Signal Processing and Communications Group
Department of Signal Theory and Communications
Technical University of Catalonia

Allocation Designs for Massive Multiple Access with Interference Cancellation

A PhD dissertation submitted in partial fulfilment
of the requirements for the Doctoral Degree in
Signal Theory and Communications

Author: Francesc Molina Oliveras
Advisor: Dr. Josep Sala Álvarez

Barcelona, July 2021

All models are wrong, but some are useful.
- George Edward Pelham Box

A la Laura i als meus pares.

Agraïments

Un dia més i un dia menys. No tinc més que paraules d'agraïment per a tothom qui ha contribuït en fer possible, d'una manera o d'una altra, la realització d'aquesta tesi.

En primer lloc, al Josep Sala, director d'aquesta tesi, ja que sense la teva inestimable col·laboració i companyerisme no hagués estat possible. No vull tancar aquest capítol sense donar-te les gràcies per tot el que m'has ensenyat. En segon lloc, voldria donar les gràcies a la resta de membres del WINTER. Us estic molt agraït a tots per haver-me tractat com un més del grup: Xavi Villares, gràcies per motivar-me; Francesc Rey, fantàstic gestor, cuiner i millor persona; Gregori Vázquez, gràcies per les converses i pels consells; Xell Lamarca, molt agraït per obrir-me la porta del teu despatx per resoldre qualsevol dubte; i Jaume Riba, és un plaer parlar amb tu (quan et deixes veure). En tercer lloc, també voldria agraïr el companyerisme de més professors amb qui he tingut oportunitat d'interactuar: Jaume Comellas, Enric Monte, Montse Pardàs; i a la resta de professors amb qui he compartit docència: Marga Cabrera, Sisco Vallverdú i Josep Salavedra.

Tinc més paraules d'agraïment i múltiples records dels companys que també estaven fent la tesi en paral·lel: Jordi Borràs, Ferran De Cabrera, Miquel India i Sergi Liesegang. Celebro els dinars i viatges que hem pogut fer junts, així com les discussions que hem compartit.

Finalment, el més important, el suport que he rebut per part de la Laura, de familiars i d'amics. Sense vosaltres, tot hagués estat molt més difícil.

- Francesc Molina, Juliol 2021

This thesis has been supported by:

(i) Secretary for Universities and Research of the "Generalitat de Catalunya" and the European Social Fund through the predoctoral Fellowship for the recruitment of new research staff FI 2018. File numbers corresponding to each year: 2018FI_B_00198, 2019FI_B1_00085, and 2020FI_B2_00062;

(ii) "Agencia Estatal de Investigación" (Spanish administration) and European Regional Development Fund under projects WINTER (TEC2016-76409-C2-1-R) and RODIN (PID2019-105717RB-C22); and

(iii) Agency for Management of University and Research Grants (AGAUR, Catalan administration) under Grant 2017 SGR 578.

Abstract

In the transition towards the next generation of wireless technology systems, the increasing number of devices curbs the potential of current wireless networks to cope with such increases in network density. Prompted by the innovations in satellite technology, wireless communications via satellite constitute a cost-effective option to achieve high transmission reliability in remote areas or to create resilient networks to be used in emergency situations. To counterbalance the growing network density, one of the main goals in the uplink is to increase the network spectral efficiency. Three non-exclusive technical approaches are identified in favour of the latter goal:

(i) the application of non-orthogonal multiple access techniques, to overcome the limited availability of orthogonal resources required in conventional access,

(ii) the exploitation of the collision domain, through interference cancellation, and

(iii) the utilisation of multibeam satellites, that, taking advantage of multiantenna technology, enable a more efficient reuse of the spatial domain.

By working on the first two points, this dissertation tackles the problem of massive multiple access. A consensual scheme that meets the main goal and the aim of reducing the interaction between devices and the satellite in the control plane is Enhanced Spread Spectrum ALOHA, which combines spreading-based short-packet transmissions with successive interference cancellation (SIC) on the receiver's side. This combination opens up several design avenues in terms of energy and code allocation to users when a certain amount of channel state information is available to them. Motivated by this scheme, this thesis re-examines previous results reported under capacity-achieving schemes and a genie-aided SIC, and studies the best allocation strategies when the SIC receiver operates nonideally. The former analyses are extended on two fronts: firstly, by adopting decoding and cancellation policies for short-length codes; and secondly, by exploring the unbalance of energy, rate, and reliability. With regard to the first point, this dissertation investigates a system model for a SIC receiver that, inspired by the demodulator adopted in the Enhanced Spread Spectrum ALOHA system, deals with the problems of user ordering and iterative decoding with short packets. With regard to the second point, this dissertation delves into the user-asymptotic regime and into the application of the calculus of variations to derive the stationary point equations corresponding to the optimal allocation rules. One of the main contributions of this thesis is the thorough investigation of *discontinuous* (piecewise continuously differentiable) functions as a class of ordered user-energy distributions to maximise spectral efficiency; an approach which has proved overwhelmingly successful.

More specifically, the system model derived in the present thesis incorporates, progressively, practical aspects of the adopted cancellation receiver in three independent chapters:

1. The first part of this thesis investigates the impact of nonideal decoding and imperfect cancellation on the first iteration of a SIC receiver aided by redundancy-check error control. The system model characterises both non-idealities using known *univariate* functions of the signal-to-interference-plus-noise ratio (SINR) under the assumption of Gaussian interference: the packet error rate (PER) and the residual energy functions. The propagation of packet decoding success/failure events throughout the SIC receiver stages is circumvented in the user-asymptotic regime, since the system model takes a deterministic form. The asymptotically optimal energy and rate allocation is studied for a wide variety of cases including finitely and infinitely many coded modulation schemes for short and unboundedly large packets.

2. The second part of this thesis investigates an iterative SIC receiver and extends the allocation designs derived previously to iterations beyond the first. The derivation of a system model is challenging, since each iteration of the receiver operates with memory with respect to the previous ones, and due to the fact that the decoding operations for the same user in different iterations are statistically dependent. This thesis motivates and states a system model that solves said difficulties by adding minimal complexity to the one adopted previously. More specifically, the model leverages *multivariate* PER functions of the SINRs that each user experiences throughout iterations, and defines bijections to relate the indices of users that remain decoded unsuccessfully at each iteration. The user-asymptotic regime is investigated to reveal mathematical forms to the above model that allow for a thorough understanding of the adopted receiver. Finally, research is conducted to designing smooth allocation functions with free endpoints exploiting the above user-asymptotic model.

3. The third and last part of this thesis studies the user-ordering problem for a SIC receiver to which the strengths received from all users are unknown. This thesis derives an accurate system model for a large-user SIC receiver, which orders users after estimating their symbol energies at the initial stage through preamble cross-correlations. Analytical findings are determined in the user-asymptotic regime, in which system performance is governed by a known kernel. The optimal energy allocation derived for asymptotically many users is shown to obey, in contrast to the practically exponential user-energy distributions obtained before, a piecewise constant function; fact that entails great computational advantages of its application.

Resumen

En la transición hacia la próxima generación de sistemas tecnológicos inalámbricos, el creciente número de dispositivos frena el potencial de las redes inalámbricas actuales para hacer frente a esos aumentos en la densidad de red. Impulsadas por las innovaciones en tecnología satelital, las comunicaciones inalámbricas vía satélite constituyen una opción rentable para lograr una alta fiabilidad de transmisión en zonas remotas o para crear redes reservadas para situaciones de emergencia. Para contrarrestar la creciente densidad de la red, uno de los objetivos principales en el enlace ascendente es aumentar la eficiencia espectral de la misma. En favor de este objetivo, se identifican tres técnicas no excluyentes:

(i) la aplicación de técnicas de acceso múltiple no ortogonal, para hacer frente a la limitada disponibilidad de recursos ortogonales requeridos en el acceso múltiple convencional,

(ii) la explotación del dominio de colisión por el receptor, mediante la cancelación de interferencias, y

(iii) la utilización de satélites multihaz, que, usando la tecnología multiantena, permiten una reutilización más eficiente del dominio espacial.

Esta tesis aborda el problema de acceso múltiple masivo trabajando en los dos primeros puntos. Un esquema consensuado que cumple con el objetivo principal y con el fin de reducir la interacción entre los dispositivos y el satélite en el plano de control es *Enhanced Spread Spectrum ALOHA*, que combina transmisiones de paquetes cortos basadas en el ensanchamiento de la señal con la cancelación sucesiva de interferencias (SIC) en recepción. Esta combinación abre diversas vías para la asignación de energía y código a los usuarios cuando estos disponen de cierta información sobre el estado del canal. Motivado por el esquema anterior, esta tesis reexamina resultados previos bajo análisis teóricos de capacidad y cancelación perfecta, y estudia las mejores estrategias de asignación cuando el receptor SIC opera de forma no ideal. Los análisis anteriores se amplían en dos frentes: en primer lugar, adoptando políticas de decodificación y cancelación adaptadas para paquetes cortos; y, en segundo lugar, explorando el desequilibrio de energía, tasa de transmisión y fiabilidad. Con respecto al primer punto, esta tesis investiga un modelo de sistema para un receptor SIC que, inspirado en el demodulador adoptado en el sistema *Enhanced Spread Spectrum ALOHA*, aborda los problemas de ordenación de usuarios y decodificación iterativa con paquetes cortos. En cuanto al segundo punto, esta tesis se adentra en el régimen asintótico de usuarios y en la aplicación del cálculo de variaciones para derivar las ecuaciones de punto estacionario correspondientes a las funciones de asignación óptimas. Una de las principales contribuciones de esta tesis es el descubrimiento de funciones discontinuas

(continuamente diferenciables a trozos) como una clase de distribuciones de energía ordenada para maximizar la eficiencia espectral; un enfoque que ha demostrado ser abrumadoramente exitoso.

En concreto, el modelo derivado en la presente tesis incorpora, progresivamente y a lo largo de tres capítulos independientes, aspectos prácticos del cancelador de interferencias adoptado:

1. La primera parte de esta tesis investiga el impacto de la decodificación no ideal y de la cancelación imperfecta en la primera iteración de un receptor SIC asistido por control de errores. El modelo de sistema caracteriza ambas no idealidades utilizando funciones conocidas de la relación señal-a-ruido-más-interferencia (SINR) bajo la suposición de interferencia gaussiana: las funciones tasa de error de paquete (PER) y energía residual. La propagación de los eventos de éxito/fracaso en la decodificación de paquetes a lo largo de las etapas del receptor SIC se sortea en el régimen asintótico de usuarios, puesto que el modelo de sistema adopta expresiones deterministas. La asignación de energía y código se estudia en el régimen asintótico de usuarios para una amplia variedad de casos, incluyendo conjuntos formados por un número finito o infinito de esquemas de modulación y corrección de errores para paquetes de longitud finita e infinita.

2. La segunda parte de esta tesis investiga un receptor SIC iterativo y extiende las asignaciones derivadas anteriormente para iteraciones del SIC más allá de la primera. La derivación de un modelo para tal sistema supone un reto, ya que cada iteración del receptor opera con memoria respecto a las anteriores y porque las operaciones de decodificación para un mismo usuario en distintas iteraciones son estadísticamente dependientes. Esta tesis propone justificadamente un modelo de sistema que resuelve dichas dificultades añadiendo complejidad mínima al adoptado con anterioridad. En concreto, el modelo usa funciones PER *multivariable*, cuyos argumentos corresponden a las SINRs que experimenta un usuario a lo largo de las iteraciones del receptor, y define biyecciones para relacionar los índices de los usuarios que permanecen decodificados sin éxito en cada iteración. Se investiga el régimen asintótico de usuarios para revelar expresiones matemáticas del modelo anterior que permitan un completo entendimiento del receptor adoptado. Por último, se investiga el diseño de funciones de asignación continuamente diferenciables con extremos libres haciendo uso del modelo asintótico anterior.

3. La tercera y última parte de esta tesis estudia el problema de ordenación de usuarios en un receptor SIC que desconoce las potencias recibidas de todos ellos. La tesis deriva un modelo de sistema para un receptor SIC que gestiona un gran número de usuarios y los ordena tras estimar sus energías en la etapa inicial mediante correlaciones de preámbulo. En el régimen asintótico de usuarios, se obtienen resultados analíticos en los que el rendimiento del sistema se rige por un *kernel* conocido. Se demuestra que, contrariamente a las distribuciones prácticamente exponenciales obtenidas anteriormente, la asignación óptima de energía derivada para un número infinito de usuarios obedece una función constante a trozos; hecho que conlleva grandes ventajas computacionales en su aplicación.

Resum

En la transició cap a la pròxima generació de sistemes tecnològics sense fils, el creixent nombre de dispositius frena el potencial de les xarxes sense fils actuals per fer front a tals augments en la densitat de xarxa. Impulsades per les innovacions en tecnologia de satèl·lits, les comunicacions sense fils via satèl·lit constitueixen una opció rentable per assolir una fiabilitat de transmissió alta en zones remotes o per crear xarxes que puguin ser utilitzades en situacions d'emergència. Per contrarestar la creixent densitat de la xarxa, un dels objectius principals en l'enllaç ascendent és augmentar l'eficiència espectral d'aquesta. S'identifiquen tres tècniques no excloents en pro d'aquest objectiu:

(i) l'aplicació de tècniques d'accés múltiple no ortogonal, per fer front a l'escassa disponibilitat de recursos ortogonals requerits en l'accés múltiple convencional,

(ii) l'explotació del domini de col·lisió per part del receptor, mitjançant la cancel·lació d'interferències, i

(iii) la utilització de satèl·lits multifeix, que, utilitzant la tecnologia multiantena, permeten una reutilització més eficient del domini espacial.

Aquesta tesi aborda el problema d'accés múltiple massiu treballant en els dos primer punts. Un esquema consensuat que acompleix amb l'objectiu principal i amb la fita de reduir la interacció entre els dispositius i el satèl·lit en el pla de control és *Enhanced Spread Spectrum ALOHA*, que combina transmissions de paquets curts basades en l'eixamplament del senyal amb la cancel·lació successiva d'interferències (SIC) en recepció. Aquesta combinació obre diverses vies per l'assignació d'energia i codi als diferents usuaris quan aquests disposen de certa informació sobre l'estat del canal. Motivat per l'esquema anterior, aquesta tesi reexamina resultats anteriors sota anàlisis teòrics de capacitat i cancel·lació perfecta, i estudia les millors estratègies d'assignació quan el receptor SIC opera de forma no ideal. Els anàlisis anteriors s'amplien sota dos fronts: en primer lloc, adoptant polítiques de descodificació i cancel·lació adaptades per paquets curts; i, en segon lloc, explorant el desequilibri d'energia, taxa de transmissió i fiabilitat. Respecte el primer punt, aquesta tesi investiga un model de sistema per un receptor SIC que, inspirat en el desmodulador adoptat en el sistema *Enhanced Spread Spectrum ALOHA*, aborda els problemes d'ordenació d'usuaris i de descodificació iterativa amb paquets curts. Pel que fa al segon punt, aquesta tesi s'endinsa en el règim asimptòtic d'usuaris i en l'aplicació del càlcul de variacions per derivar les equacions de punt estacionari corresponents a les funcions d'assignació òptimes. Una de les principals contribucions d'aquesta tesi és el descobriment de funcions discontínues (contínuament diferenciables a trossos) com una classe de distribucions d'energia ordenada per

maximitzar l'eficiència espectral; un enfocament que ha demostrat ser profundament exitós.

Concretament, el model de sistema derivat en aquesta tesi incorpora, de forma progressiva i al llarg de tres capítols independents, aspectes pràctics del cancel·lador d'interferències adoptat:

1. La primera part d'aquesta tesi investiga l'impacte de la descodificació no ideal i de la cancel·lació imperfecta en la primera iteració d'un receptor SIC assistit per control d'errors. El model de sistema proposat caracteritza ambdues no idealitats fent ús de funcions conegudes de la relació senyal-a-soroll-més-interferència (SINR) sota la hipòtesi d'interferència gaussiana: les funcions taxa d'error de paquet (PER) i energia residual. La propagació dels esdeveniments d'èxit/fracàs en la descodificació de paquets al llarg de les etapes del SIC s'aborda en el règim asimptòtic d'usuaris, donat que el model de sistema adopta expressions deterministes. Les funcions d'assignació s'estudien en el règim asimptòtic d'usuaris per una àmplia varietat de casos, incloent conjunts formats per un nombre finit o infinit d'esquemes de modulació i correcció d'errors per paquets de longitud finita i infinita.

2. La segona part de la tesi investiga un receptor SIC iteratiu i estén les assignacions derivades anteriorment per a iteracions del SIC més enllà de la primera. La derivació d'un model per a tal sistema suposa un repte, ja que cada iteració del receptor opera amb memòria respecte a iteracions anteriors i degut a que les operacions de descodificació per a un mateix usuari en iteracions diferents són estadísticament dependents. Aquesta tesi proposa justificadament un model de sistema que resol tals dificultats afegint complexitat mínima al model adoptat amb anterioritat. En concret, el model utilitza funcions PER *multivariable* amb arguments corresponents a les SINRs que experimenta un usuari al llarg de les iteracions i defineix bijeccions per a relacionar els índexs dels usuaris que romanen descodificats sense èxit en cada iteració. El règim asimptòtic d'usuaris s'investiga amb l'objectiu de revelar expressions matemàtiques del model anterior que permetin la completa comprensió del receptor adoptat. Per últim, s'investiga el disseny de funcions d'assignació contínuament diferenciables amb extrems lliures fent ús del model asimptòtic anterior.

3. La tercera i última part d'aquesta tesi estudia el problema d'ordenació d'usuaris aplicat a un receptor SIC que desconeix les potències rebudes de tots ells. La tesi deriva un model de sistema per a un receptor SIC que gestiona un gran nombre d'usuaris i els ordena després d'estimar les energies de tots ells en l'etapa inicial a través de correlacions de preamble. En el règim asimptòtic d'usuaris, s'obtenen resultats analítics en els que el rendiment del sistema ve determinat per un *kernel* conegut. Es demostra que, contràriament a les distribucions pràcticament exponencials obtingudes anteriorment, l'assignació d'energia òptima derivada per un nombre infinit d'usuaris presenta una estructura constant a trossos; fet que comporta grans avantatges computacionals en la seva aplicació.

List of Acronyms

3GPP	3rd Generation Partnership Project.
ACRDA	Asynchronous Contention Resolution Diversity ALOHA.
ASE	Asymptotic Spectral Efficiency.
AWGN	Additive White Gaussian Noise.
CDMA	Code-Division Multiple Access.
CRC	Cyclic Redundancy Check.
CRDSA	Contention Resolution Diversity Slotted ALOHA.
CV	Calculus of Variations.
DAMA	Demand Assignment Multiple Access.
DoF	Degree of Freedom.
DS	Direct Sequence or Dense Spreading.
DVB	Digital Video Broadcasting.
DVB-RCS2	Digital Video Broadcasting - Return Channel Satellite; Second Generation.
DVB-RCS	Digital Video Broadcasting - Return Channel Satellite.
E-SSA	Enhanced Spread Spectrum ALOHA.
FBL	Finite Blocklength.
FDMA	Frequency-Division Multiple Access.
FLCV	Fundamental Lemma of the Calculus of Variations.
GEO	Geostationary Orbit.
IBL	Infinite Blocklength.
IEEE	Institute of Electrical and Electronics Engineers.
LDS	Low Density Spreading.
LEO	Low Earth Orbit.
MAC	Multiple Access Channel.
MEO	Medium Earth Orbit.
MMSE	Minimum Mean Square Error.
MUD	Multiuser Detector.

NOMA Non-Orthogonal Multiple Access.

NTNs Non-Terrestrial Networks.

OMA Orthogonal Multiple Access.

PER Packet Error Rate.

PSR Packet Success Rate.

QPSK Quadrature Phase-Shift Keying.

RA Random Access.

RE Residual Energy.

SatComs Satellite Communications.

SE Spectral Efficiency.

SINR Signal-to-Interference-plus-Noise Ratio.

S-MIM S-band Mobile Interactive Multimedia.

SIC Successive Interference Cancellation.

SNR Signal-to-Noise Ratio.

TDMA Time-Division Multiple Access.

Notation

Elemental Notation

$\mathbf{x}$	Column vector.
$\mathbf{x}^H$	Hermitian of vector $\mathbf{x}$.
$\mathbf{0}_p$	Allzeros vector of p elements.
$\mathbf{1}_p$	Allones vector of p elements.
λ^{-1}	Reciprocal of scalar λ.
$f(x)$	Scalar function of the one-dimensional argument x.
$f(\mathbf{x})$	Scalar function of the n-dimensional argument $\mathbf{x}$.
$\dot{f}(x), f'(x)$	First derivative of $f(x)$.
$\ddot{f}(x), f''(x)$	Second derivative of $f(x)$.
$f^{-1}(y)$	Inverse function of $y = f(x)$.
$\ln(\cdot)$	Natural logarithm.
$\log(\cdot)$	Logarithm to the base 2.
$\Pr[\cdot]$	Probability.

Special Functions

$\delta(x)$	Dirac's delta function.
$Q_m(a, b)$	Marcum Q function of order m.
$\mathcal{Q}(x)$	Gaussian Q function.

Operators

∇_t	Gradient operator under variable t.
∂_x	Partial differentiation operator under variable x.
$\mathbb{E}[\cdot]$	Expectation operator.
∇_{EL}	Euler-Lagrange differentiation operator.
$\odot$	Hadamard product operator.

Symbolism

$\sim$	Statistically distributed as.
$\triangleq$	Defined as.
$\approx$	Approximately as.

Random Variables

$f_X(x)$	Probability density function of the random variable X.
$F_X(x)$	Probability distribution function of the random variable X.
$\bar{F}_X(x)$	Tail distribution function of the random variable X.
$\mathcal{N}(\mu, \sigma^2)$	Real Scalar Gaussian distribution with mean μ and variance σ^2.
$\mathcal{CN}(\mu, \sigma^2)$	Complex Scalar Gaussian distribution with mean μ and variance σ^2.
$\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$	Multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$.
$\mathcal{X}_r^2(\lambda)$	Non-central chi-squared random variable with r degrees of freedom and non-centrality parameter λ.

Functional Analysis

$J[x(t)]$	Functional J in $x(t)$.
δJ	First variation of the functional J.
$\delta^2 J$	Second variation of the functional J.
$\mathcal{C}[a, b]$	Space of continuously differentiable functions defined in $[a, b]$.
$\mathcal{C}_p[a, b]$	Space of piecewise continuously differentiable functions defined in $[a, b]$ and comprising p pieces.

Specific Notation

γ_{x}	Transmitted symbol energy over noise power spectral density ratio.
γ	Received symbol energy over noise power spectral density ratio.
$\hat{\gamma}$	Estimated symbol energy over noise power spectral density ratio.
Γ	Signal-to-interference-plus-noise ratio.
$\mathrm{PER}[\Gamma, R]$	Packet error rate versus SINR function of the coding scheme of rate R.
$\mathrm{PSR}[\Gamma, R]$	Packet success rate versus SINR function of the coding scheme of rate R.
$\varepsilon[\Gamma, R]$	Residual energy versus SINR function of the coding scheme of rate R.
$\Phi[\Gamma, R]$	Decoding-cancellation characteristic of the coding system of rate R.

Table of Contents

1 Introduction

Over the years, the forecasts for growth in subscriptions and traffic have been successfully achieved. Thus, the explosive growth of wireless devices is now a widespread trend. In the last forecast report developed by Ericsson [1] in November 2020, it is stated that "5G subscription uptake is expected to be significantly faster than for 4G", and it is shown that the order of magnitude of these new subscriptions is in the billion. In this context, the way these devices communicate is going to be substantially different from how it has developed so far. It is expected that such a high number of devices will be communicating with each other (machine-to-machine) or towards more complex aggregation nodes. Communication links are going to be set up intermittently, during very short periods of time, and without centralised coordination [2]. This new paradigm raises the challenge of reliable communication with short packets.

These emerging innovative settings put current wireless communications up against the wall, and focus the world's eyes on both industry and academia to devise architectures and protocols able to handle such a volume of data without prejudicing the requirements of devices, but rather the opposite, with higher demands on capacity, reliability and low latency. This, in turn, implies the use of more complex hardware architectures and the more efficient use of the available channel resources. One of the options to respond to this demand is to complement the terrestrial cellular infrastructure with non-terrestrial networks (NTNs) [3,4], as for instance, satellite communications (SatComs) or airborne networks. In both cases, the underlying idea is to add some intermediate nodes, such as satellites or aircraft platforms, that connect many devices with a ground station either directly or via space-to-space links.

This dissertation focuses on SatComs, and more concretely, devotes its content to the study of allocation strategies for very many users accessing a satellite with asynchronous random access and the receiver performing interference cancellation decoding. The following sections provide an overall picture of SatComs, which include satellite technology in Section 1.1, and, in Section 1.2, the advanced random access schemes that can counteract the inherent drawbacks of the satellite channel in an effective way. The rest of the chapter describes the contribution of this dissertation: Section 1.3 delimits its scope; Section 1.4 revises the main tools used to design such allocation strategies; and finally, Section 1.5 outlines the technical contributions resulted from this thesis, and its organisation from this chapter onwards.

1.1 Relevant Aspects of Satellite Communications Technology

The key drivers that motivate the push for the integrated satellite terrestrial networks are due to the following positive features affecting services [3]:

1. **Service continuity**. Terrestrial systems cannot guarantee uninterrupted service, as it may be blocked due to high traffic flow in peak hours or service interruption caused by emergency situations (e.g. in natural disasters). In these cases, satellite networks are very prompt to offload, from the current network infrastructure, the part of the traffic which does not demand stringent latency constraints. The latter fits in the context of massive machine-type communications.
2. **Service availability**. As the coverage of satellites is usually large, they are very effective in providing coverage to remote areas where terrestrial networks cannot be deployed for reasons of infrastructure efficiency and cost.
3. **Scalability**. One of the options that make SatComs feasible is the scalability they offer. That is, once the satellite is launched, they allow greater flexibility to support increases in network density.

These positive aspects, however, are set against the intrinsic drawbacks of the satellite channel, which appear, essentially, due to the long distance between devices located on the Earth's surface and the satellite. Recall that the GEO orbit is around 36.000 kilometres altitude and the propagation delay is close to half a second. The LEO orbits are found at an altitude of some thousands of kilometres. The MEO orbits are found in between. This fact somewhat restricts or focuses the main applicability of SatComs to delay-tolerant applications or to the creation of resilient wireless networks to be used in emergency situations. These pros and cons have also motivated the research community to enhance the use of satellite networks further by addressing the technical challenge of increasing their performance, which can generally be addressed by the following three non-exclusive techniques:

1. **Multi-beam satellite technology**. Currently, one of the most prominent techniques is the use of the multi-antenna technology in space-ground communications, which enables high throughput coverage in dense scenarios through multiple beams from the same antenna (array-fed reflectors) that are used to direct transmission to the intended users [5–7].
2. **Non-orthogonal multiple access**. Conventional multiple access schemes are not suitable for satellite scenarios due to the inherent characteristics of the satellite channel, and because the excessively large number of devices poses many difficulties to coordinate the multiple access. Such are the cases of time-, frequency- and code-based channelisation protocols in orthogonal multiple access (OMA), which are difficult to implement in large systems due to the lack of sufficient orthogonal resources available to users [8, 9]. Another example is the classical demand assignment multiple access (DAMA) protocol adopted in the digital video broadcasting - return channel satellite (DVB-RCS) standard, where a scheduler allocates dedicated resources to network subscribers, and which performs inefficiently due to the large round trip time of transport control protocol signalling [10] (e.g. three-way handshake and acknowledgement packets).

In both multi-beam and single-antenna satellites, each beam still covers a large Earth area. In a super-dense scenario, nevertheless, this implies that multiple access needs to exploit the available degrees of freedom more efficiently. Along this line, one of the most competitive techniques is non-orthogonal multiple access (NOMA). It makes possible to counteract the shortage of orthogonal resources required for conventional OMA, while allowing simultaneous access by a larger number of devices sharing one or several degrees of freedom. At the same time, however, this implies that users interfere with each other, as in the multiple access channel, but only in the shared degrees of freedom. This issue can be tackled effectively relying on interference cancellation at the receiving node, as described in the subsequent point.

The mostly dominant NOMA techniques separate users in the power or the code domain [11]. Power-domain NOMA is mostly used in downlink, where the transmitter (usually, the base station) overlaps in time and frequency the signals intended to different users with disparate power levels, and the destination nodes resort to successive interference cancellation (SIC) [12]. The use of dissimilar power levels facilitates the task of detecting the strongest signal in every user [13]. Code-domain NOMA is used either in uplink or downlink. Its main principle prompts the use of spreading-based multiple access through linear codes from a dense or sparse non-orthogonal basis. Multiple access works in the same way as in conventional code-division multiple access (CDMA), where users sign their packets after data encoding, and transmit on the same time and frequency resource. At the receiver, the processing gain is high due to the low cross-correlation between the different spreading waveforms, which can be further improved by adopting SIC. The most well-known code-domain NOMA techniques are as follows:

a) **Dense spreading (DS)** [14, 15]. Signature waveforms are constituted by repetition codes with the second moment of their cross-correlation inversely proportional to the lengths of the employed spreading codes.

b) **Low-density spreading (LDS)** [16, 17]. Extends the DS format by forcing zeros in some dimensions of the previous spreading codes.

Other relevant NOMA techniques in this area combine several domains, such as the code and frequency domains in sparse code NOMA [18].

3. **Exploitation of the collision domain**. Since users share the same collision domain, they inevitably interfere with each other at the receiving node. One feasible option to mitigate this multiple access interference is by resorting to some multiuser detection (MUD) technique. Since this implies an increase in computation complexity, current satellite manufacturers are very sceptical about introducing complex on-board digital signal processing, and is therefore preferable to move the increase in computational complexity to gateway stations [19]. With the focus on code-domain NOMA, several MUD receivers can be adopted for both DS and LDS depending on the required performance/complexity ratio. The optimum detector with maximum-likelihood sequence detection achieves the first-rate performance but its complexity is exponential in the number of users. Less complex MUD schemes can be implemented, such as the belief propagation message-passing

algorithm [20], the classical SIC algorithm after a bank of matched filters or a bank of linear minimum mean square error filters [21], or a linear MUD receiver [15, 16] when computation complexity constitutes a relevant factor. Anyway, the key point is to adopt an architecture capable of resolving packet collisions.

1.2 Advanced Random Access Schemes

In more realistic satellite networks the number of users accessing the channel is usually random. To handle this issue, the scientific community has devised many random access (RA) protocols that involve minimum coordination between transmitting devices and the central node, even leading to their autonomous operation [22–24].

Conventional RA protocols have been widely investigated in the past [25, 26]. The ALOHA system [27] is the pioneering packet-based RA protocol where users transmit on an asynchronous fashion in the same time and frequency resource. The receiver usually operates with a working window of T seconds and, due to the users' non-coordinated activity, the number of users in T seconds follows a Poisson distribution with mean $\lambda > 0$ packets per second, such that

$$\Pr\{k \text{ users in } T \text{ seconds}\} = \frac{\lambda T}{k!} \exp(-\lambda T). \tag{1.1}$$

In this system, packet collision occurs when more than one packet arrive at the central node with partial overlapping, in which case both may be lost. Due to the non-coordinated activity between the different network devices and the high collision probability at increasing traffic loads, the peak throughput is achieved at $0.5e^{-1} \approx 18.4\%$ of packets received without collision. Slotted ALOHA [28] improves the pure ALOHA system by introducing time slot synchronisation at the central node, such that, user packets are enforced to arrive at the beginning of each time slot, instead of randomising the access time. The performance of slotted ALOHA is doubled (that is, $e^{-1} \approx 36.8\%$), but the counterpart is that it requires time slot synchronisation. The latter is, in general, impracticable in large networks since it usually requires the imposition of stringent coordination constraints. Moreover, its application generates controversy in the scientific community because the peak throughput does not provide reliable multiple access to many devices, a fact that, according to the key drivers of the new era of communication, constitutes a limiting factor. Other RA protocols adopt carrier sensing mechanisms to determine when the channel is idle, as in IEEE 802.11 [29], where collision avoidance operates with specific messages to reserve the utilisation of the channel before transmission. Broadly speaking, carrier sensing techniques such as collision avoidance and also collision detection do not perform satisfactorily since the long propagation delay reduces the effective usage of the channel.

In view of the above-mentioned limitations in incorporating the pioneering RA schemes in satellite communications, advanced schemes have been developed on the basis of the earlier ones with the aim of overcoming these shortcomings and achieving substantial performance gains. The promising candidates are the Contention Resolution Diversity Slotted ALOHA introduced in the second generation DVB [30] and in the quasi-synchronous access of interactive services S-MIM [31], and the Enhanced Spread Spectrum ALOHA adopted in the asynchronous return link of S-MIM [32]. Both advanced RA schemes are reviewed in the following subsections.

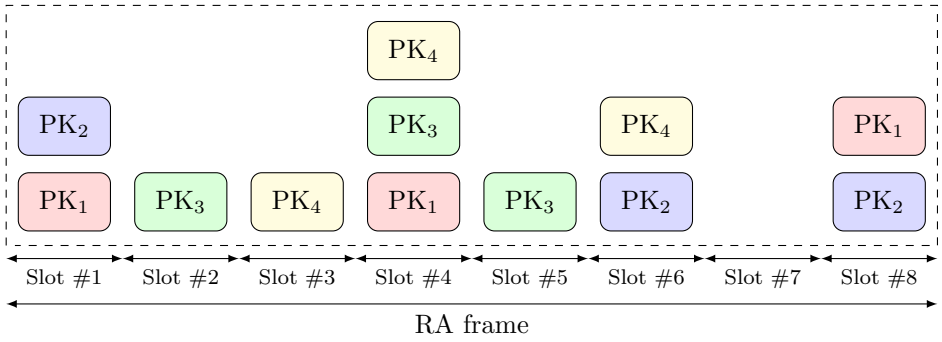

Figure 1.1: CRDSA RA frame with 8 slots per frame and $N_{\text{rep}} = 3$.

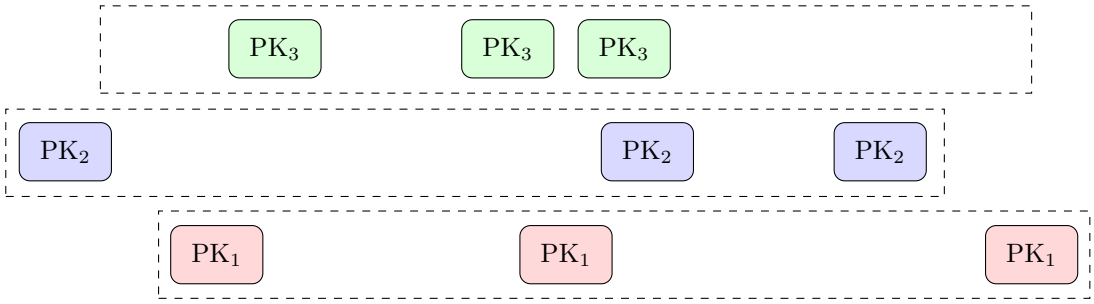

Figure 1.2: ACRDA received frames with 8 slots per frame and $N_{\text{rep}} = 3$.

1.2.1 Contention Resolution Diversity Slotted ALOHA

Contention Resolution Diversity Slotted ALOHA (CRDSA) is an esteemed RA protocol with time slot synchronisation. The interesting technique used by CRDSA is the combination of temporal diversity by transmitters with the exploitation of interference cancellation by the receiving node [33]. A slotted multiple access is implemented at the receiver to divide its operation in frames composed by a number of slots, and users are allowed to transmit N_{rep} replicas of the same packet in different slots chosen randomly. CRDSA takes advantage of different replicas of the same packet to provide more chances to the receiver to decode it successfully. If so, the packet is reconstructed and cancelled in the other time slots where it is present, so that the interference level existing in the working frame is mitigated sequentially as the interference cancellation algorithm progresses, as in the four-user example drawn in Figure 1.1. In this example, high performance is attained if users 3 and 4 are respectively decoded from slots 2 and 3, and later on, subtracted from slot 4 to obtain user 1. Finally, user 2 is decoded from slot 1 after cancelling user 1. System performance highly depends on the distribution of powers received from all users and on the specific decoding order, which opens up many possibilities for system optimisation. With regard to the first, the packet-power unbalance can be exploited if power control is enabled [34], wherein system performance substantially outperformed. CRDSA's successful performance has motivated its last introduction in the standard DVB-RCS2 [30], as well as the possibility of transmitting uplink signalling (e.g. logon and control information) and data traffic in the same payload.

A variant of CRDSA corresponds to the elimination of time slot synchronisation, as proposed in asynchronous CRDA (ACRDA), which allows for a better application in dense scenarios since

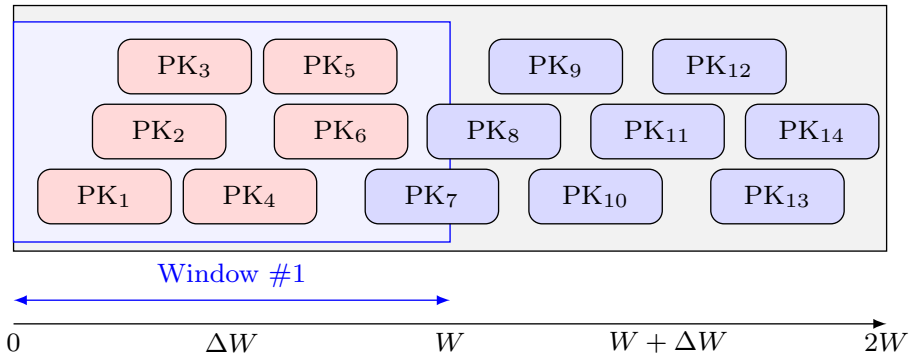

Figure 1.3: E-SSA memory buffer example: first window.

it suppresses the coordination between devices and the receiving station [35]. In this case, RA frames are set local to the transmitters and totally asynchronous from the receiver's perspective, as shown in Figure 1.2 for three users. ACRDA achieves better performance than CRDSA while operating in a truly asynchronous mode [34]. In fact, the RA nature of incoming frames introduces an artificial unbalance exploited by the interference cancellation receiver. Reported analyses in [34] evidence 35% superior throughput with respect to the best-performing CRDSA.

There are also other variants of contention resolution ALOHA, albeit less analysed throughout the literature. Enhanced Contention Resolution ALOHA is such an example that prompts the use of classical combining techniques (selection combining or maximal-ratio combining) to resolve collision patterns that interference cancellation algorithms are not able to handle [36,37].

1.2.2 Enhanced Spread Spectrum ALOHA

In general terms, spreading-based multiple access has a number of advantages for today's satellite communications [38]: firstly, it encourages totally asynchronous transmissions whereby devices can choose either bandwidth or time expansion; secondly, it allows multiplexing a high number of users in the same time-frequency resource; thirdly, multiple access interference at the output of the symbol matched filter resembles Gaussian noise [14], albeit symbols of the same packet are statistically dependent; and fourthly, system performance with large spreading gains and high traffic loads practically attains, by using less complex channel decoding techniques, the performance achieved by very low-rate coding schemes without spreading.

Continuing along the line of uncoordinated RA, Enhanced Spread Spectrum ALOHA (E-SSA) is the most prominent spreading-based RA protocol, and owes its stunning performance to a powerful packet-level demodulator at the heart of which operates an iterative SIC decoding algorithm. E-SSA was initially proposed as a messaging scheme for satellite communications, more concretely, for the return satellite link [39, 40]. It proved to be appropriate for low complexity terminals due to the low power requirements and its high throughput operation. The underlying idea is to enable multiuser access in the absence of coordination through the use of a powerful channel encoder followed by spreading signatures with large time-bandwidth products. On the receiver side, the demodulator operates under a memory buffer where packets from the different users arrive asynchronously. The demodulation and decoding power of the E-SSA system relies on a sliding window-based iterative SIC scheme.

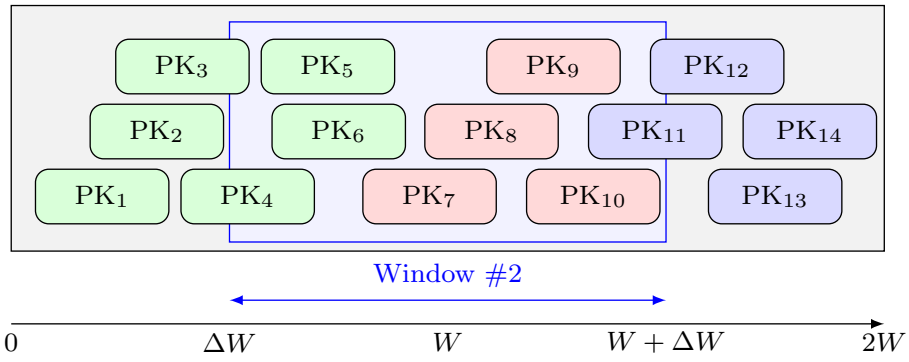

Figure 1.4: E-SSA memory buffer example: second window.

Since the iterative decoding algorithm implemented in the E-SSA system is of great relevance to this thesis, the remaining of this section summarises and exemplifies this algorithm. Firstly, samples are stored in a memory buffer after down frequency conversion and decimation, and a sliding window of size W samples, exceeding the packet duration (in samples), operates by being shifted in steps of ΔW. Figure 1.3 depicts an example of the first window of the E-SSA memory buffer. The following iterative approach is performed to process users in every window [40]:

1. **Packet detection**. Perform packet detection through preamble-based cross-correlations jointly with a comparative threshold, and perform a packet ranking $1, \ldots, K_i$ based on the estimated signal-to-interference-plus-noise ratios (SINRs): $\Gamma[1] \geq \cdots \geq \Gamma[K_i]$.
2. **Decoding and cancellation**. Select the packet with highest SINR, and perform channel decoding. Compute and check the cyclic redundancy check (CRC) to determine if packet error has occurred. If the CRC checks out, regenerate the packet received from that user with the parameters: complex amplitude, delay, and carrier frequency; estimated using the successfully recovered packet. Afterwards, the user is cancelled from the memory buffer.
3. **Iterative procedure**. Return to step 1 and proceed with the decoding-cancellation algorithm in step 2. A halting criterion stops the algorithm when packet detection does not discover new users and all users detected are processed, or when a maximum number of iterations is reached.

After processing the first window through the algorithm above, the sliding window is shifted ΔW samples to process users in $[\Delta W, W+\Delta W]$, as shown for the second window in Figure 1.4.

Note that, since the interference level is mitigated as SIC advances in stages, new users can be detected (step 1.) at each iteration of the demodulation process. Naturally, the computation complexity strongly depends on the number of iterations performed, as well as on the packet detection algorithm. In this respect, the traffic load conditions the number of iterations needed to achieve a minimum quality-of-service requirement [40]. Reported analyses show that the maximum performance is practically achieved by iterating the SIC algorithm few times (between three and five), and that the first iterations are critical to system performance.

1.3 Thesis Scope

This thesis takes up the E-SSA framework, which considers a large number of users connected to the same receiver via satellite and transmitting under long spreading signatures, and designs optimal allocation strategies in terms of energy and code allocated to users when the receiver performs a packet-based SIC scheme with known decoding-cancellation implementation function. For the above purpose, it is considered that users can accurately and individually estimate their channel power gains assuming that the satellite broadcasts a downlink signal employed by the users to obtain their local channel gains. This information allows, since the network has a large number of devices, the deterministic design of allocation functions in accordance with the known asymptotic distribution of channel power gains from all users. The analysis undertaken in this thesis simplifies its practical application since only requires users to use their channel power gain as the argument of the designed allocation function, to obtain the transmitted symbol energy and coding rate; an approach that provides accurate and successful results.

The complexity of the analysis lies in the accurate mathematical modelling of the adopted SIC receiver when some practical features of its operation are considered: the nonideal decoding due to the use of short packets, and the dynamic decode ordering. This thesis covers the statistical analysis of the SIC receiver in non-iterative and iterative operating modes. More specifically, this work elaborates a theoretical framework to compute the performance of the previous systems by analysing the user-asymptotic regime. The model is therefore used to optimise the performance of the aforementioned network in terms of energy, code and conjoint energy and code allocation via the calculus of variations with the goal on the spectral efficiency maximisation. The unbalance between energy and coding rate for different users of the system, already known in theoretical foundations of communications, is explored in depth by incorporating the last analytical findings for short-length codes, which extends them to the energy, code, reliability, and packet decoding latency sets.

1.4 Mathematical Preliminaries to the Calculus of Variations

One of the innovative contributions of this thesis is the application of some tools from the calculus of variations, or their equivalent forms in terms of the Hamiltonian equations [41], to design piecewise continuously differentiable allocations for a network that handles asymptotically many users. This section elaborates a summary of this mathematical tool from the engineering perspective of this author.

The objective is to optimise an application $J : \mathcal{S} \to \mathbb{R}$ from a function space $\mathcal{S}$ to the real field $\mathbb{R}$, denoted *functional*, and defined as

$$J[x(t)] = \int_a^b F[x(t), \dot{x}(t), t]\mathrm{d}t, \tag{1.2}$$

with $F : \mathbb{R}^3 \longrightarrow \mathbb{R}$ a known three-variate function. The functional is defined generally by considering dependence on the sought function $x(t)$, its derivative $\dot{x}(t)$, and the independent variable t. Recall that the over-dot notation (or Newton's notation) indicates differentiation.

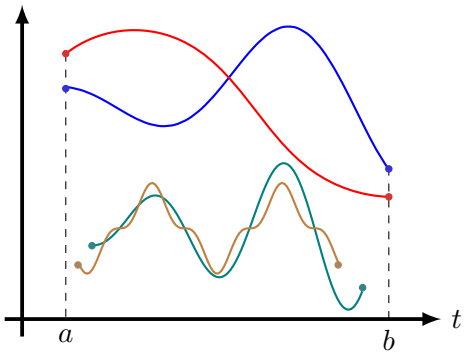

Figure 1.5: Smooth functions with fixed end points $t = a$ and $t = b$ (blue, red), and with free end points (teal, brown).

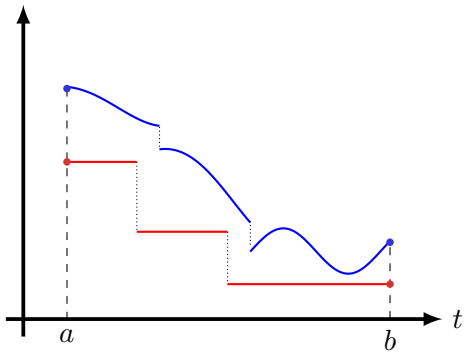

Figure 1.6: Piecewise continuously differentiable functions with fixed end points $t = a$ and $t = b$.

1.4.1 Definition of a Function Space

The main difference between the calculus of variations, an optimisation over an infinite-dimensional space, and vector calculus over a finite-dimensional space, is that the first needs the definition of such a function space from which candidate functions belonging to that space are chosen to satisfy the necessary equations for optimality [41, Chapter 1.2]. Finding such a function space is a major concern since there is usually no function space generic enough to take as a starting point. Only experience and evaluation of the results obtained can dispel this doubt, or at least increase confidence that the function space adopted provides competitive solutions, even if the globally optimal solution does not belong to it. The adoption of an unsuitable function space may lead to finding no solution to the problem or to determining suboptimal solutions. The experience gained in this work tells us that, in general, standard functionals usually admit at most discontinuous solutions (piecewise continuous), and this thesis is an example of such a case. However, there may be singular cases in which even more singular functions solve the target optimisation. The function spaces considered in the different chapters of the thesis are described below, and are exemplified in Figures 1.5 and 1.6:

1. **Continuously differentiable functions with fixed end points**. The most commonly used function space considers continuously differentiable candidates $x(t) \in \mathcal{C}[a, b]$ whose end points lie on the vertical lines $t = a$ and $t = b$. That is, smooth functions on the interval $a \leq t \leq b$ with continuous derivatives, as shown for some examples in Figure 1.5.

2. **Continuously differentiable functions with free end points**. A more general function space results from considering smooth functions with continuous derivatives ($\mathcal{C}[a, b]$) to which, additionally, the end points a, b may vary freely or between predefined limits, rather than lying on the vertical lines $t = a$ and $t = b$. Two examples are shown in Figure 1.5.

3. **Piecewise continuously differentiable functions**. An even more general function space consists of piecewise continuously differentiable functions with a *finite* number of pieces p[1]. Each piece comprises a smooth function with continuous derivatives. We denote this function space $\mathcal{C}_p[a, b]$, whereby $\mathcal{C}_1[a, b] \equiv \mathcal{C}[a, b]$. Some examples are drawn in Figure 1.6.

[1]The mathematicians authors of [41,42] were not able to extend the study to the cases where infinitely many discontinuities may be present.

Note that 1. is a particular case of 2. in which the end points are fixed, and that 1. constitutes a particular case of 3. in which $p = 1$.

1.4.2 Necessary Conditions for Standard Functionals

The fundamental unit in the calculus of variations is called *variation*: a perturbation of infinitesimal magnitude at one point of $\mathcal{S}$ $x(t)$, as

$$x(t) + v(t). \tag{1.3}$$

This definition allows to define the increment of the functional $J[x(t)]$ in $\mathcal{C}[a, b]$ at $x(t)$ as

$$\Delta J = \int_a^b F[x(t)+v(t), \dot{x}(t)+\dot{v}(t), t]\mathrm{d}t - \int_a^b F[x(t), \dot{x}(t), t]\mathrm{d}t, \tag{1.4}$$

and the necessary conditions to find extremals of J are derived by imposing conditions over the linear part of ΔJ in the variation unit. The linear part of ΔJ, δJ, denoted *first variation* can be obtained either with Fréchet differentiation, by expanding the first term of ΔJ up to the first order as

$$\delta J = \int_a^b F_x v(t)\mathrm{d}t + \int_a^b F_{\dot{x}} \dot{v}(t)\mathrm{d}t, \tag{1.5}$$

or with the Gateaux derivative as

$$\delta J = \lim_{c\to 0} \nabla_c J[x(t) + c{\cdot}v(t)] = \int_a^b F_x v(t)\mathrm{d}t + \int_a^b F_{\dot{x}} \dot{v}(t)\mathrm{d}t, \tag{1.6}$$

where $F_x \equiv \partial_x F[x, \dot{x}, t]$ and $F_{\dot{x}} \equiv \partial_{\dot{x}} F[x, \dot{x}, t]$ denote the partial derivatives of the three-variate function $F[x, \dot{x}, t]$ with respect to the first and second arguments, indicated with subscripts.

Both procedures result in the same form for the first variation δJ, which can be still simplified by integrating the last term (linear in $\dot{v}(t)$) by parts, which finally reads

$$\delta J = \int_a^b (F_x - \nabla_t F_{\dot{x}}) v(t)\mathrm{d}t + F_{\dot{x}}\Big|_{t=b} v(b) - F_{\dot{x}}\Big|_{t=a} v(a). \tag{1.7}$$

By identifying terms: (i) the first term corresponds to variations of $J[x(t)]$ when $x(t)$ holds the fixed extrema $x(a) = x_0$ and $x(b) = x_1$, and thus, $v(a) = v(b) = 0$; and (ii) the second and third terms correspond to additional variations over $x(a)$ and $x(b)$. This concludes the following equivalence

$$\max_{x(t)} \int_a^b F[x(t), \dot{x}(t), t]\mathrm{d}t \quad \Longleftrightarrow \quad \max_{x_0, x_1} \max_{x(t)} \int_a^b F[x(t), \dot{x}(t), t]\mathrm{d}t \tag{1.8a}$$

$$\text{s.t. } x(a) = x_0 \quad ; \quad x(b) = x_1 \tag{1.8b}$$

Therefore, a necessary condition for $x(t)$ to be a stationary point of the functional $J[x(t)]$ is that the first variation

$$\delta J = 0 \tag{1.9}$$

vanishes *for any* admissible variation $v(t)$, defined by the perturbations $x(t)+v(t)$ that satisfy the constraints of the problem. This first-order necessary condition works similarly as vanishing the first derivative in vector optimisation. As $\delta J = 0$ holds for every admissible variation, it must be also satisfied for the following particular cases:

1. Null variations on the extrema $v(a) = v(b) = 0$, which gives

$$\int_a^b (F_x - \nabla_t F_{\dot{x}}) v(t) \mathrm{d}t = 0, \tag{1.10}$$

   and by invoking the Fundamental Lemma of the Calculus of Variations[2], it results that

$$F_x - \nabla_t F_{\dot{x}} = 0 \quad \text{in} \quad a < t < b. \tag{1.11}$$

   This equation can still be simplified by defining the Euler-Lagrange operator $\nabla_{\mathrm{EL}}(\cdot) \triangleq \partial_x(\cdot) - \nabla_t \partial_{\dot{x}}(\cdot)$ and by applying it to the integrand F as $\nabla_{\mathrm{EL}} F = 0$.

2. Null variations $v(t)$ in $a < t \leq b$, which results in $F_{\dot{x}} v(a) = 0$ in $t = a$.

3. Null variations $v(t)$ in $a \leq t < b$, which results in $F_{\dot{x}} v(b) = 0$ in $t = b$.

Remarkably, $\nabla_{\mathrm{EL}} F = 0$ (1.11) constitutes a second-order differential equation in $a < t < b$ whose initial values are found by computing $F_{\dot{x}}|_{t=a} v(a) = 0$ and $F_{\dot{x}}|_{t=b} v(b) = 0$.

Truly, the extremals $x_*(t)$ that satisfy $\delta J = 0$ constitute stationary points of J. The above procedure can be extended further by including the second variation of the functional J, actually, the quadratic part of ΔJ, $\delta^2 J$. A posterior analysis shall be performed to guarantee that $x_*(t)$ constitutes a maximum or minimum of $J[x_*(t)]$. This is respectively satisfied when $\delta^2 J < 0$ or $\delta^2 J > 0$. For the interest of the reader, more information can be found in [41].

1.4.3 Necessary Conditions for Broken Extremals

The previous derivation is the main basis of the calculus of variations for "standard" functionals defined in the function space $\mathcal{C}[a, b]$. In some cases, nevertheless, the standard functional J does not admit stationary points in $\mathcal{C}[a, b]$. This singular situations usually occur when the stationary points lie in the boundary of $\mathcal{C}[a, b]$, and thus, a broader function space shall be considered [41, Chapter 3.15]. Such is the case of the function space of piecewise continuously differentiable functions $\mathcal{C}_p[a, b]$, also referred to as broken extremals, where the redefinition of $J[x(t)]$ in the new function space leads to splitting the integrals into p pieces as

$$J = \sum_{k=1}^{p} \int_{t_{k-1}^+}^{t_k^-} F[x(t), \dot{x}(t), t] \mathrm{d}t, \tag{1.12}$$

with $t_0^+ = a$ and $t_p^- = b$. Following the same rationales as before, the first variation of J reads

$$\delta J = \sum_{k=1}^{p} \int_{t_{k-1}^+}^{t_k^-} (F_x - \nabla_t F_{\dot{x}}) v(t) \mathrm{d}t + \sum_{k=1}^{p} \left[F_{\dot{x}}\Big|_{t=t_k^-} v(t_k^-) - F_{\dot{x}}\Big|_{t=t_{k-1}^+} v(t_{k-1}^+) \right]. \tag{1.13}$$

[2] This Lemma constitutes one of the main pillars of the calculus of variations since it allows to establish the necessary conditions for obtaining stationary functions of target functionals.

Likewise as in the previous section, a necessary condition for $x(t)$ to be a stationary point of J in $\mathcal{C}_p[a,b]$ is that it satisfies $\delta J = 0$ for any admissible variation. This concludes:

1. the p equations $\nabla_{\mathrm{EL}} F = F_x - \nabla_t F_{\dot{x}} = 0$ in $t_{k-1}^+ < t < t_k^-$ for $k = 1, \ldots, p$; and
2. the $2p$ conditions $F_{\dot{x}}\big|_{t=t_k^-} v(t_k^-) = 0$ and $F_{\dot{x}}\big|_{t=t_{k-1}^+} v(t_{k-1}^+) = 0$ for $k = 1, \ldots, p$.

Remarkably, the same equations derived for standard functionals are still valid in each piece $k = 1, \ldots, p$, and the $2p$ conditions must hold where the stationary functions have a corner.

1.5 Thesis Organisation and Research Contributions

The remaining of this section is devoted to summarising each of the chapters of the present thesis, as well as to listing the research contributions (journals, letters, and conference proceedings) derived in each of them. For easy readability, the technical contributions are listed in descending chronological order, beginning with the most recent article.

- **Chapter 2: Theoretical Analysis on the Massive Multiple Access Problem**

  The chapter provides a comprehensive overview of the massive multiple access problem from an information-theoretic point of view. The first part of the chapter reviews the capacity of the Gaussian multiple access channel with many users and summarises the sum-rate achievable with OMA and several spreading-based NOMA techniques. The latter includes analyses of DS and LDS formats in the user-asymptotic regime as a convenient tool to analyse massively populated systems. The rest of the chapter presents the practical aspects affecting the successive decoding receiver so as to illuminate the work carried out in the following chapters. The main non-idealities appear, from the transmitter's point of view, by the use of short packets and, from the receiver's point of view, by the unknown identities and strengths received from all users, in which case the receiver proceeds with the SIC receiver algorithm according to energy estimates. The first issue motivates the two next chapters, and the second and last issue motivates the study of the SIC receiver operation under dynamically ordered users.

- **Chapter 3: Optimal Allocation Designs for Cancellation Receivers**

  The scope of the chapter is the design of optimal energy and rate allocation policies for a massive number of DS spread spectrum users that transmit towards a common central node, which relies on a practical SIC scheme to deal with the decoding of short packets. The first part of the chapter conducts the analysis of such a SIC receiver in terms of a simple system model that characterises statistically the effects of channel decoding and imperfect SIC cancellation through univariate functions of the SINR. The user-asymptotic regime is investigated to overcome the randomness present in the previous system model. Within the E-SSA framework, this chapter deals with the derivation of asymptotically optimal allocations accounting for the first iteration of the adopted SIC receiver. The different designs correspond to energy, rate, and conjoint energy-rate allocations under two methodologies that vary the number of encoders users may have available: firstly, for an infinite number; and secondly, for a few of them. The main contributions of the

chapter are the successful adoption of smooth allocations with variable end points, and the characterisation of the stationary point analytic forms to the allocation problems. At the simulation level, reported results present the unbalance of energy, rate and reliability for all users.

The following journal paper and conference proceedings are derived in this chapter:

[J1] **F. Molina**, J. Sala-Álvarez, F. Rey, and J. Villares, "Asymptotically optimal energy and rate allocation for massive multiple access with interference cancellation," *Submitted to IEEE Trans. Commun.*, 2021.

