate under this channel model and both precoding schemes is equal to that of a channel without antenna correlation. For comparison, we depict in Fig. 3.15 also the achievable rates and their approximations for this channel model. Interestingly, while the shapes of the curves for both precoders are similar to those without antenna correlation, it becomes clear that low rank correlation matrices severely degrade the performance.

3.3.6 Conclusions

We have provided a unified analysis of the UL/DL performance of linear detectors/precoders in multicell multiuser TDD systems. Assuming a large system limit, we have derived asymptotically tight approximations of achievable UL/DL-rates under a very general channel model which accounts for channel estimation, pilot contamination, path loss and individual antenna correlation. These approximations were shown to be accurate for realistic system dimensions and enable, consequently, future studies of realistic effects, such as antenna correlation, spacing and aperture, without the need for simulations. Our results are also directly applicable in the context of large distributed antenna systems. For a simplified channel model, we have observed that the performance depends mainly on the physical DoF per UT the channel offers and the effective SNR. Moreover, we have determined how many antennas are needed to achieve $\eta\,\%$ of the ultimate performance limit and how many more antennas are needed with MF/BF to achieve MMSE/RZF performance. Simulations for a more realistic system model suggest that MMSE/RZF can reduce the number of antennas by one order of magnitude to achieve the performance of the simple MF/BF schemes. Since massive MIMO TDD-systems are a promising network architecture, it seems necessary to verify the theoretical performance predictions by channel measurements and prototypes.

3.4 Double-scattering channels

Abstract: *We consider a MIMO MAC, where the channel between each transmitter and the receiver is modeled by the double-scattering channel model. Based on the concept of iterative deterministic equivalents as detailed in Section 2.4, we derive deterministic approximations of the mutual information, the SINR at the output of the MMSE detector and the sum-rate with MMSE detection, which are almost surely tight in the large system limit. Moreover, we derive the asymptotically optimal transmit covariance matrices. Our simulation results show that the asymptotic analysis provides very close approximations for realistic system dimensions.*

3.4.1 Introduction

Consider a discrete-time MIMO MAC from K transmitters, equipped with n_k, $k = 1, \dots, K$, antennas, respectively, to a receiver with N antennas. The channel output vector $\mathbf{y} \in \mathbb{C}^N$ reads

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{x}_k + \mathbf{n} \tag{3.68}$$

where $\mathbf{H}_k \in \mathbb{C}^{N\times n_k}$ and $\mathbf{x}_k \in \mathbb{C}^{n_k}$ are the channel matrix and the transmit vector associated with the kth transmitter, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_N)$ is a noise vector and $\rho = \frac{1}{\sigma^2} > 0$ denotes the SNR. We assume Gaussian signaling, i.e., $\mathbf{x}_k = [x_{k,1}, \ldots, x_{k,n_k}]^\mathsf{T} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_k)$, where $\mathbf{Q}_k \in \mathbb{C}^{n_k\times n_k}$. The channel matrices $\mathbf{H}_k$ are modeled by the double-scattering model [91]

$$\mathbf{H}_k = \frac{1}{\sqrt{N_k n_k}}\mathbf{R}_k^{\frac{1}{2}}\mathbf{W}_{1,k}\mathbf{S}_k^{\frac{1}{2}}\mathbf{W}_{2,k}\mathbf{T}_k^{\frac{1}{2}} \tag{3.69}$$

where $\mathbf{R}_k \in \mathbb{C}^{N\times N}$, $\mathbf{S}_k \in \mathbb{C}^{N_k\times N_k}$ and $\mathbf{T}_k \in \mathbb{C}^{n_k\times n_k}$ are deterministic correlation matrices, while $\mathbf{W}_{1,k} \in \mathbb{C}^{N\times N_k}$ and $\mathbf{W}_{2,k} \in \mathbb{C}^{N_k\times n_k}$ are independent standard complex Gaussian matrices. Since the distributions of $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$ are unitarily invariant, we can assume $\mathbf{S}_k = \mathrm{diag}(s_{k,1}, \ldots, s_{k,N_k})$ to be diagonal matrices, without loss of generality for the statistics of $\mathbf{y}$.

The double-scattering model [91] was motivated by the observation of low-rank channel matrices, despite low antenna correlation at the transmitter and receiver, see e.g., [152, 153]. A special case of the double-scattering model is the keyhole channel [154, 155], which exhibits null antenna correlation, i.e., $\mathbf{R}_k = \mathbf{I}_N$ and $\mathbf{T}_k = n_k$ for all k, but only a single degree of freedom. The existence of such channels (under laboratory conditions) was confirmed by measurements in [155]. Several theoretical works have studied the double-scattering model so far. The authors of [156] derive capacity upper-bounds for the general model and a closed-form expression for the keyhole channel. An asymptotic study of the outage capacity of the multi-keyhole channel was presented in [157]. The diversity order of the double-scattering model was considered in [158] and it was shown that a MIMO system with t transmit antennas, r receive antennas and s scatterers achieves the diversity of order $trs/\max(t, r, s)$. A closed-from expression of the diversity-multiplexing trade-off (DMT) was derived in [159]. Beamforming along the strongest eigenmode over Rayleigh product MIMO channels, i.e., the double-scattering model without any form of correlation, was considered in [160]. Here, the authors derive exact expressions of the cumulative distribution function (cdf) and the probability density function (pdf) of the largest eigenvalue of the Gramian of the channel matrix and compute closed-form results for the ergodic capacity, outage probability and SINR distribution. In a later paper [161], the MIMO MAC with double-scattering fading is analyzed. The authors obtain closed-form upper-bounds on the sum-capacity and prove that the transmitters should send their signals along the eigenvectors of the transmit correlation matrices in order to achieve capacity. Despite the significant interest in the double-scattering channel model, little work has been done to study its asymptotic performance when the channel dimensions grow large. We are only aware of [153], in which a model without transmit and receive correlation is studied relying on tools from free probability theory. Implicit expressions of the asymptotic mutual information and the SINR with MMSE detection are found therein.

In the following, we provide deterministic equivalents of the mutual information, the SINR with MMSE-detection and the associated sum-rate. In addition, we derive the precoders which maximize the deterministic equivalent of the mutual information and provide a simple algorithm for their computation. The key idea behind the following proofs is that the double-scattering channel can be interpreted as a Kronecker channel [83] with a *random* receive correlation matrix, which itself is modeled by the Kronecker model. This observation allows us to build upon [83] which provides an asymptotic analysis of the performance of Kronecker channels with deterministic correlation matrices (see Theorem 13). Based on the Fubini theorem, we extend this work by allowing the correlation matrices to be random. A similar technique can be applied to more involved channel models, such as channels with line-of-sight components or MIMO product channels with an arbitrary number of matrices.

Denote $I_N(\sigma^2)$ the instantaneous normalized mutual information of the channel (3.68), defined as [42]

$$I_N\left(\sigma^2\right) = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k=1}^{K}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^{\mathsf{H}}\right). \tag{3.70}$$

Moreover, denote $\gamma_{k,j}^{N}\left(\sigma^2\right)$ the SINR at the output of the MMSE detector related to the transmit symbol $x_{k,j}$, given by [162]

$$\gamma_{k,j}^{N}\left(\sigma^2\right) = \mathbf{h}_{k,j}^{\mathsf{H}}\left(\sum_{i=1}^{K}\mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} - \mathbf{h}_{k,j}\mathbf{h}_{k,j}^{\mathsf{H}} + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{h}_{k,j}. \tag{3.71}$$

We define the normalized sum-rate $R_N\left(\sigma^2\right)$ with MMSE detection as

$$R_N\left(\sigma^2\right) = \frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{n_k}\log\left(1+\gamma_{k,j}^{N}\left(\sigma^2\right)\right). \tag{3.72}$$

3.4.2 Asymptotic analysis

The notation "$N \to \infty$" will be used to denote that N and all N_k, n_k grow infinitely large, satisfying $0 < \liminf \frac{N_k}{N} \leq \limsup \frac{N_k}{N} < \infty$ and $0 < \liminf \frac{n_k}{N} \leq \limsup \frac{n_k}{N} < \infty$. Additionally, we need the following technical assumption:

A 4. $\limsup_N \|\mathbf{R}_k\| < \infty$*,* $\limsup_N \|\mathbf{S}_k\| < \infty$*,* $\limsup_N \|\mathbf{T}_k\mathbf{Q}_k\| < \infty\ \forall k$*.*

Remark 20. *This assumption implies in particular that the antenna correlation at the transmitter and receiver side cannot grow with the system size, as it would be the case for very dense antenna arrays [163]. Amendments to relax this assumption can be made, following the work in [83]. Moreover, the last constraint,* $\limsup_N \|\mathbf{T}_k\mathbf{Q}_k\| < \infty$*, implies that no transmitter is allowed to focus an increasing amount of transmit power in a single direction.*

Theorem 22 (i) introduces a set of $3K$ implicit equations which uniquely determines some quantities $(g_k, \bar{g}_k, \delta_k)$ $(1 \leq k \leq K)$. These are needed later on to provide deterministic equivalents of $I_N(\sigma^2)$, $\gamma_{k,j}^{N}\left(\sigma^2\right)$, and $R_N\left(\sigma^2\right)$. A deterministic equivalent of the (ergodic) mutual information is provided by Theorem 22 (ii). The following result allows us to compute the asymptotically optimal precoding matrices $\mathbf{Q}_k$ which maximize $\bar{I}_N\left(\sigma^2\right)$ under individual transmit power constraints.

Theorem 34 (Optimal power allocation)**.** *The solution to the following optimization problem:*

$$\begin{aligned}\left(\bar{\mathbf{Q}}_1^{\star},\ldots,\bar{\mathbf{Q}}_K^{\star}\right) &= \arg\max_{\mathbf{Q}_1,\ldots,\mathbf{Q}_k} \bar{I}_N\left(\sigma^2\right)\\ &\quad s.t. \quad \frac{1}{n_k}\operatorname{tr}\mathbf{Q}_k \leq P_k\ \forall k\end{aligned}$$

where $\bar{I}_N(\sigma^2)$ *is defined in Theorem 22, is given as* $\bar{\mathbf{Q}}_k^\star = \mathbf{U}_k \bar{\mathbf{P}}_k^\star \mathbf{U}_k^\mathsf{H}$*, where* $\mathbf{U}_k \in \mathbb{C}^{n_k \times n_k}$ *is defined by the spectral decomposition of* $\mathbf{T}_k = \mathbf{U}_k \mathrm{diag}(t_{k,1}, \ldots, t_{k,n_k}) \mathbf{U}_k^\mathsf{H}$ *and* $\bar{\mathbf{P}}_k^\star = \mathrm{diag}(\bar{p}_{k,1}^\star, \ldots, \bar{p}_{k,n_k}^\star)$ *is given by the water-filling solution:*

$$\bar{p}_{k,j}^\star = \left(\mu_k - \frac{1}{g_k^\star t_{k,j}}\right)^+ \tag{3.73}$$

where μ_k *is chosen to satisfy* $\frac{1}{n_k} \operatorname{tr} \bar{\mathbf{P}}_k^\star = P_k$ *and* $g_k^\star = g_k$ *is given by Theorem 22 (i) for* $\mathbf{Q}_k = \bar{\mathbf{Q}}_k^\star$.

Proof. The proof is provided in Appendix 3.7.6. ☐

Remark 21. *The optimal power allocation matrices* $\bar{\mathbf{P}}_k^\star$ *can be calculated by the iterative water-filling Algorithm 1 (see [83, Remark 2] and [85, Remark 3] for a discussion on the convergence of this algorithm).*

Algorithm 1 Iterative water-filling algorithm

1: Let $\epsilon > 0$, $n = 0$ and $\bar{p}_{k,j}^{\star,0} = P_k$ for all k, j.
2: **repeat**
3: For all k, compute $g_k^{\star,n} = g_k$ according to Theorem 22 (i) with matrices $\mathbf{Q}_k = \mathbf{U}_k \mathrm{diag}\left(\bar{p}_{k,j}^{\star,n}\right) \mathbf{U}_k^\mathsf{H}$.
4: For all k, j, calculate $\bar{p}_{k,j}^{\star,n+1} = \left(\mu_k - \frac{1}{g_k^{\star,n} t_{k,j}}\right)^+$, with μ_k such that $\frac{1}{n_k}\sum_{j=1}^{n_k} \bar{p}_{k,j}^{\star,n+1} = P_k$.
5: $n = n + 1$
6: **until** $\max_{k,j} |\bar{p}_{k,j}^{\star,n} - \bar{p}_{k,j}^{\star,n-1}| \le \epsilon$

Remark 22. *Denote by* $(\mathbf{Q}_1^\star, \ldots, \mathbf{Q}_K^\star)$ *the precoding matrices which maximize* $\mathbb{E}\left[I_N(\sigma^2)\right]$ *for a given set of power constraints. If the condition* $\limsup \|\mathbf{T}_k \mathbf{Q}_k^\star\| < \infty$ *holds for all* k*, then* $\mathbb{E}\left[I_N(\sigma^2, \mathbf{Q}_1^\star, \ldots, \mathbf{Q}_K^\star)\right] - \bar{I}_N(\sigma^2, \bar{\mathbf{Q}}_1^\star, \ldots, \bar{\mathbf{Q}}_K^\star) \to 0$*, by Theorem 22 and the strict concavity of* $\bar{I}_N(\sigma^2)$ *and* $I_N(\sigma^2)$ *in the matrices* $\mathbf{Q}_k$*. However, this condition is difficult to verify and is outside the scope of this thesis. See [164] for such a technical discussion in the case of Rician fading channels.*

Next, we provide deterministic equivalents of the SINR $\gamma_{k,j}^N(\sigma^2)$ at the output of the MMSE detector and the associated sum-rate $R_N(\sigma^2)$.

Theorem 35 (SINR of the MMSE detector)**.** *Let* $\mathbf{Q}_k = \mathrm{diag}(p_{k,1}, \ldots, p_{k,n_k})$ *and* $\mathbf{T}_k = \mathrm{diag}(t_{k,1}, \ldots, t_{k,n_k})$ *for all* k*. Assume that* ***A 4*** *holds. Then,*

$$\gamma_{k,j}^N(\sigma^2) - \bar{\gamma}_{k,j}^N(\sigma^2) \xrightarrow{\text{a.s.}} 0$$

where $\bar{\gamma}_{k,j}^N(\sigma^2) = p_{k,j} t_{k,j} g_k$ *and* g_k *is by given by Theorem 22 (i).*

Proof. The proof is provided in Appendix 3.7.7. ☐

Remark 23. *Note that the theorem is also valid under the more general assumption* $\mathbf{T}_k = \mathbf{U}_k \mathrm{diag}(t_{k,1}, \ldots, t_{k,n_k}) \mathbf{U}_k^\mathsf{H}$ *and* $\mathbf{Q}_k = \mathbf{U}_k \mathrm{diag}(p_{k,1}, \ldots, p_{k,n_k}) \mathbf{U}_k^\mathsf{H}$.

Corollary 5 (Sum-rate with MMSE decoding)**.** *Let* $\mathbf{Q}_k = \operatorname{diag}(p_{k,1}, \ldots, p_{k,n_k})$ *and* $\mathbf{T}_k = \operatorname{diag}(t_{k,1}, \ldots, t_{k,n_k})$ *for all* k*. Assume that* ***A 4*** *holds. Then,*

$$\begin{aligned} (i) \quad & R_N\left(\sigma^2\right) - \bar{R}_N(\sigma^2) \xrightarrow{\text{a.s.}} 0 \\ (ii) \quad & \mathbb{E}\left[R_N\left(\sigma^2\right)\right] - \bar{R}_N(\sigma^2) \xrightarrow{\text{a.s.}} 0 \end{aligned}$$

where

$$\bar{R}_N(\sigma^2) = \frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{n_k} \log\left(1 + \bar{\gamma}_{k,j}^N\left(\sigma^2\right)\right)$$

and the $\bar{\gamma}_{k,j}^N\left(\sigma^2\right)$ *are given by Theorem 35.*

Proof. The proof is provided in Appendix 3.7.8. □

Remark 24. *Careful inspection of Theorem 22 (ii) reveals that the third term of* $\bar{I}_N\left(\sigma^2\right)$ *equals* $\bar{R}_N(\sigma^2)$ *since*

$$\frac{1}{N}\sum_{k=1}^{K} \log\det\left(\mathbf{I}_{n_k} + g_k \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}}\right) = \frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{n_k} \log\left(1 + p_{k,j} t_{k,j} g_k\right). \quad (3.74)$$

Thus, all other terms in $\bar{I}_N\left(\sigma^2\right)$ *correspond consequently to the gains of successive interference cancellation (SIC) [15] over simple MMSE detection.*

A special case of the double-scattering channel is the Rayleigh product MIMO channel [160] which does not exhibit any form of correlation between the transmit and receive antennas or the scatterers. For this model, Theorems 22 (i), (ii), and 35 can be given in closed form as shown in the next corollary.

Corollary 6 (Rayleigh product channel)**.** *For all* k*, let* $N_k = S$*,* $n_k = N$ *and assume* $\mathbf{T}_k = \mathbf{I}_N$*,* $\mathbf{S}_k = \mathbf{I}_S$*,* $\mathbf{R}_k = \mathbf{I}_N$*, and* $\mathbf{Q}_k = \mathbf{I}_N$*. Then* $\bar{I}_N\left(\sigma^2\right)$ *and* $\bar{\gamma}_{k,j}^N\left(\sigma^2\right)$ *as defined in Theorems 22 and 35 can be given in closed form as*

$$\begin{aligned} \bar{I}_N\left(\sigma^2\right) = & \log\left(1 + \frac{NK}{\sigma^2 S}\bar{g}\left(\bar{g} + \frac{S}{N} - 1\right)\right) - \frac{KS}{N}\log\left(1 + \frac{N}{S}\left(\bar{g} - 1\right)\right) \\ & - K\log\left(\bar{g}\right) - 2K\left(1 - \bar{g}\right) \end{aligned}$$

and

$$\bar{\gamma}_{k,j}^N\left(\sigma^2\right) = \frac{1 - \bar{g}}{\bar{g}}$$

where $\bar{g}$ *is the unique root to*

$$\bar{g}^3 - \bar{g}^2\left(2 - \frac{S}{N} - \frac{1}{K}\right) + \bar{g}\left(1 - \frac{S}{N} - \frac{1}{K} + \frac{S}{NK}\left(1 + \sigma^2\right)\right) - \frac{S}{NK}\sigma^2 = 0 \quad (3.75)$$

such that $\bar{g} \in \left(1 - \min\left[\frac{1}{K}, \frac{S}{N}\right], 1\right)$.

Proof. The proof is provided in Appendix 3.7.9. □

Note that similar expressions for the asymptotic mutual information and MMSE-SINR have been obtained in [153] by means of free probability theory. However, these results require the numerical solution of a third order differential equation.

3.4.3 Numerical results

As a first numerical example, we consider the “multi-keyhole channel”, i.e., $K = 1$, $\mathbf{S}_1 = \mathbf{I}_{N_1}$, $\mathbf{R}_1 = \mathbf{I}_N$, $\mathbf{T}_1 = \mathbf{Q}_1 = \mathbf{I}_{n_1}$, for $N = n_1 = 4$. Fig. 3.16 depicts the normalized ergodic mutual information $\mathbb{E}\left[I_N(\sigma^2)\right]$ and its asymptotic approximation $\bar{I}_N(\sigma^2)$ versus SNR for different numbers of “keyholes” $N_1 \in \{1, 2, 3, 4, 100\}$. Surprisingly, the match between both results is almost perfect although the channel dimensions are very small. As one expects, the multiplexing gain increases linearly with N_1 until $N_1 \geq N = 4$. Larger values of N_1 only change the statistical distribution of the channel matrix while the degrees of freedom are limited by the number of antennas (for $N_1 \to \infty$, $\mathbf{H}_1$ becomes a standard Rayleigh fading channel [91]).

As a second example, we consider a MAC from $K = 3$ transmitters, assuming the double-scattering model in [91]. Under this model, the correlation matrices are given as $\mathbf{R}_k = \mathbf{G}(\phi_{r,k}, d_{r,k}, N_k)$, $\mathbf{S}_k = \mathbf{G}(\phi_{s,k}, d_{s,k}, N_k)$ and $\mathbf{T}_k = \mathbf{G}(\phi_{t,k}, d_{t,k}, N_k)$, where $\mathbf{G}(\phi, d, n)$ is defined as

$$[\mathbf{G}(\phi, d, n)]_{k,l} = \frac{1}{n} \sum_{j=\frac{1-n}{2}}^{\frac{n-1}{2}} \exp\left(\mathbf{i}2\pi d(k-l)\sin\left(\frac{j\phi}{1-n}\right)\right). \tag{3.76}$$

The values $\phi_{t,k}$ and $\phi_{r,k}$ determine the angular spread of the radiated and received signals, $d_{t,k}$ and $d_{r,k}$ are the antenna spacings at the kth transmitter and receiver in multiples of the signal wavelength, N_k can be seen as the number of scatterers and $d_{s,k}$ as the spacing of the scatterers. For simplicity, we assume $N = 4$, $P_k = 1/n_k$, $N_k = 11$, $n_k = 3$, $d_{t,k} = d_{r,k} = 0.25$ and $d_{s,k} = 50$ for all k. We further assume $\phi_{r,k} = \phi_{t,k}$ for all k, with $\phi_{r,k} \in \{\pi/4, \pi/2, \pi\}$ and $\phi_{s,k} = \pi/8$. Fig. 3.17 shows $\mathbb{E}\left[I_N(\sigma^2)\right]$ and $\bar{I}_N(\sigma^2)$ with uniform and optimal power allocation versus SNR. Again, our asymptotic results yield very tight approximations, even for small system dimensions. Note that we have used the precoding matrices provided by Theorem 34 for the simulations as the optimal precoding matrices are unknown. For comparison, we also provide the sum-rate with MMSE detection $\mathbb{E}\left[R_N(\sigma^2)\right]$ and its deterministic approximation $\bar{R}_N(\sigma^2)$. We observe a good fit between both results at low SNR values, but a slight mismatch for higher values. This is due to a slower convergence of the SINR $\gamma_{k,j}^N(\sigma^2)$ to its deterministic approximation $\bar{\gamma}_{k,j}^N(\sigma^2)$, well documented in the RMT literature, e.g., [52].

3.4.4 Conclusions

In this section, we have applied the concept of iterative deterministic equivalents to the asymptotic analysis of a MIMO MAC under the double-scattering channel model. We have derived asymptotically tight deterministic approximations of information theoretic quantities of interest, such as the mutual information and the sum-rate with MMSE detection. These approximations can be easily computed by provably converging fixed-point algorithms and do not require any numerical integration. As a special case, we have consider the Rayleigh product channel in which all results are given in closed-form. Moreover, we have provided the asymptotically optimal precoding matrices and an iterative water-filling algorithm for their computation. Our simulation results suggest

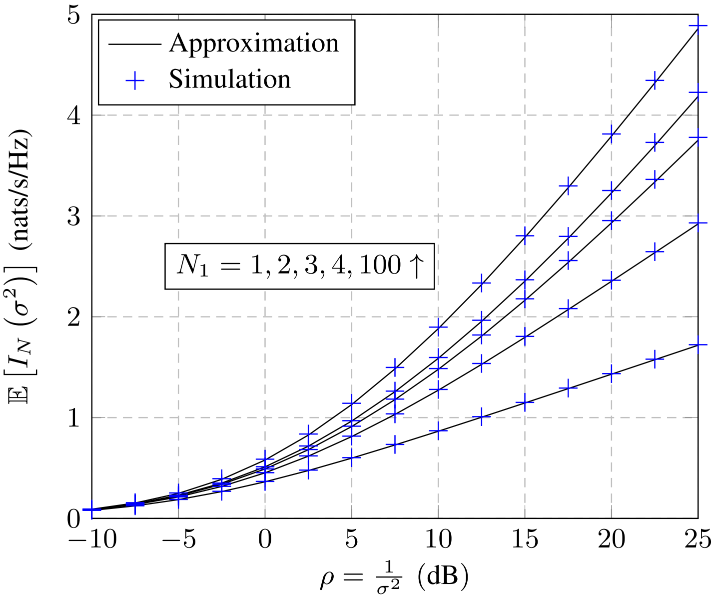

Figure 3.16: Ergodic mutual information $\mathbb{E}\left[I_N(\sigma^2)\right]$ of the multi-keyhole channel and its deterministic equivalent $\bar{I}_N(\sigma^2)$ versus $\rho = \frac{1}{\sigma^2}$.

