Universidade de Vigo

Signal Theory and Communications Department

Universida$_{\text{de}}$Vigo

DOCTORAL DISSERTATION
INTERNATIONAL MENTION

On-Ground Processing for Spectral Efficiency Improvement in Next Generation Satellite Communications

Author:

Jesús Arnau Yáñez

Directed by:

Carlos Mosquera Nartallo

2014

DOCTORAL DISSERTATION
INTERNATIONAL MENTION

On-ground processing for spectral efficiency improvement in next generation satellite communications

Author:

Jesús Arnau Yáñez

Directed by:

Carlos Mosquera Nartallo

EXAMINATION COMMITTEE

Board members: Prof. Ana García Armada (President)

Prof. Roberto López Valcárcel (Secretary)

Dr. Pantelis-Daniel Arapoglou

Thesis defense: Dec. 16, 2014.

This edition: April 2015.

Agradecementos

"Thankfulness or Gratitude is the desire or zeal springing from love, whereby we endeavour to benefit him, who with similar feelings of love has conferred a benefit on us". Baruch SPINOZA, *Ethics*, Proposition XXXIV.

Se Spinoza tiña razón ao definir a gratitude, entonces *obras son amores*, como di o refrán, e pouco valen unhas liñas como agradecemento. Pero por iso non deixan de ser necesarias, e quero aproveitar este espazo para recoñecer a algúns dos que me prestaron a súa axuda, dun xeito ou doutro, durante estes anos.

Agradézolle en primeiro lugar a Carlos Mosquera toda a súa axuda como titor, orientador, xefe e colaborador, sen a cal nada disto tería sido posible. Tamén a Nuria o feito de terme informado tan xenerosamente sobre as liñas de traballo do grupo cando aínda eu non rematara a carreira. Tanto a eles dous como a Roberto, Pedro e Fernando, teño que agradecerlles o excelente ambiente que existe dentro do grupo, e a influencia que a súa docencia tivo sobre min. Unha mención especial merecen Carmen Touriño e a súa dilixencia resolvendo os meus sucesivos enredos burocráticos.

Because of my involvement in different projects, I have had the opportunity to work with many fantastic people. I would like to thank the whole crew from the SatNEx and NGW projects, and specially Bertrand Devillers, Ana Pérez, Michael Bergmann, and Wilfried Gappmair, with whom we held a closer collaboration. I am also grateful I met and worked with Dimitrios Christopoulos, Symeon Chatzinotas, Prof. Ottersten, and the rest of the people from SnT.

I would also like to thank the Communications and TT&C section at ESTEC/ESA for the encouragement and support they provided during (and well beyond) my stay at their premises, and in particular to Daniel Arapoglou, Riccardo de Gaudenzi, and Alberto Ginesi.

Non me esquezo tampouco dos compañeiros de laboratorio nin da agradable compaña que foron todo este tempo: Dani, David, Alberto, Massimo, Bamrung, Jorge, Roi, Cristian, e moi especialmente Alberto Rico, con quen traballei máis.

Por último, quero agradecerlle á miña familia todo e apoio que me deron, dende sempre e en especial nestes anos, e a Patri os moitos ánimos e a súa notable paciencia comigo.

Resumen

La eterna demanda de una explotación del espectro más eficiente para los satélites de próxima generación requiere del estudio de técnicas que puedan suponer una potencial revolución. Esta tesis se centra en dos de los problemas principales que afectan a las prestaciones de los sistemas de comunicaciones vía satélite.

En la primera parte de la tesis, se estudia la detección multiusuario en el canal de retorno para satélites multihaz que empleen esquemas de reutilización de frecuencia agresivos. Se comienza comparando sus prestaciones con las de un escenario tradicional, y se estudia la posibilidad de dividir el procesamiento necesario entre la estación base y el satélite para hacer frente a ciertas limitaciones prácticas. A continuación, se estudian los límites de capacidad fundamentales de estos sistemas cuando el canal está afectado por atenuación por lluvia, y las atenuaciones experimentadas en los distintos haces están correlacionadas; se obtienen expresiones analíticas para los regímenes de SNR alta y baja. Por último, se considera la estimación imperfecta del canal y se obtiene, a partir de resultados de teoría de matrices aleatorias, una aproximación al error residual en función de los parámetros del sistema.

En la segunda parte, se hace hincapié en la adaptación del enlace en comunicaciones satélite móviles. Para el canal de retorno, se investiga el potencial de la información de canal de bucle abierto, la cuál es actual pero poco precisa; el descubrimiento de que esta información es mejor que la tradicional (de bucle cerrado) en ciertos escenarios motiva el estudio de un algoritmo adaptativo que dinámicamente pondere ambas. Por último, para el canal de ida se estudia el caso en el que se pueden usar retransmisiones, y se diseña la secuencia óptima de parámetros a emplear en las sucesivas retransmisiones de un paquete, con vistas a maximizar la tasa de transmisión manteniendo la tasa de paquetes perdidos bajo un determinado umbral.

Motivación

La demanda de datos inalámbricos no ha hecho más que crecer durante los últimos años, y se prevé que continúe así en el futuro próximo; de acuerdo con [1], durante 2013 se produjo un incremento de tráfico cercano al 81%, y se espera un incremento del 1000% de aquí al final de 2018. Para soportar este crecimiento, ha sido necesario crear nuevos estándares de comunicaciones que mejorasen la eficiencia espectral de las redes inalámbricas.

El mundo de las comunicaciones vía satélite no ha sido una excepción a este crecimiento, aún a pesar de su supuesta inferioridad frente a tecnologías de comunicaciones terrestres. Una de las razones para ello es que las redes terrestres son caras de desplegar en muchas áreas apartadas, y en consecuencia los sistemas de banda ancha del futuro prevén integrar infraestructuras tanto terrestres como satélite [2]. Otra de las razones es que los operadores de satélites buscan mantener la competitividad de sus servicios incluso frente a las modernas tecnologías inalámbricas terrestres, y ello requiere invertir para conseguir ofrecer servicios mejorados a un precio igual (o menor). Tal y como han escrito ciertos profesionales:

> "Los usuarios demandarán servicios interactivos mejorados al mismo precio [...].
> Pueden los satélites soportar esta evolución en la demanda?" [3].

La situación descrita ha influido en los programas de muchas de las instituciones involucradas, como se puede ver por ejemplo en los objetivos Horizonte 2020 de la Unión Europea, y ha estimulado la financiación de proyectos de investigación internacionales y la propuesta de nuevos estándares; como ejemplo de ello, en los últimos años se ha visto, entre otros, la transición de DVB-S a DVB-S2/DVB-RCS en difusión satélite [4], y los esfuerzos para la estandarización de S-UMTS. Sin embargo, y a pesar de que se han alcanzado importantes mejoras – a través de, por ejemplo, modulaciones, códigos de canal y filtros de transmisión más eficientes – sigue habiendo una gran necesidad de explotar mejor el espectro radioeléctrico y de mantener la competitividad de las comunicaciones vía satélite.

En esta tesis, se estudian dos soluciones orientadas a dos escenarios de comunicación por satélite distintos, pero en ambos con énfasis en el canal que va de los usuarios hacia la estación receptora (canal de retorno, RL). La primera solución aplica a servicios satélite fijos (*fixed satellite services*, FSS) en los que un único satélite sirve una población grande de usuarios a través de múltiples haces de antena; para este caso, se estudia la posibilidad de mitigar en el lado receptor la interferencia que aparece, y se proporcionan herramientas para analizar las prestaciones finales y ajustar parámetros relevantes. La segunda solución aplica a servicios satélite móviles (*mobile satellite services*, MSS) y busca adaptar dinámicamente los parámetros de la transmisión a las condiciones cambiantes del canal; para este caso, se proponen nuevas estrategias de adaptación tanto para el RL como para el FL, y se analizan sus prestaciones.

Mitigación de interferencias en sistemas satélite multihaz

Introducción

Las zonas de cobertura satélite de hoy en día están formadas fundamentalmente por pequeños haces de antena, en vez de una gran y única zona radiada. Esto permite proveer de servicios con mejor calidad en términos de relación portadora a interferencia (*carrier-to-interference ratio*, C/I), pero también introduce interferencia entre haces adyacentes debido a los lóbulos secundarios del patrón de radiación. Hasta la actualidad, la solución más común a esto ha sido partir el

ancho de banda disponible (a menudo también las polarizaciones disponibles) en distintos trozos (normalmente tres o cuatro) y asignar haces adyacentes a trozos distintos. Esta estrategia, denominada reutilización parcial de frecuencias, reduce la interferencia de forma abrupta, pero al precio de reducir también el ancho de banda disponible por haz.

Una forma posible de soslayar esta limitación es la de usar todo el ancho de banda disponible en cada haz (reutilización total de frecuencia), e intentar mitigar después la interferencia en el lado de la estación base (*gateway*). La esperanza tras esta solución es la de que, aún cuando no se pueda cancelar toda la interferencia, sí se pueda cancelar la suficiente para que el ancho de banda disponible ahora lo compense. Fijándose en el canal de retorno, emplear reutilización total y procesar conjuntamente todas las señales en el lado receptor traduce nuestro sistema a un canal MIMO multiusuario clásico. Esto permitiría aplicar la teoría de comunicaciones MIMO ya disponible en la búsqueda de una mejor explotación del canal. Por desgracia, la naturaleza del canal satélite deja pocos resultados aplicables.

En esta tesis se estudian las prestaciones del canal de retorno, tanto desde una perspectiva general como suponiendo un determinado receptor. En este último caso, las prestaciones alcanzables se obtienen de forma teórica en función de parámetros del sistema, como puedan ser el error de estimación del canal.

Trabajos previos

Se han llevado a cabo varios estudios para analizar las prestaciones de los satélites multihaz. En [5] se muestran resultados de simulaciones exhaustivas para evaluar cómo de factible es la detección multiusuario en sistemas satélite multihaz. Un análisis más teórico se presenta en [6], donde la capacidad de estos sistemas se aproxima, gracias a resultados existentes sobre matrices aleatorias, para un modelo de canal simplificado como Rice; se demuestra que la aproximación es precisa, pero está limitada a un modelo de canal muy específico y proporciona poca intuición.

En cuanto a la estimación imperfecta del canal, también [5] contiene resultados de simulación al respecto. En [7] se presentan resultados analíticos, de nuevo basados en teoría de matrices aleatorias, en forma de aproximaciones para distintas métricas de prestaciones. Este estudio usa un modelo de canal simplificado inspirado por las comunicaciones terrestres.

Contribuciones

- **C1.1 Comparación entre las prestaciones de la detección multiusuario terrestre e híbrida satélite/terrestre.** Para llevar a cabo la detección multiusuario del lado del receptor se necesita enviar todas las señales recibidas a través del correspondiente enlace. El ancho de banda necesario para esta operación puede constituir un problema, puesto que el número de elementos radiantes en la antena puede ser bastante elevado (y superior al número de haces que se forman). Para superar este problema, una posibilidad consiste en aplicar un procesado lineal y fijo en el satélite y reenviar solamente un conjunto reducido de

señales. En esta tesis, se comparan las prestaciones obtenidas mediante ambas alternativas, y se deducen de forma analítica condiciones que las hacen equivalentes.

- **C1.2 Análisis de la capacidad de outage de los sistemas satélite multihaz afectados por por atenuación por lluvia correlacionada.** El modelo de canal de un sistema satélite multihaz presenta muchas dificultades para el análisis de prestaciones. Esto se debe a que cada señal originada en la Tierra experimenta la misma atenuación en su camino hacia cada elemento radiante del satélite. Además, las particularidades del entorno de propagación obligan a usar complicadas distribuciones de probabilidad para esta atenuación. En esta tesis, se obtienen expresiones para la capacidad de *outage*, para SNR baja y SNR alta, de un sistema de comunicaciones satélite afectado por atenuación por lluvia. Se permite que dicha atenuación se encuentre espacialmente correlacionada, y se investiga la influencia de la distancia entre usuarios para una determinada topología.

- **C1.3. Análisis de prestaciones de la detección multiusuario con información imperfecta sobre el canal.** Un problema crucial cuando se aplica detección multiusuario es el de obtener información de canal precisa. Para un modelo en el que la estimación del canal consisten en la matriz de canal real, supuesta determinista, más una matriz de error aleatoria, en esta tesis se obtienen expresiones para el error cuadrático medio residual tras la detección en función de la longitud de la secuencia de entrenamiento empleada. Las suposiciones tomadas sobre la matriz de error permiten analizar distintas técnicas de estimación.

Adaptación de enlace en comunicaciones móviles por satélite

Introducción

La adaptación del enlace (*link adaptation*) es el proceso de cambiar dinámicamente los parámetros de la transmisión de acuerdo con el estado del canal. Distintos tipos de adaptación pueden ser de interés dependiendo del escenario en particular; considérese, por ejemplo, el ajuste de la potencia en canales con desvanecimientos *fading*, o la planificación del acceso de los usuarios en función de la calidad del canal.

Un problema particularmente interesante, y habitual en los modernos sistemas de comunicaciones inalámbricas, es el ajuste de la tasa de transmisión (*rate adaptation*). Los canales inalámbricos sufren desvanecimientos por naturaleza y, por ello, la tasa máxima que pueden soportar cambia con el tiempo. Un sistema de comunicaciones que pretenda maximizar su eficiencia espectral debe, de alguna manera, monitorizar los cambios del canal y ajustar su tasa de acuerdo a ellos. Este ajuste suele llevarse a cabo mediante la definición de un conjunto discreto de tasas de codificación y modulaciones (MCS), y suele llamársele *codificación y modulación adaptativas* (*adaptive coding and modulation*, ACM). La mayoría de los sitemas de comunicaciones modernos, incluyendo tecnologías celulares, redes inalámbricas de área local, y estándares

de comunicación por satélite (DVB-S2 [8], DVB-RCS [9]), permiten en uso de distintos MCS y, en consecuencia, el uso de ACM.

Para llevar a cabo el ACM se necesita conocer algún tipo de información del estado del canal (*channel state information*, CSI) en el transmisor (CSIT). El transmisor puede acceder a esta información a través de su reenvío por parte del receptor (*feedback*), en lo que se conoce como bucle cerrado *closed loop CSI*, o a través de las señales piloto presentes en la señal proveniente del otro sentido de la comunicación, siempre que haya reciprocidad en el canal, estimando el canal en bucle abierto (*open loop*). La primera forma de obtener CSI es la que más se usa en los estándares actuales, mientras que la segunda sólo se considera utilizable cuando la transmisión dúplex se lleva a cabo en el dominio del tiempo.

En los sistemas de comunicación móvil por satélite, donde el retardo de propagación extremo a extremo aumenta el retardo de la CSI y disminuye el número máximo de retransmisiones permitidas, la selección de MCS es especialmente difícil. E, incluso cuando la CSI es actual, pueden aparecer otras dificultades. Por ejemplo cuando se usa MIMO-OFDM, o se usan palabras código largas que experimentan distintos valores del canal, se sufre de la llamada gran dimensionalidad de la CSI; en estos casos se necesitaría un valor de CSI para cada portadora y flujo espacial, o para cada símbolo en tiempo. Reducir la dimensionalidad promediando los valores de SNR obtenidos no suele ser una buena solución, puesto que la SNR media es una mala medida de prestaciones en estos canales. Recuérdese que, además, la selección de MCS se complica aún más cuando la CSI es también imperfecta debido a los retardos en su transmisión.

Trabajos previos

Los trabajos previos sobre adaptación de enlace en comunicaciones por satélite se centran en seleccionar el MCS en base a umbrales aplicados al valor de SNR estimado. En [10–12], se diseñan dichos umbrales teniendo en cuenta el error sobre la CSI que causa la estimación imperfecta de SNR, pero no se tiene en cuenta el efecto del retardo en entornos móviles. La estrategia de obtenida se basa en la adaptación con histéresis, es decir, el MCS que se usa en la transmisión futura se escoge en función del MCS actual y del valor de SNR estimado. En [13], se propone el uso de un margen de retroceso para tener en cuenta la imprecisión de la estimación de SNR, incluyendo tanto el retardo como el error en la estimación. Este margen puede obtenerse durante el funcionamiento del sistema, a partir del intercambio de ACK/NAK, o puede obtenerse mediante simulaciones previas. En [14] se propone usar solamente una fracción de los MCS disponible, con el objetivo de que la adaptación sea así más robusta.

En esta tesis se siguen enfoques distintos para la adaptación en los canales de ida y de retorno. En el canal de ida, se explota información estadística sobre él para maximizar la velocidad de transmisión respetando una restricción sobre la tasa de paquetes perdidos; esto se consigue empelando distintos MCS en cada sucesiva retransmisión de un paquete. En cuanto al canal de retorno, se propone el uso de CSI de bucle abierto para obtener información actualizada sobre el estado del canal.

Contribuciones

- **C2.1 Adaptación de enlace para el canal de retorno de sistemas satélite móviles mediante el uso de CSI en bucle abierto y cerrado.** El canal de retorno en comunicaciones móviles por satélite sufre retardos de propagación de la CSI muy dilatados cuando opera en bucle cerrado. Si se opera en bucle abierto, las variaciones gruesas del canal se pueden detectar al observar el canal de descenso, puesto que los objetos que provocan este tipo de fenómeno se encuentran siempre cerca del terminal. Sin embargo, los devanecimientos por experimentar la onda distintos caminos (*multipath*) no se pueden detectar así cuando ambos canales se encuentran ubicados en distintas frecuencias. En conclusión, un tipo de adaptación superará al otro dependiendo de las circunstancias particulares del entorno.