[C1] **F. Molina**, J. Sala-Álvarez, F. Rey, and J. Villares, "Channel-aware energy allocation for throughput maximization in massive low-rate multiple access," in *2019 53rd IEEE Int. Conference on Communications*, Shanghai (China), May 2019, pp. 1-6.

[C2] **F. Molina**, J. Sala-Álvarez, J. Villares, and F. Rey, "Joint energy and rate allocation for successive interference cancellation in the finite blocklength regime," in *2018 6th IEEE Global Conference on Signal and Information Processing*, Anaheim (CA-USA), Nov. 2018, pp. 1-5.

[C3] **F. Molina**, J. Sala-Álvarez, J. Villares, and F. Rey, "Optimal power control law for equal-rate DS-CDMA networks governed by a successive soft interference cancellation scheme," in *2018 43rd IEEE Int. Conference on Acoustics, Speech and Signal Processing*, Calgary (AB-Canada), Apr. 2018, pp. 1-5.

[C4] J. Sala-Álvarez, F. Rey, J. Villares, and **F. Molina**, "Minimum PER user-energy profile for massive SIC receivers under an average energy constraint," in *2017 18th IEEE Int. Workshop on Signal Processing Advances in Wireless Communications*, Sapporo (Japan), Jul. 2017, pp. 1-6.

The author has also contributed in the following poster presentation:

- **F. Molina**, J. Sala-Álvarez, F. Rey, and J. Villares, "Channel-aware energy allocation for throughput maximization in massive multiple access," in *2019 IEEE Communication Theory Workshop*, Selfoss (Iceland), May 2019.

- **Chapter 4: Optimal Allocation Designs for Iterative Cancellation Receivers**

The chapter extends the design of energy and code allocation strategies followed in the previous chapter by incorporating a SIC receiver that persists in processing users decoded unsuccessfully in previous iterations. In this respect, the chapter addresses the iterative decoding problem present in the demodulator adopted in E-SSA (Section 1.2.2), whose specific architecture complicates its analysis since persistent user decoding is a process that operates with memory regarding previous iterations. The first part of the chapter elaborates an original system model to describe the statistical behaviour of the former iterative cancellation receiver. Afterwards, the user-asymptotic expressions of the above model are investigated as well as their usefulness for analysing the behaviour of the SIC receiver deterministically. The second part of the chapter is devoted to optimisation pur-

poses, where the optimal designs are derived for some cases. The latter part of the chapter conducts the analysis of the ultimate performance of the SIC receiver with a large number of iterations through a user-asymptotic analysis based on a fixed-point equation.

The following technical contributions are derived from this chapter:

[**J2**] **F. Molina** and J. Sala-Álvarez, "Energy allocation design for the satellite return channel of a massive NOMA system with interference cancellation," *Submitted to IEEE Trans. Green Commun. Netw.*, 2021.

[**L1**] **F. Molina** and J. Sala-Álvarez, "Average PER performance metrics of iterative successive interference cancellation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 74-77, 2020.

[**C5**] **F. Molina** and J. Sala-Álvarez, "Rate allocation in massive multiple access combining successive decoding with error control," in *2020 54th IEEE Asilomar Conference on Signals, Systems, and Computers - Virtual Event*, Pacific Groove (CA-USA), Nov. 2020, pp. 1-6.

[**C6**] **F. Molina** and J. Sala-Álvarez, "Energy allocation for short-packet massive multiple access with two-iteration successive decoding," in *2019 54th IEEE Int. Conference on Communications - Virtual Event*, Dublin (Ireland), Jun. 2020, pp. 1-6.

- **Chapter 5: The Case of Dynamically Ordered Interference Cancellation Decoding**

  The chapter analyses the performance of a SIC receiver in the presence of unknown symbol energies, which releases one of the principles of successive decoding. Within the framework of the E-SSA system outlined in Section 1.2.2, the chapter addresses the problems of energy estimation and ordering of users. The first part of the chapter studies the main scenarios where the receiver lacks of knowledge about the symbol energies received from all users. A SIC receiver that operates according to the energies estimates is adopted. The second part of the chapter states justifiably an attractive system model to analyse the performance of the adopted receiver. The model is based on statistical averages rather than on long-term averages of many Monte Carlo simulations. The third part of the chapter presents an optimisation in the user-asymptotic regime, and analyses potentially-optimal broken extremal solutions for the asymptotic performance maximisation of the adopted SIC receiver.

  The main results of the chapter are regarding the system model and for optimisation purposes, which have resulted in the following contributions:

  [**L2**] **F. Molina** and J. Sala-Álvarez, "Discontinuous user-energy distribution for dynamically ordered successive interference cancellation," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1673-1677, 2021.

  [**L3**] **F. Molina** and J. Sala-Álvarez, "Asymptotic performance analysis of successive interference cancellation with dynamic user-decoding order," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2931-2935, 2020.

The author has also contributed in the following poster presentation:

- **F. Molina**, J. Sala-Álvarez, F. Rey, and J. Villares, "An optimal control approach to the optimization of dynamically ordered successive interference cancellation," in *2021 IEEE Communication Theory Workshop - Virtual Event*, Banff (AB-Canada), Jun. 2021.

- **Chapter 6: Conclusions and Future Work**

  The chapter summarises the main contributions of this thesis chapter by chapter, and also outlines future research directions toward the extension of the work exposed herein.

- **Other contributions**

  The author of this thesis has also contributed in the following technical works:

  [**L4**] **F. Molina**, J. Villares, F. Rey, and J. Sala-Álvarez, "Decentralized random energy allocation for massive non-orthogonal code-division multiple access," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2306-2310, 2019.

  [**C7**] J. Borras, **F. Molina**, R. López-Valcarce, and J. Sala-Álvarez, "Energy-efficient analog beamforming with short packets in millimeter-wave MIMO systems", in *2020 54th Asilomar Conference on Signals, Systems, and Computers - Virtual Event*, Pacific Groove (CA-USA), Nov. 2020, pp. 1-6.

  [**C8**] **F. Molina** and J. Borras, "Low-complexity switching network design for hybrid precoding in mmWave MIMO systems," in *2019 27th European Signal Processing Conference*, A Coruña (Spain), Sep. 2019, pp. 1-5.

2 Theoretical Analysis on the Massive Multiple Access Problem

The relevant challenge of this thesis is the massive multiple access problem. This chapter provides a comprehensive overview of this problem for many users from a theoretical perspective. Section 2.1 discusses the achievable sum-rate in the Gaussian multiple access channel (MAC) with orthogonal multiple access in Section 2.1.1 and spreading-based non-orthogonal multiple access in Section 2.1.2. The latter includes the asymptotic analysis of direct-sequence spread spectrum with the optimal detector and a suboptimal detector based on successive interference cancellation (SIC). Practical features of successive decoding are analysed in Section 2.2 with the focus on the demodulator adopted in the Enhanced Spread Spectrum ALOHA (E-SSA) system [40]: firstly, with regard to practical decoding of short packets and imperfect SIC cancellation, and later with regard to the decoding order. Section 2.3 exposes the fundamental interplay between four relevant magnitudes in every communication system.

2.1 The Gaussian Multiple Access Channel

When K users simultaneously access a common receiver, the classical problem from the information theory perspective is to determine the set of achievable rates, or the capacity region of the K-user MAC. In the Gaussian case, the capacity region is defined by the enclosing

$$\bigcup_{\mathcal{K}\subseteq\{1,\ldots,K\}} \left\{ R[1\leq k\leq K]\in\mathbb{R}^{+} \text{ such that } \sum_{k\in\mathcal{K}} R[k] \leq \log\left(1+\sum_{k\in\mathcal{K}}\gamma[k]\right)\right\}, \tag{2.1}$$

with $\gamma[k]$ the ratio of symbol energy to noise power spectral density for user k, henceforth denoted symbol energy throughout this thesis, and the maximum sum-rate is attained with Gaussian inputs and via SIC, which results in

$$C_{\text{sum}}^{\text{MAC}} = \sum_{k=1}^{K} \log\left(1+\frac{\gamma[k]}{1+\sum_{i=k+1}^{K}\gamma[i]}\right) = \log\left(1+K\bar{\gamma}\right) \tag{2.2}$$

with $\bar{\gamma} \triangleq \frac{1}{K}\sum_{k=1}^{K}\gamma[k]$ the average symbol energy over all users. Naturally, it is not possible to draw this capacity region since it requires a K-dimensional space. Instead, it is possible to simplify its representation in the two-user plane $R[k]\times R[j]$, as shown in Figure 2.1.

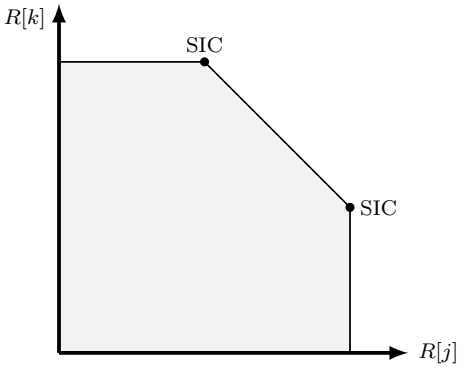

Figure 2.1: Capacity region of the K-user Gaussian MAC in the $R[k]\times R[j]$ plane.

2.1.1 Orthogonal Multiple Access

The case of orthogonal multiple access (OMA) can be analysed in a general way by considering that user k employs a fraction $0 \leq \beta_k \leq 1$ of the totally available degrees of freedom (DoFs), with $\sum_{k=1}^{K} \beta_k = 1$, and that the DoFs allocated to users are not overlapped [43]. No matter if OMA is organised in the time domain (TDMA), the frequency domain (FDMA), or the code domain as in conventional CDMA, or others. Since users are orthogonal in at least one dimension, the attainable sum-rate is the aggregate of individual rates. This gives

$$C_{\text{sum}}^{\text{OMA}} = \sum_{k=1}^{K} \beta_k \log\left(1 + \frac{\gamma[k]}{\beta_k}\right), \tag{2.3}$$

where the signal-to-noise ratio (SNR) of users per DoF is increased by a factor β_k^{-1}, equal to the reciprocal of the occupancy fraction in DoFs, either because users transmit β_k^{-1} times more power per DoF or because they capture a smaller fraction β_k of the noise present in all DoFs.

It is easy to show that $C_{\text{sum}}^{\text{OMA}} \leq C_{\text{sum}}^{\text{MAC}}$, and that its tightness depends on the distribution of received symbol energies $\gamma[1], \ldots, \gamma[K]$. If power control is additionally allowed to users, then the capacity region of the Gaussian MAC is attained at one point with OMA as long as

$$\gamma[k] = (K\bar{\gamma}) \cdot \beta_k \quad \text{for} \quad k = 1, \ldots, K \tag{2.4}$$

symbol energies are allocated proportionally to the fraction of DoFs used by users. Therefore, the channel capacity attained with OMA and optimal power control is

$$C_{\text{sum}}^{\text{OMA}} = \sum_{k=1}^{K} \beta_k \log\left(1 + K\bar{\gamma}\right) = \log\left(1 + K\bar{\gamma}\right), \tag{2.5}$$

which concludes therefore that OMA does not penalise the capacity in the Gaussian MAC.

The direct-sequence (DS) format with orthogonal sequences is a particular case of OMA where users use the same fraction $\beta_k = \frac{1}{N}$ of DoFs [15] (N is the spreading factor). In this case, the maximum sum-rate is achieved when the number of users K is equal to the number of dimensions N and the uniform energy allocation $\gamma[1] = \cdots = \gamma[K] = \bar{\gamma}$ is enforced. This

configuration attains the following capacity

$$C_{\text{sum}}^{\text{Ort. DS}} = \frac{1}{N}\sum_{k=1}^{K}\log\left(1+N\gamma[k]\right) = \frac{1}{K}\sum_{k=1}^{K}\log\left(1+K\bar{\gamma}\right) = \log\left(1+K\bar{\gamma}\right). \tag{2.6}$$

That is, when all users encode data at $\log(1+K\bar{\gamma})$ bits per symbol and transmit at rate $R[k] = \frac{1}{N}\log(1+K\bar{\gamma})$. No matter if the reader is focused on bandwidth expansion while holding the symbol time T_s as $B \approx \frac{1}{T_c} = \frac{N}{T_s}$, or on fixing the system bandwidth $B \approx \frac{1}{T_c}$ in exchange for increasing the symbol duration as $T_s \approx \frac{N}{B}$.

2.1.2 Spreading-Based Non-Orthogonal Multiple Access

The underlying idea of non-orthogonal multiple access (NOMA) is to counteract the limitation of OMA, due to the scarce availability of orthogonal resources, by enforcing users to share some DoFs. One of its advantages is that a larger number of users can simultaneously access the network since the orthogonality constraint between users is released. This fact motivates, for this thesis, the asymptotic analysis in which the number of users tends to infinity, $K \to \infty$.

Before continuing, a brief note on this regime of operation in OMA: recall that the capacity value attained by any OMA technique

$$C_{\text{sum}}^{\text{OMA}} = \log(1+K\bar{\gamma}) \tag{2.7}$$

remains invariant with respect to K as long as users make use of all DoFs without overlapping, and that, for the purpose of establishing a fair comparison with respect to the case evaluated in Section 2.1.1, $\bar{\gamma} \to 0$ as $K \to \infty$ holding fixed $P \triangleq K\bar{\gamma}$. An example of this scenario is when the bit energy over noise power spectral density ratio is fixed. This regime analysed for DS with orthogonal sequences enforces the low-rate operation on all users, which must decrease their transmission rate by a factor proportional to the number of users K, as

$$R[k] = \frac{1}{K}\log(1+K\bar{\gamma}) = \frac{1}{K}\log(1+P) \tag{2.8}$$

but leaves the sum-rate unchanged.

The interest in this operating regime, in the context of massive machine-type communications, is due to the existence of a large number of devices that operate typically under short duty cycles. Nevertheless, the requirement of resource orthogonality is very demanding and, in practice, not feasible to implement. Spreading-based NOMA techniques constitute feasible physical layer configurations for low-rate devices, as they allow access to a large number of them with much less coordination at the cost of increasing the complexity of the common receiver. From the transmitters' point of view, users spread n-symbol packets to nN chips to operate in a N times larger signal space. This enables high processing gain since users intersect each other in only a few dimensions of the signal space. The latter allows for the definition of the fundamental ratio of spreading-based NOMA techniques by taking the effective number of users per dimension K/N [15]. In the Gaussian MAC, NOMA users are separated by the different codebooks they employ; whereas in spreading-based NOMA, users are separated under signature waveforms even if they employ the same codebook. The DS and low-density spreading (LDS)

formats constitute the major representative spreading-based NOMA schemes. Both schemes prompt the use of spreading codes with a low cross-correlation, where LDS includes the exclusive use of sparse codes. In both cases, an interesting regime that provides analytical findings is the large-system limit where $N \to \infty$ as $K \to \infty$ while holding the linear relationship

$$\alpha \triangleq \lim_{K,N\to\infty} \frac{K}{N}. \tag{2.9}$$

Comparative analyses between spreading-based NOMA and OMA schemes must be done for the same product $P = K\bar{\gamma}$, which in the spreading-based NOMA case is $P = \alpha\bar{\gamma}'$, with $\bar{\gamma}'$ the average symbol energy over all users defined for the present spreading-based techniques.

Unfortunately, spreading-based NOMA techniques do not reach the capacity of the Gaussian MAC, except for some special case, because the spreading and despreading counterparts are restricted to linear operators. Its adoption decreases the computational complexity of the receiver but at the price of lower performance. The capacity value is only attained when an interference cancellation-based multiuser detector is adopted and the factor α is sufficiently high, $\alpha \to \infty$. In this respect, the optimal detector for independent randomly spread streams operates jointly with all signals after matched filter detection [15] as shown in Figure 2.2.

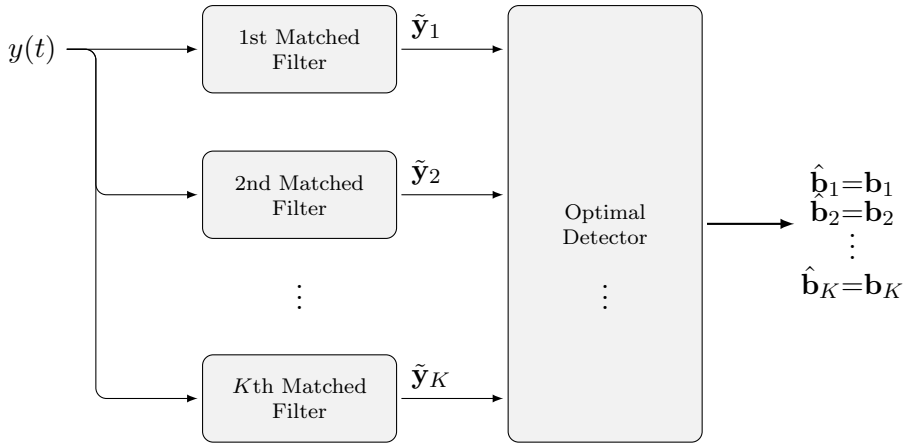

Figure 2.2: Optimal multiuser detection for randomly spread signals.

The optimal detector in the symbol-synchronous case operates as a maximum likelihood sequence detector, whereas for asynchronous streams it does so with dynamic programming after matched filtering [44]. Matched filter is used as a synonym for despreader. In both cases, the computational complexity is linear in the blocklength but exponential in the number of users when signature waveforms are known to the receiver. The spectral efficiency (SE) converges as $K \to \infty$ [15] to the quantity

$$C_{\text{sum}}^{\text{DS+Opt.}} = \alpha \log\left(1 + \bar{\gamma}' - \frac{\mathcal{F}(\bar{\gamma}', \alpha)}{4}\right) + \log\left(1 + \alpha\bar{\gamma}' - \frac{\mathcal{F}(\bar{\gamma}', \alpha)}{4}\right) - \frac{\log(e)}{4\bar{\gamma}'}\,\mathcal{F}(\bar{\gamma}', \alpha) \tag{2.10}$$

with

$$\mathcal{F}\left(\bar{\gamma}', \alpha\right) \triangleq \left[\sqrt{1 + \bar{\gamma}'\left(1 + \sqrt{\alpha}\right)^2} \;-\; \sqrt{1 + \bar{\gamma}'\left(1 - \sqrt{\alpha}\right)^2}\,\right]^2. \tag{2.11}$$

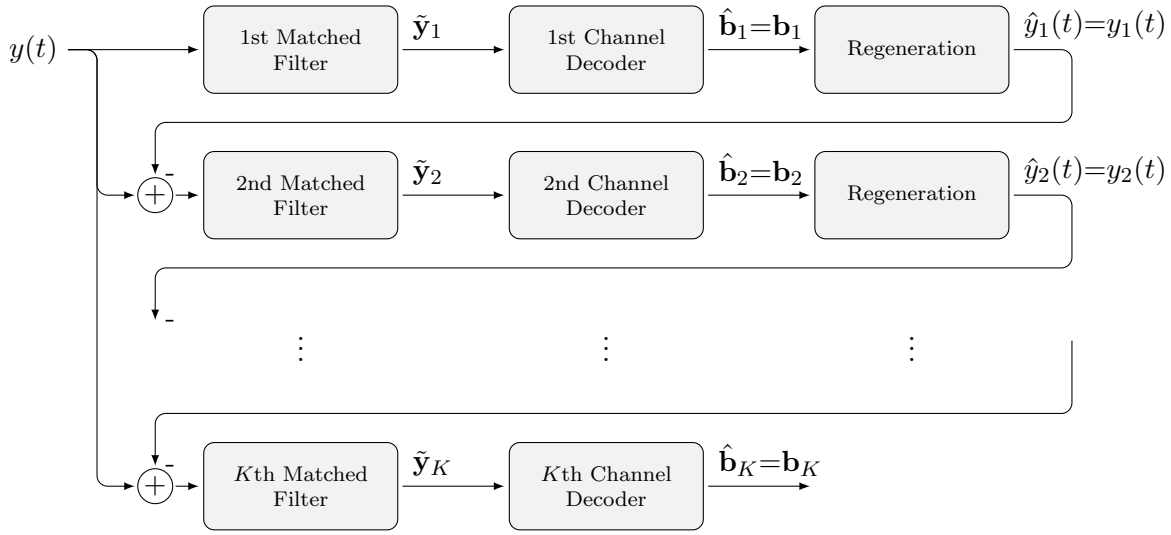

Figure 2.3: SIC algorithm after the bank of K matched filters.

2.1.3 Successive Decoding for Randomly Spread Signals

Since a large number of users will be transmitting on the satellite return link, the optimal detector is not considered to be implementable since for hundreds of users, its computational complexity is huge. A number of multiuser detectors can be adopted if performance is exchanged for computational complexity [14]. A less complex option analysed throughout the literature is the incorporation of a linear equaliser based on the minimum mean square error (MMSE) criterion between the bank of matched filters and the SIC demodulator [21, 45, 46]. The architecture adopted in [45] consists of an iterative loop of a MMSE detector followed by a bank of channel decoders. Its analysis is carried out through a simplified model for the decoder, consisting of a Gaussian approximation of the extrinsic log-likelihood ratios and a known lookup table of the corresponding bit error rates. In [21], the authors express the more complex MMSE detector as a weighted polynomial detector with as many stages as the number of users. In all cases, the performance of such systems is maximised by determining the optimal received symbol energy distribution and assuming that symbol energies from all users are perfectly known to the receiver.

Some practical aspects challenge the performance/complexity ratio of the previous detector, specially so, when the receiver operates nonideally or when symbol energies received from all users need to be estimated. The demodulator adopted in the E-SSA system analyses the low-SNR regime of the MMSE detector, in which case SIC operates after a bank of matched filters [40, 47]. It has the advantages of not requiring the computation of matrix inverses, and decreases the assumption of having perfect knowledge of received symbol energies. The present thesis adopts this scheme as its starting point. More generally, when devices use coded modulation schemes optimal up to the first order in the large blocklength expansion (capacity-achieving), then the best allocation strategy enables reliable communication with a one-iteration SIC. The SIC's algorithm is shown in Figure 2.3. Specifically, the k-th channel decoder operates with the output of the matched filter $\tilde{\mathbf{y}}_k$ to recover the transmitted bit stream $\hat{\mathbf{b}}_k = \mathbf{b}_k$ free of errors, after which, the signal received from that user is reconstructed as $\hat{y}_k(t) = y_k(t)$, and is cancelled from the signal waveform $y(t)$.

A tractable (mathematical) model for the output of the single-user matched filter is the long-code model [14, 15], which assumes that spreading codes with large time-bandwidth products behave equivalently as pseudonoise sequences such that:

(i) signals from other users can be treated as uncorrelated Gaussian noise at the output of the matched filter even if practical (non-Gaussian) modulations are employed, and

(ii) the second moment of the cross-correlation between spreading waveforms used by different users is proportional to the length of signatures employed. For the synchronous case, the latter takes the value $\frac{1}{N}$ whereas for the asynchronous case it equals $\frac{\theta}{N}$, with $0 \leq \theta \leq 1$ a decorrelation factor modelling time-mismatching with respect to the symbol synchronous case ($\theta = 1$). The last remark concerns the introduction of this variable to model practical systems. For the packet-based demodulator outlined in Section 1.2.2, the uncoordinated user-transmissions arriving at the E-SSA's burst demodulator are analysed under $\theta = 0.5$ when the window frames are sufficiently long relative to the packet duration [47].

In the general case, the aggregate SE for K users simultaneously accessing the channel

$$\mathrm{SE} = \frac{1}{N} \sum_{k=1}^{K} \log \left(1 + \frac{\gamma[k]}{1 + \frac{\theta}{N} \sum_{i=k+1}^{K} \gamma[i]} \right) \tag{2.12}$$

is taken as the reference magnitude to evaluate system performance. In this case, as system performance strongly depends on the received energy distribution $\gamma[1], \ldots, \gamma[K]$, the SE (2.12) can be further improved if we let the users vary their transmitted symbol energies so that they arrive at the SIC receiver with the best distribution that satisfies the average constraint

$$\bar{\gamma} = \frac{1}{K} \sum_{k=1}^{K} \gamma[k]. \tag{2.13}$$

In massively populated settings, the large system limit can be duly employed to approximate the performance of a massive access system or to determine the limit expressions, in the number of users K, of the allocation designs. Both are actually within the scope of the present work. In these regime, $K, N \to \infty$ while keeping the traffic load $\alpha = \frac{K}{N}$ fixed, and the discrete user indexing $k = 1, ..., K$ is turned into the continuous user index

$$t \triangleq \lim_{K \to \infty} \frac{k}{K} \quad \text{in} \quad 0 \leq t \leq 1. \tag{2.14}$$

Thus, user-variables indexed by k are substituted by functions of the variable t [48, 49]. In this asymptotic study, users transmit at an asymptotically vanishing rate $\frac{1}{N} R[k] \to 0$ using coded modulation schemes with rate

$$R(t) = \lim_{K \to \infty} R[Kt] = \log \left(1 + \frac{\gamma(t)}{1 + \alpha\theta \int_t^1 \gamma(\tau) \mathrm{d}\tau} \right). \tag{2.15}$$

The SE in the user-limit case, denoted asymptotic SE (ASE), converges as $K \to \infty$ to

$$\mathrm{ASE} = \alpha \int_0^1 R(t) \mathrm{d}t = \alpha \int_0^1 \log \left(1 + \frac{\gamma(t)}{1 + \alpha\theta \int_t^1 \gamma(\tau) \mathrm{d}\tau} \right) \mathrm{d}t. \tag{2.16}$$

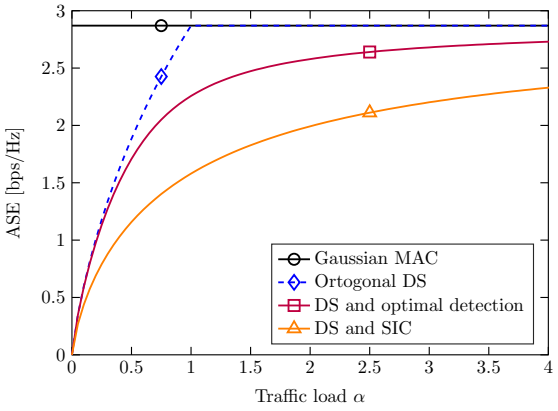

Figure 2.4: ASE of many multiple access techniques in the synchronous case and $P = 8$dB.

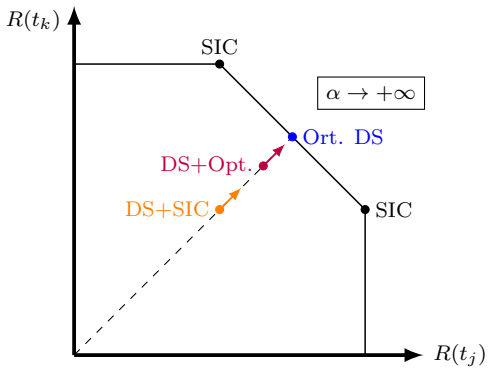

Figure 2.5: The $(K \to \infty)$-user capacity region of the Gaussian MAC in the $R(t_k) \times R(t_j)$ plane.

In this regime of infinite users, vector calculus is meaningless as ASE depends on smooth functions of the variable t. Therefore, the following variational calculus problem is stated in order to obtain the continuously differentiable $\gamma(t) \in \mathcal{C}[0,1]$ that maximises ASE under the average energy constraint $\bar{\gamma}$ (2.13), taken to the limit of users $K \to \infty$:

$$\max_{\gamma(t)} \text{ ASE} \quad \text{s.t.} \quad \bar{\gamma} = \int_0^1 \gamma(t) \mathrm{d}t. \tag{2.17}$$

Making use of the calculus of variations (see Section 1.4 for more details)[1], the maximum ASE is attained with the uniform SINR profile $\Gamma(t) = \bar{\Gamma}$

$$\Gamma(t) = \frac{\gamma(t)}{1 + \alpha\theta \int_t^1 \gamma(\tau)\mathrm{d}\tau} \quad \text{with} \quad \bar{\Gamma} \triangleq \alpha^{-1} \ln(1 + \alpha\theta\bar{\gamma}), \tag{2.18}$$

which leads to an exponential distribution of all users' energies, such that, users are decoded with the same signal-to-interference-plus-noise ratio (SINR) as SIC progresses throughout stages. Then, users lie in the bisector of the capacity region as they encode information at the same rate $R(t) = \log(1 + \bar{\Gamma})$. The attained ASE admits the following closed-form expression

$$C_{\text{sum}}^{\text{DS+SIC}} = \alpha \log(1 + \alpha^{-1} \ln(1 + \alpha\theta\bar{\gamma})), \tag{2.19}$$

which converges as $\alpha \to \infty$ to the maximum sum-rate of the Gaussian MAC ASE $\to \log(1 + \alpha\bar{\gamma})$ in the symbol synchronous case and $P = \alpha\bar{\gamma}$ fixed.

2.1.4 System Performance Evaluation of OMA and NOMA Techniques

The competitiveness of each multiple access technique is shown in Figure 2.4 in terms of their achievable ASEs. The resulting capacity region is drawn in Figure 2.5, where black lines indicate the achievable bounds of the Gaussian MAC with many users. Recall that the optimal strategy in OMA with orthogonal DS is that users perform channel inversion and arrive at the central

[1]The extremal of the functional ASE (2.16) can be found straightforwardly under $z(t) \triangleq \int_t^1 \gamma(\tau)\mathrm{d}\tau$ and the initial value $z(0) = \bar{\gamma}$, by solving the equation $\partial_z R(t) - \nabla_t \partial_{\dot{z}} R(t) = 0$ with $R(t)$ defined in (2.15).

node with an equal power and rate allocation, in which case the capacity region results symmetric and users lie in the bisector (black dashed lines). Rather, the optimal allocation strategy with non-orthogonal DS and a SIC receiver is that users transmit at the same rate and arrive with exponentially distributed symbol energies. The optimal transmission strategy preserves fairness in rate allocation but at the price of unbalancing the energies received from different users. In all cases, as the traffic load α increases, users' allocation rates move on the bisector of the capacity region until they attain the capacity of the Gaussian MAC as $\alpha \to \infty$. Broadly speaking, the performance achieved by different spreading-based NOMA techniques when the blocklength goes to infinity is found to be far from the capacity of the Gaussian MAC, except when the traffic load α is sufficiently high or the average symbol energy $\bar{\gamma}$ is sufficiently low. Its application, nonetheless, provides high computational benefits, specially so, when the number of users is high and the transmission rates are low, and also because its application enables high processing gains at low signal-to-noise ratios.

2.2 Successive Interference Cancellation: Practical Features

Recall that the main principles of successive decoding are that transmitters and the receiver agree a decode ordering, and that channel decoding operates free of errors. After processing one user, the receiver can always proceed to subsequent users after cancelling the last user decoded [43]. Both principles affect the cancellation receiver adopted in the E-SSA system [40], summarised in Section 1.2.2. Actually, the major part of this thesis devotes its content to analysing the impact of the previous features in practical machine-type communication scenarios where the user-decoding order may not be known to the receiver, and/or channel decoding fails due to the operation with short packets. In a broader context, both features constitute critical steps that put in doubt the previous studies based on the underlying principles of successive decoding. The following subsections develop in more detail the previous ideas.

2.2.1 Decode Ordering

The capacity region of the K-user Gaussian MAC is found to have $K!$ corners of equal capacity attainable by means of SIC, each corresponding to a unique allocation of rate and power to every user. Recall that the primary assumption in all capacity analyses is that the receiver knows the user decoding order. Nevertheless, the previous assumption may not hold in all settings. More concretely, in scenarios where the channel varies sufficiently rapid or in systems constrained by low coordination, such an assumption may be meaningless. Such is the case of the satellite framework envisaged in E-SSA [40]. In this case, the decoding order is not known to the receiver, and thus, it bases the SIC implementation on SINR estimates from preamble cross-correlations. This particular feature is analysed extensively in Chapter 5.

2.2.2 The Finite Blocklength Regime

All the previous theoretical analyses are only valid when the K devices make use of random codebooks with infinite-length $n \to \infty$ and power-constrained codewords $\frac{1}{n}\|\mathbf{x}_m^k\|_2^2 = \gamma_k$ for

$m = 1, \ldots, 2^{nR_k}$, and adjust their coding rate R_k as a function of: the SNR γ_k in the point-to-point Gaussian channel as $R_k = \log(1+\gamma_k)$; and the SINR Γ_k in the Gaussian MAC as $R_k = \log(1+\Gamma_k)$. Failing that, the above expressions remain as valid if the transmitted packets are sufficiently long. The machine-to-machine context, however, prompts the transmission of packets with finite (non-asymptotic) payload. This may cause that the adopted channel decoder does not always succeed, combined with the fact that signals received from all users cannot be perfectly reconstructed neither perfectly cancelled from the input signal [50,51]. With the focus on the capacity region of the Gaussian MAC, the latter rationales entail that a few corners, possibly a unique corner, achieve superior performance relative to the rest.

2.2.2.1 The Point-to-Point Gaussian Channel

How far are the former analyses from the performance attainable when transmitters make use of finite-length coding systems? Broadly speaking, the finite blocklength (FBL) regime challenges the decoding performance of the well-known channel encoders. One of the main differences is that the power constraint used in the infinite blocklength (IBL) regime can be understood, in the FBL regime, as two power constraints: the short-term or per-codeword power constraint $\frac{1}{n}\|\mathbf{x}_m\|_2^2 = \gamma$ for all $m = 1, \ldots, M$; or the long-term or average power constraint over all the M equiprobable codewords $\frac{1}{n}\frac{1}{M}\sum_{m=1}^{M}\|\mathbf{x}_m\|_2^2 = \gamma$. Recent literature has gone to great lengths to determine the maximum channel coding rate defined for n-symbol codewords as $R = \sup \frac{\log M}{n}$, so that a codebook achieving the *average* packet error probability ε can be constructed. In this case, unlike in the IBL regime, Gaussian inputs do not constitute optimal distributions under FBL constraints [52]. The most important result was derived in [53], wherein the authors show that an achievable result to the channel coding rate $\frac{\log M}{n}$ at blocklength n with an average packet error probability ε and a per-codeword power constraint of γ is

$$R = \frac{\log M}{n} \approx \log(1+\gamma) - \sqrt{\frac{V(\gamma)}{n}} \cdot \mathcal{Q}^{-1}(\varepsilon), \tag{2.20}$$

with $V(\gamma) \triangleq (1-(1+\gamma)^{-2})\log^2 e$ and $\mathcal{Q}(x)$ the tail distribution of the standard Gaussian density function. Note that the classical expression of the Shannon's rate achieved with error-free decoding is recovered as $n \to \infty$. If codebooks are constrained to be Gaussian i.i.d., then, $V(\gamma)$ shall be substituted by $V_{\text{iid}}(\gamma) = \log^2 e \cdot 2\gamma/(1+\gamma)$ [54]. The interpretation of (2.20) indicates that to transmit at a fraction η of the channel capacity $C(\gamma) \triangleq \log(1+\gamma)$ as $R = \eta C$, then the blocklength shall be proportional to V/C^2 for a given packet error probability [53]

$$n \propto \frac{1}{(1-\eta)^2} \cdot \frac{V(\gamma)}{C^2(\gamma)}. \tag{2.21}$$

In general, the maximum attainable rate in the point-to-point Gaussian channel with a blocklength constraint results severely degraded with respect to the classical Shannon's rate. The procedure the authors in [53] follow to obtain (2.20) is based on the analysis of the dispersion of mutual information, an study absolutely non trivial. Generically, the authors thought of the behaviour of a communication channel as a "bit pipe" of random length $\mathcal{H}$, measured in bits [55]. In n channel uses, the distribution of $\mathcal{H}$ is Gaussian with mean $nC(\gamma)$ bits and variance $nV(\gamma)$.

An error occurs when the number of transmitted bits $\log M$ in n channel uses is higher than $\mathcal{H}$. Then, the average packet error probability ε is

$$\Pr[\mathcal{H} < \log M] = \mathcal{Q}\left(\frac{nC(\gamma) - \log M}{nV(\gamma)}\right) = \varepsilon. \tag{2.22}$$

After some straightforward manipulations, it is easy to obtain (2.20), which approximates very well the behaviour of optimal coding schemes up to the second order in the large-n expansion for blocklengths $n \geq 100$. For the interest of the reader, the expressions corresponding to even lower blocklengths are derived in [56].

From a more practical point of view, the second order coding rate (2.20) can be closely achieved by using, instead of the randomly generated codebooks in the IBL regime, more complex coding systems. As an example, the authors in [57] show that the performance achieved by a tail-biting convolutional code with 2^{14} states (increasing memory) and by a low-density parity check code constructed over a finite field of 256 elements (increasing alphabets) are very close to the best family of codes described by (2.20).

2.2.2.2 The Gaussian Multiple Access Channel

Another interesting analysis is how the former expressions are translated to the Gaussian MAC. The problem is actually very challenging and no closed form expressions for the *exact* value of its maximal channel coding rate are known to date. Without entering into extensive mathematical details, one of the difficulties of the problem is that codewords transmitted by independent users satisfy $\frac{1}{n}\|\mathbf{x}_{m_k}^k\|_2^2 = \gamma_k$ for $m_k = 1, \ldots, M_k$, whereas this does not imply, necessarily, that the joint message satisfies $\frac{1}{n}\|\sum_k \mathbf{x}_{m_k}^k\|_2^2 = \sum_k \gamma_k$. Different approaches have been followed in order to delimit, at least, the region of maximum sum-rate for the case studied. In this sense, some authors have extended the result obtained for the point-to-point Gaussian channel to the Gaussian MAC by turning the univariate channel dispersion function $V(\gamma)$ to a positive-definite dispersion matrix $\mathbf{V}$ of dimension $(K+1) \times (K+1)$ [52, Section III.A]. In particular, for the two-user setting, an achievable pair of rates $(\frac{\log M_1}{n}, \frac{\log M_2}{n})$ to the maximum sum-rate is

$$\begin{bmatrix} \frac{\log M_1}{n} \\ \frac{\log M_2}{n} \\ \frac{\log M_1}{n} + \frac{\log M_2}{n} \end{bmatrix} \in \begin{bmatrix} C(\gamma_1) \\ C(\gamma_2) \\ C(\gamma_1+\gamma_2) \end{bmatrix} - \frac{1}{\sqrt{n}}\mathcal{Q}^{-1}\left(\varepsilon; \mathbf{V}(\gamma_1, \gamma_2)\right). \tag{2.23}$$

The term $\mathcal{Q}^{-1}\left(\varepsilon; \mathbf{V}(\gamma_1, \gamma_2)\right)$ produces a three-dimensional vector from the probability of the three-variate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{V}(\gamma_1, \gamma_2))$ being higher than $1-\varepsilon$.

The approach followed in [58] resulted into the attractive outer bound derived for the two-user case

$$\frac{\log M_1}{n} + \frac{\log M_2}{n} < \log(1+\gamma_1+\gamma_2) - \sqrt{\frac{V(\gamma_1+\gamma_2)}{n}}\mathcal{Q}^{-1}(\varepsilon), \tag{2.24}$$

which, for a fair comparison with the point-to-point case, it must be fulfilled that $\gamma_1+\gamma_2 = \gamma$. The importance of (2.24) is the assertion that the best communication strategy in the Gaussian MAC under blocklength constraints experiences a performance degradation, relative to $C(\gamma_1+\gamma_2)$,

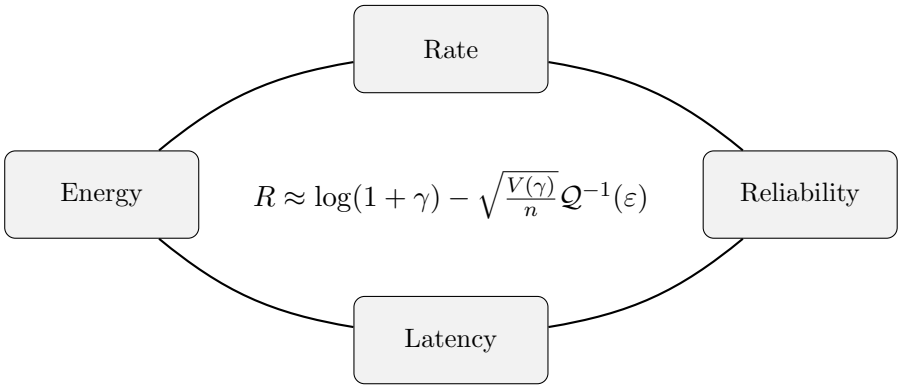

Figure 2.6: The fundamental interplay.

higher than that experienced in the point-to-point channel under the power constraint $\gamma_1+\gamma_2$.

Remarkably, joint message coding is better than independent message coding plus successive interference cancellation. Equivalently, in the Gaussian MAC, independent message encoding is as good as joint message coding only in the IBL regime. As a last curiosity, the resulting achievable region obeys curved shapes rather than the sharp corners present in the pentagonal region obtained in the IBL regime.

2.3 The Fundamental Interplay in Successive Interference Cancellation

The use of interference cancellation-based decoding requires the introduction of such a power and code allocation scheme that enforces certain unbalance between power and rate of users. It is known throughout the literature that, the maximum sum-rate under a SIC receiver is achieved by enforcing a sensitive interplay between the energy and rate of the different users at the receiver's input. As an example, the above unbalance can be straightforwardly evidenced on the basis of the following two study cases: firstly, when users transmit at the same rate $\bar{R}$ (in bits per channel use), the symbol energy distribution that maximises the network sum-rate under reliable communication is the exponential distribution

$$\gamma[k] = \beta \cdot 2^{-k\bar{R}} \quad \text{for} \quad k = 1, \ldots, K \tag{2.25}$$

with $\beta \triangleq 2^{K\bar{R}}(2^{\bar{R}} - 1)$; and secondly, when users arrive with equal symbol energies $\gamma[1] = \cdots = \gamma[K] = \bar{\gamma}$, in which case the rate transmitted by every user reads

$$R[k] = \log\left(1 + \frac{\overline{\gamma}}{1 + (K-k)\overline{\gamma}}\right) \quad \text{for} \quad k = 1, \ldots, K, \tag{2.26}$$

and which behaves as the allocation $R[k] \approx \bar{\gamma}/(1 + (K-k)\bar{\gamma})$ nats/channel use in the low-SNR regime, and as $R[k] \approx \log(1 + \frac{1}{K-k})$ nats/channel use in the high-SNR regime.

The most important analyses regarding the channel coding rate in the FBL regime (2.20) have extended the former interplay between energy and rate for the case when SIC deals with packets of finite length, according to four *fundamental magnitudes*: energy, rate, packet decoding latency (blocklength) and reliability. This thesis provides a theoretical framework to assess the impact of

the practical effects described in Section 2.2 on the performance of SIC. Throughout the thesis, the above problems affecting the SIC receiver are duly studied, and theoretical designs for the energy and code allocation to many users are constructed taking into account the interaction between the magnitudes above.

And now, without further ado, the work carried out by the author of this thesis begins from this point onwards.

3 Optimal Allocation Designs for Cancellation Receivers

One of the most important challenges in wireless multiple access systems is how to transmit short packets efficiently, since the concept of reliable communication, first introduced in pioneering information-theoretic analyses, should be substituted by the more adequate concept: *practical reliability*. In the Gaussian multiple access channel (MAC), the introduction of finite block-length constraints heavily degrades the maximum attainable rate even for the best performing family of codes. This chapter investigates the combination of finite-length channel codes with successive interference cancellation (SIC), which involves the introduction of some tweaks in the classical SIC algorithm illustrated in Figure 2.3. More concretely, this chapter is inspired by the demodulator implemented in Enhanced Spread Spectrum ALOHA (E-SSA) [40], and aims at providing a theoretical framework (model) to analyse an imperfect SIC receiver, without having to resort to lengthy Monte Carlo simulations for validation. The model is used in Section 3.5 to design the best transmission strategy in terms of energy and code allocated to asymptotically many users. The design is effectuated first for a dense number of encoders attaining the maximal channel coding rate up to the first and second order, and complemented later for a few of them for practical purposes. Numerical studies analyse in depth the unbalance of energy, rate and reliability for different users. Last but not least, the practicality of the former allocation is explored intensively for a practical SIC implementation simulated at low level.

3.1 Problem Statement

First of all, to contextualize the problem, the scenario of interest considers very many nodes transmitting to a satellite which relays the aggregate signal towards a ground station as shown in Figure 3.1. A possible context for such a scenario is when several low-complexity devices, e.g. groups of sensors that collect information of many magnitudes, transmit data to aggregation nodes that forward the information towards a central node via a satellite link. The analysis in this chapter focuses on a satellite service for the network of a large number K of aggregation nodes, (henceforth, users) connected to the gateway through slowly time-varying channels with instantaneous channel power gains

$$h[1] \geq \cdots \geq h[k] \geq \cdots \geq h[K]. \tag{3.1}$$

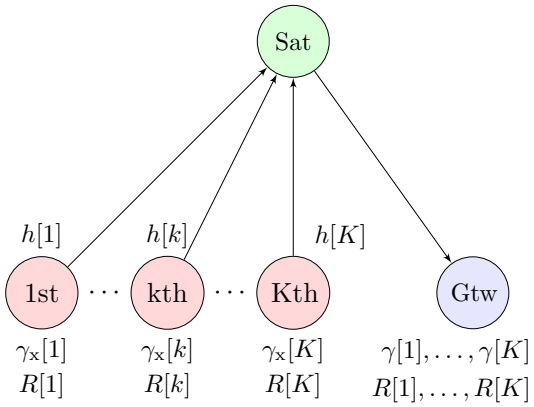

Figure 3.1: An example of the envisaged satellite return link.

Users are endowed with medium/high complexity terminals with directive antennas. Taking advantage of sufficiently high directive antennas, channel power gains are dominated by strong line-of-sight paths, which can be obtained from the directions of arrival of users' signals.

In this MAC, the K users have available p coded modulation systems, or simply encoders, $\mathcal{R} \triangleq \{\mathcal{R}_1, \ldots, \mathcal{R}_p\}$ with n_e-symbol blocklength constraints. The i-th encoder $\mathcal{R}_i$ maps k_i bits to n_e symbols, so that the coding rate is $R_i = k_i/n_e$ bits per symbol. $f_i(\cdot)$ denotes the encoding function of the i-th encoder and $f_i^{-1}(\cdot)$ its decoding function. The k-th user operates as follows: (i) encodes the information vector $\mathbf{b}_k$ as $f_i(\mathbf{b}_k)$ to generate the packet $\mathbf{s}_k$ consisting of n_e coded symbols and n_o preamble symbols, with $n \triangleq n_o + n_e$. The coding rate of user k is $R[k] \in R_{1 \leq i \leq p}$; and (ii) adopts the direct-sequence (DS) spread spectrum format to modulate each symbol of the preamble and the payload. For the latter, and inspired by E-SSA, user k takes the spreading code $\mathbf{c}_k$ from a long sequence with period far exceeding the symbol duration to generate the unit-energy signature waveform for symbol m as

$$c_{k,m}(\mathrm{t}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{c}_k[i + mN] p\left(\mathrm{t} - \frac{i}{N} T\right). \tag{3.2}$$

N is the spreading factor, T is the symbol time and $p(\mathrm{t})$ is the chip pulse. The non-italic variable t denotes time. The second moment of the cross-correlation between different waveforms is $\frac{\theta}{N}$ with $0 \leq \theta \leq 1$. Then, the baseband signal transmitted by user k is

$$x_k(\mathrm{t}) = \sqrt{\gamma_{\mathrm{x}}[k]} \sum_{m=1}^{n} \mathbf{s}_k[m] c_{k,m}(\mathrm{t} - mT), \tag{3.3}$$

where $\gamma_{\mathrm{x}}[k]$ denotes its transmitted energy per symbol (over noise energy per unit bandwidth).

When the channel of each user varies sufficiently slowly so as to assume that transmitted packets are affected by stationary flat-fading (in practice, slowly time-varying channels with multiplicative power gains $h[k]$ (3.1)), the baseband signal received at the gateway is

$$y(\mathrm{t}) = \sum_{k=1}^{K} \sqrt{\gamma_{\mathrm{x}}[k] h[k]}\; e^{j\Omega_k} \sum_{m=1}^{n} \mathbf{s}_k[m] c_{k,m}(\mathrm{t} - mT - \tau_k) + w(\mathrm{t}). \tag{3.4}$$

$\Omega_{1\leq k\leq K}$ are the respective (random) carrier phases, $\tau_{1\leq k\leq K}$ are the end-to-end delays, and $w(\mathrm{t})$ is additive white Gaussian noise.

This chapter analyses a simplified version of the demodulator implemented in E-SSA [40], in which the K users are processed resorting to a SIC approach constituted by one iteration and one stage per user. The nonideal decoding and cancellation operations of the imperfect SIC receiver are analysed in depth. Other SIC approaches more close to [40] are analysed in Chapters 4 and 5. As interference cancellation-based strategies exploit system performance under appropriate allocation designs, this chapter investigates the best energy and rate allocation for all users when they estimate accurately their channel power gains from a downlink pilot. Clearly, depending on the implemented SIC strategy, the allocation design may be problematic due to randomness involved in the successful or unsuccessful decoding of users. This thesis circumvents this problem in the user-asymptotic regime since, as K increases, the allocation problem can be addressed in a deterministic way by deriving an asymptotic allocation $g(h)$ as a function of the asymptotic distribution of channel power gains $\lim_{K\to\infty} h[1],\ldots,h[K]$. Then, the designed allocation is used by user k to determine its transmitted symbol energy $\gamma_{\mathrm{x}}[k]$ and coding rate $R[k]$ in an autonomous way, as

$$(\gamma_{\mathrm{x}}[k], R[k]) = g(h[k]). \tag{3.5}$$

In this chapter, both the transmitted symbol energy and rate distributions $\gamma_{\mathrm{x}}[1],\ldots,\gamma_{\mathrm{x}}[K]$ and $R[1],\ldots,R[K]$ are designed so as to maximise the user-aggregate spectral efficiency (SE) subject to an average energy constraint over all users $\bar{\gamma}_{\mathrm{x}}$, as

$$\max_{R[k],\gamma_{\mathrm{x}}[k]} \frac{1}{N}\sum_{k=1}^{K} R[k]\,(1-\mathrm{per}[k]) \quad \text{s.t. } \frac{1}{K}\sum_{k=1}^{K}\gamma_{\mathrm{x}}[k] = \bar{\gamma}_{\mathrm{x}}, \tag{3.6}$$

with $\mathrm{per}[k]$ the packet error probability corresponding to user k. $1-\mathrm{per}[k]$ is its reliability.

This chapter first overviews in Section 3.2 the state of the art relative to practical SIC algorithms with decoders for short packets, and the allocations already designed in both terrestrial and satellite settings. Subsequently, the modelling of a practical SIC is addressed in Section 3.3 for a finite number of users, and, in Section 3.4, for asymptotically many users. The rest of this chapter devotes Section 3.5 to designing, in the user-asymptotic case, the allocation functions for a variety of cases. Explicit derivations are moved, for clarity of explanations, to the Appendix 3.A located at the end of this chapter. Finally, conclusions are offered in Section 3.6.

3.2 State of the Art

The most relevant aspect to consider is the architecture adopted for the multiuser receiver, which for the optimal detector consists of a joint decoding algorithm such as belief propagation or maximum-likelihood sequence detection after a bank of matched filters [15]. A computationally lighter architecture is a linear minimum mean square error (MMSE) equaliser followed by SIC [45,59], which provides near-optimal performance at low traffic loads, and moderate performance in high activity load situations. In principle, its complexity is lower, but it requires the perfect knowledge of received symbol energies from all users and, if the number of users is high, inverting

a matrix of large dimension $K \times K$. Some approaches have agreed to analyse the suboptimal MMSE detector as a multi-stage detector [21,60] as it allows for obtaining some analytical results. In practice, nonetheless, E-SSA [40] considers the low-SNR regime of the MMSE equaliser by adopting a bank of matched filters followed by SIC, whose specific implementation for practical settings with short-packet decoding proves of substantial relevance.

The best known SIC strategies are distinguished by the adopted cancellation policy. For a hard SIC, the binary outputs after hard decision are used for cancellation [61], whereas a soft SIC operates with the symbol reliabilities just before the hard (binary) decision. Soft cancellation [45] outperforms the hard cancellation since it performs a more conservative cancellation when the decoding is less reliable [62], specially so at low SNRs. Figure 3.2 depicts its block diagram.

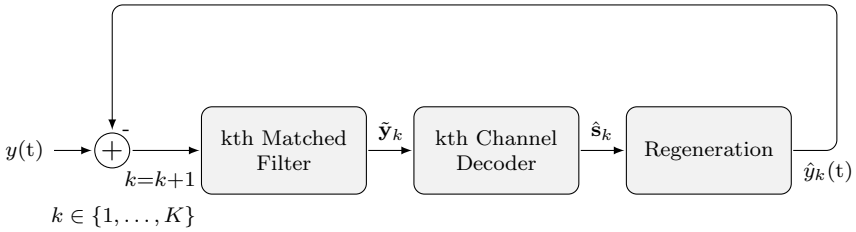

Figure 3.2: Block diagram of the successive decoding receiver with hard/soft cancellation.

Another well-known scheme is that of a SIC receiver operating under error control, in which case the receiver takes advantage of error control mechanisms already encapsulated with data, such as cyclic redundancy check (CRC), and it only reconstructs and cancels the decoded packet when the received CRC matches that calculated from the decoded data. This is the case of the cancellation system incorporated in the already mentioned E-SSA demodulator [40], which will be analysed throughout this work. Figure 3.3 depicts the block diagram corresponding to this scheme. Note that, due to CRC checkout, cancellation is only produced in case that the CRC test is passed. Naturally, since the CRC provides very high reliability, the reconstruction of the signal $y_k(\mathrm{t})$ received from user k uses the hard outputs of the channel decoder after correct CRC. Remarkably, the system operates as a hard SIC receiver aided by packet error detection.

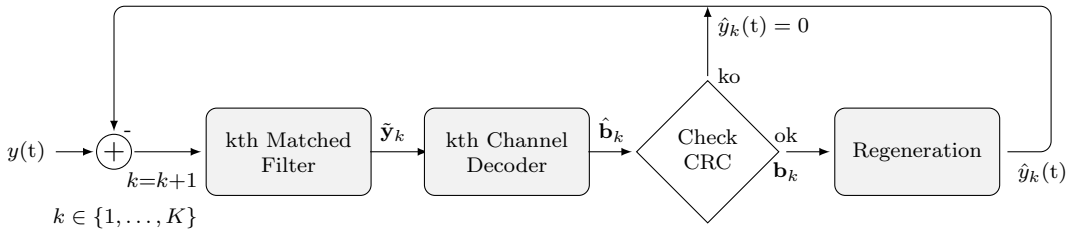

Figure 3.3: Block diagram of the successive decoding receiver with error control.

Several allocation designs for the above SIC architectures have been derived in the literature, however, limited to studying the interaction of only two of the four magnitudes: rate, energy, reliability and packet decoding latency. The literature in the context of this thesis has analysed more extensively the interplay between energy and other magnitudes. The pioneering work [48] derives, under asymptotically many DS spread spectrum users and a genie-aided SIC receiver that operates after the bank of matched filters, that the best transmission strategy is when

users transmit at the same rate and perform channel inversion so as to arrive with exponentially distributed energies. In view of the previous result, other analyses have derived the best energy allocation strategy for a practical setting in which users employ the same physical layer [40,47]. Finally, an iterative approach to obtain the power control law of all users at a target reliability is proposed [61]. Remarkably, cancellation receivers benefit from a packet-power unbalance with an exponential trend [63]. The rest of the interplays have been analysed mostly in other contexts. As an example, the authors in [64] analyse the interplay between power and blocklength under second-order coding rates and a rough approximation of full interference. Other analyses study non-orthogonal multiple access access under latency constraints [65].

The following study analyses the *exact* behaviour of a SIC received aided by error control, and serves as a starting point for later sections.

3.3 The User-Finite Model

This thesis works on the rigorous analysis of a packet-oriented SIC policy aided by error control, to then derive the allocation strategies that maximise the performance of a massively populated network. The starting point is the baseband signal captured through the receiver's antenna

$$y(\mathrm{t}) = \sum_{k=1}^{K} \sqrt{\gamma[k]}\; e^{j\Omega_k} \sum_{m=1}^{n} \mathbf{s}_k[m] c_{k,m}(\mathrm{t} - mT - \tau_k) + w(\mathrm{t}). \tag{3.7}$$

$\gamma[k] \triangleq \gamma_{\mathrm{x}}[k]h[k]$ is the symbol energy received from user k. Carrier frequency offsets are obviated in this chapter and in the rest of the thesis. The first operation carried out by the receiver is to detect the K users, to proceed afterwards with successive decoding following a pre-established decode ordering. The natural order for such decoding must be set as a function of the pair $(\gamma[k], R[k])$ of all users $1 \leq k \leq K$. Since both variables are part of the optimisation problem, it is not evident to see which combination of symbol energies and rates constitute the best user-decoding order to maximise SE for arbitrary distributions $\gamma[1], \ldots, \gamma[K]$ and $R[1], \ldots, R[K]$ [66]. One way to circumvent the ordering problem is to rank users in non-increasing order of signal-to-interference-plus-noise ratio (SINR) estimates, which coincides, under accurate estimations[1], with the order established by the distribution of received symbol energies, as

$$\gamma[1] \geq \cdots \geq \gamma[k] \geq \cdots \geq \gamma[K]. \tag{3.8}$$

The case when estimates are not accurate is addressed in Chapter 5. Since the above ordering is part of the optimisation process addressed in later sections, this chapter considers an equivalent ordering from channel power gains

$$h[1] \geq \cdots \geq h[k] \geq \cdots \geq h[K], \tag{3.9}$$

a fact that is validated via simulations after the allocation design.

In this respect, this chapter investigates the problem of energy and rate allocation for a large

[1]Under accurate SINR estimations $\Lambda[1 \leq k \leq K]$ at the initial stage, it is easy to obtain $\Lambda[k] - \Lambda[k+1] = \rho(\gamma[k] - \gamma[k+1])$ with $\rho > 0$. Consequently, the order of symbol energies preserves the order of SINRs.

number of users when all of them are detected, and the receiver proceeds on the basis of the ordering (3.9). The SIC receiver performs K stages and operates as follows to process user k:

1. **Matched filtering.** The single-user detector operates with the aggregate input signal using the matched filter (despreader) for user k. The output of the symbol matched filter

$$\tilde{\mathbf{y}}_k = \sqrt{\gamma[k]}\; e^{j\Omega_k}\; \mathbf{s}_k + \sqrt{N_t[k]}\; \mathbf{w}_k \tag{3.10}$$

   comprises the useful signal from user k contaminated by the additive Gaussian term $\mathbf{w}_k$. The term $\mathbf{w}_k$ contains Gaussian uncorrelated samples. Recall that $\mathbb{E}[\mathbf{s}_k^H \mathbf{s}_k] = 1$ and $\mathbb{E}[\mathbf{w}_k^H \mathbf{w}_k] = 1$. The last term in (3.10) contains the thermal noise plus the cross-correlation between the signature of user k and the waveforms received from interfering users. The magnitude (in power) of noise plus interference affecting user k is the factor $N_t[k]$.