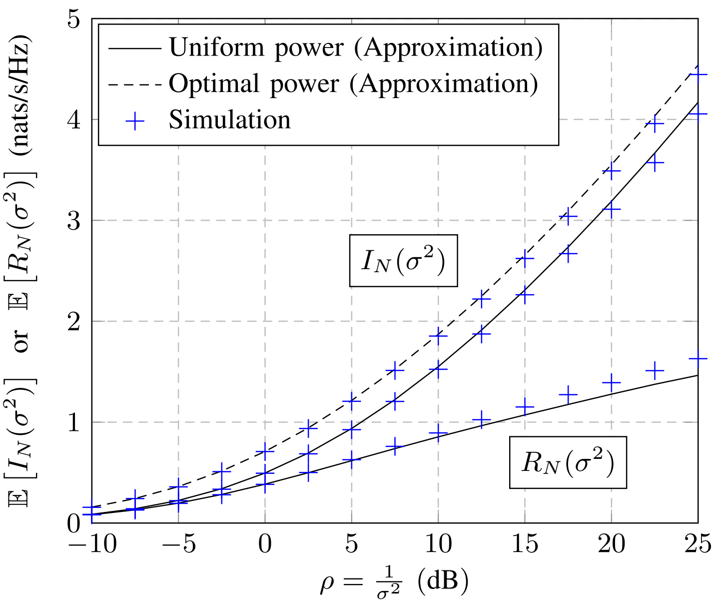

Figure 3.17: Ergodic mutual information $\mathbb{E}\left[I_N(\sigma^2)\right]$ and sum-rate $\mathbb{E}\left[R_N(\sigma^2)\right]$ of the multiple access channel and their asymptotic approximations $\bar{I}_N(\sigma^2)$ and $\bar{R}_N(\sigma^2)$ versus $\rho = \frac{1}{\sigma^2}$.

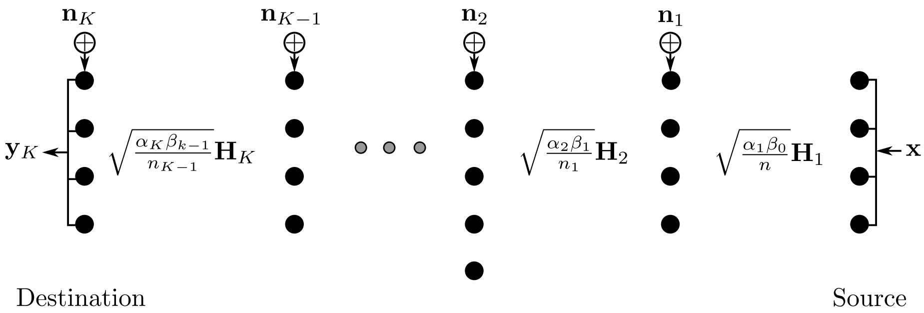

Figure 3.18: Multi-hop amplify-and-forward MIMO relay channel.

that the asymptotic performance approximations are very accurate for finite system dimensions with only a few antennas at each node.

3.5 Multihop amplify-and-forward MIMO relay channels

Abstract: *We derive a deterministic equivalent for the mutual information of the K-hop AF MIMO relay channel, assuming that the number of antennas at each node grows infinitely large. In contrast to previous works, we consider noise at each of the relays such that the system performance rapidly diminishes as K grows. The analysis is facilitated by the concept of iterative deterministic equivalents and fundamentally based on Theorem 24. The result is given by a simple recursive algorithm and shown to provide tight performance approximations for a small number of antennas at each node.*

3.5.1 Introduction and system model

Consider a multi-hop AF MIMO relay channel where a source node communicates via $K-1$ relays with a destination node. There is no direct link between the source and the destination and each relay can only receive data from the preceding hop. This is for example the case if the nodes follow a time-division multiple access (TDMA) protocol where only one node is transmitting at any given time and the path loss between relay k and $k-2$ is large. Thus each data symbol reaches the destination after K channel uses. The source and destination are respectively equipped with n and n_K antennas while the kth relay has n_k antennas. The relays operate an AF-protocol where each node simply transmits a scaled version of its received signal to the next hop. We will consider a large system limit where $n, n_1, \ldots, n_K$ grow infinitely large at the same speed. Define the following quantities:

$$\begin{aligned} c_1 &= \frac{n}{n_1} \\ c_k &= \frac{n_{k-1}}{n_k}, \qquad k = 2, \ldots, K. \end{aligned} \tag{3.77}$$

The notation "$n \to \infty$" must be understood from now on as $n \to \infty$, such that $0 < \liminf_n c_k \leq \limsup_n c_k < \infty$ for all k. We denote $\mathbf{y}_k \in \mathbb{C}^{n_k}$ the received base-band signal vector at the kth hop, given by

$$\begin{aligned}\mathbf{y}_1 &= \sqrt{\alpha_1}\mathbf{H}_1\sqrt{\frac{\beta_0}{n}}\mathbf{x}+\mathbf{n}_1\\ \mathbf{y}_k &= \sqrt{\alpha_k}\mathbf{H}_k\sqrt{\frac{\beta_{k-1}}{n_{k-1}}}\mathbf{y}_{k-1}+\mathbf{n}_k, \qquad k=2,\ldots,K \end{aligned} \tag{3.78}$$

where $\mathbf{H}_k \in \mathbb{C}^{n_k\times n_{k-1}}$ is a standard complex Gaussian matrix (let $n_0 \triangleq n$), $\mathbf{x} \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_n)$ is the channel input vector, $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_{n_k})$ is a noise vector, α_k is a path loss factor, and the parameter β_k is chosen to normalize the transmit power of the kth node according to its power budget $\rho_k > 0$, i.e.,

$$\begin{aligned}\beta_0 &= \frac{\rho_0}{\frac{1}{n}\operatorname{tr}\mathbb{E}\left[\mathbf{x}\mathbf{x}^{\mathsf{H}}\right]} = \rho_0\\ \beta_k &= \frac{\rho_k}{\frac{1}{n_k}\operatorname{tr}\mathbb{E}\left[\mathbf{y}_k\mathbf{y}_k^{\mathsf{H}}\right]}, \qquad k=1,\ldots,K-1.\end{aligned} \tag{3.79}$$

The expectation in the last equation is with respect to the transmit and noise vectors only.[6] The channel matrices $\mathbf{H}_k$ and path loss factors α_k are assumed to be known to the relays and the destination. Since the received signal at each relay is corrupted by noise, the system suffers from noise accumulation. This is in addition to the linear rate loss $\frac{1}{K}$ related to the TDMA protocol. Thus, the capacity decreases quickly with the number of hops K. Note that our system model is different from existing works which consider either no noise [165], or noise only at the destination [44]. An exception is [45], in which the authors consider a similar system model, but do not provide closed-form expressions of the asymptotic mutual information. Several other works deal with the asymptotic capacity of the dual-hop relay channel [166, 167]. Recently, an exact expression of the mutual information of the dual-hop channel for finite channel dimensions was derived in [168]. Here, we will provide an explicit deterministic equivalent of the mutual information at each relay for the general model (3.78).

Let us introduce the following, recursively defined matrices $\mathbf{R}_k\left(\boldsymbol{\beta}_{k-1}\right)$:

$$\begin{aligned}\mathbf{R}_0 &= \mathbb{E}\left[\mathbf{x}\mathbf{x}^{\mathsf{H}}\right] = \mathbf{I}_n\\ \mathbf{R}_k\left(\boldsymbol{\beta}_{k-1}\right) &= \mathbb{E}\left[\mathbf{y}_k\mathbf{y}_k^{\mathsf{H}}\right] = \mathbf{I}_{n_k} + \frac{\alpha_k\beta_{k-1}}{n_{k-1}}\mathbf{H}_k\mathbf{R}_{k-1}\left(\boldsymbol{\beta}_{k-2}\right)\mathbf{H}_k^{\mathsf{H}}, \qquad k=1,\ldots,K\end{aligned} \tag{3.80}$$

and the functionals $\mathcal{J}_k\left(x,\boldsymbol{\beta}_{k-1}\right)$, $x>0$, $k\in\{1,\ldots,K\}$, which are defined as

$$\mathcal{J}_k\left(x,\boldsymbol{\beta}_{k-1}\right) = \frac{1}{n_k}\log\det\left(\mathbf{I}_{n_k} + x\frac{\alpha_k\beta_{k-1}}{n_{k-1}}\mathbf{H}_k\mathbf{R}_{k-1}\left(\boldsymbol{\beta}_{k-2}\right)\mathbf{H}_k^{\mathsf{H}}\right) \tag{3.81}$$

where $\boldsymbol{\beta}_k = [\beta_0,\cdots,\beta_k]$. With these definitions, we can express the normalized mutual information $\mathcal{I}_k(\boldsymbol{\beta}_{k-1})$ between $\mathbf{y}_k$ and $\mathbf{x}$ as

$$\mathcal{I}_k(\boldsymbol{\beta}_{k-1}) = \frac{1}{K}\left(\mathcal{J}_k(1,\boldsymbol{\beta}_{k-1}) - \mathcal{J}_k(1,\boldsymbol{\beta}'_{k-1})\right) \tag{3.82}$$

where $\boldsymbol{\beta}'_k = [0,\beta_1,\cdots,\beta_k]$. Next, we demonstrate by a simple example that (3.82) holds.

[6] Under a long-term power constraint, the expectation could be taken also with respect to the matrices $\mathbf{H}_k$. Asymptotically, both constraints are equivalent (see Lemma 21).

Example 1 (2-hop Relay-channel)**.** *The normalized mutual information $\mathcal{I}_2(\boldsymbol{\beta}_1)$ between $\mathbf{x}$ and the channel output after the second hop $\mathbf{y}_2$ is given as*

$$\begin{aligned}
\mathcal{I}_2(\boldsymbol{\beta}_1) &= \frac{1}{Kn_2}\log\det\left(\mathbf{I}_{n_2} + \left(\mathbf{I}_{n_2} + \frac{\alpha_2\beta_1}{n_1}\mathbf{H}_2\mathbf{H}_2^{\mathsf{H}}\right)^{-1}\frac{\alpha_2\beta_1\alpha_1\beta_0}{n_1 n}\mathbf{H}_2\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}}\mathbf{H}_2^{\mathsf{H}}\right) \\
&= \frac{1}{Kn_2}\log\det\left(\mathbf{I}_{n_2} + \frac{\alpha_2\beta_1}{n_1}\mathbf{H}_2\left(\mathbf{I}_n + \frac{\alpha_1\beta_0}{n}\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}}\right)\mathbf{H}_2^{\mathsf{H}}\right) \\
&\quad - \frac{1}{Kn_2}\log\det\left(\mathbf{I}_{n_2} + \frac{\alpha_2\beta_1}{n_1}\mathbf{H}_2\mathbf{H}_2^{\mathsf{H}}\right) \\
&= \frac{1}{Kn_2}\log\det\left(\mathbf{I}_{n_2} + \frac{\alpha_2\beta_1}{n_1}\mathbf{H}_2\mathbf{R}_1(\beta_0)\mathbf{H}_2^{\mathsf{H}}\right) \\
&\quad - \frac{1}{Kn_2}\log\det\left(\mathbf{I}_{n_2} + \frac{\alpha_2\beta_1}{n_1}\mathbf{H}_2\mathbf{R}_1(0)\mathbf{H}_2^{\mathsf{H}}\right) \\
&= \frac{1}{K}\left(\mathcal{J}_2(1,\boldsymbol{\beta}_1) - \mathcal{J}_2(1,\boldsymbol{\beta}_1')\right). \qquad (3.83)
\end{aligned}$$

3.5.2 Asymptotic analysis

In Theorem 24, we have derived deterministic equivalents $\bar{\mathcal{J}}_k(x,\bar{\boldsymbol{\beta}}_{k-1})$ of the quantities $\mathcal{J}_k(x,\boldsymbol{\beta}_{k-1})$. The recursive definition of the matrices $\mathbf{R}_k(\boldsymbol{\beta}_{k-1})$ in (3.80) allows us to calculate iterative deterministic equivalents of the mutual information after each hop. This is achieved by treating the matrix $\mathbf{R}_{k-1}(\boldsymbol{\beta}_{k-2})$ as deterministic and deriving a deterministic equivalent of $\mathcal{J}_k(x,\boldsymbol{\beta}_{k-1})$ with respect to the matrix $\mathbf{H}_k$. This process can be iterated for $\mathbf{R}_{k-2}(\boldsymbol{\beta}_{k-3}), \mathbf{R}_{k-3}(\boldsymbol{\beta}_{k-4}), \ldots$ and $\mathbf{H}_{k-1}, \mathbf{H}_{k-2}, \ldots$ until the deterministic matrix $\mathbf{R}_0$ is reached. Let us first derive deterministic equivalents $\bar{\beta}_k$ of the power normalization factors β_k:

Lemma 21 (Asymptotic power normalization)**.** *Let $\beta_0 = \bar{\beta}_0 = \rho_0$. Then,*

$$\beta_k \xrightarrow[n\to\infty]{a.s.} \bar{\beta}_k = \frac{\rho_k}{1+\alpha_k\rho_{k-1}}, \qquad k = 1,\ldots,K-1.$$

Proof. Recall the definition of $\beta_k = \frac{\rho_k}{\frac{1}{n_k}\operatorname{tr}\mathbf{R}_k}$, where $\mathbf{R}_k = \mathbf{R}_k\left(\boldsymbol{\beta}_{k-1}\right)$. For $k \geq 1$, we have

$$\begin{aligned}
\frac{1}{n_k}\operatorname{tr}\mathbf{R}_k &= 1 + \frac{\alpha_k\beta_{k-1}}{n_k n_{k-1}}\operatorname{tr}\mathbf{H}_k\mathbf{R}_{k-1}\mathbf{H}_k^{\mathsf{H}} \\
&\overset{(a)}{=} 1 + \alpha_k\beta_{k-1}\frac{1}{n_k}\sum_{j=1}^{n_k}\frac{1}{n_{k-1}}\tilde{\mathbf{h}}_{k,j}^{\mathsf{H}}\mathbf{R}_{k-1}\tilde{\mathbf{h}}_{k,j} \\
&\overset{(b)}{\asymp} 1 + \alpha_k\frac{\rho_{k-1}}{\frac{1}{n_{k-1}}\operatorname{tr}\mathbf{R}_{k-1}}\frac{1}{n_{k-1}}\operatorname{tr}\mathbf{R}_{k-1} \\
&= 1 + \alpha_k\rho_{k-1} \qquad (3.84)
\end{aligned}$$

where (a) is obtained by denoting $\tilde{\mathbf{h}}_{k,j} \in \mathbb{C}^{n_{k-1}}$ the jth row vector of $\mathbf{H}_k$ and (b) is due to Lemma 13 and Lemma 19 and the definition of β_{k-1}. By the continuous mapping theorem, we finally have

$$\beta_k = \frac{\rho_k}{\frac{1}{n_k}\operatorname{tr}\mathbf{R}_k} \xrightarrow[n\to\infty]{\text{a.s.}} \frac{\rho_k}{1+\alpha_k\rho_{k-1}}. \qquad (3.85)$$

□

Applying Theorem 24 (*ii*) and Lemma 21 to (3.82) yields the following corollary which provides a deterministic equivalent of the mutual information $\mathcal{I}_k(\boldsymbol{\beta}_{k-1})$:

Corollary 7 (Asymptotic mutual information of the K-hop AF MIMO Relay channel)**.**

$$\mathcal{I}_k\left(\boldsymbol{\beta}_{k-1}\right) - \bar{\mathcal{I}}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right) \xrightarrow[n\to\infty]{a.s} 0, \qquad k = 1, \ldots, K$$

where

$$\bar{\mathcal{I}}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right) = \frac{1}{K}\left(\bar{\mathcal{J}}_k(1, \bar{\boldsymbol{\beta}}_{k-1}) - \bar{\mathcal{J}}_k(1, \bar{\boldsymbol{\beta}}'_{k-1})\right)$$

with $\bar{\boldsymbol{\beta}}_{k-1} = [\bar{\beta}_0 \cdots \bar{\beta}_{k-1}]$, $\bar{\boldsymbol{\beta}}'_{k-1} = [0\ \bar{\beta}_1 \cdots \bar{\beta}_{k-1}]$ *as given by Lemma 21, and* $\bar{\mathcal{J}}_k(x, \bar{\boldsymbol{\beta}}_{k-1})$ *and* $\bar{\mathcal{J}}_k(x, \bar{\boldsymbol{\beta}}'_{k-1})$ *as given by Theorem 24* (*ii*).

3.5.3 Numerical results

We would now like to verify our analysis by some numerical results. To this end, we consider a system with three relays, i.e., $K = 4$. We assume $n = n_4 = 4$, $n_1 = n_3 = 8$, $n_2 = 12$, $\rho_1 = \rho_3 = 0.7\rho_0$ and $\rho_2 = 0.5\rho_0$. The last assumption allows us to control the transmit power of all nodes by the transmit SNR ρ_0 of the source node. We further assume the path loss factors $\alpha_1 = 1$, $\alpha_2 = \alpha_4 = 0.7$, $\alpha_3 = 0.5$. Fig. 3.19 shows the average normalized mutual information $\mathbb{E}\left[\frac{n_k}{n}\mathcal{I}_k\left(\boldsymbol{\beta}_{k-1}\right)\right]$ after each hop ($k = 1, \ldots, 4$) versus the transmit power ρ_0 of the source node. Note that we have re-normalized all results by $\frac{n_k}{n}$ to put them on a common ground for comparison. The deterministic equivalents $\frac{n_k}{n}\bar{\mathcal{I}}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)$ as provided by Corollary 7 are drawn by solid lines, simulation results are represented by markers. The error bars represent one standard deviation of the simulation results in each direction. We can observe a very good fit between the asymptotic approximations and the simulation results for all k and the entire range of ρ_0. As expected, the performance decreases rapidly with each hop.

3.5.4 Conclusions

In this section, we have used the method of iterative deterministic equivalents for the performance analysis of a multihop AF MIMO Relay channel with noise at each relay. This approach has allowed us to derive easily computable recursive expressions for the asymptotic mutual information after each hop. We have verified our analysis by simulations which also demonstrate the accuracy of our approximations for small system dimensions. Finally, we would like to remark that, although we have considered a rather simple channel model with neither antenna correlation nor precoding at the nodes, more involved channel models can be treated in a straightforward fashion with the same techniques.

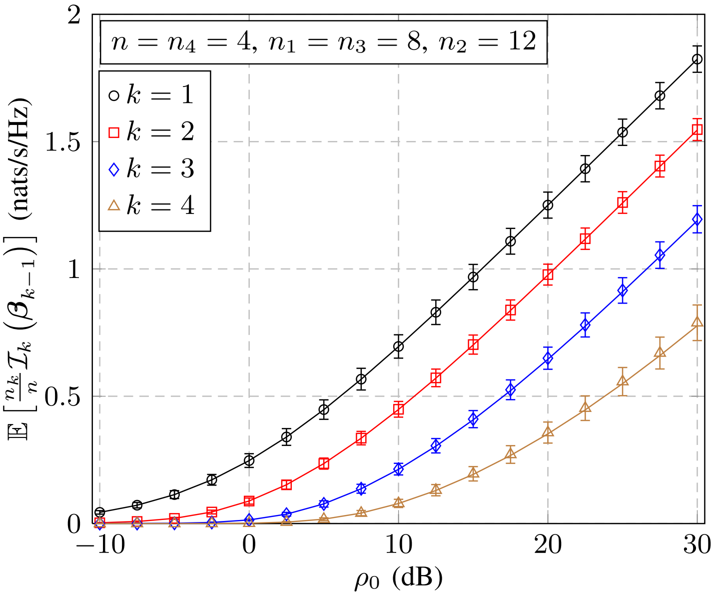

Figure 3.19: Average normalized mutual information $\mathbb{E}\left[\frac{n_k}{n}\mathcal{I}_k(\boldsymbol{\beta}_{k-1})\right]$ after the kth hop versus the transmit SNR ρ_0 of the source node. The deterministic equivalents $\frac{n_k}{n}\bar{\mathcal{I}}_k\left(\bar{\boldsymbol{\beta}}_{k-1}\right)$ are drawn by solid lines, the simulation results by markers. The error bars correspond to one standard deviation of the simulation results in each direction.

3.6 Random beamforming over quasi-static and fading channels

Abstract: *In this section, we study the performance of random isometric precoding over quasi-static and correlated fading channels. We derive deterministic approximations of the mutual information and the SINR at the output of the MMSE receiver and provide simple provably converging fixed-point algorithms for their computation. Although these approximations are only proven exact in the asymptotic regime with infinitely many antennas at the transmitters and receivers, simulations suggest that they closely match the performance of small-dimensional systems. In contrast to previous works, our analysis does not rely on arguments from free probability theory which allows us to consider random matrix models for which asymptotic freeness does not hold. Thus, our results are also a novel contribution to the field of random matrix theory and are shown to be applicable to a wide spectrum of practical systems. We specifically characterize the performance of multi-cellular communication systems, MIMO MACs, and MIMO interference channels.*

3.6.1 Introduction

Consider the following discrete time wireless channel model

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{W}_k \mathbf{P}_k^{\frac{1}{2}} \mathbf{x}_k + \mathbf{n} \tag{3.86}$$

where

(i) $\mathbf{y} \in \mathbb{C}^N$ is the channel output vector,

(ii) $\mathbf{H}_k \in \mathbb{C}^{N \times N_k}$, $k \in \{1, \ldots, K\}$, are complex channel matrices, satisfying either of the following properties:

  (ii-a) The matrix $\mathbf{H}_k \in \mathbb{C}^{N \times N_k}$ is deterministic. In this case, we will denote $\mathbf{R}_k = \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}}$.

  (ii-b) The matrix $\mathbf{H}_k \in \mathbb{C}^{N \times N_k}$ is a random channel matrix whose jth column vector $\mathbf{h}_{kj} \in \mathbb{C}^N$ is modeled as

$$\mathbf{h}_{kj} = \mathbf{R}_{kj}^{\frac{1}{2}} \mathbf{z}_{kj}, \qquad j \in \{1, \ldots, N_k\} \tag{3.87}$$

  where $\mathbf{R}_{kj} \in \mathbb{C}^{N \times N}$ are Hermitian nonnegative definite matrices and the vectors $\mathbf{z}_{kj} \in \mathbb{C}^N$ have independent and i.i.d. elements with zero mean, variance $1/N$ and $4+\epsilon$ moment of order $\mathcal{O}(1/N^{2+\varepsilon/2})$, for some common $\epsilon > 0$.

(iii) $\mathbf{W}_k \in \mathbb{C}^{N_k \times n_k}$, $k \in \{1, \ldots, K\}$, are complex (signature or precoding) matrices which contain each $n_k < N_k$ orthonormal columns of independent $N_k \times N_k$ Haar-distributed random unitary matrices,

(iv) $\mathbf{P}_k \in \mathbb{R}^{n_k \times n_k}$, $k \in \{1, \ldots, K\}$, are diagonal (power loading) matrices with nonnegative entries,

(v) $\mathbf{x}_k \sim \mathcal{CN}(0, \mathbf{I}_{n_k})$, $k \in \{1, \ldots, K\}$, are random independent Gaussian transmit vectors,

(vi) $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$ is a white Gaussian noise vector.

In addition, we define the ratios of the matrix dimensions $c_i \triangleq \frac{n_i}{N_i}$ and $\bar{c}_i \triangleq \frac{N_i}{N}$ for $i \in \{1, \dots, K\}$.