  En esta tesis se propone un método para combinar ambos tipos de CSI de forma automática, partiendo solamente del intercambio de ACK/NAK. El algoritmo de adaptación resultante se obtiene como la solución por gradiente estocástico a un problema de optimización sin restricciones que ajusta el valor de FER observado a un valor objetivo dado.

- **C2.2 Adaptación del enlace para el canal de ida en canales móviles satélites con ARQ.** La CSI adquirida en el canal de ida en comunicaciones móviles por satélites es imprecisa a causa del retardo de propagación. Para un satélite geoestacionario, por ejemplo, este retardo es de aproximadamente medio segundo. En muchos escenarios, e incluso a velocidades bajas, el tiempo de coherencia del canal es mucho menor que este valor. Además de este problema, la variación del canal es mayor que en los sistemas inalámbricos terrestres debido a los bloqueos en la línea de visión directa del enlace. En consecuencia, se pueden sufrir cambios profundos de SNR en períodos cortos de tiempo.

  En esta tesis se propone llevar a cabo la adaptación del enlace de ida basándose en información estadística sobre el canal, explotando la diversidad creada por los cambios en la SNR. Se consideran sistemas que permiten el uso de retransmisiones para incorporar diversidad temporal. Se presenta un enfoque que permite explotar los buenos momentos del canal aún con restricciones sobre la probabilidad de outage del sistema. El método se basa en el uso de distintos MCS en cada una de las sucesivas retransmisiones de un paquete. A pesar de que se basa en un conocimiento estadístico del canal, también se presenta un método para obtener este conocimiento a través del intercambio de ACK/NAK.

Notación

Salvo que se indique explícitamente lo contrario, se utiliza negrita mayúscula para denotar matrices, negrita minúscula para vectores (columna), y $\mathbf{A}^H$ para la matriz hermítica de $\mathbf{A}$; $\mathrm{diag}(\mathbf{A})$ es un vector columna formado por los elementos de la diagonal de la matriz $\mathbf{A}$, mientras que $\mathrm{diag}(\mathbf{a})$ es una matriz diagonal cuyos elementos están dados por $\mathbf{a}$. $\log(x)$ denota el logaritmo

neperiano de x; el logaritmo en base b se denota mediante $\log_b(x)$. $\mathbb{E}[X]$ es el operador experanza sobre la variable aleatoria X.

Estructura

El contenido principal de esta tesis se divide en 7 capítulos, agrupados en dos partes. La primera parte consta de los capítulos 2, 3 y 4, y la segunda de los capítulos 5, 6, 7 y 8. El capítulo 2 presenta una introducción a los sistemas multihaz y al procesado multiusuario, ensaya las dificultades asociadas al estudio de sus prestaciones, y compara entre una arquitectura con procesado solamente en tierra y otra con procesado híbrido (contribución 1.1). El capítulo 3 analiza de forma teórica las prestaciones de los sitemas multihaz estudiando su capacidad y su capacidad de outage (contribución 1.2). El capítulo 4 se centra en el uso de información imperfecta sobre el canal y obtiene aproximaciones a las prestaciones finales a través de teoría de matrices aleatorias (contribución 1.3). El capítulo 5 introduce el escenario para las comunicaciones móviles por satélite, y hace especial hincapié en el problema de monitorizar las prestaciones en canales fuertemente variantes en el tiempo. Los capítulos 6 y 7 estudian la viabilidad de la CSI en bucle abierto y proponen un algoritmo adaptativo para usarla junto con la de bucle cerrado (contribución 2.1). El capítulo 8 hace hincapié en el canal de ida, y presenta un método para obtener los parámetros de cada retransmisión con el fin de maximizar la eficiencia.

Abstract

The eternal demand of improved spectrum exploitation for next generation satellite communications requires the study of techniques that are potentially able of a breakthrough. In this thesis we focus on two of the main issues hindering the performance of satellite communications.

In the first part of the thesis, we study the use of multi-user detection techniques in the return link of multibeam satellite systems operated with an aggressive frequency reuse strategy. We start by comparing their performance with respect to a traditional scenario, and study the possibility of splitting the processing between the gateway and the satellite to cope with practical constraints. We then study the fundamental performance limits of such a system when spatially correlated rain attenuation affects the users; analytical expressions are obtained for the high and low SNR regimes. Finally, we consider non perfect channel state information and obtain, through random matrix theory, an approximation of the residual error as a function of the system's parameters.

In the second part, we focus on link adaptation for mobile satellite communications. For the return link, we investigate the potential of open-loop channel state information, which is timely but scarcely accurate; the finding that this information is better than the traditional closed-loop one depending on the channel conditions motivated the study of an adaptive algorithm to dynamically balance both. Finally, in the forward link we considered the case where retransmissions are available, and designed the optimum modulation and coding scheme sequence for the different retransmissions of a packet in terms of throughput.

Publications

The following is a list of journal and conference publications that have been produced as a result of the work on this thesis.

Journal publications

1. **Jesus Arnau**, Dimitrios Christopoulos, Symeon Chatzinotas, Carlos Mosquera, and Björn Ottersten, "Performance of the multibeam satellite return link with correlated rain attenuation", *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6286 - 6299, Nov. 2014.

2. Dimitrios Christopoulos, **Jesus Arnau**, Symeon Chatzinotas, Carlos Mosquera, and Björn Ottersten, "MMSE performance analysis of generalized multibeam satellite channels", *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1332-1335, Jul. 2013.

3. **Jesus Arnau**, Betrand Devillers, Carlos Mosquera, and Ana Pérez-Neira, "Performance study of multiuser interference mitigation schemes for hybrid broadband multibeam satellite architectures", *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, pp. 132, 2012.

Conference publications

1. Alberto Rico-Alvariño, **Jesus Arnau** and Carlos Mosquera, "Balancing closed and open loop CSI in mobile satellite link adaptation", in *Proc. ASMS & SPSC*, Livorno, Italy, Sep. 2014, **best paper award**.

2. **Jesus Arnau** and Carlos Mosquera, "Dissection of multibeam satellite communications with a large-scale antenna system toolbox," in *Proc. EW2014*, Barcelona, Spain, May 2014.

3. Alberto Rico-Alvariño, **Jesus Arnau** and Carlos Mosquera, "Statistical cross layer adaptation in fast fading mobile satellite channels", in *Proc. IEEE GLOBECOM*, Atlanta, GA, Dec. 2013.

4. **Jesus Arnau** and Carlos Mosquera, "Open loop adaptive coding and modulation for mobile satellite return links", in *Proc. AIAA ICSSC*, Florence, Italy, Oct. 2013.

5. **Jesus Arnau** and Carlos Mosquera, "Multiuser detection performance in multibeam satellite links under imperfect CSI", in *Proc. ASILOMAR*, Pacific Grove, California, Nov. 2012.

6. **Jesus Arnau**, Alberto Rico-Alvariño, and Carlos Mosquera, "Adaptive transmission techniques for mobile satellite links", in *Proc. AIAA ICSSC*, Ottawa, Canada, Sep. 2012. **Best Professional Paper Award**

7. **Jesus Arnau** and Carlos Mosquera, "Performance analysis of multiuser detection for multibeam satellites under rain fading", in *Proc. ASMS & SPSC*, Baiona, Spain, Sep. 2012.

8. **Jesus Arnau** *et. al.*, "Hybrid space-ground processing for high-capacity multi-beam satellite systems", in *Proc. IEEE GLOBECOM*, Houston, Texas, Oct. 2011.

Book chapters

1. **Jesus Arnau**, Carlos Mosquera, and Rodrigo de Lamare, "Multibeam joint detection", in S. Chatzinotas, R. de Gaudenzi and B. Ottersten, editors, *Co-operative and cognitive satellite systems*, Elsevier, 2014. To appear.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Contents

1.1 Motivation

In the last years there has been an increasing demand of wireless data, and it is forecast to keep growing in the next years; according to [1], there has been an increase of around 81 % in traffic during 2013, and a ten fold increase is expected by 2018. To accommodate this growth, modern wireless standards have been deployed to increase the spectral efficiency of wireless networks.

The world of satellite communication systems has been no exception, despite their purported inferiority in capabilities when compared with terrestrial technologies. One of the reasons for this is that terrestrial communications are costly to deploy in numerous isolated areas, and as a consequence future broadband system are envisaged to integrate satellite and terrestrial infrastructures [2]. Another reason is that satellite operators seek to maintain the competitiveness of their services even with current terrestrial technologies, which definitely requires offering improved services at the same (or lower) cost. As some professionals once put it:

> "Users will require enhanced interactive services for the same price [...]. Can satellites
> cope with the evolving user demand?" [3].

The described situation influenced the plans of many involved institutions, as for example the Horizion 2020 goals of the European Union, and fostered the funding of international research projects and the proposal of new standards; as an example, during the last four years, we have seen, among others, the transition form DVB-S2 to DVB-S2x/DVB-RCS2 in satellite broadcasting [4], and the efforts for standardizing S-UMTS. But, even though many improvements have been made –through, for example, improved modulations and channel codes, and more efficient transmit filters– there is still the need to better exploit the spectrum and maintain the competitiveness of satellite communications.

In this thesis, we study two solutions aimed at two different satellite communication scenarios, with a special emphasis on the return link (RL), which is the one originating at the user side and ending at the gateway Earth station (hub) through the satellite. The first solution applies to fixed satellite services (FSS) where a single satellite serves a large population of users through the use of multiple radiation beams; for this case, we study the possibility of mitigating the existing interference at the receiver side, and provide tools for assessing the achievable performance and dimensioning relevant system parameters. The second solution applies to mobile satellite services (MSS) and aims at dynamically adapting the transmission parameters to the changing conditions of the communications channel; for this case, we propose novel adaptation strategies for both forward and return links, and analyze their performance.

1.2 Interference mitigation in multibeam satellite systems

1.2.1 Introduction

Today's satellite coverages for interactive services mostly consist of multiple small antenna beams, rather than on a single, large radiation area. This allows the provision of services with a higher quality in terms of carrier-to-noise ratio (C/N), but also introduces interference among adjacent beams due to the side lobes of the antenna pattern. So far, the most common solution consists in splitting the available bandwidth (and probably the two orthogonal polarizations available) into different pieces (usually three or four) and allocating adjacent beams to different portions. This strategy, called *partial frequency reuse*, sharply decreases the overall interference, but at the price of dramatically cutting down the available bandwidth per beam with respect to an ideal full-reuse system.

A possible way of overcoming this limitation is by using all the available bandwidth in each beam (*full frequency reuse*), and then trying to mitigate the existing interference at the gateway end. The idea behind this solution is that, even if not all the interference can be effectively mitigated, the large bandwidth increase could compensate for it. Focusing on the return link, applying full frequency reuse and jointly processing the received signals at the gateway side

translates our system into a classic multiple-input multiple output (MIMO), multiuser channel. This allows applying the existing theory of MIMO communications in the pursue of better channel exploitation. Unfortunately, the nature of the satellite channel leaves us with few applicable theoretical results.

Since optimal strategies for this channel are prohibitively complex, linear and non-linear sub-optimal processors that mitigate interference will be considered. These techniques, that we will refer to as *multiuser detectors*, are based on well-known statistical tools, and had been used and analyzed before in the context of code division multiple-access (CDMA) communications [16]. To successfully apply them, all the signals have to be relayed to the same gateway, which must have an accurate estimate of the channel and of the noise statistics. The first condition may be difficult to meet, as it could require a very high feeder link bandwidth (specially in the case of broadband FSS). The second condition also introduces problems, since a poor estimation might render the detector inapplicable.

In this thesis we study the performance of the multiuser return link, both from a capacity perspective and by assuming a specific detector. In the latter case, the achievable performance will be obtained theoretically as a function of system elements like the estimation error.

1.2.2 Prior work

Many efforts have been put in analyzing the performance of multibeam satellite systems. In [5], extensive simulations were performed to assess the feasibility of multiuser detection in multibeam satellite communications. Theoretical analysis were carried out in [6], where the capacity of such systems was approximated using random matrix results for a simplified channel model with Rician fading; the approximation proved to be tight, but was limited to a specific channel model and provided little insight.

Regarding imperfect channel state information, in [5] simulation results are shown illustrating its influence. For analytical results, [7] obtained random-matrix based approximations of different performance metrics, but for a simple channel model inspired by wireless terrestrial communications.

1.2.3 Contributions from part I

- **C1.1) Performance comparison between on-ground and hybrid on-board on-ground multiuser detection in multibeam satellite systems.** Performing multiuser detection at the gateway requires relaying all the signals received at the satellite through the feeder link. Since the number of antenna elements at the satellite can be larger than the number of users on Earth, the amount of bandwidth required in the feeder link may become an issue. A possibility to overcome this problem is to perform a fixed linear processing on-board and then relay only a reduced set of signals. In this thesis we compare the performance achieved by both alternatives, and analytically obtain conditions under

which they are equivalent.

- **C1.2) Analysis of the outage capacity of a multibeam satellite system affected by correlated rain attenuation.** The channel model of a multibeam satellite system poses many difficulties in the analysis of its performance. This is because each signal from Earth undergoes the same channel coefficient towards all the antenna elements at the satellite. Additionally, the particularities of the propagation environment require the use of complicated statistical distributions. In this thesis, we obtain high and low SNR expressions of the outage capacity of a multibeam satellite system impaired by rain attenuation. We allow the different attenuation coefficients to be spatially correlated, and investigate the influence of the inter-user distance for a given topology.

- **C1.3) Performance analysis of multiuser detection with imperfect CSI.** A crucial issue when applying multiuser detection techniques is obtaining accurate channel state information. For a model where the estimate consists of the actual channel matrix, that is assumed deterministic, plus a random error matrix, in this thesis we obtain simple expressions of the residual mean-squared error after multiuser detection as a function of the training sequence length employed. The assumptions made on the error matrix allow to accommodate different estimation techniques.

1.3 Link adaptation in mobile satellite communications

1.3.1 Introduction

Link adaptation is the process of dynamically changing the transmission parameters according to the channel state. Depending on the particular scenario, different adaptations might result of interest; consider, for example, changing the power in fading channels to keep a constant bit rate, or scheduling the users depending on the channel quality.

One task that is particularly interesting, and that is usually performed in modern wireless communication systems, is rate adaptation. Wireless channels are of fading nature and, therefore, the rate they can support changes over time. A communication system that wants to maximize its spectral efficiency should track the channel changes and adapt its transmission rate accordingly. This rate selection is usually performed by the definition of a discrete set of available coding rates and modulations, or modulation and coding schemes (MCS). Rate adaptation is usually referred to as adaptive coding and modulation (ACM). Most modern communication systems, including cellular technologies, wireless local area networks, and satellite communication standards (DVB-S2 [8], DVB-RCS2 [9]) support different MCS and, therefore, enable the use of ACM.

Performing ACM requires the knowledge of some sort of channel state information (CSI) at the transmitter (CSIT). The transmitter can gain access to this information by means of feedback from the receiver, in what is known as *closed loop* CSI, or by exploiting the pilots present in the

incoming signal if there is channel reciprocity, and estimating the channel in an *open loop* way. The former CSI acquisition technique is widely supported by modern communication standards, while the latter is regarded as usable only when the duplexing is performed in the time domain.

MCS selection is particularly difficult in mobile satellite communications, where the large round trip time increases the CSI delay and decreases the maximum allowable number of retransmissions. And, even when CSI is timely, there can be other difficulties. For example, systems using MIMO-OFDM, or large codewords that experience different channel values, suffer from the large CSI dimensionality: one CSI value per carrier and spatial stream, or per symbol in time, would be required. Reducing the dimensionality by averaging the SNR values is usually not a good approach, since average SNR is not a good quality metric in these systems. The difficulty of performing MCS selection is increased when CSI is imperfect due to limited feedback.

1.3.2 Prior work

Prior work on satellite communications focused on selecting the MCS based on thresholds on the estimated SNR value. In [10–12] thresholds for MCS selection were designed taking into account the CSI error caused by imperfect SNR estimation, but the effect of delay in a mobile environment was not analyzed. The obtained strategy is based on adaptation with hysteresis, i.e., the MCS to be used in the next transmission is selected depending on the estimated SNR value and the current MCS. In [13] it is proposed to use a backoff margin to account for the inaccuracy of the SNR estimate, including both delay and estimation error. This backoff margin can be obtained online, from the ACK/NAK interchange, or can be obtained by simulations. In [14] it is proposed to use only a fraction of the available MCS to perform link adaptation with the objective of increasing the robustness of the adaptation.

In this thesis we follow different approaches for adaptation in forward and return links. In the forward link we exploit statistical information about the channel to maximize the throughput subject to a packet error rate constraint through the use of different MCS in each retransmission. In the return link we propose the use of open loop CSI to obtain timely information about the channel.

1.3.3 Contributions from part II

- **C2.1) Link adaptation for the return link of mobile satellite channels exploiting open loop and closed loop CSI.** The return link of mobile satellite channels suffers from a very long CSI propagation delay if operating in closed loop mode. If adaptation is performed in open loop, large channel variations can be detected by observing the forward channel, as the objects that cause blockage are always placed next to the terminal; fading due to multipath, however, is not detectable by observing the incoming singal if duplexing is performed on the frequency domain. Thus, one alternative will outperform the other depending on the particular case environment.

In this thesis we propose a method to automatically combine both CSI values depending on the ACK/NAK observed. The link adaptation algorithm is derived as the stochastic gradient descent solution of an unconstrained optimization problem which sets the observed FER to a desired target value.