2. **Decoding and cancellation.** The channel decoder guesses the bit stream transmitted by user k by making use of the decoding function of the i-th decoding system as $\hat{\mathbf{b}}_k = f_i^{-1}(\tilde{\mathbf{y}}_k)$. The CRC is then computed and compared with the received CRC to determine whether the decoded bit stream is correct or not. The used error control code is considered sufficiently powerful so as to assume that the probability of getting a bit stream with false alarm CRC is sufficiently low. If the CRC checks out, the transmitted packet is regenerated as $\mathbf{s}_k = f_i(\mathbf{b}_k)$, and its complex amplitude $\hat{A}[k]$ and timing $\hat{\tau}_k$ are improved using the n symbols of $\mathbf{s}_k$. The signal received from user k is regenerated as

$$\hat{y}_k(\mathrm{t}) = \hat{A}[k] \sum_{m=1}^{n} \mathbf{s}_k[m] c_{k,m}(\mathrm{t} - mT - \hat{\tau}_k), \tag{3.11}$$

   and cancelled at the waveform level from the received signal $y(\mathrm{t})$ as $y(\mathrm{t}) - \hat{y}_k(\mathrm{t})$. Then, the subsequent user, $k \to k{+}1$, is processed until and including user $k = K$.

The adopted SIC policy corresponds to a simple version of the packet decoding strategy implemented in the E-SSA system [40]. We adopt one-iteration SIC rather than iterative SIC, and a sufficiently long sliding window that does not overlap or, failing that, without sliding window. More sophisticated receivers are analysed in posterior chapters.

Naturally, the above model does not allow a straightforward evaluation of the adopted SIC receiver. For this very reason, a simpler and mathematically more tractable model is derived next. Since long spreading codes make multiple access interference Gaussian[2] at the output of the matched filter even if users do not employ Gaussian codebooks, the expressions above are derived in terms of the SINR variable Γ, whereby interference contributes as an added uncorrelated Gaussian term[3]. Then, the SINR after despreading user k is

$$\Gamma[k] = \frac{\gamma[k]}{N_t[k]}. \tag{3.12}$$

The assumption of Gaussian interference after every matched filter leads to assess the magnitude

[2] Rapidly converging to a Gaussian distribution when the spreading factor N increases.

[3] Strictly speaking, the Gaussian assumption does not consider interfering symbols of the same user statistically independent. The reason is because codewords maintain the dependence introduced by the encoder employed.

of imperfection of the decoding and cancellation operations by adopting known functions of the variable SINR Γ. For channel decoding, we consider that the packet error rate (PER) versus SINR curve associated with every encoder $\mathcal{R}_{1\leq i\leq p}$ is known and denoted $\mathrm{PER}[\Gamma, R_i]$. Analogously, the residual energy (RE) fraction to model imperfect cancellation is the known curve $\varepsilon[\Gamma, R_i]$. The advantage of adopting such functions to model the above operations is that both can be obtained through independent simulations of the joint decoding-cancellation system.

The estimated PER of the i-th coding system $\mathcal{R}_i$ at SINR $\Gamma = \gamma/N_t$ is obtained as (3.13a): by computing the average number of unsuccessful decoding trials of a noisy vector $\mathbf{y} = \sqrt{\gamma}e^{j\Omega}f_i(\mathbf{b}) + \sqrt{N_t}\mathbf{w}$, constituted by: a random message $\mathbf{b}$ encoded as $f_i(\mathbf{b})$; a uniformly distributed random phase Ω; and Gaussian noise $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I})$. The RE versus SINR curve is computed by estimating amplitude of *only* the successfully decoded packets, and by evaluating (3.13b):

$$\mathrm{PER}[\Gamma, R_i] = \Pr\left[f_i^{-1}(\mathbf{y}) \neq \mathbf{b}\right], \tag{3.13a}$$

$$\varepsilon[\Gamma, R_i] = \frac{1}{\gamma}\,\mathbb{E}\left[\|(\sqrt{\gamma}e^{j\Omega} - \hat{A})f_i(\mathbf{b})\|_2^2\right]. \tag{3.13b}$$

With the former definitions, the denominator of the SINR $\Gamma[k] = \gamma[k]/N_t[k]$ (3.12) is the noise plus interference term affecting user k after despreading (or equivalently, at the channel decoder input), given by

$$N_t[k] = 1 + \frac{\theta}{N}\sum_{i=1}^{k-1}\epsilon[i]\gamma[i] + \frac{\theta}{N}\sum_{i=k+1}^{K}\gamma[i]. \tag{3.14}$$

In this expression, the second term corresponds to the interference of processed users and the third term to that of users not processed yet. $\frac{\theta}{N}$ is the effective decorrelation factor for asynchronous DS users ($\theta = 1$ for synchronous users and $\theta = 0.5$ for totally asynchronous users), and $\epsilon[i]$ is the binary random variable that models the decoding success of user i:

1. $\epsilon[i] = 1$ with probability $\mathrm{PER}[\Gamma[i], R[i]]$ when the CRC does not check out (packet error),
2. $\epsilon[i] = \varepsilon[\Gamma[i], R[i]]$ complementarily, with probability $\mathrm{PSR}[\Gamma[i], R[i]] \triangleq 1 - \mathrm{PER}[\Gamma[i], R[i]]$, when the CRC checks out (packet successful). PSR stands for packet success rate.

The analysis assumes that $\epsilon[1\leq k\leq K]$ are statistically independent, since the decoder operates practically independently with independent messages subject to an uncorrelated Gaussian term.

Under this model, each user can be decoded correctly or incorrectly and, since the system evolves sequentially, the probability tree reaches 2^k states at the end of stage k. Of course, as it has already been mentioned, if users are allocated to sufficiently low error probabilities, then the above model turns out to be deterministic in practice, although actually, the model is random. In this chapter, however, since the configuration for the error probabilities depends in its turn on the allocation design, the goal of this thesis, such a simplification is considered invalid and the chapter continues with the previous random expressions.

3.4 The Asymptotic Model

A simple way to deal with the randomness of the above model is by resorting to the user-asymptotic regime. As demonstrated below, this allows to replace the random variables by their expectations; thus, turning the (random) system model in Section 3.3 into a deterministic one in exchange for operating in a function space rather than in a vector space.

To that aim, firstly, the following user index is adopted

$$t \triangleq \lim_{K\to\infty} \frac{k}{K}, \tag{3.15}$$

which must not be confused with the time variable t (non-italic) used in Section 3.3. This definition considers a dense number of users in the (continuous) interval $0 \leq t \leq 1$, wherein the strongest user $k = 1$ corresponds to $t = 0$ and the weakest user $k = K$ to $t = 1$. In this case, the decoding order is given by the non-increasing channel function

$$h(t) = \lim_{K\to\infty} h[Kt] \tag{3.16}$$

which, moreover, if $h(t)$ is smooth in t, defines a continuously differentiable function also called *profile*. The latter will cause the allocation functions to also be smooth.

Secondly, the above is used to determine the behaviour of the SIC system in this regime. We multiply and divide the last two addends of (3.14) by K and replace the sums by integrals under the differential $\mathrm{d}t \triangleq \lim_{K\to\infty} K^{-1}$. Therefore, the user-asymptotic SINR profile is

$$\Gamma(t) = \frac{\gamma(t)}{N_t(t)} \quad \text{in} \quad 0 \leq t \leq 1, \tag{3.17}$$

and as SIC progresses, $N_t(t)$ evolves according to the continuous expression in the variable t

$$N_t(t) = 1 + \alpha\theta \int_0^t q[\Gamma(\tau), R(\tau)]\gamma(\tau)\mathrm{d}\tau + \alpha\theta \int_t^1 \gamma(\tau)\mathrm{d}\tau, \tag{3.18}$$

where $q[\Gamma, R] \triangleq 1 - (1 - \varepsilon[\Gamma, R])\mathrm{PSR}[\Gamma, R]$ is, in fact, the expectation of ϵ [4]. It is still possible to come up with a simpler expression for $N_t(t)$. It can be characterised, using the decoding-cancellation characteristic $\Phi[\Gamma, R] \triangleq \theta(1-\varepsilon[\Gamma, R])\Gamma\cdot\mathrm{PSR}[\Gamma, R]$, either by the differential equation (3.19) or by its equivalent integral form (3.20)

$$\frac{\dot{N}_t(t)}{N_t(t)} = -\alpha\Phi\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right], \tag{3.19}$$

$$N_t(t) = N_t(0)\exp\left(-\alpha\int_0^t \Phi\left[\frac{\gamma(\tau)}{N_t(\tau)}, R(\tau)\right]\mathrm{d}\tau\right), \tag{3.20}$$

with the initial value

$$N_t(0) = 1 + \alpha\theta\int_0^1 \gamma(\tau)\mathrm{d}\tau. \tag{3.21}$$

[4] When $\gamma(t)$ and $R(t)$ are smooth, every interval Δ sufficiently small of users contains many independent and (practically) identically distributed random variables. This justifies substituting them by their expectations.

3.4.1 Advantage of the User-Asymptotic Model

One of the main advantages of analysing the SIC system with asymptotically many users is that its random behaviour for finite K users becomes deterministic. For a finite number of users, the more users considered, the more valid the result. The second advantage is that many equivalent expressions for the mitigation of interference described by the noise plus interference term can be obtained. Specifically, the following equivalent differential/integral expressions in the continuous variable t and $\gamma(t) = \gamma_{\mathrm{x}}(t)h(t)$ are derived

$$1) \quad N_{t,1}(t) = 1 + \alpha\theta \int_0^t q\left[\frac{\gamma(\tau)}{N_{t,1}(\tau)}, R(\tau)\right]\gamma(\tau)\mathrm{d}\tau + \alpha\theta\int_t^1 \gamma(\tau)\mathrm{d}\tau, \tag{3.22a}$$

$$2) \quad \frac{\dot{N}_{t,2}(t)}{N_{t,2}(t)} = -\alpha\Phi\left[\frac{\gamma(t)}{N_{t,2}(t)}, R(t)\right] \quad \text{and} \quad N_{t,2}(0) = N_t(0), \tag{3.22b}$$

$$3) \quad N_{t,3}(t) = N_t(0)\exp\left(-\alpha\int_0^t \Phi\left[\frac{\gamma(\tau)}{N_{t,3}(\tau)}, R(\tau)\right]\mathrm{d}\tau\right). \tag{3.22c}$$

Subscripts are added only in this subsection to distinguish the above expressions. Clearly, in the user-asymptotic regime, the above expressions are equal $N_{t,1}(t) = N_{t,2}(t) = N_{t,3}(t)$. However, since the expressions are not given in explicit form, they need to be computed numerically by partitioning t into M intervals, e.g. $t_0, \ldots, t_{M-1}$ with $t_0 = 0$ and $|t_i - t_j| = \frac{1}{M}$, and then by deriving discrete versions for the above equations (differential equations as finite differences and integrals as Riemann sums). The term associated with the interval $t_{i\geq 1}$ is computed as:

$$1) \quad N_{t,1}(t_i) = 1 + \frac{\alpha\theta}{M}\sum_{k=0}^{i-1} q\left[\frac{\gamma(t_k)}{N_{t,1}(t_k)}, R(t_k)\right]\gamma(t_k) + \frac{\alpha\theta}{M}\sum_{k=i+1}^{M-1}\gamma(t_k), \tag{3.23a}$$

$$2) \quad N_{t,2}(t_i) = N_{t,2}(t_{i-1})\left(1 - \frac{\alpha}{M}\Phi\left[\frac{\gamma(t_{i-1})}{N_{t,2}(t_{i-1})}, R(t_{i-1})\right]\right), \tag{3.23b}$$

$$3) \quad N_{t,3}(t_i) = N_{t,3}(t_{i-1})\exp\left(-\frac{\alpha}{M}\Phi\left[\frac{\gamma(t_{i-1})}{N_{t,3}(t_{i-1})}, R(t_{i-1})\right]\right), \tag{3.23c}$$

with the initial term resulting in

$$N_{t,1}(t_0) = N_{t,2}(t_0) = N_{t,3}(t_0) = 1 + \frac{\alpha\theta}{M}\sum_{k=0}^{M-1}\gamma(t_k). \tag{3.24}$$

Indeed, the mismatch between the above three computations depends on the number of intervals M. The asymptotic model allows for choosing the most conservative version to compute the noise plus interference term. It is straightforward to show that 1) and 2) are in fact equivalent, and that both correspond to the first-order Taylor expansion of 3). Since $\Phi[\Gamma, R] > 0$, the latter constitutes a worst case for computing $N_t(t_i)$ and the chosen one for the numerical implementation. The only point that remains to be solved is whether random variables can be substituted by their expectations in the system model described in Section 3.3 rather than resorting to the user-asymptotic regime. What is certain and has been observed by the author of this thesis is that the asymptotic computation followed by a numerical implementation provides higher accuracy, especially so, if the system operates at moderate PERs, where the error propagation along

the SIC receiver stages becomes more relevant. In this respect, some examples that evidence the impact of M on system performance for K users are shown in Section 3.5.4.4.

3.5 Asymptotically Optimal Allocation Designs

This section addresses the energy and rate allocation design to maximise asymptotic spectral efficiency (ASE). We exploit the previous asymptotic system model and conduct the allocation design in the asymptotic large-user regime leveraging tools from the calculus of variations within a predefined function space. The objective is to study the triple interplay between the rate, energy and reliability magnitudes of different users. To finally derive the conjoint energy and rate allocation design with optimal reliability, some simplified cases are first addressed by fixing one of the above magnitudes. More specifically, the cases of fair reliability, fair transmitted power and fair encoding rate are analysed, respectively, throughout Sections 3.5.1–3.5.3. Finally, the long-awaited energy and rate allocation with optimal reliability is analysed in Section 3.5.4. In general and in anticipation of the solution, since allocation functions do not always show explicit forms, simulations are attached to each section for better understanding.

3.5.1 Energy and Rate Allocation with Fair Reliability

This section investigates the conjoint energy and rate allocation with fair constrained reliability, with the focus on highlighting the unbalance between energy and rate of the different users when they are subject to *equal* blocklength and individual reliability constraints. More concretely, the model adopted considers, generically, the blocklength represented under the variable n_e and users subject to the individual reliability (packet success probability) constraint $0 < \rho \leq 1$.

The first part of this section studies the case of uniform unitary channel power gains $h(t) = 1$ in $0 \leq t \leq 1$. In this case, the best-performing allocation with zero channel unbalance $\gamma_{\mathrm{x}}(t) = \gamma(t)$ can be also interpreted as the most favourable channel-independent distribution to SIC.

Recall that, in the infinite blocklength (IBL) regime $n_e \to \infty$ with a genie-aided SIC receiver that operates under perfect cancellation $\varepsilon = 0$, the best transmission strategy attains reliable communication $\rho = 1$ when users are allocated to the same transmission rate and they arrive at the receiver's front end with exponentially distributed energies as

$$\gamma(t) = \bar{\Gamma}(1 + \alpha\theta\bar{\gamma}_{\mathrm{x}}) \exp(-\alpha\theta\bar{\Gamma}t) \quad \text{in} \quad 0 \leq t \leq 1, \tag{3.25}$$

$$\bar{\Gamma} \triangleq \alpha^{-1} \ln(1 + \alpha\theta\bar{\gamma}_{\mathrm{x}}), \tag{3.26}$$

such that, users are decoded with equal SINRs $\Gamma(t) = \bar{\Gamma}$. The attained ASE is $\alpha \log(1 + \bar{\Gamma})$.

The first analysis that follows extends the previous study: firstly, from the transmitters' point of view, when $\mathcal{R}$ consists of infinitely many coding schemes with rates $R_i \in r(\Gamma)$ corresponding to second-order coding rates (in the large-n_e expansion) [53]

$$r(\Gamma) = \log(1 + \Gamma) - \sqrt{\frac{V(\Gamma)}{n_e}} \mathcal{Q}^{-1}(1-\rho); \tag{3.27}$$

and secondly, from the receiver's point of view, when each cancellation is subject to a non-

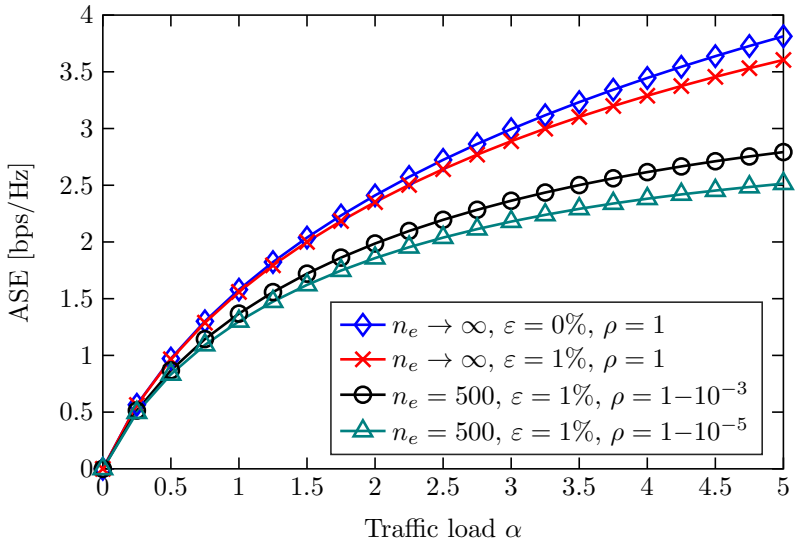

Figure 3.4: ASE versus traffic load α.

committal value $\varepsilon(\Gamma) = \varepsilon$. This case corresponds to $\Phi[\Gamma, R] \equiv \Phi[\Gamma] = \theta(1-\varepsilon)\rho\Gamma$. Admittedly, because n_e is not infinite, it is not possible to generate a continuous coding rate since, in practice, $R_i = k_i/n_e$ belongs to a discrete set. Nevertheless, we adopt a model in which we consider the previous possible ($R_i \in \mathbb{R}^+$) as an approximation to the practical case. The allocation problem is formulated in the space of discontinuous functions $\mathcal{C}[0, t_*]$ as:

$$\max_{0<t_*\leq 1} \max_{\gamma_{\mathrm{x}}(t)} \alpha \int_0^{t_*} r\left(\frac{\gamma_{\mathrm{x}}(t)}{N_t(t)}\right) \mathrm{d}t \quad \text{s.t.} \int_0^{t_*} \gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}} \ ; \ (3.19) \ ; \ (3.21). \tag{3.28}$$

By the calculus of variations, the above optimisation problems can be solved jointly. The interpretation of the obtained results reveals that the optimal SINR profile must again be uniform in the variable t, $\Gamma(t) = \bar{\Gamma}$ in $0 \leq t \leq t_*$. In this case, however, the value of $\bar{\Gamma}$ is in general far from the analytical value (3.26) found previously, which now reads

$$\bar{\Gamma} = \frac{1}{\alpha t_* \theta(1-\varepsilon)\rho} \ln\left(\frac{1+\alpha\theta\bar{\gamma}_{\mathrm{x}}}{1+\alpha\theta\bar{\gamma}_{\mathrm{x}}(1-(1-\varepsilon)\rho)}\right). \tag{3.29}$$

The received symbol energy profile is the discontinuous function

$$\gamma(t) = \begin{cases} \bar{\Gamma}(1+\alpha\theta\bar{\gamma}_{\mathrm{x}})\exp(-\alpha\theta(1-\varepsilon)\rho\bar{\Gamma}t) & \text{if} \quad 0 \leq t \leq t_* \\ 0 & \text{if} \quad t_* < t \leq 1 \end{cases}. \tag{3.30}$$

Moreover, the optimal $0 < t_* \leq 1$ guarantees that $\Gamma(t) \geq \Gamma_*$ in $0 \leq t \leq t_*$, with Γ_* the *unique* solution to $r(\Gamma_*) = r'(\Gamma_*)\Gamma_*$, for which the attained network performance is $\mathrm{ASE} = \alpha\rho t_* \cdot r(\bar{\Gamma})$.

The per-user reliability constraint ρ hardly penalises the SINR $\bar{\Gamma}$ but, instead, the coding rate $r(\bar{\Gamma})$ results severely degraded, specially so, at low blocklengths. A short simulation is undertaken to evaluate numerically the penalty loss in ASE due to the use of short-length packets. Figure 3.4 shows the ASE attained with the above uniform SINR profile sweeping several traffic loads. Comparative performance analyses are carried out with $\bar{\gamma}_{\mathrm{x}} = 8\mathrm{dB}$ and $\theta = 1$. Clearly, the finite blocklength (FBL) regime strongly downgrades ASE, specially so, under demanding reliability

constraints. The energy distribution that mostly favours SIC performance is shown to be uniform for $\alpha \to 0$, which coincides with the optimal distribution for the conventional matched filter in absence of SIC. The energy unbalance increases with the offered traffic, since the first users must compensate for higher interference whereas the last users must decrease their transmitted energy to satisfy the average energy constraint.

The second part of this chapter investigates the impact produced by the known channel $h(t)$ on the allocation designs. The goal is to find the best energy unbalance at transmission $\gamma_{\mathrm{x}}(t)$ subject to the average energy constraint over all users $\bar{\gamma}_{\mathrm{x}}$. ASE is the benchmark figure of merit. We have the following optimisation:

$$\max_{0<t_*\leq 1} \max_{\gamma_{\mathrm{x}}(t)} \alpha \int_0^{t_*} r\left(\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}\right) \mathrm{d}t \quad \text{s.t.} \int_0^{t_*} \gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}} \ ; \ (3.19) \ ; \ (3.21). \tag{3.31}$$

The solution to this problem considers discontinuous functions $\gamma_{\mathrm{x}}(t) \in \mathcal{C}[0, t_*]$, where $0 < t_* \leq 1$ is optimised as well. The stationary point equation that governs optimality, derived in point 1 of Appendix 3.A.1, is

$$\lambda = \frac{r'(\Gamma(t))}{N_t(t)\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}c - \alpha\theta(1-\varepsilon)\rho I_{\mathrm{x}}(t)} \quad \text{in} \quad 0 \leq t \leq t_* \tag{3.32}$$

with $c \triangleq 1 + \alpha\theta\bar{\gamma}(1-(1-\varepsilon)\rho)$, $\bar{\gamma} \triangleq \int_0^{t_*} \gamma_{\mathrm{x}}(\tau)h(\tau)\mathrm{d}\tau$ and $I_{\mathrm{x}}(t) \triangleq \int_t^{t_*} \gamma_{\mathrm{x}}(\tau)\mathrm{d}\tau$. In this case, the optimal user-index $0 < t_* \leq 1$ guarantees that $\Gamma(t) \geq \Gamma_*$ in $0 \leq t \leq t_*$, with Γ_* the unique solution to $r(\Gamma_*) = r'(\Gamma_*)\Gamma_*$. Specifically, if the user-admission is $t_* < 1$, then $\Gamma(t_*) = \Gamma_*$ and users $t_* < t \leq 1$ remain silent $\Gamma(t_* < t \leq 1) = 0$. Remarkably, although the optimal $\Gamma(t)$ is not given in explicit form, it is not uniform in t, in contrast to the case of uniform unitary channel gains $h(t) = 1$. The unbalance of the SINR profile is then a function of the unbalance of the channel gain profile $h(t)$. More specifically, the slope of the optimal $\Gamma(t)$ profile, $\dot{\Gamma}(t)$, is roughly close to $\nabla_t \ln h(t)$. Nonetheless, the exact results must be evaluated numerically. The following lines describe a numerical implementation to compute, exactly, the optimal $\Gamma(t)$ from (3.32).

Numerical Resolution: The proposed algorithm adopts a partition of t in M intervals $t_0, \ldots, t_{M-1}$ of the same length and $t_0 = 0$, and performs bisection searches over the parameters $\Gamma(t_0)$ and $\bar{\gamma}$. The algorithm performs M steps for each pair of them. At the first step $i = 0$, $\lambda > 0$ is computed by evaluating (3.32) at $t = t_0$ with $N_t(t_0) = 1+\alpha\theta\bar{\gamma}$ and $I_{\mathrm{x}}(t_0) = \bar{\gamma}_{\mathrm{x}}$. Next, the symbol energies are computed as $\gamma(t_0) = \Gamma(t_0)N_t(t_0)$ and $\gamma_{\mathrm{x}}(t_0) = \gamma(t_0)/h(t_0)$, and the cumulative integrals are updated as $I_{\mathrm{x}}(t_1) = I_{\mathrm{x}}(t_0) - \frac{1}{M}\gamma_{\mathrm{x}}(t_0)$ and $N_t(t_1) = N_t(t_0)\exp(-\frac{\alpha}{M}\Phi[\Gamma(t_0)])$. Now, at step $i \geq 1$: (i) $\Gamma(t_i)$ is computed as the SINR that satisfies (3.32) for the computed λ. If $\Gamma(t_i) < \Gamma_*$ the algorithm is stopped and $t_* = t_{i-1}$ is set; (ii) $\gamma(t_i) = \Gamma(t_i)N_t(t_i)$ and $\gamma_{\mathrm{x}}(t_i) = \gamma(t_i)/h(t_i)$ are computed; and (iii) $I_{\mathrm{x}}(t_{i+1}) = I_{\mathrm{x}}(t_i) - \frac{1}{M}\gamma_{\mathrm{x}}(t_i)$ and $N_t(t_{i+1}) = N_t(t_i)\exp(-\frac{\alpha}{M}\Phi[\Gamma(t_i)])$ are updated. After computing the last step $i = M-1$, $\bar{\gamma}$ assumed known and the computed profile $\gamma_{\mathrm{x}}(t_0), \ldots, \gamma_{\mathrm{x}}(t_{M-1})$ are verified to satisfy the initial constraints, computed as

$$\bar{\gamma}_{\mathrm{x}} = \frac{1}{M}\sum_{i=0}^{M-1} \gamma_{\mathrm{x}}(t_i) \quad \text{and} \quad \bar{\gamma} = \frac{1}{M}\sum_{i=0}^{M-1} \gamma_{\mathrm{x}}(t_i)h(t_i), \tag{3.33}$$

within an acceptable tolerance. If not, $\Gamma(t_0)$ and $\bar{\gamma}$ are updated, and the procedure is repeated.

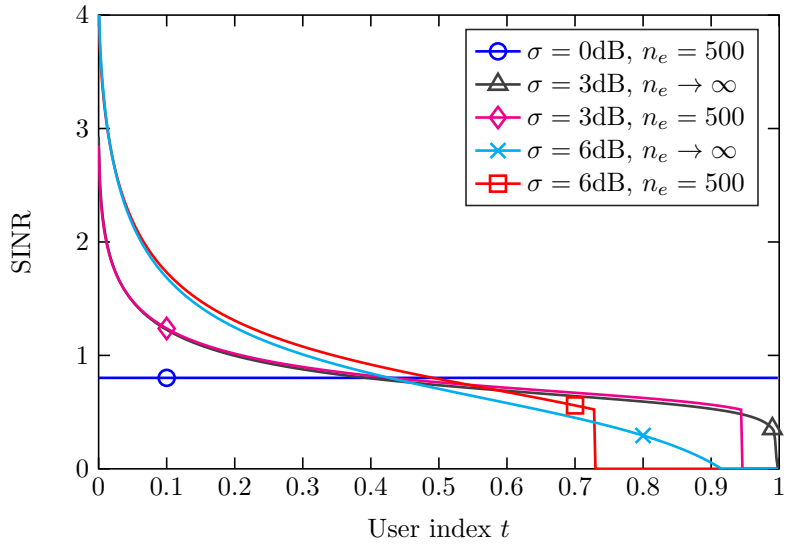

Figure 3.5: Optimal SINR profile $\Gamma(t)$, computed under $M = 1000$ points, at $\alpha = 4.0$.

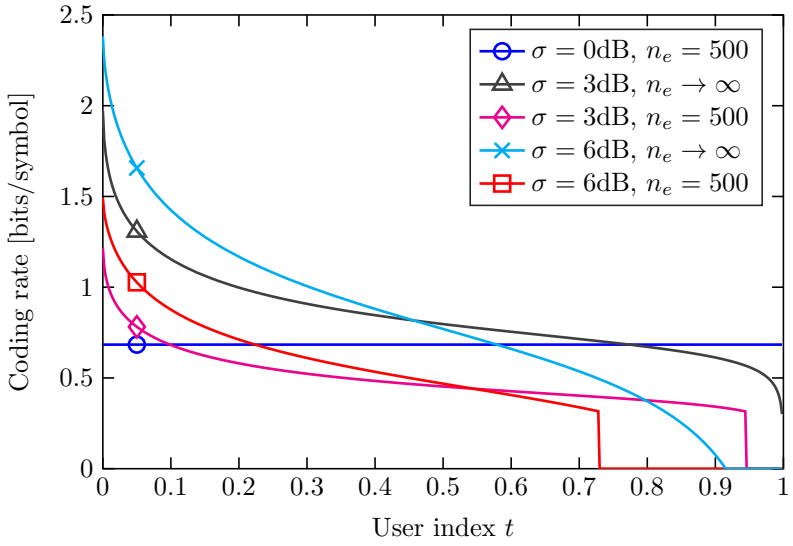

Figure 3.6: Optimal coding rate profile $R(t)$, computed under $M = 1000$ points, at $\alpha = 4.0$.

Figures 3.5 and 3.6 depict the optimal SINR and rate allocation profiles computed under user channel power gains log-normally distributed with unit mean and deviation σ. As shown, effectively, when the channel is not unbalanced ($\sigma = 0$dB), the optimal profiles are uniform in the variable t. In contrast, when the channel is unbalanced $\sigma > 0$dB, they are shown to be non-increasing in t. The latter is achieved when strongest users exploit high SINRs and transmission rates, and the weakest users remain silent rather than perform channel inversion. The optimal profiles in the IBL and FBL regimes attain practically the same SINRs except for their respective user-admission indices (see Figure 3.5). Rather, the respective coding rates depicted in Figure 3.6 substantially differ. With regard to network performance, the ASE achieved is higher when the channel profile $h(t)$ provides sufficient unbalance for the offered traffic. Contrarily, ASE is penalised when $h(t)$ is highly unbalanced.

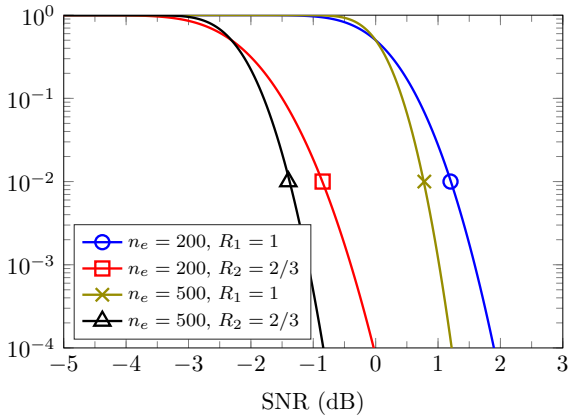

Figure 3.7: PER versus SNR curve.

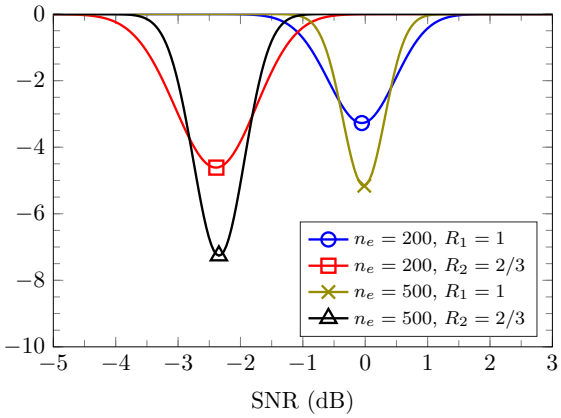

Figure 3.8: First derivative of the PER curve.

3.5.2 Rate Allocation with Fair Transmitted Power

This subsection investigates the best rate allocation strategy when all users transmit equal symbol energy $\bar{\gamma}(t) = \bar{\gamma}_{\mathrm{x}}$, in which case the received symbol energy distribution preserves the unbalance of the known (non-increasing) channel profile $h(t)$, as

$$\gamma(t) = \bar{\gamma}_{\mathrm{x}} h(t). \tag{3.34}$$

Recall that if users employ capacity-achieving coding schemes (IBL regime), then the optimal allocation shall be adjusted according to $R(t) = \log(1 + \gamma(t)/N_t(t))$. The interesting case, however, is when users employ coding schemes for short packets. The analysis conducted in this section considers infinitely many coding schemes achieving second-order coding rates in the long blocklength expansion. In this case, one option to optimise the coding rate of each user t, $R(t)$, is by letting it vary freely and by accounting for the effective transmission rate

$$R(t) \cdot \mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}, R(t)\right], \tag{3.35}$$

with the function $\mathrm{PSR}[\Gamma, R] \triangleq 1 - \mathrm{PER}[\Gamma, R]$ computed from the second-order normal approximation of the maximal channel coding rate (dropping the term $\frac{\log n}{n}$) as

$$\mathrm{PER}[\Gamma, R] = \mathcal{Q}\left((\log(1+\Gamma) - R)\sqrt{\frac{n_e}{V(\Gamma)}}\right), \tag{3.36}$$

and with the noise plus interference term computed as

$$\frac{\dot{N}_t(t)}{N_t(t)} = -\alpha\Phi\left[\frac{\bar{\gamma}_{\mathrm{x}}h(t)}{N_t(t)}, R(t)\right] \quad \text{and} \quad N_t(0) = 1 + \alpha\theta\bar{\gamma}_{\mathrm{x}}\int_0^1 h(t)\mathrm{d}t. \tag{3.37}$$

Note that in (3.35), sufficiently high coding rates will produce low PSR values. For each coding scheme, the known PER versus SINR curve must be smooth and non-increasing in Γ (actually, a reasonable and verifiable assumption for the PER versus SINR curves of typical coding schemes). It is easy to check that $\mathrm{PER}[\Gamma, R]$ in (3.36) is smooth, non-increasing in Γ, and increasing in R.

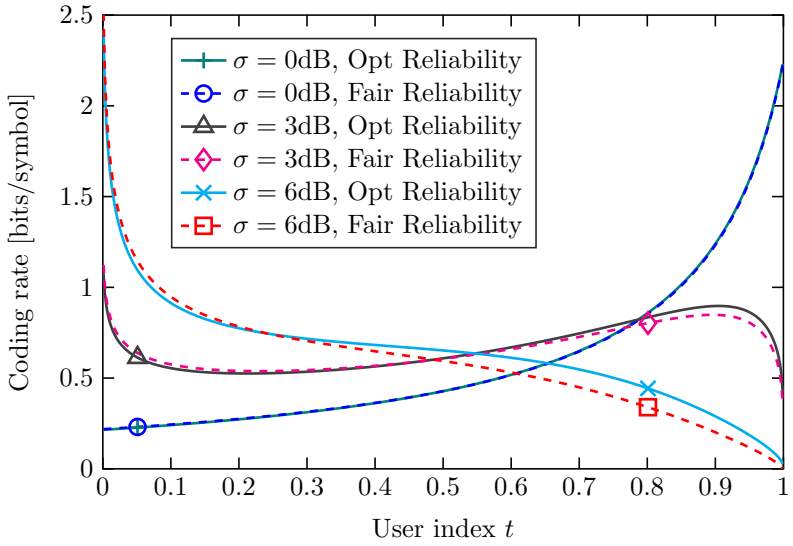

Figure 3.9: Optimal coding rate profile $R(t)$, computed under $M = 1000$ points, at $\alpha = 4.0$.

Two examples are shown in Figure 3.7. The first derivatives are drawn in Figure 3.8.

The allocation problem is to determine the continuously differentiable $R(t) \in \mathcal{C}[0,1]$ that maximises ASE

$$\max_{R(t)} \ \alpha \int_0^1 R(t) \cdot \mathrm{PSR}\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right] \mathrm{d}t \quad \text{s.t. (3.37).} \tag{3.38}$$

As discussed in point 2 of Appendix 3.A.1, the optimal $R(t)$ profile may be found by solving

$$\beta(t)\Phi_R\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right] = \mathrm{PSR}\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right] + R(t)\mathrm{PSR}_R\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right] \tag{3.39}$$

in $0 \le t \le 1$ jointly with the following differential equation and the initial value $\beta(1) = 0$:

$$\dot{\beta}(t) = \alpha R(t)\mathrm{PSR}_\Gamma\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right]\frac{\gamma(t)}{N_t(t)} - \alpha\beta(t)\Phi_\Gamma\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right]\frac{\gamma(t)}{N_t(t)}. \tag{3.40}$$

Numerical Resolution: The procedure to compute the optimal $R(t)$ from (3.39)–(3.40) considers a partition of $0 \le t \le 1$ into M intervals $t_0, \ldots, t_{M-1}$ with $t_0 = 0$. A temporarily known $\beta(t_0) < 0$ is considered. Since $\gamma(t)$ is known, at step i, $R(t_i)$ is computed from (3.39) and $\dot{\beta}(t_i)$ from (3.40), then $N_t(t_i)$ and $\beta(t_i)$ are updated respectively, under finite differences, as $N_t(t_{i+1}) = N_t(t_i)\exp(-\frac{\alpha}{M}\Phi[\gamma(t_i)/N_t(t_i), R(t_i)])$ and $\beta(t_{i+1}) = \beta(t_i) + \frac{1}{M}\dot{\beta}(t_i)$. The initial value for $\beta(t_0) < 0$ is obtained by bisection search until satisfying $\beta(t_{M-1}) = 0$ within an acceptable tolerance.

Figure 3.9 illustrates some examples of the optimal rate allocation profiles when channel power gains of all users are lognormally distributed with unit mean and deviation σ, and users transmit packets of $n_e = 500$ symbols. The rate allocation with optimal reliability derived in this subsection is compared with that under fair reliability for the same average reliability. The optimal $R(t)$ profiles result very close to those generated under fair reliability allocation except when the traffic load is high and the channel is sufficiently unbalanced. In the case evaluated

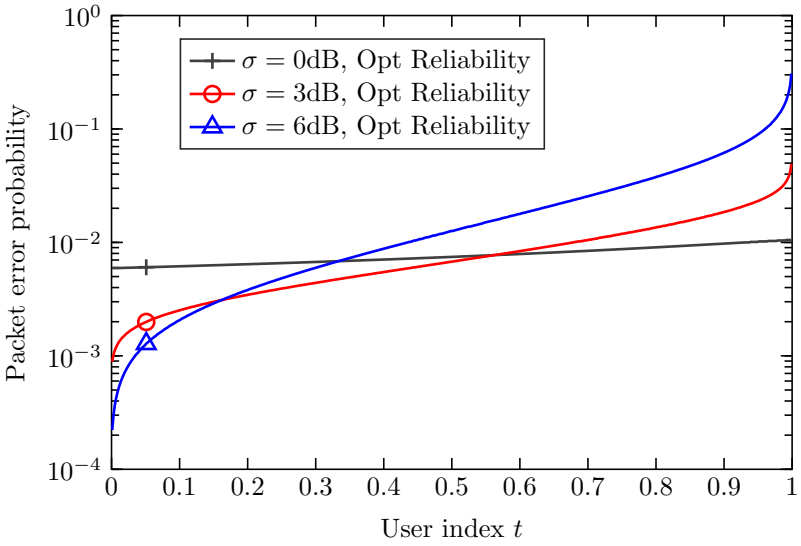

Figure 3.10: Optimal user-PER profile $\mathrm{PER}[\Gamma(t), R(t)]$, computed under $M = 1000$ points, at $\alpha = 4.0$.

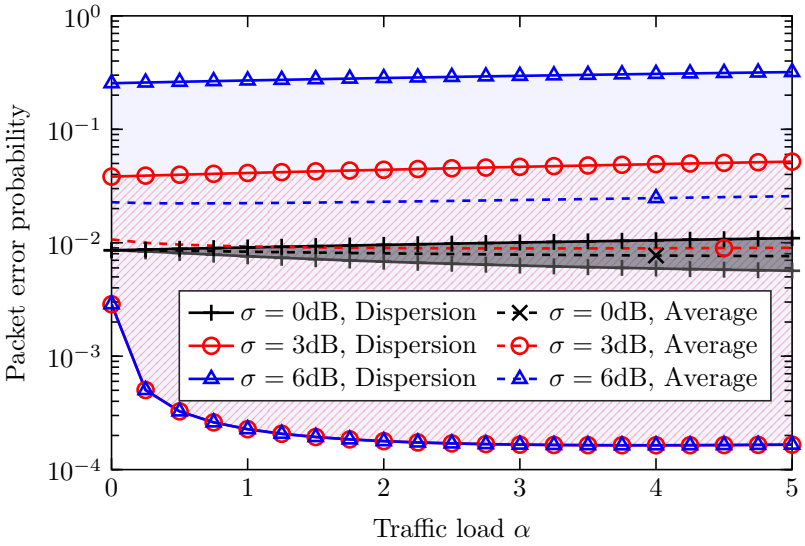

Figure 3.11: Dispersion of the optimal user-PER profile (minimum and maximum PER values), and the average PER over all users.

in this section, in which users do not enable power control, the coding rates are adjusted to generate, together with the SINR of each user, a non-decreasing user-PER profile as shown in Figure 3.10. The objective is to minimise the error[5] propagation along the stages of SIC, which is achieved by allocating more reliability to the strongest users and less reliability to the weakest users. Remarkably, the optimal operation of SIC is found when users are allocated to non-decreasing PERs roughly in the range $[10^{-4}, 1]$. The dispersion of the optimal user-PER profile is depicted in Figure 3.11, which shows higher dispersion as the received symbol energy distribution becomes more unbalanced.

[5]interference from users cancelled imperfectly plus that from users decoded unsuccessfully.

3.5.3 Energy Allocation with Fair Encoding Rate

This section investigates the energy allocation problem when $\mathcal{R}$ consists of a single coding scheme ($p = 1$). For brevity of notation, the coding rate is denoted R, and a single-argument PER curve $\mathrm{PER}[\Gamma]$ is adopted to characterise the error performance of the former coding scheme. This system configuration facilitates the computation of the noise plus interference profile, which can be simplified using the univariate function $\Phi[\Gamma] \triangleq \theta(1-\varepsilon[\Gamma])\Gamma\cdot\mathrm{PSR}[\Gamma]$ as

$$\frac{\dot{N}_t(t)}{N_t(t)} = -\alpha\Phi\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}\right] \quad \text{and} \quad N_t(0) = 1 + \alpha\theta\int_0^1 \gamma_{\mathrm{x}}(t)h(t)\mathrm{d}t. \tag{3.41}$$

When users employ a capacity-achieving coding scheme and the SIC cancellation system is characterised either by a slight slope or by a SINR-independent cancellation error magnitude $\varepsilon[\Gamma] \approx \varepsilon$, communication is reliable for users $0 \leq t \leq t_*$ as long as they are allocated above the SINR $\Gamma_{\mathrm{th}} \triangleq 2^R - 1$ (if R is given in bits per channel use). In this case, the received symbol energy profile $\gamma(t)$ is such that it generates the uniform SINR profile $\Gamma(t) = \Gamma_{\mathrm{th}}$ in $0 \leq t \leq t_*$, and it is cast as (3.42a) with the initial noise plus interference term (3.42b):

$$\gamma(t) = \Gamma_{\mathrm{th}} N_t(0)\exp(-\alpha\Phi[\Gamma_{\mathrm{th}}]t) \quad \text{in} \quad 0 \leq t \leq t_*, \tag{3.42a}$$

$$N_t(0) = \frac{1-\varepsilon}{\exp(-\alpha\Phi[\Gamma_{\mathrm{th}}]t_*) - \varepsilon}. \tag{3.42b}$$

The optimal t_* is computed as

$$t_* \triangleq \max\left\{ t_0 \in (0,1] \quad , \quad \int_0^{t_0} \frac{\gamma(t)}{h(t)}\mathrm{d}t \leq \bar{\gamma}_{\mathrm{x}} \right\}, \tag{3.43}$$

and the attained network performance is $\mathrm{ASE} = \alpha t_* R$. If active users operate with a finite-length coded modulation scheme and they are additionally constrained by fair reliability $0 < \rho \leq 1$, the previous results are still applicable by just substituting Γ_{th} by the SINR that achieves the target reliability $\mathrm{PSR}[\Gamma_{\mathrm{target}}] = \rho$. The attained network performance is $\mathrm{ASE} = \alpha t_* \rho R$.

The case when users employ a short-length coding scheme and with no reliability constraints is studied next. This case was analysed in our conference paper [67]. We seek for the smooth transmitted symbol energy profile $\gamma_{\mathrm{x}}(t) \in \mathcal{C}[0, t_*]$ (with a possible discontinuity at $t = t_* \leq 1$) that maximises ASE subject to an average energy constraint, as follows:

$$\max_{0<t_*\leq 1} \max_{\gamma_{\mathrm{x}}(t)} \alpha R \int_0^{t_*} \mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}\right]\mathrm{d}t \quad \text{s.t.} \int_0^{t_*} \gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}} \ ; \ (3.41). \tag{3.44}$$

The adopted figure of merit accounts for the average number of successful packet transmissions over all users (network reliability). The stationary point equation (point 3 of Appendix 3.A.1) is cast in terms of such a $\gamma_{\mathrm{x}}(t)$ that satisfies

$$\lambda = \frac{\mathrm{PSR}'[\Gamma(t)]}{N_t(t)\left(\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}\right) - \alpha(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}} I(t))\Phi'[\Gamma(t)]} \quad \text{in} \quad 0 \leq t \leq t_*, \tag{3.45a}$$

$$I_{\mathrm{x}}(t) \triangleq \int_t^{t_*} \gamma_{\mathrm{x}}(\tau)\mathrm{d}\tau \quad , \quad I(t) \triangleq \int_t^{t_*} \gamma_{\mathrm{x}}(\tau)h(\tau)\mathrm{d}\tau. \tag{3.45b}$$

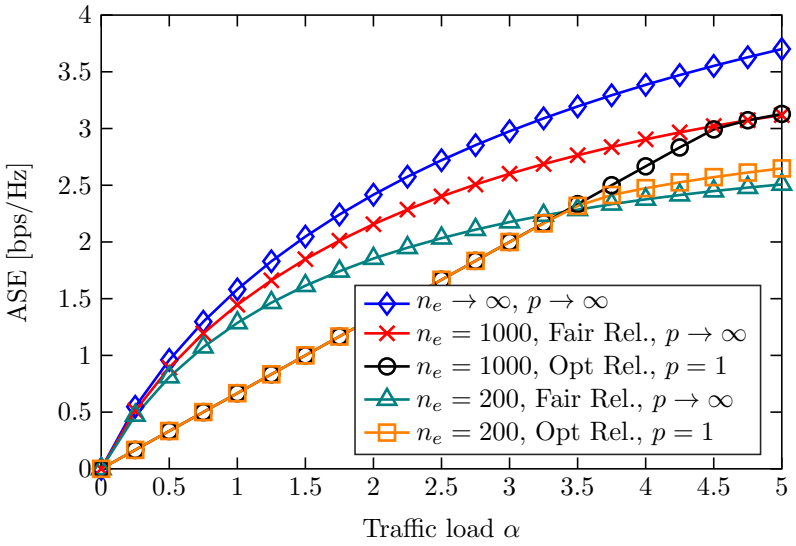

Figure 3.12: ASE versus traffic load α. Simulations compare: multiple encoders ($p \to \infty$) and a single encoder ($p = 1$) of rate 2/3; and fair reliability $\rho = 1-10^{-3}$ (Fair Rel.) and optimal reliability (Opt Rel.).

The optimal t_* is either $t_* = 1$ if $\Gamma(0 \leq t \leq 1) \geq \Gamma_*$, or such $0 < t_* \leq 1$ that it allocates $\Gamma(t_*) = \Gamma_*$, with Γ_* computed as the solution to $\mathrm{PSR}[\Gamma_*] = \Gamma_* \cdot \mathrm{PSR}'[\Gamma_*]$. In most of cases, Γ_* is a unique point located at the right of the inflexion point $\mathrm{PSR}''[\Gamma_{\mathrm{ip}}] = 0$. If Γ_* is not unique, all solutions must be taken into account during the optimization procedure. The concluding remark is that the optimal $\Gamma(t)$ is not uniform in t. Since the solution depends on the PSR derivative function, the expected result is that it becomes more unbalanced as $\mathrm{PSR}[\Gamma]$ is smoother. Nonetheless, its exact computation must be addressed numerically.

Numerical Resolution: The procedure carried out to compute the optimal $\Gamma(t)$ from (3.45a) considers the same partition for the user-variable t as in previous cases: M intervals of the same length $t_0, \ldots, t_{M-1}$ with $t_0 = 0$. M steps are followed: at step $i = 0$, $\lambda > 0$ is computed from (3.45a) evaluated at $t = t_0$ with $N_t(t_0) = 1 + \alpha\theta I(t_0)$, the symbol energies are computed as $\gamma(t_0) = \Gamma(t_0)N_t(t_0)$ and $\gamma_{\mathrm{x}}(t_0) = \gamma(t_0)/h(t_0)$, and cumulative integrals are updated as $I_{\mathrm{x}}(t_1) = I_{\mathrm{x}}(t_0) - \frac{1}{M}\gamma_{\mathrm{x}}(t_0)$, $I(t_1) = I(t_0) - \frac{1}{M}\gamma(t_0)$ and $N_t(t_1) = N_t(t_0)\exp(-\frac{\alpha}{M}\Phi[\Gamma(t_0)])$. Now, at step $i \geq 1$: (i) $\Gamma(t_i)$ is computed by solving (3.45a) for λ. If $\Gamma(t_i) < \Gamma_*$, the algorithm is stopped and t_* is set to $t_* = t_{i-1}$; (ii) $\gamma(t_i) = \Gamma(t_i)N_t(t_i)$ and $\gamma_{\mathrm{x}}(t_i) = \gamma(t_i)/h(t_i)$ are computed; and (iii) $I_{\mathrm{x}}(t_{i+1}) = I_{\mathrm{x}}(t_i) - \frac{1}{M}\gamma_{\mathrm{x}}(t_i)$, $I(t_{i+1}) = I(t_i) - \frac{1}{M}\gamma(t_i)$ and $N_t(t_{i+1}) = N_t(t_i)\exp(-\frac{\alpha}{M}\Phi[\Gamma(t_i)])$ are updated. Finally, bisection searches over $\Gamma(t_0)$ and $I(t_0)$ are performed.

Approximate Numerical Resolution: An approximate resolution, that holds to a sufficient accuracy, can be carried out when users operate at sufficiently high SINRs so as to assume $\Phi[\Gamma] = \theta(1-\varepsilon[\Gamma])\Gamma$ $\mathrm{PSR}[\Gamma] \approx \theta(1-\varepsilon[\Gamma])\Gamma$ in $\Gamma > \Gamma_*$, case in which its first derivative can be approximated by $\Phi'[\Gamma] \approx \theta(1-\varepsilon)$. In this situation, the stationary point equation (3.45a) can be simplified and the optimal SINR profile can be computed more efficiently.

As shown in Figure 3.12, when users transmit short packets the ASE achieved with a single encoder can surpass that with multiple encoders constrained by fair reliability. Besides, in practice, when the blocklength is sufficiently high the same ASE is achieved. The reason is because fair reliability allocates very different packet error probabilities to users relative to the

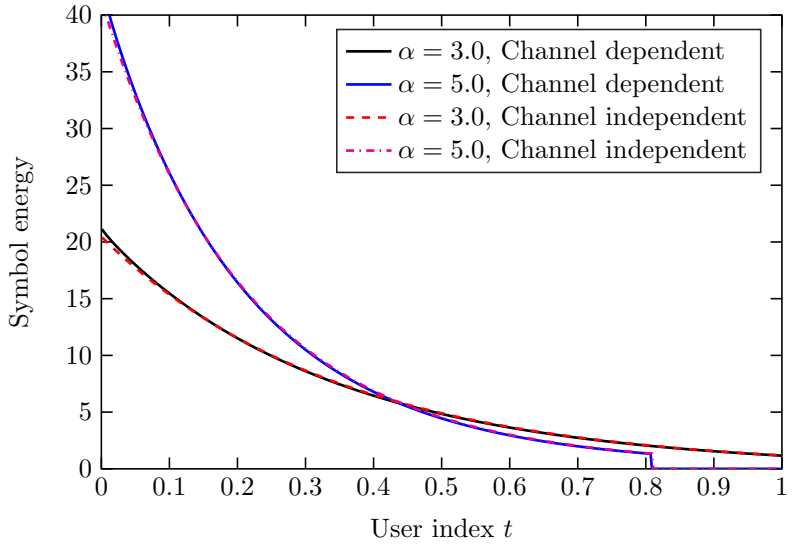

Figure 3.13: Optimal received symbol energy profile $\gamma(t)$ at the blocklength $n_e = 200$.

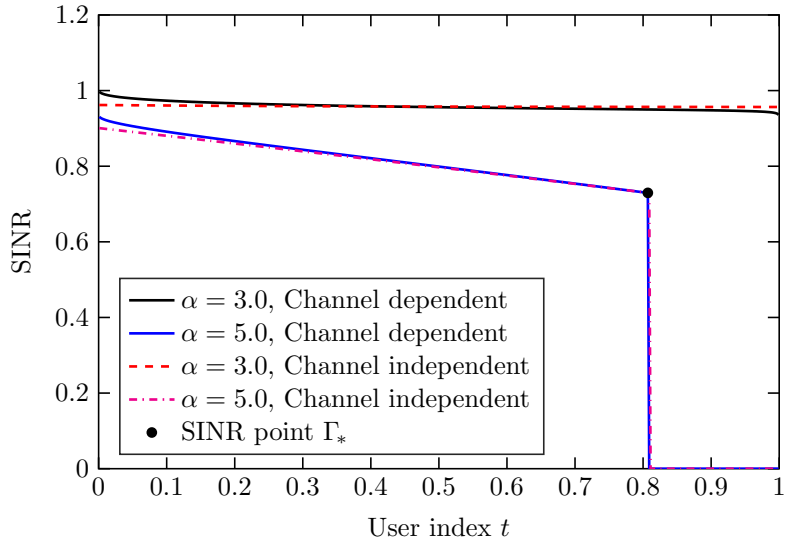

Figure 3.14: Optimal SINR profile $\Gamma(t)$ at the blocklength $n_e = 200$.

case of optimal reliability, which allocates non-decreasing error probabilities to all users. The optimal (channel dependent) received energy profiles $\gamma_{\mathrm{x}}(t)$ are shown in Figure 3.13 together with the profiles that mostly benefit SIC (channel independent), computed under $h(t) = 1$ and the same average received symbol energy. This simulation reveals that the best transmission strategy is that users perform channel inversion to generate a symbol energy profile $\gamma(t) = \gamma_{\mathrm{x}}(t)h(t)$ very close to the most favourable distribution at reception. Small mismatching is evidenced at the first and the last users. The more the offered traffic increases, the more unbalanced $\gamma(t)$ becomes; since the first users must deal with high interference levels. The respective SINR profiles are drawn in Figure 3.14. As shown, when the PER curve is not steep (as in the IBL regime), the SINR profile is not flat neither. Generally, $\Gamma(t)$ is non-increasing and its slope depends on the steepness of PER$[\Gamma]$. At low traffic loads $\Gamma(t)$ is practically uniform, $\Gamma(0)/\Gamma(t_0) \approx 1$,

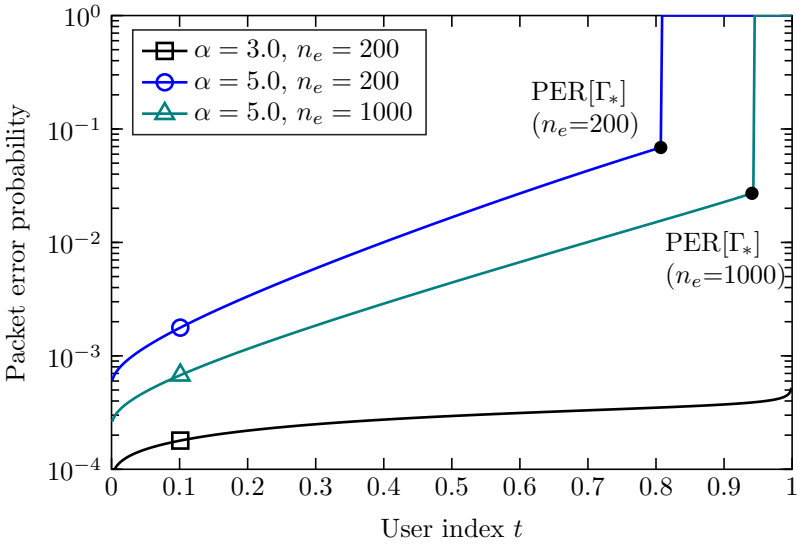

Figure 3.15: Optimal user-PER profile PER[$\Gamma(t)$].

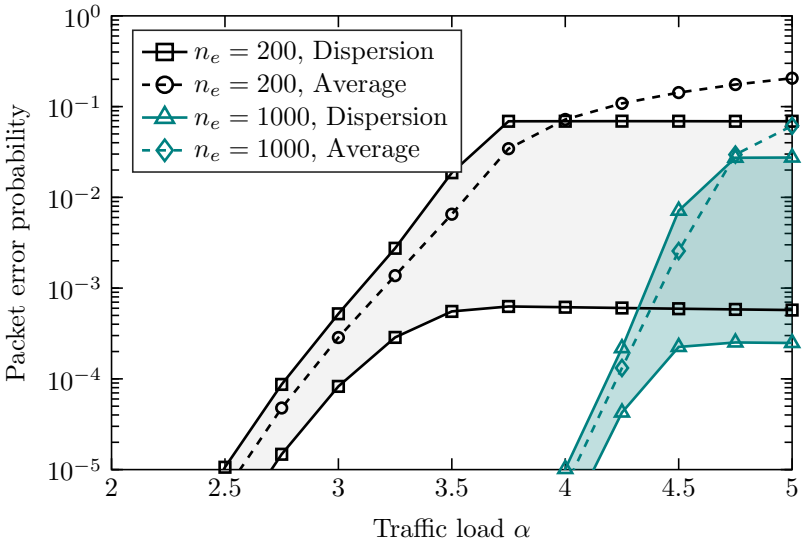

Figure 3.16: Dispersion of the optimal user-PER profile. This corresponds to the points of minimum PER (first user, PER[$\Gamma(0)$]) and maximum PER (user t_*, PER[$\Gamma(t_*)$]) excluding users remaining silent. The average PER over all users (active and non-active users) is drawn in dashed lines.

whereas at high loads it becomes substantially unbalanced, $\Gamma(0)/\Gamma(t_0) > 1$. A consequence of the above result is that the packet error probability of each user is a non-decreasing function of the decoding order, as shown in Figure 3.15. Another interesting analysis presented in Figure 3.16 is the dispersion of the optimal user-PER profile. At sufficiently low traffic loads, even if the optimal SINR profile is practically uniform, users experience a dispersion in PER values of less than one order of magnitude. The reason is because users are allocated to high SINRs, where the PER derivative function (evaluated at such SINRs) is close to zero, and then, the optimal user-PER profile is practically dispersion-less. Rather, as the load increases users are allocated to lower SINRs, where the PER derivative function produces more relevant results. The PER range for saturated traffic loads (i.e. $t_* < 1$) is about two orders of magnitude.