Remark 25. *The statistical model (3.87) of the channel $\mathbf{H}_k$ under assumption (ii-b) generalizes several well-known fading channel models of interest (see [70, 169] for examples). These models comprise in particular the Kronecker channel model with transmit and receive correlation matrices [170, 83], where the matrices $\mathbf{H}_k$ are given by*

$$\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{Z}_k \mathbf{T}_k^{\frac{1}{2}} \tag{3.88}$$

with $\mathbf{Z}_k \in \mathbb{C}^{N \times N_k}$ a random matrix whose elements are independent $\mathcal{CN}(0, 1/N)$ and $\mathbf{R}_k \in \mathbb{C}^{N \times N}$, $\mathbf{T}_k \in \mathbb{C}^{N_k \times N_k}$ antenna correlation matrices. Since both $\mathbf{Z}_k$ and $\mathbf{W}_k$ are unitarily invariant, we can assume without loss of generality for the statistical properties of $\mathbf{y}$ that $\mathbf{T}_k = \mathrm{diag}(t_{k1}, \dots, t_{kN_k})$. Defining the matrices $\mathbf{R}_{kj} = t_{kj} \mathbf{R}_k$ for $j \in \{1, \dots, N_k\}$, we fall back to the channel model in (3.87)*. Taking instead all $\mathbf{R}_{kj}$ to be diagonal matrices makes the entries of $\mathbf{H}_k$ independent with $[\mathbf{H}_k]_{ij}$ of zero mean and variance $[\mathbf{R}_{kj}]_{ii}/N$. This corresponds to a centered variance profile model, studied extensively in [74, 51, 164].*

It is our objective to study the performance of the communication channel (3.86) in the large dimensional regime where $N, N_1, \dots, N_K, n_1, \dots, n_K$ are simultaneously large. In the following, we will consider both the *quasi-static channel scenario* which assumes hypotheses (i), (ii-a), (iii)-(vi), and the *fading channel scenario* which assumes (i), (ii-b), (iii)-(vi). The study of the latter naturally arises as an extension of the study of the quasi-static channel scenario. The respective application contexts of both scenarios are described below.

Quasi-static channel scenario (hypothesis (ii-a))

Possible applications of the channel model (3.86) under assumptions (i), (ii-a), (iii)-(vi) arise in the study of direct-sequence (DS) or multi-carrier (MC) CDMA systems with isometric signatures over frequency-selective fading channels or space division multiple access (SDMA) systems with isometric precoding matrices over flat-fading channels. More precisely, for DS-CDMA systems, the matrices $\mathbf{H}_k$ are either Toeplitz or circular matrices (if a cyclic prefix is used) constructed from the channel impulse response; for MC-CDMA, the matrices $\mathbf{H}_k$ are diagonal and represent the channel frequency response on each sub-carrier; for flat fading SDMA systems, the matrices $\mathbf{H}_k$ can be of arbitrary form and their elements represent the complex channel gains between the transmit and receive antennas. In all cases, the diagonal entries of the matrices $\mathbf{P}_k$ determine the transmit power of each signature (CDMA) or transmit stream (SDMA).

The large system analysis of random i.i.d. and random orthogonal precoded systems with optimal and sub-optimal linear receivers has been the subject of numerous publications. The asymptotic performance of MMSE receivers for the channel model (3.86) for the case $K = 1, \mathbf{P}_1 = \mathbf{I}_{n_1}$ and $\mathbf{H}_1$ diagonal with i.i.d. elements has been studied in [72] relying on results from free probability theory. This result was extended to frequency-selective fading channels and sub-optimal receivers in [141]. Although not published, the associated mutual information was evaluated in [171] (this result is recalled in [71, Theorem 4.11]). The case of i.i.d. and isometric MC-CDMA over Rayleigh fading channels with multiple signatures per user terminal, i.e., $K \geq 1$ and $\mathbf{H}_k$ diagonal with i.i.d. complex Gaussian entries, was considered in [172], where approximate solutions of the SINR at the output of the MMSE receiver were provided. Asymptotic expressions for the spectral efficiency of the same model were then derived in [173]. DS-CDMA over flat-fading channels, i.e., $K \geq 1$, $n_k = N$ and $\mathbf{H}_k = \mathbf{I}_N$ for all k, was studied in [174], where the authors derived deterministic equivalents of the Shannon- and η-transform based on the asymptotic freeness [71, Section 3.5] of the matrices $\mathbf{W}_k\mathbf{P}_k\mathbf{W}_k^{\mathsf{H}}$. Besides, a sum-rate maximizing power-allocation algorithm was proposed. Finally, a different approach via incremental matrix expansion [76] led to the exact characterization of the asymptotic SINR of the MMSE receiver for the general channel model (3.86). However, the previously mentioned works share the underlying assumption that the spectral distributions of the matrices $\mathbf{H}_k$ and $\mathbf{P}_k$ converge to some limiting distributions or the matrices $\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}$ are jointly diagonalizable.[7] Also, the computation of the asymptotic SINR requires the computation of rather complicated implicit equations. These can be solved in most cases by standard fixed-point algorithms but a proof of convergence to the correct solution was not provided. Finally, a closed-form expression for the asymptotic spectral efficiency is missing, although an approximate solution which requires numerical integration was presented in [173].

The above works assume non-random communication channels and can therefore be only applied to the performance analysis of static or slow fading channels. Turning the matrices $\mathbf{H}_k$ into random matrices instead allows for the study of the ergodic performance of fast fading channels with isometric precoders. The next section discusses the practical applications in this broader context.

Fading channel scenario (hypothesis (ii-b))

The second scenario considers the channel model (3.68) under assumptions (i), (ii-b), (iii)-(vi). In contrast to the first scenario, the $\mathbf{H}_k$ matrices are now assumed to be random. Thus, we aim at evaluating both the instantaneous performance for a random channel realization and the ergodic performance of these channels. These are appropriate performance measures in fast fading environments.

Of particular interest in this setting is the evaluation of the MIMO channel capacity under random beamforming. In point-to-point MIMO channels, the ergodic channel capacity has been the object of numerous works and is by now well understood [14, 175]. However, the ergodic sum-rate of more involved models, such as the MIMO MAC [83] under individual or sum power constraints, has been studied only recently through the scope of random matrix theory. As a by-product, we will extend the results of [83] to the transmit covariance optimization in the class of scaled identity matrices under sum power constraints. More fundamental is the capacity of MIMO channels with co-channel interference, for which much less is known about the optimal transmission strategies [176, 177].

The first interesting question relates to the problem of how many antennas

[7] That is, there exists a unitary matrix $\mathbf{V}$ such that $\mathbf{V}\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}\mathbf{V}^{\mathsf{H}}$ is diagonal for all k.

should be used for transmission and how many independent data streams should be sent, which are the same problem when the channels have i.i.d. entries. With transmit antenna correlation, however, it makes a difference which antennas are selected for transmission and the question of the optimal number of antennas to be used becomes a combinatorial problem. To circumvent this issue, random beamforming can be used. The remaining question is then how many orthogonal streams should be sent, using *all* available antennas. This is one of our key motivations as our results enable the evaluation of the sum-rate of systems composed of multiple transmitter-receiver pairs, each applying random isotropic beamforming.

In summary, regardless of the specific application scenario of the model (3.68), unitary precoders have gained significant interest in wireless communications [46] (see also the work on spatial multiplexing systems [47] and limited feedback beamforming solutions in future wireless standards [48]). Thus, the performance evaluation of isometric precoded systems is compulsory and a field of active research [49].

Contributions

The object of this section is to propose a new framework for the analysis of large random matrix models involving Haar matrices using the *Stieltjes-transform method*. The analysis is fundamentally based on a trace lemma for Haar matrices first provided in [72] and recalled in Lemma 15. Unlike previous contributions, we dismiss most of the practical constraints of free probability theory, combinatorial and incremental matrix expansion methods, such as the need for spectral limits of the deterministic matrices in the model to exist, or the need for the matrices $\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}$ to be diagonalizable in a common eigenvector basis. The expressions we derive appear to be very similar to previously derived expressions when the precoding matrices $\mathbf{W}_k$ have i.i.d. entries instead of being Haar distributed (see in particular Remark 26). This allows for a unified understanding of both models with i.i.d. or Haar matrices.

Before summarizing our main contributions, we introduce some definitions which will be of repeated use. The central object of interest is the matrix $\mathbf{B}_N \in \mathbb{C}^{N\times N}$, defined as

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{H}_k\mathbf{W}_k\mathbf{P}_k\mathbf{W}_k^{\mathsf{H}}\mathbf{H}_k^{\mathsf{H}}. \tag{3.89}$$

We denote by $I_N(\sigma^2)$ the normalized mutual information of the channel (3.68), given by [42]

$$I_N(\sigma^2) = \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{B}_N\right). \tag{3.90}$$

We further denote by $\gamma_{kj}^N(\sigma^2)$ the SINR at the output of the linear MMSE detector for the jth component of the transmit vector $\mathbf{x}_k$, which reads [136]

$$\gamma_{kj}^N(\sigma^2) = p_{kj}\mathbf{w}_{kj}^{\mathsf{H}}\mathbf{H}_k^{\mathsf{H}}\left(\mathbf{B}_{N(k,j)} + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{H}_k\mathbf{w}_{kj} \tag{3.91}$$

where $\mathbf{B}_{N(k,j)} = \mathbf{B}_N - p_{kj}\mathbf{H}_k\mathbf{w}_{kj}\mathbf{w}_{kj}^{\mathsf{H}}\mathbf{H}_k^{\mathsf{H}}$ and $\mathbf{w}_{kj}$ is the jth column of $\mathbf{W}_k$. We then define the normalized sum-rate with MMSE detection as

$$R_N(\sigma^2) = \frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{n_k}\log\left(1+\gamma_{kj}^{N}(\sigma^2)\right). \tag{3.92}$$

Our technical contributions are as follows: In Theorems 18 and 23 (*ii*) we provide deterministic equivalents of the mutual information for the quasi-static and fading channel model, respectively. We further establish deterministic equivalents of the SINR and the sum-rate with MMSE detection. The expressions are easy to compute as they are shown to be the limits of simple (provably converging) fixed-point algorithms, they are given in closed form and do not require any numerical integration, and they require only very general conditions on the matrices $\mathbf{H}_k$ and $\mathbf{P}_k$.

We then present several applications of our results to wireless communications. First, we consider a cellular uplink orthogonal SDMA communication model with inter-cell interference, assuming independent codes in adjacent cells and quasi-static channels at all communication pairs. We then study a MIMO MAC from several multi-antenna transmitters to a multi-antenna receiver under the fading channel scenario (hypothesis (ii-b)). The transmitters are unaware of the channel realizations and send an arbitrary number of independent data streams using isometric random beamforming vectors. The receiver is assumed to be aware of all instantaneous channel realizations and beamforming vectors. Under this setting, we derive the optimal power allocation under individual or sum-power constraints which can be computed by an iterative water-filling algorithm. Finally, we address the problem of finding the optimal number of independent streams to be transmitted in a two-by-two interference channel. Although the use of deterministic approximations in this context requires an exhaustive search over all possible stream-configurations, it is computationally much less expensive than Monte Carlo simulations. Extensions to more than two transmit-receive pairs and possible different objective functions, e.g., weighted sum-rate or sum-rate with MMSE decoding, are straightforward and not presented. For all these applications, numerical simulations show that the deterministic approximations are very tight even for small system dimensions. In the interference channel model, these simulations suggest in particular that, at low SNR, it is optimal to use all streams while, at high SNR, stream-control, i.e., transmitting less than the maximal number of streams, is beneficial.

Our work also constitutes a novel contribution to the field of random matrix theory, as we introduce new proof techniques based on the Stieltjes transform method in the context of random isometric matrices. Namely, we provide in Theorem 15 a deterministic equivalent $\bar{F}_N$ of the empirical spectral distribution (e.s.d.) F_N of $\mathbf{B}_N$. Although deterministic equivalents of e.s.d. are by now more or less standard and have been developed for rather involved random matrix models [74, 83, 70], results for the case of isometric (Haar) matrices are still an exception. In particular, most results on Haar matrices are based on the assumption of *asymptotic freeness* of the underlying matrices, a requirement which is rarely met for the matrices in the channel model (3.86) of interest here. The approach taken in this section is therefore novel as it does not rely on free probability theory [178, 179] and we do not require any of the matrices in (3.86) to be asymptotically free. Interestingly, a very recent extension of free probability theory, coined free deterministic equivalents [180], has come as a response to the present article in which free probability tools are developed to tackle the aforementioned limitations.

3.6.2 Main results

We will distinguish the results for the quasi-static and the fading channel scenarios. Since we will make limiting considerations as the system dimensions grow large, some technical assumptions will be necessary:

A1 The notation $N \to \infty$ denotes the simultaneous growth of N, N_i, n_i for all i, in such a way that $0 \leq \liminf_N c_i \leq \limsup_N c_i < 1$ and $0 < \liminf_N \bar{c}_i = \frac{N_i}{N} \leq \limsup_N \bar{c}_i < \infty$.

In order to control the power loading matrices as the system grows large, we need the following assumption:

A2 There exists $P > 0$ such that, for all k, $\limsup_N \|\mathbf{P}_k\| \leq P$.

Under (ii-a), the channel gains will need to remain bounded for all large N:

A3-a There exists $R > 0$ such that $\max_k \limsup_N \|\mathbf{R}_k\| \leq R$, where we recall that $\mathbf{R}_k = \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}}$.

The equivalent constraint under (ii-b) is that the channel correlations remain bounded for all large N:

A3-b There exists $R > 0$ such that $\limsup_N \|\mathbf{R}_{kj}\| \leq R$ for all j, k.

Due to some technical issues, it will be sometimes necessary to require the following condition:

A4 For all random matrices $\mathbf{H}_k$ within a set of probability one, there exists $M > 0$ such that $\max_k \|\mathbf{H}_k \mathbf{H}_k^{\mathsf{H}}\| < M$ for all large N.

Assumption **A4** is met in particular in the situation when there exists $m > 0$, such that for all k, j, N, $\mathbf{R}_{kj} \in \mathcal{R}_N$ with $\mathcal{R}_N$ a discrete set of cardinality $|\mathcal{R}_N| < m$ for all N (see the arguments in [83]). For example, this holds true for a common correlation matrix at each receiver, i.e., $\mathbf{R}_{kj} = \bar{\mathbf{R}}_k \, \forall j$.

Fundamental equations and mutual information

The fundamental equations for the quasi-static and the fading channel model are provided in Theorem 15 (i) and Theorem 23 (i), respectively. These Theorems provide a set of deterministic quantities which are needed for the computation of the deterministic equivalents of the mutual information and the SINR.

Remark 26. *Assume $\bar{c}_i = 1$ for every i (e.g., when $\mathbf{H}_i$ is a Toeplitz matrix as in the CDMA case). Then, extending every $\mathbf{P}_i \in \mathbb{C}^{n_i \times n_i}$ into $N \times N$ matrices filled with zeros, we can assume $c_i = 1$ without affecting the final result. In this* *scenario, the fundamental equations under (ii-a) (see Theorem 15 (i)) become for* $z = -x$

$$\bar{e}_i = \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(e_i \mathbf{P}_i + [1 - e_i \bar{e}_i] \mathbf{I}_N\right)^{-1} \tag{3.93}$$
$$e_i = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{e}_j \mathbf{R}_j + x \mathbf{I}_N \right)^{-1}.$$

This can be compared to the scenario where the matrices $\mathbf{W}_i$*, instead of being Haar matrices, have i.i.d. entries of variance* $1/N$*. The fundamental equations of this model are provided in Theorem 13 (i) and are given as follows:*

$$\underline{\bar{e}}_i = \frac{1}{N} \operatorname{tr} \mathbf{P}_i \left(\underline{e}_i \mathbf{P}_i + \mathbf{I}_N\right)^{-1} \tag{3.94}$$
$$\underline{e}_i = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \underline{\bar{e}}_j \mathbf{R}_j + x \mathbf{I}_N \right)^{-1}$$

such that $\underline{e}_i$ *is positive for all* i*. The scalars* $\underline{e}_i$ *and* $\underline{\bar{e}}_i$ *are also defined as the limits of a classical fixed-point algorithm. The only difference between the two sets of equations lies in the additional term* $-e_i \bar{e}_i \mathbf{I}_N$ *in* (3.93)*, not present in* (3.94)*.*

Deterministic equivalents of the (ergodic) mutual information are provided in Theorem 18 for the quasi-static and in Theorem 23 (*ii*) for the fading channel model. Next, we will provide approximations of the SINR at the output of the MMSE receiver.

Theorem 36 (SINR of the MMSE detector under (ii-a))**.** *Consider the system model* (3.86) *under assumptions (i), (ii-a), (iii)-(vi) and, for* $\sigma^2 > 0$*, denote*

$$\gamma_{kj}^N(\sigma^2) = p_{kj} \mathbf{w}_{kj}^{\mathsf{H}} \mathbf{H}_k^{\mathsf{H}} \left(\mathbf{B}_{N(k,j)} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{H}_k \mathbf{w}_{kj}. \tag{3.95}$$

Assume **A1***,* **A2***, and* **A3***-a. Then, as* $N \to \infty$*,*

$$\gamma_{kj}^N(\sigma^2) - \bar{\gamma}_{kj}^N(\sigma^2) \xrightarrow{\text{a.s.}} 0$$

where

$$\bar{\gamma}_{kj}^N(\sigma^2) = \frac{p_{kj} e_k}{\bar{c}_k - e_k \bar{e}_k}$$

with $e_k = e_k(-\sigma^2)$ *and* $\bar{e}_k = \bar{e}_k(-\sigma^2)$ *defined in Theorem 15 (i).*

Proof. The proof is provided in Appendix 3.7.10. □

As an (almost immediate) corollary, we have

Corollary 8. *Under the conditions of Theorem 36, denote*

$$R_N(\sigma^2) = \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \log\left(1 + \gamma_{kj}^N(\sigma^2)\right).$$

Then,

$$
\begin{aligned}
(i) \quad & \mathbb{E}\left[R_N(\sigma^2)\right] - \bar{R}_N(\sigma^2) \to 0 \\
(ii) \quad & R_N(\sigma^2) - \bar{R}_N(\sigma^2) \xrightarrow{\text{a.s.}} 0
\end{aligned}
$$

where

$$
\bar{R}_N(\sigma^2) = \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \log\left(1 + \bar{\gamma}_{kj}^N(\sigma^2)\right).
$$

Proof. The proof is provided in Appendix 3.7.11. □

Theorem 37 (SINR of the MMSE detector under (ii-b))**.** *Consider the system model* (3.86) *under assumptions (i), (ii-b), (iii)-(vi) and, for $\sigma^2 > 0$, denote*

$$
\gamma_{kj}^N(\sigma^2) = p_{kj} \mathbf{w}_{kj}^{\mathsf{H}} \mathbf{H}_k^{\mathsf{H}} \left(\mathbf{B}_{N(k,j)} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{H}_k \mathbf{w}_{kj}.
$$

Assume **A1**, **A2**, **A3**-*b, and* **A4**. *Then, as $N \to \infty$,*

$$
\gamma_{kj}^N(\sigma^2) - \bar{\gamma}_{kj}^N(\sigma^2) \xrightarrow{\text{a.s.}} 0
$$

where

$$
\bar{\gamma}_{kj}^N(\sigma^2) = \frac{p_{kj} b_k}{\bar{c}_k - b_k \bar{b}_k}
$$

with b_k and $\bar{b}_k$, given by Theorem 23 (i) for $x = \sigma^2$.

Proof. The proof is provided in Appendix 3.7.12. □

Similar to the quasi-static channel scenario, we also have the following corollary.

Corollary 9. *Under the conditions of Theorem 37, denote*

$$
R_N(\sigma^2) = \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \log\left(1 + \gamma_{kj}^N(\sigma^2)\right).
$$

Then,

$$
\begin{aligned}
(i) \quad & \mathbb{E}\left[R_N(\sigma^2)\right] - \bar{R}_N(\sigma^2) \to 0 \\
(ii) \quad & R_N(\sigma^2) - \bar{R}_N(\sigma^2) \xrightarrow{\text{a.s.}} 0
\end{aligned}
$$

where

$$
\bar{R}_N(\sigma^2) = \frac{1}{N} \sum_{k=1}^{K} \sum_{j=1}^{n_k} \log\left(1 + \bar{\gamma}_{kj}^N(\sigma^2)\right).
$$

Proof. The proof is provided in Appendix 3.7.13 □

Remark 27. *Surprisingly, the fundamental equations of Theorems 15 (i) and 23 (i) cannot be solved with the proposed fixed-point algorithms for the case $c_k = 1$ (recall that assumption (iii) of the model imposes $c_k < 1$). Moreover, the proof of Theorem 15 cannot be easily extended to this case. However, if $\mathbf{P}_k = p_k \mathbf{I}_{N_k}$, for some $p_k > 0$, the random matrix $\mathbf{B}_N$ reduces to*

$$\mathbf{B}_N = \sum_{k=1}^{K} p_k \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}} = \sum_{k=1}^{K} p_k \mathbf{R}_k. \tag{3.96}$$

For the quasi-static channel scenario, $\mathbf{B}_N$ is thus entirely deterministic. A careful inspection of the fixed-point equations of Theorem 15 reveals that $\bar{e}_k$, with definition extended to $c_k = 1$, has two solutions in the adherence of $[0, \bar{c}_k/e_k)$, i.e., $\bar{e}_k = \frac{\bar{c}_k}{e_k}$ or $\bar{e}_k = p_k$. Simulations suggest that, in this scenario, the fixed-point algorithm proposed in Theorem 15 may converge to either of the solutions depending on the choice of the system parameters. Note that, for $\bar{e}_k = p_k$, Theorem 18 reduces to

$$\bar{I}_N(\sigma^2) = \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k=1}^{K} p_k \mathbf{R}_k\right)$$

as it should be. As for $\bar{e}_k = \frac{\bar{c}_k}{e_k}$, this cannot lead to a correct solution as $\bar{I}_N(\sigma^2)$ would be independent of p_k. These observations are consistent with the condition $\bar{e}_k < \frac{\bar{c}_k}{e_k}$. Similarly, $\bar{b}_k$ in Theorem 23 has the same two possible solutions in this scenario. With $\bar{b}_k = p_k$, the asymptotic mutual information reduces to

$$\bar{I}_N(\sigma^2) = \bar{V}_N(\sigma^2)$$

which is the asymptotic mutual information of a channel with a generalized variance profile as provided in Theorem 14. Thus our results are consistent for the case $c_k = 1$ and $\mathbf{P}_k = p_k \mathbf{I}_{N_k}$. However, if the entries of $\mathbf{P}_k$ are not all equal and $c_k = 1$, we cannot easily infer the solutions of $\bar{e}_k$, $\bar{b}_k$ and the proposed fixed point algorithms may not converge to the correct solutions.