- **C2.2) Link adaptation techniques for the forward link of mobile satellite channels using ARQ.** CSI acquisition in the forward link of mobile satellite channels is rather inaccurate due to the large round trip delay. For a geostationary satellite, for example, this delay is around half a second. In many scenarios, and even for low speeds, the channel coherence time is lower than this value. On top of this problem, the channel variation is larger than in terrestrial systems due to the sensitivity to blockages of the line of sight propagation path. Thus, it is possible to suffer very large SNR variations in a small period of time.

  In this thesis we propose to perform link adaptation based on statistical channel information, and try to exploit the diversity created by the large SNR variations. We consider systems that allow the use of retransmissions to incorporate time diversity. We present an approach that allows exploiting the good channel instants while incorporating an outage constraint on the system.The method is based on the use of different MCS in different retransmissions. This technique is based on statistical information on the channel, but we present a method to estimate the necessary parameters from the exchange of ACK/NAK.

1.4   Notation

Bold uppercase face $\mathbf{A}$ denotes matrices, bold lowercase $\mathbf{a}$ denotes (column) vectors, and $\mathbf{A}^H$ is the hermitean of $\mathbf{A}$; $\mathrm{diag}(\mathbf{A})$ is a column vector formed by the elements in the diagonal of matrix $\mathbf{A}$, while $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with elements given by $\mathbf{a}$. $\log(x)$ stands for the neperian logarithm of x; base-b logarithms are denoted as $\log_b(x)$. $\mathbb{E}\left[X\right]$ is the expectation operator over a random variable X.

1.5   Structure

The main content in this thesis is structured in seven chapters, grouped in two parts. The first part consists of chapters 2, 3 and 4, and the second of chapters 5, 6, 7 and 8.

Chapter 2 introduces multibeam systems and multiuser processing, sketches the difficulties of their performance analysis, and compares between on-ground beamforming and hybrid processing (C1.1). Chapter 3 theoretically analyzes the performance of a multibeam satellite system through the outage capacity (C1.2). Chapter 4 focuses on the use of imperfect channel state information and obtains approximations to the performance in this case (C1.3).

Chapter 5 introduces the mobile satellite communications scenario, with a focus on the

problem of performance monitoring in time varying channels. Chapters 6 and 7 assess the feasibility of exploiting open-loop CSI, and propose an adaptive algorithm to use it along the closed-loop CSI (C2.1). Chapter 8 focuses on the forward link and shows a method to obtain the transmission parameters for each retransmission of a packet.

Part I

Interference mitigation in multibeam satellite systems

Chapter 2

Introduction to multibeam satellite systems

Contents

2.1   Introduction

As sketched in Chapter 1, the use of multiple spot beams in modern communication satellites is a consolidated technology which leverages the spatial reuse of spectral resources to increase the throughput and, consequently, reduce the cost per bit. The same frequency band is shared by different beams to provide an overall higher throughput as long as the intrinsic multiuser interference can be kept under control. This interference, related to the degree of reuse of the spectrum, is caused by the non-null side lobes of the beams radiation patterns, and it is conventionally minimized by excluding adjacent beams from using the same portion of spectrum (or *color*), as illustrated in Figure 2.1. However, more aggressive frequency reuse alternatives have been proposed [17–19] in an effort to push forward the overall spectral efficiency. The adopted color scheme, that is, the number of frequency bands which are assigned to the different beams, determines the spectral efficiency cap, and requires the use of joint processing such as multiuser detection (MUD) for those more aggressive frequency reuse factors. The current frequency reuse standard[1] in today's broadband multi-beam satellites is four, although MSS satellites with on-board digital transparent processor can implement more flexible and non uniform color schemes. The interference introduced by the higher frequency reuse can be mitigated by some advanced processing at the transmit or receive side for the forward and return link, respectively [20], [21].

In this chapter we will present the main considerations on the use of multiuser interference mitigation schemes for the return link of multibeam communication satellites. We will show an example of the current baseline scenario and illustrate its performance. Next we will review some of the most common multiuser detection techniques, and sketch the fundamental performance limits of this multiuser channel taking into account the specifics of the multibeam satellite setting. Before finishing the channel, we will theoretically analyze the implications of having a fixed beamforming network on-board.

2.2   System model

This first part of the thesis focuses on a multi-user uplink, where multiple single-antenna terminals communicate towards a satellite equipped with a multi-feed reflector. We propose a beam pattern model and include the possibility of correlation among the rain attenuations experienced by different users –both dependent on the system geometry and inter-user distance.

Let us consider a multi-user channel with K single-antenna terminals transmitting towards a single satellite equipped with the same number of antennas (Figure 2.2), so that the signal model reads as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{s} \in \mathcal{C}^{K \times 1}$ is the transmitted signal vector, such that $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right] = \mathbf{I}$, $\mathbf{y} \in \mathcal{C}^{K \times 1}$ is the received

---

[1]Polarization poses an additional degree of freedom to reuse frequency bands while avoiding the associated interbeam interference.
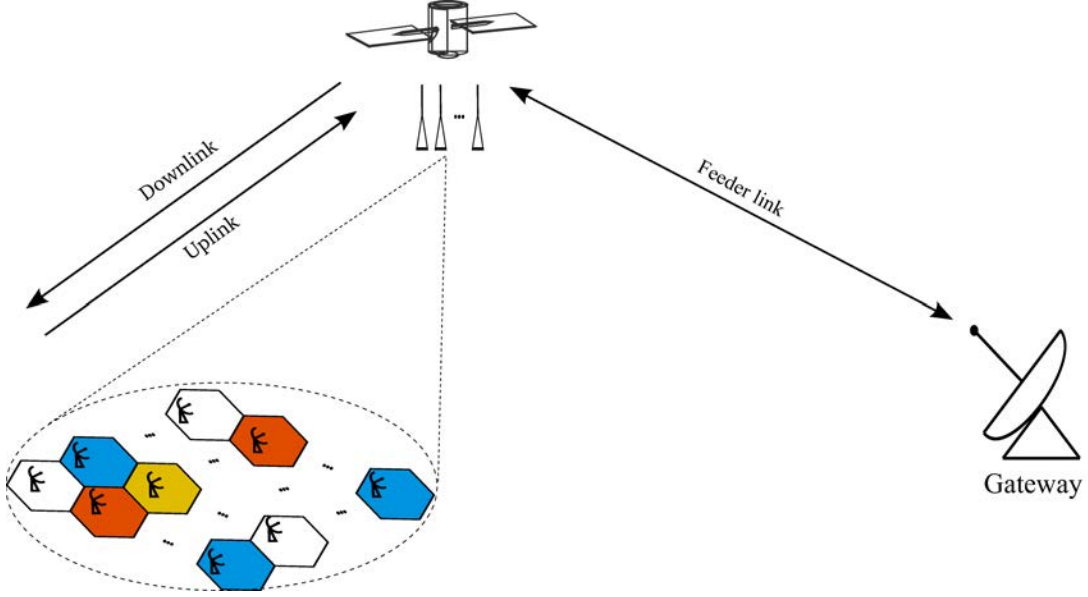
Figure 2.1: System model under consideration, depicting a partial frequency reuse plan.

signal vector, $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, N_0\mathbf{I}\right)$ is the complex noise vector, and N_0 is the noise power. Matrix $\mathbf{H} \in \mathcal{C}^{K \times K}$ represents the complex-valued channel and can be expressed as [6, 22]

$$\mathbf{H} = \mathbf{GD} \tag{2.2}$$

where $\mathbf{G} \in \mathcal{C}^{K \times K}$ is a full column-rank matrix containing the antenna radiation pattern, and $\mathbf{D} = \mathrm{diag}(\boldsymbol{\delta})$, $\mathbf{D} \in \mathcal{C}^{K \times K}$, is a diagonal matrix of random entries modeling the rain attenuation coefficients.

2.2.1   Beam radiation pattern

Matrix $\mathbf{G} \in \mathcal{C}^{K \times K}$ accounts for the values of the beam radiation pattern for the positions of the users within the beam spots. In consequence, $\mathbf{G}$ will change as a function of the location of the active users at a given time. In most practical studies, $\mathbf{G}$ is obtained numerically from specialized antenna design software. However, a number of references have also opted for modeling a tapered-aperture antenna by a combination of Bessel functions[2].

Numerical model

A customary way of obtaining the antenna pattern is to use some specific software and obtain a numerical array of values. A pattern provided by ESA in the context of the SatNEx III project [15] will be used in repeated occasions. This pattern, designed to serve 100 beams with an antenna with 155 feeds, can be factored as $\mathbf{G} = \mathbf{BG}_2$, where $\mathbf{B} \in \mathcal{C}^{100 \times 155}$ represents an

---

[2]Even simpler models are possible as shown in [6], where a Gaussian characteristic is assumed for the main lobe of the antenna radiation pattern.

$$y = Hs + n$$
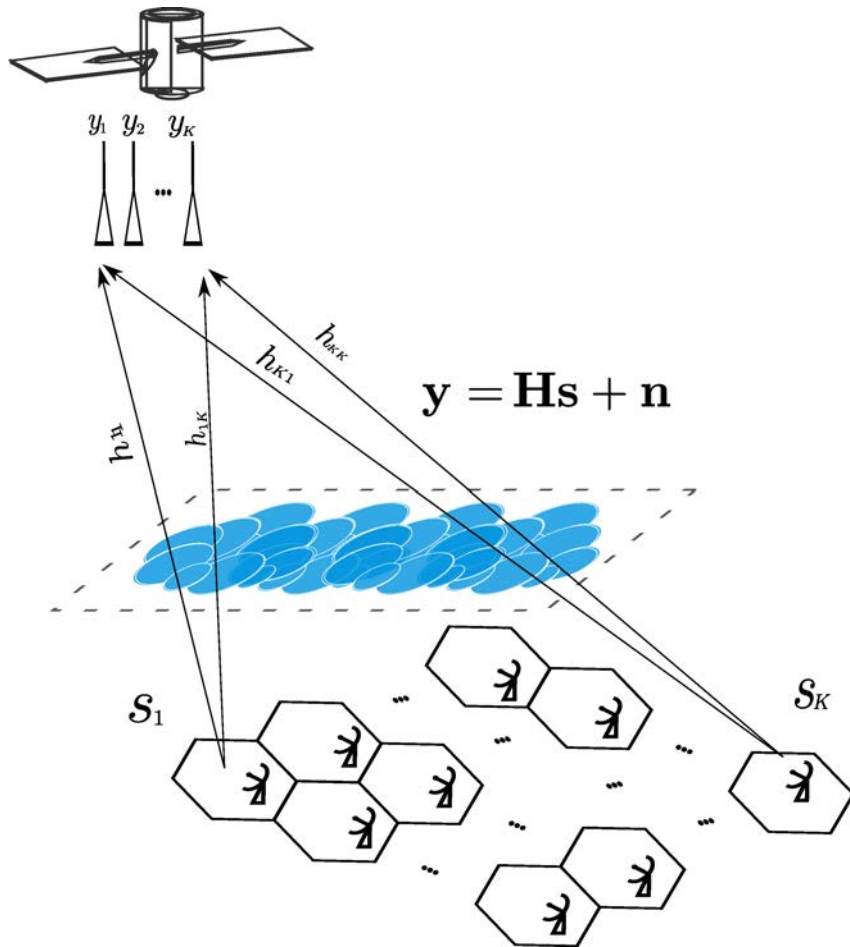
Figure 2.2: Diagram of the satellite return link under study.
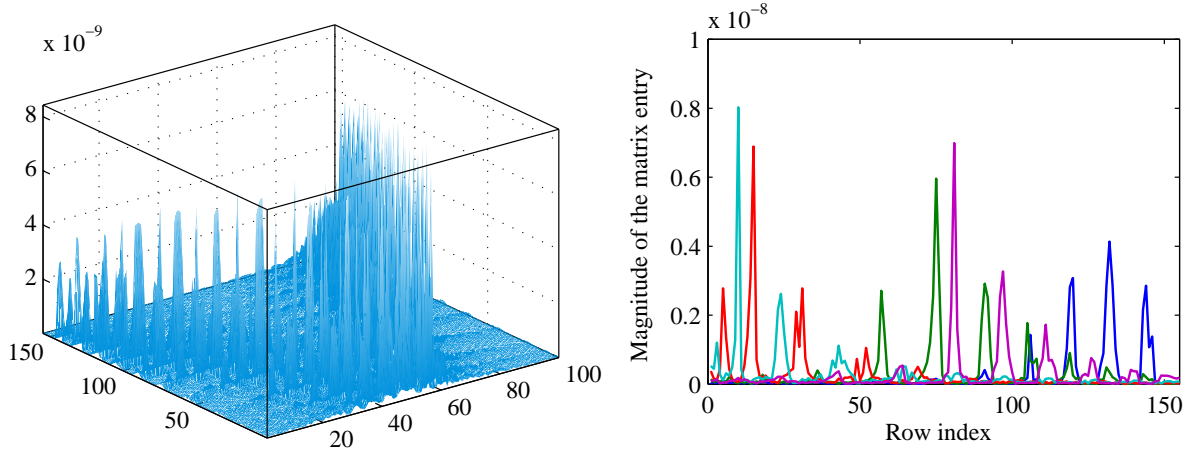
Figure 2.3: Antenna pattern provided in the context of [15], with free space path losses already included. Left represents the magnitude for every entry, right for only one column.
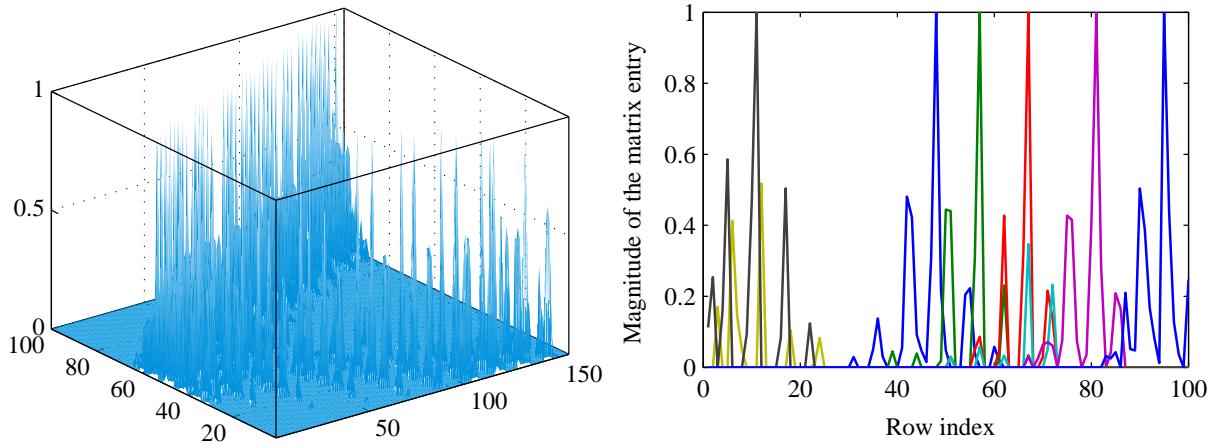
Figure 2.4: On-board beamforming provided in the context of [15]. Left represents the magnitude for every entry, right for only one column.

on-board beamforming matrix (Figure 2.4) and $\mathbf{G}_2 \in \mathcal{C}^{155 \times 100}$ is the actual beam radiation pattern (Figure 2.3). We will explain the implications of an on-board beamformer later on.

In these cases, and for simplicity, we will assume that all the deterministic losses per beam, including the path losses, are already included into this matrix.

Analytical model

As explained, it is also possible to model the antenna by a combination of Bessel functions. If d_a is the diameter of the aperture, T is the aperture edge taper (with $T = 0$ the antenna aperture field is uniform), and $p = 0, 1, 2, ...$ is related to the decay of the field, then [23, p. 184], [24]

$$g_{ij}(\theta_{ij}) = G_{\max} \left[\frac{(p+1)(1-T)}{(p+1)(1-T)+T} \left(2J_1(u_{ij})/u_{ij} + 2^{p+1}p! \frac{T}{1-T} \frac{J_{p+1}(u_{ij})}{u_{ij}^{p+1}} \right) \right]^2, \qquad (2.3)$$

with $J_p(u)$ the Bessel function of the first kind and order p, $G_{\max}$ is the maximum axis gain of the antenna, θ_{ij} is the off-axis angle with respect to the beam's boresight, and

$$u_{ij} = \frac{\pi d_{\mathrm{a}}}{\lambda} \sin \theta_{ij}. \tag{2.4}$$

The value at beam contour is usually between 3 and 4 dB.

2.2.2 Fading

In addition to the directivity of the satellite antennas, represented by $\mathbf{G}$, Equation (2.2) captures the random propagation effects in matrix $\mathbf{D}$. Note that this matrix is diagonal, as opposed to conventional MIMO models used in terrestrial systems, to denote that one user signal undergoes the same channel as seen from the different antennas [6, 22]. This is due to the relatively small separation of the antenna elements with respect to the satellite altitude, and constitutes a major difference with respect to terrestrial systems. Also note that, for a simplified analysis, any known deterministic attenuation that might be present, as for example the free-space path losses, will be often embedded into matrix $\mathbf{G}$.

In fixed satellite systems operating above 10 GHz, and especially in bands like Ka and above, atmospheric fading, and specifically rain attenuation, dominate the propagation budget [25]. There are many statistical models for the rain attenuation coefficients, which have been obtained by fitting experimental data with different distributions. In particular, ITU-R P. 1853 assumes a log-normal distribution for these coefficients when expressed in dB [26]; mathematically speaking, this implies that the attenuation coefficients δ_ℓ would follow

$$-20 \log_{10}(\delta_\ell) \sim \mathcal{LN}(\mu_\ell, \sigma_\ell) \tag{2.5}$$

where μ_ℓ and σ_ℓ are the log-normal location and scale parameters, respectively, expressed in dB. Time and spatial correlation of the rain attenuation can be also characterized given their impact on fade and multiuser mitigation techniques [27–29], as we will highlight later on. For illustration purposes, Figure 2.5 shows the empirical complementary cumulative distribution of the rain attenuation for the cities of Rome and Madrid; measurements were performed at 30 GHz and provided in [15].