3.5.4 Energy and Rate Allocation with Optimal Reliability

This section addresses the much anticipated energy and rate allocation problem with optimal reliability. Recall that, as evidenced in the previous section, the allocation of equal reliability to all users downgrades network performance, specially so, when users transmit very short packets. This section conducts the allocation design by letting the energy, rate and reliability vary freely so as to maximise ASE now defined as

$$\text{ASE} = \alpha \int_0^1 R(t) \cdot \text{PSR}\left[\frac{\gamma_{\text{x}}(t)h(t)}{N_t(t)}, R(t)\right] \text{d}t. \tag{3.46}$$

For this case, we recover the initial expression for the noise plus interference profile with $\Phi[\Gamma, R] = \theta(1-\varepsilon[\Gamma, R])\Gamma\cdot\text{PSR}[\Gamma, R]$ and $\text{PSR}[\Gamma, R] = 1-\text{PER}[\Gamma, R]$, as

$$\frac{\dot{N}_t(t)}{N_t(t)} = -\alpha\Phi\left[\frac{\gamma_{\text{x}}(t)h(t)}{N_t(t)}, R(t)\right] \quad \text{with} \quad N_t(0) = 1 + \alpha\theta\int_0^1 \gamma_{\text{x}}(t)h(t)\text{d}t. \tag{3.47}$$

The remaining of this section is organised in four subsections. Section 3.5.4.1 aims at studying the theoretical limits of SIC under optimal reliability, which are used in Section 3.5.4.2 to conduct a more practical allocation design under a finite number of encoders. Section 3.5.4.3 compares the performance achieved by both systems. The last section (Section 3.5.4.4) analyses the practicality of the proposed allocation, and validates the accuracy of the theoretical results for a low-level SIC implementation inspired in the first iteration of the demodulator adopted in E-SSA [40], where, additionally, the single-encoder setting is extended to multiple encoders.

3.5.4.1 Allocation Designs with Asymptotically Many Encoders

This section considers, for all users, a pool $\mathcal{R}$ of infinite, $p \to \infty$, channel encoders achieving optimal second order coding rates with a per-codeword power constraint [53], case in which the PER characteristic of each coding scheme results the bivariate function of the SINR Γ and the rate R also adopted in (3.36). On the basis of previous analyses, we shall consider discontinuous allocations $\gamma_{\text{x}}(t), R(t) \in \mathcal{C}[0, t_*]$ as candidates to satisfy the optimization problem:

$$\max_{0<t_*\leq 1} \max_{\gamma_{\text{x}}(t), R(t)} \alpha\int_0^{t_*} R(t)\text{PSR}\left[\frac{\gamma_{\text{x}}(t)h(t)}{N_t(t)}, R(t)\right]\text{d}t \quad \text{s.t.} \int_0^{t_*}\gamma_{\text{x}}(t)\text{d}t = \bar{\gamma}_{\text{x}} \;;\; (3.47). \tag{3.48}$$

It should be emphasised that this analysis corresponds to a more general study than that of unboundedly large packets $n_e \to \infty$, since a generic blocklength n_e is considered. Recall that, in the IBL regime, we shall particularise $\text{PSR}[\Gamma, R] = 1$ and $R(t) = \log(1+\gamma_{\text{x}}(t)h(t)/N_t(t))$. In the case studied herein, the optimal $\gamma_{\text{x}}(t), R(t)$ profiles must satisfy the following stationary point equations in $0 \leq t \leq t_*$ (the proof is sketched in Appendix 3.A.1)

$$\lambda = \frac{\alpha R(t)\text{PSR}_\Gamma[\Gamma(t), R(t)]}{N_t(t)\left(\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\text{x}}\right) - \alpha(I_{\text{x}}(t) + \alpha\theta\bar{\gamma}_{\text{x}}I(t))\Phi_\Gamma[\Gamma(t), R(t)]}, \tag{3.49a}$$

$$\lambda = -\frac{\text{PSR}[\Gamma(t), R(t)] + R(t)\text{PSR}_R[\Gamma(t), R(t)]}{(I_{\text{x}}(t) + \alpha\theta\bar{\gamma}_{\text{x}}I(t))\Phi_R[\Gamma(t), R(t)]}, \tag{3.49b}$$

and the optimal $0 < t_* \leq 1$ guarantees that active users $0 \leq t \leq t_*$ are allocated above to a minimum SINR $\Gamma(t) \geq \Gamma_*$ and rate $R(t) \geq R_*$, with the pair (Γ_*, R_*) the unique solution to

$$\Gamma_* \cdot \mathrm{PSR}_\Gamma[\Gamma_*, R_*] = \mathrm{PSR}[\Gamma_*, R_*], \tag{3.50a}$$

$$-R_* \cdot \mathrm{PSR}_R[\Gamma_*, R_*] = \mathrm{PSR}[\Gamma_*, R_*]. \tag{3.50b}$$

As the blocklength tends to infinity, this implicit constraint vanishes, $(\Gamma_*, R_*) \to (0, 0)$. The minimum allocation values guarantee that fractional variations in rate and SINR do not incur in fractional PSR decrements of higher magnitude. The solution for the stationary points $\Gamma(t), R(t)$ needs to be obtained numerically following a similar procedure to the one presented in the previous section.

3.5.4.2 Allocation Designs with Finitely Many Encoders

Another interesting configuration is the case studied in this subsection, where users have available p coded modulation schemes denoted by the set $\mathcal{R} \triangleq \{\mathcal{R}_1, \ldots, \mathcal{R}_p\}$. The analysis herein pursues the determination of the best transmission strategy based on known coded modulation schemes, rather than evaluating theoretical limits as in the previous case. Then, we shall consider that users are decoded in groups of users that employ the same coding scheme, for which we assume a partition of $0 \leq t \leq 1$ into the following $p + 1$ intervals:

$$\{t_0^+, t_1^-\} \cup \{t_1^+, t_2^-\} \cup \cdots \cup \{t_{k-1}^+, t_k^-\} \cup \cdots \cup \{t_{p-1}^+, t_p^-\} \cup \{t_p^+, t_{p+1}^-\}. \tag{3.51}$$

Superscripts indicate the left and the right limits at every point t_k: t_k^- and t_k^+. Then, three optimization problems are posed to determine the best energy/code allocation strategy that, taking advantage of the unbalance introduced by the known distribution of channel power gains, maximises ASE. The study pursued herein corresponds to an extension of our conference paper [68]. The allocation problem is formulated in the function space of piecewise continuously differentiable functions $\gamma_{\mathrm{x}}(t), R(t) \in \mathcal{C}_p[t_0^+, t_p^-]$, as:

$$\max_{\mathcal{P}(\mathcal{R})} \max_{t_1,\ldots,t_p} \max_{\gamma_{\mathrm{x}}(t)} \alpha \sum_{k=1}^{p} \int_{t_{k-1}^+}^{t_k^-} R_k \mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}, R_k\right] \mathrm{d}t \tag{3.52a}$$

$$\text{s.t.} \quad \sum_{k=1}^{p} \int_{t_{k-1}^+}^{t_k^-} \gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}} \tag{3.52b}$$

$$\text{s.t.} \quad \frac{\dot{N}_t(t)}{N_t(t)} = -\alpha\Phi\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}, R(t)\right] \tag{3.52c}$$

$$\text{s.t.} \quad N_t(0) = 1 + \alpha\theta \int_0^1 \gamma_{\mathrm{x}}(t)h(t)\mathrm{d}t \tag{3.52d}$$

Outward from the expression inside: firstly, the optimal allocation $\gamma_{\mathrm{x}}(t)$ is determined by fixing the partition $t_1, \ldots, t_p$ (with a little abuse of notation, t_k denotes t_k^- and t_k^+) and $\mathcal{R}$; secondly, the optimal indices $t_1, \ldots, t_p$ are determined by fixing the other magnitudes; and thirdly, permutations over $\mathcal{R}$ are addressed. The solution derived in Appendix 3.A.1 is summarized below:

1. The stationary solution to the inner optimisation results into $k = 1, \ldots, p$ equations, each

of which is used to obtain the k-th piece of $\gamma_{\mathrm{x}}(t)$ in the respective interval, as

$$\lambda = \frac{\alpha R_k \mathrm{PSR}_\Gamma[\Gamma(t), R_k]}{N_t(t)\left(\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}\right) - \alpha(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}} I(t))\Phi_\Gamma[\Gamma(t), R_k]} \quad \text{in} \quad t_{k-1}^+ < t \leq t_k^-. \tag{3.53}$$

2. The optimisation of the user indices $t_1, \ldots, t_p$ conclude that the optimal allocation profile $\gamma_{\mathrm{x}}(t)$ also satisfies the following equation at the corners $t_1, \ldots, t_{p-1}$

$$\lambda = \alpha \cdot \frac{R_k \mathrm{PSR}[\Gamma(t_k^-), R_k] - R_{k+1}\mathrm{PSR}[\Gamma(t_k^+), R_{k+1}]}{F(t_k^-, R_k) - F(t_k^+, R_{k+1})} \tag{3.54}$$

with $F(t, R) = \gamma(t)(\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}) - \alpha(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}} I(t))\Phi[\Gamma(t), R]$, and the optimal index t_p^- ensures $\Gamma(t) \geq \Gamma_*$ with Γ_* computed as $\mathrm{PSR}[\Gamma_*, R_p] = \Gamma_* \cdot \mathrm{PSR}'[\Gamma_*, R_p]$. Then, it equals either $t_p^- = 1$ or such a value that satisfies $\Gamma(t_p^-) = \Gamma_*$.

In general, the optimisation problem is non-convex due to the $p!$ permutations over the set $\mathcal{R}$ and the shape of the error characteristics of each coding scheme. Nonetheless, if the PER curves corresponding to the different coding schemes $\mathrm{PER}[\Gamma, R_{1\leq i\leq p}]$ and the characteristics $\Phi[\Gamma, R_{1\leq i\leq p}]$ are not intertwined with each other in $\Gamma \geq \Gamma_*$, then the solution can be obtained in a computationally affordable way by means of the following algorithm.

Numerical resolution: The number of encoders employed by users and the fraction of users using each encoder shall be determined along with the allocation. We then propose an algorithm that dynamically constructs the set of encoders used by users and that achieves the optimal system performance. To that aim, the stationary point equations are solved together, with the interval $0 \leq t \leq 1$ discretised as $\tau_0, \ldots, \tau_{M-1}$ with $|\tau_i - \tau_{i+1}| = \frac{1}{M}$, and the following three variables temporarily known: $\mathcal{R}_1$, $\Gamma(\tau_0)$ and $I(\tau_0)$. The algorithm performs M steps. At step $i = 0$, $\lambda > 0$ is computed by evaluating (3.53) at $t = \tau_0$ as

$$\lambda = \frac{\alpha R_1 \mathrm{PSR}_\Gamma[\Gamma(\tau_0), R_1]}{N_t(\tau_0)\left(\frac{1}{h(\tau_0)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}\right) - \alpha(I_{\mathrm{x}}(\tau_0) + \alpha\theta\bar{\gamma}_{\mathrm{x}} I(\tau_0))\Phi_\Gamma[\Gamma(\tau_0), R_1]} \tag{3.55}$$

with $I_{\mathrm{x}}(\tau_0) = \bar{\gamma}_{\mathrm{x}}$ and $N_t(\tau_0) = 1 + \alpha\theta I(\tau_0)$, and $N_t(\tau_1) = N_t(\tau_0)\exp(-\frac{\alpha}{M}\Phi[\Gamma(\tau_0), R_1])$, $I(\tau_1) = I(\tau_0) - \frac{1}{M}\gamma_{\mathrm{x}}(\tau_0)h(\tau_0)$ and $I_{\mathrm{x}}(\tau_1) = I_{\mathrm{x}}(\tau_0) - \frac{1}{M}\gamma_{\mathrm{x}}(\tau_0)$ are updated. At step $i = 1$ and for each encoder $\mathcal{R}_{1\leq j\leq p}$, the SINR $\Gamma(\tau_1)|\mathcal{R}_j$ is computed from (3.53) with $N_t(\tau_1)$, $I_{\mathrm{x}}(\tau_1)$ and $I(\tau_1)$. Also, for each encoder $\mathcal{R}_{j\neq 1}$, $\lambda|\mathcal{R}_j$ (the conditional on $\mathcal{R}_j$ is due to notation) is computed from (3.54)

$$\lambda|\mathcal{R}_j = \alpha \cdot \frac{R_1 \mathrm{PSR}[\Gamma(\tau_0), R_1] - R_j \mathrm{PSR}[\Gamma(\tau_1)|\mathcal{R}_j, R_j]}{F(\tau_0, R_1) - F(\tau_1, R_j)}. \tag{3.56}$$

If some $\lambda|\mathcal{R}_j$ coincides with λ within a tolerance margin, $R(\tau_1) = R_j$ is allocated and $\Gamma(\tau_1)|\mathcal{R}_j \to \Gamma(\tau_1)$ is considered. Otherwise, $R(\tau_1) = R_1$ is allocated and $\Gamma(\tau_1)|\mathcal{R}_1 \to \Gamma(\tau_1)$ is considered. Next, $N_t(\tau_2) = N_t(\tau_1)\exp(-\frac{\alpha}{M}\Phi[\Gamma(\tau_1), R(\tau_1)])$, $I(\tau_2) = I(\tau_1) - \frac{1}{M}\gamma_{\mathrm{x}}(\tau_1)h(\tau_1)$ and $I_{\mathrm{x}}(\tau_2) = I_{\mathrm{x}}(\tau_1) - \frac{1}{M}\gamma_{\mathrm{x}}(\tau_1)$ are updated. At step $i = i+1$, the steps described above for $i = 1$ are followed, choosing as reference the encoder employed by the last user. Firstly, the SINRs $\Gamma(\tau_i)|\mathcal{R}_{1\leq j\leq p}$ are computed from (3.53) with the precomputed $N_t(\tau_i)$, $I_{\mathrm{x}}(\tau_i)$ and $I(\tau_i)$. Secondly, $\lambda|\mathcal{R}_j$ is computed from (3.54), and the corresponding $N_t(\tau_i)$, $I_{\mathrm{x}}(\tau_i)$ and $I(\tau_i)$ are updated.

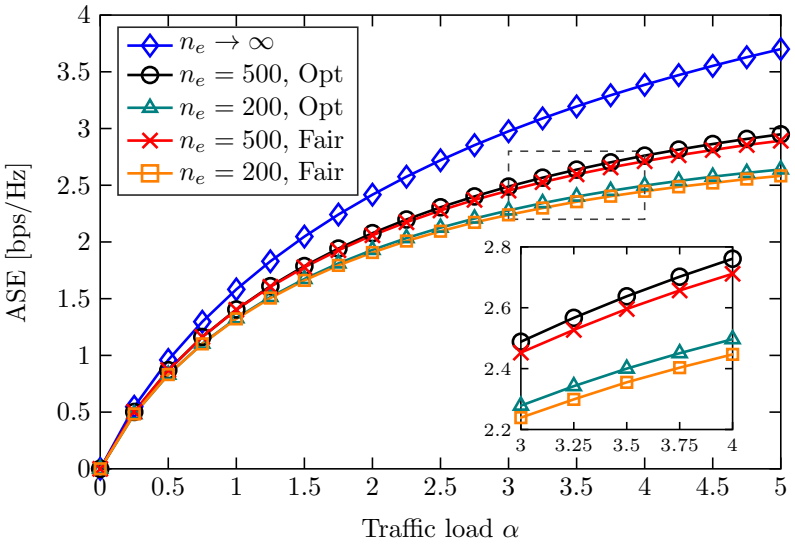

Figure 3.17: ASE versus offered traffic α with optimal and fair reliability $\rho = 1{-}10^{-2}$.

3.5.4.3 Performance Evaluation

The ASE is used as the key performance indicator. For simulation purposes, we consider users subject to a log-normal channel power unbalance of unit mean and deviation $\sigma = 3$dB. To simulate a more realistic case, we suppress the very high channel power gains from the log-normal theoretical model. More specifically, only the values h with $\Pr[H < h] \leq 1{-}10^{-3}$ are used, and the resulting distribution is then scaled to have unit mean. The cancellation system operates as $\varepsilon[\Gamma, R] = 1\%$, and the factor θ is set to 1. The energy constraint is $\bar{\gamma}_{\mathrm{x}} = 8$dB.

The attained ASE with finite blocklength constraints is drawn in Figure 3.17 and compared with the performance achieved when users are constrained by fair reliability. The first analysis considers for all users, infinitely many coding schemes optimal up to second order (in the large blocklength expansion). As shown, both systems achieve similar ASEs when users transmit long packets, which corresponds to coding systems with PER curves sufficiently steep (in Figure 3.17, this corresponds to $n_e = 500$). Rather, when the PER curves are less steep, which corresponds to shorter packet lengths, the difference between the two ASEs is greater and increases notably with the offered traffic. The need of an adequate allocation becomes more relevant as the load increases since the receiver operates in a more interference-limited regime. Besides, a common point between optimal and fair reliability is that they achieve ASEs far from the optimal ASE in the IBL regime. The main reason is the use of finite-length codes, since they require more SINR to achieve high reliability relative to the study under capacity-achieving schemes. In general, fair reliability penalises ASE since the optimal performance is attained unbalancing the reliabilities of users. Fair reliability allocates zero energy to more users and devotes the rest of energy to compensate for the non-unbalanced reliability. The latter is shown in Figure 3.18. Effectively, when users transmit short packets $n_e = 200$, the best transmission strategy under optimal reliability activates the transmission of approximately 10% more users. This analysis results more noticeable as the reliability constraint is more demanding. Instead, optimal and fair reliability admit practically the same number of users for increasing blocklengths.

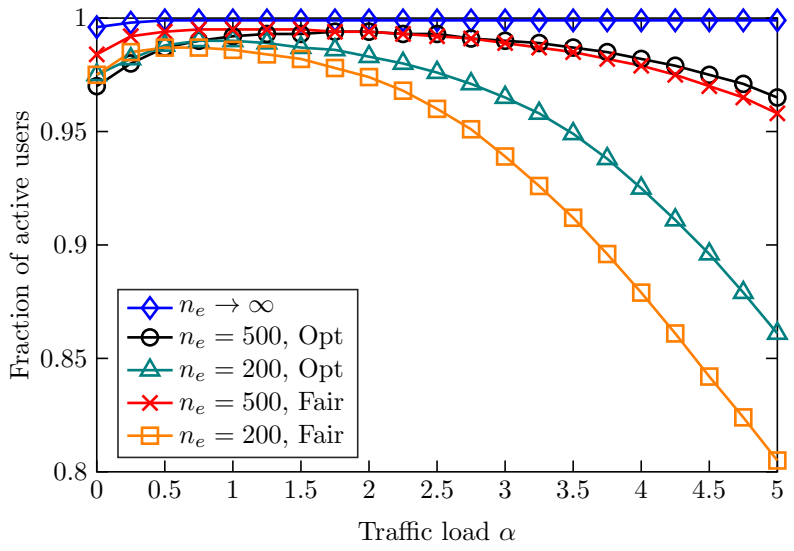

Figure 3.18: Fraction of active users $0 < t_* \leq 1$ with optimal and fair reliability $\rho = 1-10^{-2}$.

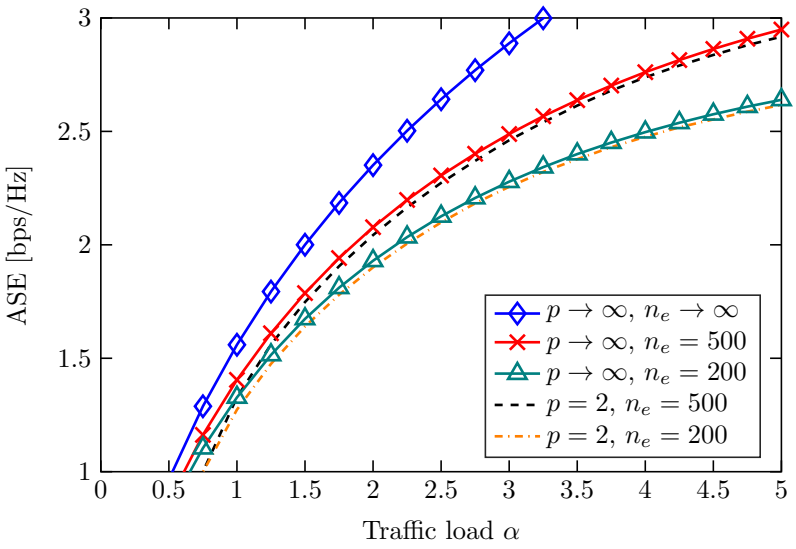

Figure 3.19: ASE versus traffic load α for infinitely ($p \to \infty$) and finitely many ($p = 2$) encoders. The case of $p = 2$ considers second-order coding schemes [53] with $R_1 = 4/3$ and $R_2 = 2/3$ bits/symbol.

The above analysis considers, as a theoretical limit, that users have available infinitely many coding schemes, $p \to \infty$. Figure 3.19 depicts the results from Section 3.5.4.2 for $p = 2$ coding schemes. As shown, regardless of blocklength and offered traffic, the proposed allocation practically attains the ASE achieved with asymptotically many encoders with *only* two of them with sufficiently far apart rates. The latter is interesting for practical scenarios since a finite and very small number of encoders allows for an overwhelming performance as long as users enable loose power control to compensate for the poor combination of coding rates. Figure 3.20 depicts the fraction of users using every encoder. On a practical level, the following code allocation policy is followed. At low traffic loads, users mostly exploit high-rate codes and, as the offered traffic increases, the use of such codes is increasingly restricted to only users with better channels, while

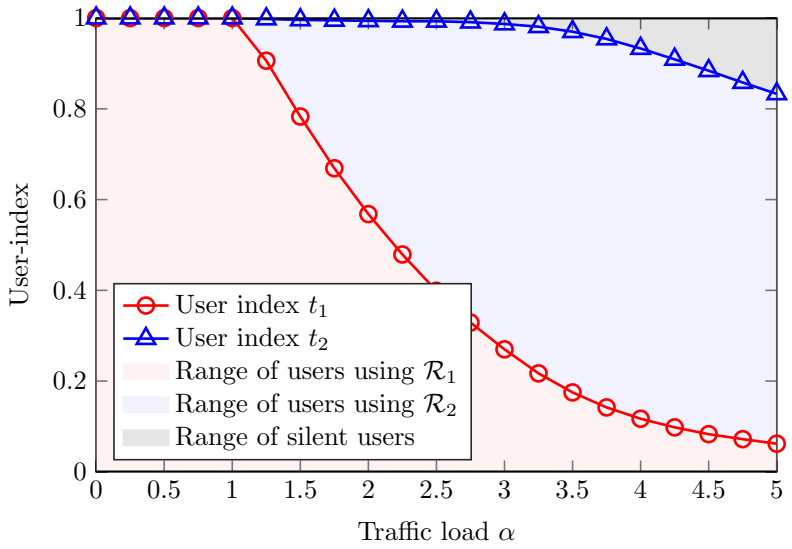

Figure 3.20: Range of users employing every encoder. Simulations correspond to $n_e = 500$ symbols.

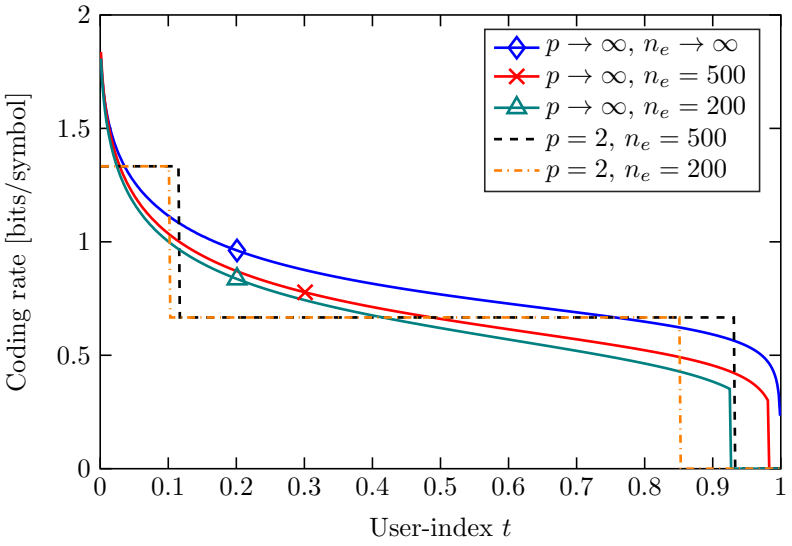

Figure 3.21: Coding rates (in bits/symbol) versus traffic load α for $p \to \infty$ and $p = 2$ encoders.

users with moderate channel gains use the more protective encoder, and users with worst channels remain silent. Examples of the optimal coding rate of each user are shown in Figure 3.21. The coding rate of each user is non-increasing in t. In the IBL regime with perfect cancellation $\varepsilon[\Gamma, R] = 0$, the decoding order does not penalise ASE as long as energy and rate are allocated adequately. When practical decoding and cancellation features are considered, however, users suffer due to the interference from previous decoding failures and imperfectly cancelled users. In this case, it is preferable that strong users transmit at high rates and that the rate loss is suffered more by the latter users, which must employ more protective encoders. Remarkably, the same ordering is evidenced for finitely many coding schemes, which indeed simplifies the optimisation in Section 3.5.4.2.

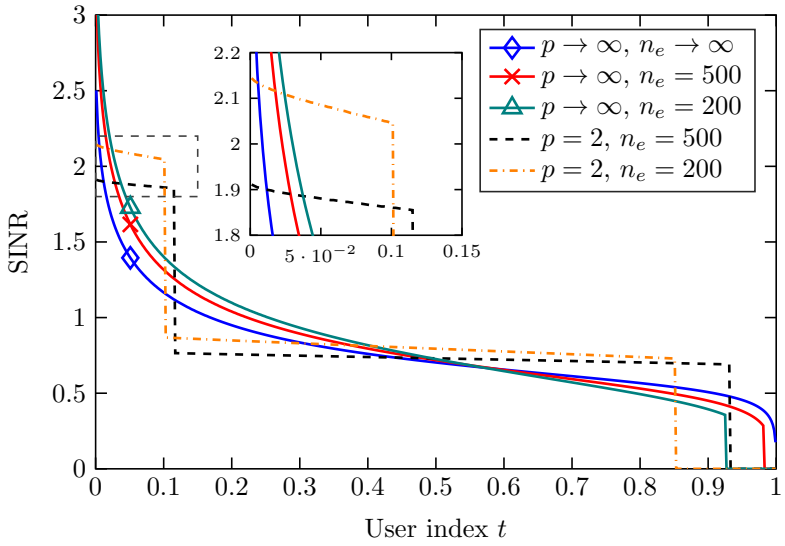

Figure 3.22: ASE versus traffic load α for infinitely ($p \to \infty$) and finitely many ($p = 2$) encoders. The case of $p = 2$ considers optimal second-order coding schemes of rates $R_1 = 4/3$ and $R_2 = 2/3$ bits/symbol.

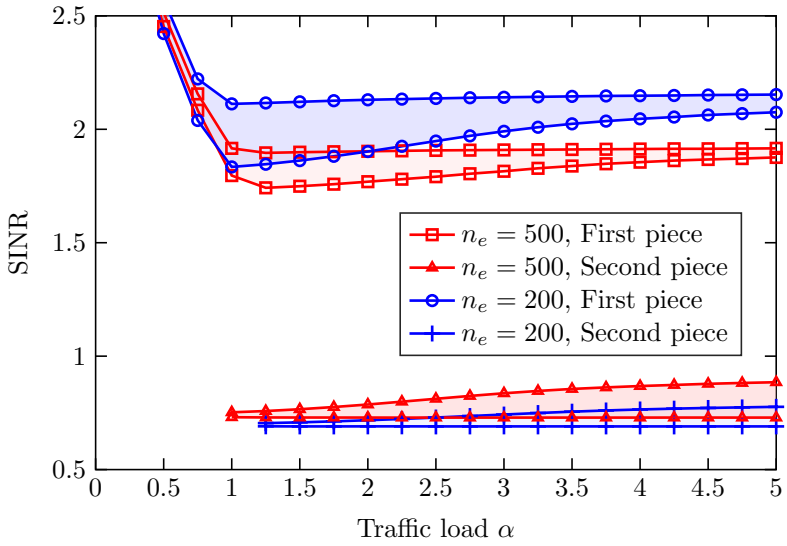

Figure 3.23: ASE versus traffic load α for infinitely ($p \to \infty$) and finitely many ($p = 2$) encoders. The case of $p = 2$ considers optimal second-order coding schemes of rates $R_1 = 4/3$ and $R_2 = 2/3$ bits/symbol.

Figure 3.22 reveals the optimal SINRs allocated to all users. While for infinitely many coding schemes the SINR of each active user follows a continuous profile, no such profile exists when users have only a few of them. The result is not surprising. The optimal SINR profile bears a similar relationship to the profile obtained in Section 3.5.3, where all users employ the same coded modulation scheme. In our case, the optimal SINR profile is practically uniform in each piece of the function. It is also worth noting that the last user assigned with energy is allocated to a SINR equal to the SINR point Γ_* of the most protective code. The dispersion of the optimal SINR profile for the active users of each piece is drawn in Figure 3.23, which shows an almost constant dispersion throughout the offered traffic. This fact indicates that the best

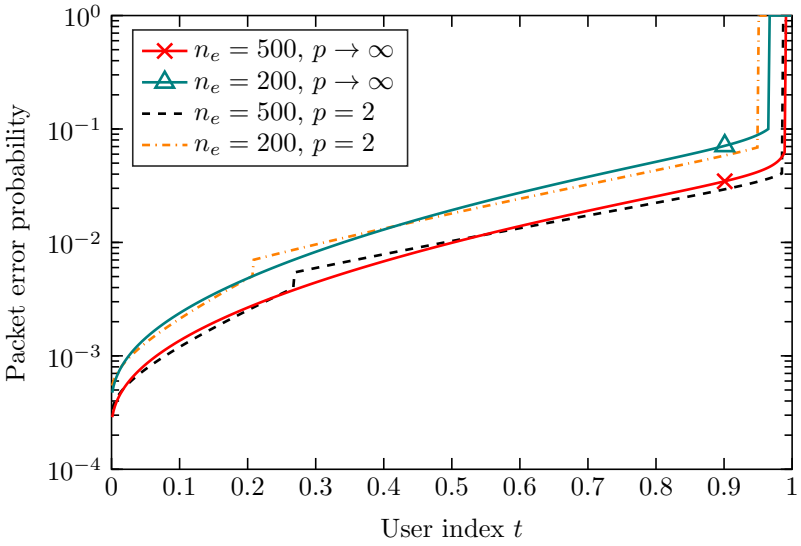

Figure 3.24: Optimal user-PER profile $\mathrm{PER}[\Gamma(t), R(t)]$ for $p \to \infty$ and $p = 2$ encoders. The evaluated traffic load is $\alpha = 3.0$.

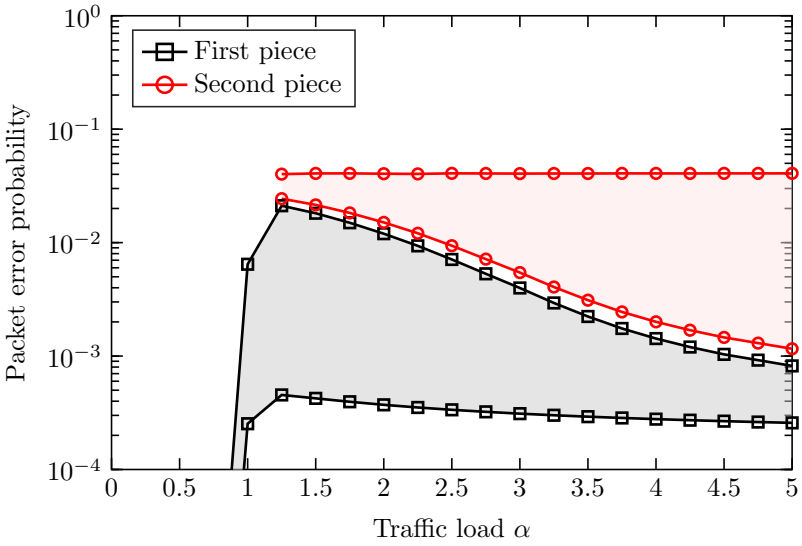

Figure 3.25: Dispersion of the optimal user-PER profile versus traffic load. Simulations correspond to the case $p = 2$ and the blocklength $n_e = 500$ symbols.

transmission strategy allocates non-decreasing packet error probabilities to users. Although the optimal SINR profile is close to a piecewise constant function, the optimal user-PER profile shows a substantially higher dispersion across active users. Some examples are shown in Figure 3.24 where, effectively, the first users are allocated to lower packet error probabilities whereas the last users are allocated to increasing error probabilities. The reason is to minimize the impact of error propagation along the stages of SIC while improving ASE. The dispersion of the optimal user-PER profile versus traffic load is drawn in Figure 3.25, showing dispersions of more than one order of magnitude over each piece. The total dispersion is around two orders of magnitude.

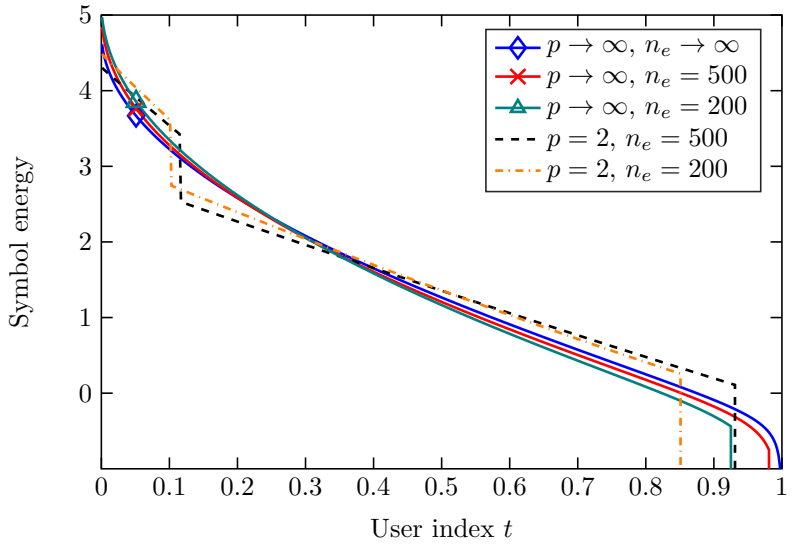

Figure 3.26: Received symbol energy profile $\gamma_{\mathrm{x}}(t)$ at $\alpha = 3.0$ for $p \to \infty$ and $p = 2$ encoders.

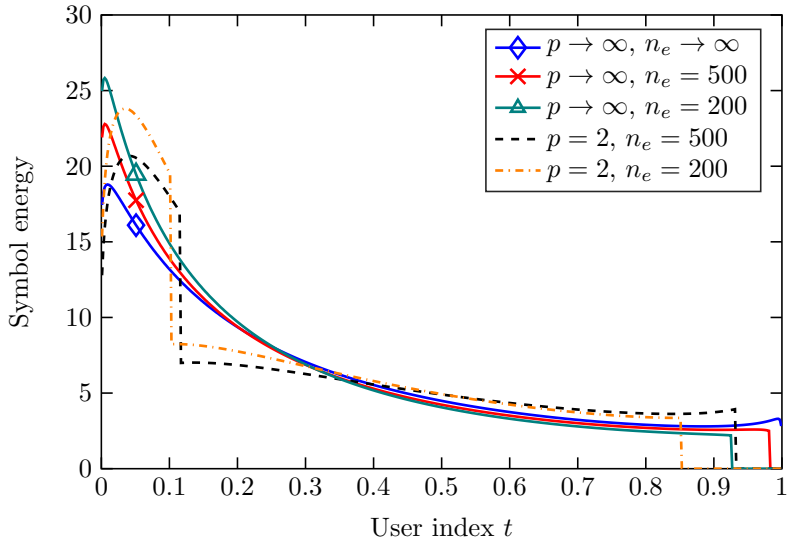

Figure 3.27: Transmitted symbol energy profile $\gamma_{\mathrm{x}}(t)$ at $\alpha = 3.0$ for $p \to \infty$ and $p = 2$ encoders.

The last part of this subsection draws the optimal received and transmitted symbol energy profiles in Figures 3.26 and 3.27, respectively. Regardless of the number of encoders users may have available, the optimal symbol energies allocated to users are adjusted so as to create a distribution at reception with an exponential trend. This corresponds, depending on the channel unbalance, to non-monotonic transmitted energy profiles $\gamma_{\mathrm{x}}(t)$, unlike the previous profiles.

3.5.4.4 The Practical Implementation and Performance Assessment

Although previous allocation designs are derived on the basis of a strong theoretical model and under some assumptions for the specific SIC demodulator, the practicality of the above allocations has not been addressed yet. The following lines address this purpose. Recall that,

we only need that the computation of the PER and RE versus SINR functions is carried out using the same algorithms as the SIC implementation will then use. This section assesses the asymptotic results for many users. More specifically, it elaborates on the relationship between the number of users K and the number of points M needed to solve the stationary point equations derived from asymptotically many users. As described in Section 3.3, we assume that users are pre-loaded with the optimal allocation and that they are able to estimate with high accuracy their individual channel power gains. To that aim, the central node estimates the power of thermal noise, computes the optimal distribution from the variational calculus-based approach addressed in this chapter, and transmits, on a low-rate control channel, a broadcast signal containing a compressed table with the optimal energy and rate allocation and the asymptotic distribution of user-channel power gains. The k-th user employs its individual and accurate estimate $\hat{h}[k]$ from the downlink pilot to determine, based on the known profile $h(t)$, the position it occupies in the overall ordering as

$$\hat{t}_k = h^{-1}\left(\hat{h}[k]\right). \tag{3.57}$$

This index $\hat{t}_k$ is then used to obtain the coding rate and the transmitted symbol energy as $R(\hat{t}_k)$ and $\gamma_{\mathrm{x}}(\hat{t}_k)$. At the simulation level, we consider a satellite scenario with a (simplified) theoretical model for the channel power gain profile $h(t)$. From simulating a satellite scenario with European coverage, we have obtained that the channel attenuations from of all users plus the antenna radiation pattern can be fit into a lognormal distribution. Chip sequences are binary, and the square root raised cosine pulse with roll-off factor 0.35 is used for pulse shaping. The coded modulation scheme combines QPSK with the standardised 3GPP turbo code [69]. We evaluate the coding rates 2/3 and 1/3 with payloads $n_e = 498$ and $n_e = 504$ symbols, and we consider preambles of $n_o = 50$ symbols. The PER versus SINR curve corresponding to each coding scheme is simulated by estimating PER over some SINR points, and by interpolating and smoothing the resulting curve. The first derivative of te PER versus SINR curve is computed through numerical differentiation. During simulations, we use them as lookup tables to obtain accurate findings for the PER and its first derivative at any SINR. The cancellation system estimates amplitude and phase of users decoded successfully (with perfect CRC detection) by correlating the reconstructed packet with the input signal. Carrier frequency offsets are discarded. The RE versus SINR curve is computed likewise as the PER curve, under the above cancellation system.

As stated along the present chapter, the numerical resolution to determine the optimal allocation needs some numerical implementation for the noise plus interference term at each stage of SIC. Our first study evaluates, precisely, the impact that different numerical implementations produce on the individual reliabilities of all users. These correspond to expressions in 2) and 3) exhibited in Section 3.4.1 to compute the term associated with the interval t_i:

$$2) \quad N_{t,2}(t_i) = N_{t,2}(t_{i-1})\left(1 - \frac{\alpha}{M}\,\Phi\left[\frac{\gamma(t_{i-1})}{N_{t,2}(t_{i-1})}, R(t_{i-1})\right]\right), \tag{3.58a}$$

$$3) \quad N_{t,3}(t_i) = N_{t,3}(t_{i-1})\exp\left(-\frac{\alpha}{M}\,\Phi\left[\frac{\gamma(t_{i-1})}{N_{t,3}(t_{i-1})}, R(t_{i-1})\right]\right). \tag{3.58b}$$

We show the obtained results in Figure 3.28 and Table 3.1. As demonstrated, the computation

Table 3.1: Average PER over active users at $\alpha = 3.0$. Theoretical analyses ($K \to \infty$) are computed under $M = 1200$ points. Empirical computations assess the optimal profiles for K users.

Average PER ($\times 10^{-3}$)	$K \to \infty$	$K = 600$	$K = 300$	$K = 150$
Computation as 2)	13.60	23.49	20.35	17.38
Computation as 3)	13.98	10.60	10.37	9.48

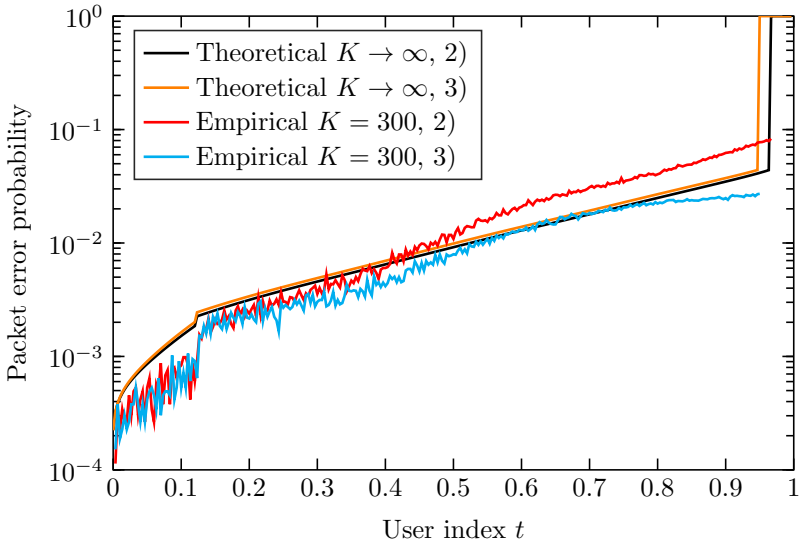

Figure 3.28: Theoretical and empirical user-PER profile for different numerical implementations of the noise plus interference term.

according to 3) results in a more conservative numerical implementation that concludes, theoretically, in a lower ASE while maintaining practically the average PER over active users. This is achieved by preventing the transmission of more users. The computations under method 2) conclude higher mismatching between theoretical and empirical computations. The reason is that for a finite number of users, the decoding and imperfect cancellation errors exhibit a non deterministic behaviour that is propagated along SIC stages, and which can be prevented, to a greater extend, as long as the asymptotic model is computed through a more conservative version of the noise plus interference term. The remaining of this chapter adopts the version 3) (integral form) to compute numerically the noise plus interference term.

The analysis that follows simulates the optimal allocation at a number of K and M pairs and the traffic load $\alpha = 3.0$. Results are shown in Table 3.2. Broadly speaking, the use of the most conservative version for the calculation of the noise plus interference term results in lower average PERs than those predicted theoretically ($K \to \infty$). The above results show less variability as the number of points M increases. Conversely, if the number of points M is low, system performance is not well predicted by the asymptotic computations. This is due to the fact that as M decreases, the proposed algorithm results in a conservative allocation that results, later on, in much lower empirical average PERs. Furthermore, in these cases, system performance is highly sensitive to the number of users K and to the sampling of $0 \leq t \leq 1$ used to compute the theoretical allocation.

Table 3.2: Average PER over active and over all (active / all) users at $\alpha = 3.0$.

PER ($\times 10^{-3}$)	$M = 300$	$M = 600$	$M = 1200$
Theoretical $K \to \infty$	13.57 / 92.48	13.26 / 74.11	13.11 / 64.10
Empirical $K = 1200$	00.85 / 84.96	02.46 / 65.64	06.93 / 58.24
Empirical $K = 600$	10.66 / 61.77	01.14 / 84.35	10.60 / 61.71
Empirical $K = 300$	02.55 / 82.34	07.47 / 67.02	10.37 / 59.85
Empirical $K = 150$	02.53 / 62.38	05.68 / 65.34	09.48 / 55.70

3.6 Concluding Remarks

This chapter has performed an extensive treatment of the allocation problem for a massively populated satellite network in which short length-packets are transmitted, preamble plus payload are modulated following direct-sequence spread spectrum, and the receiver employs an interference cancellation-based multiuser receiver after the conventional matched filter. This chapter has tackled the allocation design accounting for the first iteration of the iterative cancellation receiver proposed in E-SSA, for which the following key system designs have been considered. Firstly, by analysing the random access behaviour of many machine-type communication transmissions under asymptotically many users. Secondly, by accounting for the known asymptotic distribution of channel power gains from all users. Thirdly, by considering many encoders with finite blocklength constraints, each of them characterised by its perfectly known packet error rate versus signal-to-interference-plus-noise ratio curve.

The adoption of a successive decoding strategy aided by redundancy-check error control has conducted its analysis in the asymptotic large-user regime to deal with randomness in packet decoding success and failures propagated throughout the stages. Furthermore, the joint energy and rate allocation has been analysed by tackling simpler problems that allow a better understanding of the triple interplay between energy, rate and reliability, and, at simulation level, by adopting a log-normal theoretical model for the distribution of channel power gains. The stationary point allocations derived via the calculus of variations no longer exhibit the form of continuously differentiable functions. In fact, the optimal profiles are discontinuous in the user variable by silencing users with the worst channels. The best strategy for active users is to protect the weakest users from the increasing interference, propagated throughout the stages, from users cancelled imperfectly and from users decoded unsuccessfully. The latter is achieved, generally, by allocating non-increasing symbol energies (at reception), rates, SINRs and user-reliability. The most relevant simulations disclose that, when users have available few coded modulation schemes, the best energy allocation strategy results very close to the best distribution at the receiver's input, and that the performance loss incurred due to the low number of encoders can be practically compensated if users enable loose power control. Finally, low level simulations of a real SIC implementation reveal the very accurate predictions of our theoretical computations.

Appendix 3.A Proofs

This appendix provides the details of the optimization problems stated in the present chapter. The calculus of variations is used as the optimisation tool. For the reader's interest, the main background in optimization over function spaces can be found in [41, 70]. More concretely, this appendix only focuses on the derivation of the stationary point equations corresponding to the stated problems. Their resolution is dealt with in the respective sections.

All the stationary point equations presented in Section 3.5 can be obtained by evaluating particular cases of the stationary solution computed under asymptotically many coding schemes subject to finite blocklength transmission constraints. The optimisation problem is

$$\max_{0<t_*\leq 1} \max_{\gamma_{\mathrm{x}}(t),R(t)} \alpha \int_0^{t_*} R(t)\mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}, R(t)\right] \mathrm{d}t \quad \text{s.t.} \int_0^{t_*} \gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}}, \tag{3.59}$$

where $N_t(t)$ depends on $\gamma_{\mathrm{x}}(t)$ and $R(t)$ through the dynamic equation and its initial value:

$$\frac{\dot{N}_t(t)}{N_t(t)} = -\alpha\Phi\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}, R(t)\right], \tag{3.60a}$$

$$N_t(0) = 1 + \alpha\theta \int_0^{t_*} \gamma_{\mathrm{x}}(t)h(t)\mathrm{d}t. \tag{3.60b}$$

The way to proceed is firstly, to adopt a function space with smooth and discontinuous elements $\gamma_{\mathrm{x}}(t), R(t) \in \mathcal{C}[0, t_*]$, and secondly, to consider smooth and infinitesimal variations

$$\gamma_{\mathrm{x}}(t) + v_{\mathrm{x}}(t) \tag{3.61a}$$

$$R(t) + v_R(t) \tag{3.61b}$$

in $0 \leq t \leq t_*$. Differentiability of functionals is measured through the norm

$$\|y(t)\| = \max_{0\leq t\leq t_*} |y(t)| + \max_{0<t<t_*} |\dot{y}(t)| + \max_{0<t<t_*} |\ddot{y}(t)| + \dots. \tag{3.62}$$

We impose several conditions on the above variations. On the one hand, the variation $v_{\mathrm{x}}(t)$ may not vary freely as a consequence of the superimposed average energy constraint in the right side of (3.59). More concretely, if we evaluate the constraint at any $\gamma_{\mathrm{x}}(t) \in \mathcal{C}[0, t_*]$, then the admissibility condition for such a smooth variation $v_{\mathrm{x}}(t)$ is

$$\int_0^{t_*} \gamma_{\mathrm{x}}(t)\mathrm{d}t + \int_0^{t_*} v_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}}. \tag{3.63}$$

On the other hand, since there is no constraint imposed over $R(t)$, $v_R(t)$ is only constrained by those smooth perturbations that produce finite norm $\|R(t) + v_R(t)\| < \infty$.

The above does not help us to solve the problem itself but it provides the necessary background to later on understand the operations applied in the adopted function space.

Appendix 3.A.1 Optimal Allocation Designs with Infinitely Many Encoders

The starting point is the Lagrangian

$$\mathcal{L}[\gamma_{\mathrm{x}}(t), R(t)] = \alpha \int_0^{t_*} R(t)\mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}, R(t)\right] - \lambda\left(\int_0^{t_*}\gamma_{\mathrm{x}}(t)\mathrm{d}t - \bar{\gamma}_{\mathrm{x}}\right) \tag{3.64a}$$

$$- \int_0^{t_*}\beta(t)\left(\frac{\dot{N}_t(t)}{N_t(t)} + \alpha\Phi\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}, R(t)\right]\right)\mathrm{d}t. \tag{3.64b}$$

We consider that when $\gamma_{\mathrm{x}}(t), R(t)$ vary as $\gamma_{\mathrm{x}}(t)+v_{\mathrm{x}}(t)$ and $R(t)+v_R(t)$, the noise plus interference profile does it as $N_t(t)+\delta N_t(t)$. Then, if we expand $\mathcal{L}[\gamma_{\mathrm{x}}(t)+v_{\mathrm{x}}(t), R(t)+v_R(t)]$ up to the first order at any point $\gamma_{\mathrm{x}}(t)$ that satisfies the constraint $\int_0^{t_*}\gamma_{\mathrm{x}}(\tau)\mathrm{d}\tau = \bar{\gamma}_{\mathrm{x}}$, we have

$$\mathcal{L} \sim \alpha\int_0^{t_*} R(t)\mathrm{PSR}[\cdot]\mathrm{d}t \tag{3.65a}$$

$$+ \alpha\int_0^{t_*}\left(\mathrm{PSR}[\cdot] + R(t)\mathrm{PSR}_R[\cdot] - \beta(t)\Phi_R[\cdot]\right)v_R(t)\mathrm{d}t \tag{3.65b}$$

$$+ \int_0^{t_*}\left(\alpha R(t)\mathrm{PSR}_\Gamma[\cdot]\frac{h(t)}{N_t(t)} - \lambda - \alpha\beta(t)\Phi_\Gamma[\cdot]\frac{h(t)}{N_t(t)}\right)v_{\mathrm{x}}(t)\mathrm{d}t \tag{3.65c}$$

$$- \int_0^{t_*}\left(\alpha R(t)\mathrm{PSR}_\Gamma[\cdot]\frac{\Gamma(t)}{N_t(t)} - \beta(t)\frac{\dot{N}_t(t)}{N_t^2(t)} - \alpha\beta(t)\Phi_\Gamma[\cdot]\frac{\Gamma(t)}{N_t(t)}\right)\delta N_t(t)\mathrm{d}t \tag{3.65d}$$

$$- \int_0^{t_*}\frac{\beta(t)}{N_t(t)}\delta\dot{N}_t(t)\mathrm{d}t. \tag{3.65e}$$

The functions $\mathrm{PSR}[\Gamma, R], \Phi[\Gamma, R]$ and their derivatives are written obviating their explicit arguments, as $\mathrm{PSR}[\cdot]$, $\Phi[\cdot]$ and $\mathrm{PSR}_\Gamma[\cdot]$, $\mathrm{PSR}_R[\cdot]$, $\Phi_\Gamma[\cdot]$, $\Phi_R[\cdot]$. The first term of $\mathcal{L}$ corresponds to the zeroth-order term, and the rest to the first variation of $\mathcal{L}$, denoted $\delta\mathcal{L}$.

The last term is now integrated by parts with $u = \beta(t)/N_t(t)$ and $\mathrm{d}v = \delta\dot{N}_t(t)\mathrm{d}t$, as

$$\int_0^{t_*}\frac{\beta(t)}{N_t(t)}\delta\dot{N}_t(t)\mathrm{d}t = \frac{\beta(t)}{N_t(t)}\delta N_t(t)\Big|_{t=0}^{t=t_*} - \int_0^{t_*}\left(\frac{\dot{\beta}(t)}{N_t(t)} - \frac{\beta(t)\dot{N}_t(t)}{N_t^2(t)}\right)\delta N_t(t)\mathrm{d}t, \tag{3.66}$$

for which the first variation yields

$$\delta\mathcal{L} = \alpha\int_0^{t_*}\left(\mathrm{PSR}[\cdot] + R(t)\mathrm{PSR}_R[\cdot] - \beta(t)\Phi_R[\cdot]\right)v_R(t)\mathrm{d}t \tag{3.67a}$$

$$+ \int_0^{t_*}\left(\alpha R(t)\mathrm{PSR}_\Gamma[\cdot]\frac{h(t)}{N_t(t)} - \lambda - \alpha\beta(t)\Phi_\Gamma[\cdot]\frac{h(t)}{N_t(t)}\right)v_{\mathrm{x}}(t)\mathrm{d}t \tag{3.67b}$$

$$- \int_0^{t_*}\left(\alpha R(t)\mathrm{PSR}_\Gamma[\cdot]\frac{\Gamma(t)}{N_t(t)} - \frac{\dot{\beta}(t)}{N_t(t)} - \alpha\beta(t)\Phi_\Gamma[\cdot]\frac{\Gamma(t)}{N_t(t)}\right)\delta N_t(t)\mathrm{d}t \tag{3.67c}$$

$$+ \frac{\beta(0)}{N_t(0)}\delta N_t(0) - \frac{\beta(t_*)}{N_t(t_*)}\delta N_t(t_*) \tag{3.67d}$$

with the variations $\delta N_t(0)$ expressed as a function of those $v_{\mathrm{x}}(t)$ as

$$\delta N_t(0) = \alpha\theta\int_0^{t_*} h(t)v_{\mathrm{x}}(t)\mathrm{d}t. \tag{3.68}$$

Now, $\delta\mathcal{L} = 0$ for *every* admissible variation is necessary for first-order optimality. In particular:

1. for $v_{\mathrm{x}}(t) = \delta N_t(t) = 0$ and admissible $v_R(t) \neq 0$, and invoking the Fundamental Lemma of the Calculus of Variations (FLCV), we get

$$\mathrm{PSR}[\cdot] + R(t)\mathrm{PSR}_R[\cdot] - \beta(t)\Phi_R[\cdot] = 0 \quad \text{in} \quad 0 \leq t \leq t_*; \tag{3.69}$$

2. for $v_R(t) = \delta N_t(t) = 0$ and admissible variations $v_{\mathrm{x}}(t) \neq 0$, the FLCV results into the following stationary point equation in $0 \leq t \leq t_*$

$$\alpha R(t)\mathrm{PSR}_\Gamma[\cdot]\frac{h(t)}{N_t(t)} - \lambda - \alpha\beta(t)\Phi_\Gamma[\cdot]\frac{h(t)}{N_t(t)} + \beta(0)\frac{\alpha\theta}{N_t(0)}h(t) = 0; \tag{3.70}$$

3. analogously to the previous case, for $v_R(t) = v_{\mathrm{x}}(t) = 0$, $\delta N_t(t) \neq 0$ in $0 < t < t_*$ and null variations over the end points $\delta N_t(0) = \delta N_t(t_*) = 0$, we have

$$\alpha R(t)\mathrm{PSR}_\Gamma[\cdot]\Gamma(t) - \dot{\beta}(t) - \alpha\beta(t)\Phi_\Gamma[\cdot]\Gamma(t) = 0 \quad \text{in} \quad 0 \leq t \leq t_*. \tag{3.71}$$

   The boundary condition is $\beta(t_*) = 0$.