Remark 28. *Based on the previous remark, under scenario (ii-b) with $K = 1$, $\mathbf{P}_1 = \mathbf{I}_{n_1}$, $N_1 = n_1 = N$, and $\mathbf{R}_{1j} = \mathbf{I}_N$ for all j, the set of implicit equations in Theorem 23 reduces to:*

$$\bar{b}(\sigma^2) = 1, \qquad g(\sigma^2) = \frac{\zeta(\sigma^2)}{1 + \zeta(\sigma^2)}, \qquad \zeta(\sigma^2) = \frac{1}{\frac{1}{1+\zeta(\sigma^2)} + \sigma^2}$$

which has a unique solution satisfying $\zeta(\sigma^2) \geq 0$ and that can be given in closed-form:

$$\zeta(\sigma^2) = \frac{-1 + \sqrt{1 + \frac{4}{\sigma^2}}}{2}.$$

We recognize that $\zeta(\sigma^2)$ is the Stieltjes transform of the Marčenko-Pastur law with scale parameter 1 [71, Equation (3.20)] evaluated on the negative real

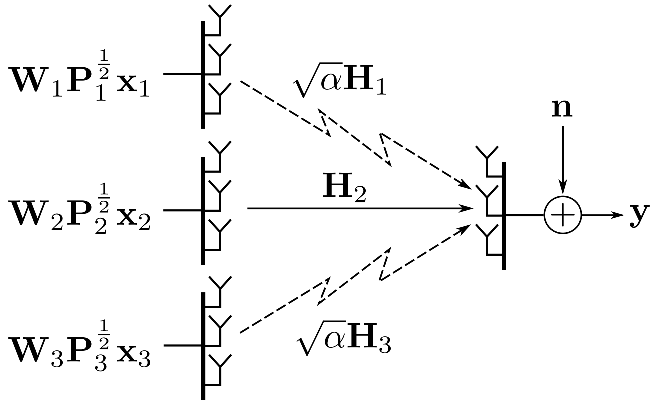

Figure 3.20: Three-cell example: The BS in the center cell decodes the n streams from the UT in its own cell while treating the other signals as interference.

axis. This result is consistent with our expectations since $\mathbf{B}_N = \mathbf{Z}_1\mathbf{Z}_1^{\mathsf{H}}$, *where* $\mathbf{Z}_1 \in \mathbb{C}^{N\times N}$ *has i.i.d. entries with zero mean and variance* $1/N$. *Moreover, the expression of the normalized asymptotic mutual information as given in Theorem 23* (*ii*) *reduces to*

$$\bar{I}_N(\sigma^2) = \bar{V}_N(\sigma^2) = \log\left(1 + \zeta(\sigma^2) + 1/\sigma^2\right) - \frac{\zeta(\sigma^2)}{1+\zeta(\sigma^2)}$$

which is consistent with the asymptotic spectral efficiency of a Rayleigh-fading $N\times N$ *MIMO channel [80, Equation (9)] (see also [71, Section 13.2.2]). Equivalently, the asymptotic SINR of the MMSE detector and the associated normalized sum-rate can be given as (cf. [80, Proposition VI.1]):*

$$\bar{\gamma}_j^N = \zeta(\sigma^2), \qquad \bar{R}_N(\sigma^2) = \log(1+\zeta(\sigma^2)).$$

Remark 29. *Technically, the results obtained for the quasi-static scenario unfold from the Stieltjes transform framework very similar to [83], [74]. However, some new tools are introduced which simplify the analysis made in these papers, such as the method of standard interference functions to prove existence and uniqueness of the derived deterministic equivalents. As for the results in the fading channel scenario, they unfold from the conjugation of the results obtained in the quasi-static scenario and the results obtained in [70] (recalled Theorem 14) for a channel model similar to* (3.86) *but without the presence of the* $\mathbf{W}_k$ *matrices. The central tool to allow this conjugation is the Fubini theorem on the product probability space engendering both the* $\mathbf{W}_k$ *and* $\mathbf{H}_k$ *matrices.*

3.6.3 Numerical results

Uplink orthogonal SDMA with inter-cell interference

In this first example, we apply the theoretical results of Section 3.6.2 under the quasi-static channel scenario (hypothesis (ii-a)) to the uplink channel of an orthogonal SDMA scheme with inter-cell interference. We consider a three cell system with one active UT per cell. The UT in cell k is equipped with N_k

transmit antennas. We focus on the central cell, whose BS is equipped with N antennas, and assume that the the signals received from neighboring cells are treated as noise. This setup is schematically depicted in Figure 3.20. The received signal $\mathbf{y}$ at the BS reads

$$\mathbf{y} = \mathbf{H}_2\mathbf{W}_2\mathbf{P}_2^{\frac{1}{2}}\mathbf{x}_2 + \underbrace{\sqrt{\alpha}\mathbf{H}_1\mathbf{W}_1\mathbf{P}_1^{\frac{1}{2}}\mathbf{x}_1 + \sqrt{\alpha}\mathbf{H}_3\mathbf{W}_3\mathbf{P}_3^{\frac{1}{2}}\mathbf{x}_3 + \mathbf{n}}_{\triangleq\mathbf{z}} \tag{3.97}$$

with $\mathbf{H}_i \in \mathbb{C}^{N\times N_i}$ the channel matrix from UT i to the BS , $\mathbf{x}_i \sim \mathcal{CN}(0, \mathbf{I}_{n_i})$ the transmit symbol of UT i, $\mathbf{W}_i \in \mathbb{C}^{N_i\times n_i}$ the isometric precoding vectors composed of n_i orthogonal streams and $0 < \alpha < 1$ an inter-cell interference factor. The vector $\mathbf{z} \in \mathbb{C}^N$ combines the inter-cell interference and the thermal noise. The covariance matrix $\mathbf{Z} \in \mathbb{C}^{N\times N}$ of $\mathbf{z}$ is given as

$$\mathbf{Z} = \mathbb{E}\left[\mathbf{z}\mathbf{z}^{\mathsf{H}}\right] = \alpha\left[\mathbf{H}_1\mathbf{W}_1\mathbf{P}_1\mathbf{W}_1^{\mathsf{H}}\mathbf{H}_1^{\mathsf{H}} + \mathbf{H}_3\mathbf{W}_3\mathbf{P}_3\mathbf{W}_3^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\right] + \sigma^2\mathbf{I}_N. \tag{3.98}$$

We assume a SDMA system with channel matrices $\mathbf{H}_k \in \mathbb{C}^{N\times N_k}$ generated as realizations of a random standard Gaussian matrix with entries of zero mean and variance $1/N_k$. For simplicity, we further assume that each UT uses $n_k = n$ different transmit signatures to which it assigns equal unit power, i.e., $\mathbf{P}_k = \mathbf{I}_n$. Under these assumptions, the mutual information $I_N(\sigma^2)$ of the central cell when the interference is treated as noise is given by

$$\begin{aligned} I_N(\sigma^2) &= \frac{1}{N}\log\det\left(\mathbf{I}_N + \mathbf{Z}^{-\frac{1}{2}}\mathbf{H}_2\mathbf{W}_2\mathbf{W}_2^{\mathsf{H}}\mathbf{H}_2\mathbf{Z}^{-\frac{1}{2}}\right) \\ &= \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k=1}^{3}\mathbf{H}_k\mathbf{W}_k\mathbf{W}_k^{\mathsf{H}}\mathbf{H}_k\right) \\ &\qquad - \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{\substack{k=1\\k\neq 2}}^{3}\mathbf{H}_k\mathbf{W}_k\mathbf{W}_k^{\mathsf{H}}\mathbf{H}_k\right). \end{aligned} \tag{3.99}$$

According to Theorem 8, the spectral norm of $\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}$ is almost surely uniformly bounded. For such channel realizations, we are therefore in the conditions of Theorem 15. As a consequence, $I_N(\sigma^2) - \bar{I}_N(\sigma^2) \xrightarrow{\text{a.s.}} 0$, with $\bar{I}_N$ defined in Theorem 15 (*ii*). An approximation of the SINR at the output of the MMSE receiver for the jth entry of $\mathbf{x}_2$ can also be computed directly by Theorem 36. We assume $\alpha = 0.25$, $N = 16$, $N_1 = N_2 = N_3 = 8$ and define $\text{SNR} = 1/\sigma^2$. We consider a single random realization of the matrices $\mathbf{H}_k$, which is assumed to be static and therefore deterministically known.

Figure 3.21 depicts $I_N(\sigma^2)$ and the deterministic equivalent $\bar{I}_N(\sigma^2)$ versus SNR for different values of $n \in \{1, 4, 8\}$, scaled to bits/s/Hz instead of nats. Note that for the case $n = 8$, the matrix $\mathbf{B}_N$ and, thus, the mutual information are deterministic (see Remark 28). We observe a very accurate fit between both results over the full range of SNR and n. This validates the deterministic approximation of the mutual information for systems of even small dimensions.

In Figure 3.22, we compare the per-receive antenna sum rate $R_N(\sigma^2)$ with single-stream MMSE-detection to the associated deterministic equivalent $\bar{R}_N(\sigma^2)$, for the same system conditions as in Figure 3.21. The sum rate $R_N(\sigma^2)$ is ex-

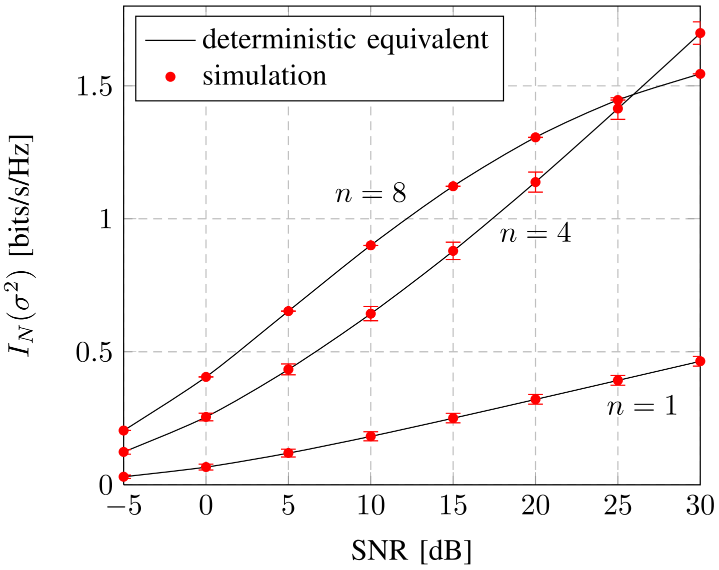

Figure 3.21: Mutual information $I_N(\sigma^2)$ versus SNR for different numbers of transmit signatures n, $N = 16$, $N_i = 8$, $\mathbf{P}_i = \mathbf{I}_n$, $\alpha = 0.5$. Error bars represent one standard deviation on each side.

plicitly given by

$$R_N(\sigma^2) = \frac{1}{N}\sum_{k=1}^{n} \log\left(1 + \gamma_{2,k}^N(\sigma^2)\right)$$

with $\gamma_{ij}^N(\sigma^2)$ defined in (3.95). As for $\bar{R}_N(\sigma^2)$, from Corollary 8, it reads

$$\bar{R}_N(\sigma^2) = c_2\bar{c}_2 \log\left(1 + \frac{e_2(-\sigma^2)}{\bar{c}_2 - e_2(-\sigma^2)\bar{e}_2(-\sigma^2)}\right)$$

with $e_2(-\sigma^2)$ and $\bar{e}_2(-\sigma^2)$ defined in Theorem 15 (i). For the case $n = 8$, we have used $\bar{e}_k = 1$ to compute the deterministic equivalents (see Remark 28). Similar to the previous observations, the deterministic equivalent provides an accurate approximation for all values of SNR and n, although the precision is slightly less than for the mutual information in Figure 3.21.

Multiple access channel

In this and the following example, we apply the theoretical results of Section 3.6.2 under the fading channel scenario (hypothesis (ii-b)). We consider a MAC from three transmitters to a single receiver as shown in Figure 3.23. The channel from each transmitter to the receiver is modeled by the Kronecker model (see Remark 25) with individual transmit and receive covariance matrices $\mathbf{T}_k$ and $\mathbf{R}_k$ and we assume additionally a different path loss $\alpha_k > 0$ on each link. The received signal vector $\mathbf{y}$ for this model reads

$$\mathbf{y} = \sum_{k=1}^{3} \sqrt{\alpha_k}\mathbf{R}_k^{\frac{1}{2}}\mathbf{Z}_k\mathbf{T}_k^{\frac{1}{2}}\mathbf{W}_k\mathbf{P}_k^{\frac{1}{2}}\mathbf{x}_k + \mathbf{n} \tag{3.100}$$

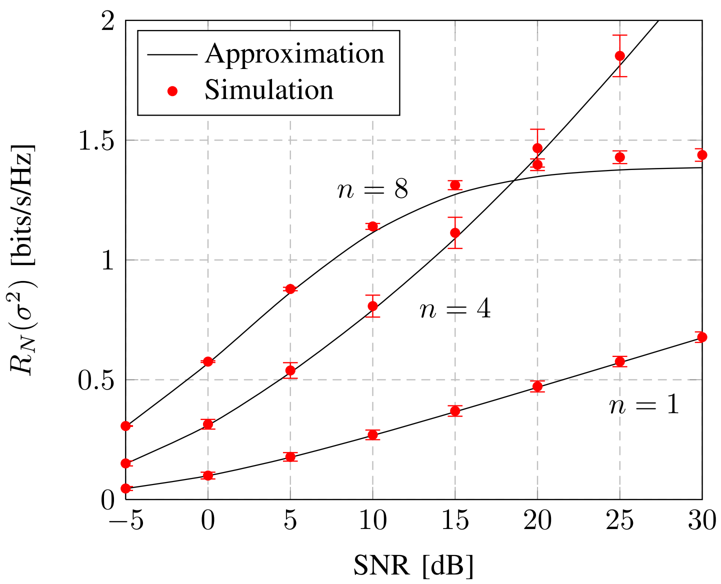

Figure 3.22: Sum rate $R_N(\sigma^2)$ at the output of the MMSE decoder for user 2 versus SNR for different numbers of transmit signatures n, $N = 16$, $N_i = 8$, $\mathbf{P}_i = \mathbf{I}_n$, $\alpha = 0.5$. Error bars represent one standard deviation on each side.

where $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_k})$ and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_N)$. We create the correlation matrices according to a generalization of Jakes' model with non-isotropic signal transmission, see e.g., [181, 182, 183], where the elements of $\mathbf{T}_k$ and $\mathbf{R}_k$ are given as

$$[\mathbf{T}_k]_{ij} = \frac{1}{\theta_{\max}^{t,k} - \theta_{\min}^{t,k}} \int_{\theta_{\min}^{t,k}}^{\theta_{\max}^{t,k}} \exp\left(\frac{\mathbf{i}2\pi}{\lambda} d_{ij}^{t,k} \cos(\theta)\right) d\theta \tag{3.101}$$

$$[\mathbf{R}_k]_{ij} = \frac{1}{\theta_{\max}^{r,k} - \theta_{\min}^{r,k}} \int_{\theta_{\min}^{r,k}}^{\theta_{\max}^{r,k}} \exp\left(\frac{\mathbf{i}2\pi}{\lambda} d_{ij}^{r} \cos(\theta)\right) d\theta \tag{3.102}$$

where $(\theta_{\min}^{t,k}, \theta_{\max}^{t,k})$ and $(\theta_{\min}^{r,k}, \theta_{\max}^{r,k})$ determine the azimuth angles over which useful signal power for the kth transmitter is radiated or received, $d_{ij}^{t,k}$ and d_{ij}^{r} are the distances between the antenna elements i and j at the kth transmitter and receiver, respectively, and λ is the signal wavelength. We assume uniform power allocation for all k, and define SNR $= 1/\sigma^2$. All other parameters are summarized in Table 3.1.

Figure 3.24 compares the normalized mutual information $I_N(\sigma^2)$ and the normalized rate with MMSE decoding $R_N(\sigma^2)$, averaged over 10,000 different realizations of the matrices $\mathbf{H}_k$ and $\mathbf{W}_k$, against their deterministic approximations $\bar{I}_N(\sigma^2)$ and $\bar{R}_N(\sigma^2)$ by Theorem 23 (*ii*) and Corollary 9. Although we have chosen small dimensions for all matrices (see Table 3.1), the match between both results is almost perfect. Also the fluctuations of $I_N(\sigma^2)$ and $R_N(\sigma^2)$ are rather small as can be seen from the error bars representing one standard deviation in each direction.

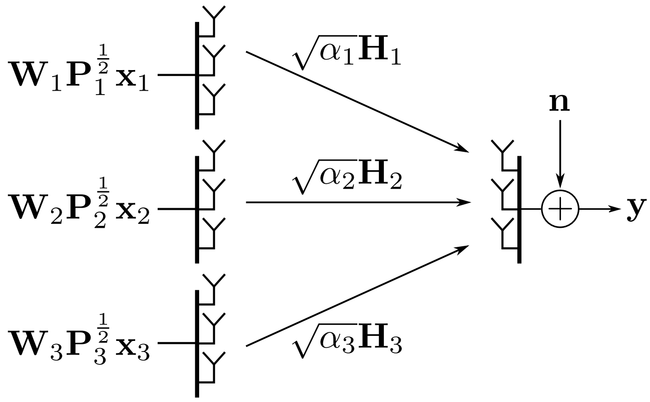

Figure 3.23: MIMO MAC from three transmitters ($k = 1, 2, 3$) with N_k antennas to a receiver with N antennas. Each transmitter sends n_k streams with precoding matrix $\mathbf{W}_k$ and power allocation $\mathbf{P}_k$ over the channel $\sqrt{\alpha_k}\mathbf{H}_k$.

Table 3.1: Simulation parameters for Figure 3.24: $N = 10$, $d^r_{ij} = 8\lambda(i - j)$

k	N_k	n_k	$\theta^{t,k}_{\min}$	$\theta^{t,k}_{\max}$	$\theta^{r,k}_{\min}$	$\theta^{r,k}_{\max}$	$d^{t,k}_{ij}$	α_k
1	10	8	0	$\pi/2$	$-\pi/4$	0	$4\lambda(i-j)$	1
2	5	4	$-\pi/4$	$\pi/4$	0	$\pi/3$	$4\lambda(i-j)$	1/2
3	5	4	$-\pi/2$	0	$-\pi/3$	$\pi/3$	$4\lambda(i-j)$	1/2

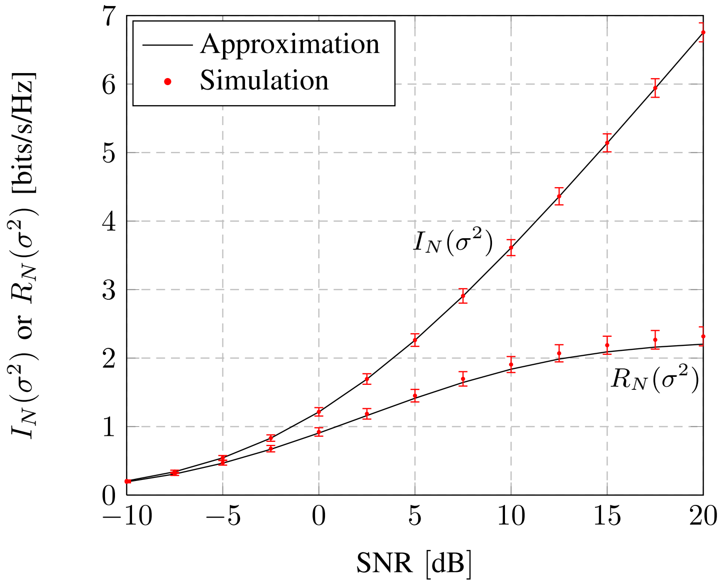

Figure 3.24: Comparison of the average normalized mutual information $I_N(\sigma^2)$ and the normalized rate with MMSE decoding $R_N(\sigma^2)$ with their deterministic approximations $\bar{I}_N(\sigma^2)$ and $\bar{R}_N(\sigma^2)$. Error bars represent one standard deviation in each direction.

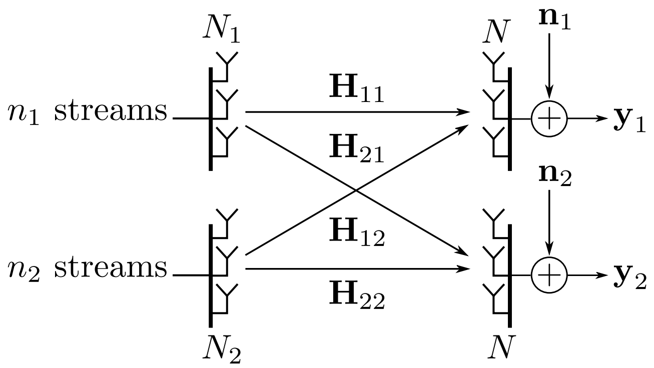

Figure 3.25: Interference channel from two transmitters with N_k ($k = 1, 2$) antennas, respectively, to two receivers with N antennas each. Each transmitter sends n_k independent data streams to its respective receiver.

Stream-control in interference channels

Our last example considers a MIMO interference channel consisting of two transmitter-receiver pairs as depicted in Figure 3.25. The received signal vectors $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{C}^N$ are respectively given as

$$\mathbf{y}_1 = \mathbf{H}_{11}\mathbf{W}_1\mathbf{P}_1^{\frac{1}{2}}\mathbf{x}_1 + \mathbf{H}_{12}\mathbf{W}_2\mathbf{P}_2^{\frac{1}{2}}\mathbf{x}_2 + \mathbf{n}_1 \tag{3.103}$$

$$\mathbf{y}_2 = \mathbf{H}_{21}\mathbf{W}_1\mathbf{P}_1^{\frac{1}{2}}\mathbf{x}_1 + \mathbf{H}_{22}\mathbf{W}_2\mathbf{P}_2^{\frac{1}{2}}\mathbf{x}_2 + \mathbf{n}_2 \tag{3.104}$$

where $\mathbf{H}_{qk} \in \mathbb{C}^{N\times N_k}$, $\mathbf{W}_k \in \mathbb{C}^{N_k\times N_k}$, $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_k})$, $\mathbf{P}_k \in \mathbb{R}_+^{N_k\times N_k}$ satisfying $\frac{1}{N_k}\operatorname{tr}\mathbf{P}_k = 1$, and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_N)$, for $q, k \in \{1, 2\}$. Assuming that the receivers are aware of both precoding matrices and their respective channels but treat the interfering transmission as noise, the normalized mutual informations between $\mathbf{x}_1$ and $\mathbf{y}_1$, and $\mathbf{x}_2$ and $\mathbf{y}_2$, are respectively given as

$$\begin{aligned} I_1(\sigma^2) &= \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k=1}^{2}\mathbf{H}_{1k}\mathbf{W}_k\mathbf{P}_k\mathbf{W}_k^{\mathsf{H}}\mathbf{H}_{1k}^{\mathsf{H}}\right) \\ &\quad - \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{H}_{12}\mathbf{W}_2\mathbf{P}_2\mathbf{W}_2^{\mathsf{H}}\mathbf{H}_{12}^{\mathsf{H}}\right) \end{aligned} \tag{3.105}$$

$$\begin{aligned} I_2(\sigma^2) &= \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k=1}^{2}\mathbf{H}_{2k}\mathbf{W}_k\mathbf{P}_k\mathbf{W}_k^{\mathsf{H}}\mathbf{H}_{2k}^{\mathsf{H}}\right) \\ &\quad - \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{H}_{21}\mathbf{W}_1\mathbf{P}_1\mathbf{W}_1^{\mathsf{H}}\mathbf{H}_{21}^{\mathsf{H}}\right). \end{aligned} \tag{3.106}$$

We adopt the same channel model as for the last example where the channel matrices $\mathbf{H}_{qk}$ are given as

$$\mathbf{H}_{qk} = \mathbf{R}_{qk}^{\frac{1}{2}}\mathbf{Z}_{qk}\mathbf{T}_k^{\frac{1}{2}} \tag{3.107}$$

where $\mathbf{Z}_{qk} \in \mathbb{C}^{N\times N_k}$ have independent $\mathcal{CN}(0, 1/N)$ entries and $\mathbf{T}_k$ and $\mathbf{R}_{qk}$ are calculated according to (3.101) and (3.102), respectively. We assume that no channel state information is available at the transmitters, so that the matrices $\mathbf{P}_k$ are simply used to determine the number of independently transmitted streams:

$$\mathbf{P}_k = \frac{N_k}{n_k}\mathrm{diag}\left(\underbrace{1,\ldots,1}_{n_k},\underbrace{0,\ldots,0}_{N_k-n_k}\right). \tag{3.108}$$

We will now apply the previously derived results to find the optimal number of streams $(n_1^\star, n_2^\star)$ maximizing the normalized ergodic sum-rate of the interference channel above. That is, we seek to find

$$(n_1^\star, n_2^\star) = \max_{n_1,n_2} \mathbb{E}\left[I_1(\sigma^2) + I_2(\sigma^2)\right] \tag{3.109}$$

$$\text{s.t. } 1 \leq n_1 \leq N_1,\ 1 \leq n_2 \leq N_2 \tag{3.110}$$

where the expectation is with respect to both channel and precoding matrices. Due to the complexity of the random matrix model, this optimization problem appears intractable by exact analysis. At the same time, any solution based on an exhaustive search in combination with Monte Carlo simulations becomes quickly prohibitive for large N_1, N_2, since $N_1 \times N_2$ possible combinations need to be tested. Relying on Theorem 23 *(ii)*, we can calculate an approximation of $\mathbb{E}\left[I_1(\sigma^2) + I_2(\sigma^2)\right]$ to find an approximate solution which becomes asymptotically exact as N_1 and N_2 grow large. Thus, we determine $(\bar{n}_1^\star, \bar{n}_2^\star)$ as the solution to

$$(\bar{n}_1^\star, \bar{n}_2^\star) = \max_{n_1,n_2} \bar{I}_1(\sigma^2) + \bar{I}_2(\sigma^2) \tag{3.111}$$

$$\text{s.t. } 1 \leq n_1 \leq N_1,\ 1 \leq n_2 \leq N_2 \tag{3.112}$$

where $\bar{I}_1(\sigma^2), \bar{I}_2(\sigma^2)$ are calculated based on a direct application of Theorem 23 *(ii)* to each of the two log-det terms in $I_1(\sigma^2)$ and $I_2(\sigma^2)$, respectively. The optimal values $(\bar{n}_1^\star, \bar{n}_2^\star)$ are then found by an exhaustive search over all possible combinations. Although we still need to compute $N_1 \times N_2$ values, this is computationally much cheaper than Monte Carlo simulations. Although Theorem 23 *(ii)* does not hold for the case $n_i = N_i$ in general, we can compute a deterministic equivalent of the mutual information by letting $\bar{b}_k = 1$ since $\mathbf{P}_k = \mathbf{I}_{N_k}$ (see Remark 28). In this case, the matrices $\mathbf{W}_k$ vanish and $\bar{I}_N(\sigma^2)$ reduces to the deterministic equivalent of the mutual information of a channel with a variance profile as given by Theorem 14.