2.2.3 Why diagonal fading

Diagonal fading, as understood through this work, appears when all the received signals undergo the same channel attenuation. As already explained, this comes from the fact that the satellite antennas are close to each other in comparison with the distance between the satellite and the Earth.
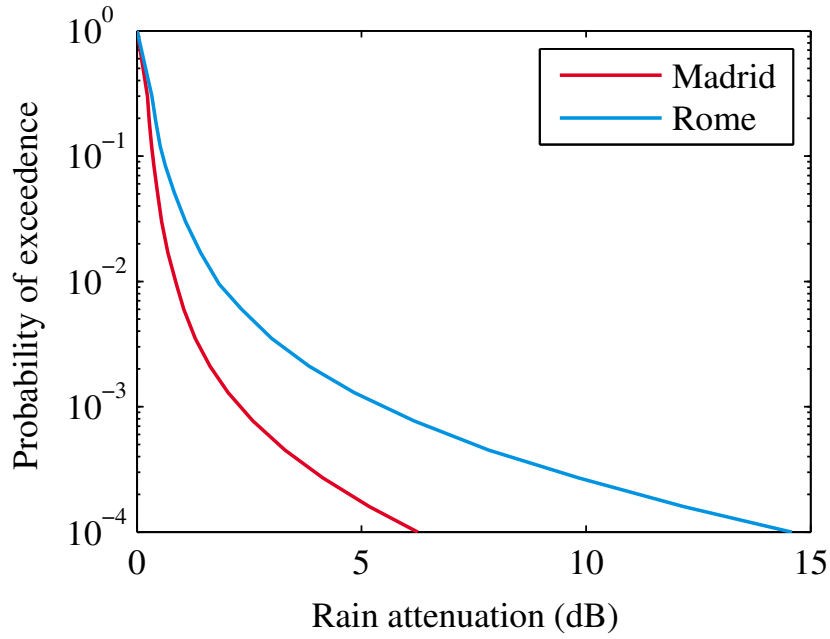
Figure 2.5: Complementary cumulative distribution functions of the rain attenuation for the cities of Madrid and Rome.

This model can be easily derived following the lines of [22]. Start by considering

$$\mathbf{H} = \mathbf{G} \odot \mathbf{F} \tag{2.6}$$

with $\mathbf{G}$ the antenna pattern matrix and $\odot$ the entry-wise product, such that each element of the antenna pattern would be scaled by the corresponding fading coefficient in matrix $\mathbf{F}$.

Assume now the common Kronecker correlation model, so that $\mathbf{F} = \mathbf{R}_{\mathrm{rx}}^{1/2} \tilde{\mathbf{F}} \mathbf{R}_{\mathrm{tx}}^{1/2}$. If there is total correlation at the receiving end, then $\mathbf{F} = \mathbf{1} \tilde{\mathbf{F}} \mathbf{R}_{\mathrm{tx}}^{1/2}$. The resulting rank-one matrix has constant columns. Since element-by-element multiplication by a matrix with constant columns is equivalent to scaling each column, we conclude that $\mathbf{H}$ can be expressed as $\mathbf{GD}$ with $\mathbf{D}$ a diagonal matrix.

2.3 Baseline scenario

Before explaining the multiuser detection techniques considered, we will describe the usual way in which multibeam satellite systems are operated nowadays, which is through a partial frequency reuse plan. We will describe an example of such a system in detail and illustrate its performance through simulations.
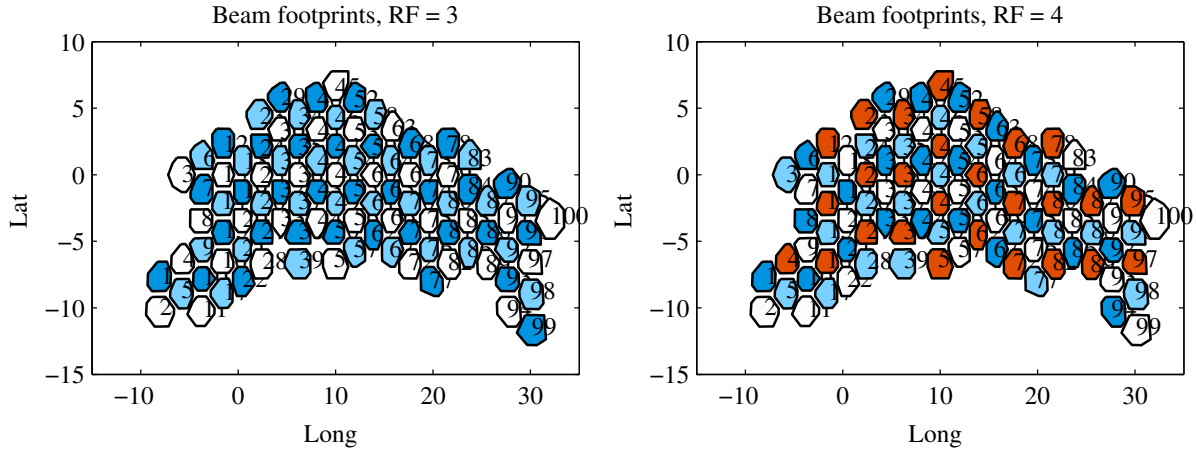
Figure 2.6: Coverage diagram with frequency reuse factors 3 (left) and 4 (right). Geographical coordinates are fictitious.

2.3.1   Partial frequency reuse

To reduce interference among adjacent beams, it is customary to split the available bandwidth in three or four bands, and allocate interfering beams to different bands; for the beam pattern provided by ESA in [15], Figure 2.6 depicts two examples of frequency allocation for these frequency reuse factors, where different colors represent different portions of the spectrum.

2.3.2   Scenario description

The user link operates at 30 GHz (Ka-band), and is based on the DVB-RCS2 standard [30]. The baudrate is 4 Msymb/s and the guardbands amount to 11% of the carrier bandwidth [31].

Table 2.1: System parameters

Atmospheric fading	City of Rome	Center frequency	30 GHz
Number of beams	100	System bandwidth	500 MHz
UTs location	Uniformly distributed	Carrier rate	4 MBaud
Rec. noise figure (sat.)	2.5 dB	Filter rolloff factor	0.25
Total rec. noise (sat.) T	517 K	Guard band percentage	11 %

The user link has a total available bandwidth of 500 MHz; color schemes with frequency reuse factor equal to 3 and 4 were studied, corresponding to 166 MHz and 125 MHz available bandwidth per beam, respectively. For illustrations purposes, simulation results cover a large range of transmit powers, although it is important to stress that the most extreme values do not correspond to practical cases.

Results have been averaged for a total of 1000 channel realizations, with the exception of those showing the average probability of non-availability, which required 10000 iterations to

Table 2.2: DVB-RCS2 MODCOD description.

ModCod	Spectral efficiency (bps/Hz)	Required E_s/N_0 (dB)
QPSK_13	0.53	−0.45
QPSK_12	0.8	1.80
QPSK_23	1.07	3.75
QPSK_34	1.2	4.85
QPSK_56	1.33	6.10
8PSK_23	1.6	7.60
8PSK_34	1.8	8.90
8PSK_56	2	10.30
16QAM_34	2.4	11.20
16QAM_56	2.67	12.20

yield a reasonable confidence interval. Apart from the fading, the randomness of the channel is due to the position of the users, which are assumed to be uniformly distributed within each spot. For each realization, the SINR for each user after interference mitigation is computed, and its throughput is then inferred according to the specifications of DVB-RCS2 (see Table 2.2).

2.3.3 Numerical examples

We will now show simulation results illustrating the performance obtained by these schemes.

We can see that, even though factor 4 shows slightly better SINR values (Figure 2.7), performance is very similar in terms of outage, as shown on Figure 2.8 on the right. However, the achievable throughput is greater with factor 3 because of the extra bandwidth available. For this reason, factor 3 will be considered as baseline in this part of the thesis.

2.4 Multiuser detection techniques

2.4.1 Overview

We consider again the signal model at the symbol level (2.1). The processing taking place at the gateway has to extract the symbols **s** from the received samples **y**; these symbols correspond to the users transmitting from the beams which are served by this gateway, which could coexist with other gateways. Each gateway is said to serve a group of beams known as cluster; for the time being, let us assume a single gateway architecture with users in all beams managed by the same gateway.
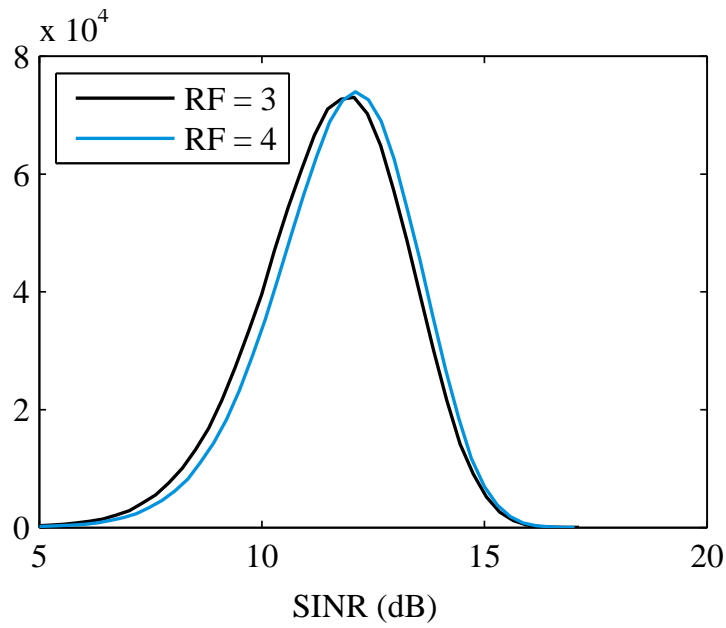
Figure 2.7: Histogram of the simulated SINR with EIRP = 40 dBW.
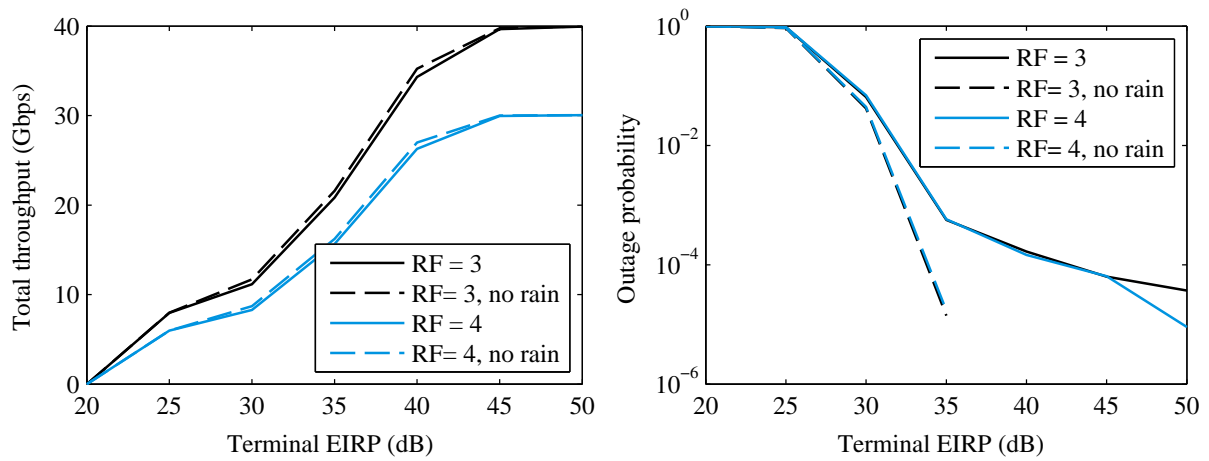
Figure 2.8: Throughput and outage as a function of EIRP for different values of RF.

The optimal maximum likelihood (ML) detector is described, if noise is Gaussian, by

$$\hat{\mathbf{s}}_{\mathrm{ML}} = \arg \min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 \tag{2.7}$$

where the $K \times 1$ data vector $\mathbf{s}$ contains the data symbols of the K users. The ML detector has a cost that is exponential in the number of users K and the modulation order, and which is too complex to be implemented in gateways associated to a large number of simultaneous users. Even though the ML solution can be alternatively computed using sphere decoder (SD) algorithms [32,33] that are very efficient for systems with a small number of antennas, the cost of SD algorithms depends on the noise variance, the number of data streams to be detected and the signal constellation, resulting in high computational costs for low signal-to-noise ratios, high-order constellations and a large number of users.

The high computational complexity of the ML detector and the SD algorithms in the satellite scenarios suggests the deployment of alternative strategies for joint detection, which often rely on signal processing with receive filters. The key advantage of these approaches is that the cost is typically not dependent on the specific modulation, and the gateway can compute the receive filter only once per data packet and perform detection. In the following we will describe the most relevant architectures that seek to accomplish this task.
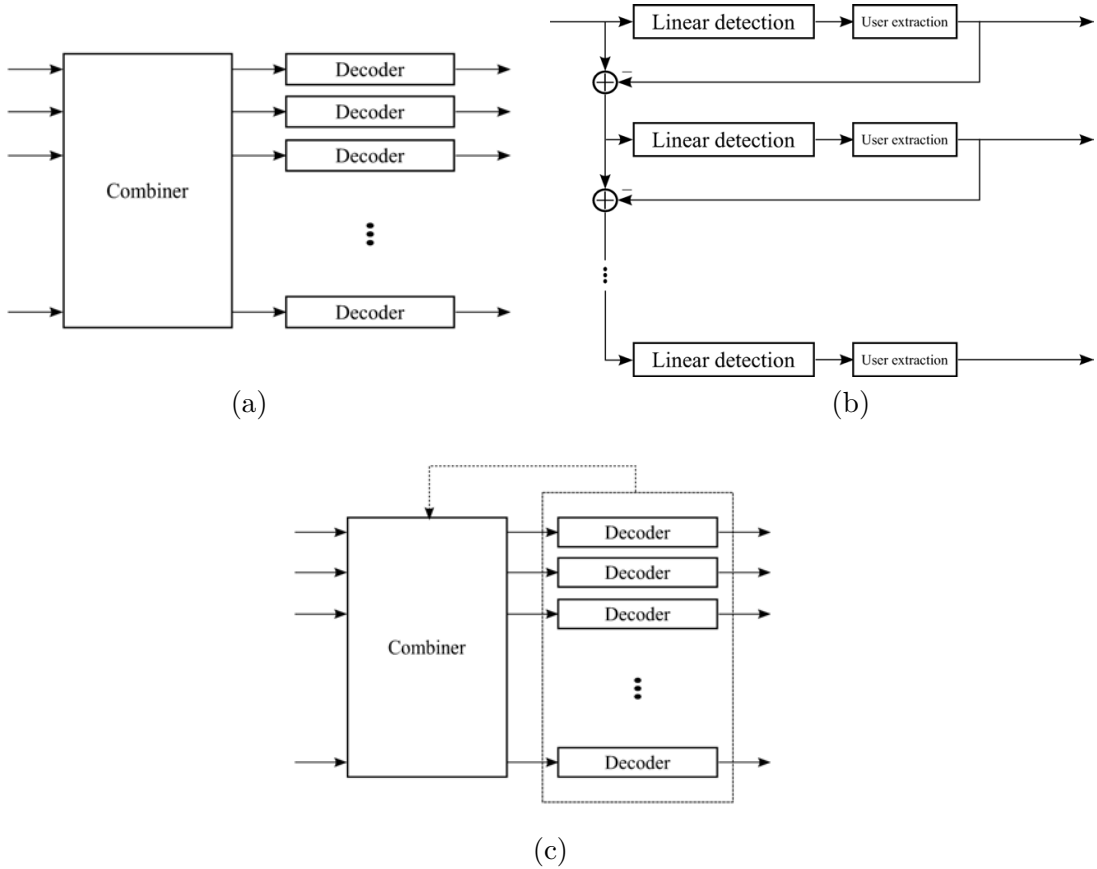
Figure 2.9: Diagram of common receiver architectures.

- **Linear processors** obtain their outputs as linear transformations of the user's input signals, which are also affected by noise. Decoding of the symbols, performed separately for each user's stream, is decoupled from the multiuser detection linear scheme, as shown in Figure 2.9 (a). These receivers, including the zero-forcing (ZF) and minimum mean-squared error (MMSE) detectors, will be reviewed in Section 2.4.2.

- The structure of the data is such that **non-linear detection** can improve the performance of linear algorithms, despite the fact that we are assuming a linear model for the receive signal. Successive interference cancellation (SIC) approaches used in the Vertical-Bell Laboratories Layered Space-Time (VBLAST) systems [34, 35], and decision feedback (DF) [36, 37] detectors, are techniques that can offer attractive trade-offs between performance and complexity. In particular, SIC with MMSE filtering (MMSE-SIC) is known to achieve the best possible performance in terms of sum rate for a multiuser MIMO channel [38]. These techniques, which operate as shown on Figure 2.9 (b), will be reviewed in Section 2.4.2.

- If additional complexity can be afforded, the structure of the coded data can be exploited to improve the MUD processing. Following the turbo-principle, soft-input soft-output (SISO) detection and SISO decoding can be combined, as detailed in Section 2.4.3. This **Iterative Detection and Decoding** (IDD) process is illustrated in Figure 2.9 (c).

In the next section, we will provide a more detailed description of each of the categories above.