It is possible to end up with simpler expressions for the stationary point equations. Multiplying (3.70) by $\gamma_{\mathrm{x}}(t)$ and subtracting (3.71) from it, we obtain the relationship

$$\dot{\beta}(t) = \lambda\gamma_{\mathrm{x}}(t) + \beta(0)\frac{\alpha\theta}{N_t(0)}\gamma_{\mathrm{x}}(t)h(t). \tag{3.72}$$

After some straightforward manipulations, one can obtain the closed-form expression $\beta(0) = \lambda\bar{\gamma}_{\mathrm{x}}N_t(0)$, after which, substituting it into (3.72), we get

$$\dot{\beta}(t) = \lambda\gamma_{\mathrm{x}}(t)\left(1 + \alpha\theta\bar{\gamma}_{\mathrm{x}}h(t)\right), \tag{3.73}$$

$$\beta(t) = -\lambda\left(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}}I(t)\right), \tag{3.74}$$

with

$$I_{\mathrm{x}}(t) \triangleq \int_t^{t_*}\gamma_{\mathrm{x}}(\tau)\mathrm{d}\tau \quad \text{and} \quad I(t) \triangleq \int_t^{t_*}\gamma_{\mathrm{x}}(\tau)h(\tau)\mathrm{d}\tau. \tag{3.75}$$

Substituting the results into (3.70), we finally obtain:

1. the stationary point equation for the optimal $\gamma_{\mathrm{x}}(t)$ for a given coding rate profile $R(t)$:

$$\lambda = \frac{\alpha R(t)\mathrm{PSR}_\Gamma[\Gamma(t), R(t)]}{N_t(t)\left(\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}\right) - \alpha(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}}I(t))\Phi_\Gamma[\Gamma(t), R(t)]}, \tag{3.76}$$

2. the stationary point equation for the optimal $R(t)$ for a given energy profile $\gamma_{\mathrm{x}}(t)$:

$$\lambda = \frac{\mathrm{PSR}[\Gamma(t), R(t)] + R(t)\mathrm{PSR}_R[\Gamma(t), R(t)]}{\left(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}}I(t)\right)\Phi_R[\Gamma(t), R(t)]}. \tag{3.77}$$

Moreover, the outer optimization in (3.59) can be found differentiating $\mathcal{L}$ with respect to the

user index t_* when $t_* < 1$. Otherwise, $t_* = 1$. The first gives

$$\alpha R(t_*)\mathrm{PSR}[\Gamma(t_*), R(t_*)] = \lambda\left(\gamma_{\mathrm{x}}(t_*) + \alpha\theta\bar{\gamma}_{\mathrm{x}}\gamma(t)\right), \tag{3.78}$$

which combined with (3.76)–(3.77) at $t = t_*$, energy and rate allocation are produced above the points $\Gamma(t_*)$ and $R(t_*)$ solutions of

$$\Gamma(t_*)\cdot\mathrm{PSR}_{\Gamma}[\Gamma(t_*), R(t_*)] = \mathrm{PSR}[\Gamma(t_*), R(t_*)], \tag{3.79a}$$

$$-R(t_*)\cdot\mathrm{PSR}_{R}[\Gamma(t_*), R(t_*)] = \mathrm{PSR}[\Gamma(t_*), R(t_*)]. \tag{3.79b}$$

Appendix 3.A.2 Particular Cases of the Optimal Allocation Designs

As stated at the beginning of this Appendix, the stationary point equations presented along this chapter correspond to particular cases of the equations (3.76)–(3.77), each of which is reviewed in the sequel:

1. Section 3.5.1 addresses the optimal allocation design constrained by fair reliability. The following problem to determine the optimal $\gamma_{\mathrm{x}}(t) \in \mathcal{C}[0, t_*]$ needs to be addressed:

$$\max_{0<t_*\leq 1}\ \max_{\gamma_{\mathrm{x}}(t)}\ \alpha\int_0^{t_*} r\left(\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}\right)\mathrm{d}t \quad \text{s.t.} \int_0^{t_*}\gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}}. \tag{3.80}$$

   Note that, the above problem corresponds to (3.59) but substituting $\mathrm{PSR}[\Gamma, R] \to r(\Gamma)$ and $\Phi[\Gamma, R] \to \Phi[\Gamma] = \theta(1-\varepsilon)\Gamma\rho$, and addressing the optimisation under $\gamma_{\mathrm{x}}(t)$ and not in $R(t)$. Therefore, the solution concludes the stationary point equation (3.76) particularized at the above values:

$$\lambda = \frac{r'(\Gamma(t))}{N_t(t)\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}c - \alpha\theta(1-\varepsilon)\rho I_{\mathrm{x}}(t)} \quad \text{in} \quad 0 \leq t \leq t_*, \tag{3.81}$$

   with $c = 1 + \alpha\theta I(0)(1-(1-\varepsilon)\rho)$. The optimization under t_* concludes that if $t_* < 1$, then $r(\Gamma(t_*)) = r'(\Gamma(t_*))\cdot\Gamma(t_*)$.

2. Section 3.5.2 finds the best coding rate profile $R(t) \in \mathcal{C}[0, 1]$ in case of fair transmitted powers. The following optimization problem needs to be addressed:

$$\max_{R(t)}\ \alpha\int_0^1 R(t)\cdot\mathrm{PSR}\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right]\mathrm{d}t. \tag{3.82}$$

   The above problem corresponds to (3.59) where the constraint is suppressed and the optimisation is addressed only in $R(t)$ with $t_* = 1$. In this case, the stationary point equation for $R(t)$ can be obtained directly from (3.69), which gives

$$\beta(t)\Phi_R\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right] = \mathrm{PSR}\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right] + R(t)\mathrm{PSR}_R\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right] \tag{3.83}$$

   and with $\beta(t)$ computed, together with the end point $\beta(1) = 0$, as

$$\dot{\beta}(t) = \alpha R(t)\mathrm{PSR}_{\Gamma}\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right]\frac{\gamma(t)}{N_t(t)} - \alpha\beta(t)\Phi_{\Gamma}\left[\frac{\gamma(t)}{N_t(t)}, R(t)\right]\frac{\gamma(t)}{N_t(t)}. \tag{3.84}$$

3. Section 3.5.3 derives the best energy profile $\gamma_{\mathrm{x}}(t) \in \mathcal{C}[0, t_*]$ when all users employ the same coded modulation system. We have the following variational calculus problem:

$$\max_{0<t_*\leq 1} \max_{\gamma_{\mathrm{x}}(t)} \alpha R \int_0^{t_*} \mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}\right] \mathrm{d}t \quad \text{s.t.} \int_0^{t_*} \gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}}. \tag{3.85}$$

Note that, the above problem corresponds to (3.59) where $R(t) = R$, $\mathrm{PSR}[\Gamma, R] \to \mathrm{PSR}[\Gamma]$ and $\Phi[\Gamma, R] \to \Phi[\Gamma] = \theta(1 - \varepsilon[\Gamma])\Gamma{\cdot}\mathrm{PSR}[\Gamma]$. Then, the stationary point equation results (3.76) with the above considerations, as

$$\lambda = \frac{\mathrm{PSR}'[\Gamma(t)]}{N_t(t)\left(\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}\right) - \alpha(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}} I(t))\Phi'[\Gamma(t)]} \quad \text{in} \quad 0 \leq t \leq t_*. \tag{3.86}$$

The optimal t_* is either $t_* = 1$ or such $0 < t_* \leq 1$ that sets $\mathrm{PSR}[\Gamma(t_*)] = \Gamma(t_*)\mathrm{PSR}'[\Gamma(t_*)]$.

Appendix 3.A.3 Optimal Allocation Designs with Finitely Many Encoders

This section solves the optimal allocation problem with finitely many encoders:

$$\max_{t_1,\ldots,t_p} \max_{\gamma_{\mathrm{x}}(t)} \alpha \sum_{k=1}^{p} \int_{t_{k-1}^+}^{t_k^-} R_k \mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t(t)}, R_k\right] \mathrm{d}t \tag{3.87a}$$

$$\text{s.t.} \quad \sum_{k=1}^{p} \int_{t_{k-1}^+}^{t_k^-} \gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}}. \tag{3.87b}$$

In this case, it is straightforward to show, by following the same rules as in the previous section, that the inner optimisation results in the same stationary point equation (3.76) evaluated for the given coding schemes $\mathcal{R}_1, \ldots, \mathcal{R}_p$ and user indices $t_0^+, t_1^- \ldots, t_{p-1}^+, t_p^-$. We get the $k = 1, \ldots, p$ stationary point equations

$$\lambda = \frac{\mathrm{PSR}_\Gamma[\Gamma(t), R_k]}{N_t(t)\left(\frac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}\right) - \alpha(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}} I(t))\Phi_\Gamma[\Gamma(t), R_k]} \quad \text{in} \quad t_{k-1}^+ \leq t \leq t_k^-. \tag{3.88}$$

The optimal user index at the corner t_k is obtained by differentiating $\mathcal{L}$ under t_k. This leads, for the firsts $k = 1, \ldots, p-1$ corners, to

$$\lambda = \alpha \cdot \frac{R_k \mathrm{PSR}[\Gamma(t_k^-), R_k] - R_{k+1}\mathrm{PSR}[\Gamma(t_k^+), R_{k+1}]}{F(t_k^-, R_k) - F(t_k^+, R_{k+1})}, \tag{3.89}$$

with

$$F(t, R) = \gamma(t)(\tfrac{1}{h(t)} + \alpha\theta\bar{\gamma}_{\mathrm{x}}) - \alpha(I_{\mathrm{x}}(t) + \alpha\theta\bar{\gamma}_{\mathrm{x}} I(t))\Phi[\Gamma(t), R]. \tag{3.90}$$

The last corner t_p^- is optimised when $t_p^- = 1$ or when the following equation is satisfied:

$$\mathrm{PSR}[\Gamma(t_p^-), R_p] = \Gamma(t_p^-)\mathrm{PSR}_\Gamma[\Gamma(t_p^-), R_p]. \tag{3.91}$$

4 Optimal Allocation Designs for Iterative Cancellation Receivers

This chapter continues the research in the previous chapter. Recall that, when users employ optimal coding schemes up to the first-order (capacity-achieving), the successive interference cancellation (SIC) receiver can decode *all* transmitting users successfully without exception. As demonstrated previously, this strategy contrasts with the fact that, in emerging wireless networks, users employ coding schemes for short packets, that is, with non-infinite bit-payload. The latter implies that, from a strictly mathematical point of view, reliable communication is not guaranteed to all users regardless of the adopted SIC scheme. This chapter is devoted, precisely, to analysing this problem. Continuing along this line, one option to increase decoding capabilities of the SIC receiver is to iterate the decoding algorithm over the users decoded unsuccessfully after the first iteration. As is thoroughly explained in this chapter, this small modification of the classical SIC algorithm, applied many times, provides almost reliable communication to users in exchange for an increase in the receiver's complexity. Its exact characterisation, however, is challenging from a statistical point of view since the new iterative system results in a chain of dynamic subsystems linked by operations that are statistically dependent. The chapter analyses this iterative cancellation system in detail, and puts forth a system model that greatly simplifies the analysis of this complex receiver. In our view, the main contribution of this chapter is the system model proposed in the following publication

[L1] **F. Molina** and J. Sala-Álvarez, "Average PER performance metrics of iterative successive interference cancellation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 74-77, 2020

which, to the author's knowledge, is the first one to have reported an exact analysis per iteration of a SIC receiver in the infinite-user regime.

This chapter first describes in Section 4.1 the state of the art relative to iterative cancellation systems. For a better understanding of the complexity of this receiver, the system model already published by the author of this thesis [71] ([L1]) is analysed step by step: firstly, investigating a SIC receiver that persists twice in decoding users (two-iteration SIC) in Sections 4.2 and 4.3; and secondly, generalising the model to an iterative SIC receiver that performs an arbitrary number of iterations in Sections 4.5 and 4.6. The allocation designs are investigated under two-iteration SIC in Section 4.4. Conclusions are offered in Section 4.7.

4.1 State of the Art

One of the aspects that conditions, to a greater extent, the performance attainable by SIC is the error correcting power of adopted channel coding schemes. When the blocklength is infinite, $n_e \to \infty$, the received noisy codewords can all be decoded without errors if users adapt their coding rates according to the signal-to-interference-plus-noise ratio (SINR) they experience when going a decoding attempt, or cannot if the users fail to do so. This abrupt behaviour has motivated, throughout the literature, the design of energy-code allocation strategies by considering simply the logarithm as the mapping function between energy and rate [48], or more complex functions [46]. These works assume that users transmit unboundedly long packets, or with a sufficiently long payload (in the order of thousands of symbols).

Nevertheless, when users transmit short packets the picture is different. As shown in the previous chapter, the decoder always succeeds with a certain probability at every SINR point despite the lower error correction power that achieved by decoders for short packets. Notably, the decoding behaviour of short-length decoding is smooth in the SINR. Eventually, some users may be decoded unsuccessfully after the first iteration of the SIC receiver. Chapter 3 has dealt extensively with a system model for such a SIC receiver. A competitive approach to tackle the processing of users decoded unsuccessfully is to iterate the SIC algorithm over all users [45], or only over users remaining decoded unsuccessfully [40]. Naturally, both strategies entail an increase of the receiver's complexity, but system performance is stunning. The choice of one of the latter strategies depends, strictly, on whether the receiver operates with redundancy check error control (CRC) or not. In [45], the authors consider users employing the same modulation and error correcting code, and a receiver constituted of a minimum mean square error (MMSE) multiuser detector that iterates over a bank of channel decoders. Many iterations further improve the SINR of all users, and the packet error rate versus SINR curve is adopted as a lookup table to compute system performance in the last iteration. In [40, 47], SIC is adopted after the bank of single-user matched filters, which analyses the low-SNR regime of the MMSE-based multiuser receiver. In this case, users enter CRCs to enable error control at reception and the receiver persists in decoding only users decoded unsuccessfully. Other approaches are followed to improve symbol reliabilities when error control is not available [72]. Both schemes exploit the performance of the receiver as long as users arrive with an appropriate power unbalance.

This chapter derives an accurate model for the operation of iterative SIC that operates under redundancy-check error control. The challenging task is the model adopted for the persistent decoding operations of the same user. Throughout the literature, iterative SIC has been analysed on very few circumstances. As an example, such a system is analysed in [47] by setting initial and convergence conditions in terms of the SINRs that achieve target packet error probabilities; a trigger condition is set heuristically to ensure that iterative SIC converges to low packet error probabilities (around 10^{-3} or 10^{-4}) if users are allocated, at the first iteration, to error probabilities PER $\lessapprox 0.9$. Is this universal condition too pessimistic, or can it be done better? In the opinion of this author, the necessity of introducing such a trigger criterion is due to the model already developed overlooks the memory in decoding the same user throughout iterations. This chapter sheds light on this aspect, and uses the model as the starting point to optimise system performance.

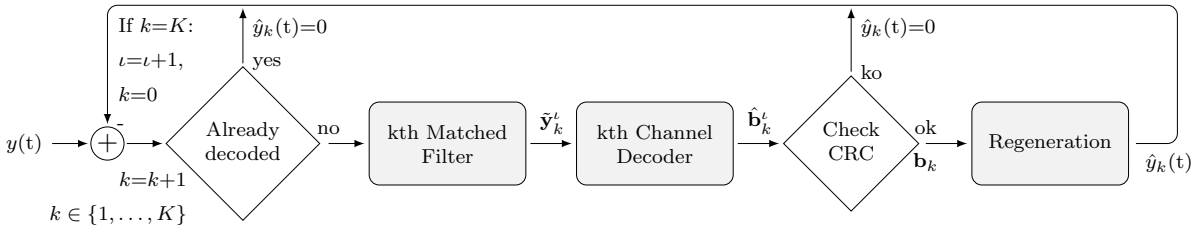

Figure 4.1: Block diagram of an iterative SIC receiver aided by redundancy-check error control.

4.2 The Model of Two-Iteration SIC

This section analyses in detail the operative of iterative SIC from a statistical point of view. The block diagram of the decoding-cancellation system is shown in Figure 4.1. Unlike the previous chapter, this section considers a SIC receiver that persists twice in decoding users. In Figure 4.1, this corresponds to $\iota \in \{1, 2\}$. The starting point of this section is the baseband signal[1] received from many direct-sequence users

$$y(\mathrm{t}) = \sum_{k_1=1}^{K} \sqrt{\gamma[k_1]}\; e^{j\Omega_{k_1}} \sum_{m=1}^{n} \mathbf{s}_{k_1}[m] c_{k_1,m}(\mathrm{t} - mT - \tau_{k_1}) + w(\mathrm{t}). \tag{4.1}$$

The index set $\mathcal{K}_1 \triangleq \{1, \ldots, K\}$ is employed to index users as $1 \leq k_1 \leq K$, where the subscript 1 is added, unlike in the previous chapter, to emphasise the notation relative to the first iteration. In the expression above, $\sqrt{\gamma[k_1]}e^{j\Omega_{k_1}}$ is the complex amplitude received from user k_1. As described in the previous chapter, the first iteration of SIC processes all users in K stages (one per user) and, so as not to lose the thread, the specific operations performed by the receiver are reviewed next. Recall that, at stage k_1, the decoder operates with the output of the matched filter (despreader)

$$\tilde{\mathbf{y}}_{k_1}^1 = \sqrt{\gamma[k_1]}\; e^{j\Omega_{k_1}}\; \mathbf{s}_{k_1} + \sqrt{N_t^1[k_1]}\; \mathbf{w}_{k_1}^1 \tag{4.2}$$

to obtain the packet $\mathbf{s}_{k_1}$ transmitted by user k_1. $\mathbf{w}_{k_1}^1$ is the Gaussian noise plus interference term affecting user k_1, and $N_t^1[k_1]$ is the noise plus interference level. The superscripts "1" mention explicitly the iteration index $\iota = 1$. Then, the i-th decoding function is used to recover the bit sequence transmitted by user k_1 as $\hat{\mathbf{b}}_{k_1}^1 = f_i^{-1}(\tilde{\mathbf{y}}_{k_1}^1)$, after which CRC checks if the decoding has been successful or not. If so, the received user is reconstructed as $\hat{y}_{k_1}(\mathrm{t})$, and is cancelled at the waveform level as $y(\mathrm{t}) - \hat{y}_{k_1}(\mathrm{t})$. The model adopted for the complete operation of SIC in the first iteration is based on the Gaussian interference assumption, which allows to compute the SINR of user k_1 and its noise plus interference term as in Chapter 3:

$$\Gamma^1[k_1] = \frac{\gamma[k_1]}{N_t^1[k_1]}, \tag{4.3}$$

$$N_t^1[k_1] = 1 + \frac{\theta}{N} \sum_{i=1}^{k_1-1} \epsilon_1[i]\gamma[i] + \frac{\theta}{N} \sum_{i=k_1+1}^{K} \gamma[i]. \tag{4.4}$$

[1] Carrier frequency offsets are discarded in our analysis.

In this expression, $\epsilon_1[1{\leq}i{<}k_1]$ are the binary random variables associated with the decoding-cancellation operation of users $i < k_1$ already processed. The joint decoding-cancellation system is analysed, as in Chapter 3, by adopting the known packet error rate (PER) and residual energy (RE) curves. For user i, $\epsilon_1[i]$ equals 1 with probability $\mathrm{PER}[\Gamma^1[i], R[i]]$, and $\varepsilon[\Gamma^1[i], R[i]]$ with probability $\mathrm{PSR}[\Gamma^1[i], R[i]] \triangleq 1{-}\mathrm{PER}[\Gamma^1[i], R[i]]$. PSR stands for packet success rate.

The above is, in fact, a summary of a part of the previous chapter, after which the first SIC iteration ends with K_2 packet errors and succeeds with $K{-}K_2$ users. The novelty introduced in this section may be found in that the K_2 users that survive decoded unsuccessfully undergo a new decoding attempt following the same SIC scheme. Bear in mind that at least one user must be decoded successfully and cancelled to proceed to the new iteration. Otherwise, further decoding attempts will fail. This section continues assuming that $K_2 < K$. Henceforth, the notation employed differs from that already published in [71] and leads to a more simple analysis.

The index set $\mathcal{K}_1 = \{1, \ldots, K\}$ is partitioned into those corresponding to the subset of users decoded successfully $\mathcal{K}_1^{\mathrm{ok}}$ and that of users decoded unsuccessfully $\mathcal{K}_1^{\mathrm{ko}}$. Recall that only users $k_1 \in \mathcal{K}_1^{\mathrm{ko}}$ are processed again. Two problems arise: firstly, that the number of users to be processed is smaller than K; and secondly, whether the matched filter outputs for the same user and different iterations are statistically dependent or not. To tackle the first issue the index set in the second iteration is $\mathcal{K}_2 \triangleq \{1, \ldots, K_2\}$, users are indexed as $1 \leq k_2 \leq K_2$, and the mappings

$$\phi_{1,2} \;:\; k_1 \in \mathcal{K}_1^{\mathrm{ok}} \longmapsto k_2 \in \{1, \ldots, K_2\}, \tag{4.5}$$

$$\phi_{2,1} \;:\; k_2 \in \{1, \ldots, K_2\} \longmapsto k_1 \in \mathcal{K}_1^{\mathrm{ok}}, \tag{4.6}$$

relate the indices of users between iterations. It is worth noting that $\phi_{1,2}$ can only be applied to the subset of users $\mathcal{K}_1^{\mathrm{ok}}$ that survive (decoded unsuccessfully) after the first iteration, while $\phi_{2,1}$ can be applied to all indices $1 \leq k_2 \leq K_2$. Thus, user k in the first iteration $\iota = 1$ is ranked at position $\phi_{1,2}[k]$ in the second iteration $\iota = 2$, and conversely, user k in the second iteration $\iota = 2$ was ranked the $\phi_{2,1}[k]$-th in the first iteration $\iota = 1$. Figure 4.2 exemplifies the above.

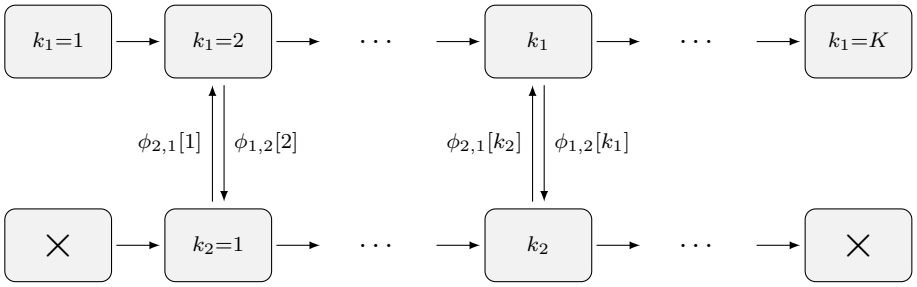

Figure 4.2: An example of the relationship between the indices of users of different iterations.

The second issue is related to the processing of users. Recall that the matched filter is only computed for users that survive after the first iteration. The model assumes Gaussian statistic for interference. The output of the matched filter for the k_2-th user in the second iteration is

$$\tilde{\mathbf{y}}_k^2 = \sqrt{\gamma[k]}\; e^{j\Omega_k}\; \mathbf{s}_k + \sqrt{N_t^2[k]}\; \mathbf{w}_k^2 \quad \text{with} \quad k = \phi_{2,1}[k_2]. \tag{4.7}$$

$N_t^2[\phi_{2,1}[k_2]]$ is variance of noise plus interference after despreading, and $\mathbf{w}_k^2$ is a unit-power

Gaussian term. In the expression above, the Gaussian term $\mathbf{w}_k^2$ is not statistically independent from the corresponding $\mathbf{w}_k^1$ (4.2) in the first iteration. Actually, both terms are related through the linear decomposition of $\mathbf{w}_k^1$ as

$$\sqrt{N_t^1[k]}\ \mathbf{w}_k^1 = \sqrt{N_t^2[k]}\ \mathbf{w}_k^2 + \sqrt{N_t^1[k]-N_t^2[k]}\ \mathbf{w}_k^c, \tag{4.8}$$

where $\mathbf{w}_k^2$ is the noise plus interference term affecting user k after subtracting $\mathbf{w}_k^c$. The latter term corresponds to the interference cancelled between the stage k in the first iteration and the stage $\phi_{1,2}[k]$ in the second iteration. Note that $\mathbf{w}_k^1$ (in the first SIC iteration) comprises two Gaussian independent terms, whereas in the second iteration (4.7) is only valid if a packet error event occurred in the first iteration. That said, the i-th decoding function is applied on $\tilde{\mathbf{y}}_k^2$ to determine the bit sequence $\hat{\mathbf{b}}_k^2 = f_i^{-1}(\tilde{\mathbf{y}}_k^2)$, and cancellation is produced when CRC checks out.

The SINR after despreading user k_2 in the second iteration is computed as

$$\Gamma^2[k] = \frac{\gamma[k]}{N_t^2[k]} \quad \text{with} \quad k = \phi_{2,1}[k_2]. \tag{4.9}$$

The difficulty of analysing this second iteration is the model chosen for channel decoding since, as shown before, the despreader outputs of different iterations are not statistically independent. The conclusion drawn from the previous rationale is that the same PER characteristic used in the first iteration evaluated at the new SINR as $\mathrm{PER}[\Gamma^2, R_i]$, does not model the magnitude of packet error decoding correctly[2]. Note that the above does not include any variable relative to the previous iteration. The following lines conduct an explanation to the model adopted for channel decoding in the second iteration. Contrasting the two expressions for the output of the matched filter of the same user k in different iterations

$$\tilde{\mathbf{y}}_k^1 = \sqrt{\gamma[k]}\ e^{j\Omega_k}\ \mathbf{s}_k + \sqrt{N_t^2[k]}\ \mathbf{w}_k^2 + \sqrt{N_t^1[k]-N_t^2[k]}\ \mathbf{w}_k^c \tag{4.10}$$

$$\tilde{\mathbf{y}}_k^2 = \sqrt{\gamma[k]}\ e^{j\Omega_k}\ \mathbf{s}_k + \sqrt{N_t^2[k]}\ \mathbf{w}_k^2, \tag{4.11}$$

intuition leads to think that channel decoding will succeed more frequently when a high amount of interference ($N_t^1[k]-N_t^2[k]$) is cancelled between iterations. The latter means, essentially, that $N_t^2[k]$ shall be much lower than $N_t^1[k]$ for the channel decoder to operate with a substantially different noise realisation, or, in practice, with an independent noise realisation. The application of this result to user k_2 entails that the second SIC iteration exploits, performance-wise, the decoder's capabilities when $\Gamma^2[k_2]$ is sufficiently high when compared to $\Gamma^1[\phi_{2,1}[k_2]]$. Remarkably, the important finding is that the packet error probability for persisting decoding operations depends on the current and previous SINRs of the same user. Continuing along this line, the novelty introduced herein is the modelling of channel decoding for the second iteration through the *multivariate* PER function

$$\mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i], \tag{4.12}$$

[2]It is easy to check that the model is incorrect. Assume a trivial case where no interference is cancelled. Then $\Gamma^2 = \Gamma^1$, and therefore, the decoder would give the same result as in the first iteration, that is, a packet failure event. In contrast, the model $\mathrm{PER}[\Gamma^2, R_i]$ produces a packet error probability notably different.

which stands for the packet error probability for a user with SINR Γ^2 that has been unsuccessfully decoded in the previous iteration under SINR Γ^1, and which generalises the definition of the PER versus SINR characteristic well-known in the literature as an instance of block error rate [73], packet loss ratio [47], frame error rate [45] or simply, packet error rate [49]. Recall that the last argument "R_i" is only used to identify the encoder, and thus, the above characteristic is now a bivariate function of the SINR at each iteration.

The PER functions associated with the i-th coding scheme can be computed together: firstly, by generating vector $\mathbf{b}$ and encoding it as $\sqrt{\gamma}e^{j\Omega}f_i(\mathbf{b})$ (Ω is uniformly distributed in $[-\pi,+\pi]$); secondly, by subjecting communication to Gaussian noises $\mathbf{w}^2,\mathbf{w}^c\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$; thirdly, by generating the signals

$$\mathbf{y}^1=\sqrt{\gamma}e^{j\Omega}f_i(\mathbf{b})+\sqrt{N_t^2}\ \mathbf{w}^2+\sqrt{N_t^1-N_t^2}\ \mathbf{w}^c, \tag{4.13}$$

$$\mathbf{y}^2=\sqrt{\gamma}e^{j\Omega}f_i(\mathbf{b})+\sqrt{N_t^2}\ \mathbf{w}^2, \tag{4.14}$$

with the SINRs $\Gamma^1=\gamma/N_t^1$ and $\Gamma^2=\gamma/N_t^2$; and finally, by computing the following expressions

$$\mathrm{PER}[\Gamma^1,R_i]=\Pr\left[f_i^{-1}(\mathbf{y}^1)\neq\mathbf{b}\right], \tag{4.15}$$

$$\mathrm{PER}_2[\Gamma^1,\Gamma^2,R_i]=\Pr\left[f_i^{-1}(\mathbf{y}^2)\neq\mathbf{b}\mid f_i^{-1}(\mathbf{y}^1)\neq\mathbf{b}\right]. \tag{4.16}$$

The introduction of such a function (4.16) allows to describe the *exact* behaviour of channel decoding conditioned on previous packet decoding failures. The counterpart is that its calculation is computationally exhaustive since it constitutes a bivariate function of (Γ^1,Γ^2).

Under the above definitions, the noise plus interference term affecting user k_2 in the second iteration reads as follows, and whose terms are described in the sequel:

$$N_t^2[\phi_{2,1}[k_2]]=1+\xi_1+\frac{\theta}{N}\sum_{i=1}^{k_2-1}\epsilon_2[i]\gamma[\phi_{2,1}[i]]+\frac{\theta}{N}\sum_{i=k_2+1}^{K_2}\gamma[\phi_{2,1}[i]], \tag{4.17}$$

1. The **second term** ξ_1 corresponds to the aggregate interference from users successfully decoded and imperfectly cancelled in the first iteration. Mathematically, this term is computed by adding the contributions of users whose decoding random variable $\epsilon_1[1\leq k_1\leq K]$ denotes a successful decoding event.

2. The **third term** corresponds to the aggregate interference from users $i<k_2$ processed in the second iteration before stage k_2. For its computation, recall that $k=\phi_{2,1}[i]$ produces the user index of the i-th decoding in the first ordering. In the expression above, $\epsilon_2[i]$ is a random variable associated with the decoding and cancellation of user i, which equals 1 if packet decoding fails with probability $\mathrm{PER}_2[\Gamma^1[i],\Gamma^2[i],R[i]]$ and equals $\varepsilon[\Gamma^2[i]]$ with complementary probability $\mathrm{PSR}_2[\Gamma^1[i],\Gamma^2[i],R[i]]\triangleq 1-\mathrm{PER}_2[\Gamma^1[i],\Gamma^2[i],R[i]]$. The random variables $\epsilon_2[1\leq k_2\leq K_2]$ are statistically independent of each other.

3. The **fourth term** is the aggregate interference from users $k_2<i\leq K_2$ still not processed at stage k_2.

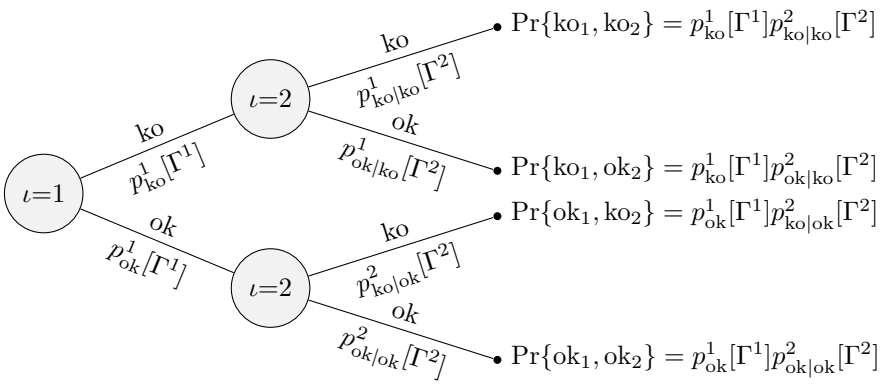

Figure 4.3: Probability tree of user undergoing a second decoding attempt regardless of the first outcome.

4.2.1 Analytical Findings for the PER Function

The above model provides, in contrast to prior works, mathematical expressions that allow for the *exact* analysis of two-iteration SIC based on a multivariate decoding function. Nevertheless, the major weakness of the above model is that the calculation of such a function is computationally intense, and it can only be obtained by simulating a particular coding and modulation scheme. This subsection delves into the understanding of this multivariate function, and derives analytical findings for such a function that allow its approximation and also the evaluation of two-iteration SIC when theoretical coded modulation schemes are employed, provided that the classical PER versus SINR curve PER[Γ] is known.

Figure 4.3 depicts the probability tree corresponding to the persistent decoding of one user. Two strategies are compared: firstly, when the receiver operates with error control, in which case the second iteration $\iota = 2$ only proceeds if the first decoding fails. The packet error probability after two decoding attempts results $p^1_{\mathrm{ko}}[\Gamma^1]p^2_{\mathrm{ko|ko}}[\Gamma^2]$; and secondly, when the receiver does not operate with error control, in which case the second iteration is performed regardless of the outcome of the first attempt since the latter is unknown to the receiver. The packet error probability for the latter case is $p^1_{\mathrm{ko}}[\Gamma^1]p^2_{\mathrm{ko|ko}}[\Gamma^2] + p^1_{\mathrm{ok}}[\Gamma^1]p^2_{\mathrm{ko|ok}}[\Gamma^2]$. Clearly, the first approach achieves lower error probability and the performance improvement is due to the error detection in the intermediate step. Now, substituting the previous notation by the PER functions defined previously as $p^1_{\mathrm{ko}}[\Gamma^1] = \mathrm{PER}[\Gamma^1, R_i]$, $p^2_{\mathrm{ko|ko}}[\Gamma^2] = \mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i]$ and $p^1_{\mathrm{ko}}[\Gamma^1]p^2_{\mathrm{ko|ko}}[\Gamma^2] + p^1_{\mathrm{ok}}[\Gamma^1]p^2_{\mathrm{ko|ok}}[\Gamma^2] = \mathrm{PER}[\Gamma^2, R_i]$, the following relationship is stated for a channel decoding operation governed by error control:

$$\mathrm{PER}[\Gamma^1, R_i]\mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i] \leq \mathrm{PER}[\Gamma^2, R_i]. \tag{4.18}$$

A last remark: this bound shall be substituted by an equality when the algorithm does not operate with error control in the intermediate step. The latter equation sets the upper bound

$$\mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i] \leq \frac{\mathrm{PER}[\Gamma^2, R_i]}{\mathrm{PER}[\Gamma^1, R_i]} \tag{4.19}$$

which evidences a lot of insight to understand the behaviour of an iterative cancellation receiver.

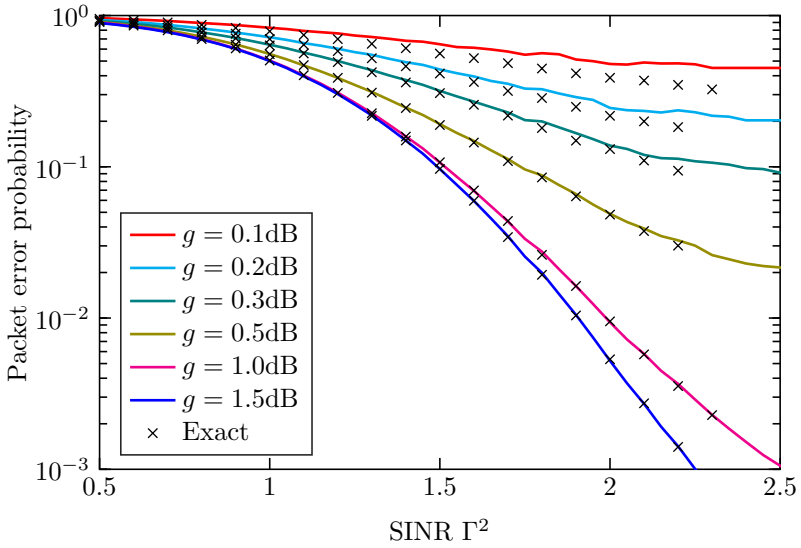

Figure 4.4: Contour lines of $\mathrm{PER}_2[\Gamma^1, \Gamma^2]$ at the SINR gain $g = \Gamma^2/\Gamma^1$. The exact computations (cross markers) are drawn over the theoretical upper-bound (4.19) (solid lines). Simulations are carried out for the turbo code standardised in DVB-RCS [74] with 440 bits, coding rate 1/2, QPSK modulation, and 10 max-log-map iterations. The symbol payload is $n_e = 440$.

Firstly, the second decoding attempt does not succeed, $\mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i] \approx 1$, when low interference is cancelled between iterations, and thus $\Gamma^2 \approx \Gamma^1$. Secondly, the receiver exploits the decoding performance when $\Gamma^2 \gg \Gamma^1$.

Figure 4.4 depicts the results of a short simulation, run to evidence the tightness of the above bound for a standardised coded modulation system (details about its parameters may be found in the caption of Figure 4.4). The results are not surprising; the bound is very tight when Γ^2 is sufficiently larger than Γ^1, since the equivocation probability $p^2_{\mathrm{ko|ok}}[\Gamma^2]$ turns out to be very low at high Γ^2/Γ^1 ratios. Notably, it is very unlikely that the decoder will fail in decoding a packet previously decoded successfully unless the SINR gain $g = \Gamma^2/\Gamma^1$ is very low. As shown in the attached figure, poor SINR gains, for instance $g = 0.1$dB or $g = 0.2$dB, entail the inaccurate tightness of the above upper bound. Remarkably, simulations reveal that SINR gains of less than half a dB are enough to use the bound as a strict equality.

4.3 The Asymptotic Model of Two-Iteration SIC

As in the previous chapter, the system is in general random, since users can be decoded correctly or incorrectly at each stage. For the case studied herein, randomness in decoding users is propagated throughout SIC stages of every iteration even more than in the previous chapter, where users are subjected to a single decoding operation. This fact complicates the obtention of a tractable form for its understanding. Rather, a mathematically tractable form to address the analysis of such an iterative cancellation system may be found by resorting to the user-asymptotic regime, in which case the system model defined in Section 4.2 takes a very attractive form. The asymptotic model for each of the iterations is described below.

4.3.1 The First Iteration of Two-Iteration SIC

The asymptotic model corresponding to the first iteration is the one described in Section 3.4, and reviewed below. More concretely, the continuous user indexing

$$t = \lim_{K\to\infty} \frac{k_1}{K} \tag{4.20}$$

identifies users in the first SIC iteration, at the same time the traffic load $\alpha \triangleq K/N$ is asymptotically held constant. The SINR of user t after despreading reads

$$\Gamma^1(t) = \frac{\gamma(t)}{N_t^1(t)}, \tag{4.21}$$

with the noise plus interference term $N_t^1(t)$ computed from any of the following equivalent forms:

$$1) \quad N_{t,1}^1(t) = 1 + \alpha\theta \int_0^t q_1\left[\frac{\gamma(\tau)}{N_{t,1}^1(\tau)}, R(\tau)\right]\gamma(\tau)\mathrm{d}\tau + \alpha\theta\int_t^1 \gamma(\tau)\mathrm{d}\tau, \tag{4.22a}$$

$$2) \quad \frac{\dot{N}_{t,2}^1(t)}{N_{t,2}^1(t)} = -\alpha\Phi_1\left[\frac{\gamma(t)}{N_{t,2}^1(t)}, R(t)\right] \quad \text{and} \quad N_{t,2}^1(0) = N_t^1(0), \tag{4.22b}$$

$$3) \quad N_{t,3}^1(t) = N_t^1(0)\exp\left(-\alpha\int_0^t \Phi_1\left[\frac{\gamma(\tau)}{N_{t,3}^1(\tau)}, R(\tau)\right]\mathrm{d}\tau\right). \tag{4.22c}$$

Herein, $q_1[\Gamma^1, R_i] \triangleq 1-(1-\varepsilon[\Gamma^1, R_i])\mathrm{PSR}[\Gamma^1, R_i]$ and $\Phi_1[\Gamma^1, R_i] \triangleq \theta(1-\varepsilon[\Gamma^1, R_i])\Gamma^1\cdot\mathrm{PSR}[\Gamma^1, R_i]$, and the initial noise term is

$$N_t^1(0) = N_{t,1}^1(0) = N_{t,2}^1(0) = N_{t,3}^1(0) = 1 + \alpha\theta\int_0^1 \gamma_{\mathrm{x}}(t)h(t)\mathrm{d}t. \tag{4.23}$$

The average number of users decoded unsuccessfully after the first iteration is computed in the asymptotic large-user regime as

$$\overline{\mathrm{per}}_1 = \lim_{K\to\infty} \frac{1}{K}\sum_{k_1=1}^{K} \mathrm{PER}[\Gamma^1[k_1], R[k_1]] = \int_0^1 \mathrm{PER}[\Gamma^1(t), R(t)]\mathrm{d}t. \tag{4.24}$$

4.3.2 The Second Iteration of Two-Iteration SIC

This subsection analyses the second SIC iteration, which processes users previously decoded unsuccessfully. In this new iteration, the number of users to be processed is considered, again, asymptotically large. The assessment of such a model for a finite number of users is addressed in later sections. Under the user-asymptotic assumption, the indexing in the second iteration is

$$t_2 = \lim_{K_2\to\infty} \frac{k_2}{K_2}, \tag{4.25}$$

and analogously as in the first iteration, the traffic load in the second iteration $K_2/N \to \alpha_2$ is asymptotically held constant. As evidenced next, the adoption of this asymptotic large-user regime allows to predict the decoding power of two-iteration SIC in a comfortable and simple

way. One of the most interesting results is the tractable analytic relation for the mappings $\phi_{2,1}$ and $\phi_{1,2}$ in Section 4.2, that are now defined as

$$\phi_{1,2} \ : \ t \in [0,1] \longmapsto t_2 \in [0,1] \tag{4.26}$$

$$\phi_{2,1} \ : \ t_2 \in [0,1] \longmapsto t \in [0,1] \tag{4.27}$$

Specifically, under the asymptotic model it is true that each interval of length $\mathrm{d}t$ in the first iteration can be expressed as the union of two disjoint intervals, whose lengths add up to $\mathrm{d}t$, as

$$\mathrm{d}t = \mathrm{PER}[\Gamma^1(t), R(t)]\mathrm{d}t + \mathrm{PSR}[\Gamma^1(t), R(t)]\mathrm{d}t. \tag{4.28}$$

$\mathrm{PER}[\Gamma^1(t), R(t)]\mathrm{d}t$ is the interval of users decoded unsuccessfully and $\mathrm{PSR}[\Gamma^1(t), R(t)]\mathrm{d}t$ is that of users decoded successfully[3]. Since only users in the first interval undergo a second decoding attempt, each differential $\mathrm{d}t$ contains an interval $\mathrm{d}\phi_{1,2}(t)$ of users decoded unsuccessfully

$$\mathrm{d}\phi_{1,2}(t) = \frac{1}{\overline{\mathrm{per}}_1} \ \mathrm{PER}[\Gamma^1(t), R(t)]\mathrm{d}t, \tag{4.29}$$

the fraction of users decoded unsuccessfully in the interval $\mathrm{d}t$ over the total of those decoded unsuccessfully in the first iteration. The initial value problem $\dot{\phi}_{1,2}(t) = \frac{1}{\overline{\mathrm{per}}_1}\mathrm{PER}[\Gamma^1(t), R(t)]$ with $\phi_{1,2}(0) = 0$ can be solved straightforwardly as

$$\phi_{1,2}(t) = \frac{1}{\overline{\mathrm{per}}_1} \int_0^t \mathrm{PER}[\Gamma^1(\tau), R(\tau)]\mathrm{d}\tau. \tag{4.30}$$

Note that, since $\phi_{1,2}(0) = 0$ and $\phi_{1,2}(1) = 1$, $t_2 = \phi_{1,2}(t)$ defines a warping of $0 \leq t \leq 1$ onto itself: $0 \leq t_2 \leq 1$. The latter definitions allow to start defining the SINR after despreading for the user t_2 in the second iteration, previously given by (4.9), as

$$\Gamma^2(t) = \frac{\gamma(t)}{N_t^2(t)} \quad \text{with} \quad t = \phi_{2,1}(t_2), \tag{4.31}$$

with the noise plus interference term affecting the same user t_2 given by

$$\begin{aligned} N_t^2(\phi_{2,1}(t_2)) = 1 + \xi_1 + \alpha_2\theta \int_0^{t_2} q_2\Big[\Gamma^1(\phi_{2,1}(\tau)), \Gamma^2(\phi_{2,1}(\tau)), R(\phi_{2,1}(\tau))\Big]\gamma(\phi_{2,1}(\tau))\mathrm{d}\tau \\ + \alpha_2\theta \int_{t_2}^1 \gamma(\phi_{2,1}(\tau))\mathrm{d}\tau, \end{aligned} \tag{4.32}$$

and with $q_2[\Gamma^1, \Gamma^2, R_i] \triangleq 1 - (1 - \varepsilon[\Gamma^2, R_i])(1 - \mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i])$. It is still possible to obtain simpler equations of the noise plus interference level evolving with SIC in terms of a differential equation or its integral version. The formal proof is extended in Appendix 4.B.1, and concludes the following equivalent continuous expressions in the variable t:

$$1) \quad N_{t,1}^2(t) = 1 + \xi_1 + \alpha\theta \int_0^t q_2 \left[\frac{\gamma(\tau)}{N_{t,1}^1(\tau)}, \frac{\gamma(\tau)}{N_{t,1}^2(\tau)}, R(\tau)\right] \mathrm{PER}\left[\frac{\gamma(\tau)}{N_{t,1}^1(\tau)}, R(\tau)\right] \gamma(\tau)\mathrm{d}\tau \tag{4.33a}$$

[3]The model is only valid if $R(t)$ is smooth in $0 \leq t \leq 1$, or failing that, if $R(t)$ is smooth except at some discontinuity points.

$$+\alpha\theta\int_t^1 \gamma(\tau)\mathrm{PER}\left[\frac{\gamma(\tau)}{N_{t,1}^1(\tau)}, R(\tau)\right]\mathrm{d}\tau,$$

$$2)\quad \frac{\dot{N}_{t,2}^2(t)}{N_{t,2}^2(t)} = -\alpha\Phi_2\left[\frac{\gamma(t)}{N_{t,2}^1(t)}, \frac{\gamma(t)}{N_{t,2}^2(t)}, R(t)\right] \quad \text{and} \quad N_{t,2}^2(0) = N_{t,2}^1(1), \tag{4.33b}$$

$$3)\quad N_{t,3}^2(t) = N_{t,3}^1(1)\exp\left(-\alpha\int_0^t \Phi_2\left[\frac{\gamma(\tau)}{N_{t,3}^1(\tau)}, \frac{\gamma(\tau)}{N_{t,3}^2(\tau)}, R(\tau)\right]\mathrm{d}\tau\right). \tag{4.33c}$$

The major changes have been the substitution of α_2 by α, the introduction of the decoding-cancellation function $\Phi_2[\Gamma^1, \Gamma^2, R_i] = \theta(1-\varepsilon[\Gamma^2, R_i])\Gamma^2(1-\mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i])\mathrm{PER}[\Gamma^1, R_i]$, and the initial noise plus interference term $N_{t,2}^2(0) = N_{t,3}^2(0) = N_t^1(1)$. In the expressions above, numerical subscripts and superscripts are included for notation purposes. The computation of the above terms can be extended by combining different versions of the terms $N_t^1(t)$ and $N_t^2(t)$. Note that both the differential and the integral expressions reveal that the initial noise plus interference level at the second iteration $N_t^2(0) = N_t^1(1)$ coincides with the endpoint of the first iteration. The second iteration continues with what was left at the end of the first iteration. Furthermore, unlike when handling a finite number of users and only users decoded unsuccessfully are processed again, the interest in analysing the system in the asymptotic large-user regime is that the calculations are equivalent to processing all user-indices, again, and weighting their cancellation by $\mathrm{PER}[\Gamma^1, R_i]$. Moreover, if we also make use of the approximation $\mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i] \approx \mathrm{PER}[\Gamma^2, R_i]/\mathrm{PER}[\Gamma^1, R_i]$, then $\Phi_2[\Gamma^1, \Gamma^2, R_i]$ can be simplified as

$$\Phi_2[\Gamma^1, \Gamma^2, R_i] \approx \theta(1-\varepsilon[\Gamma^2, R_i])\Gamma^2(\mathrm{PER}[\Gamma^1, R_i] - \mathrm{PER}[\Gamma^2, R_i]). \tag{4.34}$$

The mitigation of noise plus interference in the second SIC iteration depends on the magnitude

$$\mathrm{PER}[\Gamma^1, R_i] - \mathrm{PER}[\Gamma^2, R_i]. \tag{4.35}$$

Thus, the iterative cancellation receiver exploits its decoding performance only when Γ^2/Γ^1 is sufficiently high so that $\mathrm{PER}[\Gamma^2, R_i] \ll \mathrm{PER}[\Gamma^1, R_i]$.

4.4 Allocation Designs for Two-Iteration SIC

The design of optimal allocations is addressed herein with the aim of maximising asymptotic spectral efficiency (ASE). The spectral efficiency (SE) achieved in the first SIC iteration aggregates the effective transmission rates of all users. Analogously, the SE achieved in the second SIC iteration aggregates the effective transmission rates of all users decoded unsuccessfully in the first iteration. They are respectively given by

$$\mathrm{SE}_1 = \frac{1}{N}\sum_{k_1=1}^{K} R[k_1]\ \mathrm{PSR}[\Gamma^1[k_1], R[k_1]], \tag{4.36}$$

$$\mathrm{SE}_2 = \frac{1}{N}\sum_{k_2=1}^{K_2} R[\phi_{2,1}[k_2]]\ \mathrm{PSR}_2[\Gamma^1[\phi_{2,1}[k_2]], \Gamma^2[\phi_{2,1}[k_2]], R[\phi_{2,1}[k_2]]]. \tag{4.37}$$

After some manipulations, explicitly shown in Appendix 4.B.2, the ASE after two iterations is cast as the compact expression

$$\mathrm{ASE} = \mathrm{ASE}_1 + \mathrm{ASE}_2 = \alpha \int_0^1 R(t)\ \mathrm{PSR}\left[\frac{\gamma(t)}{N_t^2(t)}, R(t)\right] \mathrm{d}t, \tag{4.38}$$

which is, actually, the figure of merit to be maximised in the following. Recall that, in the infinite-user regime, the noise plus interference terms $N_t^1(t)$ and $N_t^2(t)$ can be computed from any of the equations (4.22a)–(4.22c) and (4.33a)–(4.33c).

4.4.1 Rate Allocation with Fair Transmitted Power

This subsection solves the rate allocation problem when users transmit the same power and the receiver performs two-iteration SIC. Recall that, in this case, the received symbol energy distribution $\gamma(t) = \bar{\gamma}_{\mathrm{x}} h(t)$ presents the unbalance produced by the distribution of the channel power gains from all users. The cases of fair and optimal reliability are also studied.

To evaluate the performance gain of two-iteration SIC versus the one-iteration SIC, we first address a naive study in which the rate allocation already designed accounting for the one-iteration SIC in Section 3.5.2 (of the previous chapter) is run under two-iteration SIC. The ASE (in Table 4.1) hardly varies whereas the average PER over all users (in Table 4.2) is highly reduced. This leads to think that the more iterations performed, the more iterative SIC exploits its decoding power and allows to allocate vanishing error probabilities. Note that the average PERs decay more than two orders of magnitude when the offered traffic is sufficiently high. The highest ASE gains are obtained for short packets since the allocations designed in Section 3.5.2 allocate users to higher error probabilities and to lower system throughputs. Remarkably, the potential benefit of iterative SIC is in coding systems having softer PER curves.

Table 4.1: ASE for several blocklengths and traffic loads. The optimal rate allocation designed under one-iteration SIC (1-SIC) is compared with the performance achieved by the same allocation under two-iteration SIC (2-SIC).

Blocklength / Load	$\alpha = 2.0$	$\alpha = 3.0$	$\alpha = 4.0$	$\alpha = 5.0$
1-SIC $n_e = 500$, Optimal	2.01	2.40	2.65	2.83
2-SIC $n_e = 500$	2.03	2.41	2.67	2.85
1-SIC $n_e = 200$, Optimal	1.86	2.18	2.37	2.50
2-SIC $n_e = 200$	1.88	2.21	2.41	2.54

Table 4.2: Average PER over all users for several blocklengths and traffic loads. The optimal rate allocation designed under one-iteration SIC (1-SIC) is compared with the performance achieved by the same allocation under two-iteration SIC (2-SIC).

Blocklength / Load	$\alpha = 2.0$	$\alpha = 3.0$	$\alpha = 4.0$	$\alpha = 5.0$
1-SIC $n_e = 500$, Optimal	$1.69 \cdot 10^{-2}$	$1.72 \cdot 10^{-2}$	$1.76 \cdot 10^{-2}$	$1.81 \cdot 10^{-2}$
2-SIC $n_e = 500$	$2.47 \cdot 10^{-5}$	$4.95 \cdot 10^{-6}$	$1.12 \cdot 10^{-6}$	$2.64 \cdot 10^{-7}$
1-SIC $n_e = 200$, Optimal	$2.37 \cdot 10^{-2}$	$2.43 \cdot 10^{-2}$	$2.50 \cdot 10^{-2}$	$2.59 \cdot 10^{-2}$
2-SIC $n_e = 200$	$1.15 \cdot 10^{-4}$	$3.57 \cdot 10^{-5}$	$1.28 \cdot 10^{-5}$	$4.98 \cdot 10^{-6}$

Another approach that attains higher performance is when all users are subject to the per-user reliability constraint $0 < \rho \leq 1$ (at the last SIC iteration). The rate allocation problem is to determine such $R(t)$ profile that achieves the required reliability, as:

$$R(0 \leq t \leq 1) > 0 \in \mathcal{C}[0,1] \quad , \quad \text{PSR}\left[\frac{\gamma(t)}{N_t^I(t)}, R(t)\right] = \rho. \tag{4.39}$$

In this case, the attained network performance is $\text{ASE} = \alpha\rho\int_0^1 R(t)\mathrm{d}t$.

The above expression considers the general term $N_t^I(t)$ for the noise plus interference level according to the I iterations of the SIC receiver. For one-iteration SIC ($I = 1$), $N_t^1(t)$ is computed from one of the expressions (4.22a)–(4.22c), and for two-iteration SIC ($I = 2$), $N_t^2(t)$ from one of the expressions (4.33a)–(4.33c). A last remark, when users operate in the infinite blocklength (IBL) regime, $n_e \to \infty$, reliable communication, $\rho = 1$, is possible performing one-iteration SIC as long as user t encodes data using a capacity-achieving coding scheme with rate $R(t) = \log(1 + \gamma(t)/N_t^1(t))$. In the finite blocklength (FBL) regime, no explicit form for such $R(t)$ profile that solves the above problem is found, in which case the following numerical procedures are proposed:

1. *Numerical resolution for one-iteration SIC*: The allocation design is straightforward. At step $0 \leq i \leq M-1$, the $R(t_i)$ that satisfies $\text{PSR}[\gamma(t_i)/N_t^1(t_i), R(t_i)]=\rho$ is determined. The initial noise term is $N_t^1(t_0) = 1 + \frac{\alpha\theta}{M}\sum_{i=0}^{M-1}\gamma(t_i)$, and $N_t^1(t_i)$ is updated at each step as $N_t^1(t_{i+1}) = N_t^1(t_i)\exp(-\frac{\alpha}{M}\Phi_1[\gamma(t_i)/N_t^1(t_i), R(t_i)])$.
2. *Numerical resolution for two-iteration SIC*: The value $N_t^2(t_0)$ is assumed known temporarily, and found after M steps by bisection search. At step $0 \leq i \leq M-1$, $R(t_i)$ is found by solving $\text{PSR}[\gamma(t_i)/N_t^2(t_i), R(t_i)] = \rho$, after which the noise plus interference terms $N_t^1(t_i)$ and $N_t^2(t_i)$ are updated as $N_t^1(t_{i+1}) = N_t^1(t_i)\exp(-\frac{\alpha}{M}\Phi_1[\gamma(t_i)/N_t^1(t_i), R(t_i)])$ and $N_t^2(t_{i+1}) = N_t^2(t_i)\exp(-\frac{\alpha}{M}\Phi_2[\gamma(t_i)/N_t^1(t_i), \gamma(t_i)/N_t^2(t_i), R(t_i)])$.

In both cases, the expressions (4.22c) and (4.33c) have been adopted for the calculation of the noise plus interference terms. Nonetheless, the above algorithm can be modified by combining other expressions.

Lastly, it is still possible to improve on the previous design if, instead of imposing the same reliability constraint to all users, the per-user reliability is allowed to vary freely. The case of one-iteration SIC corresponds exactly to the problem solved in Section 3.5.2. The following lines concentrate on designing the rate allocation profile $R(t)$ when the receiver performs two-iteration SIC. The optimisation problem is

$$\max_{R(t)} \ \alpha\int_0^1 R(t)\ \text{PSR}\left[\frac{\gamma(t)}{N_t^2(t)}, R(t)\right]\mathrm{d}t, \tag{4.40}$$

where the term $N_t^2(t)$ depends on $R(t)$ through any of the expressions (4.22a)–(4.22c) and (4.33a)–(4.33c). We proceed as in Section 3.5.2, by deriving the stationary point equation and

by proposing a numerical method to obtain the solution. The stationary $R(t)$ profile satisfies

$$\begin{aligned} R(t)\mathrm{PSR}_R\left[\frac{\gamma(t)}{N_t^2(t)}, R(t)\right] + \mathrm{PSR}\left[\frac{\gamma(t)}{N_t^2(t)}, R(t)\right] \; + \\ -\beta_1(t)(\Phi_1)_R\left[\frac{\gamma(t)}{N_t^2(t)}, R(t)\right] - \beta_2(t)(\Phi_2)_R\left[\frac{\gamma(t)}{N_t^1(t)}, \frac{\gamma(t)}{N_t^2(t)}, R(t)\right] = 0 \end{aligned} \tag{4.41}$$

in $0 \leq t \leq 1$, with $\beta_1(t)$ and $\beta_2(t)$ computed by solving the following differential equations with the boundaries $\beta_1(1) = \beta_2(0)$ and $\beta_2(1) = 0$:

$$\dot{\beta}_1(t) = -\alpha\frac{\gamma(t)}{N_t^1(t)}\left(\beta_1(t)(\Phi_1)_{\Gamma^1}\left[\frac{\gamma(t)}{N_t^1(t)}, R(t)\right] + \beta_2(t)(\Phi_2)_{\Gamma^1}\left[\frac{\gamma(t)}{N_t^1(t)}, \frac{\gamma(t)}{N_t^2(t)}, R(t)\right]\right), \tag{4.42}$$

$$\dot{\beta}_2(t) = \alpha\frac{\gamma(t)}{N_t^2(t)}\left(R(t)\mathrm{PSR}_{\Gamma^2}\left[\frac{\gamma(t)}{N_t^2(t)}, R(t)\right] - \beta_2(t)(\Phi_2)_{\Gamma^2}\left[\frac{\gamma(t)}{N_t^1(t)}, \frac{\gamma(t)}{N_t^2(t)}, R(t)\right]\right). \tag{4.43}$$

We have validated that the use of a simple and fast numerical resolution, such as the ones implemented in Chapter 3, leads to numerical inaccuracies. The reason is that, in this case, the stationary point equations are very sensitive to the computation of the terms $N_t^1(t)$ and $N_t^2(t)$. In previous implementations, $N_t^1(t_{i+1})$ and $N_t^2(t_{i+1})$ are computed explicitly as a function of the previous terms $N_t^1(t_i)$ and $N_t^2(t_i)$ (see Section 3.4.1 for further details). There exist, mainly, two ways to increase the accuracy of these computations: by increasing the number of points M, or by resorting to more accurate, and therefore, more complex solving methods. The rate allocation problem solved herein follows the second approach. Specifically, the above variational calculus problem is discretised into M intervals as

$$\max_{R(t_0),\ldots,R(t_{M-1})} \; \frac{\alpha}{M}\sum_{i=0}^{M-1} R(t_i)\ \mathrm{PSR}\left[\frac{\gamma(t_i)}{N_t^2(t_i)}, R(t_i)\right], \tag{4.44}$$

and we solve it using sequential quadratic programming [75], where the noise plus interference term at each step t_i corresponds to the solution of the non-linear equations

$$N_t^2(t_i) = N_t^2(t_{i-1})\exp\left(-\frac{\alpha}{M}\Phi_2\left[\frac{\gamma(t_i)}{N_t^1(t_i)}, \frac{\gamma(t_i)}{N_t^2(t_i)}, R(t_i)\right]\right) \quad \text{for} \quad i = 1, \ldots, M-1 \tag{4.45a}$$

$$N_t^1(t_i) = N_t^1(t_{i-1})\exp\left(-\frac{\alpha}{M}\Phi_1\left[\frac{\gamma(t_i)}{N_t^1(t_i)}, R(t_i)\right]\right) \quad \text{for} \quad i = 1, \ldots, M-1 \tag{4.45b}$$

$$N_t^2(t_0) = N_t^1(t_{M-1}) \quad , \quad N_t^1(t_0) = 1 + \frac{\alpha\theta}{M}\sum_{i=0}^{M-1}\gamma(t_i). \tag{4.45c}$$

To speed up computations and get higher numerical accuracy, the sequential quadratic programming algorithm is aided by the analytical expression of the gradient of the cost function. The latter corresponds, exactly, to the left part of (4.41) evaluated at each $t = t_i$. Additionally, $\beta_1(t)$ and $\beta_2(t)$ are calculated numerically, starting from the $(M-1)$-th term $\beta_2(t_{M-1}) = 0$ down to $\beta_2(t_0)$. $\beta_2(t_i)$ for $i = 0, \ldots, M-2$ are computed recursively as $\beta_2(t_i) = \beta_2(t_{i+1}) - \frac{1}{M}\dot{\beta}_2(t_{i+1})$. The computation of $\beta_1(t)$ is analogous starting from $\beta_1(t_{M-1}) = \beta_2(t_0)$.