Figure 3.26 and Figure 3.27 show $\mathbb{E}\left[I_1(\sigma^2) + I_2(\sigma^2)\right]$ and the deterministic approximation $\bar{I}_1(\sigma^2) + \bar{I}_2(\sigma^2)$, by Theorem 23 *(ii)*, as a function of (n_1, n_2) for the simulation parameters as given in Table 3.2. We have assumed $\mathrm{SNR} = 0\,\mathrm{dB}$ and $\mathrm{SNR} = 40\,\mathrm{dB}$ in Figure 3.26 and Figure 3.27, respectively. In both figures, the solid grid represents simulation results and the markers the deterministic approximations. We observe here again an almost perfect overlap between both sets of results for all values of (n_1, n_2). The optimal values $(n_1^\star, n_2^\star)$ and $(\bar{n}_1^\star, \bar{n}_2^\star)$ coincide for both values of SNR and are indicated by large crosses. At low SNR, both transmitters should send as many independent streams as transmit antennas, i.e., $n_1 = n_2 = 10$. At high SNR, one transmitter should use only a single stream ($n_2 = 1$) and the other transmitter $n_1 = N - 1 = 9$ streams. These results are in line with the observations of [177].

Table 3.2: Simulation parameters for Figure 3.26 and 3.27: $N = 10$, $d_{ij}^{r,k} = 4\lambda(i-j)$, $d_{ij}^{t,k} = 4\lambda(i-j)$

(q,k)	N_k	$\theta_{\min}^{t,k}$	$\theta_{\max}^{t,k}$	$\theta_{\min}^{r,q,k}$	$\theta_{\max}^{r,q,k}$
(1,1)	10	0	$\pi/2$	$-\pi/4$	0
(1,2)	10	$-\pi/2$	0	0	$\pi/4$
(2,1)	10	0	$\pi/2$	$-\pi/3$	0
(2,2)	10	$-\pi/2$	0	0	$\pi/3$

Obviously, the last result is highly unfair and better solutions can be achieved by using different objective functions, such as weighted sum-rate maximization. Also optimal stream-control with MMSE decoding could be carried out in a similar manner. Although we would still need to perform and exhaustive search over all possible combinations of n_1, n_2, the computations based on deterministic equivalents are significantly faster than simulation-based approaches. The development of more intelligent algorithms to determine $(\bar{n}_1^\star, \bar{n}_2^\star)$ is left to future work. The extension to more than two transmitter-receiver pairs is straightforward.

3.6.4 Conclusions

We have studied a class of wireless communication channels with random unitary signature or precoding matrices over quasi-static and fast fading channels and with multiple users or cells. We have provided deterministic approximations of the mutual information, the SINR at the output of the MMSE receiver and the associated sum-rate, which are asymptotically accurate as the system dimensions grow large. Simulations in the contexts of multi-cell SDMA, MIMO-MAC, and interference channels verify the accuracy of the approximations even for systems of small dimensions.

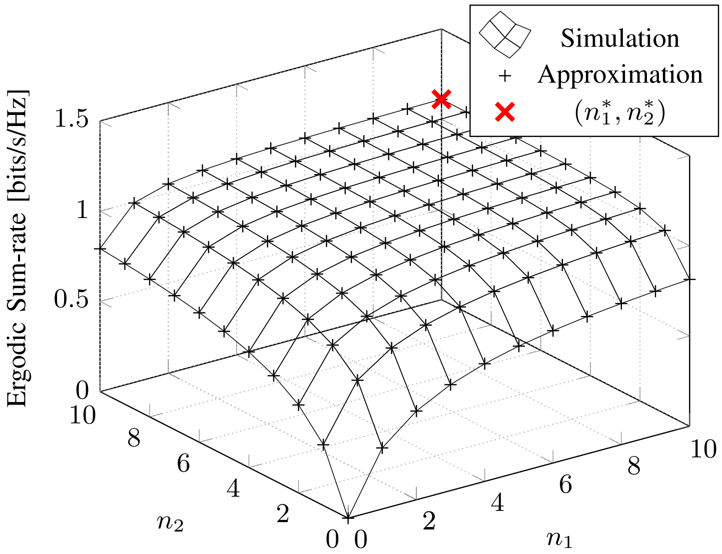

Figure 3.26: Sum-rate versus number of transmitted data-streams (n_1, n_2) for SNR $= 0\,\mathrm{dB}$ and all other parameters as provided in Table 3.2. Solid lines correspond to simulation results, markers to the deterministic approximation by Theorem 23 (*ii*). As expected, both transmitters should send the maximum number of independent streams.

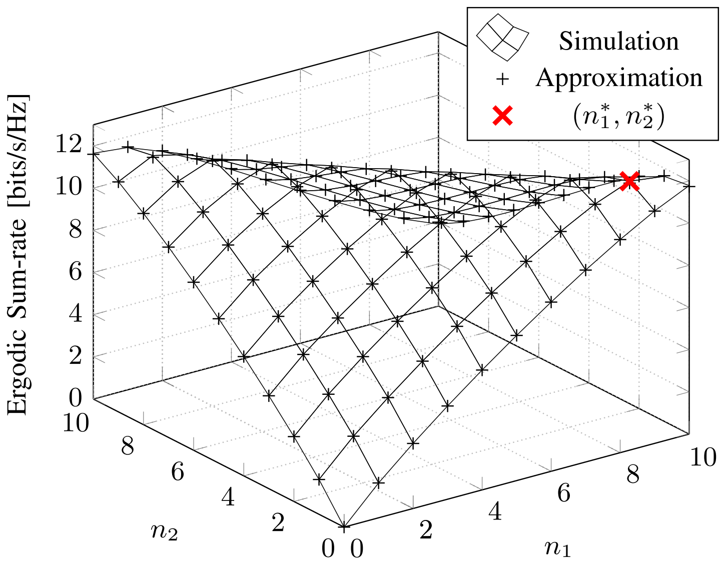

Figure 3.27: Sum-rate versus number of transmitted data-streams (n_1, n_2) for SNR $= 40\,\mathrm{dB}$ and all other parameters as provided in Table 3.2. Solid lines correspond to simulation results, markers to the deterministic approximation by Theorem 23 (*ii*). As co-channel interference is dominant there is a clear gain of limiting the number of transmitted streams.

3.7 Appendices

3.7.1 Proof of Theorem 26

We start by defining the following auxiliary variables $\delta_j = \frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P,\ j = 1,\dots,K$. Using this definition, we can re-write $\overline{R}(\tau)$ in (3.22) as

$$\overline{R}(\tau) = \frac{1}{N}\sum_{j=1}^{K}\left[\log(1+\delta_j) - \frac{\delta_j}{1+\delta_j}\right] - \frac{1}{N}\log\det\left(\frac{L}{KP}\mathbf{T}_P\right). \tag{3.113}$$

We define $\delta_j' = \frac{d\,\delta_j}{d\,\tau} = \frac{1}{K}\operatorname{tr}\mathbf{D}_j'(\tau)\mathbf{T}_P + \frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P'$, where $\mathbf{T}_P' = \frac{d}{d\,\tau}\mathbf{T}_P$. Taking the derivative of $\overline{R}(\tau)$ with respect to τ yields

$$\begin{aligned}\overline{R}'(\tau) &= \frac{1}{N}\sum_{j=1}^{K}\left[\frac{\delta_j'}{1+\delta_j} - \frac{\delta_j'(1+\delta_j)-\delta_j\delta_j'}{(1+\delta_j)^2}\right] - \frac{1}{N}\operatorname{tr}\mathbf{T}_P^{-1}\mathbf{T}_P' \\ &= \frac{1}{N}\sum_{j=1}^{K}\left[\frac{\delta_j\delta_j'}{(1+\delta_j)^2}\right] - \frac{1}{N}\operatorname{tr}\mathbf{T}_P^{-1}\mathbf{T}_P'. \end{aligned}\tag{3.114}$$

This expression can be further simplified by re-writing the definition of $\mathbf{T}_P$ as a function of δ_j:

$$\mathbf{T}_P = \left(\frac{L}{KP}\mathbf{I}_N + \frac{1}{K}\sum_{j=1}^{K}\frac{\mathbf{D}_j(\tau)}{1+\delta_j}\right)^{-1}. \tag{3.115}$$

Using this expression, we have

$$\begin{aligned}\operatorname{tr}\mathbf{T}_P^{-1}\mathbf{T}_P' &= -\operatorname{tr}\mathbf{T}_P^{-1}\mathbf{T}_P\ \frac{d}{d\,\tau}\left(\frac{L}{KP}\mathbf{I}_N + \frac{1}{K}\sum_{j=1}^{K}\frac{\mathbf{D}_j(\tau)}{1+\delta_j}\right)\mathbf{T}_P \\ &= -\operatorname{tr}\mathbf{T}_P\left(\frac{1}{K}\sum_{j=1}^{K}\frac{(1+\delta_j)\mathbf{D}_j'(\tau)-\delta_j'\mathbf{D}_j(\tau)}{(1+\delta_j)^2}\right) \\ &= \sum_{j=1}^{K}\frac{\delta_j'\delta_j - (1+\delta_j)\frac{1}{K}\operatorname{tr}\mathbf{D}_j'(\tau)\mathbf{T}_P}{(1+\delta_j)^2}\end{aligned}\tag{3.116}$$

Plugging this expression into (3.114) and replacing δ_j by $\frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P$ leads to

$$\overline{R}'(\tau) = \frac{1}{N}\sum_{j=1}^{K}\frac{\frac{1}{K}\operatorname{tr}\mathbf{D}_j'(\tau)\mathbf{T}_P}{1+\frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P}. \tag{3.117}$$

In [51, Proposition 5.3], it is proved that

$$\left(\frac{L}{KP} + \max_{i,j}\overline{v}_{ij}(\tau)\right)^{-1} \le [\mathbf{T}_P]_{ii} \le \frac{KP}{L}. \tag{3.118}$$

Since both $\overline{v}_{ij}(\tau)$ and $\overline{v}_{ij}'(\tau)$ are positive for $\tau, P > 0$, it follows from (3.118) that $\frac{1}{K}\operatorname{tr}\mathbf{D}_j'(\tau)\mathbf{T}_P > 0$ and $\frac{1}{K}\operatorname{tr}\mathbf{D}_j(\tau)\mathbf{T}_P > 0$. This implies $\overline{R}'(\tau) > 0$ which concludes the proof.

3.7.2 Proof of Theorem 27

We want to show that $\overline{R}''(\tau) < 0$. Under the assumption of a doubly regular variance profile matrix $\overline{\mathbf{V}}(\tau)$, the implicit matrix equation $\mathbf{T}(z)$ (2.215) of Theorem 25 (i) reduces to a scalar equation, such that $\mathbf{T}(z) = t(z)\mathbf{I}_N$, where

$$t(z) = \frac{1}{-z + \frac{\mathcal{K}(\tau)}{1+\mathcal{K}(\tau)t(z)}}. \tag{3.119}$$

The unique solution to this equation (such that $t(z) \in \mathcal{S}(\mathbb{R}^+)$) can be given in closed-form as

$$t(z) = \frac{\sqrt{1 - \frac{\mathcal{K}(\tau)}{z}} - 1}{2\mathcal{K}(\tau)}. \tag{3.120}$$

Let $t_P = t(-\frac{L}{KP})$. By Theorem 26, the first derivative of $\overline{R}(\tau)$ can be written as

$$\begin{aligned}\overline{R}'(\tau) &= \frac{1}{N}\sum_{j=1}^{N} \frac{\frac{1}{N}\operatorname{tr}\mathbf{D}_j'(\tau)t_P}{1+\frac{1}{N}\operatorname{tr}\mathbf{D}_j(\tau)t_P} \\ &= \frac{t_P\mathcal{K}'(\tau)}{1+t_P\mathcal{K}(\tau)}\end{aligned} \tag{3.121}$$

where $\mathcal{K}'(\tau) = \frac{d}{d\tau}\mathcal{K}(\tau)$. Similarly, the second derivative writes

$$\begin{aligned}\overline{R}''(\tau) &= \frac{[t_P'\mathcal{K}'(\tau) + t_P\mathcal{K}''(\tau)][1+t_P\mathcal{K}(\tau)] - t_P\mathcal{K}'(\tau)[t_P'\mathcal{K}(\tau) + t_P\mathcal{K}'(\tau)]}{[1+t_P\mathcal{K}(\tau)]^2} \\ &= \frac{t_P'\mathcal{K}'(\tau) + t_P\mathcal{K}''(\tau)[1+t_P\mathcal{K}(\tau)] - [t_P\mathcal{K}'(\tau)]^2}{[1+t_P\mathcal{K}(\tau)]^2}.\end{aligned} \tag{3.122}$$

We now need to verify that the numerator of the last equation is negative. One can easily verify from (3.24) and (3.25) that $\mathcal{K}'(\tau) > 0$ and it follows from (3.118) that $t_P > 0$. It remains thus to check that $t_P' < 0$ and $\mathcal{K}''(\tau) < 0$. Write therefore t_P as

$$\begin{aligned}t_P &= \frac{\sqrt{1+\frac{KP}{L}\mathcal{K}(\tau)} - 1}{2\mathcal{K}(\tau)} \\ &= \frac{KP}{2L\left(\sqrt{1+\frac{KP}{L}\mathcal{K}(\tau)} + 1\right)}\end{aligned} \tag{3.123}$$

which is a strictly decreasing function of τ since $\mathcal{K}'(\tau) > 0$. Hence, we have that $t_P' < 0$. In order to show that $\mathcal{K}''(\tau) < 0$, define the two auxiliary functions $\hat{\mathcal{K}}(\tau) = \frac{1}{N}\sum_{i=1}^{N}\hat{v}_{ij}(\tau)$ and $\tilde{\mathcal{K}}(\tau) = \frac{1}{N}\sum_{i=1}^{N}\tilde{v}_{ij}(\tau)$ which are independent of the column index j. It is a simple exercise to verify that $\hat{v}_{ij}(\tau)$ are positive increasing concave functions and $\tilde{v}_{ij}(\tau)$ are positive decreasing convex functions. Due to the regularity conditions of the variance profile, one can verify from (3.7) that the quantization noise σ_i^2 is the same for all BS-antennas, i.e., $\sigma_i = \sigma^2$. Thus,

$$\begin{aligned}\mathcal{K}(\tau) &= \frac{1}{N}\sum_{i=1}^{N}\overline{v}_{ij}(\tau)\\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{v}_{ij}(\tau)}{1+\sigma^2+\frac{PN}{L}\tilde{\mathcal{K}}(\tau)}\\ &= \frac{\hat{\mathcal{K}}(\tau)}{1+\sigma^2+\frac{PN}{L}\tilde{\mathcal{K}}(\tau)}\end{aligned} \tag{3.124}$$

Since both $\hat{\mathcal{K}}(\tau)$ and $(1+\sigma^2+\frac{PN}{L}\tilde{\mathcal{K}}(\tau))^{-1}$ are positive increasing concave functions, it follows from [132, Exercise 3.32 (b)] that the same holds also for their product. Hence, $\mathcal{K}''(\tau) < 0$ and, thus, $\overline{R}''(\tau) < 0$.

3.7.3 Proof of Theorem 28

We start by expanding the difference $R_{\text{net}}(\tau^*) - R_{\text{net}}(\overline{\tau}^*)$ as follows:

$$\begin{aligned}&R_{\text{net}}(\tau^*) - R_{\text{net}}(\overline{\tau}^*)\\ &= \left[R_{\text{net}}(\tau^*) - \overline{R}_{\text{net}}(\tau^*)\right] + \left[\overline{R}_{\text{net}}(\tau^*) - \overline{R}_{\text{net}}(\overline{\tau}^*)\right] + \left[\overline{R}_{\text{net}}(\overline{\tau}^*) - R_{\text{net}}(\overline{\tau}^*)\right].\end{aligned} \tag{3.125}$$

From Theorem 25 (ii), we have that the first and last term of the RHS of (3.125) vanish asymptotically, i.e.,

$$R_{\text{net}}(\tau^*) - \overline{R}_{\text{net}}(\tau^*) \xrightarrow[K\to\infty]{} 0 \quad \text{and} \quad \overline{R}_{\text{net}}(\overline{\tau}^*) - R_{\text{net}}(\overline{\tau}^*) \xrightarrow[K\to\infty]{} 0. \tag{3.126}$$

By the definition of τ^* and $\overline{\tau}^*$, we have for the LHS of (3.125) and the second term on the RHS of (3.125)

$$R_{\text{net}}(\tau^*) - R_{\text{net}}(\overline{\tau}^*) \geq 0 \quad \text{and} \quad \overline{R}_{\text{net}}(\tau^*) - \overline{R}_{\text{net}}(\overline{\tau}^*) \leq 0. \tag{3.127}$$

Equations (3.125), (3.126) and (3.127) together imply that

$$R_{\text{net}}(\tau^*) - R_{\text{net}}(\overline{\tau}^*) \xrightarrow[K\to\infty]{} 0 \tag{3.128}$$

$$\overline{R}_{\text{net}}(\tau^*) - \overline{R}_{\text{net}}(\overline{\tau}^*) \xrightarrow[K\to\infty]{} 0. \tag{3.129}$$

Equation (3.128) together with Theorem 25 (ii) proves part (i).

Assume now that $\overline{\mathbf{V}}(\tau)$ is a doubly regular matrix. Since $\overline{R}_{\text{net}}(\tau)$ is by Theorem 27 a strictly concave function which takes its unique maximum at point $\overline{\tau}^*$, (3.129) implies that $\tau^* - \overline{\tau}^* \xrightarrow[K\to\infty]{} 0$. This proves part ($ii$).

3.7.4 Proof of Theorem 31

We start by dividing the denominator and numerator of γ_{jm}^{dl} by $\frac{1}{N}$.

Signal power: Straight-forward computations yield

$$\begin{aligned}\frac{1}{N}\lambda_j\left|\mathbb{E}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\hat{\mathbf{h}}_{jjm}\right]\right|^2 &= \frac{1}{\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\hat{\mathbf{h}}_{jjk}^{\mathsf{H}}\hat{\mathbf{h}}_{jjk}\right]}\left|\mathbb{E}\left[\frac{1}{N}\hat{\mathbf{h}}_{jjm}^{\mathsf{H}}\hat{\mathbf{h}}_{jjm}\right]\right|^2\\ &= \bar{\lambda}_j\left(\frac{1}{N}\operatorname{tr}\boldsymbol{\Phi}_{jjm}\right)^2\end{aligned} \tag{3.130}$$

where $\bar{\lambda}_j = \left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjk}\right)^{-1}$. By **A 1** and **A 2**, we have $0 < \liminf \bar{\lambda}_j \leq \limsup \bar{\lambda}_j < \infty$.
Interference power: As a direct consequence of Lemma 11, we have

$$\frac{\lambda_j}{N}\operatorname{var}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\hat{\mathbf{h}}_{jjm}\right] = \bar{\lambda}_j \mathbb{E}\left[\left|\frac{1}{N}\mathbf{h}_{jjm}^{\mathsf{H}}\hat{\mathbf{h}}_{jm} - \frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjm}\right|^2\right] \xrightarrow[N\to\infty]{} 0. \quad (3.131)$$

For the remaining terms, we have by (3.50)

$$\begin{aligned}
&\frac{1}{N}\lambda_l \mathbb{E}\left[\left|\mathbf{h}_{ljm}^{\mathsf{H}}\hat{\mathbf{h}}_{llk}\right|^2\right] \\
&= \bar{\lambda}_l \mathbb{E}\left[\frac{1}{N^2}\left|\mathbf{h}_{ljm}^{\mathsf{H}}\mathbf{R}_{llk}\mathbf{Q}_{lk}\left(\sum_{i=1}^{L}\mathbf{h}_{lik} + \frac{1}{\sqrt{\rho_\tau}}\mathbf{n}_{lk}^{\tau}\right)\right|^2\right] \\
&= \bar{\lambda}_l \begin{cases} \frac{1}{N^2}\operatorname{tr}\mathbf{R}_{ljm}\mathbf{\Phi}_{llk} & , k \neq m \\ \mathbb{E}\left[\left|\frac{1}{N}\mathbf{h}_{ljm}\mathbf{R}_{llm}\mathbf{Q}_{lm}\mathbf{h}_{ljm}\right|^2\right] + \frac{1}{N^2}\operatorname{tr}\mathbf{R}_{ljm}\left(\mathbf{\Phi}_{llm} - \mathbf{\Phi}_{ljm}\mathbf{Q}_{lm}\mathbf{R}_{llm}\right) & , k = m \end{cases}.
\end{aligned} \quad (3.132)$$

By Lemma 12, $\mathbb{E}\left[\left|\frac{1}{N}\mathbf{h}_{ljm}\mathbf{R}_{llm}\mathbf{Q}_{lm}\mathbf{h}_{ljm}\right|^2\right] - \left|\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{ljm}\right|^2 \to 0$. Combining all results yields

$$\begin{aligned}
&\sum_{(l,k)\neq(j,m)} \frac{1}{N}\lambda_l \mathbb{E}\left[\left|\mathbf{h}_{ljm}^{\mathsf{H}}\hat{\mathbf{h}}_{llk}\right|^2\right] \\
&\quad - \left[\frac{1}{N}\sum_{l,k}\bar{\lambda}_l \frac{1}{N}\operatorname{tr}\mathbf{R}_{ljm}\mathbf{\Phi}_{llk} + \sum_{l\neq j}\bar{\lambda}_l\left|\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{ljm}\right|^2\right] \xrightarrow[N\to\infty]{} 0.
\end{aligned} \quad (3.133)$$

Note that we have neglected the terms $\frac{1}{N^2}\operatorname{tr}\mathbf{R}_{ljm}\mathbf{\Phi}_{ljm}\mathbf{Q}_{lm}\mathbf{R}_{llm}$ which appear only $L-1$ times and therefore vanish asymptotically. Moreover, we have added the single term $\bar{\lambda}_j \frac{1}{N^2}\operatorname{tr}\mathbf{R}_{jjm}\mathbf{\Phi}_{jjk}$ which is also negligible for large N. Replacing the deterministic equivalents for the useful signal and the interference power in (3.57) finally yields

$$\gamma_{jm}^{\mathrm{dl}} - \frac{\bar{\lambda}_j\left(\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{jjm}\right)^2}{\frac{1}{\rho_{\mathrm{dl}}N} + \frac{1}{N}\sum_{l,k}\bar{\lambda}_l\frac{1}{N}\operatorname{tr}\mathbf{R}_{ljm}\mathbf{\Phi}_{llk} + \sum_{l\neq j}\bar{\lambda}_l\left|\frac{1}{N}\operatorname{tr}\mathbf{\Phi}_{ljm}\right|^2} \xrightarrow[N\to\infty]{} 0. \quad (3.134)$$