2.4.2   Linear and nonlinear detection

Recall that the input-output relation describing the received values is given by

$$\mathbf{y} = \mathbf{Hs} + \mathbf{n}. \tag{2.8}$$

This model is widely used in multiuser, multi-antenna and multiple-access communications, and numerous results are reported in the literature. The properties of the channel matrix $\mathbf{H}$ are instrumental to characterize the performance of the different receiver algorithms, and in most instances this matrix is supposed to be known by the receiver -perfect CSI assumption. Some considerations will be included later in the chapter on the estimation of $\mathbf{H}$.

The use of Equation (2.8) is implicitly assuming a centralized reception of the user signals sharing the spectrum along different beams. This single-gateway assumption may prove realistic in narrow-band systems for which the aggregated throughput of all users can be channelized through a single feeder-link. In broadband systems a more feasible solution is to deploy a number of gateways which reuse feeder link frequencies. Each gateway will be able to handle the signals in a group of beams known as cluster, and full frequency reuse will cause inter-cluster interference in addition to intra-cluster interference. The signal processing complexity at each

gateway is also reduced with respect to that in a centralized gateway. Actually, the complexity of optimal detectors is exponential in the number of users K; we will review a few well-known suboptimum approaches to detect $\mathbf{s}$ in (2.8), first for the case of a centralized reception of all signals in a single gateway, only to drop later this assumption and consider a blind approach to combat the inter-cluster interference.

Linear detectors

The most common family of suboptimal detectors to cancel interference in (2.8) are those performing only linear operations to the samples at the receive antennas, and which implement the detection rule [39]

$$\hat{\mathbf{s}} = Q(\mathbf{W}^H \mathbf{y}) \tag{2.9}$$

where $Q(\cdot)$ represents the slicer associated with the optimal decision regions used for detection and $\mathbf{W}$ is a $K \times K$ matrix characterizing the linear detector.

The minimization of the mean square error,

$$\mathbf{W} = \arg \min_{\mathbf{W}} \mathbb{E} \left[\|\mathbf{s} - \mathbf{W}^H \mathbf{y}\|_2^2 \right], \tag{2.10}$$

gives the following explicit expression for the Minimum MSE **(MMSE)** receive filter:

$$\mathbf{W}_{\text{MMSE}}^H = (\mathbf{H}^H \boldsymbol{\Sigma}^{-1} \mathbf{H} + \mathbf{I})^{-1} \mathbf{H}^H \boldsymbol{\Sigma}^{-1}. \tag{2.11}$$

where $\mathbf{H}$ is the channel matrix, and recall that $\boldsymbol{\Sigma}$ is the noise covariance matrix, as defined in Section 2.2. This receiver is known to maximize the post-filtering SINR [40, 41], and constitutes a popular choice. In addition to the channel matrix $\mathbf{H}$, the noise statistics need to be known.

Another well-known detector results from neglecting the noise and canceling the interference. The corresponding zero-forcing receiver reads as

$$\mathbf{W}_{\text{ZF}}^H = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \tag{2.12}$$

where the knowledge of the noise statistics is not required; this detector is known as the decorrelator detector in the context of CDMA systems.

The covariance matrix of the MMSE detection error is given by [41]

$$\begin{aligned} \mathbf{Q}_{\text{MMSE}} &\doteq \mathbb{E} \left[(\mathbf{s} - \hat{\mathbf{s}}) (\mathbf{s} - \hat{\mathbf{s}})^H \right] \\ &= \left(\mathbf{I} + \mathbf{H}^H \boldsymbol{\Sigma}^{-1} \mathbf{H} \right)^{-1} \end{aligned} \tag{2.13}$$

which is convenient for obtaining each user's individual SINR as a function of the corresponding diagonal element of $\mathbf{Q}_{\text{MMSE}}$:

$$\text{sinr}_\ell = \frac{1}{[\mathbf{Q}_{\text{MMSE}}]_{\ell\ell}} - 1. \tag{2.14}$$

The corresponding snr_ℓ for the ZF receiver, assuming ideal cancellation of the interference, is given by

$$\mathrm{snr}_\ell = \frac{1}{\left[(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{\Sigma}^{-1} \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1} \right]_{\ell\ell}} \tag{2.15}$$

which, when $\mathbf{\Sigma} = N_0 \mathbf{I}$, simplifies to

$$\mathrm{snr}_\ell = \frac{1}{N_0 \left[(\mathbf{H}^H \mathbf{H})^{-1} \right]_{\ell\ell}}. \tag{2.16}$$

Both estimators behave similarly for high SINR situations as easily inferred from their respective expressions.

Note that both cases apply first the term $\mathbf{H}^H$ (preceded by the noise whitening operation in the MMSE receiver), that is, the **spatial matched filter**. The complexity of detection is polynomial in the number of users K, since in both cases a $K \times K$ matrix needs to be inverted. Iterative strategies have been developed to avoid this costly operation, [42]. These recursions can be implemented as multi-stage detection in the form of serial or parallel interference cancellation. If, for simplicity, we assume $\mathbf{\Sigma} = N_0 \mathbf{I}$ in (2.11), we can write

$$\hat{\mathbf{s}} = \mathbf{M}^{-1} \mathbf{H}^H \mathbf{y}, \tag{2.17}$$

with $\mathbf{M} = \mathbf{H}^H \mathbf{H} + N_0 \mathbf{I}$ for the MMSE detector and $\mathbf{M} = \mathbf{H}^H \mathbf{H}$ for the ZF detector. As an example, $\mathbf{M}^{-1}$ can be expressed as a Taylor series,

$$\mathbf{M}^{-1} = \sum_{n=0}^{\infty} (\mathbf{I} - \mathbf{M})^n, \tag{2.18}$$

as long as the spectral radius of $\mathbf{M}$ is less than 2. The practical value of this series depends on the number of terms which need to be used for a good approximation. Many well-known successive detectors can be related to expressions such as (2.18), which, in particular, can be implemented through the following recursion [43]:

$$\hat{\mathbf{s}}^{(n+1)} = \mathbf{H}^H \mathbf{y} + (\mathbf{I} - \mathbf{M}) \hat{\mathbf{s}}^{(n)}, \quad n = 0, 1, \ldots \tag{2.19}$$

with $\hat{\mathbf{s}}^0 = 0$, and which happens to be a parallel interference canceler or, equivalently, the Jacobi's method for solving a linear algebraic system. There are many alternative forms of interference cancellation based on different decompositions of the matrix $\mathbf{M}$. For example, matrix $\mathbf{M}$ can be split in another form giving rise to a serial interference canceler. If $\mathbf{M} \doteq \mathbf{D} + \mathbf{L} + \mathbf{L}^H$, with $\mathbf{D}$ diagonal and $\mathbf{L}$ strictly lower-triangular, then we can rewrite the recursion as

$$\hat{\mathbf{s}}^{(n+1)} = \mathbf{D}^{-1} \left(\mathbf{H}^H \mathbf{y} - \mathbf{L}^H \hat{\mathbf{s}}^{(n)} - \mathbf{L} \hat{\mathbf{s}}^{(n+1)} \right), \quad n = 0, 1, \ldots \tag{2.20}$$

where interference estimates of each user are canceled as soon as they are available. This recursion corresponds to the Gauss-Seidel method. Non-linear detection schemes, introduced in

the next section, can be obtained from the use of non-linear operations such as hard decisions at the output of each cancellation stage.

We have been assuming the perfect knowledge of the channel matrix $\mathbf{H}$ in (2.8) for the derivation of different detection schemes. This knowledge is never perfect in practice, and it is based on the exchange of training sequences, which may not be possible if, for example, users are handled by different isolated gateways. Even in the absence of this information, a single user can be extracted if the corresponding column $\mathbf{h}_i$ in $\mathbf{H}$ is known. Thus, the MMSE detector can be considered as a bank of single user detectors, with

$$\hat{s}_i = \mathbf{w}_i^H \mathbf{y}, \quad i = 1, \ldots, K \tag{2.21}$$

and the single user detector response $\mathbf{w}_i$ for user i given by

$$\mathbf{w}_i = (\mathbf{H}\mathbf{H}^H + N_0 \mathbf{I})^{-1} \mathbf{h}_i, \quad i = 1, \ldots, K, \tag{2.22}$$

assuming again that the noise variance is uniform and uncorrelated along the different antennas. Note that the matrix in the expression of $\mathbf{w}_i$ is the input covariance matrix, which can be estimated from the input data when other users' channel responses are unknown [44]. Alternatively, and given that the MMSE solution coincides with that obtained from the minimization of the output energy, vector $\mathbf{w}_i$ can be expressed as

$$\mathbf{w}_i = \mathbf{h}_i + \mathbf{p}_i \tag{2.23}$$

where $\mathbf{p}_i$ needs to be adapted to become orthogonal to $\mathbf{h}_i$, thus minimizing the output energy. This is known as the **Blind Minimum Output Energy**, which turns out to be quite sensitive to imperfections in the knowledge of $\mathbf{h}_i$, and admits different variants to avoid undesired cancellation [44]. As a remark, this method belongs to the family of subspace-based blind methods.

The formulation in (2.22) can be exploited for the multiple gateway case, for which a given gateway tries to detect a group of users while fighting the interference from adjacent beams managed by other gateways, all this without gaining access to the signals which are being relayed to other gateways. This problem has only started to receive attention recently for multibeam satellites due to their potential for broadband services. Here we sketch a decoupled approach without resorting to the exchange of information among gateways; such exchange would surely help, as pointed out for the forward link in [45].

Let us start by defining a splitting of the signal model as follows:

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{N_{\mathrm{gw}}} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_{N_{\mathrm{gw}}} \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_{N_{\mathrm{gw}}} \end{pmatrix} + \mathbf{n} \tag{2.24}$$

where N_{gw} is the number of gateways, each serving K_n beams. The channel matrix affecting the signals traveling to the n-th gateway, $\mathbf{H}_n$, can be expressed as

$$\mathbf{H}_n = \begin{bmatrix} \tilde{\mathbf{H}}_n & \bar{\mathbf{H}}_n \end{bmatrix} \qquad n = 1, 2, \cdots, N_{\mathrm{gw}} \tag{2.25}$$

where $\tilde{\mathbf{H}}_n \in \mathcal{C}^{K_n \times K_n}$ comprises the channels from the users served by the n-th gateway, and $\bar{\mathbf{H}}_n \in \mathcal{C}^{K_n \times (K - K_n)}$ the rest of them.

Each gateway can estimate $\tilde{\mathbf{H}}_n$, but not $\bar{\mathbf{H}}_n$, which would have to be communicated by other gateways if needed. However, it is possible to reformulate (2.22) as

$$\mathbf{w}_i = \hat{\mathbf{R}}_{\mathrm{y}}^{-1} \tilde{\mathbf{h}}_i \tag{2.26}$$

where $\hat{\mathbf{R}}_{\mathrm{y}}$ is an estimate of the covariance matrix of $\mathbf{y}$. In this way, and as long as the individual gateways can obtain reasonably good estimations of $\hat{\mathbf{R}}_{\mathrm{y}}$, inter-cluster interference could be partially mitigated without gateway intercommunication.

Nonlinear detectors

Performance of linear joint detection can be improved if some non-linear processing is implemented, by either applying some form of successive cancellation, or by approximating the optimum detector. Actually, the successive cancelers can be considered as variants of the iterative implementations shown above, whereas the approximations to the optimum detector can trade performance gain by implementation complexity.

Nonlinear tentative decision functions can be included at the output of each stage of successive cancelers; although initially developed as ad-hoc schemes, they can be related to the recursive linear updates shown above. Thus, iteration (2.19) can be generalized to

$$\hat{\mathbf{s}}^{(n+1)} = \psi \left(\mathbf{H}^H \mathbf{y} + (\mathbf{I} - \mathbf{M})\hat{\mathbf{s}}^{(n)} \right), n = 0, 1, \ldots \tag{2.27}$$

that is, a non-linear operation ψ after each stage of parallel interference cancellation. The function ψ can be implemented as a tentative decision $Q(\cdot)$, in what is known as **parallel interference cancellation (PIC)**. On the other side, if we include a decision step in (2.20), we have

$$\hat{\mathbf{s}}^{(n+1)} = Q \left(\mathbf{D}^{-1} \left(\mathbf{H}^H \mathbf{y} - \mathbf{L}^H \hat{\mathbf{s}}^{(n)} - \mathbf{L}\hat{\mathbf{s}}^{(n+1)} \right) \right), n = 0, 1, \ldots \tag{2.28}$$

a **successive interference cancellation (SIC)** scheme (see Figure 2.10). Different variants of these schemes can be found in the state of the art.

Order of detection.   Note that the way users are ordered in $\mathbf{s}$ affects how interference is subtracted, since at each stage users being decoded are subtracted before processing the remaining users. For the MMSE-SIC receiver, the ordering is irrelevant in terms of capacity, but has an
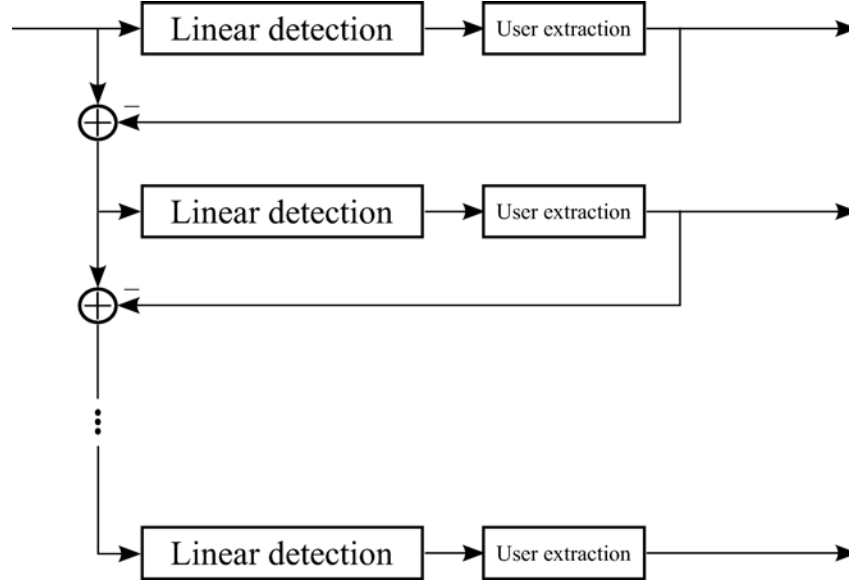
Figure 2.10: Implementation of SIC detection.

effect on the bit error rate, as these systems suffer from error propagation. Since obtaining the optimum ordering would require trying a very large number of combinations, other suboptimal approaches are often employed. For example, ordering by the user's channel norm can be done with very little computational complexity. However, ordering by the output SINR at each state is known to perform better. For the case of ZF-SIC, ordering by the best user at each state is known to be optimal [46].

In this sense, the **multi-branch successive interference cancellation (MB-SIC)** algorithm generalizes the standard SIC algorithm using multiple SIC algorithms in parallel branches [47]. The MB-SIC algorithm relies on different ordering patterns and produces multiple candidates for detection to get a performance closer to that of the ML detector. The ordering of the first branch is identical to a standard SIC algorithm, whereas the remaining branches are ordered by cyclic shifts relative to the first branch.

Complexity considerations

The computational complexity of the different detectors can be obtained as a function of the number of users K. In Table 2.3 we offer an approximate comparison, in the fashion of [48, Table 2]. For this table, the number of points in the constellation is denoted by $|\mathcal{S}|$, and we have assumed that the cost of a matrix inversion is cubic with K, as for example when Gauss-Jordan elimination is used. Note that, for the MMSE-SIC, reduced-complexity implementations exist; see for example [49] for the so-called square-root MMSE-SIC implementation.

Table 2.3:   Approximate computational complexity of different detectors.

	Inverse	Product
MMSE	$\overbrace{2K^3}$ +	$\overbrace{K^2}$
MMSE-SIC	$K(2K^3 + K^2)$	
MAP	$K^2|\mathcal{S}|^K$	

2.4.3   Other alternatives: iterative detection

Iterative detection and decoding (IDD) schemes have received considerable attention following the discovery of turbo codes [50], and the use of the turbo principle for mitigation of several sources of interference [20, 50–53]. More recently, work on IDD schemes has been extended to low-density parity-check codes (LDPC) [52, 54] and their variants.

As anticipated in Section 2.4.1, IDD systems combine an efficient soft-input soft-output detection algorithm and decoding technique. The detector produces log-likelihood ratios associated with the encoded bits, and these serve as input to the decoder which, in a second phase, generates a posteriori probabilities, which are then fed back; the whole process is repeated for a number of iterations.

For practical reasons, iterative detection schemes were not analyzed further in this thesis; notice that, because of the large number of users present in multibeam satellite communications, the computational complexity of these schemes is very high.

2.5   Analyzing performance

In this section we review the ultimate performance limits experienced by the return link of multibeam satellite systems, which can be considered as an instance of a multiuser MIMO (MU-MIMO) communication setting. We will briefly address the behavior of the sum rate, and the implications of having a large number of terminals and antennas. A more detailed theoretical study of the channel is left for Chapter 3.

For notational convenience, we will make the signal-to-noise ratio explicit in the following equations. To do this, let us define the equivalent signal model

$$\mathbf{y} = \sqrt{\gamma}\mathbf{H}\mathbf{s} + \mathbf{n} \tag{2.29}$$

where now $\mathbf{n}$ is assumed to have unit variance, and γ represents the transmit power over receiver noise (SNR).