In general, it is easy to obtain a non-smooth function due to the high dimensionality of the adopted vector space, a known problem in the literature. So as to obtain a smooth so-

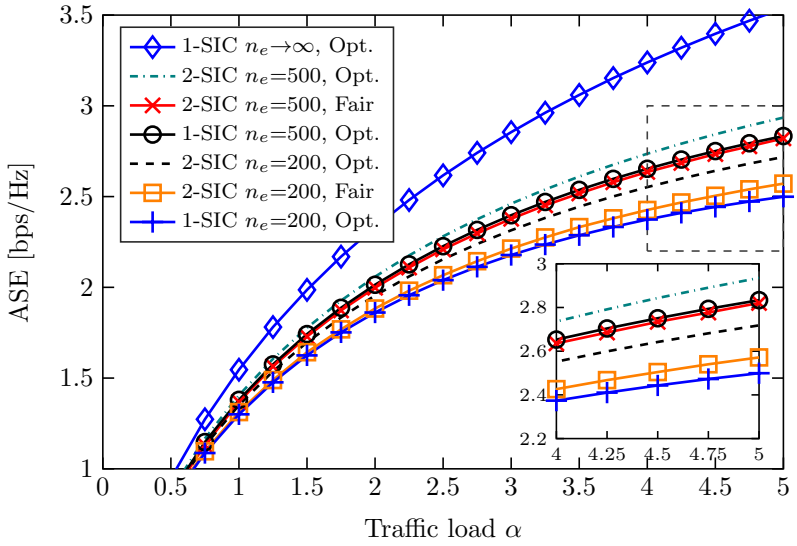

Figure 4.5: ASE versus traffic load α. Computations compare: (i) the IBL regime $n_e \to \infty$; (ii) the optimal ASE under one-iteration (1-SIC) and two-iteration (2-SIC) SIC; and (iii) 2-SIC with fair reliability ρ, set with the same average reliability as that resulted under optimal 2-SIC.

lution with high accuracy, we propose an incremental method in traffic load that computes the optimal rate allocation profile at $\alpha \to 0$, whose analytic form solves $\mathrm{PSR}\left[\gamma(t_i), R(t_i)\right] = -R(t_i)\mathrm{PSR}_R\left[\gamma(t_i), R(t_i)\right]$ in $i = 0, \ldots, M-1$, and uses it to initialise the problem (4.44) at $\alpha = \Delta\alpha$. The resulting rate allocation is then used to initialise (4.44) at $\alpha = 2\Delta\alpha$, and the process is repeated at $\alpha = n\Delta\alpha$ until reaching the target traffic load.

The remaining of this section compares the optimal allocations designed herein and their performance with those in Section 3.5.2 (of Chapter 3) designed under one-iteration SIC. Simulation parameters are the same as in Chapter 3. The average symbol energy is $\bar{\gamma}_{\mathrm{x}} = 8\mathrm{dB}$ and users are subject to channel power gains lognormally distributed with unit mean and deviation $\sigma = 3\mathrm{dB}$. The uncanceled energy factor is $\varepsilon(\Gamma) = 1\%$ and the decorrelator factor θ is set to 1. Simulations are carried out under $M = 1000$ points. Figure 4.5 depicts ASE versus traffic load for optimal and fair-constrained reliability holding the same average reliability. As shown, two-iteration SIC outperforms one-iteration SIC, specially so, at low blocklengths. Also, the benefit of its use increases with α. This is because the more interference between iterations is cancelled, the more the decoding performance is exploited, which occurs more frequently the larger α is. The cases evaluated under fair reliability and two-iteration SIC do not always improve the ASE achieved with optimal reliability and one-iteration SIC. In fact, ASE is only shown to be superior when the blocklength is low. Broadly speaking, the non-unbalanced reliability of users results in a penalising factor to ASE. In all cases, however, ASE in the FBL regime with one-iteration and two-iteration SIC are shown to be far from the theoretical limit in the IBL regime since the transmission of short packets strongly degrades the maximum attainable rate relative to the capacity of the Gaussian channel.

Some examples of the optimal rate allocation profiles are illustrated in Figure 4.6. As shown, rate allocation constrained by fair reliability allows to increase only the coding rate of the first users in detriment of the last users' coding rate. Contrarily, the transmission rate of practically

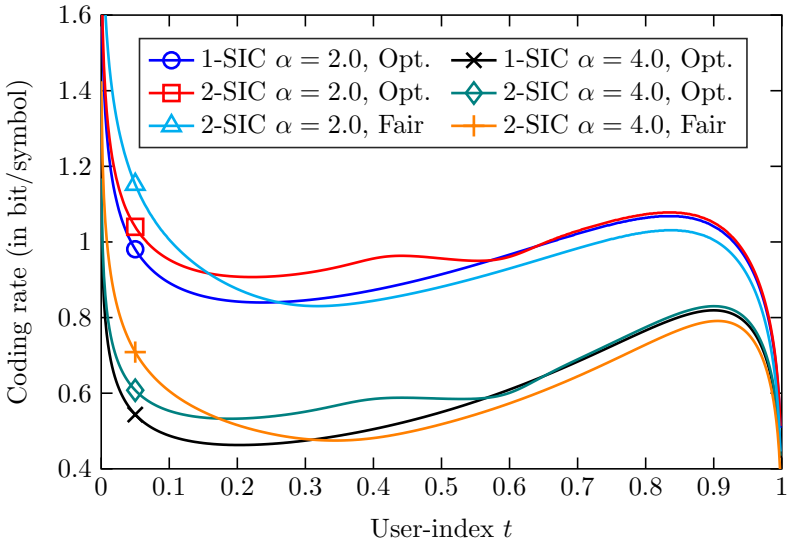

Figure 4.6: Optimal rate allocation profile $R(t)$ with optimal and fair-constrained reliability. The block-length evaluated is $n_e = 200$ symbols.

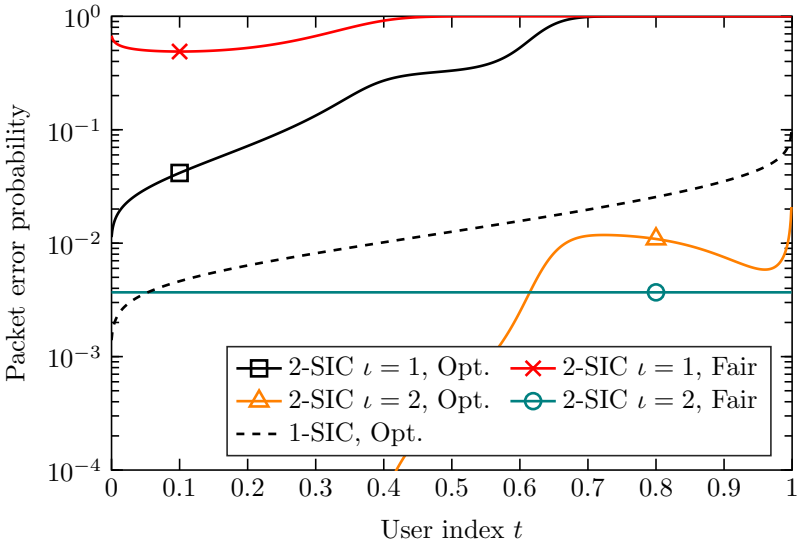

Figure 4.7: User-PER profile $\mathrm{PER}[\Gamma^{\iota}(t), R(t)]$ at the traffic load $\alpha = 4.0$ and $n_e = 200$ symbols. The iteration index is $\iota \in \{1, 2\}$.

all users can be increased provided that no fair reliability constraint is imposed. Additionally, how users benefit, individually, from two-iteration SIC in terms of packet error probability is depicted in Figure 4.7 and addressed below. The optimal rate allocation under two-iteration SIC allocates high PERs to users in the first SIC iteration. In the depicted example, the first iteration decodes 51% of the traffic load, leaving the rest for the second iteration, which succeeds in decoding approximately 99% of the remaining load. As observed, the rate allocation under two-iteration SIC with optimal reliability increases the individual performance of the strongest users providing them with almost reliable communication. Rather, the case of fair reliability allocation substantially differs from the previous case, even if both systems are compared under

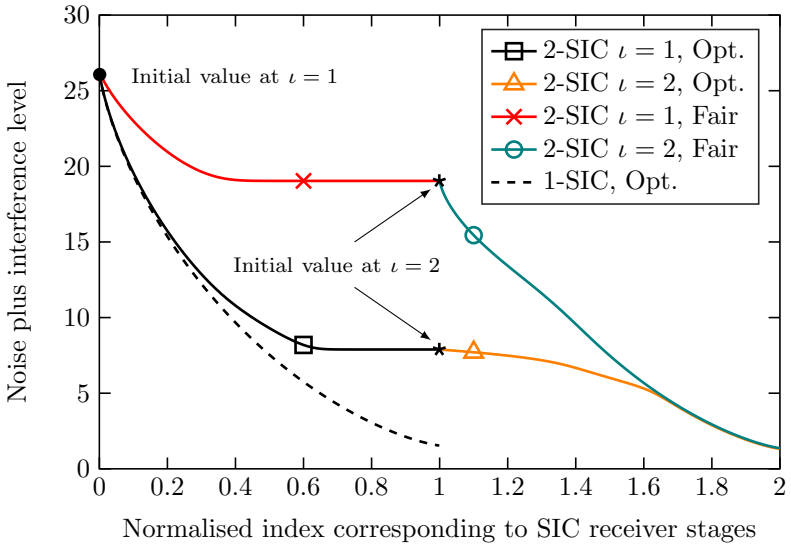

Figure 4.8: Evolution of the noise plus interference level throughout SIC receiver stages. Its evolution throughout the first iteration corresponds to the abscissae range 0 to 1, and that of the second iteration to the range 1 to 2. The traffic load is $\alpha = 4.0$ and the blocklength is $n_e = 200$ symbols.

the same average reliability. In the case of fair reliability, the first iteration cancels 16% of the traffic load and the second iteration deals successfully with 99% of the remaining load. In any case, an interesting conclusion is that the optimal user-PER profile is non-monotonic, unlike the optimal user-PER profile designed under one-iteration SIC. This conclusions slightly differs from that in our conference paper [76] due to the computation method used for the optimal $R(t)$ profile. Notably, user-reliabilities are very sensitive to the computation method.

The previous behaviour is also present in Figure 4.8, albeit in a different way. It depicts the evolution of noise plus interference level as SIC progresses throughout stages. The optimal allocation constrained by fair reliability leaves the major part of interference to the second SIC iteration. Rather, the case of optimal reliability deals with a big part of the interference in the first iteration and leaves weaker users to the second iteration. Both approaches contrast with the results shown for one-iteration SIC (dashed), which removes as much interference as possible in the first decoding attempt.

4.4.2 Energy Allocation with Fair Encoding Rate

This section addresses the design of the best transmission strategy in terms of energy allocated to users when they share the same coded modulation system. The described setting is closely related with the physical layer configuration proposed in Enhanced Spread Spectrum ALOHA [40], in which iterative SIC exploits system performance. This subsection, nevertheless, investigates a two-iteration SIC. Further iterations of the adopted receiver are analysed in later sections. Herein, the PER versus SINR curve, PER[Γ], obviating the argument R is adopted, for which the scope of the present section is to determine the best energy allocation rule $\gamma_{\mathrm{x}}(t) \in \mathcal{C}[0, t_*]$ subject to an average energy constraint, as:

$$\max_{0<t_*\leq 1} \max_{\gamma_{\mathrm{x}}(t)} \alpha R \int_0^{t_*} \mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t^2(t)}\right] \mathrm{d}t \tag{4.46a}$$

$$\text{s.t. } \int_0^{t_*} \gamma_{\mathrm{x}}(t)\mathrm{d}t = \bar{\gamma}_{\mathrm{x}} \tag{4.46b}$$

$$\text{s.t. } N_t^2(t) = N_t^1(t_*) \exp\left(-\alpha \int_0^t \Phi_2\left[\frac{\gamma_{\mathrm{x}}(\tau)h(\tau)}{N_t^1(\tau)}, \frac{\gamma_{\mathrm{x}}(\tau)h(\tau)}{N_t^2(\tau)}\right] \mathrm{d}\tau\right) \tag{4.46c}$$

$$\text{s.t. } N_t^1(t) = N_t^1(0) \exp\left(-\alpha \int_0^t \Phi_1\left[\frac{\gamma_{\mathrm{x}}(\tau)h(\tau)}{N_t^1(\tau)}\right] \mathrm{d}\tau\right) \tag{4.46d}$$

$$\text{s.t. } N_t^1(0) = 1 + \alpha\theta \int_0^{t_*} \gamma_{\mathrm{x}}(t)h(t)\mathrm{d}t. \tag{4.46e}$$

Note that we seek again for a discontinuous allocation in $0 \leq t \leq 1$, which allocates null energy to users $t_* < t \leq 1$. For the design above, we proceed in the same way as in the previous section. The inner problem is solved numerically through sequential quadratic programming, where the above asymptotic expressions are sampled at the indices $t_0, \ldots, t_{M-1}$. The following optimization needs to be addressed numerically

$$\max_{0\leq M_*\leq M-1} \max_{\gamma(t_0),\ldots,\gamma(t_{M_*})} \frac{\alpha R}{M} \sum_{i=0}^{M_*} \mathrm{PSR}\left[\frac{\gamma(t_i)}{N_t^2(t_i)}\right] \tag{4.47a}$$

$$\text{s.t.} \quad \bar{\gamma}_{\mathrm{x}} = \frac{1}{M} \sum_{i=0}^{M_*} \frac{\gamma(t_i)}{h(t_i)} \tag{4.47b}$$

$$\text{s.t.} \quad N_t^1(t_i) = N_t^1(t_{i-1}) \exp\left(-\frac{\alpha}{M} \Phi_1\left[\frac{\gamma(t_i)}{N_t^1(t_i)}\right]\right) \quad i \geq 1 \tag{4.47c}$$

$$\text{s.t.} \quad N_t^2(t_i) = N_t^2(t_{i-1}) \exp\left(-\frac{\alpha}{M} \Phi_2\left[\frac{\gamma(t_i)}{N_t^1(t_i)}, \frac{\gamma(t_i)}{N_t^2(t_i)}\right]\right) \quad i \geq 1 \tag{4.47d}$$

$$\text{s.t.} \quad N_t^1(t_0) = 1 + \frac{\alpha\theta}{M} \sum_{i=0}^{M_*} \gamma(t_i) \quad ; \quad N_t^2(t_0) = N_t^1(t_{M_*}) \tag{4.47e}$$

where, additionally, to speed up computations, the gradients of the cost function and constraints are computed as sampled versions of the variational derivatives of the original variational calculus problem. Appendix 4.B.4 derives explicit expressions for them.

Simulations are attached in the sequel. Simulations are carried out taking the same simulation parameters as in Section 4.4.1. Firstly, the ASE of many systems is compared versus traffic load. The case of a single channel encoder practically attains the optimal ASE achieved when us-

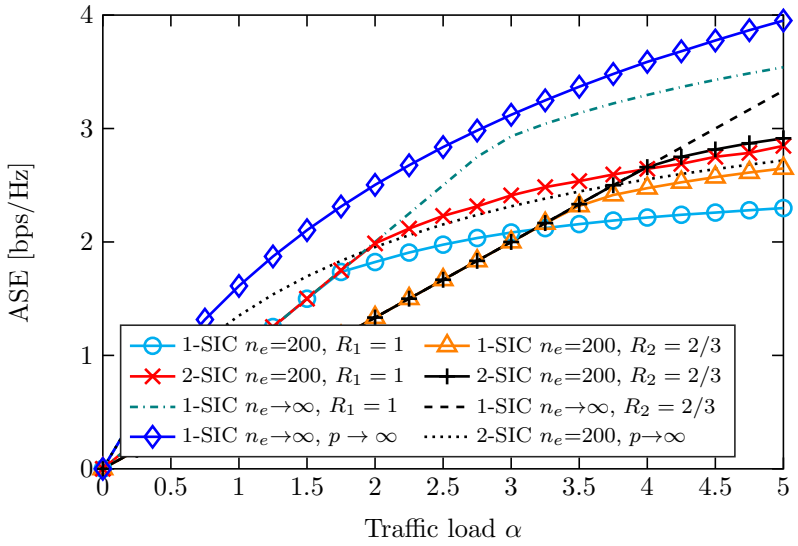

Figure 4.9: ASE versus traffic load α. The performance of the following systems is compared: (i) the IBL regime $n_e \to \infty$ with infinitely many coding systems $p \to \infty$ and with a single coding system with rate $R_1 = 1$ or $R_2 = 2/3$ bits/symbol; (ii) one-iteration (1-SIC) and two-iteration (2-SIC) SIC with a short-length encoder of $n_e = 200$ symbols; and (iii) the optimal rate allocation derived in Section 4.4.1.

ing infinitely many coding systems, $p \to \infty$, and rates adjusted according to $R(t) = \log(1+\Gamma(t))$. This advantageous result is of practical interest since the rate allocation problem can be reduced to a few encoding systems with low penalty on ASE. The interesting problem, however, is when SIC deals with short packets. As shown for packets of $n_e = 200$ symbols, the performance attained under one-iteration SIC can be substantially improved by providing a second decoding attempt to packet decoding failures in the first iteration, as in two-iteration SIC. As is evident from subsequent simulations, this is achieved primarily by allowing the transmission to more users. The latter differs substantially from the optimal behaviour under one-iteration SIC, which forces users not reaching the minimum SINR requirement to be silent. The results are also compared with the optimal rate allocation derived in Section 4.4.1. At sufficiently high traffic loads, ASE can be further increased by enabling energy allocation rather than rate allocation. The reason is that, through energy allocation, one can increase the transfer of energy, thus enabling higher ASE. This is not possible only through rate allocation since the received energy distribution remains fixed.

The energy unbalance needed at reception is studied next through the energy profiles depicted in Figure 4.10. Recall that the optimal energy allocation under one-iteration SIC results practically exponential. Two-iteration SIC allows for a substantial reduction of the unbalance in the received symbol energy distribution when all users are admitted in the system, compared with the optimal allocation under one-iteration SIC. The latter occurs primarily at low activity loads. In contrast, at high activity loads, it is preferable to practically maintain the unbalance of optimal distributions in exchange for allowing the transmission to more users (see the optimal distributions depicted under square and crossing markers).

The next analysis studies the unbalance of per-user packet error probability. Figure 4.11 shows the optimal user-PER profiles associated with the respective receivers at the low block-length $n_e = 200$ symbols. In contrast to what occurs under one-iteration SIC, the optimal

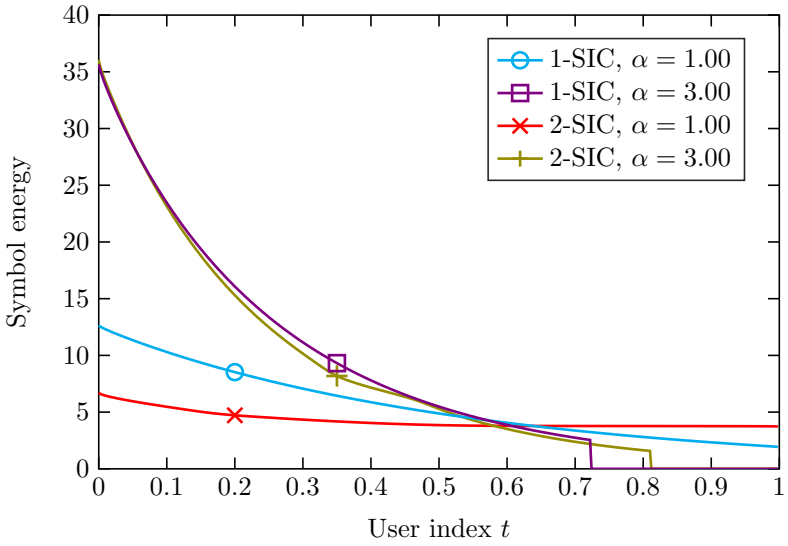

Figure 4.10: Received energy profile $\gamma(t)$ for one-iteration (1-SIC) and two-iteration (2-SIC) SIC. Simulations correspond to the theoretical encoder of $n_e = 200$ symbols and rate $R_1 = 1$ bit/symbol.

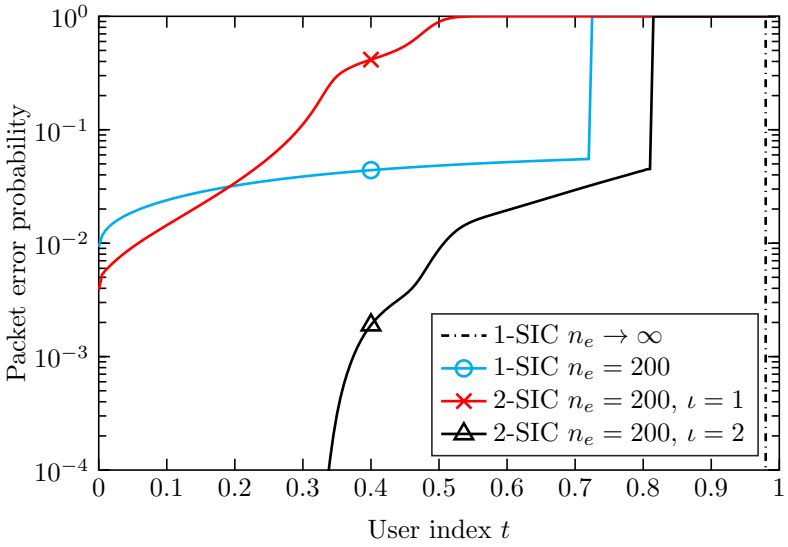

Figure 4.11: User-PER profile for one-iteration (1-SIC) and two-iteration (2-SIC) SIC. The iteration index is $\iota \in \{1, 2\}$, and the simulated traffic load is $\alpha = 3.0$.

operation under two-iteration SIC decodes most of the interference in the first iteration leaving the rest to the second iteration. Moreover, it allocates vanishing error probabilities to the strongest users, which attain in practice reliable communication. In the depicted simulation, all users experience lower PERs, relative to the optimal one-iteration SIC, thanks to two-iteration SIC. The fraction of active users is also evidenced in the same figure. As drawn, one-iteration SIC accepts approximately 70% of users, whereas two-iteration SIC enables communication to 80% of users. Remarkably, unlike one-iteration SIC, two-iteration SIC takes advantage of unbalancing user reliabilities. The strongest users are allocated to very low error rates and the weakest active users to higher PERs.

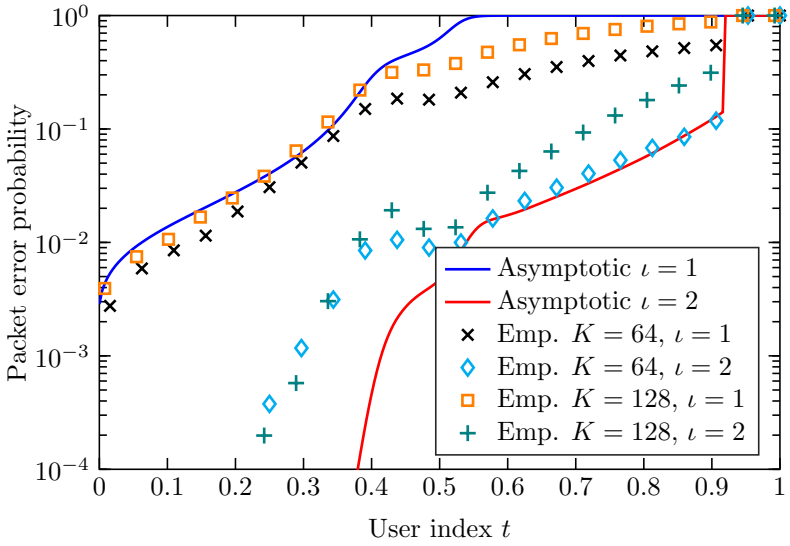

Figure 4.12: Asymptotic and empirical user-PER profile at $\alpha = 2.0$. Empirical results are evaluated under the standardised DVB-RCS turbo code [74] with QPSK modulation, 440 bits and coding rate 1/2. Empirical computations are averaged under 10^5 Monte Carlo runs.

The only study that remains still not analysed is the validation of the performance achieved by the optimal allocation design for a practical coded modulation scheme. The simulation is depicted in Figure 4.12. Specifically, we adopt the standardised turbo code from digital video broadcasting (the same coded modulation scheme as the simulation in Figure 4.4). The asymptotic computations $K, N \to \infty$ predict very well the real behaviour of the SIC system at the first iteration $\iota = 1$ for hundreds of users. The first users experience results slightly better (the solid blue line practically matches the square and the "$\times$" markers). The rest of users experience packet error probabilities even better than those predicted by the asymptotic model. Higher accuracy is obtained as the number of users increase. This behaviour is firstly, due to the computation of the asymptotic functions by means of a more conservative version of the noise plus interference term, which computes the i-th term by solving the non-linear equation $N_t^1(t_i) = N_t^1(t_{i-1}) \exp(-\frac{\alpha}{M}\Phi_1[\gamma(t_i)/N_t^1(t_i)])$, and secondly, since the better error rates experienced by first users cause less interference to other users. The matching between the asymptotic and the empirical results in the second iteration is slightly worse. As predicted, the very first users experience very low packet error probabilities (in the attached figure, they are shown to be lower than 10^{-4}). The rest of users experience error probabilities higher than those predicted by the asymptotic computations. This is due to the combination of small mismatching in the first iteration, which is propagated throughout the stages of SIC, and the fact that the number of users undergoing a second decoding attempt (in the second iteration) is random, as it depends on the effectiveness of the decoding success in the first iteration. A conclusive result is that, even with the difficulties encountered in analysing iterations beyond the first one, the proposed asymptotic model predicts the behaviour of the system well, and under a very complexity affordable method.

4.5 The Generalised Model of Iterative SIC

This section extends the above model to an iterative SIC system that performs an arbitrary number of iterations I. Recall from the previous section that, in order to proceed to subsequent iterations, at least one user must be successfully decoded and cancelled from the input signal. To generalise the above model, the following definitions are taken:

1. As already seen, the number of users remaining in the system evolves dynamically throughout SIC iterations starting from K down to zero. Without loss of generality, users are re-indexed at the beginning of each iteration ι. More specifically, the index set of users processed in the iteration ι is $\mathcal{K}_\iota \triangleq \{1,\ldots,K_\iota\}$, where for the purpose of generalisation, the number of users to be processed in the ι-th iteration is denoted K_ι ($K_1 = K$ is the initial number of users) and the user-index used in the same iteration is $k_\iota \in \mathcal{K}_\iota$. Consequently, the number of users in the system evolves iteration-by-iteration as

$$K_1 > K_2 > \cdots > K_{I-1} > K_I. \tag{4.48}$$

2. At the end of the ι-th iteration, the index set $\mathcal{K}_\iota$ can be partitioned into the disjoint subsets $\mathcal{K}_\iota^{\text{ok}}$ and $\mathcal{K}_\iota^{\text{ko}}$ respectively associated with users decoded successfully and unsuccessfully in the current iteration. Only users $k_\iota \in \mathcal{K}_\iota^{\text{ko}}$ are processed in the next iteration $\iota+1$. The relationships between user-indices of consecutive iterations ι and $\iota+1$ are

$$\phi_{\iota,\iota+1} \;:\; k_\iota \in \mathcal{K}_\iota^{\text{ko}} \longmapsto k_{\iota+1} \in \{1,\ldots,K_{\iota+1}\}, \tag{4.49}$$

$$\phi_{\iota+1,\iota} \;:\; k_{\iota+1} \in \{1,\ldots,K_{\iota+1}\} \longmapsto k_\iota \in \mathcal{K}_\iota^{\text{ko}}, \tag{4.50}$$

and they can be composed to determine the relationship between user-indices of iterations ι and ν, $\phi_{\iota,\nu} : k_\iota \longmapsto k_\nu$, as

$$\phi_{\iota,\nu>\iota} \triangleq \phi_{\nu-1,\nu} \circ \phi_{\nu-2,\nu-1} \circ \cdots \circ \phi_{\iota+1,\iota+2} \circ \phi_{\iota,\iota+1}, \tag{4.51}$$

$$\phi_{\iota,\nu<\iota} \triangleq \phi_{\nu+1,\nu} \circ \phi_{\nu+2,\nu+1} \circ \cdots \circ \phi_{\iota-1,\iota-2} \circ \phi_{\iota,\iota-1}. \tag{4.52}$$

3. The channel decoding model corresponding to the i-th coding scheme considers, at iteration ι, the packet error probability conditional on unsuccessful decoding in all preceding iterations as the multivariate function

$$\text{PER}_\iota[\Gamma^1,\ldots,\Gamma^\iota,R_i]. \tag{4.53}$$

The computation of the previous function is an extension of the algorithm described in Section 4.2 to compute $\text{PER}_2[\Gamma^1,\Gamma^2,R_i]$. The explicit computation is detailed in Appendix 4.A. Note that the proposed decoding model captures the system's memory through the sequence of non-decreasing SINRs of each user in the current ι and the previous $1 \leq j < \iota$ iterations $\Gamma^1 < \cdots < \Gamma^\iota$. This function applied to model the average number of packet decoding errors of a user k_ι with SINRs

$$\boldsymbol{\Gamma}^\iota[k_\iota] \triangleq [\ \Gamma^1[\phi_{\iota,1}[k_\iota]],\Gamma^2[\phi_{\iota,1}[k_\iota]],\ldots,\Gamma^\iota[\phi_{\iota,1}[k_\iota]]\ ] \tag{4.54}$$

considers that channel decoding does not succeed with probability $\mathrm{PER}_\iota[\mathbf{\Gamma}^\iota[k_\iota], R_i]$ (used for brevity of notation), and that succeeds with complementary probability.

The interest in this generalised model is that it allows an accurate description of the behaviour of an iterative SIC iteration-by-iteration. This characterisation was not possible before the contribution of this thesis (described in [71]). Albeit the accuracy of the system model, its difficulty is the calculation of such a multivariate function. One way to approximate its computation is by considering that $\mathrm{PER}_\iota[\mathbf{\Gamma}^\iota, R_i]$ highly depends on the SINRs associated with the latest iterations, so that multivariate PER can be approximated by the latest two SINRs:

$$\mathrm{PER}_\iota[\mathbf{\Gamma}^\iota, R_i] \approx \mathrm{PER}_2[\Gamma^{\iota-1}, \Gamma^\iota, R_i]. \tag{4.55}$$

Rigorously speaking, the accuracy of the above model applied to iterative SIC can only be guaranteed for the first two or three SIC iterations. Beyond the third iteration, the accuracy of the previous approximation highly depends on the combination of SINRs evaluated. For the purpose of generalisation, the subsequent analysis does not particularise $\mathrm{PER}_\iota[\mathbf{\Gamma}^\iota, R_i]$ until later.

The same description line as in Section 4.2 is followed to define the k_ι-th user SINR

$$\Gamma^\iota[\phi_{\iota,1}[k_\iota]] = \frac{\gamma[\phi_{\iota,1}[k_\iota]]}{N_t^\iota[\phi_{\iota,1}[k_\iota]]}, \tag{4.56}$$

with the noise plus interference term affecting user k_ι, $N_t^\iota[\phi_{\iota,1}[k_\iota]]$, computed considering Gaussian interfering signals added to thermal noise, as

$$N_t^\iota[\phi_{\iota,1}[k_\iota]] = 1 + \xi_{\iota-1} + \frac{\theta}{N}\sum_{i=1}^{k_\iota-1}\epsilon_\iota[i]\gamma[\phi_{\iota,1}[i]] + \frac{\theta}{N}\sum_{i=k_\iota+1}^{K_\iota}\gamma[\phi_{\iota,1}[i]]. \tag{4.57}$$

In the previous expression, $\xi_{\iota-1}$ is the aggregate interference from users decoded successfully and imperfectly cancelled in preceding iterations. Herein, $\epsilon_\iota[1 \leq i < k_\iota]$ are the binary random variables modelling the joint decoding-cancellation of users already processed before stage k_ι. Note that, as the SIC progresses throughout stages, more interference is removed and less interference remains in the system. Even if users increase their SINRs at each iteration, the channel decoding success depends on the SINR gain $\Gamma^\iota/\Gamma^{\iota-1}$. Remarkably, although the sequence of SINRs is non-decreasing, that of SINR gains does not necessarily increase.

4.6 The Asymptotic Generalised Model of Iterative SIC

The first consideration to enable analysis under this model is to assume an asymptotic number of users at the beginning of the first iteration as well as at the beginning of further iterations. The above model is consistent in the asymptotic large-user regime since, at each iteration, only a fraction of the dense number of users processed is decoded successfully. For a more practical case, in which the number of users is finite, the model is certainly true for the firsts few iterations as long as the system handles hundreds of users, but it is perhaps questionable for iterations beyond the firsts. Nonetheless, the adoption of this asymptotic model facilitates the understanding of iterative SIC sensitivity to system parameters.

This section takes the system model in Section 4.5 to the asymptotic large-user regime. The asymptotic system model is summarised next:

1. The sequence of asymptotic traffic loads (as $N \to \infty$) at every iteration

$$\alpha_1 \triangleq \frac{K_1}{N} \;>\; \cdots \;>\; \alpha_\iota \triangleq \frac{K_\iota}{N} \;>\; \cdots \;>\; \alpha_I \triangleq \frac{K_I}{N} \tag{4.58}$$

is defined assuming that the system processes asymptotically many users at every iteration $(K_1, \ldots, K_I) \longrightarrow \infty$. The traffic loads α_1 and α are used interchangeably. Moreover, users in every iteration $1 \leq \iota \leq I$ are indexed using the continuous indexing

$$t_\iota \triangleq \lim_{K_\iota \to \infty} \frac{k_\iota}{K_\iota} \quad \text{with} \quad 0 \leq t_\iota \leq 1. \tag{4.59}$$

In the same manner as in previous sections, the user variables indexed by k_ι are turned into asymptotic functions in the continuous variable t_ι. Considerations for a finite number of users are addressed in latter sections. Continuing under the asymptotic large-user regime, the vector of non-decreasing SINRs for a user t_ι after despreading from iteration ι down to $\iota = 1$ is (4.60), where the SINR of user t_ι at iteration ι is defined by (4.61):

$$\mathbf{\Gamma}^\iota(t) = \left[\frac{\gamma(t)}{N_t^1(t)}, \ldots, \frac{\gamma(t)}{N_t^\iota(t)}\right] \quad \text{with} \quad t = \phi_{\iota,1}(t_\iota), \tag{4.60}$$

$$\Gamma^\iota(\phi_{\iota,1}(t_\iota)) = \frac{\gamma(\phi_{\iota,1}(t_\iota))}{N_t^\iota(\phi_{\iota,1}(t_\iota))}. \tag{4.61}$$

In the above expressions, $\phi_{\iota,1}(t_\iota)$ computes the index of user t_ι of the ι-th iteration, in the first iteration. The indices t_1 and t are used interchangeably. Equivalent expressions for the noise plus interference profile $N_t^\iota(\phi_{\iota,1}(t_\iota))$ are discussed in the fourth point.

2. The asymptotic average number of users unsuccessfully decoded in iteration ι is

$$\overline{\text{per}}_\iota = \int_0^1 \text{PER}_\iota[\mathbf{\Gamma}^\iota(\phi_{\iota,1}(\tau)), R(\phi_{\iota,1}(\tau))]\text{d}\tau, \tag{4.62}$$

for which the asymptotic traffic loads of consecutive iterations ι and $\iota - 1$ are computed as $\alpha_\iota = \alpha_{\iota-1}\overline{\text{per}}_{\iota-1}$, or as a function of the initial traffic load as $\alpha_\iota = \alpha_1 \prod_{k=1}^{\iota-1} \overline{\text{per}}_k$.

3. In this regime, it is possible to derive a very interesting result for the asymptotic relationship between user indices of iteration ι and the first iteration $\iota = 1$. To that aim, the proof associated with $\phi_{1,2}$ in Section 4.3.2 is taken as a reference. In the case studied herein, after the K_ι stages, each user interval $\text{d}t_\iota$ can be decomposed as [4]

$$\text{d}t_\iota = \text{PER}_\iota[\mathbf{\Gamma}^\iota(\phi_{\iota,1}(t_\iota)), R(\phi_{\iota,1}(t_\iota))]\text{d}t_\iota + \text{PSR}_\iota[\mathbf{\Gamma}^\iota(\phi_{\iota,1}(t_\iota)), R(\phi_{\iota,1}(t_\iota))]\text{d}t_\iota. \tag{4.63}$$

The next iteration $\iota + 1$ deals with all the intervals $\text{PER}_\iota[\mathbf{\Gamma}^\iota(\phi_{\iota,1}(t_\iota)), R(\phi_{\iota,1}(t_\iota))]\text{d}t_\iota$. The

[4] Recall that $R(t)$ shall be a smooth, or failing that, a piecewise smooth function.

following differential equation is formulated and solved

$$\dot{\phi}_{\iota,\iota+1}(t_\iota) = \frac{1}{\overline{\text{per}}_\iota}\ \text{PER}_\iota[\mathbf{\Gamma}^\iota(\phi_{\iota,1}(t_\iota)), R(\phi_{\iota,1}(t_\iota))], \tag{4.64}$$

$$\phi_{\iota,\iota+1}(t_\iota) = \frac{1}{\overline{\text{per}}_\iota}\ \int_0^{t_\iota} \text{PER}_\iota[\mathbf{\Gamma}^\iota(\phi_{\iota,1}(\tau)), R(\phi_{\iota,1}(\tau))]\text{d}\tau. \tag{4.65}$$

Finally, making use of the relationships (4.51)–(4.52) the following relationships between user-indices of the first iteration and the iteration $\iota + 1$ are obtained:

$$\dot{\phi}_{1,\iota+1}(t) = \frac{1}{\Pi_{k=1}^{\iota}\overline{\text{per}}_k}\ \prod_{k=1}^{\iota} \text{PER}_k[\mathbf{\Gamma}^k(t), R(t)], \tag{4.66}$$

$$\phi_{1,\iota+1}(t) = \frac{1}{\Pi_{k=1}^{\iota}\overline{\text{per}}_k}\ \int_0^{t} \prod_{k=1}^{\iota} \text{PER}_k[\mathbf{\Gamma}^k(\tau), R(\tau)]\text{d}\tau. \tag{4.67}$$

4. The noise plus interference term for a user t_ι ($t = \phi_{\iota,1}(t_\iota)$) is computed as

$$N_t^\iota(t) = 1 + \xi_{\iota-1} + \alpha_\iota\theta\int_0^t q_\iota\left[\mathbf{\Gamma}^\iota(\tau), R(\tau)\right]\dot{\phi}_{1,\iota}(\tau)\gamma(\tau)\text{d}\tau + \alpha_\iota\theta\int_t^1 \dot{\phi}_{1,\iota}(\tau)\gamma(\tau)\text{d}\tau. \tag{4.68}$$

with $q_\iota[\mathbf{\Gamma}^\iota, R_i] \triangleq 1 - (1 - \varepsilon[\Gamma^\iota, R_i])\text{PSR}_\iota[\mathbf{\Gamma}^\iota, R_i]$.

It should be noted that the above model requires the multivariate function $\text{PER}_\iota[\Gamma^1, \ldots, \Gamma^\iota, R_i]$ for each iteration $1 \leq \iota \leq I$ and each coding scheme. One way to simplify the above model while accepting some reduction in accuracy, is, as discussed above, to adopt the following model for the multivariate function

$$\text{PER}_\iota[\mathbf{\Gamma}^\iota, R_i] \approx \text{PER}_2[\Gamma^{\iota-1}, \Gamma^\iota, R_i] \approx \frac{\text{PER}[\Gamma^\iota, R_i]}{\text{PER}[\Gamma^{\iota-1}, R_i]}. \tag{4.69}$$

In this case, the noise plus interference level can be calculated in the form of a differential or integral equation through the expressions

$$\frac{\dot{N}_t^\iota(t)}{N_t^\iota(t)} = -\alpha_1\Phi_\iota\left[\frac{\gamma(t)}{N_t^\iota(t)}, \frac{\gamma(t)}{N_t^\iota(t)}, R(t)\right], \tag{4.70}$$

$$N_t^\iota(t) = N_t^\iota(0)\exp\left(-\alpha_1\int_0^t \Phi_\iota\left[\frac{\gamma(\tau)}{N_t^{\iota-1}(\tau)}, \frac{\gamma(\tau)}{N_t^\iota(\tau)}, R(\tau)\right]\text{d}\tau\right). \tag{4.71}$$

with $\Phi_\iota[\Gamma^{\iota-1}, \Gamma^\iota, R_i] \triangleq \theta(1-\varepsilon[\Gamma^\iota, R_i])\Gamma^\iota(\text{PER}[\Gamma^{\iota-1}, R_i] - \text{PER}[\Gamma^\iota, R_i])$, and the initial noise plus interference term $N_t^\iota(0) = N_t^{\iota-1}(1)$.

The average packet error rate after ι iterations is computed as

$$\overline{\text{per}} = \prod_{k=1}^{\iota}\overline{\text{per}}_k \tag{4.72a}$$

$$= \prod_{k=1}^{\iota}\int_0^1 \text{PER}_k[\mathbf{\Gamma}^k(\phi_{k,1}(\tau)), R(\phi_{k,1}(\tau))]\text{d}\tau, \tag{4.72b}$$

after which, by applying the approximation (4.69) and using the expressions derived in point 3

it results in

$$\overline{\text{per}} \approx \int_0^1 \text{PER}[\Gamma^\iota(t), R(t)]\text{d}t. \tag{4.73}$$

As mentioned before, the above asymptotic model allows to evaluate the asymptotic performance of an iterative canceller in a very convenient way. For a practical case, where the system handles K users, the model predicts accurately the behaviour of the real system at the first few iterations. For further iterations, it does not constitute a strict match, since the initial assumption of having an asymptotically large number of users to process in each iteration fails to be an accurate assumption. To the author's best knowledge, the accurate matching of this model to a practical case needs a halting policy that stops the operation of iterative SIC when some criterion that relates the number of users K with $N_t^\iota(0) - N_t^\iota(1)$ is met.

In any case, the above model relies on the assumption that the receiver operates under perfect packet error detection, which, for practical purposes, is not always true. The latter is only possible if users encapsulate a sufficiently long CRC to underestimate packet detection errors. Throughout the thesis, this problem has been circumvented by assuming perfect packet error detection. The impact of finite CRC on system performance remains to be analysed.

4.6.1 A Fixed Point Equation for Iterative SIC

The investigation in this chapter is continued in the asymptotic large-user regime so as to evidence the behaviour of iterative SIC for over two iterations. In this regime, it is always possible to continue the analysis pursued above for many iterations since: firstly, at each iteration a fraction of non-zero users is decoded successfully; and secondly, at the end of each iteration a fraction of non-zero users is decoded unsuccessfully, and thus, left to be decoded.

Roughly speaking, one can understand the iterative SIC system in the asymptotic large-user regime as a system that, at each iteration, operates with an input $N_t^\iota(0)$ and that outputs $N_t^\iota(1)$. The specific input-output relationship is given by the evolution of the noise plus interference level at each stage t in (4.71). This interpretation allows to analyse the mitigation of the noise plus interference term throughout iterations by means of a fixed-point equation. At each iteration, the noise plus interference term does not increase, and the system converges (does not cancel substantially inter-user interference) when the input and output magnitudes are the same. For the previous scenario where all users use the same channel encoder, we perform a heuristic optimisation by superimposing an exponential packet-power distribution at the receiver's input, in order to evidence the noise plus interference mitigation over the iterations. In Figure 4.13 we plot the initial and final noise plus interference levels for a SIC receiver of I iterations. More concretely, the optimisations for receivers with 3, 5 and 10 iterations are compared. The vertical lines correspond to the mitigation of the noise level plus interference. From top to bottom, the initial and final levels are depicted. As can be seen, when the receiver performs few iterations, it is beneficial to concentrate the highest cancellation magnitude in the first SIC receiver stages whereas when the receiver performs a higher number of iterations, it is better to split the SIC cancellation power in the different iterations.

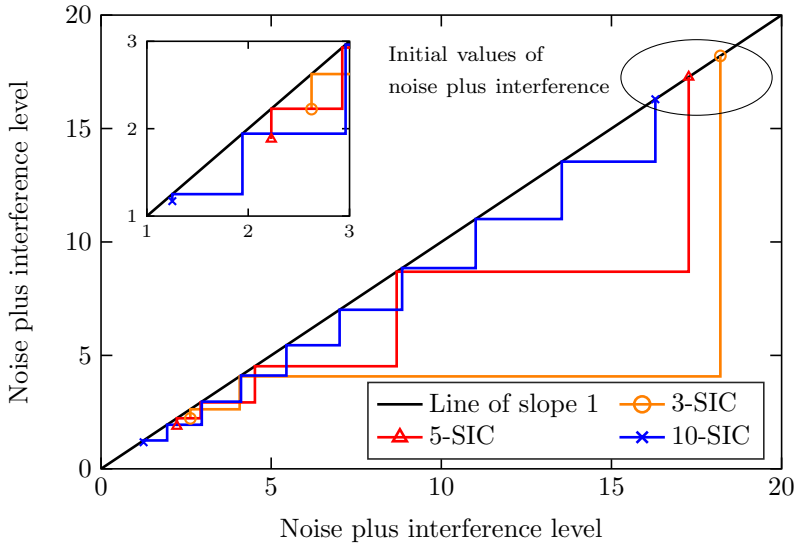

Figure 4.13: Evolution of the noise plus interference level throughout iterations of the SIC receiver. Simulations correspond to a theoretical encoder with $n_e = 200$ symbols and $R = 1$ bit/symbol. The traffic load is $\alpha = 2.4$.

4.7 Concluding Remarks

This chapter has dealt with an important feature of the demodulator adopted in the Enhanced Spread Spectrum ALOHA system: the iterative successive decoding policy that iterates the multiuser decoding strategy over users decoded unsuccessfully. The above is employed to overcome the nonideal decoding behaviour of coded modulation systems for short packets. The model for this receiver is challenging due to two important facts: firstly, the receiver handles a lesser number of users at each iteration; and secondly, the decoding operations operate with memory with respect to previous decoding attempts for the same user. All this focuses the investigation in this chapter to model the noise plus interference level of iterative SIC in the infinite-user regime as a deterministic function of previous packet success/failure decoding errors. The novelty introduced in this chapter is twofold; firstly, the usage of invertible mappings to relate users in the different iterations of SIC; and secondly, the modelling of the magnitude of packet error failures in the second iteration using a multivariate packet error rate function of the signal-to-interference-plus-noise ratios experienced by a user in all its processed stages.

The energy and rate allocation functions are designed leveraging the attractive form that the above system model shows in the infinite-user regime. The optimisation under two-iteration SIC concludes that the optimal allocations substantially outperform the optimisations under one-iteration SIC when the packet error rate characteristic curves of employed coding systems are smoother, and users are not constrained by fair reliability. The optimal allocations are also discontinuous in the user decoding variable, and enable the transmission to a high number of users. Later simulations reveal that, using an heuristic criteria with many decoding iterations, iterative SIC can practically overcome the sum-rate loss of SIC due to the use of finite-length coding as long as the number of users is sufficiently high.

Appendix 4.A Computation of the Multivariate PER

The computation of the multivariate PER functions associated with the i-th coding scheme for an iterative SIC system constituted by I iterations can be generalised from the algorithm in Section 4.2: firstly, generate vector $\mathbf{b}$ and encode it as $\sqrt{\gamma}e^{j\Omega}f_i(\mathbf{b})$; secondly, subject communication to the statistically independent Gaussian noises $\mathbf{w}^1,\ldots,\mathbf{w}^I \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$; thirdly, generate the received signals

$$\mathbf{y}^k = \sqrt{\gamma}e^{j\Omega}f_i(\mathbf{b}) + \sqrt{N_t^I}\ \mathbf{w}^I + \sum_{\iota=k}^{I-1}\sqrt{N_t^\iota - N_t^{\iota+1}}\ \mathbf{w}^\iota \quad \text{for} \quad k=1,\ldots,I \tag{4.74}$$

with $N_t^1 \geq N_t^2 \geq \cdots \geq N_t^I$ and $\Gamma^\iota = \gamma/N_t^\iota$; and finally, compute the multivariate PER at each iteration $1 \leq \iota \leq I$ as

$$\mathrm{PER}[\Gamma^1, R_i] = \Pr\left[f_i^{-1}(\mathbf{y}^1) \neq \mathbf{b}\right], \tag{4.75}$$

$$\mathrm{PER}_\iota[\Gamma^1,\ldots,\Gamma^\iota, R_i] = \Pr\left[f_i^{-1}(\mathbf{y}^\iota) \neq \mathbf{b} \mid f_i^{-1}(\mathbf{y}^1) \neq \mathbf{b},\ldots,f_i^{-1}(\mathbf{y}^{\iota-1}) \neq \mathbf{b}\right]. \tag{4.76}$$

Appendix 4.B Proofs

Appendix 4.B.1 Noise Plus Interference Profile of Two-Iteration SIC

This appendix derives two equivalent expressions for the noise plus interference term affecting user t_2 given by

$$\begin{aligned} N_t^2(\phi_{2,1}(t_2)) = 1 + \xi_1 + \alpha_2\theta\int_0^{t_2} q_2\Big[\Gamma^1(\phi_{2,1}(\tau)), \Gamma^2(\phi_{2,1}(\tau)), R(\phi_{2,1}(\tau))\Big]\gamma(\phi_{2,1}(\tau))\mathrm{d}\tau \\ + \alpha_2\theta\int_{t_2}^{1}\gamma(\phi_{2,1}(\tau))\mathrm{d}\tau \end{aligned} \tag{4.77}$$

with $q_2[\Gamma^1, \Gamma^2, R] = 1 - (1 - \varepsilon[\Gamma^2, R])(1 - \mathrm{PER}_2[\Gamma^1, \Gamma^2, R])$. The next steps are followed:

1. The change of variable $u = \phi_{2,1}(\tau)$ is applied to both integrals. The differential is

$$\mathrm{d}\tau = \dot{\phi}_{1,2}(u)\mathrm{d}u = \frac{\mathrm{PER}[\Gamma^1(u), R(u)]}{\overline{\mathrm{per}}_1}\,\mathrm{d}u. \tag{4.78}$$

   In the first integral, the limits are $\phi_{2,1}(0) = 0$ and $\phi_{2,1}(t_2)$. In the second integral, the limits are $\phi_{2,1}(t_2)$ and $\phi_{2,1}(1) = 1$. The result gives

$$\begin{aligned} N_t^2(\phi_{2,1}(t_2)) = 1 + \xi_1 + \alpha_2\theta\int_0^{\phi_{2,1}(t_2)} q_2\Big[\Gamma^1(u), \Gamma^2(u), R(u)\Big]\gamma(u)\mathrm{d}\phi_{1,2}(u) \\ + \alpha_2\theta\int_{\phi_{2,1}(t_2)}^{1}\gamma(u)\mathrm{d}\phi_{1,2}(u). \end{aligned} \tag{4.79}$$

2. The noise plus interference term $N_t^2(t)$ can be defined substituting argument $\phi_{2,1}(t)$ by t, expanding $\mathrm{d}\phi_{2,1}(t)$, and substituting $\alpha_2 = \alpha\overline{\mathrm{per}}_1$, as

$$\begin{aligned} N_t^2(t) = 1 + \xi_1 + \alpha\theta\int_0^{t} q_2\Big[\Gamma^1(u), \Gamma^2(u), R(u)\Big]\mathrm{PER}[\Gamma^1(u), R(u)]\gamma(u)\mathrm{d}u \\ + \alpha\theta\int_{t}^{1}\gamma(u)\mathrm{PER}[\Gamma^1(u), R(u)]\mathrm{d}u. \end{aligned} \tag{4.80}$$

3. Differentiating $N_t^2(t)$ with respect to variable t and arranging terms using $\Phi_2[\Gamma^1, \Gamma^2, R] \triangleq \theta(1 - \varepsilon[\Gamma^2, R])\Gamma^2(1 - \mathrm{PER}_2[\Gamma^1, \Gamma^2, R])\mathrm{PER}[\Gamma^1, R]$, the following differential equation and its equivalent integral form is obtained:

$$\frac{\dot{N}_t^2(t)}{N_t^2(t)} = -\alpha\Phi_2\left[\frac{\gamma(t)}{N_t^1(t)}, \frac{\gamma(t)}{N_t^2(t)}, R(t)\right], \tag{4.81}$$

$$N_t^2(t) = N_t^2(0)\exp\left(-\alpha\int_0^t \Phi_2\left[\frac{\gamma(\tau)}{N_t^1(\tau)}, \frac{\gamma(\tau)}{N_t^2(\tau)}, R(\tau)\right]\mathrm{d}\tau\right) \tag{4.82}$$

   with the initial noise plus interference term $N_t^2(0)$ equal to the noise plus interference term left at the last step of the first iteration

$$N_t^2(0) = N_t^1(1) = N_t(0)\exp\left(-\alpha\int_0^t \Phi_1\left[\frac{\gamma(\tau)}{N_t^1(\tau)}, R(\tau)\right]\mathrm{d}\tau\right). \tag{4.83}$$

Appendix 4.B.2 Closed-Form Expression for Asymptotic Spectral Efficiency

This appendix derives a closed-form expression for the ASE after two SIC iterations

$$\mathrm{ASE} = \mathrm{ASE}_1 + \mathrm{ASE}_2. \tag{4.84}$$

ASE_1 and ASE_2 denote the ASEs achieved in each iteration, defined by

$$\mathrm{ASE}_1 = \alpha \int_0^1 R(t)\ \mathrm{PSR}[\Gamma^1(t), R(t)]\mathrm{d}t, \tag{4.85a}$$

$$\mathrm{ASE}_2 = \alpha_2 \int_0^1 R(\phi_{2,1}(t))\ \mathrm{PSR}_2[\Gamma^1(\phi_{2,1}(t)), \Gamma^2(\phi_{2,1}(t)), R(\phi_{2,1}(t))]\mathrm{d}t. \tag{4.85b}$$

The change of variable $u \triangleq \phi_{2,1}(t)$ is applied on ASE_2. The differential is

$$\mathrm{d}\tau = \dot{\phi}_{1,2}(u)\mathrm{d}u = \frac{\mathrm{PER}[\Gamma^1(u), R(u)]}{\overline{\mathrm{per}}_1}\ \mathrm{d}u, \tag{4.86}$$

and the integral limits are $\phi_{2,1}(0) = 0$ and $\phi_{2,1}(1) = 1$. This gives

$$\mathrm{ASE}_2 = \alpha \int_0^1 R(u)\ \mathrm{PSR}_2[\Gamma^1(u), \Gamma^2(u), R(u)]\mathrm{PER}[\Gamma^1(u), R(u)]\mathrm{d}u, \tag{4.87}$$

where we have used that $\alpha_2 = \alpha\overline{\mathrm{per}}_1$. Now, making use of $\mathrm{PSR}_2[\Gamma^1, \Gamma^2, R_i] = 1 - \mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i]$ and using the upper-bound $\mathrm{PER}_2[\Gamma^1, \Gamma^2, R_i] \leq \mathrm{PER}[\Gamma^2, R]/\mathrm{PER}[\Gamma^1, R_i]$ with strict equality, the following is obtained after some straightforward manipulations

$$\mathrm{ASE}_2 = \alpha \int_0^1 R(u)\ \mathrm{PSR}[\Gamma^2(u), R(u)]\mathrm{d}u - \alpha \int_0^1 R(u)\ \mathrm{PSR}[\Gamma^1(u), R(u)]\mathrm{d}u. \tag{4.88}$$

Finally,

$$\mathrm{ASE} = \mathrm{ASE}_1 + \mathrm{ASE}_2 = \alpha \int_0^1 R(u)\ \mathrm{PSR}[\Gamma^2(u), R(u)]\mathrm{d}u. \tag{4.89}$$

Appendix 4.B.3 Optimal Rate Allocation with Fair Transmitted Power

This appendix derives the stationary point equations of the rate allocation problem under two-iteration SIC. The functional to be maximised under $R(t)$ is

$$\alpha \int_0^1 R(t)\ \mathrm{PSR}\left[\frac{\gamma(t)}{N_t^2(t)}, R(t)\right]\mathrm{d}t, \tag{4.90}$$

with $\gamma(t) = \bar{\gamma}_{\mathrm{x}} h(t)$ a given smooth function, and the noise plus interference $N_t^2(t)$ computed as

$$\frac{\dot{N}_t^2(t)}{N_t^2(t)} = -\alpha\Phi_2\left[\frac{\gamma(t)}{N_t^1(t)}, \frac{\gamma(t)}{N_t^2(t)}, R(t)\right] \quad , \quad \frac{\dot{N}_t^1(t)}{N_t^1(t)} = -\alpha\Phi_1\left[\frac{\gamma(t)}{N_t^1(t)}, R(t)\right] \tag{4.91}$$

under the initial terms $N_t^1(0) = 1 + \alpha\theta\int_0^1\gamma(t)\mathrm{d}t$ and $N_t^2(0) = N_t^1(1)$. The Lagrangian is

$$\begin{aligned}\mathcal{L} = \alpha\int_0^1 R(t)\ \mathrm{PSR}\left[\frac{\gamma(t)}{N_t^2(t)}, R(t)\right]\mathrm{d}t - \int_0^1\beta_1(t)\left(\frac{\dot{N}_t^1(t)}{N_t^1(t)} + \alpha\Phi_1\left[\frac{\gamma(t)}{N_t^1(t)}, R(t)\right]\right)\mathrm{d}t \\ - \int_0^1\beta_2(t)\left(\frac{\dot{N}_t^2(t)}{N_t^2(t)} + \alpha\Phi_2\left[\frac{\gamma(t)}{N_t^1(t)}, \frac{\gamma(t)}{N_t^2(t)}, R(t)\right]\right)\mathrm{d}t.\end{aligned} \tag{4.92}$$