3.7.5 Proof of Theorem 33

Define the following matrices for $j = 1,\ldots,L$ and $k = 1,\ldots,K$:

$$\mathbf{\Sigma}_j = \left(\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^{\mathsf{H}} + \mathbf{Z}_j^{\mathrm{dl}} + N\varphi_j^{\mathrm{dl}}\mathbf{I}_N\right)^{-1} \quad (3.135)$$

$$\mathbf{\Sigma}_{jk} = \left(\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^{\mathsf{H}} - \hat{\mathbf{h}}_{jjk}\hat{\mathbf{h}}_{jjk}^{\mathsf{H}} + \mathbf{Z}_j^{\mathrm{dl}} + N\varphi_j^{\mathrm{dl}}\mathbf{I}_N\right)^{-1}. \quad (3.136)$$

Signal power: We divide the denominator and numerator of γ_{jm}^{dl} by $\frac{1}{N}$. Thus,

$$\begin{aligned}
\sqrt{\frac{\lambda_j}{N}}\mathbf{h}_{jjm}^{\mathsf{H}}\mathbf{w}_{jm} &= \sqrt{\frac{\lambda_j}{N}}\mathbf{h}_{jjm}\boldsymbol{\Sigma}_j\hat{\mathbf{h}}_{jjm} \\
&\overset{(a)}{=} \sqrt{\frac{K}{N}}\sqrt{\frac{1}{\mathbb{E}\left[\operatorname{tr}\boldsymbol{\Sigma}_j\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^{\mathsf{H}}\boldsymbol{\Sigma}_j\right]}}\frac{\mathbf{h}_{jjm}\boldsymbol{\Sigma}_{jm}\hat{\mathbf{h}}_{jjm}}{1+\mathbf{h}_{jjm}\boldsymbol{\Sigma}_{jm}\hat{\mathbf{h}}_{jjm}} \\
&\overset{(b)}{\asymp} \sqrt{\frac{K}{N}}\sqrt{\frac{1}{\mathbb{E}\left[\operatorname{tr}\boldsymbol{\Sigma}_j - \operatorname{tr}\left(\mathbf{Z}_j^{\text{dl}}+N\varphi_j^{\text{dl}}\mathbf{I}_N\right)\boldsymbol{\Sigma}_j^2\right]}}\frac{\frac{1}{N}\operatorname{tr}\boldsymbol{\Phi}_{jjm}\mathbf{T}_j}{1+\frac{1}{N}\operatorname{tr}\boldsymbol{\Phi}_{jjm}\mathbf{T}_j} \\
&\overset{(c)}{\asymp} \sqrt{\frac{K}{N}}\sqrt{\frac{1}{\frac{1}{N}\operatorname{tr}\mathbf{T}_j - \frac{1}{N}\operatorname{tr}\left(\frac{\mathbf{Z}_j^{\text{dl}}}{N}+\varphi_j^{\text{dl}}\mathbf{I}_N\right)\bar{\mathbf{T}}'}}\frac{\frac{1}{N}\operatorname{tr}\boldsymbol{\Phi}_{jjm}\mathbf{T}_j}{1+\frac{1}{N}\operatorname{tr}\boldsymbol{\Phi}_{jjm}\mathbf{T}_j} \\
&\overset{(d)}{=} \sqrt{\bar{\lambda}_j}\frac{\delta_{jm}}{1+\delta_{jm}}
\end{aligned} \tag{3.137}$$

where (a) follows from Lemma 6, (b) follows from Lemma 13, Lemma 8 and Theorem 14 applied to the term $\mathbf{h}_{jjm}\boldsymbol{\Sigma}_{jm}\hat{\mathbf{h}}_{jjm}$ and (c) results from Theorem 14 applied to $\operatorname{tr}\boldsymbol{\Sigma}_j$ and Theorem 21 applied to $\operatorname{tr}\left(\mathbf{Z}_j^{\text{dl}}+N\varphi_j^{\text{dl}}\mathbf{I}_N\right)\boldsymbol{\Sigma}_j^2$. Note that both theorems do not only imply almost sure convergence but also convergence in the mean. In the last step, we have used the definitions $\delta_{jm} = \frac{1}{N}\operatorname{tr}\mathbf{R}_{jjm}\mathbf{T}_j$ and $\bar{\lambda}_j = \frac{K}{N}\left(\frac{1}{N}\operatorname{tr}\mathbf{T}_j - \frac{1}{N}\operatorname{tr}\left(\frac{\mathbf{Z}_j^{\text{dl}}}{N}+\varphi_j^{\text{dl}}\mathbf{I}_N\right)\bar{\mathbf{T}}'\right)^{-1}$. By the dominated convergence theorem and the continuous mapping theorem, it is straight-forward to show that

$$\frac{\lambda_j}{N}\left|\mathbb{E}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\boldsymbol{\Sigma}_j\hat{\mathbf{h}}_{jjm}\right]\right|^2 - \bar{\lambda}_j\frac{\delta_{jm}^2}{\left(1+\delta_{jm}\right)^2} \xrightarrow[N\to\infty]{} 0. \tag{3.138}$$

Interference power: Define the following quantities:

$$a = \hat{\mathbf{h}}_{jjm}^{\mathsf{H}}\boldsymbol{\Sigma}_j\hat{\mathbf{h}}_{jjm}, \qquad \bar{a} = \mathbb{E}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\boldsymbol{\Sigma}_j\hat{\mathbf{h}}_{jjm}\right], \qquad b = \tilde{\mathbf{h}}_{jjm}^{\mathsf{H}}\boldsymbol{\Sigma}_j\hat{\mathbf{h}}_{jjm}. \tag{3.139}$$

By Lemma 6, we have $0 \le a, \bar{a} \le 1$. Moreover, $\mathbb{E}[b] = 0$ and $\mathbb{E}[ab] = \mathbb{E}[ab^*] = 0$. Thus,

$$\begin{aligned}
\operatorname{var}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\boldsymbol{\Sigma}_j\hat{\mathbf{h}}_{jjm}\right] &= \mathbb{E}\left[|a-\bar{a}+b|^2\right] \\
&= \mathbb{E}\left[(a-\bar{a})(a+\bar{a})\right] + \mathbb{E}\left[|b|^2\right] \\
&\le 2\ \mathbb{E}\left[|a-\bar{a}|)\right] + \mathbb{E}\left[|b|^2\right].
\end{aligned} \tag{3.140}$$

We have shown in (3.137) (b) that $a - \frac{\delta_{jm}}{1+\delta_{jm}} \xrightarrow[N\to\infty]{\text{a.s}} 0$. Since $a, \bar{a}$ are bounded, it follows from the dominated convergence theorem that $\mathbb{E}\left[|a-\bar{a}|\right] \xrightarrow[N\to\infty]{} 0$. Moreover, one can show that $\mathbb{E}\left[|b|^2\right] \le \frac{1}{N}\left(\frac{1}{\varphi_j^{\text{dl}}}\right)^2\|\mathbf{R}_{jjm}\|^2 \xrightarrow[N\to\infty]{} 0$. Thus,

$$\frac{1}{N}\lambda_j \operatorname{var}\left[\mathbf{h}_{jjm}^{\mathsf{H}}\boldsymbol{\Sigma}_j\hat{\mathbf{h}}_{jjm}\right] \xrightarrow[N\to\infty]{} 0. \tag{3.141}$$

Consider now the terms $|\mathbf{h}_{jlm}^{\mathsf{H}}\mathbf{w}_{lk}|^2$:

$$
\begin{aligned}
|\mathbf{h}_{jlm}^{\mathsf{H}}\mathbf{w}_{lk}|^2 &\overset{(a)}{=} \frac{\hat{\mathbf{h}}_{llk}^{\mathsf{H}}\boldsymbol{\Sigma}_{lk}\mathbf{h}_{jlm}\mathbf{h}_{jlm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lk}\hat{\mathbf{h}}_{llk}}{\left(1+\hat{\mathbf{h}}_{llk}^{\mathsf{H}}\boldsymbol{\Sigma}_{lk}\hat{\mathbf{h}}_{llk}\right)^2} \\
&\overset{(b)}{\asymp} \frac{1}{\left(1+\delta_{lk}\right)^2}\begin{cases}\mathbf{h}_{ljm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lk}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lk}\mathbf{h}_{ljm} & ,\ k\neq m \\ |\vartheta_{ljm}|^2 & ,\ k=m\end{cases}
\end{aligned}
$$

where (a) is due to Lemma 6, (b) follows from Lemmas 13, 8, Theorem 14, and where we have used the definitions $\delta_{lk} = \frac{1}{N}\operatorname{tr}\boldsymbol{\Phi}_{llk}\mathbf{T}_l$ and $\vartheta_{ljm} = \frac{1}{N}\operatorname{tr}\boldsymbol{\Phi}_{ljm}\mathbf{T}_l$. In order to treat the terms for $k \neq m$ further, we need the following identity from Lemma 7:

$$
\boldsymbol{\Sigma}_{lk} = \boldsymbol{\Sigma}_{lkm} - \frac{\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm}\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}}{1+\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm}}, \qquad k\neq m \tag{3.142}
$$

where $\boldsymbol{\Sigma}_{lkm} = \left(\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^{\mathsf{H}} - \hat{\mathbf{h}}_{jjk}\hat{\mathbf{h}}_{jjk}^{\mathsf{H}} - \hat{\mathbf{h}}_{jjm}\hat{\mathbf{h}}_{jjm}^{\mathsf{H}} + \mathbf{Z}_j^{\mathrm{dl}} + N\varphi_j^{\mathrm{dl}}\mathbf{I}_N\right)^{-1}$. Note that $\boldsymbol{\Sigma}_{lkm}$ is independent of $\mathbf{h}_{jlm}$ while $\boldsymbol{\Sigma}_{lk}$ is not. Using (3.142), we can write

$$
\begin{aligned}
\mathbf{h}_{ljm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lk}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lk}\mathbf{h}_{ljm} &= \mathbf{h}_{ljm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lkm}\mathbf{h}_{ljm} \\
&\quad + \frac{\left|\mathbf{h}_{ljm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm}\right|^2\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm}}{\left(1+\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm}\right)^2} \\
&\quad - 2\Re\left\{\frac{\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\mathbf{h}_{ljm}\mathbf{h}_{ljm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lk}\hat{\mathbf{h}}_{llm}}{1+\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm}}\right\}.
\end{aligned} \tag{3.143}
$$

As already shown above, we have $\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm} \asymp \delta_{lm}$ and $\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\mathbf{h}_{ljm} \asymp \vartheta_{ljm}^*$. From Lemmas 13, 8 and Theorem 21, we can similarly obtain

$$
\mathbf{h}_{ljm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lkm}\mathbf{h}_{ljm} \asymp \frac{1}{N^2}\operatorname{tr}\mathbf{R}_{ljm}\mathbf{T}_{lk}' \tag{3.144}
$$

$$
\hat{\mathbf{h}}_{llm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm} \asymp \frac{1}{N^2}\operatorname{tr}\boldsymbol{\Phi}_{llm}\mathbf{T}_{lk}' = \frac{\delta_{lmk}'}{N} \tag{3.145}
$$

$$
\mathbf{h}_{ljm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lkm}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lkm}\hat{\mathbf{h}}_{llm} \asymp \frac{1}{N^2}\operatorname{tr}\boldsymbol{\Phi}_{ljm}\mathbf{T}_{lk}' = \frac{\vartheta_{ljmk}'}{N} \tag{3.146}
$$

where $\mathbf{T}_{lk}' = \mathbf{T}'(\varphi_l^{\mathrm{dl}})$ and $\boldsymbol{\delta}_{lk}' = [\delta_{l1k}'\dots\delta_{lKk}']^{\mathsf{T}} = \boldsymbol{\delta}'(\varphi_l^{\mathrm{dl}})$ are given by Theorem 21 for $\mathbf{S}_N = \mathbf{Z}_l^{\mathrm{dl}}/N$, $\boldsymbol{\Theta}_N = \boldsymbol{\Phi}_{llk}$, $\mathbf{D}_N = \mathbf{I}_N$ and $\mathbf{R}_k = \boldsymbol{\Phi}_{llk}$ for all k. Combining the last results yields to

$$
\begin{aligned}
\mathbf{h}_{ljm}^{\mathsf{H}}\boldsymbol{\Sigma}_{lk}\boldsymbol{\Phi}_{llk}\boldsymbol{\Sigma}_{lk}\mathbf{h}_{ljm} &\asymp \frac{\operatorname{tr}\mathbf{R}_{ljm}\mathbf{T}_{lk}'}{N^2} - \frac{2\Re\left\{\vartheta_{ljm}^*\vartheta_{ljmk}'\right\}(1+\delta_{jm}) - |\vartheta_{ljm}|^2\delta_{lmk}'}{N\left(1+\delta_{lm}\right)^2} \\
&= \frac{\mu_{ljmk}}{N}.
\end{aligned} \tag{3.147}
$$

Note now that

$$\sum_{(l,k)\neq(j,m)} \frac{\lambda_l}{N} |\mathbf{h}_{jlm}^{\mathsf{H}} \mathbf{w}_{lk}|^2 \leq \sum_l \bar{\lambda}_l \mathbf{h}_{jlm}^{\mathsf{H}} \mathbf{\Sigma}_l \hat{\mathbf{H}}_{ll} \hat{\mathbf{H}}_{ll}^{\mathsf{H}} \mathbf{\Sigma}_l^{\mathsf{H}} \mathbf{h}_{jlm}$$
$$\leq \sum_l \bar{\lambda} \mathbf{h}_{jlm}^{\mathsf{H}} \mathbf{\Sigma}_{lm} \mathbf{h}_{ljm}. \quad (3.148)$$

Since $\mathbf{h}_{jlm}^{\mathsf{H}} \mathbf{\Sigma}_{lm} \mathbf{h}_{ljm} \asymp \frac{1}{N} \operatorname{tr} \mathbf{R}_{ljm} \mathbf{T}_l$, $\mathbb{E}\left[\mathbf{h}_{jlm}^{\mathsf{H}} \mathbf{\Sigma}_{lm} \mathbf{h}_{ljm}\right] - \frac{1}{N} \operatorname{tr} \mathbf{R}_{ljm} \mathbf{T}_l \to 0$ by Lemmas 13, 8 and Theorem 14, and $\frac{1}{N} \operatorname{tr} \mathbf{R}_{ljm} \mathbf{T}_l \leq \frac{1}{\varphi_l^{\mathrm{dl}}} \|\mathbf{R}_{ljm}\|$, we have by dominated convergence arguments

$$\sum_{(l,k)\neq(j,m)} \frac{\lambda_l}{N} \mathbb{E}\left[|\mathbf{h}_{jlm}^{\mathsf{H}} \mathbf{w}_{lk}|^2\right] - \sum_{l,k} \frac{\bar{\lambda}_l}{N} \frac{\mu_{ljmk}}{(1+\delta_{lk})^2} - \sum_{l\neq j} \bar{\lambda}_j \frac{|\vartheta_{ljm}|^2}{(1+\delta_{lm})^2} \xrightarrow[N\to\infty]{} 0 \quad (3.149)$$

where we have also subtracted the asymptotically negligible term $\frac{\bar{\lambda}_j}{N} \frac{\mu_{jjmm}}{(1+\delta_{jm})^2}$. Combining (3.138), (3.141) and (3.149) concludes the proof.

3.7.6 Proof of Theorem 34

The proof follows closely those of [164, Proposition 5] and [83, Proposition 3]. We first recall the following property of concave functions (see e.g. [132]):

Property 4. *A function $f(\mathbf{Q}_1, \ldots, \mathbf{Q}_K)$ is strictly concave in the Hermitian nonnegative matrices $\mathbf{Q}_1, \ldots, \mathbf{Q}_K$, if and only if, for any couples $(\mathbf{Q}_{1_a}, \mathbf{Q}_{1_b}), \ldots, (\mathbf{Q}_{K_a}, \mathbf{Q}_{K_b})$ of Hermitian nonnegative matrices, the function*

$$\phi(\lambda) = f(\lambda \mathbf{Q}_{1_a} + (1-\lambda)\mathbf{Q}_{1b}, \ldots, \lambda \mathbf{Q}_{K_a} + (1-\lambda)\mathbf{Q}_{Kb}), \qquad \lambda \in [0,1]$$

is strictly concave.

Consider now $\bar{I}_N(\sigma^2)$ seen as a function of λ for $\mathbf{Q}_k = \lambda \mathbf{Q}_{k_a} - (1-\lambda)\mathbf{Q}_{k_b}$, where $\mathbf{Q}_{k_a}, \mathbf{Q}_{k_b}$ are Hermitian nonnegative definite matrices, for $k = 1, \ldots, K$. Thus, by the chain rule of differentiation,

$$\frac{d\bar{I}_N(\sigma^2)}{d\lambda} = \frac{\partial \bar{I}_N(\sigma^2)}{\partial \lambda} + \sum_{k=1}^K \frac{\partial \bar{I}_N(\sigma^2)}{\partial \bar{g}_k} \frac{\partial \bar{g}_k}{\partial \lambda} + \frac{\partial \bar{I}_N(\sigma^2)}{\partial g_k} \frac{\partial g_k}{\partial \lambda} + \frac{\partial \bar{I}_N(\sigma^2)}{\partial \delta_k} \frac{\partial \delta}{\partial \lambda}. \quad (3.150)$$

One can verify that the partial derivatives of $\bar{I}_N\left(\sigma^2\right)$ with respect to $g_k, \bar{g}_k, \delta_k$, respectively, satisfy

$$\frac{\partial \bar{I}_N\left(\sigma^2\right)}{\partial g_k} = \frac{\partial \bar{I}_N\left(\sigma^2\right)}{\partial \bar{g}_k} = \frac{\partial \bar{I}_N\left(\sigma^2\right)}{\partial \delta_k} = 0, \quad (3.151)$$

due to the defining relation (2.14). Thus,

$$\begin{aligned}
&\frac{d\bar{I}_N(\sigma^2)}{d\lambda} \\
&= \frac{\partial \bar{I}_N(\sigma^2)}{\partial \lambda} \\
&= \sum_{k=1}^{K} \frac{1}{N} \operatorname{tr} \mathbf{T}_k^{\frac{1}{2}} \left(\mathbf{I}_{n_k} + g_k \mathbf{T}_k^{\frac{1}{2}} \left(\lambda \mathbf{Q}_{k_a} + (1-\lambda) \mathbf{Q}_{k_b} \right) \mathbf{T}_k^{\frac{1}{2}} \right)^{-1} \mathbf{T}_k^{\frac{1}{2}} \left(\mathbf{Q}_{k_a} - \mathbf{Q}_{k_b} \right).
\end{aligned} \tag{3.152}$$

The second derivative therefore reads

$$\begin{aligned}
&\frac{d^2\bar{I}_N(\sigma^2)}{d\lambda^2} = \\
&- \sum_{k=1}^{K} \frac{1}{N} \operatorname{tr} \left[\underbrace{\mathbf{T}_k^{\frac{1}{2}} \left(\mathbf{I}_{n_k} + g_k \mathbf{T}_k^{\frac{1}{2}} \left(\lambda \mathbf{Q}_{k_a} + (1-\lambda) \mathbf{Q}_{k_b} \right) \mathbf{T}_k^{\frac{1}{2}} \right)^{-1} \mathbf{T}_k^{\frac{1}{2}}}_{\triangleq \mathbf{A}_k} \underbrace{\left(\mathbf{Q}_{k_a} - \mathbf{Q}_{k_b} \right)}_{\triangleq \mathbf{B}_k} \right]^2
\end{aligned} \tag{3.153}$$

where $\mathbf{A}_k$ are Hermitian nonnegative definite and $\mathbf{B}_k$ are Hermitian. Let $\mathbf{A}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{U}_k^{\mathsf{H}}$ be the eigenvalue decomposition of $\mathbf{A}_k$, where $\mathbf{U}_k \in \mathbb{C}^{n_k \times n_k}$ are unitary matrices and $\mathbf{D}_k = \operatorname{diag}(d_{k,1}, \ldots, d_{k,n_k}) \succeq 0$. Moreover, denote $\mathbf{Z}_k = \mathbf{B}_k \mathbf{U}_k = [\mathbf{z}_{k,1} \ldots \mathbf{z}_{k,n_k}]$. Then,

$$\begin{aligned}
\frac{1}{N} \operatorname{tr} \left[\mathbf{A}_k \mathbf{B}_k \right]^2 &= \frac{1}{N} \operatorname{tr} \mathbf{D}_k \mathbf{U}^{\mathsf{H}} \mathbf{B}_k \mathbf{A}_k \mathbf{B}_k \mathbf{U} \\
&= \frac{1}{N} \operatorname{tr} \mathbf{D}_k \mathbf{Z}_k^{\mathsf{H}} \mathbf{A}_k \mathbf{Z}_k \\
&= \frac{1}{N} \sum_{j=1}^{n_k} d_{k,j} \mathbf{z}_{k,j}^{\mathsf{H}} \mathbf{A}_k \mathbf{z}_{k,j} \geq 0.
\end{aligned} \tag{3.154}$$

If $\mathbf{T}_k \succ 0$, or equivalently, if $\mathbf{T}_k$ is invertible, we have $\mathbf{A}_k \succ 0$ and (3.154) holds with strict inequality for $\mathbf{Q}_{k_a} \neq \mathbf{Q}_{k_b}$. Thus, if any of the matrices $\mathbf{T}_k$ is invertible and $\mathbf{Q}_{k_a} \neq \mathbf{Q}_{k_b}$, we have

$$\sum_{k=1}^{K} \frac{1}{N} \operatorname{tr} \left[\mathbf{A}_k \mathbf{B}_k \right]^2 > 0 \tag{3.155}$$

and hence $\frac{d^2\bar{I}_N(\sigma^2)}{d\lambda^2} < 0$. Thus, $\bar{I}_N(\sigma^2)$ is strictly concave in the matrices $\mathbf{Q}_k$. Due to (3.151) it is then sufficient to maximize

$$\log\det \left(\mathbf{I}_{n_k} + g_k \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k \mathbf{T}_k^{\frac{1}{2}} \right) \tag{3.156}$$

with respect to $\mathbf{Q}_k$ and with the constraint $\frac{1}{n_k} \operatorname{tr} \mathbf{Q}_k \leq P_k$. The solution to this problem is well-known [42] and given by the water-filling solution stated in the theorem. However, since g_k depends on $\mathbf{Q}_k$, this solution must be computed iteratively by Algorithm 1.

3.7.7 Proof of Theorem 35

Similar to the proof of Theorem 22 (ii), let us consider the matrix model

$$\mathbf{H}_k = \frac{1}{\sqrt{n_k}} \mathbf{Z}_k \mathbf{W}_{2,k} \mathbf{T}_k^{\frac{1}{2}} \mathbf{Q}_k^{\frac{1}{2}} \tag{3.157}$$

where

$$\mathbf{Z}_k = \frac{1}{\sqrt{N_k}} \mathbf{R}_k^{\frac{1}{2}} \mathbf{W}_{k,1} \mathbf{S}_k^{\frac{1}{2}}. \tag{3.158}$$

Then,

$$\begin{aligned} \gamma_{k,j}^N &= p_{k,j} t_{k,j} \frac{1}{n_k} \mathbf{w}_{2,k,j}^{\mathsf{H}} \mathbf{Z}_k^{\mathsf{H}} \\ &\times \left(\sum_{i=1}^K \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} - p_{k,j} t_{k,j} \frac{1}{n_k} \mathbf{Z}_k \mathbf{w}_{2,k,j} \mathbf{w}_{2,k,j} \mathbf{Z}_k^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{Z}_k \mathbf{w}_{2,k,j}. \end{aligned} \tag{3.159}$$

It was shown in the proof of Theorem 22 (ii) that, almost surely, $\limsup_N \|\mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}}\| < 0$. From the Fubini theorem, Lemma 13 and Lemma 8, we therefore have

$$\gamma_{k,j}^N - p_{k,j} t_{k,j} \frac{1}{n_k} \operatorname{tr} \mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}} \left(\sum_{i=1}^K \mathbf{H}_i \mathbf{H}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_{n_k} \right) \xrightarrow{\text{a.s.}} 0. \tag{3.160}$$

Applying Theorem 13 (ii) leads to

$$\gamma_{k,j}^N - p_{k,j} t_{k,j} \frac{1}{n_k} \operatorname{tr} \mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}} \left(\sum_{i=1}^K \bar{e}_i \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_{n_k} \right)^{-1} \xrightarrow{\text{a.s.}} 0 \tag{3.161}$$

where $\bar{e}_i$ are given as the unique solutions to (2.174). Notice now from (2.174) that

$$e_k = \frac{1}{n_k} \operatorname{tr} \mathbf{Z}_k \mathbf{Z}_k^{\mathsf{H}} \left(\sum_{i=1}^K \bar{e}_i \mathbf{Z}_i \mathbf{Z}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_{n_k} \right)^{-1} \tag{3.162}$$

and that $\max_k |e_k - g_k| \xrightarrow{\text{a.s.}} 0$ by (2.201). This finally implies

$$\gamma_{k,j}^N - p_{k,j} t_{k,j} g_k \xrightarrow{\text{a.s.}} 0. \tag{3.163}$$

3.7.8 Proof of Corollary 5

Part (i) is a consequence of Theorem 35 and the continuous mapping theorem.