At a given time instant, the sum of the rates of the users, R_ℓ, is known to be bounded
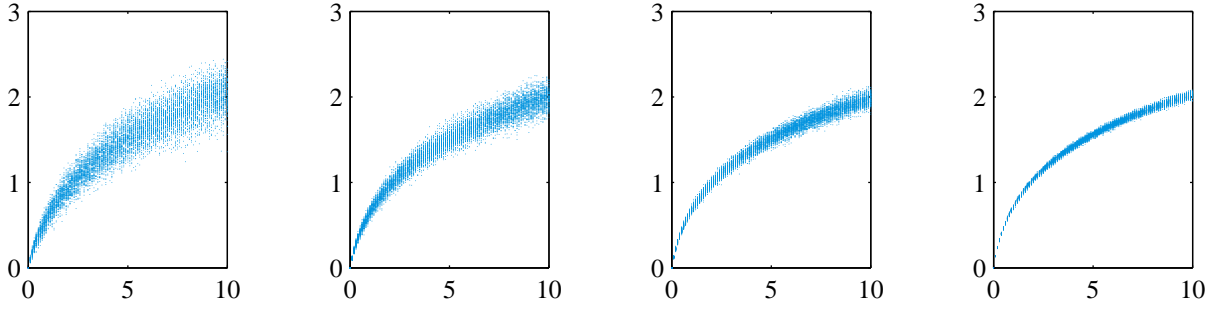
Figure 2.11: Realizations of $1/\tilde{K} \cdot \log \det \left(\mathbf{I} + \gamma \tilde{\mathbf{H}}^H \tilde{\mathbf{H}}\right)$ as a function of γ; from left to right, $\tilde{K} \in \{3, 7, 19, 100\}$.

by [55]

$$C_{\text{sum}} \doteq \sum_{\ell=1}^{K} R_\ell \leq \log_2 \det \left(\mathbf{I} + \gamma \mathbf{D}^2 \mathbf{G}^H \mathbf{G}\right)$$
$$= \sum_{\ell=1}^{K} \log_2 \left(1 + \gamma \lambda_\ell \left\{\mathbf{D}^2 \mathbf{G}^H \mathbf{G}\right\}\right) \tag{2.30}$$

where $\lambda_\ell \{\}$ denotes the ℓ-th largest eigenvalue of a matrix.

The quantity above, which is the mutual information conveyed by the channel at a given time instant, is of paramount importance for determining the theoretical performance limits of a communication system. However, characterizing it is a difficult task because of the peculiar channel model: modeling C_{sum} requires knowing the eigenvalues of the matrix product $\mathbf{D}^2 \mathbf{G}^H \mathbf{G}$, where $\mathbf{D}$ is random diagonal and $\mathbf{G}$ has a complicated expression. Recall that this is a major difference with respect to terrestrial multiuser communications, where fading and atmospheric attenuations cannot be put into a diagonal matrix.

The field of random matrix theory (RMT) [56, 57] can shed some light into this problem. When K grows large, the empirical distribution of the eigenvalues has been proven to converge to a deterministic function for many channel models. If this is the case, it is possible to rewrite (2.30) as [56]

$$\lim_{K \to \infty} C_{\text{sum}} = \lim_{K \to \infty} \sum_{\ell=1}^{K} \log_2 \left(1 + \gamma \lambda_\ell \left\{\mathbf{D}^2 \mathbf{G}^H \mathbf{G}\right\}\right)$$
$$= K \mathbb{E}_{\mathbf{D}} \left[\log_2 \left(1 + \gamma \lambda \left\{\mathbf{D}^2 \mathbf{G}^H \mathbf{G}\right\}\right)\right] \tag{2.31}$$
$$= K \int \log_2 \left(1 + \gamma x\right) \, \mathrm{d} F_{\mathbf{D}^2 \mathbf{G}^H \mathbf{G}}(x)$$

where $F_{\mathbf{D}^2 \mathbf{G}^H \mathbf{G}}(x)$ is the cumulative distribution function (CDF) of the eigenvalues of matrix $\mathbf{D}^2 \mathbf{G}^H \mathbf{G}$. This is illustrated in Figure 2.11, which shows different realizations of $1/\tilde{K} \cdot \log \det \left(\mathbf{I} + \gamma \tilde{\mathbf{H}}^H \tilde{\mathbf{H}}\right)$ as a function of γ. $\tilde{\mathbf{H}}$ are square submatrices of $\mathbf{H}$ of length $\tilde{K}$. We can see that the depicted quantity tends to be deterministic as $\tilde{K}$ grows large.

The equation above implies that, as K grows large, C_{sum} converges to an expectation, which is a deterministic value. Unfortunately, this value is still difficult to compute, and for some models of interest convergence may not hold, or will be precarious for common values of K. Among the attempts made through random matrix theory, [58] obtained C_{sum} for a model similar to the one used here, but with *asymptotically free* [56] matrices $\mathbf{G}$ and $\mathbf{D}$, and [6] resorted to an approximation of the eigenvectors of $\mathbf{G}$; in either case the results offer little insight.

Alternatively, it is possible to obtain tight bounds, or high and low SNR approximations, exploiting the large dimensions of the system. To provide some insight, in Appendix 2.A we will give some results illustrating the former option; high and low SNR approximations, with and without correlation in the entries of $\mathbf{D}$, will require a chapter of their own.

2.6 A comparison between hybrid and full on-ground mutiuser detection

Within the SatNEx III (Satellite Network of Experts) framework [15], funded by the European Space Agency, and building on previous results such as those cited in the introduction, the use of multiuser interference mitigation schemes together with ground-based beamforming (GBBF) was analyzed to evaluate its potential improvement with respect to more classic on-board beamforming settings. The main achievements of this research, which have been partly presented in [17, 59], are reported in detail in this section.

2.6.1 Description of a hybrid architecture

A key objective of the study was to evaluate the performance of hybrid on-board on-ground architectures. The main idea is depicted in Figure 2.12. The satellite is equipped with N feeds, with $N \geq K$ and K the number of users. Since relaying all N signals to Earth would require a high feeder link bandwidth, we consider the possibility of applying a fixed linear processing at the satellite, represented by matrix $\mathbf{B} \in \mathbb{C}^{N \times K'}$, that outputs only K' signals, so that $K \leq K' < N$; the reduced set of K' signals is then sent to the gateway, which applies multiuser processing to obtain K signals. We shall remark that the processing is considered fixed because making it reconfigurable increases the payload complexity (the latter can be done, for example, in mobile satellite systems using a much smaller bandwidth [60]).

Throughout this work, only the case $K' = K$ was considered. We will now describe briefly some terminology before reporting the obtained results.

- We will use the expressions *full on-ground processing* or *feed processing* when all the N signals are sent to the gateway. In that case, we will say that signals are processed at *feed level*.

- When fixed beamforming is present, we will say that there is *hybrid processing* or *beam*
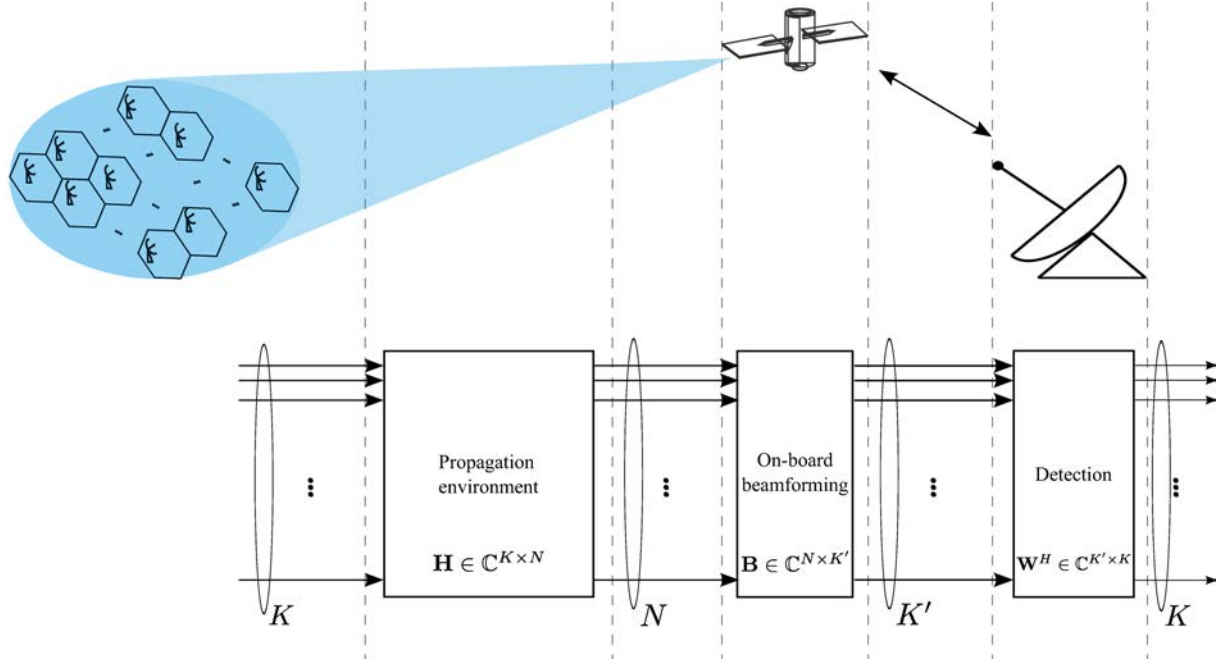
Figure 2.12: Graphical description of an architecture with a different number of feeds and beams

processing, or that the signals are processed at *beam level*.

In the following subsections, we will describe the differences in performance obtained by both options. It will be shown that full on-ground architectures tend to outperform hybrid architectures with fixed on-board weights and on-ground multiuser interference mitigation. This conclusion will be analytically supported under perfect channel state information, and a sufficient condition for the two architectures to be equivalent will be given. Moreover, detailed simulations will provide some first intuition on the behavior of both architectures when imperfect channel knowledge is assumed, and also when the analog beamforming suffers from miscalibration issues.

2.6.2 Perfect CSI at the gateway

As a first step, let us assume that the gateway has perfect knowledge of the channel state. Throughout this section, we will establish measures of performance and show that, under linear combining, feed processing outperforms beam processing.

Let us briefly explain the model used. $\mathbf{B}$ will denote the fixed on-board beamforming matrix, as explained in Section 2.2.1. We will denote $\mathbf{H}$ the channel matrix at feed level and $\mathbf{H}_b = \mathbf{B}\mathbf{H}$ the channel matrix at beam level. For clarity, through this section we will add a subindex to the identity matrix, $\mathbf{I}_t$, denoting that it is of size $t \times t$.

For the return link, and assuming on-ground feed processing, the MMSE combiner that yields $\hat{\mathbf{s}} = \mathbf{W}^H \mathbf{y}$ is

$$\mathbf{W}^H = \mathbf{H}^H \left(N_0 \mathbf{I}_N + \mathbf{H}\mathbf{H}^H \right)^{-1} = \left(N_0 \mathbf{I}_K + \mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \tag{2.32}$$

whereas the processing of the beams would entail $\hat{\mathbf{s}} = \mathbf{W}_b^H \mathbf{y}_b$ with

$$\mathbf{W}_b^H = \left(\mathbf{I}_K + \mathbf{H}_b^H \mathbf{\Sigma}^{-1} \mathbf{H}_b \right)^{-1} \mathbf{H}_b^H \mathbf{\Sigma}^{-1}. \tag{2.33}$$

Recall that a key objective of this study is to compare the performance of these two approaches. To accomplish this task, we will make use of the mean-squared error (MSE) after combining, which is defined as $\mathbf{E}\{||\mathbf{s} - \hat{\mathbf{s}}||^2\}$. Let us denote $\mathbf{Q}_f$ its covariance matrix, then it would read

$$\mathbf{Q}_f = \mathbf{R}_x - \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{yx} = \left(\mathbf{I}_K + \frac{1}{N_0}\mathbf{H}^H\mathbf{H} \right)^{-1} \tag{2.34}$$

for the case of feed processing. For the case of beam processing, it would be

$$\mathbf{Q}_b = \left(\mathbf{I}_K + \mathbf{H}_b^H \frac{\left(\mathbf{BB}^H\right)^{-1}}{N_0} \mathbf{H}_b \right)^{-1} = \left(\mathbf{I}_K + \frac{1}{N_0}\mathbf{H}^H\mathbf{PH} \right)^{-1} \tag{2.35}$$

with $\mathbf{P} = \mathbf{B}^H \left(\mathbf{BB}^H \right)^{-1} \mathbf{B}$.

Since the SINR for the ith user is given by $1/\mathbf{Q}_{ii} - 1$, it makes sense to use the total MSE, given by $\sum_{i=1}^K \mathbf{Q}_{ii} = \text{trace}\{\mathbf{Q}\}$, as a performance metric. Then, it can be shown that

$$\text{trace}\{\mathbf{Q}_b\} \geq \text{trace}\{\mathbf{Q}_f\} \tag{2.36}$$

as follows. Let us express both traces as

$$\text{trace}\{\mathbf{Q}_f\} = \sum_{i=1}^K \lambda_i(\mathbf{Q}_f) = \sum_{i=1}^K \frac{N_0}{N_0 + \lambda_i(\mathbf{H}^H\mathbf{H})} \tag{2.37}$$

and

$$\text{trace}\{\mathbf{Q}_b\} = \sum_{i=1}^K \lambda_i(\mathbf{Q}_b) = \sum_{i=1}^K \frac{N_0}{N_0 + \lambda_i(\mathbf{H}^H\mathbf{PH})} \tag{2.38}$$

where $\lambda_i(\mathbf{H}^H\mathbf{H})$ denotes the i-th largest eigenvalue of $\mathbf{H}^H\mathbf{H}$. We have that (2.36) is an immediate consequence of the following, stronger result.

Theorem 1. *Let* $\mathbf{H}$ *and* $\mathbf{B}^H$ *be two tall matrices of the same size with full column rank. Let* $\mathbf{P}$ *be a projection matrix of* $\mathbf{B}$*, that is,* $\mathbf{P} = \mathbf{B}^H \left(\mathbf{BB}^H \right)^{-1} \mathbf{B}$*. Then, it holds that* $\sigma_i(\mathbf{H}) \geq \sigma_i(\mathbf{PH})$*, with* $\sigma_i(\mathbf{A})$ *denoting the ith largest singular value of matrix* $\mathbf{A}$*.*

Proof. See Appendix 2.B.   □

A sufficient condition for the traces to be equal is that $\mathbf{PH} = \mathbf{H}$. Since $\mathbf{P}$ is a projection matrix, this will happen whenever $\text{range}(\mathbf{B}^H) = \text{range}(\mathbf{H})$. However, since $\mathbf{B}$ is fixed and $\mathbf{H}$ time-varying, it seems not possible to meet such condition. Recall that, even if the fading is negligible, the users are assumed to be located randomly into their beam spots. Thus, assuming

constant $\mathbf{H}$ would require the feed pattern to be constant over each cell; it would also require users near the border of their cells to experiment almost the same interference as if they were located close to the center. On account of all these facts, ensuring range($\mathbf{B}^H$) = range($\mathbf{H}$) is not realistic.

2.6.3   Non-perfect CSI at the gateway

In a realistic scenario, the gateway does not know the actual values in the channel matrix $\mathbf{H}$, but has only an estimate of them. The type and quality of these estimates, commonly based on the use of training sequences, will have an effect on the ultimate performance of the system. In this section, the estimation of the matrix $\mathbf{H}$ will be introduced. Some degree of uncertainty on the beamforming matrix $\mathbf{B}$ will also be discussed, while the noise power N_0 will be assumed perfectly known.

Imperfect knowledge of the channel matrix

The origins and implications of imperfect channel estimation will be discussed in more detail in Chapter 4. For the time being, let us simply consider the following models for the estimated channel matrix:

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E} = \mathbf{H} + \sqrt{\frac{N_0}{L}}\mathbf{W} \tag{2.39}$$

without fixed beamforming (that is, when processing is performed at feed level) where L is the training sequence length and $\mathbf{W}$ is a matrix with independent zero-mean unit-variance Gaussian entries. For the on-board beamforming case, it would read

$$\hat{\mathbf{H}}_b = \mathbf{H}_b + \mathbf{E}_b = \mathbf{H} + \sqrt{\frac{N_0}{L}}\mathbf{B}\mathbf{W}. \tag{2.40}$$

Analog miscalibration

For the case of beam processing, the gateway must be aware of the exact beamforming weights that are set on board, since they will be necessary to compute the noise covariance matrix in the MMSE combiner (2.33). Even though these are subject to calibration, it is very likely that their actual values will experiment some minor changes through time, mainly because of the non-ideal nature of the analog circuitry. As a consequence, the information at the gateway can be seen as an estimate of the actual beamformer. The following mathematical model is proposed:

$$\hat{\mathbf{B}} = \mathbf{B} + \Delta\mathbf{B} \tag{2.41}$$

where the entries in $\Delta\mathbf{B}$ are of the form $\Delta\mathbf{B}_{ij} = b_{ij}r_{ij}$ and r_{ij} are independent, real, zero-mean Gaussian random variables. This models a variation, both in the real and imaginary parts of the weights, that is random with given variance, but proportional to the original value.
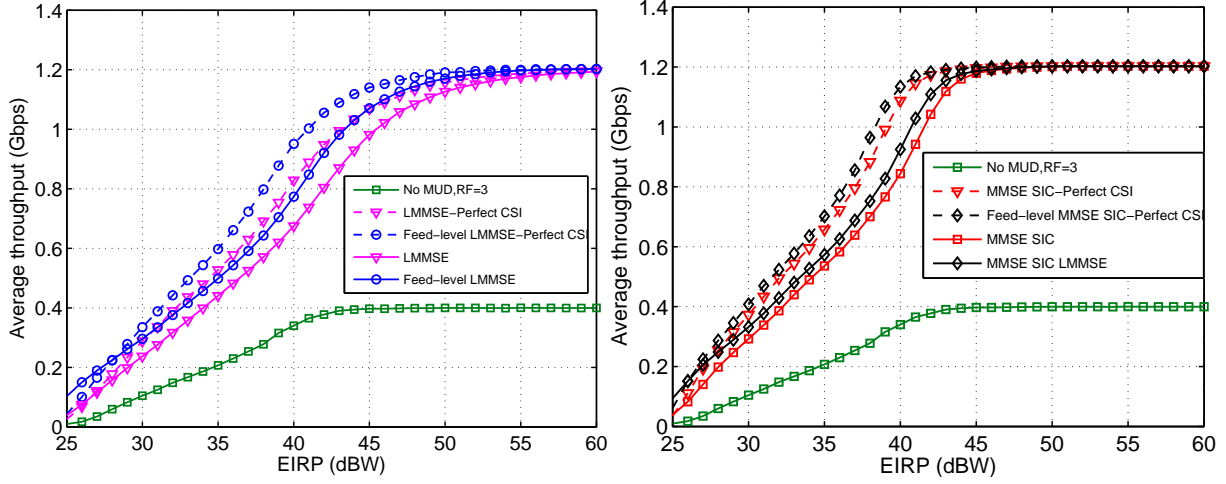
Figure 2.13: Evolution of throughput as a function of terminal EIRP with (right) and without (left) nonlinear processing.