From this point onwards, the explicit arguments of functions PSR Φ_1, and Φ_2 are obviated. The space of smooth functions with continuous first derivative $R(t) \in \mathcal{C}[0,1]$ is adopted. The stationary point equation is found by considering variations $R(t) + \delta R(t)$ and by computing the first variation of the Lagrangian, with respect to the variation $\delta R(t)$, as

$$\delta\mathcal{L} = \alpha\int_0^1 \left(R(t)\ \mathrm{PSR}_R[\cdot] + \mathrm{PSR}[\cdot] - \beta_1(t)(\Phi_1)_R[\cdot] - \beta_2(t)(\Phi_2)_R[\cdot]\right)\delta R(t)\mathrm{d}t. \tag{4.93}$$

The same result is obtained if $R(t)$ does not vary at $t = 0$ and $t = 1$: $\delta R(0) = \delta R(1) = 0$. Then, using the Fundamental Lemma of the Calculus of Variations the stationary point solution satisfies

$$R(t)\ \mathrm{PSR}_R[\cdot] + \mathrm{PSR}[\cdot] - \beta_1(t)(\Phi_1)_R[\cdot] - \beta_2(t)(\Phi_2)_R[\cdot] = 0 \quad \text{in} \quad 0 \le t \le 1. \tag{4.94}$$

An identical approach can be followed to determine the corresponding $N_t^1(t)$ and $N_t^2(t)$. To avoid a lengthy mathematical development, a simplified approach is described next:

1. With regard to $N_t^2(t)$, by applying the Euler-Lagrange operator $(\partial_{N_t^2} - \nabla_t\partial_{\dot{N}_t^2})$ over the integrand of $\mathcal{L}$. The following differential equation is obtained for $\beta_2(t)$

$$\dot{\beta}_2(t) = \alpha\frac{\gamma(t)}{N_t^2(t)}\left(R(t)\mathrm{PSR}_{\Gamma^2}[\cdot] - \beta_2(t)(\Phi_2)_{\Gamma^2}[\cdot]\right). \tag{4.95}$$

2. With regard to $N_t^1(t)$, by applying the Euler-Lagrange operator $(\partial_{N_t^1} - \nabla_t\partial_{\dot{N}_t^1})$ over the integrand of $\mathcal{L}$. The following differential equation in $\beta_1(t)$ is obtained

$$\dot{\beta}_1(t) = -\alpha\frac{\gamma(t)}{N_t^1(t)}\left(\beta_1(t)(\Phi_1)_{\Gamma^1}[\cdot] + \beta_2(t)(\Phi_2)_{\Gamma^1}[\cdot]\right). \tag{4.96}$$

3. The initial values for the differential equations are $\beta_2(1) = 0$ and $\beta_1(1) = \beta_2(0)$.

Appendix 4.B.4 Optimal Energy Allocation with Fair Encoding Rate

This appendix derives the stationary point equations of the energy allocation problem under two-iteration SIC. The functional to be maximised under $\gamma_{\mathrm{x}}(t)$ is

$$\alpha R\int_0^{t_*}\mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t^2(t)}\right]\mathrm{d}t - \lambda\left(\int_0^{t_*}\gamma_{\mathrm{x}}(t)\mathrm{d}t - \bar{\gamma}_{\mathrm{x}}\right), \tag{4.97}$$

with the noise plus interference term $N_t^2(t)$ computed as

$$\frac{\dot{N}_t^2(t)}{N_t^2(t)} = -\alpha\Phi_2\left[\frac{\gamma(t)}{N_t^1(t)}, \frac{\gamma(t)}{N_t^2(t)}, R(t)\right] \quad , \quad N_t^2(0) = N_t^1(t_*) \tag{4.98}$$

$$\frac{\dot{N}_t^1(t)}{N_t^1(t)} = -\alpha\Phi_1\left[\frac{\gamma(t)}{N_t^1(t)}, R(t)\right] \quad , \quad N_t^1(0) = 1 + \alpha\theta\int_0^{t_*}\gamma_{\mathrm{x}}(t)h(t)\mathrm{d}t. \tag{4.99}$$

The Lagrangian is

$$\mathcal{L} = \int_0^{t_*}\mathrm{PSR}\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t^2(t)}\right]\mathrm{d}t - \lambda\left(\int_0^{t_*}\gamma_{\mathrm{x}}(t)\mathrm{d}t - \bar{\gamma}_{\mathrm{x}}\right) \tag{4.100a}$$

$$-\int_0^{t_*}\beta_1(t)\left(\frac{\dot{N}_t^1(t)}{N_t^1(t)} + \alpha\Phi_1\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t^1(t)}\right]\right)\mathrm{d}t \tag{4.100b}$$

$$-\int_0^{t_*}\beta_2(t)\left(\frac{\dot{N}_t^2(t)}{N_t^2(t)} + \alpha\Phi_2\left[\frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t^1(t)}, \frac{\gamma_{\mathrm{x}}(t)h(t)}{N_t^2(t)}\right]\right)\mathrm{d}t. \tag{4.100c}$$

Following the rationales in Appendix 4.B.3, the first variation reads

$$\delta\mathcal{L} = \int_0^{t_*}\left(\partial_\gamma\mathcal{L} - \nabla_t\partial_{\dot{\gamma}}\mathcal{L}\right)v_\gamma(t)\mathrm{d}t + \left(\partial_{\dot{\gamma}}\mathcal{L}\right)\Big|_{t=t_*}v_\gamma(t_*) - \left(\partial_{\dot{\gamma}}\mathcal{L}\right)\Big|_{t=0}v_\gamma(0) \tag{4.101a}$$

$$+\int_0^{t_*}\left(\partial_{N_t^1}\mathcal{L} - \nabla_t\partial_{\dot{N}_t^1}\mathcal{L}\right)\delta N_t^1(t)\mathrm{d}t + \left(\partial_{\dot{N}_t^1}\mathcal{L}\right)\Big|_{t=t_*}\delta N_t^1(t_*) - \left(\partial_{\dot{N}_t^1}\mathcal{L}\right)\Big|_{t=0}\delta N_t^1(0) \tag{4.101b}$$

$$+\int_0^{t_*}\left(\partial_{N_t^2}\mathcal{L} - \nabla_t\partial_{\dot{N}_t^2}\mathcal{L}\right)\delta N_t^2(t)\mathrm{d}t + \left(\partial_{\dot{N}_t^2}\mathcal{L}\right)\Big|_{t=t_*}\delta N_t^2(t_*) - \left(\partial_{\dot{N}_t^2}\mathcal{L}\right)\Big|_{t=0}\delta N_t^2(0). \tag{4.101c}$$

In our case, we have $\partial_{\dot{\gamma}}\mathcal{L} = 0$, $\partial_{\dot{N}_t^1}\mathcal{L} = -\frac{\beta_1(t)}{N_t^1(t)}$, $\delta N_t^1(0) = \alpha\theta\int_0^{t_*}h(t)v_\gamma(t)\mathrm{d}t$, $\partial_{\dot{N}_t^2}\mathcal{L} = -\frac{\beta_2(t)}{N_t^2(t)}$, and $\delta N_t^2(0) = \delta N_t^1(t_*)$. Then, the above yields

$$\delta\mathcal{L} = \int_0^{t_*}\left(\partial_\gamma\mathcal{L} - \nabla_t\partial_{\dot{\gamma}}\mathcal{L} + \alpha\theta\frac{\beta_1(0)}{N_t^1(0)}h(t)\right)v_\gamma(t)\mathrm{d}t \tag{4.102a}$$

$$+\int_0^{t_*}\left(\partial_{N_t^1}\mathcal{L} - \nabla_t\partial_{\dot{N}_t^1}\mathcal{L}\right)\delta N_t^1(t)\mathrm{d}t + \left(\partial_{\dot{N}_t^1}\mathcal{L}\right)\Big|_{t=t_*}\delta N_t^1(t_*) - \left(\partial_{\dot{N}_t^2}\mathcal{L}\right)\Big|_{t=0}\delta N_t^1(t_*) \tag{4.102b}$$

$$+\int_0^{t_*}\left(\partial_{N_t^2}\mathcal{L} - \nabla_t\partial_{\dot{N}_t^2}\mathcal{L}\right)\delta N_t^2(t)\mathrm{d}t + \left(\partial_{\dot{N}_t^2}\mathcal{L}\right)\Big|_{t=t_*}\delta N_t^2(t_*). \tag{4.102c}$$

We obtain the following stationary point equation in $0 \le t \le t_*$:

$$\mathrm{PSR}_\Gamma[\cdot]\frac{h(t)}{N_t^2(t)} + \alpha\theta\frac{\beta_1(0)}{N_t^1(0)}h(t) = \tag{4.103a}$$

$$\alpha\beta_1(t)(\Phi_1)_\Gamma[\cdot]\frac{h(t)}{N_t^2(t)} + \alpha\beta_2(t)\left[(\Phi_2)_{\Gamma^1}[\cdot]\frac{h(t)}{N_t^1(t)} + (\Phi_2)_{\Gamma^2}[\cdot]\frac{h(t)}{N_t^2(t)}\right] + \lambda \tag{4.103b}$$

with the functions $\beta_1(t)$ and $\beta_2(t)$ computed as

$$\dot{\beta}_1(t) = \alpha\beta_1(t)(\Phi_1)_\Gamma[\cdot]\Gamma^1(t) + \alpha\beta_2(t)(\Phi_2)_{\Gamma^1}[\cdot]\Gamma^1(t) \tag{4.104}$$

$$\dot{\beta}_2(t) = \alpha\beta_2(t)(\Phi_2)_{\Gamma^2}[\cdot]\Gamma^2(t) - \mathrm{PSR}_\Gamma[\cdot]\Gamma^2(t) \tag{4.105}$$

with the boundaries $\beta_1(t_*) = \beta_2(0)$ and $\beta_2(0) = \beta_1(1)$.

5 The Case of Dynamically Ordered Interference Cancellation Decoding

The energy and code allocations derived in the previous chapters are based on the assumption that the receiver knows the decoding order of users, prior to performing successive interference cancellation (SIC). An example of such a scenario is when the receiver perfectly knows the strengths received from all users or, failing that, when it can estimate them accurately. In these cases, SIC can operate by decoding the best user at each stage. The present chapter, however, analyses the case when the strengths (symbol energies or powers) received from all users are unknown to the receiver. Thus, received symbol energies are first estimated, after which the SIC receiver proceeds in non-increasing order of the former estimates. We refer to this receiver, as in some context-aware settings, as a dynamically ordered SIC. This chapter presents an original analysis to one of the aspects that, relative to the demodulator adopted in the Enhanced Spread Spectrum ALOHA system outlined in Section 1.2.2, remains to be analysed: the impact of estimation-based decode ordering on system performance. As described afterwards, the scenarios of interest may correspond to wireless setting where the large number of users challenges the receiver's capability to obtain a perfect estimation of the strengths received from all users, or to cases in which the identities of users are unknown to the receiver.

The main results derived in this chapter are condensed in the following publications:

[**L2**] **F. Molina** and J. Sala-Álvarez, "Discontinuous user-energy distribution for dynamically ordered successive interference cancellation," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1673-1677, 2021,

[**L3**] **F. Molina** and J. Sala-Álvarez, "Asymptotic performance analysis of successive interference cancellation with dynamic user-decoding order," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2931-2935, 2020.

The remaining of this chapter starts in Section 5.1 with the state of the art relative to dynamically ordered SIC. The system models for many users and asymptotically many users [77] are presented, respectively, in Section 5.2 and Section 5.3. The best energy allocation strategy, derived in the asymptotic large-user regime [78], is addressed in Section 5.4. Finally, conclusions are offered in Section 5.5. Appendix 5.A contains the mathematical derivations corresponding to the present chapter.

5.1 State of the Art

The previous chapters of this thesis have evaluated and optimised the performance of a network whose receiver operates under iterative or non-iterative SIC policies, considering that the decoding order of users is known to the receiver. With regard to the latter consideration, the analyses undertaken previously were not exceptions. Most theoretical studies have adopted such an argument as it constitutes one of the theoretical foundations of SIC [43, 79]. However, down to more practical settings, more recent analyses under the umbrella of non-orthogonal multiple access (NOMA) have alleviated the need of agreeing on a decoding order between transmitting users and the receiver station.

To go deep into this matter, let us put first our analysis into the context of machine-type communications. In the uplink of a massive NOMA system, it is common that large-scale fading coefficients are partially known, or totally unknown to the receiver. The reason is largely due to the dense dissemination of users in certain areas and their low activity periods, which challenges the task of the receiver to obtain the channel state information of all users. In these cases, as users may transmit with different powers and be affected by dissimilar channel attenuations, it is a common practice to consider, for system model analysis, users subject to (statistically) independent non-identically distributed large-scale fading [80]. Mainly, two approaches have been followed to set a user-decoding order before proceeding to successive decoding. The first approach consists in establishing a decoding order based on a deterministic parameter, which allows to apply the already known results to our study case; for instance, when the receiver knows the expectation of the large-scale fading coefficients affecting all users [80, Section III.B]. A second approach consists in setting the decoding order based on the instantaneous power levels received from all users, in which case one can consider that the received powers are obtained accurately [47] [80, Section III.A], or need to be estimated. Analyses of cancellation receivers that follow the second approach, also referred to as dynamically ordered SIC, have been, admittedly, very few [81–83]. All of them derive outage expressions for two or three users by calculating all possible combinations of orderings. However, these analyses may be infeasible for many users.

This chapter elaborates on the second approach, for a scenario comprising very many users. In the satellite communications context, the demodulator implemented in Enhanced Spread Spectrum ALOHA [40] constitutes a representative example of a practical system that deals with the problem of ordering users from the estimation of their strengths via preamble cross-correlations. This scenario may extend the unsourced random access paradigm [84–86], since the receiver does not know the identities of all users, and no channel state information is given. The last remark has reference to the fact, in these scenarios, a feasible configuration that facilitates the ordering task is that users employ the same coded modulation scheme.

5.2 System Model

As stated in the introductory section of the present chapter, the novelty introduced herein is the adoption of a SIC receiver to which the strengths (powers or symbol energies) received from all users are unknown. This chapter is intended to be mostly self-contained, and thus, some concepts already discussed throughout previous chapters of this thesis are reviewed as well.

The setting evaluated can be contextualized, as in some context-aware settings, in the satellite scenario envisaged in Enhanced Spread Spectrum ALOHA, where K direct-sequence users transmit on the uplink. The starting point of this chapter is the same K-user multiple access set-up adopted in the previous chapters, except for the peculiarity that the identities of all users are unknown to the receiver and that no channel state information is provided to it. Since the receiver needs to estimate the strengths of all users, we deem it appropriate, as inspired by the unsourced random access context [84–86], that the same codebook be adopted by all users[1]. In this respect, $f(\cdot)$ is the encoding function used to generate, from the binary sequence $\mathbf{b}_k$, the packet transmitted by user k, $\mathbf{s}_k = f(\mathbf{b}_k)$, consisting of a n_e-symbol payload and a n_o-symbol preamble. Direct-sequence spread spectrum is adopted together with long spreading signatures $c_{k,m}(t)$ for each symbol m of the transmitted packet. The spreading gain is N.

The baseband signal at the receiver's front-end is

$$y(\mathrm{t}) = \sum_{k=1}^{K} \sqrt{\gamma_{\mathrm{x}}[k]h[k]}\; e^{j\Omega_k} \sum_{m=1}^{n_o+n_e} \mathbf{s}_k[m] c_{k,m}(\mathrm{t} - mT - \tau_k) + w(\mathrm{t}), \tag{5.1}$$

with $\gamma_{\mathrm{x}}[k]$ the energy per symbol transmitted by user k, $h[k]$ its slowly time-varying channel power gain, and Ω_k the random carrier phase term. T is the symbol period, τ_k the end-to-end delay, and $w(\mathrm{t})$ Gaussian noise.

The symbol energy received from user k is $\gamma[k] = \gamma_{\mathrm{x}}[k]h[k]$, and thus, the distribution of received energies is $\gamma[1], \ldots, \gamma[K]$, which, for the convenience of analysis, is considered ordered non-increasingly according to the indexing $1 \leq k \leq K$. The analysis that follows studies a SIC receiver to which the symbol energies $\gamma[1], \ldots, \gamma[K]$ are unknown. In this sense, the demodulator adopted in the Enhanced Spread Spectrum ALOHA system has dealt with a similar, albeit more complex, problem [40]. It detects users and estimates their strengths combining preamble cross-correlations with an iterative SIC approach. This chapter, however, focuses only on the estimation-based ordering of users (at the initial stage) based on preamble cross-correlations. To that aim, it is considered that all users are detected, and the chapter deepens into evaluating the impact of noisy estimations on the ordering process and on system performance.

More specifically, at each time slot or window frame, generically denoted n, the implementation carried out by the receiver first estimates the symbol energies from all users, and then performs SIC. The system model derived herein distinguishes between: (i) the user-index $1 \leq k \leq K$ set in non-increasing order of the (real) received symbol energies; and (ii) the user-order $1 \leq k' \leq K$ set in non-increasing order of energy estimates[2], as:

$$\gamma[1] \;\geq\; \cdots \;\geq\; \gamma[k] \;\;\geq\; \cdots \;\geq\; \gamma[K], \tag{5.2}$$

$$\hat{\gamma}[1] \;\geq\; \cdots \;\geq\; \hat{\gamma}[k'] \;\geq\; \cdots \;\geq\; \hat{\gamma}[K]. \tag{5.3}$$

After ordering users, one can make use of the system model derived in Chapter 3 by defining

[1]Note that if users, due to different quality-of-service requirements, employ different encoders, the receiver needs to estimate: (i) the strengths received from all users; and (ii) the encoder employed by every user. With regard to the second task, users must encapsulate some additional information inside payload to allow, in the absence of coordination between users and the receiver, the identification of the encoder employed by each user.

[2]The ordering produced by energy estimates coincides with that set by signal-to-interference-plus-noise ratio estimates, when interference is stationary along the processing window frame.

the mapping function between indexings k' and k: $\phi_n : k' \longrightarrow k$. Note that the last definition preserves the subscript n to emphasise that ϕ_n changes at every processing time unit n. Then, the system starts by processing user $\phi_n[k' = 1]$ down to user $\phi_n[k' = K]$. At stage k' (out of K), the receiver operates with the output of the matched filter (despreader)

$$\tilde{\mathbf{y}}_k = \sqrt{\gamma[k]}\; e^{j\Omega_k}\; \mathbf{s}_k + \sqrt{N_t[k']}\; \mathbf{w}_k \quad \text{with} \quad k = \phi_n[k']. \tag{5.4}$$

Note that $\phi_n[k']$ produces the index of the k'-th ordered user, at time n, in the ordering based on the received symbol energies. The term $\mathbf{w}_k$ is the sampled waveform associated with the cross-correlation between the spreading signature of user $\phi_n[k']$ and others, and $N_t[k']$ is the variance of noise plus interference at stage k'. The advantage of using direct-sequence spread spectrum is that, at the despreader's output, the statistical distribution of interfering signals resembles a Gaussian distribution even if users do not employ Gaussian codebooks (actually, the case evaluated herein). Then, by treating interference as Gaussian signals that contribute as an addition to thermal noise, the signal-to-interference-plus-noise ratio (SINR) corresponding to the despreader's output at stage k' is

$$\Gamma[k'] = \frac{\gamma[\phi_n[k']]}{N_t[k']}, \tag{5.5}$$

where its denominator denotes the variance of the noise plus interference level affecting the user processed at stage k', computed as

$$N_t[k'] = 1 + \frac{\theta}{N} \sum_{i=1}^{k'-1} \epsilon[i]\gamma[\phi_n[i]] + \frac{\theta}{N} \sum_{i=k'+1}^{K} \gamma[\phi_n[i]]. \tag{5.6}$$

In the above expression, $\frac{\theta}{N}$ is the average decorrelation factor between symbol signatures in the long-code model [14], and $0 \leq \theta \leq 1$ is a known average factor that models the dispersion of user-delays $\tau_{1 \leq k \leq K}$. As in the previous chapters, the adopted SIC policy is aided by redundancy-check error control that operates at stage k' as follows: firstly, channel decoding "determines" the bit stream transmitted by user k as $\hat{\mathbf{b}}_k = f^{-1}(\tilde{\mathbf{y}}_k)$; and secondly, the CRC is checked out, after which cancellation is only produced if packet decoding succeeds. The respective success/failure packet decoding event to model the processing of user $k = \phi_n[k']$ is modelled through the binary random variable $\epsilon[k']$, which equals 1 when $\hat{\mathbf{b}}_k \neq \mathbf{b}_k$, an event that occurs with probability $\mathrm{PER}[\Gamma[\phi_n[k']]$; and equals $\varepsilon[\Gamma[\phi_n[k']]]$ complementarily when $\hat{\mathbf{b}}_k = \mathbf{b}_k$. In the previous line:

(i) $\mathrm{PER}[\Gamma]$ is the known packet error rate (PER) versus SINR curve associated with the employed decoding system, adopted to model the error performance of its operation, and

(ii) $\varepsilon[\Gamma]$ is the known residual energy (RE) curve, used to model the average fraction of energy that remains after cancelling a user.

At this point, and after the K stages of the cancellation receiver, our concern is to evaluate its performance. This chapter adopts the (user-aggregate) spectral efficiency figure of merit that adds the effective transmission rates of all users, defined by

$$\mathrm{SE} = \frac{K}{N} R\; (1 - \overline{\mathrm{per}}_{\mathrm{lt}}), \tag{5.7}$$

with R the rate of the coded modulation scheme employed by all users, and $\overline{\text{per}}_{\text{lt}}$ the (long term) average number of packet decoding errors averaged over sufficiently large time slots. In this chapter, nonetheless, since users employ the same physical layer and transmission protocol, the above figure of merit is uniquely determined by the behaviour of channel decoding after operating with each user. Motivated by relevant performance figures of merit sensitive to system model parameters from a network- or user-centric perspective, this chapter also evaluates the individual and network long-term average packet error probabilities:

$$\overline{\text{per}}_{\text{lt}}[k] = \lim_{L\to\infty} \frac{1}{L}\sum_{n=1}^{L} \text{PER}\left[\Gamma[\phi_n^{-1}[k]]\right], \tag{5.8}$$

$$\overline{\text{per}}_{\text{lt}} = \frac{1}{K}\sum_{k=1}^{K} \overline{\text{per}}_{\text{lt}}[k]. \tag{5.9}$$

Certainly, the above (long-term) model simplifies the performance evaluation of such a receiver and achieves high accuracy, but still lacks sufficient elements to provide insights into the behaviour of dynamically ordered SIC without descending into the simulation level. The following subsection addresses this purpose, where a system model based on statistical averages rather than on long-term averages is proposed and reasonably justified.

5.2.1 Estimating the Strengths of All Users

The energy estimates corresponding to all users are analysed herein. Recall that user strengths are estimated at the initial SIC stage and before proceeding to successive decoding. Actually, our study adopts a simpler approach than that in [40]. In our case, the task of ordering users is decoupled from the decoding process. More specifically, after detecting all users, a bank of single-user detectors compute preamble cross-correlations[3], each of them to estimate the symbol energy received from a user. The SINR corresponding to the k-th user energy estimator is

$$\Lambda[k] = \frac{\gamma[k]}{1 + \frac{\theta}{N}\sum_{i\neq k}\gamma[i]}. \tag{5.10}$$

That is, users are subject to the same noise plus interference level at the despreader's output, except for their own contribution. Nonetheless, when the number of users K is sufficiently large, the denominator of the former expression can be accurately approximated by a constant term by adding, additionally, the contribution of user k to the SINR of its estimator, as

$$\Lambda[k] \approx \frac{\gamma[k]}{1 + \alpha\theta\bar{\gamma}}. \tag{5.11}$$

$\alpha \triangleq K/N$ is the system load, and $\bar{\gamma} \triangleq \frac{1}{K}\sum_{k=1}^{K}\gamma[k]$ is the average symbol energy received from all users[4]. Remarkably, as the number of users increases, user-energy estimations are affected by the same noise plus interference level.

[3] Cross-correlation constitutes the basis for the maximum-likelihood single-user estimation of the energy of a signal in presence of Gaussian interfering signals treated as noise and known preambles.

[4] The validity of posterior results is subject to having $\bar{\gamma} < \infty$, according to [14, Ch. 2].

The analysis that follows investigates the statistical behaviour of symbol energy estimates. Since the estimates are computed from the same signal $y(\mathrm{t})$ after the respective despreading operators, energy estimations are subject to noise plus interference terms that are not statistically independent from the strictly mathematical point of view. Nonetheless, one can consider, for analysis purposes, energy estimates to be statistically independent provided that: firstly, the preambles employed by all users are independent; secondly, that long-spreading codes after despreading largely change the noise plus interference terms $\mathbf{w}_{1\leq k\leq K}$ affecting all estimators; and thirdly, that users are subject to non-synchronous interfering signals. Remarkably, the latter considerations allow to assume, to a sufficiently high accuracy, that different users are subject to terms $\mathbf{w}_k$ statistically independent. Thus, estimated symbol energies are modelled as independent random variables. Specifically, the estimated symbol energy from user k follows a scaled non-centred chi-squared distribution

$$\hat{\gamma}[k] \sim \frac{1}{\beta}\mathcal{X}_2^2\left(\beta\gamma[k]\right), \tag{5.12}$$

with 2 degrees of freedom and non-centrality parameter $\beta\gamma[k] = 2n_o\Lambda[k]$. In the expression above, $\beta \triangleq 2n_o(1+\alpha\theta\bar{\gamma})^{-1}$ is a known factor independent of k.

The previous estimations can be further enhanced if it is considered: firstly, that during many consecutive time slots or working frames, the same users transmit without varying their transmitted symbol energies; and secondly, that the channel distribution varies sufficiently slowly so that the received energy distribution $\gamma[1],\ldots,\gamma[K]$ stays practically unchanged. Then, energy estimations can be improved by averaging those obtained in m previous time slots or window frames. Then, symbol energy estimates are distributed as

$$\hat{\gamma}[k] \sim \frac{1}{\beta}\mathcal{X}_{2m}^2\left(\beta\gamma[k]\right), \tag{5.13}$$

a non-central chi-square distribution with $2m$ degrees of freedom and non-centrality parameter $\beta\gamma[k] = 2mn_o\Lambda[k]$, with $\beta \triangleq 2mn_o(1+\alpha\theta\bar{\gamma})^{-1}$. A last remark concerns the straightforward prove that the estimator is consistent since $\hat{\gamma}[k] \to \gamma[k]$ as $m \to \infty$.

5.2.2 Statistical Average System Model

As stated previously, our objective is to gain further knowledge on the behaviour of dynamically ordered SIC with estimation-based ordering of users. To that aim, this subsection assumes that the same users are transmitting during many time slots, and that the distribution of received symbol energies $\gamma[1],\ldots,\gamma[K]$ stays practically constant within a sufficiently long period of time. The latter assumption holds as long as the channel of each user varies sufficiently slowly (as found in the land mobile satellite channel) and users do not vary their transmitted powers. Therefore, instead of operating with instantaneous ordering expressions, one can make use of statistics and turn long-term expressions into statistical averages. The explanation in this section provides a more extensive discussion relative to that in our article [77].

Some key points are first stated to illuminate the system model proposed herein, which differs substantially from those adopted in the previous chapters and also from the recent literature. Getting down to the substance, since users are ordered at different positions in different time

units or window frames, the adoption of the univariate model (in previous chapters) in which user k is processed at stage k is meaningless. Instead, for instance, the proposed model considers the SINR of user k when is ordered and processed at stage k', defined as the ratio between the symbol energy received from user k to the noise plus interference level:

$$\bar{\Gamma}[k,k'] = \frac{\gamma[k]}{\bar{N}_t[k,k']}. \tag{5.14}$$

The overline notation is added to distinguish the variables from those adopted in the long-term system model described previously. The difficulty lies in the calculation of the term $\bar{N}_t[k,k']$, the noise plus interference term corresponding to user k when is ordered the k'-th.

5.2.2.1 Calculation of the noise plus interference term

The model adopted for $\bar{N}_t[k,k']$ is based on a comprehensive large-user analysis of the term $N_t[k']$ (5.6), whose expression is reproduced below

$$N_t[k'] = 1 + \frac{\theta}{N}\sum_{i=1}^{k'-1}\epsilon[i]\gamma[\phi_n[i]] + \frac{\theta}{N}\sum_{i=k'+1}^{K}\gamma[\phi_n[i]]. \tag{5.15}$$

The corresponding user processed at stage k' experiences an interference level consisting of the interference from users already processed (first summation) plus that from users yet to be processed (second summation). Regarding the first summation, one can consider that if the number of users already processed $k'-1$ is sufficiently large, the summation is equivalent to a summation where all users $u \neq k$ other than $k = \phi_n[k']$ are processed in all previous stages $i < k'$, and their respective contributions are weighted by their probabilities. In this respect, the factor $p_{u,i|k,k'}$ is used to denote the probability that a user u is ordered the i-th conditioned on the fact that user k is ordered the k'-th:

$$p_{i,u|k,k'} \triangleq \Pr[\phi_n^{-1}[i]{=}u \mid \phi_n^{-1}[k]{=}k']. \tag{5.16}$$

Inspired by this model, the average interference from users already processed when user k is ordered the k'-th is computed as

$$\bar{\xi}_{\text{prv}}[k,k'] \triangleq \frac{\theta}{N}\sum_{i=1}^{k'-1}\sum_{u\neq k} q[\bar{\Gamma}[u,i]]\gamma[u]p_{u,i|k,k'}, \tag{5.17}$$

where $q[\Gamma] = 1-(1-\varepsilon[\Gamma])\text{PSR}[\Gamma]$ is the expectation of the random variable $\epsilon[i]$ in (5.15), which subtracts the fraction $(1-\varepsilon[\Gamma])\text{PSR}[\Gamma]$ of the user's energy after processing it. $\text{PSR}[\Gamma] \triangleq 1-\text{PER}[\Gamma]$ is the packet success rate function. Similar rationales can be followed with regard to the second summation in (5.15), to finally compute the interference level from users remaining unprocessed at state k' when user k is ordered the k'-th, as

$$\bar{\xi}_{\text{rem}}[k,k'] \triangleq \frac{\theta}{N}\sum_{i=k'+1}^{K}\sum_{u\neq k}\gamma[u]p_{u,i|k,k'}. \tag{5.18}$$

The reader must note that the large-user model adopted consists of an extension of the models described in the previous chapters, where univariate functions are turned into bivariate functions of the user index k and the user order k', and where, at each SIC stage, we assume that the system works equivalently as when all users different from the evaluated one are processed. Joining the above computations, we have the expression for the new noise plus interference term

$$\bar{N}_t[k,k'] = 1 + \frac{\theta}{N}\sum_{i=1}^{k'-1}\sum_{u\neq k} q[\bar{\Gamma}[u,i]]\gamma[u]p_{u,i|k,k'} + \frac{\theta}{N}\sum_{i=k'+1}^{K}\sum_{u\neq k}\gamma[u]p_{u,i|k,k'}. \tag{5.19}$$

The challenge is to find an expression for the probability $p_{u,i|k,k'}$ since it contains four arguments and, therefore, its expression intuitively can make the final analysis difficult. We strive for a simplification and derive a simpler model. Let us observe the conditional imposed on the probability $p_{u,i|k,k'}$, k,k', which is a common term to all addends of the above summations. In the many-user regime, the conditional becomes more and more irrelevant. That is, the fact that user k is ordered in a specific position does not condition, to a large extent, the order of the rest of users, and as $K\to\infty$ the conditional turns out to be independent. Under this assumption and considering many users, the probability $p_{u,i|k,k'}$ can be approximated by obviating the conditional

$$p_{u,i|k,k'} \approx p_{u,i} \triangleq \Pr[\phi_n^{-1}[u] = i] \tag{5.20}$$

that is, by the probability that user u is ordered the i-th. It is easy to illustrate the magnitude of this approximation with a simple example. Consider that all K users are estimated with the same SINR, which corresponds to the case when the received symbol energy distribution $\gamma[1] = \cdots = \gamma[K]$ is uniform. Then, all users have the same probability of being ordered at a given position. Keeping the subscript notation in the probability terms, we have $p_{u,i} = \frac{1}{K}$. Moreover, the previous probability conditional on user k already ordered is $p_{u,i|k,k'} = \frac{1}{K-1}$, term that converges as $K\to\infty$ to $p_{u,i}$. For the moment, we will leave each of the terms $p_{k,k'}$ without an analytical expression. Their computations are addressed in Section 5.2.2.3 for a generic received symbol energy distribution.

The approximation (5.20) allows to compute accurately the interference corresponding to users already processed and that to users remaining unprocessed, as

$$\bar{\xi}_{\text{prv}}[k,k'] = \frac{\theta}{N}\sum_{i=1}^{k'-1}\sum_{u\neq k} q[\bar{\Gamma}[u,i]]\gamma[u]p_{u,i}, \tag{5.21}$$

$$\bar{\xi}_{\text{rem}}[k,k'] = \frac{\theta}{N}\sum_{i=k'+1}^{K}\sum_{u\neq k}\gamma[u]p_{u,i}. \tag{5.22}$$

Joining both terms, the model for the noise plus interference term seen by user k when is ordered at position k' is

$$\bar{N}_t[k,k'] = 1 + \frac{\theta}{N}\sum_{i=1}^{k'-1}\sum_{u\neq k} q[\bar{\Gamma}[u,i]]\gamma[u]p_{u,i} + \frac{\theta}{N}\sum_{i=k'+1}^{K}\sum_{u\neq k}\gamma[u]p_{u,i}. \tag{5.23}$$

5.2.2.2 Calculation of individual and network key performance indicators

As we have commented in previous sections, the focus of the present chapter is also the evaluation of individual and global utilities, more specifically, the ones defined in (5.8)–(5.9). The key performance indicators are the packet error probability associated with user k and the average packet error probability over all users, which were defined previously through long-term expectations. Recall that, the ordering is established on the basis of estimates but, once the ordering is given, system performance depends on the symbol energies received from all users. The model adopted addresses the above peculiarity by computing the performance indicators based on statistical averages rather than on long-term computations. The following expressions are obtained:

$$\overline{\text{per}}[k] = \sum_{k'=1}^{K} \text{PER}\left[\frac{\gamma[k]}{\bar{N}_t[k,k']}\right] p_{k,k'}, \tag{5.24}$$

$$\overline{\text{per}} = \frac{1}{K}\sum_{k=1}^{K}\sum_{k'=1}^{K} \text{PER}\left[\frac{\gamma[k]}{\bar{N}_t[k,k']}\right] p_{k,k'}. \tag{5.25}$$

5.2.2.3 Computation of User-Order Probabilities

The only computation that remains yet to be found is the probability that user k is ordered the k'-th: $p_{k,k'}$ (we have adopted the initial notation for the user-index k and the user-order k'). The challenge is to find an expression for the probability matrix

$$\begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,K} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,K} \\ \vdots & \vdots & \ddots & \\ p_{K,1} & p_{K,2} & \cdots & p_{K,K} \end{bmatrix}. \tag{5.26}$$

An expression for each of the former probabilities is provided next, which are shown to take the same form except for some parameters.

We focus the following analysis on the user's energy estimate k, $\hat{\gamma}[k]$, and we first analyse the distribution of $\phi_n^{-1}[k]$ conditioned on $\hat{\gamma}[k] = x$, $\phi_n^{-1}[k]|\{\hat{\gamma}[k]{=}x\}$. The order of user k can be obtained as the sum of $K-1$ Bernoulli random variables $\mathcal{B}_{i\neq k}$ equal to 1 when $\hat{\gamma}[i] \geq \hat{\gamma}[k]$, the energy estimated from user i is higher than that estimated from user k $\hat{\gamma}[k] = x$, as

$$\phi_n^{-1}[k]|\{\hat{\gamma}[k]{=}x\} = 1 + \mathcal{B}_1 + \cdots + \mathcal{B}_{k-1} + \mathcal{B}_{k+1} + \cdots + \mathcal{B}_K. \tag{5.27}$$

The success probability of the i-th Bernoulli random variable $\mathcal{B}_i$ is computed from the tail distribution of a non-central chi-squared distribution, as

$$q_i = \Pr[\hat{\gamma}[i] \geq x] = \Pr\left[\mathcal{X}^2_{2m}\left(\beta\gamma[k]\right) \geq \beta x\right] = Q_m\left(\sqrt{\beta\gamma[k]}, \sqrt{\beta x}\right). \tag{5.28}$$

$Q_m(a,b)$ is the Marcum-Q function of order m (m is an integer).

The distribution of $\phi_n^{-1}[k]|\{\hat{\gamma}[k]{=}x\}$ is Poisson Binomial, and approaches the normal distri-

bution $\mathcal{N}(\mu_k(x), \sigma_k^2(x))$ as K increases. The mean and variance are computed as

$$\mu_k(x) \triangleq 1 + \sum_{i \neq k} q_i \quad \text{and} \quad \sigma_k^2(x) \triangleq \sum_{i \neq k} q_i(1 - q_i). \tag{5.29}$$

The probability of a user being ordered the k'-th can be then computed approximately by sampling the Gaussian density $f_{\mathcal{N}}(k'; \mu_k(x), \sigma_k^2(x))$ at the order point k', and by averaging the result according to the density of energy estimates $f_{\hat{\gamma}}(x; \gamma[k])$ associated with user k, as

$$p_{k,k'} \approx \int_0^\infty f_{\mathcal{N}}\left(k'; \mu_k(x), \sigma_k^2(x)\right) f_{\hat{\gamma}}(x; \gamma[k]) \mathrm{d}x. \tag{5.30}$$

A last remark concerns the above approximation, which simplifies the computation of the user-order probabilities through the probability mass function of the Poisson Binomial random variable with that of a Gaussian density. Certainly, the result is only valid for a sufficiently large number of users K, in practice, in the order of hundreds of users.

5.3 The Model for Asymptotically Many Users

As seen throughout the present document, this thesis elaborates on the analysis of the SIC receiver in the asymptotic large-user regime. This section provides an amenable tractable form to the above model, divided in four subsections. Firstly, a brief analysis on the asymptotic user-order probabilities is addressed in Section 5.3.1. Secondly, the system model described in Section 5.2.2 is turned into the asymptotic large-user regime in Section 5.3.2. Next, Section 5.3.3 comments briefly on the extension of the model derived herein to the more complex iterative receiver adopted in the E-SSA system. Finally, Section 5.3.4 presents the simulation results.

The user-asymptotic behaviour of the system model described in Section 5.2.2 is analysed first. As considered in the previous chapters, it is advantageous to adopt a more convenient notation to index users in the user-asymptotic regime. Two asymptotic indexings are defined:

$$t = \lim_{K \to \infty} \frac{k}{K} \quad (0 \leq t \leq 1), \tag{5.31}$$

$$t' = \lim_{K \to \infty} \frac{k'}{K} \quad (0 \leq t' \leq 1), \tag{5.32}$$

which condense a dense number of users in the interval $[0, 1]$. Henceforth, $0 \leq t \leq 1$ is denoted user-index, and $0 \leq t' \leq 1$ user-order. The latter definitions together with the asymptotic traffic load relationship

$$\alpha \triangleq \lim_{K,N \to \infty} \frac{K}{N}, \tag{5.33}$$

allow to define asymptotic and *smooth* functions of the user-variable t or the user-variable t'. Examples of such smooth functions are the distributions corresponding to the received and to the estimated symbol energies from all users:

$$\gamma(t) \triangleq \lim_{K \to \infty} \gamma[Kt] \quad \text{and} \quad \hat{\gamma}(t') \triangleq \lim_{K \to \infty} \hat{\gamma}[Kt']. \tag{5.34}$$

5.3.1 Asymptotic User-Order Probabilities

As seen above, the difficulty of the previous model lies on determining an explicit expression for the probability that user k is ordered the k'-th $p_{k,k'} = \Pr[\phi_n^{-1}[k] = k']$, which, due to the asymptotic indexings (5.31)–(5.32), is now turned into the probability that user $k \to Kt$ is ordered at $k' \to Kt'$, defined by

$$\mathrm{d}p(t,t') = \lim_{K\to\infty} \Pr[\phi_n^{-1}[Kt] = Kt']. \tag{5.35}$$

Note that $\mathrm{d}p(t,t')$ is an infinitesimal quantity since the number of possible orders for a user t increases as $K \to \infty$ in the same order.

Curiously, as proved in Appendix 5.A.1, $\mathrm{d}p(t,t')$ admits a closed-form expression in the asymptotic large-user regime. The order of user $k = Kt$ depends only on its estimated symbol energy and not on the estimates corresponding to other users. This is rigorously true in the user-asymptotic regime since the asymptotic distribution of estimated energies is known, and we only need the estimate from user $k = Kt$ to determine its position in the known asymptotic distribution. Therefore, the asymptotic user-order probabilities are cast compactly as

$$\mathrm{d}p(t,t') = a(t,t')\mathrm{d}t', \tag{5.36}$$

an asymptotic (K-independent) kernel $a(t,t')$ times the differential $\frac{1}{K} \to \mathrm{d}t'$. The proof follows from taking the expressions derived in Section 5.2.2.3 to the user-limit. Summarising the contribution in Appendix 5.A.1, the asymptotic kernel $a(t,t')$ obeys three equivalent expressions:

$$a(t,t') = \nabla_{t'} \bar{F}_{\hat{\gamma}}(\bar{\mu}^{-1}(t');\gamma(t)) \tag{5.37a}$$

$$= \dot{\bar{F}}_{\hat{\gamma}}(x;\gamma(t)) \cdot \frac{1}{\dot{\bar{\mu}}(x)}\bigg|_{x=\bar{\mu}^{-1}(t')} \tag{5.37b}$$

$$= -f_{\hat{\gamma}}(x;\gamma(t)) \cdot \frac{1}{\dot{\bar{\mu}}(x)}\bigg|_{x=\bar{\mu}^{-1}(t')}. \tag{5.37c}$$

In the expression above, $\bar{F}_{\hat{\gamma}}(x;\gamma) \triangleq \Pr[\mathcal{X}^2_{2m}(\beta\gamma) \geq \beta x]$ and $f_{\hat{\gamma}}(x;\gamma)$ in $x \geq 0$ are, respectively, the tail distribution and the density function of symbol energy estimates. Moreover, $\bar{\mu}^{-1}(t')$ is the inverse function of

$$\bar{\mu}(x) = \int_0^1 Q_m\left(\sqrt{\beta\gamma(\tau)}, \sqrt{\beta x}\right) \mathrm{d}\tau \tag{5.38}$$

applied to the user-order t', or equivalently, the x solution to $\bar{\mu}(x) = t'$. Finally, the last observation concerns the consistency of the above model in that, the probabilities $\mathrm{d}p(t,t')$ add up to 1 in both directions t and t'.

5.3.2 Asymptotic System Model

In previous sections, the system's behaviour has been explored for many users and, in some cases, leveraging some approximations based on the asymptotic large-user regime. This subsection derives the user-asymptotic form of the model adopted before, to enable full understanding of

dynamically ordered SIC. The SINR (5.14) now turns into the bivariate function of the user-index t and the user-order t'

$$\bar{\Gamma}(t,t') = \frac{\gamma(t)}{\bar{N}_t(t')}, \tag{5.39}$$

where the noise plus interference profile $N_t(t')$ comes from considering that, as $K \to \infty$, the term $\bar{N}_t[k,k'] = \bar{N}_t[Kt,Kt'] \to N_t(t')$, obviating the absence of contribution of the evaluated user in the noise plus interference level. In this regime, the asymptotic noise plus interference function corresponding to the adopted SIC policy is

$$\bar{N}_t(t') = 1 + \alpha\theta\int_0^1\int_0^{t'} q[\bar{\Gamma}(u,\tau)]\gamma(u)\mathrm{d}p(u,\tau)\mathrm{d}u + \alpha\theta\int_0^1\int_{t'}^1 \gamma(u)\mathrm{d}p(u,\tau)\mathrm{d}u. \tag{5.40}$$

Remarkably, inner integrals operate under the differential $\mathrm{d}p(u,\tau) = a(u,\tau)\mathrm{d}\tau$, and outer integrals under $\mathrm{d}u$. Note that the above expressions correspond, simply, to an extension of the model derived in Chapter 3 for one-iteration SIC with known decoding order, where single integrals have been substituted by double integrals that average the contribution of all users at each stage. In this user-asymptotic regime, the model is easy to be understand: if we observe any position t'_i of the interval $0 \leq t' \leq 1$ over a window $\Delta t'$ as $[t'_i - \frac{\Delta t'}{2}, t'_i + \frac{\Delta t'}{2}]$, we may find: a fraction $\frac{\Delta p(t_1,t'_i)}{\Delta t'}$ of users with energy $\gamma(t_1)$; a fraction $\frac{\Delta p(t_2,t'_i)}{\Delta t'}$ of users with energy $\gamma(t_2)$; etc. The smaller the window length $\Delta t'$, the truer the result. Generalising the result, we have that, at each interval $\mathrm{d}t'$, we may find the fraction $\frac{\mathrm{d}p(t,t')}{\mathrm{d}t'} = a(t,t')$ of users received with the same energy $\gamma(t)$, albeit ordered randomly within the interval $\mathrm{d}t'$. We have adopted a model that ignores the ordering of users in each infinitesimal interval, by considering that all users contribute as a fraction of the mitigation of interference at each stage.

With regard to the noise plus interference term (5.40), it is possible to come up with a simpler expression in terms of a dynamic equation (differential equation) that can be obtained by differentiating (5.40) with respect to the user-order t', and by dividing the result by $\bar{N}_t(t')$. The following differential equation and initial condition are obtained:

$$\frac{\dot{\bar{N}}_t(t')}{\bar{N}_t(t')} = -\alpha\int_0^1 \Phi\left[\frac{\gamma(u)}{\bar{N}_t(t')}\right] a(u,t')\mathrm{d}u, \tag{5.41}$$

$$\bar{N}_t(0) = 1 + \alpha\theta\int_0^1 \gamma(u)\mathrm{d}u, \tag{5.42}$$

with $\Phi[\Gamma] \triangleq \theta(1-\varepsilon[\Gamma])\Gamma\cdot\mathrm{PSR}[\Gamma]$ the same decoding-cancellation characteristic defined in Chapter 3. Now, using (5.37c) for the term $a(u,t')$ and defining the term $N_t(t')$ using as argument the energy corresponding to the user-order t', by means of substituting $t' = \bar{\mu}(x)$ as $\tilde{N}_t(x) = \bar{N}_t(\bar{\mu}(x))$, the following differential equation and initial condition are obtained:

$$\frac{\dot{\tilde{N}}_t(x)}{\tilde{N}_t(x)} = \alpha\int_0^1 \Phi\left[\frac{\gamma(u)}{\tilde{N}_t(x)}\right] f_{\hat{\gamma}}(x;\gamma(u))\mathrm{d}u, \tag{5.43}$$

$$\tilde{N}_t(\infty) = 1 + \alpha\theta\int_0^1 \gamma(u)\mathrm{d}u. \tag{5.44}$$

The interest in obtaining this form of differential equation is that (5.43) contains explicitly the density $f_{\hat{\gamma}}(x;\gamma)$. Note that, whereas the user order goes from $t'=0$ down to $t'=1$, the order in the variable x goes from the highest estimated symbol energy $x\to\infty$ down to the lowest $x\to 0$.

With all this said, the last expressions that remain to be determined are the asymptotic forms of the packet error probabilities at individual (5.24) and network (5.25) level. With regard to the individual figure of merit, the asymptotic expression for the packet error probability averages the contribution of a user in all possible orderings. This corresponds to one of the following two equivalent expressions

$$\overline{\text{per}}(t) = \int_0^1 \text{PER}\left[\frac{\gamma(t)}{\bar{N}_t(\tau)}\right] a(t,\tau)\mathrm{d}\tau \tag{5.45a}$$

$$= \int_0^\infty \text{PER}\left[\frac{\gamma(t)}{\tilde{N}_t(x)}\right] f_{\hat{\gamma}}(x;\gamma(t))\mathrm{d}x. \tag{5.45b}$$

Along the same line, the figure of merit that averages the individual packet error probabilities over all users reads

$$\overline{\text{per}} = \int_0^1\int_0^1 \text{PER}\left[\frac{\gamma(t)}{\bar{N}_t(\tau)}\right] a(t,\tau)\mathrm{d}\tau\mathrm{d}t \tag{5.46a}$$

$$= \int_0^1\int_0^\infty \text{PER}\left[\frac{\gamma(t)}{\tilde{N}_t(x)}\right] f_{\hat{\gamma}}(x;\gamma(t))\mathrm{d}x\mathrm{d}t. \tag{5.46b}$$

5.3.3 Extension of the Asymptotic System Model

The limitation of the proposed analysis is the adopted SIC policy, that estimates the strengths of users at the initial stage and before proceeding to successive decoding. Certainly, the estimation of energies can be improved if, as the system progresses by cancelling users, symbol energies of remaining users are re-estimated to reduce the variance of previous estimations. The latter corresponds to a more complex SIC receiver. However, the present analysis, even evaluating a naive version of a more sophisticated receiver, should be understood as a starting point to enable, following the same methodology, the analysis of more intricate schemes. As an example, to analyse the demodulator adopted in the Enhanced Spread Spectrum ALOHA system [40], the processes of estimation and ordering of users can not be decoupled as in our study. In such a case, intuition leads to analyse the joint system as two intertwined dynamical systems. Nevertheless, its extensive analysis is reserved for future research.

5.3.4 Empirical Evaluation

So as to evaluate the accuracy of the proposed statistical average system model relative to the long-term average model, the individual and network performance figures of merit are compared. The simulation carried out comprises K users transmitting packets of payload $n_e = 450$ symbols encoded with a theoretical coded modulation system of rate 1 bit/symbol. In addition, it is also considered that users are received with the exponential distribution $\gamma(t) = 10\exp(-1.5t)$ sampled as $t = k/K$ for $k = 1,\ldots,K$, and that symbol energies are estimated through preambles

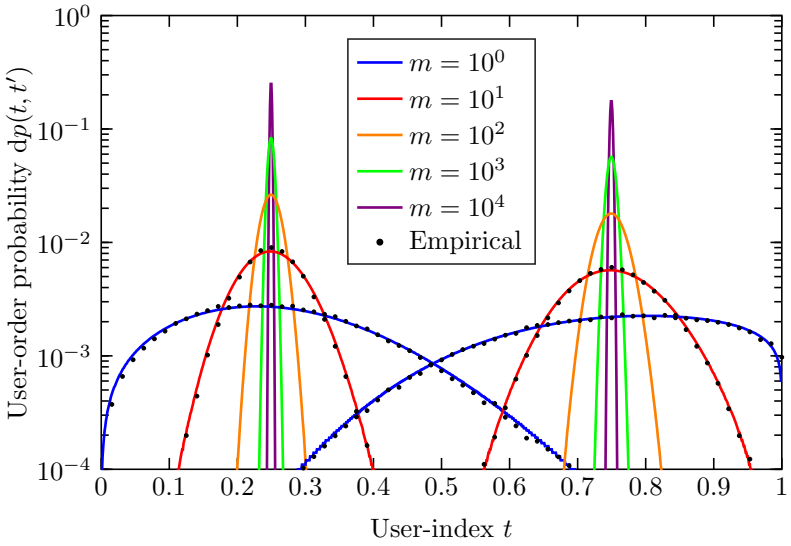

Figure 5.1: Asymptotic and empirical user-order probabilities for users $t_1 = 0.25$ and $t_2 = 0.75$. The simulated traffic load is $\alpha = 1.50$. The empirical computations are computed for $K = 64$ users.

of $n_o = 50$ symbols. With regard to the SIC receiver, the residual energy factor is set, as in the previous chapters, to $\varepsilon[\Gamma] = \varepsilon = 1\%$, and the decorrelation factor θ to 1.

Figure 5.1 depicts the empirical user-order probabilities overlapped with the asymptotic computations. Empirical computations simulate the ordering process with $K = 64$ independent estimations of symbol energies from the stated distribution $\gamma(t)$. Two observations are carried out. Firstly, at the asymptotic level, the order of user t converges to its index, $t' \to t$, set as a function of the received symbol energies, the more energy estimates are averaged. Secondly, at the empirical level, the asymptotic probabilities match the empirical computations for only $K = 64$ users. The same conclusion is obtained from simulating different distributions. Notably, the user-order probabilities are validated for a low K.

The next simulations evaluate the accuracy of the statistical average model:

(i) The packet error probability averaged over all users is analysed versus the traffic load α in Figure 5.2. As shown, the asymptotic computations $\overline{\mathrm{per}}$ (5.46a) predict very accurately the empirical behaviour $\overline{\mathrm{per}}_{\mathrm{lt}}$ (5.9) of the adopted SIC in all the regimes evaluated. The asymptotic predictions improve slightly when the empirical computations are carried out under a higher number of users. Moreover, system performance is substantially degraded when the symbol energies received from all users are estimated in the presence of high interference levels (high α). In these cases, when tenths of symbol energies are averaged, the variances of such estimations are reduced, as well as its impact on system performance.

(ii) The individual packet error probabilities experienced by all users are evaluated. To that aim, the asymptotic probability profile $\overline{\mathrm{per}}(t)$ in (5.45a) is contrasted with the (long-term) empirical computation $\overline{\mathrm{per}}_{\mathrm{lt}}[Kt]$ (5.8) for K users. As shown, the asymptotic results practically match the empirical computations when the number of users is sufficiently high. In the depicted simulation, an accurate prediction is shown when the system supports some hundreds of users.

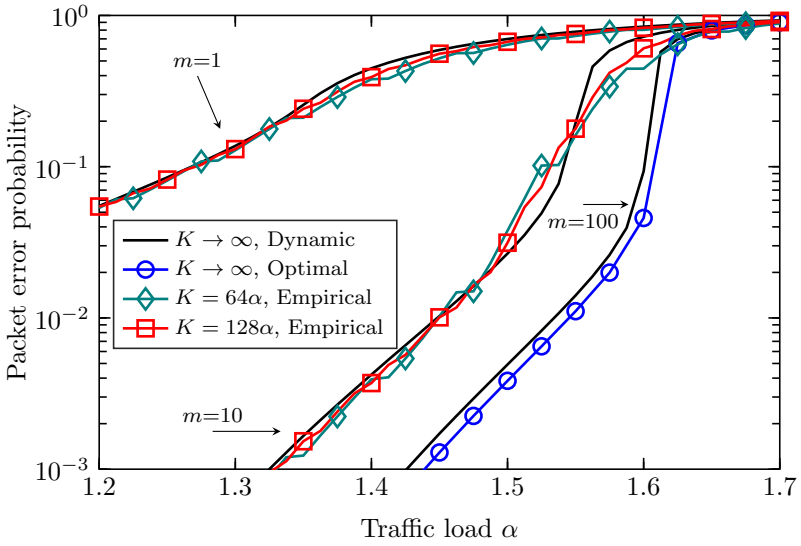

Figure 5.2: Average packet error probability over all users (network-centric figure of merit) versus traffic load α. The asymptotic computations $K \to \infty$ depict $\overline{\text{per}}$ (5.46a) with a decode ordering based on symbol energy estimates (dynamic) and on the received symbol energies (optimal). The empirical computations depict $\overline{\text{per}}_{\text{lt}}$ (5.9) averaged under 10^4 Monte Carlo runs.

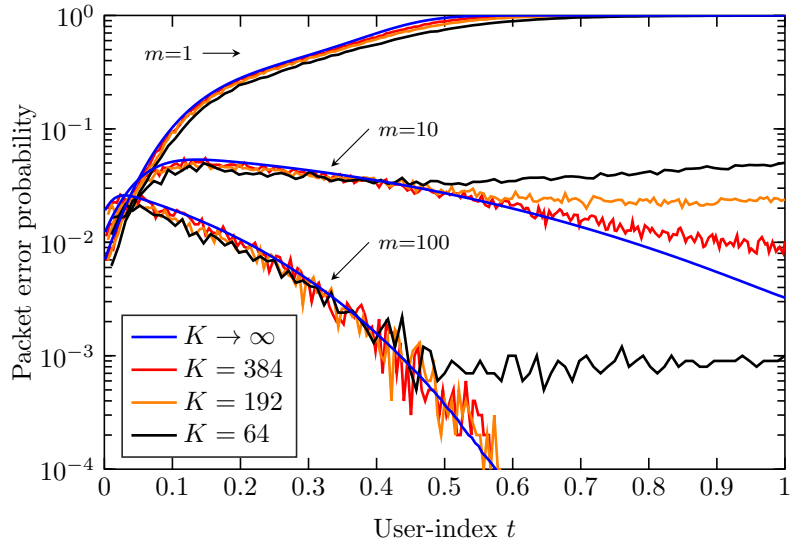

Figure 5.3: Individual packet error probabilities (user-centric figure of merit) achieved with dynamic ordering of users. The asymptotic computations $K \to \infty$ depict the user-PER profile $\overline{\text{per}}(t)$ (5.45a), and the empirical computations show the long-term figure of merit $\overline{\text{per}}_{\text{lt}}[Kt]$ (5.8) averaged under 10^4 Monte Carlo runs.

5.4 Asymptotically Optimal Energy Allocation

Similarly to previous chapters, this section devotes its content to addressing the energy allocation problem in the user-asymptotic regime accounting for the adopted SIC. According to the structure used in previous chapters, the present analysis corresponds to the energy allocation design with optimal reliability and fair coding rate. Recall that, in order to reduce the complexity of the SIC receiver with dynamic decoding order, users adopt the same coded modulation scheme. This simplifies the ordering task, as users are decoded in non-increasing order of estimated symbol energies. Getting to the crux of the matter, we assume that, contrarily to the receiver, users are capable of accurately estimating their individual channel gains without feedback from the receiver. The former holds reasonably provided that the satellite broadcasts a known pilot over a low-rate control channel and that the channel reciprocity hypothesis applies.