For Part (ii), first notice that $R_N(\sigma^2) \leq I_N(\sigma^2)$ and $\bar{R}_N(\sigma^2) \leq \bar{I}_N(\sigma^2)$. Thus,

$$\left| R_N(\sigma^2) - \frac{1}{N} \sum_{k=1}^K \sum_{j=1}^{n_k} \log\left(1 + p_{k,j} t_{k,j} g_k\right) \right| \leq I_N(\sigma^2) + \bar{I}_N(\sigma^2) \triangleq \phi_N(\sigma^2). \tag{3.164}$$

Since $\mathbb{E}\left[\phi_N(\sigma^2)\right] < \infty$ by Theorem 22 (ii) (b), it follows from dominated convergence arguments that

$$\mathbb{E}\left[R_N(\sigma^2)\right] - \bar{R}_N(\sigma^2) \to 0. \tag{3.165}$$

3.7.9 Proof of Corollary 6

Under the assumptions of the corollary, the fundamental equations in Theorem 22 (i) reduce to

$$\bar{g} = \frac{1}{1+g} \tag{3.166}$$

$$g = \frac{S}{N}\frac{\delta}{1+\bar{g}\delta} \tag{3.167}$$

$$\delta = \frac{1}{K\frac{\bar{g}g}{\delta} + \frac{S}{N}\sigma^2} \tag{3.168}$$

From (3.166), we have

$$g = \frac{1-\bar{g}}{\bar{g}}. \tag{3.169}$$

Solving (3.167) for δ and replacing g by (3.169) yields

$$\delta = \frac{1-\bar{g}}{\bar{g}\left(\bar{g}+\frac{S}{N}-1\right)}. \tag{3.170}$$

Solving (3.168) for δ and replacing g by (3.169) leads to

$$\delta = \frac{1-K(1-\bar{g})}{\frac{S}{N}\sigma^2}. \tag{3.171}$$

Equating (3.170) and (3.171) and rearranging the terms as a polynomial in $\bar{g}$ finally yields

$$\bar{g}^3 - \bar{g}^2\left(2-\frac{S}{N}-\frac{1}{K}\right) + \bar{g}\left(1-\frac{S}{N}-\frac{1}{K}+\frac{S}{NK}\left(1+\sigma^2\right)\right) - \frac{S}{NK}\sigma^2 = 0. \tag{3.172}$$

By Theorem 22 (i), only one of the roots of this polynomial satisfies $\bar{g}, g, \delta > 0$. Now, (3.169) implies $\bar{g} < 1$, (3.170) implies $\bar{g} > 1 - \frac{S}{N}$ (3.171) implies $\bar{g} > 1 - \frac{1}{K}$. Hence $\bar{g} \in \left(1 - \min\left[\frac{1}{K}, \frac{S}{N}\right], 1\right)$.

Similarly, $\bar{I}_N\left(\sigma^2\right)$ reduces under the assumptions of the corollary to

$$\bar{I}_N\left(\sigma^2\right) = \log\left(1+\frac{NK}{\sigma^2 S}\frac{\bar{g}g}{\delta}\right) + \frac{KS}{N}\log\left(1+\bar{g}\delta\right) + K\log\left(1+g\right) - 2K\bar{g}g. \tag{3.173}$$

Replacing $\frac{g}{\delta}$ by $\bar{g} + \frac{S}{N} - 1$ in the first term, δ by (3.170) in the second term, g by (3.169) in the third term and $\bar{g}g$ by $(1-\bar{g})$ in the last term leads to the desired result.

The simplification of Theorem 35 is immediate since $\bar{\gamma}_{kj}^N = p_{k,j}t_{k,j}g_k = \frac{1-\bar{g}}{\bar{g}}$ by (3.169).

3.7.10 Proof of Theorem 36

We will pursue a similar approach as for the proof of Theorem 15, but we can now take advantage of all results derived so far.

First denote d_i the unique positive solution, for $e_i > 0$, to

$$e_i = d_i \left(\bar{c}_i - \frac{1}{N} \sum_{l=1}^{n_i} \frac{p_{il} d_i}{1 + p_{il} d_i} \right) . \tag{3.174}$$

This solution exists and is unique due to the arguments given in the introduction of Step 2 of the proof of Theorem 15.

Whatever the value of c_i, we will proceed as previously by extending the matrix $\mathbf{P}_i$ to an N_i-dimensional matrix with the last $N_i - n_i$ diagonal entries filled with zeros. This way, we can write

$$e_i = d_i \left(\frac{1}{N} \sum_{l=1}^{N_i} 1 - \frac{p_{il} d_i}{1 + p_{il} d_i} \right) = \frac{1}{N} \sum_{l=1}^{N_i} \frac{d_i}{1 + p_{il} d_i}. \tag{3.175}$$

Since d_i is a continuous mapping of e_i and $e_i \leq \frac{P}{|z|}$, it follows that d_i is bounded from above.

Remember now that for $\limsup c_i < 1$ for all i and, for some $z_0 < 0$, we have that $z < z_0$ implies

$$\mathbb{E}[|f_i - e_i|^4] = \mathbb{E}\left[\left| f_i - \frac{1}{N} \sum_{l=1}^{N_i} \frac{d_i}{1 + p_{il} d_i} \right|^4 \right] \leq \frac{C}{N^2} \tag{3.176}$$

for some constant $C > 0$. Also, from (2.28),

$$\mathbb{E}\left[\left| f_i - \frac{1}{N} \sum_{l=1}^{N_i} \frac{\delta_i}{1 + p_{il} \delta_i} \right|^4 \right] \leq \frac{C_1}{N^2} \tag{3.177}$$

for some $C_1 > C$. From these two inequalities, we have

$$\mathbb{E}\left[\left| \frac{1}{N} \sum_{l=1}^{N_i} \frac{\delta_i}{1 + p_{il} \delta_i} - \frac{1}{N} \sum_{l=1}^{N_i} \frac{d_i}{1 + p_{il} d_i} \right|^4 \right] \leq \frac{16 C_1}{N^2}. \tag{3.178}$$

Also, from an immediate application of the Lemma 15, we remind that

$$\mathbb{E}\left[\left| \mathbf{w}_{il}^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}} \left(\mathbf{B}_{(i,l)} - z \mathbf{I}_N \right)^{-1} \mathbf{H}_i \mathbf{w}_{il} - \delta_i \right|^4 \right] \leq \frac{C_2}{N^2} \tag{3.179}$$

for some $C_2 > C_1$.

Together, this implies that for z small enough and for any $k \in \{1, \dots, n_k\}$,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{N}\sum_{l=1}^{N_i}\frac{d_i}{1+p_{il}d_i} - \frac{1}{N}\sum_{l=1}^{N_i}\frac{\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}}{1+p_{il}\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}}\right|^4\right] \\
&\leq 8\left[\mathbb{E}\left[\left|\frac{1}{N}\sum_{l=1}^{N_i}\frac{d_i}{1+p_{il}d_i} - \frac{1}{N}\sum_{l=1}^{N_i}\frac{\delta_i}{1+p_{il}\delta_i}\right|^4\right]\right. \\
&\quad\left. + \mathbb{E}\left[\left|\frac{1}{N}\sum_{l=1}^{N_i}\frac{\delta_i}{1+p_{il}\delta_i} - \frac{1}{N}\sum_{l=1}^{N_i}\frac{\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}}{1+p_{il}\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}}\right|^4\right]\right] \\
&= \quad 8\left[\mathbb{E}\left[\left|\frac{1}{N}\sum_{l=1}^{N_i}\frac{d_i}{1+p_{il}d_i} - \frac{1}{N}\sum_{l=1}^{N_i}\frac{\delta_i}{1+p_{il}\delta_i}\right|^4\right]\right. \\
&\quad\left. + \mathbb{E}\left[\left|\frac{1}{N}\sum_{l=1}^{N_i}\frac{\delta_i - \mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}}{(1+p_{il}\delta_i)(1+p_{il}\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik})}\right|^4\right]\right] \\
&\leq \frac{136C_2}{N^2}.
\end{aligned} \tag{3.180}$$

This ensures that for $z < z_0$,

$$\frac{1}{N}\sum_{l=1}^{N_i}\frac{d_i}{1+p_{il}d_i} - \frac{1}{N}\sum_{l=1}^{N_i}\frac{\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}}{1+p_{il}\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}} \xrightarrow{\text{a.s.}} 0 \tag{3.181}$$

irrespectively of the choice of k.

Since the function $f : x \mapsto \frac{1}{N}\sum_{l=1}^{N_i}\frac{x}{1+p_{il}x}$ is continuous and has positive derivative, it is a one-to-one continuous function. Therefore, for $\mathbf{B}_1, \mathbf{B}_2, \dots$ a realization such that the convergence of (3.181) is ensured, we also have by continuity $d_i - \mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik} \to 0$. Finally,

$$d_i - \mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik} \xrightarrow{\text{a.s.}} 0. \tag{3.182}$$

Noticing from (2.77) that $d_i = \frac{e_i}{\bar{c}_i - e_i\bar{e}_i}$, we have proved the convergence for $z < z_0$. The Vitali convergence theorem then ensures that the convergence holds true for all $z \in \mathbb{C} \setminus \mathbb{R}^+$.

Since the quantities d_i and $\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}$ are uniformly bounded for all N (a result that holds surely since we assumed the $\mathbf{H}_i$ deterministic), the dominated convergence theorem also ensures that the convergence holds in the first mean.

3.7.11 Proof of Corollary 8

In order to prove part (ii) we simply invoke the continuous mapping theorem for the function $\phi : x \mapsto \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\log(1+p_{ik}x)$ on the convergence (3.182). The convergence in the mean sense (part (i)) is obtained using the boundedness of d_i and $\mathbf{w}_{ik}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_{(i,k)} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ik}$ uniformly on N and hence the boundedness of their image by ϕ. The dominated convergence theorem then gives the result.

3.7.12 Proof of Theorem 37

The proof follows directly from (2.236), (2.237), and Theorem 36.

3.7.13 Proof of Corollary 9

The almost sure convergence (part (*ii*)) follows directly from Theorem 37 and the continuous mapping theorem.

For the convergence in mean, note first that, as a standard result of information theory, $I_N(\sigma^2) - R_N(\sigma^2) \geq 0$ for all N. Consider now the extended matrix model where $\mathbf{H}_k^L \in \mathbb{C}^{LN \times LN_k}$ is defined in (2.218), $\mathbf{P}_k^L = \mathbf{P}_k \otimes \mathbf{I}_L \in \mathbb{C}^{Ln_k \times Ln_k}$ and $\mathbf{W}_k^L \in \mathbb{C}^{LN_k \times Ln_k}$ is constructed from Ln_k columns of a $LN_k \times LN_k$ random unitary matrix. Denote $I_{N,L}(\sigma^2)$ and $R_{N,L}(\sigma^2)$ the associated mutual information and MMSE sum-rate for this channel model. One can verify that for this model and by Theorem 23 (*ii*) and the convergence of $R_N(\sigma^2) - \bar{R}_N(\sigma^2)$ in the almost sure sense, the following holds

$$I_{N,L}(\sigma^2) \xrightarrow[L\to\infty]{\text{a.s.}} \bar{I}_N(\sigma^2) \tag{3.183}$$

$$R_{N,L}(\sigma^2) \xrightarrow[L\to\infty]{\text{a.s.}} \bar{R}_N(\sigma^2). \tag{3.184}$$

Thus,

$$\begin{aligned}
&I_{N,L}(\sigma^2) - R_{N,L}(\sigma^2) \\
&= I_{N,L}(\sigma^2) - \bar{I}_N(\sigma^2) + \bar{I}_N(\sigma^2) - \bar{R}_N(\sigma^2) + \bar{R}_N(\sigma^2) - R_{N,L}(\sigma^2) \\
&\xrightarrow[L\to\infty]{\text{a.s.}} \bar{I}_N(\sigma^2) - \bar{R}_N(\sigma^2)
\end{aligned} \tag{3.185}$$

from which we can conclude that $\bar{I}_N(\sigma^2) - \bar{R}_N(\sigma^2) \geq 0$ for all N. Using this result, it follows that

$$\begin{aligned}
\left|R_N(\sigma^2) - \bar{R}_N(\sigma^2)\right| &\leq I_N(\sigma^2) + \bar{I}_N(\sigma^2) \\
&\leq I_N(\sigma^2) - \bar{I}_N(\sigma^2) + 2\sup_N \bar{I}_N(\sigma^2) \\
&\triangleq v_N.
\end{aligned} \tag{3.186}$$

Since $v_N \xrightarrow{\text{a.s.}} 2\sup_N \bar{I}_N(\sigma^2) < \infty$ and $\mathbb{E}[v_N] \to 2\sup_N \bar{I}_N(\sigma^2)$ by Theorem 23 (*ii*), it finally follows from [63, Problem 16.4 (a)] that

$$\mathbb{E}\left[R_N(\sigma^2)\right] - \bar{R}_N(\sigma^2) \to 0. \tag{3.187}$$

Chapter 4

Conclusions & Outlook

In order to cope with the exploding demand for wireless data services, advanced mobile communication systems will be characterized by a dense deployment of different types of wireless access points. Likely is a mix of low-power small-cells, indoor femto cells, and macro BSs which are possibly equipped with large arrays of smart antennas. Mitigating interference and reducing the power consumption in such networks is of paramount importance. Thus, CoMP techniques as well as self-optimization functionalities are not only a desirable but also a necessary feature. As mobile networks become more complex also the methods required for their theoretical performance analysis must evolve. This implies that they must be able to account for the most important characteristics of such networks, namely fading, path loss, interference, imperfect CSI, pilot contamination, antenna correlation, LOS conditions, and cooperation with limited data exchange. In this thesis, we have developed novel methods based on large random matrix theory which are able to take these characteristics into account. In particular, the concept of deterministic equivalents which is based on a large system assumption leads often to surprisingly simple and tight approximations of the system performance and allows one to draw important conclusions about the most relevant parameters. One can think of this method as a way to provide a *deterministic abstraction of the physical layer* which significantly reduces the system complexity. Owing to this complexity reduction, it is for example possible to optimize certain system parameters (training length, precoding matrices, etc.) which would have been intractable otherwise. This approach might also be important for the joint optimization of multiple layers of the protocol stack.

We have demonstrated the usefulness of the deterministic equivalent approach in the context of several different scenarios of practical interest, such as the performance analysis and optimization of network MIMO and large-scale MIMO systems and the study of double-scattering and multi-hop relay channels. Moreover, several novel contributions to the field of random matrix theory originate from this dissertation. The most important are the concept of iterative deterministic equivalents and the derivation of deterministic equivalents for a certain class of functionals of random unitary matrices.

Some words of caution are in order with regards to the asymptotic nature of our results. We have put an emphasis on the fact that most of the deterministic equivalents provide very accurate approximations for finite system dimensions. Sometimes the asymptotic and exact results are already indistinguishable for systems with two transmit and receive antennas. However, nothing guarantees that this is the case for all possible parameter choices. In general, the approximations get worse in the high SNR regime (see [50] for a recent discussion of this effect for the mutual information of a point-to-point MIMO channel). An important indicator of the tightness of asymptotic approximations are scaling results which describe the converge rates of certain quantities to their asymptotic limits. For example, while the variance of the normalized mutual information of certain $N \times N$ MIMO channels can be shown to decrease as $1/N^2$ [51], the variance of the SINR with an MMSE detector scales as $1/N$ [52]. Thus, the accuracy of deterministic equivalents depends not only on the system size but, more importantly, also on the random quantity under consideration.

In the following, we will outline some interesting topics for future work:

- **Random network topologies:** Throughout the entire document, we have tacitly assumed that the transmitters and receivers are located at fixed and known positions. However, this assumption is rarely met in practice. UTs naturally move around and the positions of user-deployed access points, such as femto cells, are neither known nor bound to remain static. Thus, the network topology is an additional random parameter of mobile communication systems which must be taken into consideration. A recent approach to tackle this problem is stochastic geometry [53]. In this framework, UTs and access points are seen as random point processes in space. These point processes are generally characterized by their density and their tendency to form clusters, i.e., attraction points of higher density. Although this technique has led to many interesting results on the expected performance of randomly deployed multi-tier networks [54, 55], only very recently cooperation between access points has been considered [56, 57]. However, the consideration of cooperation between nodes introduces complicated interdependencies between the points of the underlying point process (e.g., association of UTs to cell clusters) which cannot be easily resolved with existing techniques. A combination of RMT and stochastic geometry might overcome this problem. RMT could remove the randomness due to fading in such networks while stochastic geometry would average over all possible user locations. This is especially important if one wants to maximize the average system performance with respect to a certain parameter, e.g., the optimal BS-placement or clustering. We have reported first results in this direction in [184, 185].

- **Coding over finite block length:** The mutual information and the achievable rates considered in this thesis are all based on the crucial assumption of codewords of infinite length. Thus, although the channel coherence time is limited, the messages are assumed to be encoded and transmitted over infinitely many of such coherence blocks. In practice, this assumption would imply an infinite delay. Recently, there has been a growing interest in the ultimate performance limits of coding over finite block lengths which was mainly spurred by the papers [58, 59]. However, obtaining tight and explicit bounds on the error probability for a given block length is difficult for fading MIMO channels. The application of RMT to this field of information theory might lead to significant simplifications. We have provided some first results on this topic in [186].

- **Intelligent use of excess antennas:** In Section 3.3, we have analyzed the performance of large-scale MIMO systems where the BSs are equipped with a much larger number of antennas than there are UTs per cell. However, this type of large antenna arrays is more or less a brute-force approach to counter inter-cell interference and to reduce transmit powers. It would be interesting to investigate other ways to make a possibly more intelligent use of additional or "excess" antennas in a network. For example, a macro cell BS could sacrifice some of its antennas for interference cancellation to lower network tiers, such as femto cells. Assuming a reversed TDD protocol between macros and femtos [60], CSI at the macro BSs could be obtained "for free". An interesting question in this context would be if the performance loss in the macro cell is sufficiently compensated for by the resulting rate improvements in the femto network. Another interesting question in this context would be related to the optimal placement of antennas: For a given area with a given distribution of UTs, what is the best way to deploy N antennas to cover this area?

- **Time-varying channels:** The fading channel models we have used in this thesis do not exhibit any correlation over time. That is, the channel realizations at two different time instants are independent random variables. Building upon recent results on time-varying random matrices [61], [62], it would be interesting to study how the network performance varies over time for a given stochastic process. However, such research is still in its infancy and it will take some effort and time until these methods are sufficiently understood to consider matrix models which are relevant for communication systems.

Bibliography

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2010–2015," Feb. 2011, White paper. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf

[2] The Climate Group and Global e-Sustainability Initiative (GeSI), "SMART 2020: Enabling the low carbon economy in the information age," Jun. 2008. [Online]. Available: http://www.smart2020.org/

[3] C.-C. Chong, K. Hamaguchi, P. F. M. Smulders, and S.-K. Yong, "Millimeter-wave wireless communication systems: Theory and applications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, 2007.

[4] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[5] R. G. Gallager, *Low Density Parity-Check Codes.* MIT Press, Cambridge, MA, 1963.

[6] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[7] S. Sesia, I. Toufil, and M. Baker, Eds., *LTE - The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. Wiley, 2011.

[8] S. Ahmadi, *Mobile WiMAX: A Systems Approach to Understanding IEEE 802.16m Radio Access Technology.* Academic Press (Elsevier), 2011.

[9] B. L. Floch, M. Alard, and C. Berrou, "Coded orthogonal frequency division multiplex," *Proc. IEEE*, vol. 83, no. 6, Jun. 1995.

[10] L. S. Cardoso, M. Kobayashi, O. Ryan, and M. Debbah, "Vandermonde frequency division multiplexing for cognitive radio," in *Proc. IEEE 9th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jul. 2008, pp. 421–425.

[11] M. Luby, "Lt codes," in *Proc. IEEE Symposium on Foundations of Computer Science (FOCS)*, 2002, pp. 271–280.

[12] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[13] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[14] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Feb. 1999.

[15] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications.* New York, NY, USA: Cambridge University Press, 2005.

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

[17] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, 2012, to appear. [Online]. Available: http://liu.diva-portal.org/smash/get/diva2:450781/FULLTEXT01

[18] GreenTouch Consortium, "2010–2011 Annual report," 2011. [Online]. Available: http://www.greentouch.org/uploads/documents/GreenTouch_2010-2011_Annual_Report.pdf

[19] H. Huang, O. Alrabadi, J. Daly, D. Samardzija, C. Tran, R. Valenzuela, and S. Walker, "Increasing throughput in cellular networks with higher-order sectorization," in *Proc. Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2010, pp. 630–635.

[20] F. Rayal, "Why have smart antennas not yet gained traction with wireless network operators?" *IEEE Antennas Propag. Mag.*, vol. 47, no. 6, pp. 124–126, Dec. 2005.

[21] P. Zetterberg, "Performance of antenna tilting and beamforming in an urban macrocell," in *Proc. 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology (Wireless VITAE)*, May 2009, pp. 201–206.

[22] D. C. Cox, "Wireless personal communications: what is it?" *IEEE Personal Commun. Mag.*, vol. 2, no. 2, pp. 20–35, Apr. 1995.

[23] W. Webb, *Wireless Communications: The Future.* New York: Wiley, 2007.

[24] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, Apr. 2012, to appear. [Online]. Available: http://users.ece.utexas.edu/~jandrews/pubs/JSAC_FemtoSurvey.pdf

[25] H. Claussen, "The future of small cell networks," *IEEE Comm. Soc. MMTC*, vol. 5, no. 5, pp. 32–36, Sep. 2010. [Online]. Available: http://committees.comsoc.org/mmc/e-news/E-Letter-September10.pdf

[26] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[27] D. Gesbert, S. V. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, Dec. 2010.

[28] P. Marsch and G. Fettweiss, *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*, 1st ed. Cambridge University Press, 2011.

[29] G. Caire, A. Ramprashad, and C. Papadopoulos, "Rethinking network MIMO: Cost of CSIT, performance analysis, and architecture comparisons," in *Proc. Information Theory and Applications Workshop (ITA'10)*, San Diego, CA, US, May 2010.

[30] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[31] K. Loa, C.-C. Wu, S.-T. Sheu, Y. Yuan, M. Chion, D. Huo, and L. Xu, "IMT-advanced relay standards [WiMAX/LTE Update]," *IEEE Commun. Mag.*, vol. 48, no. 8, pp. 40–48, Aug. 2010.

[32] Y. Lin, L. Shao, Z. Zhu, Q. Wang, and R. K. Sabhikhi, "Wireless network cloud: Architecture and system requirements," *IBM Journal of Research and Development*, vol. 54, no. 1, pp. 4:1–4:12, Jan.–Feb. 2010.