2.6.4   Simulation results

In order to further compare the performance of the proposed and multiuser detection architectures, Monte Carlo simulations have been carried out according to the scenario already described in Table 2.1. Recall that this scenario features $K = 100$ beams covering the whole Europe area. The satellite antenna pattern was provided by ESA, and corresponds to an array fed reflector antenna with $N = 155$ feeds. Matrix **B** was also provided by ESA, as a typical beamforming matrix of current systems. It was designed such as to limit the level of interference among users in a conventional system (without interference mitigation technique but with an adequate frequency reuse pattern).

The reference scenario consists in a frequency reuse factor equal to 3, fixed beamforming and no processing at the gateway. Simulation results have been extracted for a number of interference mitigation techniques. The purpose of this is to compare the performance of both architectures in as many different situations as possible. Simulation results cover a large range of transmit powers, keeping in mind that the most extreme values do not correspond to practical cases.

Figure 2.13 depicts the evolution of the total average throughput as a function of the terminals equivalent isotropic radiated power (EIRP). Results have been averaged only for those realizations in which the link was active. To this extent, Figure 2.14 shows the average probability of non-availability for the different MUD techniques. It can be observed that a considerable increase in throughput is obtained thanks to using multiuser detection, although at the cost of some loss in availability. In fact, only SIC detection manages to reduce the outage probability with respect to the benchmark scenario.

Moreover, full on-ground processing reports higher throughput figures both with perfect and non-perfect CSI. To further investigate the potential advantage of this strategy, Figure 2.15 (left)
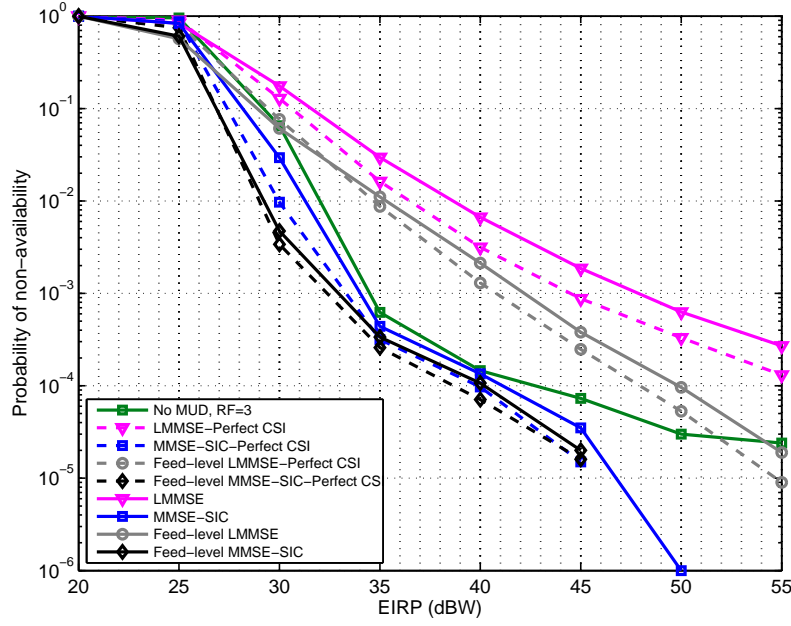
Figure 2.14: Average probability of non-availability in the return link as a function of terminal EIRP.

represents the performance gain obtained in this case with respect to the hybrid architecture, using training sequences of length 128 symbols. Let t_f and ρ_f be the average throughput and average availability, respectively, of the overall system when full on-ground feed processing is employed, and define t_b and ρ_b as the corresponding counterparts in the beam processing case. The feed combining gain is defined as

$$\gamma \doteq \frac{t_f \rho_f}{t_b \rho_b}. \tag{2.42}$$

It can be seen that, despite the existence of channel estimation errors, there are always non-negligible improvements when choosing a full on-ground architecture. Moreover, results on the hybrid architecture assumed so far perfect knowledge of the fixed beamforming matrix. Recall now the error model (2.41) for the analog calibration, given by the error matrix $\Delta\mathbf{B}_{ij} = b_{ij}r_{ij}$ and let β be the variance of the random variables r_{ij}. Following this model, Figure 2.15 (right) depicts the evolution of the feed combining gain for different values of β, that is, for different degrees of uncertainty on the analog beamforming weights, both with linear MMSE and MMSE-SIC receivers. Results account for more significant feed combining gains when some degree of uncertainty is present.

The results obtained in the previous sections show that feed-level techniques tend to out-perform beam processing ones, a fact that has been analytically proved for the case of perfect CSI. This gain may be rather small when the channel estimation errors are noticeable but the estimation of the on-board analog beamforming is accurate. On the contrary, if the uncertainty on the beamforming is high, then the advantage for using all the information from the feeds seems to be much higher: with the greatest level of uncertainty simulated, the feed combining gain reaches 20% at an EIRP 40 dBW. It is to be noticed, however, that working with the 155
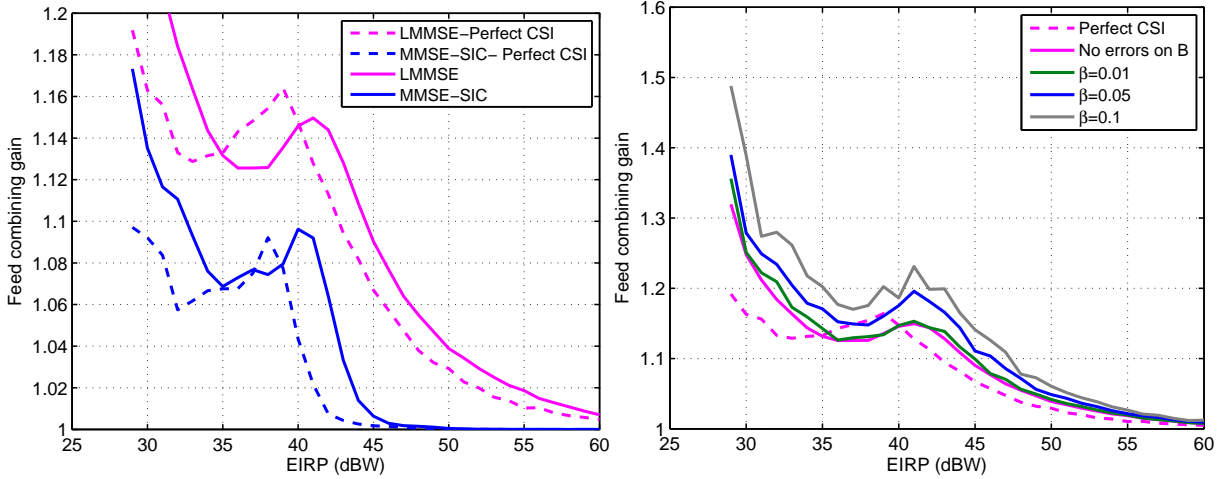
Figure 2.15: Feed combining gain γ as a function of the terminal EIRP with perfect (left) and imperfect (right) knowledge of the beamformer.

feed signals would require rather more bandwidth in the feeder link. Therefore, there exists a tradeoff between performance and feeder link requirements, and the choice of the most suitable processing architecture would need to take into account all these considerations.

2.7 Limitations in the feeder link

Before applying any on-ground processing technique, the signals received at the satellite must be relayed to the gateway on Earth. This feeder link communication is frequently considered to be transparent, although it has some limitations with important consequences for the global system performance. As an example, [60] points out the implications of applying different phase variations to the signals being relayed, even when their transmission is error free.

However, the two most significant issues from the feeder link are its required availability and its limited capacity. The feeder link needs to be designed to guarantee an available communication during most of the time, which may be put in risk in very high frequency bands such as Ka and, especially, Q/V-band, due to atmospheric fading. The Q/V-band is especially attractive due to its available 5 GHz of bandwidth, significantly more than the 2 GHz in the Ka-band; for the same reasons, the W-band (70-80 GHz) is also under study. Together with transmit power margins, other fade mitigation techniques can be required to prevent frequent systems outages, even gateway diversity to guarantee a functional feeder link at any moment. A considerable amount of work has been carried out in modeling the outage of multi-gateway systems, focusing specially on the effects of rain attenuation (see [61, 62] and references therein).

Multiple gateways can also be needed to increase the feeder link capacity. If B_b denotes the available bandwidth per beam, K the number of beams, and f_r the beam frequency reuse factor (reuse one meaning that all the beams share the same frequency band), then the feeder link requires a bandwidth of at least $K \cdot B_b / f_r$ Hz. In turn, this means that *the feeder link requires*
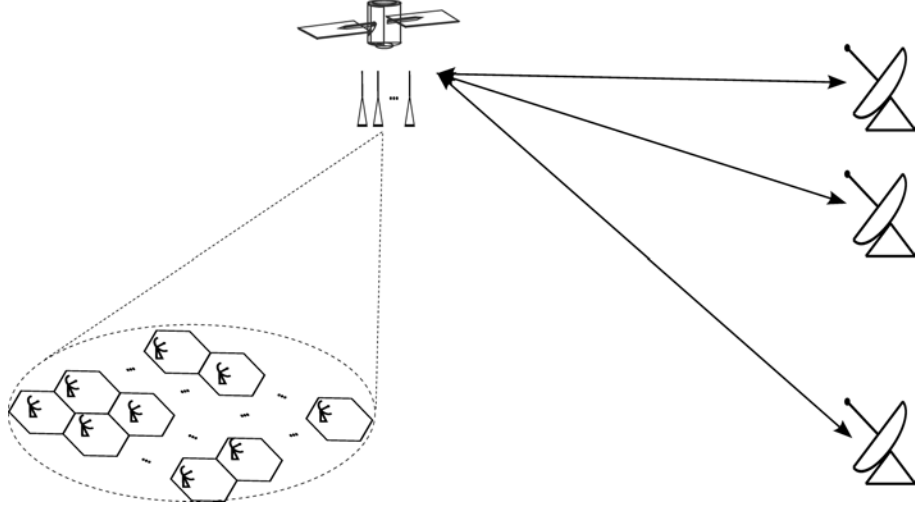
Figure 2.16: Diagram of a multibeam satellite system served by multiple gateways. A single receive antenna is usually employed in the feeder link.

an order of magnitude of bandwidth more than the user link. If the use of higher frequency bands is not enough, multiple gateways will have to be deployed as sketched on Figure 2.16. The required number of gateways would be

$$\left\lceil \frac{B_\mathrm{b} \cdot K}{f_\mathrm{r} \cdot B_\mathrm{f}} \right\rceil \tag{2.43}$$

which can be extremely large. As illustrated in Figure 2.17, for 100 beams and a total beam bandwidth of 500 MHz, 25 gateways would be required for a full frequency reuse setting.

If gateways are sufficiently separated on Earth, then interference can be avoided by means of very directive antennas (recall that gateway antennas can be very large), so the same feeder link bandwidth can be reused; the contrary would imply applying partial frequency reuse –with the consequent loss in spectral efficiency– or using interference mitigation techniques. Each gateway would manage a set of beams or cluster, and could apply multibeam detection schemes to the corresponding signals. As noticed earlier, severe inter-cluster interference might result from a full frequency reuse configuration. Coordination between gateways has been proposed to mitigate these effect, in the fashion of the terrestrial literature [45, 63]; this coordination could be eventually implemented by means of high-speed fiber-optic links. As already mentioned in Section 2.4.2, blind multiuser detection methods can also be considered to help mitigating the inter-cluster interference.

2.8   Conclusions

In this chapter we have introduced the system model for multibeam satellite communications, and explained the baseline scenario with partial frequency reuse. To overcome the limitations of this scenario, different multiuser detection techniques have been reviewed, and the ultimate
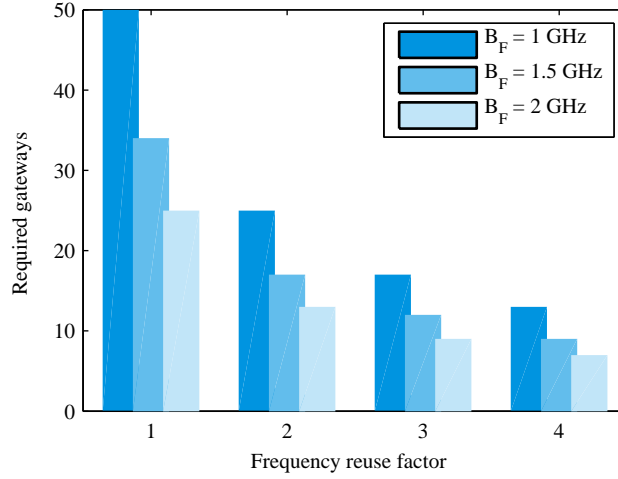
Figure 2.17: Required gateways, as a function of the frequency reuse factor, for different values of the feeder link bandwidth B_{F} and $K = 100$.

performance of such a joint processing system has been sketched. Finally, we have explained the fundamental performance limitations introduced by on-board processing, and the effects of having limited bandwidth in the feeder link.

Some of the contents in this chapter were published in the EURASIP Journal on Satellite Communications and Networking [21] and in *IEEE Communications Letters* [64]. Parts of this work were done in collaboration with Dimitrios Christopoulos, Dr. Symeon Chatzinotas, Prof. Carlos Mosquera and Prof. Björn Ottersten.

Appendix 2.A   Example of an approximation to the performance

Theorem 2. *The per user MMSE of an uplink MU SIMO system can be approximated by*

$$\hat{\epsilon}^2 = \frac{1}{1 + \gamma \exp\left(\frac{1}{K} \log \det\left(\mathbf{H}^H \mathbf{H}\right)\right)}, \tag{2.44}$$

relative to a specific channel instance.

Proof. In [65], an explicit relationship between the channel mutual information $\mathrm{I}_e = \log_2 \det\left(\mathbf{I} + \gamma \mathbf{H}^\dagger \mathbf{H}\right)$ in bits/sec/Hz and ϵ^2 was derived by differentiating with respect to the SNR. This result is easily extended for vector channels yielding [56]

$$\gamma \frac{\partial}{\partial \gamma} \mathrm{I}_e(\gamma) = K - \mathrm{trace}\left(\left(\mathbf{I}_K + \gamma \mathbf{H}^H \mathbf{H}\right)^{-1}\right) \tag{2.45}$$

$$= K(1 - \epsilon^2). \tag{2.46}$$

Furthermore, and with the use of Minkowski's inequality, a bound which is tight over the whole

SNR region and exact at high SNR,I_e, has been derived in [66]:

$$I_e \geq I_{lb} = K \log_2 \left(1 + \gamma \det \left(\left(\mathbf{H}^H \mathbf{H} \right)^{1/K} \right) \right). \tag{2.47}$$

By differentiation with respect to γ we get the following:

$$\gamma \frac{\partial}{\partial \gamma} I_{lb} = \frac{\gamma K \det \left(\left(\mathbf{H}^H \mathbf{H} \right)^{1/K} \right)}{1 + \gamma \det \left((\mathbf{H}^H \mathbf{H})^{1/K} \right)} \implies$$

$$\hat{\epsilon}^2 = \frac{1}{1 + \gamma \det \left((\mathbf{H}^H \mathbf{H})^{1/K} \right)}. \tag{2.48}$$

Finally (2.48) can be rewritten as in (2.44). $\qquad\square$

Since the differentiation does not preserve the direction of the bound, the characteristics of this approximation will be studied in more detail in the following lemma.