The figure of merit to evaluate system performance is the asymptotic spectral efficiency (ASE) defined by the following two equivalent expressions

$$\mathrm{ASE} = \alpha R \int_0^1 \int_0^1 \mathrm{PSR}\left[\frac{\gamma(t)}{\bar{N}_t(t')}\right] a(t,t')\mathrm{d}t\mathrm{d}t' \tag{5.47a}$$

$$= \alpha R \int_0^1 \int_0^\infty \mathrm{PSR}\left[\frac{\gamma(t)}{\tilde{N}_t(x)}\right] f_{\hat{\gamma}}(x;\gamma(t))\mathrm{d}x\mathrm{d}t. \tag{5.47b}$$

This chapter addresses, relative to Chapters 3 and 4, a more simplified problem in which the average energy constraint is enforced at reception, with the aim at setting the working point of the SIC receiver. Such an analysis corresponds to the case when channel power gains corresponding to all users are practically the same, or when users transmit towards a satellite transponder that fixes the operating point of the satellite's amplifier. In the latter and more reasonable case, it is understood that the constraint on the transmitted energy is also present, but that of the receiver is more restrictive. The energy allocation problem solved herein can be extended, at a larger stage, to the more general case by setting the average energy constraint to transmitters.

5.4.1 Optimal User-Energy Distribution

The maximisation of ASE (5.47b) is addressed by solving the following variational calculus problem subject to an average energy constraint $\bar{\gamma}$ (5.48b), and the dynamical equation, derived in this chapter, governing the evolution of the noise plus interference level along the SIC receiver stages (5.48c)–(5.48d). The following problem needs to be solved:

$$\max_{\gamma(t)} \ \mathrm{ASE} \tag{5.48a}$$

$$\text{s.t. } \bar{\gamma} = \int_0^1 \gamma(t)\mathrm{d}t \tag{5.48b}$$

$$\text{s.t. } \frac{\dot{\tilde{N}}_t(x)}{\tilde{N}_t(x)} = \alpha \int_0^1 \Phi\left[\frac{\gamma(t)}{\tilde{N}_t(x)}\right] f_{\hat{\gamma}}(x;\gamma(t))\mathrm{d}t \tag{5.48c}$$

$$\text{s.t. } \tilde{N}_t(\infty) = 1 + \alpha\theta\bar{\gamma} \tag{5.48d}$$

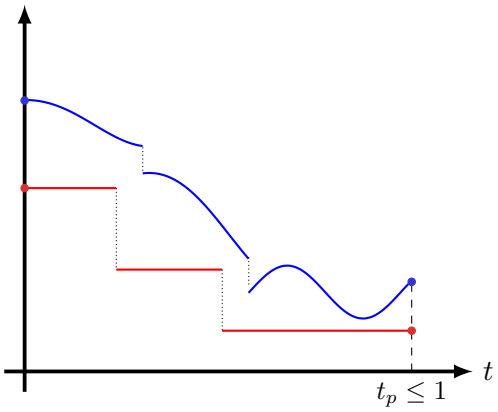

Figure 5.4: Piecewise continuous $\gamma(t)$ candidates with end points 0 and t_p. The blue curve depicts a general piecewise continuously differentiable function, and the red curve a piecewise constant function.

Unlike the variational calculus problems addressed in the previous chapters, this chapter presents an innovative point: the adoption of a broader function space. Such is the case of piecewise continuously differentiable functions

$$\gamma(t) \in \mathcal{C}_p[t_0, t_p] \quad \text{with} \quad t_0 = 0 \ , \ t_p \leq 1 \tag{5.49}$$

comprising at most $p+1$ pieces. The last piece, defined in $t_p < t \leq 1$, corresponds to users allocated to zero energy. Those functions are also denoted in the field of the calculus of variations as *broken extremals* [41, Chapter 3.15]. Some examples of the adopted candidate distributions are depicted in Figure 5.4. In this case, to impose such a function space on (5.48a)–(5.48d), we must replace the integrals defined over the interval $0 \leq t \leq 1$ by multiple integrals, such as for the average energy constraint

$$\sum_{k=1}^{p} \int_{t_{k-1}}^{t_k} \gamma(t) \mathrm{d}t = \bar{\gamma}. \tag{5.50}$$

The reader shall understand that, formally, writing $\int_0^1 (\cdot) \mathrm{d}t$ entails, implicitly, the adoption of continuously differentiable candidates. In this function space, we shall extend the above problem by addressing the optimisation of the user-indices $t_1, \ldots, t_p$ and the number of pieces $p \in \mathbb{N}^+$. Therefore, the following problem needs to be solved:

$$\max_{p \geq 1} \max_{0 < t_1, \ldots, t_p \leq 1} \max_{\gamma(t)} \alpha R \sum_{k=1}^{p} \int_{t_{k-1}}^{t_k} \int_0^{\infty} \mathrm{PSR} \left[\frac{\gamma(t)}{\tilde{N}_t(x)} \right] f_{\hat{\gamma}}(x; \gamma(t)) \mathrm{d}x \mathrm{d}t \tag{5.51a}$$

$$\text{s.t. } \bar{\gamma} = \sum_{k=1}^{p} \int_{t_{k-1}}^{t_k} \gamma(t) \mathrm{d}t \tag{5.51b}$$

$$\text{s.t. } \frac{\dot{\tilde{N}}_t(x)}{\tilde{N}_t(x)} = \alpha \sum_{k=1}^{p} \int_{t_{k-1}}^{t_k} \Phi \left[\frac{\gamma(t)}{\tilde{N}_t(x)} \right] f_{\hat{\gamma}}(x; \gamma(t)) \mathrm{d}t \tag{5.51c}$$

$$\text{s.t. } \tilde{N}_t(\infty) = 1 + \alpha \theta \bar{\gamma}. \tag{5.51d}$$

The first (outer) optimisation needs to be addressed through enumerative search. With regard

to the second and third (inner) problems, the calculus of variations provides a tool for addressing them together. Detailed calculations can be found in Appendix 5.A.2. Summarising the contribution, the stationary point equation reveals that the optimal user-energy distribution $\gamma(t)$ conforms to a piecewise constant structure regardless of the user-indices $t_1, \ldots, t_p$:

$$\gamma(t) = \gamma_k \quad \text{in} \quad t_{k-1} \leq t < t_k \quad \text{for} \quad k = 1, \ldots, p. \tag{5.52}$$

An example of such a structure can be found in Figure 5.4. It should be noted that the above result does not indicate that such a piecewise constant function with $p > 1$ must exist necessarily. The proof of existence implies finding such a function that satisfies the stationary point equation at each interval. In fact, if it does not exist, the optimal distribution would be the continuously differentiable function $\gamma(t) = \bar{\gamma}$ in $0 \leq t \leq 1$, since the function space of continuously differentiable functions is contained inside that of piecewise continuous functions. The interest in this function space applied to our analysis is that, certainly, it is possible to find piecewise continuously differentiable distributions that satisfy the stationary point equations. The unique case in which the optimal $\gamma(t)$ is continuous everywhere is when $\alpha \to 0$, in which case $\gamma(t) = \bar{\gamma}$ in $0 \leq t \leq 1$. The rest of cases need to be evaluated numerically.

Up to this point, this chapter has proved that the stationary points of the spectral efficiency maximization of dynamically ordered SIC follow piecewise constant functions. The result is clear: the exponential energy distribution is only optimal when users are processed in the correct order of received symbol energies. Whenever symbol energies from all users need to be estimated, and they are so at the initial stage, it is better to create groups of users who share the same energy. This way, all users from a given group achieve the same individual performance, and estimating the energy of a particular user becomes the energy estimation of any user from a given group.

5.4.2 Search of the Optimal User-Energy Distribution

The stationary point user-energy distributions constitute piecewise constant functions. The latter includes the global optimum, local optima and some saddle points (if they exist). Based on this statement and given the complexity of the involved equations to determine the globally optimal solution, the next motivation of this section is the proposal of a complexity-affordable search to obtain the best (if possible) piecewise constant function or, failing that, a very competitive locally-optimal solution. Recall that the amplitude (energy), length and the number of intervals of the best distribution remain to be determined. In view of the structure derived for the optimal distribution, our approach particularises the variational calculus problem (5.51a)–(5.51d) for a generic piecewise constant profile that comprises p pieces in the interval $0 \leq t \leq 1$. The following steps are followed:

1. The energy and length of each piece are denoted by vectors

$$\boldsymbol{\gamma} \triangleq [\gamma_1, \ldots, \gamma_p] \tag{5.53}$$

$$\boldsymbol{\Delta} \triangleq [\Delta_1, \ldots, \Delta_p] \quad \text{with} \quad \Delta_{1 \leq i \leq p} = t_i - t_{i-1}, \tag{5.54}$$

for which the average energy constraint $\sum_{k=1}^{p}\int_{t_{k-1}}^{t_k}\gamma(t)\mathrm{d}t=\bar{\gamma}$ (5.51b) turns out into

$$\boldsymbol{\Delta}^T\boldsymbol{\gamma}=\bar{\gamma}\quad;\quad\boldsymbol{\Delta}^T\mathbf{1}_p=1. \tag{5.55}$$

2. The dynamic equation for the noise plus interference term (5.51c) particularised for a $\gamma(t)$ that follows the structure above can also be also turned, after defining the known function

$$B[\boldsymbol{\gamma},N_t,x]\triangleq\Phi\left[\frac{\boldsymbol{\gamma}}{N_t}\right]\odot f_{\hat{\gamma}}(x;\boldsymbol{\gamma}), \tag{5.56}$$

into the integral

$$\tilde{N}_t(x)=\tilde{N}_t(\infty)\exp\left(-\alpha\int_x^\infty\boldsymbol{\Delta}^TB[\boldsymbol{\gamma},\tilde{N}_t(y),y]\mathrm{d}y\right) \tag{5.57a}$$

$$\tilde{N}_t(\infty)=1+\alpha\theta\bar{\gamma}. \tag{5.57b}$$

Even in this case, the integral must be addressed numerically. The approach proposed herein computes a Riemann sum instead of the summation in a continuous interval. The interval $x\in[0,\infty)$ is discretised into M intervals of the same length from $x_0=0$ to $x_{M-1}=x_{\max}$. Then, the term $\tilde{N}_t(x)$ is computed recursively as

$$\tilde{N}_t(x_{M-1})=1+\alpha\theta\bar{\gamma}, \tag{5.58a}$$

$$\tilde{N}_t(x_{i-1})=\tilde{N}_t(x_i)\exp\left(-\frac{\alpha}{M}\boldsymbol{\Delta}^TB[\boldsymbol{\gamma},\tilde{N}_t(x_i),x_i]\right)\quad i=1,\ldots,M-1. \tag{5.58b}$$

3. The ASE defined in (5.51a) is particularised for a $\gamma(t)$ that follows the structure of the optimal user-energy distribution. By defining

$$A[\boldsymbol{\gamma},N_t,x]\triangleq\mathrm{PSR}\left[\frac{\boldsymbol{\gamma}}{N_t}\right]\odot f_{\hat{\gamma}}(x;\boldsymbol{\gamma}), \tag{5.59}$$

and by using the same discretisation as in the second point, the ASE reads

$$\mathrm{ASE}=\frac{\alpha R}{M}\sum_{i=0}^{M-1}\boldsymbol{\Delta}^TA[\boldsymbol{\gamma},\tilde{N}_t(x_i),x_i]. \tag{5.60}$$

Finally, the resulting vector optimization constitutes two maximisations:

$$\max_{p\geq1}\ \max_{\boldsymbol{\gamma},\boldsymbol{\Delta}}\ \frac{1}{M}\sum_{i=0}^{M-1}\boldsymbol{\Delta}^TA[\boldsymbol{\gamma},\tilde{N}_t(x_i),x_i] \tag{5.61a}$$

$$\text{s.t. }\bar{\gamma}=\boldsymbol{\Delta}^T\boldsymbol{\gamma}\quad;\quad1=\boldsymbol{\Delta}^T\mathbf{1}_p \tag{5.61b}$$

$$\text{s.t. }\tilde{N}_t(x_{i-1})=\tilde{N}_t(x_i)\exp\left(-\frac{\alpha}{M}\boldsymbol{\Delta}^TB[\boldsymbol{\gamma},\tilde{N}_t(x_i),x_i]\right)\quad\text{for}\quad i=1,\ldots,M-1 \tag{5.61c}$$

$$\text{s.t. }\tilde{N}_t(x_{M-1})=1+\alpha\theta\bar{\gamma} \tag{5.61d}$$

$$\text{s.t. }\boldsymbol{\gamma},\boldsymbol{\Delta}\geq\mathbf{0}_p \tag{5.61e}$$

$$\text{s.t. }\gamma_i\geq\gamma_{i+1}\quad\text{for}\quad0\leq i\leq p-1 \tag{5.61f}$$

(i) The inner optimisation determines, for a given p, the amplitude and length of each piece. The vector optimisation is solved through sequential quadratic programming.

(ii) The optimal p is found in the outer optimization problem by enumerative search and, since the search is unbounded, a halting criterion is added to its computation. More specifically, the algorithm adopted increases p and obtains locally-optimal distributions until insignificant changes are experimented to the so-far best solution. This may happen when we obtain locally-optimal singular distributions having similar amplitudes ($\gamma_i \approx \gamma_j$ for some i, j), or some interval with practically or null length (some $\Delta_k \approx 0$).

5.4.3 Numerical Results

This section presents the results obtained from the previous algorithm. The worst case for the receiver to carry out the task of ordering users is considered: the symbol energies received from all users are estimated without averaging previous estimates. According to the nomenclature used throughout the chapter, this corresponds to m=1. With regard to the users, transmitted packets are of $n_e = 450$ symbols plus $n_o = 50$ preamble symbols. The adopted coded modulation scheme is a best-performing code of rate R that follows the second order expansion of the maximal channel coding rate. The PER versus SINR curve is

$$\mathrm{PER}[\Gamma] = \mathcal{Q}\left(\sqrt{\frac{n_e}{V(\Gamma)}}\,(C(\Gamma) - R)\right). \tag{5.62}$$

With regard to the receiver, simulations consider the same parameters as in the previous chapters. The average symbol energy is $\bar{\gamma} = 8$dB, the uncancelled energy fraction is $\varepsilon[\Gamma] = \varepsilon = 1\%$, and the decorrelator factor θ is set to 1. With regard to the search algorithm, it is true that it provides very competitive solutions at moderate/low computation complexity. Nonetheless, the obtained distributions are found to be very sensitive to the algorithm's parameters $x_{\max}$ and the number of points M, although their performance are found to be very close to each other. In this respect, the section has considered an academical point of view for these simulations in that, both values are chosen to be sufficiently high so as to obtain the best distributions, to the best of our accuracy. The values chosen are $x_{\max} = 100$ and $M = 5000$[5].

The simulation drawn in Figure 5.5 evaluates ASE versus traffic load α for optimal and dynamic decoding orders, with the aim at evaluating the loss in network performance due to performing the estimation-based ordering of users. The performance of three coded modulation schemes with rates $R_1 = 4/3$, $R_2 = 1$, and $R_3 = 2/3$ bits/symbol are compared. The ASE is shown to be highly penalised when the SIC receiver proceeds according to energy estimates. The lowest performance degradation is achieved when users employ high-rate coding schemes. The reason is because they operate at higher SINRs, and thus, enable better performance to the estimation and the ordering processes. Contrarily, the use of lowest coding rates enables the processing of users at high traffic loads but to the detriment of the performance of estimations.

The optimal number of pieces p at each traffic load is illustrated in Figure 5.6. As shown, it exhibits a complex interaction between unbalancing or not the received user-energy distribution.

[5]For more practical studies, M should be less than a thousand to speed up computations.

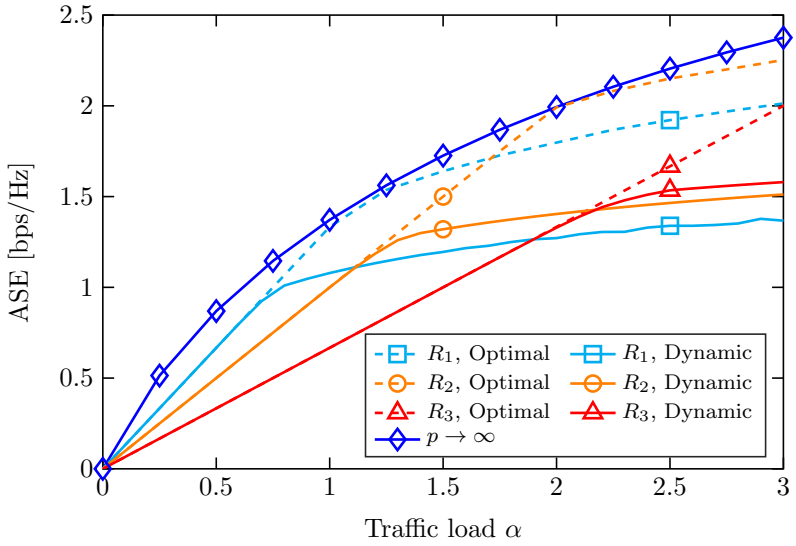

Figure 5.5: ASE with optimal and dynamic decode ordering versus traffic load α. Computations compare system performance achieved with asymptotically many encoders ($p \to \infty$) with that achieved by three theoretical coded modulation systems with rates $R_1 = 4/3$, $R_2 = 1$ and $R_3 = 2/3$ bits/symbol.

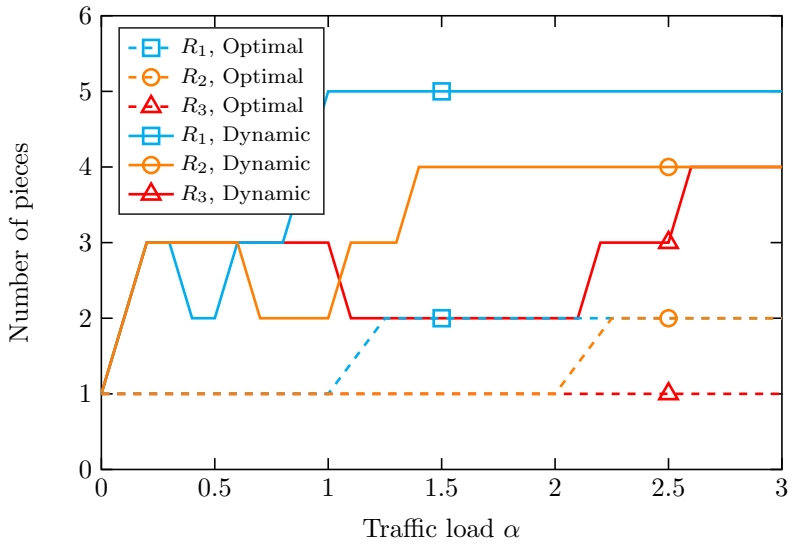

Figure 5.6: Number of pieces p versus traffic load α for optimal and dynamic decode orderings.

SIC benefits from exponentially distributed energies, whereas the impact of energy estimates on SIC performance benefits from uniform energies. At $\alpha \to 0$, the best distribution is the uniform $\gamma(t) = \bar{\gamma}$ regardless of the coding system adopted. At other α, the optimal number of pieces is more intricate; for instance, at low α, the number of pieces increases so as to approach the unbalance of the optimal distribution under optimal decode ordering. However, the SINRs of energy estimates decrease as α increases, and thus, it is better to decrease the number of pieces for a certain range of α until interference is so high that SIC demands a strong energy unbalance to succeed. Remarkably, it is beneficial to use low p to reduce the impact of imperfect estimates on ASE. Some examples of the obtained distributions are depicted in Figures 5.7 and 5.8.

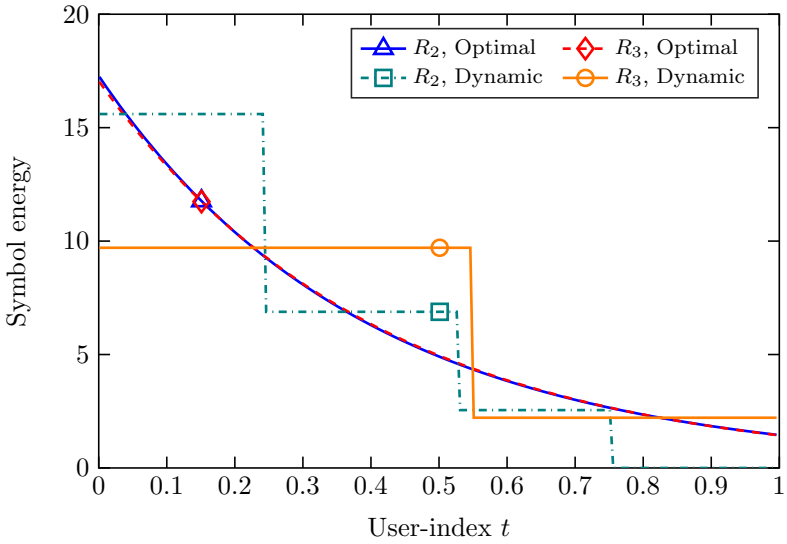

Figure 5.7: Optimal user-energy distribution at $\alpha = 2.0$. The figure compares the profiles with optimal and dynamic decode orderings for the encoders with rates $R_2 = 1$ and $R_3 = 2/3$ bits/symbol.

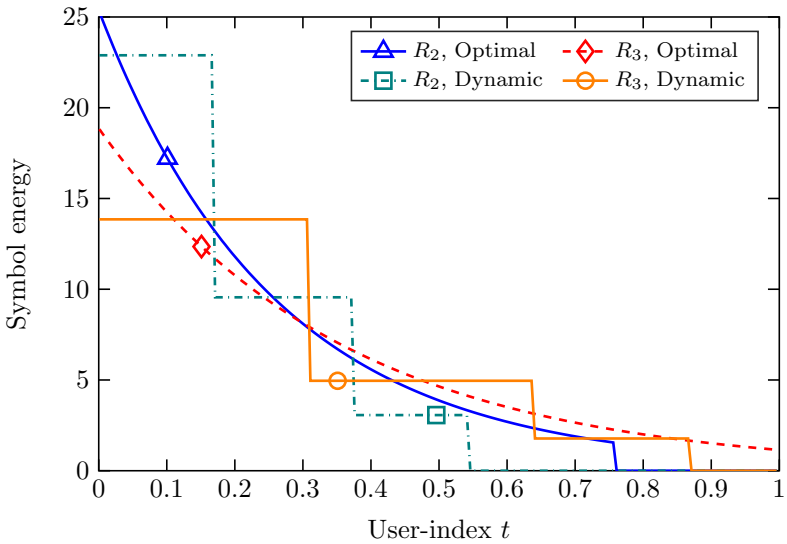

Figure 5.8: Optimal user-energy distribution at $\alpha = 3.0$. The figure compares the profiles with optimal and dynamic decode orderings for the encoders with rates $R_2 = 1$ and $R_3 = 2/3$ bits/symbol.

5.5 Concluding Remarks

This chapter has analysed the impact of dynamic ordering of users on the performance of successive interference cancellation. The work carried out in the present chapter is framed within one of the most relevant points of the demodulator adopted in the Enhanced Spread Spectrum ALOHA system [40], which selects the user estimated with highest signal-to-noise-plus-interference ratio at each decoding stage. The demodulator adopted in this chapter considers the ordering problem at the initial stage, in which all users are considered to be detected and the decode ordering is approached after estimating symbol energies from all users through preamble cross-correlations.

This chapter has proposed and validated a system model to analyse the behaviour of a dynamically ordered successive decoding receiver. The analysis was inspired by the asymptotic large-user system behaviour, in which the order of a particular user only depends on the symbol energy estimated from the same user and not from that corresponding to other users. The system model computes, at each stage, a statistical average that weights the contributions of all users by the probability that each user be ordered at the given stage. The asymptotic large-user regime has been investigated to show tractable expressions of the proposed model. The maximisation of asymptotic spectral efficiency has been tackled by means of the calculus of variations in a space of piecewise continuously differentiable functions. The most striking result is the fact that, effectively, the stationary user-energy distributions are shown to be no longer continuously differentiable, except for the case of vanishing traffic load. Rather, stationary point distributions comprise many pieces. Furthermore, a complexity-affordable search algorithm with high performance/complexity ratio is investigated to obtain competitive distributions.

Appendix 5.A Proofs

Appendix 5.A.1 Asymptotic User-Order Probabilities

This appendix derives the user-asymptotic expression of the probability that user k is ordered at position k'

$$p_{k,k'} = \Pr[\phi_n^{-1}[k] = k']. \tag{5.63}$$

Recall that, as shown in Section 5.2.2.3, the proof for a finite number K of users starts by computing the distribution of the order of user k conditioned on the estimated symbol energy $\hat{\gamma}[k] = x$. This leads, if K is large enough, to the Gaussian distribution

$$\phi_n^{-1}[k]|\{\hat{\gamma}[k]{=}x\} \sim \mathcal{N}(\mu_k, \sigma_k) \tag{5.64}$$

with the mean μ_k and variance σ_k^2 computed as

$$\mu_k = 1 + \sum_{i \neq k} \Pr[\hat{\gamma}[i] > x] \quad , \quad \sigma_k^2 = \sum_{i \neq k} \Pr[\hat{\gamma}[i] > x] \Pr[\hat{\gamma}[i] < x]. \tag{5.65}$$

As a curiosity, the convergence to Gaussianity is very fast. For some tenths of users, the previous assumption holds justifiably. Then, the probability mass function corresponding to the order of user k can be approximated: firstly, by sampling the Gaussian density; and secondly, by taking the average over all possible values of $\hat{\gamma}[k]$; as

$$p_{k,k'} = \int_0^\infty f_{\mathcal{N}}\left(k'; \mu_k(x), \sigma_k^2(x)\right) f_{\hat{\gamma}}(x; \gamma[k]) \mathrm{d}x. \tag{5.66}$$

In the asymptotic large-user regime, due to the definition of the asymptotic indices to index and to order users $t = k/K$ and $t' = k'/K$, the user-order probability $p_{k,k'}$ is turned into that of user t be ordered the t'-th

$$\mathrm{d}p(t, t') = \Pr[\phi_n^{-1}[Kt] = Kt']. \tag{5.67}$$

In addition, the bijection $\phi_n^{-1}[Kt]$ is turned into the asymptotic function $\phi_n^{-1}(t)$, which produces the order of user t in the dense interval $0 \leq t' \leq 1$, as

$$\phi_n^{-1}(t) \triangleq \lim_{K \to \infty} \frac{1}{K} \phi_n^{-1}[Kt]. \tag{5.68}$$

We take as reference the rationales exposed before and show the user-asymptotic behaviour of the expressions above. Firstly, the order of user t conditioned on $\hat{\gamma}(t){=}x$ follows the Gaussian distribution $\phi_n^{-1}(t)|\{\hat{\gamma}(t){=}x\} \sim \mathcal{N}(\bar{\mu}, \bar{\sigma}^2)$ with mean $\bar{\mu}$ and variance $\bar{\sigma}^2$ computed as

$$\bar{\mu}(x) = \lim_{K \to \infty} \frac{1}{K} \left(1 + \sum_{i \neq k} \Pr[\hat{\gamma}[i] > x]\right) = \int_0^1 Q_m\left(\sqrt{\beta\gamma(\tau)}, \sqrt{\beta x}\right) \mathrm{d}\tau, \tag{5.69}$$

$$\bar{\sigma}^2(x) = \lim_{K \to \infty} \frac{1}{K^2} \sigma_{Kt} = 0. \tag{5.70}$$

Therefore, the distribution is deterministic:

$$\phi_n^{-1}(t)|\{\hat{\gamma}(t)=x\} \sim \delta(t-\bar{\mu}(x)). \tag{5.71}$$

The result is clear. In the asymptotic large-user regime, the estimates from other users do not affect the order of a given user. The asymptotic distribution of energy estimates is known, and the receiver only needs to estimate the symbol energy received from user t, $\hat{\gamma}(t)=x$, to determine the position $\mu(x)$ it occupies in the decode ordering. The probability that user t is ordered the t'-th is computed as the differential

$$\mathrm{d}p(t,t') = \mathrm{d}t' \int_0^\infty \delta(t'-\mu(x)) f_{\hat{\gamma}}(x;\gamma(t))\mathrm{d}x. \tag{5.72}$$

Applying the change of variable $\tau \triangleq \mu(x)$ with the differential $\mathrm{d}x = \nabla_\tau \bar{\mu}^{-1}(\tau) = \frac{\mathrm{d}\tau}{\dot{\bar{\mu}}(\bar{\mu}^{-1}(\tau))}$ and the integral limits $\mu(0)=1$ and $\mu(\infty)=0$, we have

$$\begin{aligned}\mathrm{d}p(t,t') &= -\mathrm{d}t' \int_0^1 \delta(t'-\tau) f_{\hat{\gamma}}(\bar{\mu}^{-1}(\tau);\gamma(t)) \tfrac{1}{\dot{\bar{\mu}}(\bar{\mu}^{-1}(\tau))}\mathrm{d}\tau &\text{(5.73a)}\\ &= -f_{\hat{\gamma}}(\bar{\mu}^{-1}(t');\gamma(t)) \tfrac{\mathrm{d}t'}{\dot{\bar{\mu}}(\bar{\mu}^{-1}(t'))}. &\text{(5.73b)}\end{aligned}$$

Appendix 5.A.2 Derivation of the Stationary Point Equation

This appendix derives the stationary point equation associated with the spectral efficiency maximisation of dynamically ordered SIC under an average energy constraint:

$$\begin{aligned}\max_{0<t_1,\ldots,t_p\leq 1} \max_{\gamma(t)} \; & \alpha R \sum_{k=1}^{p} \int_{t_{k-1}}^{t_k} \int_0^\infty \mathrm{PSR}\left[\frac{\gamma(t)}{\tilde{N}_t(x)}\right] f_{\hat{\gamma}}(x;\gamma(t))\mathrm{d}x\mathrm{d}t &\text{(5.74a)}\\ \text{s.t. } & \bar{\gamma} = \sum_{k=1}^{p} \int_{t_{k-1}}^{t_k} \gamma(t)\mathrm{d}t &\text{(5.74b)}\\ \text{s.t. } & \frac{\dot{\tilde{N}}_t(x)}{\tilde{N}_t(x)} = \alpha \sum_{k=1}^{p} \int_{t_{k-1}}^{t_k} \Phi\left[\frac{\gamma(t)}{\tilde{N}_t(x)}\right] f_{\hat{\gamma}}(x;\gamma(t))\mathrm{d}t &\text{(5.74c)}\\ \text{s.t. } & \tilde{N}_t(\infty) = 1+\alpha\theta\bar{\gamma}. &\text{(5.74d)}\end{aligned}$$

The Lagrangian is formulated in the function space of piecewise continuous elements, as

$$\begin{aligned}\mathcal{L}[\gamma(t),\tilde{N}_t(x)] \triangleq & \sum_{k=1}^{p} \int_{t_{k-1}}^{t_k} \int_0^\infty A[\gamma(t),\tilde{N}_t(x),x]\mathrm{d}x\mathrm{d}t - \lambda\left(\sum_{k=1}^{p}\int_{t_{k-1}}^{t_k}\gamma(t)\mathrm{d}t - \bar{\gamma}\right)\\ & - \int_0^\infty \beta(x)\left(\frac{\dot{\tilde{N}}_t(x)}{\tilde{N}_t(x)} - \alpha\sum_{k=1}^{p}\int_{t_{k-1}}^{t_k} B[\gamma(t),\tilde{N}_t(x),x]\mathrm{d}t\right)\mathrm{d}x.\end{aligned} \tag{5.75}$$

with

$$A[\gamma,N_t,x] \triangleq \mathrm{PSR}\left[\frac{\gamma}{N_t}\right] f_{\hat{\gamma}}(x;\gamma) \tag{5.76}$$

$$B[\gamma,N_t,x] \triangleq \Phi\left[\frac{\gamma}{N_t}\right] f_{\hat{\gamma}}(x;\gamma) \tag{5.77}$$

The stationary point equations are found by considering variations $\gamma(t)+v(t)$ and $\tilde{N}_t(x)+v_{N_t}(x)$, and by expanding the functional J up to the first order. The first variation is cast as

$$\begin{aligned}\delta J = &\sum_{k=1}^{p}\int_{t_{k-1}}^{t_k}\left[\int_0^\infty (A_\gamma + \alpha\beta(x)B_\gamma)\,\mathrm{d}x - \lambda\right] v_\gamma(t)\mathrm{d}t \\ &+ \int_0^\infty \left[\sum_{k=1}^{p}\int_{t_{k-1}}^{t_k}(A_{N_t}+\alpha\beta(x)B_{N_t})\,\mathrm{d}t + \frac{\dot{\beta}(x)}{\tilde{N}_t(x)}\right] v_{N_t}(x)\mathrm{d}x \\ &- \frac{\beta(\infty)}{\tilde{N}_t(\infty)}v_{N_t}(\infty) + \frac{\beta(0)}{\tilde{N}_t(0)}v_{N_t}(0).\end{aligned} \tag{5.78}$$

The stationary points are found by satisfying $\delta J = 0$ for any admissible variation $v_\gamma(t)$ and $v_{N_t}(x)$. Two steps are followed:

1. Firstly, the stationary symbol energy profile $\gamma(t)$ is determined by particularising δJ at null variations $v_{N_t}(x) = 0$. We have

$$\sum_{k=1}^{p}\int_{t_{k-1}}^{t_k}\left[\int_0^\infty (A_\gamma + \alpha\beta(x)B_\gamma)\,\mathrm{d}x - \lambda\right] v_\gamma(t)\mathrm{d}t = 0 \tag{5.79}$$

   for all the admissible variations $v_\gamma(t)$. Since we have considered piecewise continuous functions, the Fundamental Lemma of the Calculus of Variations applies at each interval $t_{k-1} \le t < t_k$ for $k = 1, \ldots, p$. This gives the following set of equations

$$\int_0^\infty \Big(A_\gamma[\gamma(t), \tilde{N}_t(x), x] + \alpha\beta(x)B_\gamma[\gamma(t), \tilde{N}_t(x), x]\Big)\,\mathrm{d}x = \lambda \quad 0 \le t < t_1 \tag{5.80a}$$

$$\vdots$$

$$\int_0^\infty \Big(A_\gamma[\gamma(t), \tilde{N}_t(x), x] + \alpha\beta(x)B_\gamma[\gamma(t), \tilde{N}_t(x), x]\Big)\,\mathrm{d}x = \lambda \quad t_{k-1} \le t < t_k \tag{5.80b}$$

$$\vdots$$

$$\int_0^\infty \Big(A_\gamma[\gamma(t), \tilde{N}_t(x), x] + \alpha\beta(x)B_\gamma[\gamma(t), \tilde{N}_t(x), x]\Big)\,\mathrm{d}x = \lambda \quad t_{p-1} \le t < t_p \tag{5.80c}$$

   The interesting observation in all of them is that, all the equations only depend on $\gamma(t)$ and not on t inside their respective intervals. The conclusion is that the stationary point symbol energy profiles $\gamma(t)$ follow piecewise constant structures:

$$\gamma(t) = \gamma_k \quad \text{in} \quad t_{k-1} \le t < t_k \quad \text{for} \quad k = 1, \ldots, p. \tag{5.81}$$

2. Secondly, the stationary point for the noise plus interference term $\tilde{N}_t(x)$ is found by particularising δJ at $v_\gamma(t)T = 0$ and $v_{N_t}(\infty) = v_{N_t}(0) = 0$. This gives, after the application of the fundamental lemma of the calculus of variations

$$-\sum_{k=1}^{p}\int_{t_{k-1}}^{t_k}\Big(A_{N_t}[\gamma(t), \tilde{N}_t(x), x] + \alpha\beta(x)B_{N_t}[\gamma(t), \tilde{N}_t(x), x]\Big)\,\mathrm{d}t = \frac{\dot{\beta}(x)}{\tilde{N}_t(x)}. \tag{5.82}$$

   Since $\tilde{N}_t(\infty)$ does not vary, $v_{N_t}(\infty) = 0$. Thus, the initial value of $\dot{\beta}(x)$ is $\beta(0) = 0$.

6 Conclusions and Future Work

6.1 Conclusions

This thesis has dealt with the design of energy and code allocation strategies for the uplink of a massively populated network in which the receiver leverages interference cancellation to provide high throughput and reliable multiple access. The scenario under evaluation is contextualised in the framework of massive machine-type communications, where many devices run a latency tolerant application and rely on a transparent satellite connected to a ground-based gateway over the downlink. This thesis has addressed two main problems in this framework:

1. The derivation of system models that characterise statistically the behaviour of practical successive interference cancellation (SIC) systems. This thesis has explored the application of the key design aspects of the demodulator adopted in the Enhanced Spread Spectrum ALOHA system, which incorporates asynchronous spread spectrum transmissions together with a powerful packet-based demodulation algorithm based on iterative SIC. This demodulator fully extends the one considered in classical information-theoretic analyses, as it combines the packet detection process with the ordering of users and the iterative decoding-cancellation system implementation for short-length codes. A particular challenge has been to establish a model embodying the above features. The present thesis has approached the more complex analysis of the joint system by incorporating such features gradually along three chapters.

2. The design of allocation strategies for a large number of users. The known analytical findings corresponding to the maximisation of the network sum-rate reveal that all users must transmit at the same rate and arrive with exponentially distributed symbol energies. This thesis has explored in depth the already known unbalance between energy and coding rate of different users by incorporating analytical findings for short-length codes. In this respect, when the asymptotic distribution of channel power gains from all users is known, the analytic form of the best allocation strategy has been determined leveraging the user-asymptotic regime of the previous system model. More specifically, this thesis has addressed the problem of allocating energy and code to asymptotically many users by taking advantage of the calculus of variations over function spaces. Contrarily to vector

optimisation, the difficulty of this problem lies in considering a function space generic enough, so as to enable discontinuous candidates to be potentially optimal solutions to the stated problems. One of the main contributions of this thesis has been the bet on discontinuous (piecewise continuously differentiable) functions to solve the problems of energy and code allocation.

This thesis has been organised into five chapters, whose conclusions are summarised below chapter by chapter. The introductory part has been divided into two chapters:

1. The research challenges and overview on massive multiple access in the context of satellite communications have been summarised in Chapter 1. The use of non-orthogonal multiple access jointly with the exploitation of the collision domain by the receiver have been identified as major objectives to counterbalance the increasing network density. The analysis in this thesis has been focused on the uplink; two relevant random access protocols that solve such a problem effectively have been outlined throughout the chapter. This thesis has selected one of them for study: the Enhanced Spread Spectrum ALOHA protocol. In this scheme, users employ coding systems for short packets and non-orthogonal spreading waveforms, and the receiver handles multiple packet collisions under an innovative iterative multiuser decoding algorithm. The specific features of such a multiuser receiver have been discussed as motivation for introducing posterior chapters.
2. A theoretical perspective on Gaussian multiple access channels (MAC) under infinite and finite blocklength constraints has been analysed in Chapter 2. The first part of the chapter has reviewed pioneering analyses on orthogonal and non-orthogonal multiple access approaches to the Gaussian MAC, with specific application to broadband satellite services under spreading-based multiple access and many users. The second part of the chapter has been devoted to the performance analysis of multiuser receivers for randomly spread users with optimal detection and successive decoding. In addition, the chapter has also discussed the main practical features of successive decoding corresponding to the decoding of short packets, imperfectly cancelled users, and the user decode ordering.

The substance of the work carried out in this thesis revolves around the practical features of the iterative demodulator adopted in the Enhanced Spread Spectrum ALOHA system:

3. The first iteration of an imperfect SIC receiver has been analysed in Chapter 3. The non-idealities of channel decoding for short packets and imperfect cancellation have been studied for a SIC receiver aided by redundancy-check error control. The model adopted for such a receiver has addressed the characterisation of both features through the known packet error rate (PER) and the residual energy (RE) curves versus signal-to-interference-plus-noise ratio (SINR). Under this policy, the events of packet success and packet failure corresponding to users previously processed are propagated throughout the stages of SIC, totalling an exponential number of possible combinations. This thesis has shown that a convenient approach to address such an issue is by analysing the user-asymptotic regime, in which case the above model has resulted into a set of differential or integral equations that describe the evolution of noise plus interference throughout the processing stages.

   The allocation designs accounting for the first iteration of such a receiver have been

derived by resorting to the calculus of variations for a broad variety of energy, rate, and conjoint designs. The commonest result after all the designs is that SIC benefits much more from unbalancing energy, rate and reliability rather than from allocating uniformly one or more of them to users. The optimal allocations share a common trait: they belong to a metric space of discontinuous functions that allocate null energy to weakest users.

4. The previous analysis and the system-based optimization presented in the previous chapter have been extended in Chapter 4 by taking up an iterative approach to the adopted SIC receiver. More specifically, the iterative SIC receiver bases its operation on further iterations of the demodulator adopted in the Enhanced Spread Spectrum ALOHA system, in which users decoded unsuccessfully are processed persistently. Remarkably, the system has been modelled as a cascade of SIC schemes that operate, at each iteration, with less users and variables statistically dependent on previous iterations. For system modelling, this has entailed the introduction of multivariable functions for user decoding. The user-asymptotic regime is analysed to facilitate the mathematical treatment of the previous system model, and also, to enable the full understanding of the sensitivity of the adopted iterative receiver to system model parameters.

   The allocation designs corresponding to this chapter have explored the unbalance between energy and code allocated to users with fair and optimal reliability. Iterative SIC has been shown to outperform substantially the non-iterative SIC analysed in the previous chapter when the traffic load is sufficiently high and the coded modulation schemes employed by users are less competitive. Moreover, iterative SIC has shown an overwhelming performance in all the scenarios analysed if the number of users and iterations are sufficiently high, as it practically provides error-free decoding performance to all users.

5. The impact of estimated symbol energies has been studied throughout Chapter 5. To that aim, it has been considered that the receiver does not know the identities of all users and that no channel state information is provided to the receiver. The adopted receiver then performs SIC after estimating the symbol energies received from all users. This analysis corresponds to a simplified approach of the demodulator adopted in Enhanced Spread Spectrum ALOHA. The challenge in this chapter has been the derivation of an accurate system model for such a receiver implementation. Through an extensive development, the model is based on the behaviour of such a system in the user-asymptotic regime. Summarising the contribution, the long-term average expressions have been substituted by statistical averages that weight the contribution of all users by the known ordering probabilities. The user-asymptotic regime has been studied in depth to simplify the analytical complexity of the system model. More specifically, the analysis has been turned into deterministic, and the ordering probabilities have shown a very simple form in terms of the density of energy estimates from preamble cross-correlations.

   The asymptotic optimisation of the above system has been useful to identify the best-performing energy allocation for the user-aggregate spectral efficiency. It has been shown that the stationary point distributions have the structure of piecewise constant functions, that decrease the impact of energy estimation on the task of ordering users, at the same time that unbalance user-energies to benefit the performance of the SIC receiver.

6.2 Potential Topics for Future Research

Following the completion of this thesis, the next research lines have been identified as potential topics to continue the studies initiated in the last three chapters:

1. **An accurate system model for less users**. This thesis has focussed mainly on settings involving many users as a direct observation of current massive machine-type communication scenarios. More specifically, the system models corresponding to the adopted receivers have resorted to the asymptotic large-user regime to deal with the randomness present in packet decoding success and failure events. The limitation of this work, then, is the application of our system models and asymptotic allocations for a finite number of users. In this respect, as it has been demonstrated throughout the present thesis, we must have in the order of a few hundred users to validate the asymptotic expressions.

   A future research line would consist in introducing some corrections in the asymptotic computations to validate such expressions for even less users.

The subsequent future research lines put more attention into practical implementation aspects of the system considered:

2. **The detection of users**. Taking the demodulator adopted in the Enhanced Spread Spectrum ALOHA system as reference, one of the aspects that has been left unstudied is the packet detection process. Recall that such a demodulator combines the detection and ordering of users with the successive decoding implementation. In this respect, the present thesis has studied the above features when all transmitting users are detected.

   An interesting study reserved for future research would be the analysis and the introduction of the packet detection algorithm to the system model and to the allocation design procedure proposed in this thesis. For the one-iteration SIC discussed in Chapter 3, we believe it is relatively straightforward. The univariate PER function can be modified to include the process of detecting users. However, its analysis for iterative SIC can be more complex since the decoding algorithm can be intertwined with the detection of new users.

3. **The iterative ordering and decoding of users**. Another future research topic concerns Chapter 5, where system analysis and performance has been analysed for a simple study case in which the estimation and the ordering of users are decoupled from successive decoding. According to the implementation adopted in the Enhanced Spread Spectrum ALOHA system, the performance of the receiver can be improved further if the estimation task is combined with the decoding of users (see Section 5.3.3 for more details).

   The analysis pursued in Chapter 5 has been found to be consistent with the above policy. Then, the future work would be determining an analytical expression for the user-order probabilities corresponding to the new ordering and cancellation policy.

4. **A model for sliding-window based iterative SIC**. As described in Chapter 1, the highest performing system performs a sliding window-based interference cancellation approach that overlaps many windows within a burst frame. For simplicity, more simplified cases have been analysed throughout this thesis. A very interesting future work involves the analysis and modelling of the more complex approach described extensively in [40].

Bibliography

[1] Ericsson, "Ericsson mobility report," Nov. 2020.

[2] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 905–929, 2020.

[3] F. Rinaldi, H. L. Maattanen, J. Torsner, S. Pizzi, S. Andreev, A. Iera, Y. Koucheryavy, and G. Araniti, "Non-terrestrial networks in 5G beyond: A survey," *IEEE Access*, vol. 8, pp. 165 178–165 200, 2020.

[4] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Network*, vol. 35, no. 2, pp. 244–251, 2021.

[5] A. Wang, L. Lei, E. Lagunas, A. I. Pérez-Neira, S. Chatzinotas, and B. Ottersten, "NOMA-enabled multi-beam satellite systems: Joint optimization to overcome offered-requested data mismatches," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 900–913, 2021.

[6] J. Chu, X. Chen, C. Zhong, and Z. Zhang, "Robust design for NOMA-based multibeam LEO satellite internet of things," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1959–1970, 2021.

[7] Y. Couble, C. Rosenberg, E. Chaput, J.-B. Dupé, C. Baudoin, and A.-L. Beylot, "Two-color scheme for a multi-beam satellite return link: Impact of interference coordination," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 5, pp. 993–1003, 2018.

[8] M. Berioli, G. Cocco, G. Liva, and A. Munari, *Modern Random Access Protocols.* Now Foundations and Trends, 2016.

[9] H. Peyravi, "Medium access control protocols performance in satellite communications," *IEEE Commun. Mag.*, vol. 37, no. 3, pp. 62–71, 1999.

[10] M. Luglio, C. Roseti, and F. Zampognaro, "Performance evaluation of TCP-based applications over DVB-RCS DAMA schemes," *Int. Journal of Satellite Communications and Networking*, vol. 27, no. 6, pp. 349–349, 2009.

[11] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 2018.

[12] C. Liu and D. Liang, "Heterogeneous networks with power-domain NOMA: Coverage, throughput, and power allocation analysis," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3524–3539, 2018.

[13] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2017.

[14] S. Verdú, *Multiuser detection*. Cambridge University Press, 1998.

[15] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Info. Theory*, vol. 45, no. 2, pp. 622–640, 1999.

[16] M. T. P. Le, G. C. Ferrante, T. Q. S. Quek, and M. Di Benedetto, "Fundamental limits of low-density spreading NOMA with fading," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4648–4659, 2018.

[17] H. Asgharimoghaddam, J. Kaleva, and A. Tölli, "Capacity approaching low density spreading in uplink NOMA via asymptotic analysis," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1635–1649, 2021.

[18] J. Bao, Z. Ma, Z. Ding, G. K. Karagiannidis, and Z. Zhu, "On the design of multiuser codebooks for uplink SCMA systems," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 1920–1923, 2016.

[19] T. Cola, D. Tarchi, and A. Vanelli-Coralli, "Future trends in broadband satellite communications: Information centric networks and enabling technologies," *Int. Journal of Satellite Communications and Networking*, vol. 33, no. 5, pp. 473–490, 2015.

[20] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *2006 IEEE Information Theory Workshop (ITW)*, 2006, pp. 160–164.

[21] R. R. Müller and S. Verdú, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1429–1441, 2001.

[22] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of M2M massive access in wireless cellular networks," *Digital Communications and Networks*, vol. 1, no. 1, pp. 1–19, Feb. 2015.

[23] M. T. Islam, A. M. Taha, and S. Akl, "A survey of access management techniques in machine type communications," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 74–81, 2014.

[24] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, 2015.

[25] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Springer, 2018.

[26] R. De Gaudenzi, O. Del Río Herrero, G. Gallinaro, S. Cioni, and P.-D. Arapoglou, "Random access schemes for satellite networks: from VSAT to M2M: A survey," *Int. Journal of Satellite Communications and Networking*, vol. 36, no. 1, pp. 66–107, 2018.

[27] N. Abramson, "THE ALOHA SYSTEM: another alternative for computer communications," in *Proceedings of the November 17-19, 1970, fall joint computer conference*, vol. 37, 1970, pp. 281–285.

[28] L. Kleinrock and S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. Commun.*, vol. 23, no. 4, pp. 410–423, 1975.

[29] "IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements - Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," *IEEE Std 802.11-2007*, pp. 1–1076, Jun. 2007.

[30] EN ETSI, "Digital video broadcasting (DVB); Second generation DVB interactive satellite system (DVB-RCS2); Part 2: Lower layers for satellite standard," *ETSI EN 301 545 V1.1.1*, Jan. 2012.

[31] TS ETSI, "Satellite Earth stations and systems; Air interface for S-band mobile interactive multimedia (S-MIM); Part 4: Physical layer specification, return link synchronous access," *ETSI TS 102 721–4 V1.1.1*, Dec. 2011.

[32] ——, "Satellite Earth stations and systems; Air interface for S-band mobile interactive multimedia (S-MIM); Part 3: Physical layer specification, return link asynchronous access," *ETSI TS 102 721–3 V1.1.1*, Dec. 2011.

[33] E. Casini, R. De Gaudenzi, and O. Del Río Herrero, "Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1408–1419, 2007.

[34] O. Del Río Herrero and R. De Gaudenzi, "Generalized analytical framework for the performance assessment of slotted random access protocols," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 809–821, 2014.

[35] R. De Gaudenzi, O. Del Río Herrero, G. Acar, and E. G. Barrabés, "Asynchronous contention resolution diversity ALOHA: Making CRDSA truly asynchronous," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6193–6206, 2014.

[36] F. Clazzer and C. Kissling, "Enhanced contention resolution aloha - ECRA," in *9th Int. ITG Conference on Systems, Communication and Coding*, Jan. 2013, pp. 1–6.

[37] F. Clazzer, C. Kissling, and M. Marchese, "Enhancing contention resolution ALOHA using combining techniques," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2576–2587, 2018.

[38] O. Del Río Herrero, R. De Gaudenzi, and J. P. Vidal, "Design guidelines for advanced random access protocols," in *the Proc. of the 30-th AIAA Int. Satellite Systems Communications Conference*, Sep. 2012.

[39] A. Reichman, "Enhanced spread spectrum aloha (E-SSA), an emerging satellite return link messaging scheme," in *2014 IEEE 28th Convention of Electrical Electronics Engineers in Israel (IEEEI)*, Dec. 2014, pp. 1–4.

[40] R. De Gaudenzi, O. Del Río Herrero, and G. Gallinaro, "Enhanced spread ALOHA physical layer design and performance," *Int. Journal on Satellite Communications and Networking*, vol. 32, no. 6, pp. 457–473, 2014.

[41] I. M. Gelfand, R. A. Silverman *et al.*, *Calculus of variations*. Courier Corporation, 2000.

[42] L. M. Graves, "Discontinuous solutions in space problems of the calculus of variations," *American Journal of Mathematics*, vol. 52, no. 1, pp. 1–28, 1930.

[43] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[44] S. Verdú, "Minimum probability of error for asynchronous gaussian multiple-access channels," *IEEE Trans. Info. Theory*, vol. 32, no. 1, pp. 85–96, 1986.

[45] P. Salvo Rossi, K. Kansanen, R. R. Müller, and C. Rachinger, "Power randomization for iterative detection over random-access fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5704–5713, 2015.

[46] G. Caire, S. Guemghar, A. Roumy, and S. Verdú, "Maximizing the spectral efficiency of coded CDMA under successive decoding," *IEEE Trans. Info. Theory*, vol. 50, no. 1, pp. 152–164, 2004.

[47] F. Collard and R. De Gaudenzi, "On the optimum packet power distribution for spread ALOHA packet detectors with iterative successive interference cancelation," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6783–6794, 2014.

[48] D. V. Djonin and V. K. Bhargava, "Asymptotic analysis of the conventional decision feedback receiver in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 1066–1078, 2003.

[49] J. Sala, J. Villares, and F. Rey, "Asymptotic and finite user PER analysis of successive interference cancellation for DS-CDMA," *IEEE Commun. Let.*, vol. 15, no. 11, pp. 1145–1147, 2011.

[50] S. P. Weber, J. G. Andrews, X. Yang, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with successive interference cancellation," *IEEE Trans. Info. Theory*, vol. 53, no. 8, pp. 2799–2814, 2007.

[51] J. Blomer and N. Jindal, "Transmission capacity of wireless ad hoc networks: Successive interference cancellation vs. joint detection," in *2009 IEEE Int. Conference on Communications (ICC)*, 2009, pp. 1–5.

[52] E. MolavianJazi and J. N. Laneman, "A second-order achievable rate region for gaussian multi-access channels via a central limit theorem for functions," *IEEE Trans. Info. Theory*, vol. 61, no. 12, pp. 6719–6733, 2015.

[53] 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, 2010.

[54] J. Scarlett, V. Y. F. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive non-gaussian channels," *IEEE Trans. Info. Theory*, vol. 63, no. 1, pp. 81–92, 2017.

[55] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultra-reliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.

[56] T. Erseghe, "Coding in the finite-blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations," *IEEE Trans. Info. Theory*, vol. 62, no. 12, pp. 6854–6883, 2016.

[57] G. Liva, L. Gaudio, T. Ninacs, and T. Jerkovits, "Code design for short blocks: A survey," *ArXiv*, vol. abs/1610.00873, 2016.

[58] E. MolavianJazi and J. N. Laneman, "A random coding approach to gaussian multiple access channels with finite blocklength," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 286–293.

[59] G. Caire, R. R. Müller, and T. Tanaka, "Iterative multiuser joint decoding: optimal power allocation and low-complexity implementation," *IEEE Trans. Info. Theory*, vol. 50, no. 9, pp. 1950–1973, 2004.

[60] L. Cottatellucci, R. R. Muller, and M. Debbah, "Asynchronous CDMA systems with random spreading—part II: Design criteria," *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1498–1520, 2010.

[61] A. Agrawal, J. G. Andrews, J. M. Cioffi, and T. Meng, "Iterative power control for imperfect successive interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 878–884, 2005.

[62] F. Molina, J. Sala-Álvarez, J. Villares, and F. Rey, "Optimal power control law for equal-rate DS-CDMA networks governed by a successive soft interference cancellation scheme," in *2018 IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 1–5.

[63] R. De Gaudenzi and O. Del Río Herrero, "Advances in random access protocols for satellite networks," in *2009 Int. Workshop on Satellite and Space Communications*, Sep. 2009, pp. 331–336.

[64] Y. Xu, C. Shen, T. Chang, S. Lin, Y. Zhao, and G. Zhu, "Transmission energy minimization for heterogeneous low-latency NOMA downlink," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1054–1069, 2020.

[65] Y. Xu, C. Shen, D. Cai, and G. Zhu, "Latency constrained non-orthogonal packets scheduling with finite blocklength codes," *IEEE Trans. Veh. Tech.*, vol. 69, no. 10, pp. 12 312–12 316, 2020.

[66] A. U. Toboso, S. Loyka, and F. Gagnon, "Optimal detection ordering for coded V-BLAST," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 100–111, 2014.

[67] F. Molina, J. Sala-Alvarez, F. Rey, and J. Villares, "Channel-aware energy allocation for throughput maximization in massive low-rate multiple access," in *2019 IEEE Int. Conference on Communications (ICC)*, 2019, pp. 1–6.

[68] F. Molina, J. Sala-Álvarez, J. Villares, and F. Rey, "Joint energy and rate allocation for successive interference cancellation in the finite blocklength regime," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 1045–1049.

[69] TS ETSI, "Universal mobile telecommunications systems (UMTS); Multiplexing and channel coding (FDD)," *ETSI TS 125 212 v12.1.0 Release 12*, Jan. 2015.

[70] G. M. Ewing, *Calculus of variations with applications.* Courier Corporation, 1985.

[71] F. Molina and J. Sala-Álvarez, "Average PER performance metrics of iterative successive interference cancellation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 74–77, 2020.

[72] P. Li, R. C. d. Lamare, and R. Fa, "Iterative successive interference cancellation based on multiple feedback for multiuser MIMO systems," in *17th European Wireless 2011 - Sustainable Wireless Technologies*, 2011, pp. 1–6.

[73] L. Yuan, Z. Zheng, N. Yang, and J. Zhang, "Performance analysis of short-packet non-orthogonal multiple access with Alamouti space-time block coding," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2900–2905, 2021.

[74] ETSI EN, "Digital video broadcasting (DVB); Second generation DVB Interactive satellite system (DVB-RCS2); Part 2: Lower layers for satellite standard," *ETSI EN 301 545-2 V1.2.1*, Apr. 2014.

[75] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.

[76] F. Molina and J. Sala-Álvarez, "Rate allocation in massive multiple access combining successive decoding with error control," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, 2020, pp. 1059–1063.

[77] F. Molina and J. Sala-Álvarez, "Asymptotic performance analysis of successive interference cancellation with dynamic user-decoding order," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2931–2935, 2020.

[78] ——, "Discontinuous user-energy distribution for dynamically ordered successive interference cancellation," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1673–1677, 2021.

[79] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[80] J. Wang, B. Xia, K. Xiao, Y. Gao, and S. Ma, "Outage performance analysis for wireless non-orthogonal multiple access systems," *IEEE Access*, vol. 6, pp. 3611–3618, 2018.

[81] Y. Gao, B. Xia, K. Xiao, Z. Chen, X. Li, and S. Zhang, "Theoretical analysis of the dynamic decode ordering SIC receiver for uplink NOMA systems," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2246–2249, 2017.

[82] Y. Gao, B. Xia, Y. Liu, Y. Yao, K. Xiao, and G. Lu, "Analysis of the dynamic ordered decoding for uplink NOMA systems with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6647–6651, 2018.

[83] V.-L. Dao, L.-N. Hoang, S. Girs, and E. Uhlemann, "Outage performance of pairwise NOMA allowing a dynamic decoding order and optimal pairs of power levels," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1886–1906, 2020.

[84] X. Xie, Y. Wu, J. Gao, and W. Zhang, "Massive unsourced random access for massive MIMO correlated channels," in *2020 IEEE Global Communications Conference (GLOBECOM)*, 2020, pp. 1–6.

[85] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive access for future wireless communication systems," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 148–156, 2020.

[86] Y. Polyanskiy, "A perspective on massive random-access," in *2017 IEEE Int. Symposium on Information Theory (ISIT)*, 2017, pp. 2523–2527.