[33] "Spectrum policy task force," Federal Communications Commision, Tech. Rep. 02-135, Nov. 2002. [Online]. Available: http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-228542A1.pdf

[34] R. Tandra, A. Sahai, and S. M. Mishra, "What is a spectrum hole and what does it take to recognize one?" *Proc. IEEE*, vol. 97, no. 5, pp. 824–848, May 2009.

[35] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[36] M. Debbah, "Theoretical foundations of mobile flexible networks," *REV Journal on Electronics and Communications*, vol. 1, no. 1, Jan.–Mar. 2011.

[37] M. Jain, J. I. Choi, T. Kim, D. Bharadia, K. Srinivasan, S. Seth, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. 17th Annual International Conference on Mobile Computing and Networking (MOBICOM)*, 2011.

[38] M. R. Andrews, P. P. Mitra, and R. deCarvalho, "Tripling the capacity of wireless communications using electromagnetic polarization," *Nature*, vol. 409, pp. 316–318, 2001.

[39] M. Shafi, M. Zhang, A. L. Moustakas, P. J. Smith, A. F. Molisch, F. Tufvesson, and S. H. Simon, "Polarized MIMO channels in 3-D: Models, measurements and mutual information," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 514–527, 2006.

[40] P. Duhamel and M. Kieffer, *Joint Source-Channel Decoding*, 1st ed. Elsevier, 2010.

[41] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *Proc. IEEE International Symposium on Information Theory (ISIT'09)*, May 28 - Jul. 3 2009, pp. 2184–2188.

[42] T. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: John Wiley and Sons, Inc., 2006.

[43] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.

[44] N. Fawaz, K. Zarifi, M. Debbah, and D. Gesbert, "Asymptotic capacity and optimal precoding in MIMO multi-hop relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2050–2069, Apr. 2011.

[45] S.-P. Yeh and O. Leveque, "Asymptotic capacity of multi-level amplify-and-forward relay networks," in *Proc. IEEE International Symposium on Information Theory (ISIT'07)*, Jun. 2007, pp. 1436–1440.

[46] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[47] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[48] J. Lee, J. K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-Advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, May 2009.

[49] K. Huang, J. G. Andrews, and R. W. Heath, "Performance of orthogonal beamforming for SDMA with limited feedback," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 152–164, Jan. 2009.

[50] Y. Chen and M. R. McKay, "Coulumb fluid, Painlevé transcendants and the information theory of MIMO systems," *IEEE Trans. Inf. Theory*, submitted.

[51] W. Hachem, P. Loubaton, and J. Najim, "A CLT for information theoretic statistics of Gram random matrices with a given variance profile," *The Annals of Probability*, vol. 18, no. 6, pp. 2071–2130, Dec. 2008.

[52] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim, “A central limit theorem for the SINR at the LMMSE estimator output for large dimensional systems,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 5048–5063, Nov. 2009.

[53] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, “Stochastic geometry and random graphs for the analysis and design of wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1029–1046, Sep. 2009.

[54] J. G. Andrews, F. Baccelli, and R. K. Ganti, “A tractable approach to coverage and rate in cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[55] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, “Modeling and analysis of K-tier downlink heterogeneous cellular networks,” *IEEE J. Sel. Areas Commun.*, 2012, to appear. [Online]. Available: http://arxiv.org/pdf/1103.2177

[56] K. Huang and J. G. Andrews, “A stochastic-geometry approach to coverage in cellular networks with multi-cell cooperation,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Houston, TX, US, Dec. 2011.

[57] ——, “Characterizing multi-cell cooperation via the outage-probability exponent,” in *Proc. IEEE Int. Conf. Comm. (ICC)*, Ottawa, Canada, Jun. 2012.

[58] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[59] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[60] A. Adhikary, V. Ntranos, and G. Caire, “Cognitive femtocells: Breaking the spatial reuse barrier of cellular systems,” in *Proc. IEEE Inf. Theory and Applications Workshop (ITA)*, San Diego, CA, US, Feb. 2011.

[61] P. Smith, M. McKay, A. Giorgetti, and M. Chiani, “Random matrix processes: Quantifying rates of change in MIMO systems,” in *European Wireless Conference (EW)*, Apr. 2010, pp. 194–199.

[62] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*. Cambridge University Press, 2010, vol. 200, no. 9.

[63] P. Billingsley, *Probability and Measure*, 3rd ed. John Wiley & Sons, Inc., 1995.

[64] A. W. van der Vaart, *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, New York, 2000.

[65] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[66] Z. D. Bai and J. W. Silverstein, "On the signal-to-interference ratio of CDMA systems in wireless communications," *The Annals of Applied Probability*, vol. 17, no. 1, pp. 81–101, 2007.

[67] ——, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer Series in Statistics, New York, NY, USA, 2009.

[68] J. Hoydis, M. Kobayashi, and M. Debbah, "Asymptotic performance of linear receivers in network MIMO," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2010, pp. 942–948.

[69] Z. D. Bai and J. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional random matrices," *Ann. Probab.*, vol. 26, pp. 316–345, 1998.

[70] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *IEEE Trans. Inf. Theory*, 2012, to appear. [Online]. Available: http://arxiv.org/abs/0906.3682

[71] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011.

[72] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1293–1311, May 2003.

[73] J.-M. Chaufray, W. Hachem, and P. Loubaton, "Asymptotic analysis of optimum and suboptimum CDMA downlink MMSE receivers," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2620–2638, Nov. 2004.

[74] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.

[75] M. G. Krein and A. A. Nudel'man, *The Markov moment problem and extremal problems.* Translations of Mathematical Monographs, American Mathematical Society, Providence, 1977, vol. 50.

[76] M. J. M. Peacock, I. B. Collings, and M. L. Honig, "Eigenvalue distributions of sums and products of large random matrices via incremental matrix expansions," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2123–2138, May 2008.

[77] E. C. Titchmarsh, *The Theory of Functions.* New York, NY, USA: Oxford University Press, 1939.

[78] V. A. Marčenko and L. A. Pastur, “Distributions of eigenvalues for some sets of random matrices,” *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, April 1967.

[79] R. Couillet, J. W. Silverstein, Z. D. Bai, and M. Debbah, “Eigen-inference for energy estimation of multiple sources,” *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2420–2439, Apr. 2011.

[80] S. Verdu and S. Shamai, “Spectral efficiency of CDMA with random spreading,” *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[81] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, “A new approach for mutual information analysis of large dimensional multi-antenna channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.

[82] D. Jonsson, “Some limit theorems for the eigenvalues of a sample covariance matrix,” *J. Multivariate Anal.*, vol. 12, no. 1, pp. 1–38, 1982.

[83] R. Couillet, M. Debbah, and J. W. Silverstein, “A deterministic equivalent for the analysis of correlated MIMO multiple access channels,” *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3493–3514, Jun. 2011.

[84] S. Wagner, “MU-MIMO Transmission and Reception Techniques for the Next Generation of Cellular Wireless Standards (LTE-A),” Ph.D. dissertation, EURECOM, 2229 route des crétes, BP 193 F-06560 Sophia-Antipolis cedex, 2011. [Online]. Available: http://www.eurecom.fr/people/cifre_wagner.en.htm

[85] R. Couillet, J. Hoydis, and M. Debbah, “Random beamforming over quasi-static and fading channels: A deterministic equivalent approach,” *IEEE Trans. Inf. Theory*, Nov. 2011, accepted. [Online]. Available: http://arxiv.org/abs/1011.3717

[86] R. D. Yates, “A framework for uplink power control in cellular radio systems,” *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[87] M. J. Ablowitz and A. S. Fokas, *Complex Variables: Introduction and Applications*, 2nd ed. Cambridge University Press, 2003.

[88] L. Li, A. M. Tulino, and S. Verdú, “Design of reduced-rank MMSE multiuser detectors using random matrix methods,” *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 986–1008, Jun. 2004.

[89] J. Hoydis, A. Kammoun, J. Najim, and M. Debbah, “Outage performance of cooperative small-cell systems under Rician fading channels,” in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, San Francisco, CA, US, Jun. 2011, pp. 551–555.

[90] J. Hoydis, J. Najim, R. Couillet, and M. Debbah, "Fluctuations of the mutual information in large distributed antenna systems with colored noise," in *Proc. IEEE Allerton Conference on Communication, Control, and Computing (ALLERTON)*, Urbana-Champaign, IL, US, Sep. 2010, pp. 240–245.

[91] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[92] J. Hoydis, R. Couillet, and M. Debbah, "Iterative deterministic equivalents for the capacity analysis of communication systems," *IEEE Trans. Inf. Theory*, Dec. 2011, submitted. [Online]. Available: http://arxiv.org/abs/1112.4167

[93] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[94] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill Series in Higher Mathematics, May 1986.

[95] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large n analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[96] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers (ACSSC)*, Pacific Grove, CA, US, Nov. 2007, pp. 83–87.

[97] S. Venkatesan, H. Huang, A. Lozano, and R. Valenzuela, "A WiMAX-based implementation of network MIMO for indoor wireless systems," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, 2009.

[98] S. Shamai, O. Somekh, and B. M. Zaidel, "Multi-cell communications: An information theoretic perspective," in *Proc. Joint Worshop on Communications and Coding (JWCC'04), Donnini (Florence), Italy*, Oct. 2004.

[99] O. Somekh, O. Simeone, Y. Bar-Ness, A. Haimovich, U. Spagnolini, and S. Shamai (Shitz), *An Information Theoretic View of Distributed Antenna Processing in Cellular Systems: Open Architecture for Future Wireless Communications*. Auerbach Publications, CRC Press, New York, NY, USA, 2007.

[100] P. Marsch and G. Fettweis, "On multi-cell cooperative transmission in backhaul-constrained cellular systems," *Annals of Telecommunications*, vol. 63, no. 5–6, pp. 253–269, Jun. 2008.

[101] S. Ramprashad and G. Caire, "Cellular vs network MIMO: A comparison including channel state information overhead," in *Proc. IEEE Personal, Indoor and Mobile Radio Communications Symposium (PIMRC'09)*, Tokyo, Japan, Sep. 2009.

[102] S. Ramprashad, G. Caire, and H. Papadopoulos, “Cellular and network MIMO architectures: MU-MIMO spectral efficiency and costs of channel state information,” in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers (ACSSC’09)*, Pacific Grove, CA, US, Nov. 2009.

[103] S. V. Hanly and P. Whiting, “Information-theoretic capacity of multi-receiver networks,” *Telecommunication Systems*, vol. 1, no. 1, pp. 1–42, Dec. 1993.

[104] A. D. Wyner, “Shannon-theoretic approach to a Gaussian cellular multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, Nov. 1994.

[105] O. Somekh, B. M. Zaidel, and S. Shamai, “Sum rate characterization of joint multiple cell-site processing,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4473–4497, Dec. 2007.

[106] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, “Local base station cooperation via finite-capacity links for the uplink of linear cellular networks,” *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 190–204, Jan. 2009.

[107] S. Shamai, O. Simeone, O. Somekh, A. Sanderovich, B. M. Zaidel, and H. V. Poor, “Information-theoretic implications of constrained cooperation in simple cellular models,” in *Proc. IEEE Personal, Indoor and Mobile Radio Commun. Symp. (PIMRC)*, Cannes, France, Sep. 2008.

[108] S. Shamai, O. Somekh, O. Simeone, A. Sanderovich, B. M. Zaidel, and H. V. Poor, “Cooperative multi-cell networks: Impact of limited-capacity backhaul and inter-users links,” in *Proc. Joint Workshop on Coding and Communications (JWCC’07)*, Durnstein, Austria, Oct. 2007, pp. 14–16.

[109] A. Sanderovich, O. Somekh, and S. Shamai, “Uplink macro diversity with limited backhaul capacity,” in *Proc. IEEE International Symposium on Information Theory (ISIT’07)*, Nice, France, Jun. 2007, pp. 11–15.

[110] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, “Downlink multicell processing with limited backhaul capacity,” *EURASIP Journal on Advances in Signal Processing*, vol. 2009, 2009.

[111] R. Zakhour and D. Gesbert, “Optimized data sharing in multicell MIMO with finite backhaul capacity,” in *Proc. International Zurich Seminar on Communications (IZS’10)*, Zurich, Switzerland, Mar. 2010.

[112] N. Levy and S. Shamai, “Clustered local decoding for Wyner-type cellular models,” in *Proc. Information Theory and Applications Workshop (ITA’09)*, San Diego, CA, US, Feb. 2009, pp. 318–322.

[113] ——, “Information theroretic aspects of users’ activity in a Wyner-like cellular model,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2241–2248, May 2010.

[114] E. Aktas, J. Evans, and S. V. Hanly, “Distributed decoding in a cellular multiple-access channel,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 241–250, Jan. 2008.

[115] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, Dec. 2008.

[116] F. Boccardi, H. Huang, and A. Alexiou, "Network MIMO with reduced backhaul requirements by mac coordination," in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC'08)*, Pacific Grove, CA, US, Oct. 2008, pp. 1125–1129.

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

[118] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[119] M. Kobayashi, G. Caire, and N. Jindal, "How much training and feedback are needed in MIMO broadcast channels?" in *Proc. IEEE International Symposium on Information Theory (ISIT'08)*, Toronto, ON, Canada, Jul. 2008, pp. 2663–2667.

[120] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[121] M. Kobayashi, M. Debbah, and J.-C. Belfiore, "Outage efficient strategies for network MIMO with partial CSIT," in *Proc. IEEE International Symposium on Information Theory (ISIT'09)*, Seoul, Korea, Jul. 2009, pp. 249–253.

[122] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Distributed multicell and multiantenna precoding: Characterization and performance evaluation," in *Proc. IEEE Global Communications Conference (GLOBECOM'09)*, Honolulu, HI, US, Dec. 2009, pp. 1–6.

[123] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. E. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, 2008.

[124] P. Marsch and G. Fettweis, "On downlink network MIMO under a constrained backhaul and imperfect channel knowledge," in *Proc. IEEE Global Communications Conference (GLOBECOM'09)*, Honolulu, HI, US, Dec. 2009, pp. 1–6.

[125] ——, "Uplink CoMP under a constrained backhaul and imperfect channel knowledge," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1730–1742, Jun. 2011.

[126] D. Aktas, M. N. Bacha, J. E. Evans, and S. V. Hanly, "Scaling results on the sum capacity of cellular networks with MIMO links," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3264–3274, Jul. 2006.

[127] H. Huh, G. Caire, S.-H. Moon, and I. Lee, "Multi-cell MIMO downlink with fairness criteria: The large-system limit," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Austin, TX, US, Jun. 2010, pp. 2058–2062.

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

[129] A. Sanderovich, S. Shamai, and Y. Steinberg, "Distributed MIMO receiver: Achievable rates and upper bounds," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4419–4438, Oct. 2009.

[130] A. Del Coso and S. Simoens, "Distributed compression for MIMO coordinated networks with a backhaul constraint," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4698–4709, Sep. 2009.

[131] P. Marsch and G. Fettweis, "On uplink network MIMO under a constrained backhaul and imperfect channel knowledge," in *Proc. IEEE International Conference on Communications (ICC'09)*, Dresden, Germany, Jun. 2009.

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

[133] P. Marsch and G. Fettweis, "On base station cooperation schemes for uplink network MIMO under a constrained backhaul," in *Proc. International Symposium on Wireless Personal Multimedia Communications (WPMC'08)*, Lapland, Finland, Sep. 2008.

[134] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving massive MIMO spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, 2011, submitted. [Online]. Available: http://arxiv.org/abs/1107.3862

[135] S. Moshavi, E. G. Kanterakis, and D. L. Schilling, "Multistage linear receivers for DS-CDMA systems," *International Journal of Wireless Information Networks*, vol. 3, no. 1, pp. 1–17, Jan. 1996.

[136] S. Verdú, *Multiuser Detection.* Cambridge University Press, 1998.

[137] L. Cottatellucci and R. R. Müller, "A systematic approach to multistage detectors in multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3146–3158, Sep. 2005.

[138] R. R. Mueller and S. Verdu, "Design and analysis of low-complexity interference mitigation on vector channels," vol. 19, no. 8, pp. 1429–1441, Aug. 2001.

[139] P. Loubaton and W. Hachem, "Asymptotic analysis of reduced rank Wiener filters," in *Proc. of Information Theory Workshop (ITW'03)*, Paris, France, Mar. 31 - Apr. 4, 2003, pp. 328–331.

[140] M. L. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1928–1946, Jul. 2001.

[141] W. Hachem, “Simple polynomial detectors for CDMA downlink transmissions on frequency-selective channels,” *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 164–171, Jan. 2004.

[142] L. Cottatellucci, R. R. Müller, and M. Debbah, “Asynchronous CDMA systems with random spreading - part ii: Design criteria,” *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1498–1520, Apr. 2010.

[143] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” *IEEE Trans. Commun.*, 2012, submitted. [Online]. Available: http://arxiv.org/abs/1112.3810

[144] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, “The global footprint of mobile communications: The ecological and economic perspective,” *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 55–62, Aug. 2011.

[145] International Agency for Research on Cancer (IARC), “IARC classifies radiofrequency electromagnetic fields as possibly carcinogenic to humans,” May 2011. [Online]. Available: http://www.iarc.fr/en/media-centre/pr/2011/pdfs/pr208_E.pdf

[146] T. L. Marzetta, “How much training is required for multiuser MIMO?” in *Proc. Fortieth Asilomar Conference on Signals, Systems and Computers (ACSSC’06)*, 29 Sep.–1 Nov. 2006, pp. 359–363.

[147] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’11)*, Prague, Czech Repulic, May 22–27, 2011, pp. 3464–3467.

[148] S. K. Mohammed and E. G. Larsson, “Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: The doughnut channel,” *IEEE Trans. Wireless Commun.*, 2011, submitted. [Online]. Available: http://arxiv.org/abs/1111.3752

[149] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, “Pilot contamination and precoding in multi-cell TDD systems,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[150] S. Wei, D. Goeckel, and R. Janaswamy, “On the asymptotic capacity of MIMO systems with antenna arrays of fixed length,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1608–1621, Jul. 2005.

[151] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, “Linear pre-coding performance in measured very-large MIMO channels,” in *Proc. IEEE Vehicular Technology Conf. (VTC Fall)*, San Francisco, CA, US, Sep. 2011, pp. 1–5.

[152] R. R. Müller and H. Hofstetter, “Confirmation of random matrix model for the antenna array channel by indoor measurements,” in *Proc. IEEE Int. Symp. Antennas and Propagation Society*, vol. 1, Jul. 2001, pp. 472–475.

[153] R. R. Müller, “A random matrix model of communication via antenna arrays,” *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2495–2506, Sep. 2002.

[154] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 361–368, Apr. 2002.

[155] P. Almers, F. Tufvesson, and A. F. Molisch, "Keyhole effect in MIMO wireless channels: Measurements and theory," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3596–3604, Dec. 2006.

[156] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.

[157] G. Levin and S. Loyka, "Multi-keyhole MIMO channels: Asymptotic analysis of outage capacity," in *Proc. IEEE International Symposium on Information Theory (ISIT'06)*, Seattle, Washington, US, Jul. 2006, pp. 1305–1309.

[158] H. Shin and M. Z. Win, "MIMO diversity in the presence of double scattering," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2976–2996, Jul. 2008.

[159] S. Yang and J.-C. Belfiore, "Diversity-multiplexing tradeoff of double scattering MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2027–2034, Apr. 2011.

[160] S. Jin, M. R. McKay, K.-K. Wong, and X. Gao, "Transmit beamforming in Rayleigh product MIMO channels: capacity and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5204–5221, Oct. 2008.

[161] X. Li, S. Jin, X. Gao, and M. R. McKay, "Capacity bounds and low complexity transceiver design for double-scattering MIMO multiple access channels," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2809–2822, May 2010.

[162] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1993.

[163] D. Gesbert, T. Ekman, and N. Christophersen, "Capacity limits of dense palm-sized MIMO arrays," in *Proc. Global Telecommunications Conference (GLOBECOM '02)*, vol. 2, Taipei, Taiwan, Nov. 2002, pp. 1187–1191.

[164] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim, "On the capacity achieving covariance matrix for Rician MIMO channels: An asymptotic approach," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1048–1069, Mar. 2010.

[165] G. H. Tucci, "Spectral analysis of the amplify and forward relay network as the number of relay layers increases," in *Proc. 8th International Symposium on Wireless Communication Systems (ISWCS'11)*, Aachen, Germany, Nov. 2011.

[166] V. I. Morgenshtern and H. Bölcskei, "Random matrix analysis of large relay networks," in *Proc. 44th Allerton Annual Conference on Communications, Control and Computing*, Urbana-Champaign, IL, US, Sep. 2006, pp. 106–112.

[167] J. Wagner, B. Rankov, and A. Wittneben, "Large n analysis of amplify-and-forward MIMO relay channels with correlated Rayleigh fading," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5735–5746, Dec. 2008.

[168] S. Jin, M. R. McKay, C. Zhong, and K.-K. Wong, "Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2204–2224, May 2010.

[169] J. Hoydis, M. Debbah, and M. Kobayashi, "Asymptotic moments for interference mitigation in correlated fading channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2796–2800.

[170] C. N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[171] W. Hachem, "An expression for $\int \log(t/\sigma^2 + 1)\mu \boxplus \tilde{\mu}(dt)$," 2008, unpublished.

[172] M. J. M. Peacock, I. B. Collings, and M. L. Honig, "Asymptotic analysis of LMMSE multiuser receivers for multi-signature multicarrier CDMA in Rayleigh fading," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 964–972, Jun. 2004.

[173] ——, "Asymptotic spectral efficiency of multiuser multisignature CDMA in frequency-selective channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1113–1129, Mar. 2006.

[174] R. Couillet and M. Debbah, "Uplink capacity of self-organizing clustered orthogonal CDMA networks in flat fading channels," in *Proc. IEEE Information Theory Workshop (ITW'09)*, Taormina, Sicily, 2009.

[175] J. Dumont, W. Hachem, P. Loubaton, and J. Najim, "On the asymptotic analysis of mutual information of MIMO Rician correlated channels," in *IEEE-EURASIP International Symposium on Control, Communication, and Signal Processing (ISCCSP'06)*, Marrakech, Morocco, Mar. 2006.

[176] R. S. Blum, J. H. Winters, and N. R. Sollenberger, "On the capacity of cellular systems with MIMO," *IEEE Commun. Lett.*, vol. 6, pp. 242–244, Jun. 2002.

[177] R. S. Blum, "MIMO capacity with interference," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 793–801, Jun. 2003.

[178] D. Voiculescu, K. J. Dykema, and A. Nica, "Free random variables," *American Mathematical Society*, 1992.

[179] F. Hiai and D. Petz, "The Semicircle Law, Free Random Variables and Entropy - Mathematical Surveys and Monographs No. 77," 2006.

[180] R. Speicher, C. Vargas, and T. Mai, "Free deterministic equivalents, rectangular random matrix models, and operator-valued free probability theory," 2011. [Online]. Available: http://arxiv.org/abs/1110.1237

[181] D. Chizhik, F. Rashid-Farrokhi, J. Ling, and A. Lozano, "Effects of antenna separation on the capacity of BLAST in correlated channels," *IEEE Communications Letters*, vol. 4, pp. 337–339, Nov. 2000.

[182] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, p. 437, Mar. 2000.

[183] A. L. Moustakas, H. U. Baranger, L. Balents, A. M. Sengupta, and S. H. Simon, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, 2000.

[184] J. Hoydis, A. Müller, R. Couillet, and M. Debbah, "Analysis of multicell cooperation with random user locations via deterministic equivalents," in *Proc. Workshop on Spatial Stochastic Models for Wireless Networks (SPASWIN)*, Paderborn, Germany, May 2012.

[185] A. Müller, J. Hoydis, R. Couillet, and M. Debbah, "Optimal 3D cell planning: A random matrix theory approach," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, submitted.

[186] J. Hoydis, R. Couillet, P. Piantanida, and M. Debbah, "A random matrix approach to the finite blocklength regime of MIMO fading channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, US, Jul. 2012, submitted.

List of Figures

List of Tables