Theorem 3. *Let $\hat{\epsilon}^2$ be the derived approximation of ϵ^2, then $\hat{\epsilon}^2$ and ϵ^2 will have a crossing point, at the point where the initial bound is least tight.*

Proof. Denoting as λ_i the ordered eigenvalues of $\mathbf{H}^H \mathbf{H}$, let us define the function $D(\gamma) \doteq I_e(\gamma) - I_{lb}(\gamma)$, where $I_e(\gamma) = \sum_i \log(1 + \gamma \lambda_i)$ and $I_{lb}(\gamma) = \log \left(1 + \gamma \left(\prod_i \lambda_i \right)^{1/K} \right)$, with the following properties: 1) $D(0) = 0$. 2) For γ_o sufficiently large, but finite, we have that $D(\gamma_o) = 0$. From properties 1 and 2, a straightforward application of Cauchy's mean-value theorem yields that there will be at least one point $\gamma^* \in (0, \gamma_o)$, with zero derivative and subsequently a crossing point between the approximation and the actual function. $\qquad\square$

Theorem 4. *The average over the channel realizations, per user MMSE can be approximated by*

$$\hat{\epsilon}^2_{avg} = \frac{1}{1 + \gamma \exp \left(\frac{1}{K} \mathbb{E}_{\mathbf{H}} \left[\log \det \left(\mathbf{H}^H \mathbf{H} \right) \right] \right)}, \tag{2.49}$$

Proof. Let us assume the function $\phi(x) = (1/(1 + \exp(x)))$ which is convex for $x < 0$ and concave for $x > 0$. By applying Jensen's inequality over the two regions we get that $\mathbb{E}_{\mathbf{H}} \left[\hat{\epsilon}^2 \right] \approx \hat{\epsilon}^2_{\text{avg}}$. $\qquad\square$

Theorem 5. *The per user average MMSE spectral efficiency of a multibeam satellite system under rain fading can be approximated by the closed form formula*

$$\hat{\epsilon}^2_{avg\text{-}rain} = \frac{1}{1 + \gamma \exp \left(\frac{1}{K} \log \left(\det \mathbf{G}^2 \right) + \mu_l \right)}, \tag{2.50}$$

where μ_l the mean of the equivalent log-normal distribution.

Proof. Following the same line of reasoning as in the proof of Lemma 3:

$$\hat{\epsilon}^2_{\text{avg-rain}} \approx \frac{1}{1 + \gamma \exp\left(\frac{1}{K}\mathbb{E}_{\mathbf{H}}\left[\log\det\left(\mathbf{G}^2\mathbf{D}^2\right)\right]\right)}. \tag{2.51}$$

where $\log\det(\mathbf{G}^2\mathbf{D}^2) = \log\left(\det\mathbf{G}^2\right) + \log\det\mathbf{D}^2$ since the matrices are square and $\mathbb{E}\left[\log\left(\det\mathbf{D}^2\right)\right] = \mathbb{E}\left[\sum_{i=1}^{K}\log\delta_i^2\right] = K \cdot \mu_l$. Since $\mathbf{G}$ is a deterministic matrix, from the above analysis, (2.50) is deduced. □

Appendix 2.B   Proof of Theorem 1

The goal is to prove

$$\sigma_i(\mathbf{H}) \geq \sigma_i(\mathbf{PH}) \tag{2.52}$$

with $\sigma_i(\mathbf{A})$ representing the ith largest singular value of matrix $\mathbf{A}$.

Consider now $\mathbf{B} = \mathbf{U\Sigma V}^H$, the singular value decomposition (SVD) of $\mathbf{B}$, then

$$\mathbf{P} = \mathbf{B}^H(\mathbf{BB}^H)^{-1}\mathbf{B} = \mathbf{V\Sigma}^H(\mathbf{\Sigma\Sigma}^H)^{-1}\mathbf{\Sigma V}^H = \mathbf{V}\begin{pmatrix} \mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^H = \mathbf{V\Phi V}^H. \tag{2.53}$$

Since $\mathbf{V}$ is a unitary matrix, then it holds that

$$\sigma_i(\mathbf{VH}) = \sigma_i(\mathbf{H}) \tag{2.54}$$

and

$$\sigma_i(\mathbf{V}^H\mathbf{\Phi VH}) = \sigma_i(\mathbf{\Phi VH}). \tag{2.55}$$

Let us define now $\mathbf{A} \doteq \mathbf{VH}$. On account of the previous statements, proving (2.52) is equivalent to proving

$$\sigma_i(\mathbf{A}) \geq \sigma_i(\mathbf{\Phi A}) \tag{2.56}$$

that is, each singular value of a matrix $\mathbf{A}$ is larger or equal to that of the same matrix after setting some rows to zero. To prove this fact, we will make use of the following property: let $\mathbf{A}$ be in general any tall matrix such that $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \ldots \ \mathbf{a}_k]$ and define $\mathbf{A}_r = [\mathbf{a}_1 \ \mathbf{a}_2 \ \ldots \ \mathbf{a}_r]$, then for all r from 1 to $k-1$ it holds that [67]

$$\sigma_1(\mathbf{A}_{r+1}) \geq \sigma_1(\mathbf{A}_r) \geq \sigma_2(\mathbf{A}_{r+1}) \geq \cdots \geq \sigma_r(\mathbf{A}_{r+1}) \geq \sigma_r(\mathbf{A}_r) \geq \sigma_{r+1}(\mathbf{A}_{r+1}). \tag{2.57}$$

This interlacing property will prove useful for our purpose even though matrix $\mathbf{A}$ loses rows and not columns. Recall now that matrix $\mathbf{A}$ is of size $N \times N$. If we write

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} \tag{2.58}$$

with both block matrices of size $K \times N$, and

$$\mathbf{\Phi}\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{0} \end{pmatrix} \tag{2.59}$$

then it is possible to define

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_1^H & \mathbf{A}_2^H \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{2.60}$$

whose singular values take the form $\sigma(\tilde{\mathbf{A}}) = [\sigma_1(\mathbf{A}) \ \sigma_2(\mathbf{A}) \ \dots \ \sigma_K(\mathbf{A}) \ 0 \ 0 \ \dots \ 0]$. If we now remove columns from the right, the interlacing property tells us that

$$\sigma_i\left(\begin{pmatrix} \mathbf{A}_1 \\ \mathbf{0} \end{pmatrix}\right) \leq \sigma_i(\tilde{\mathbf{A}}) \tag{2.61}$$

which implies that

$$\sigma_i(\mathbf{\Phi}\mathbf{A}) \leq \sigma_i(\mathbf{A}) \tag{2.62}$$

and concludes the proof. $\qquad\square$

Chapter 3

Multiuser detection under rain attenuation

Contents

3.1 Introduction

Rain attenuation is known to be the major impairment affecting satellite and LOS terrestrial communications in bands above 10 GHz [25]. These links' capacity [68] has been exploited in many ways, but adaptive coding and modulation (ACM) has proven to be the most efficient one.

As we know, such capacity is becoming increasingly valuable: as increasing volumes of multimedia contents are demanded through satellite links, higher throughput and improved availability are being sought, shifting the payload to higher bands, exploiting multibeam coverages, and even envisaging joint multi-user processing [17, 21, 22, 45, 64, 69].

So far, studies on the performance of such systems have assumed uncorrelated rain attenuation among different beams [21, 45], which is considered to be accurate for large beam diameters [21, 70]. However, spatial correlation of rain attenuation is known to degrade performance of other wireless systems, both satellite and terrestrial. In satellite communications, its impact has been widely studied in the context of site diversity [61, 71, 72] and SIMO and MISO broadband transmission [73]. For terrestrial applications, available studies include dual-hop relay systems [74], MIMO broadband communications [62] and cellular systems [75].

In this chapter, we study the impact of spatially correlated rain attenuation over a multi-beam satellite return link with full frequency reuse –all the beams using the same frequency band. For a comprehensive assessment, we start by proposing an analytical model for the antenna pattern that generates the beams. Then, we briefly study the *ergodic capacity* of the link, showing that, at high signal-to-noise ratio (SNR), the loss induced by rain tends to be constant, that is, independent of the SNR value and dependent only on the rain statistics. Ergodic capacity represents the maximum rate achievable with arbitrarily low error probability as long as we can code over a sufficiently large number of channel realizations. However, this is not the case of rain-faded satellite links, because codewords usually span only a few realizations of the channel, a situation commonly known as *slow fading*; as an example, [29, Eq. 11] suggests rain fading coherence times of hundreds of seconds. In this case, ergodic capacity represents the average rate at which we can transmit with perfect channel state information (CSI) at the transmitter.

For this reason, in this chapter we focus on the *outage capacity* of the link, which is the maximum transmission rate at which the outage probability does not exceed a given value ϵ. We derive high and low SNR approximations that provide insights on the effect of key system parameters – like the inter-user distance, the satellite beam radius, or the rain intensity. Analytical results are verified through extensive Monte Carlo simulations. Additionally, the derived expressions can be easily particularized for the single-user case, providing some novel insights; note that the single-user case is relevant for state-of-the-art multibeam systems employing partial frequency reuse [70] –different portions of the available spectrum are assigned to adjacent beams to reduce interference–, where each beam is processed separately without exploiting the residual interference.

3.1.1   Summary of this chapter

The following is an outline of the contributions of the chapter.

- We explicitly derive the probability density function (PDF) of the rain attenuation in *natural units* (3.8), d_i, which is the exponentiation of a log-normal random variable (log-

Table 3.1: Summary of results involving outage capacity

		Single user		Multiple users
C_ϵ	L	$\log_2\left(1 + \gamma \cdot e^{-e^{\sigma Q^{-1}(\epsilon)+\beta}}\right)$	$C_{\text{awgn}} \cdot e^{-e^{\sigma Q^{-1}(\epsilon)+\beta}}$	$\sqrt{V_{\text{L}}} \cdot Q^{-1}(1-\epsilon) + C_{\text{low}}$
	H		$C_{\text{awgn}} - e^{\sigma Q^{-1}(\epsilon)+\beta} \cdot \log_2 e$	$C_{\text{awgn}} - e^{Q^{-1}(\epsilon)\Omega+M_{\text{H}}}$

μ, σ are the location and scale parameters of the rain distribution, γ the SNR, $\beta = \mu - 2.1617$, C_{awgn} the capacity of an unfaded channel, C_{low} is the ergodic capacity at low SNR, $\Omega, M_{\text{H}}, V_{\text{L}}$ are parameters of the obtained approximations, defined in (3.29), (3.30), (3.34), respectively.

lognormal); then, after deriving an approximation of the Gaussian error function (3.47), we obtain an analytical, closed-form expression for any moment of the attenuation, $\mathbb{E}\left[d_i^k\right], k > 0$ (3.15).

- For the multibeam antenna pattern, an analytical model based on Wyner's 2D cellular arrangement is obtained (3.2). Its isolation parameter is determined by emulating a typical antenna's radiation pattern through the method of the Bessel functions (3.6).

- The ergodic capacity of the link is briefly studied, obtaining analytical expressions at low (3.20) and high (3.21) SNR. The obtained expressions suggest a mild influence of rain attenuation on the ergodic capacity.

- Regarding the outage capacity of the link, we derive approximations for the high and low SNR regimes (3.28), (3.33), which tightly fit simulation data[1].

- The outage capacity results can be particularized to the scalar, single-user channel. In such a case, a general expression for any SNR can be found (3.39), as well as insightful particular versions for the low (3.41) and high (3.42) SNR regimes.

- Table 3.1 summarizes the results obtained regarding outage capacity (L and H stand for low and high SNR, respectively).

3.2 System model

Let us consider a multi-user channel with K single-antenna terminals transmitting towards a single satellite equipped with the same number of antennas, so that the signal model reads as

$$\mathbf{y} = \sqrt{\gamma}\mathbf{H}\mathbf{s} + \mathbf{n}, \tag{3.1}$$

as in Section 2.5.

We shall remark that γ refers to a normalized power over noise constant, which does not take interference into account. Inter-user interference is present in matrix $\mathbf{G}$, and in the attenuation

---

[1]Although the resulting expression has been particularized for the proposed antenna model, any other model could be used instead.

experienced by the different users through $\mathbf{D}$. This will be made clearer through the following paragraphs.

3.2.1 Antenna pattern

As explained in Section 2.2, in many practical instances, the antenna pattern is obtained by means of specific software, which models real antennas and components and yields the numerical entries of matrix $\mathbf{G}$.

For analytical purposes, we will rather use a mathematical model that allows to write $\mathbf{G}$ as a function of key system parameters, like the inter-user distance or the beam radius. In short, we will resort to a well-known information theoretic cellular model, namely Wyner's 2D model [76], to describe the geometry of the system. Using this model, $\mathbf{G}$ will be a function of a, the *isolation parameter* between adjacent beams; to model this parameter, we will use the well-accepted Bessel function antenna model for tapered-aperture antennas, simplifying the general formula showed in Section 2.2.1.

Wyner model

Similarly to [6], we describe the geometry of the beam coverage using Wyner's 2D arrangement of cells. This arrangement is achieved by considering L lines, each of M cells (so that $K = LM$), placed one on top of another, where *each cell is affected by a single tier of interfering cells*; this renders the channel matrix to a Toeplitz-Block-Toeplitz (TBT) matrix of the form

$$\mathbf{G} = \mathbf{T}_{2D} = \begin{pmatrix} \mathbf{T}_{1D} & \mathbf{S} & 0 & \dots & 0 \\ \mathbf{S}^H & \mathbf{T}_{1D} & \mathbf{S} & \ddots & \vdots \\ 0 & \mathbf{S}^H & \mathbf{T}_{1D} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \mathbf{S} \\ 0 & \dots & 0 & \mathbf{S}^H & \mathbf{T}_{1D} \end{pmatrix}, \tag{3.2}$$

where

$$\mathbf{S}^T = \begin{pmatrix} a & a & 0 & \dots & 0 \\ 0 & a & a & \ddots & \vdots \\ & 0 & a & \ddots & 0 \\ \vdots & & \ddots & \ddots & a \\ 0 & \dots & & 0 & a \end{pmatrix}, \mathbf{T}_{1D} = \begin{pmatrix} 1 & a & 0 & \dots & 0 \\ a & 1 & a & \ddots & \vdots \\ 0 & a & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & a \\ 0 & \dots & 0 & a & 1 \end{pmatrix} \tag{3.3}$$

where a is the so-called isolation parameter, and with $\mathbf{S} \in \mathbb{R}^{M \times M}$, $\mathbf{T}_{1D} \in \mathbb{R}^{M \times M}$ and $\mathbf{T}_{2D} \in \mathbb{R}^{LM \times LM}$.

Remark: Assuming each cell to be affected by a single tier of interfering cells is, in fact, more realistic in multibeam satellite systems than in terrestrial scenarios, since in the former interbeam interference is solely controlled by the antenna design, and not by the terrain characteristics.

A drawback of the model above is that there is no closed-form expression for the eigenvalues of $\mathbf{T}_{2D}$. However, in the limit when $L, M \to \infty$, the eigenvalues of the TBT matrix of (3.2) are given by [77]:

$$\lambda_{(l-1)M+m}(\mathbf{T}_{2D}) = 1 + 2a\left(\cos\left(\frac{2\pi l}{L}\right) + \cos\left(\frac{2\pi m}{M}\right) + \cos\left(\frac{2\pi l}{L} + \frac{2\pi m}{M}\right)\right), \qquad (3.4)$$

and we can exploit this for large systems.

Modeling a

Using the Bessel function model in Section 2.2.1 with typical tapered-aperture antenna parameters [23, p. 184], [24], the channel gain from the i-th beam towards the j-th user is given by

$$g_{ij}(\theta_{ij}) = G_{\max}\left(\frac{J_1(u_{ij})}{2u_{ij}} + 36\frac{J_3(u_{ij})}{u_{ij}^3}\right)^2 \qquad (3.5)$$

as a function of the off-axis angle with respect to the beam's boresight, $\theta_{ij} = \arctan\left(d_{ij}/D\right)$, where D is the distance from the user to the satellite and d_{ij} is the distance between the center of the i-th and j-th cells. In the equation above, $u_{ij} = 2.07123 \sin\theta_{ij}/\sin\theta_{3\mathrm{dB}}$, J_1 and J_3 are the Bessel functions of the first kind, of order one and three respectively, $G_{\max}$ is the maximum axis gain of each antenna, and $\theta_{3\mathrm{dB}}$ is the angle associated to the beam's 3 dB radius, $\theta_{3\mathrm{dB}} = \arctan\left(R/D\right)$, with R the beam's radius.

For the sake of the analysis, the following approximations can be made: we will assume that the Earth curvature is negligible, that the slant ranges among all users are identical and equal to the GEO satellite elevation distance, $D = 36000\,\mathrm{Km}$, that all users are placed in the centers of the beams, and that each beam's radiation pattern is fixed: as the distance between users changes, so will the centers of the fixed radius beams (see Figure 3.1).

Under the aforementioned assumptions, and since the distance among users is much smaller than the satellite altitude, the relative distance d is translated into an angle simply via $\theta \approx \frac{d}{D}$, and we finally have (see Figure 3.2):

$$a = \sqrt{\frac{g(\theta)}{G_{\max}}} = \frac{J_1(u)}{2u} + 36\frac{J_3(u)}{u^3}, \quad u \approx 2.07123 \cdot d/R. \qquad (3.6)$$
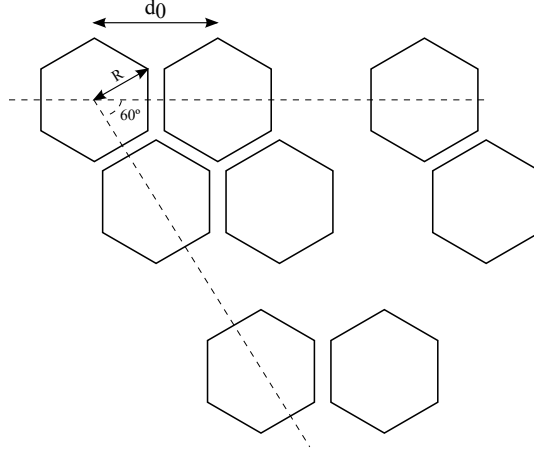
Figure 3.1: Example of fixed beam patterns with changing distance between users; d_0, a parameter that we will vary through the analysis, is the distance between the centers of adjacent beamspots.
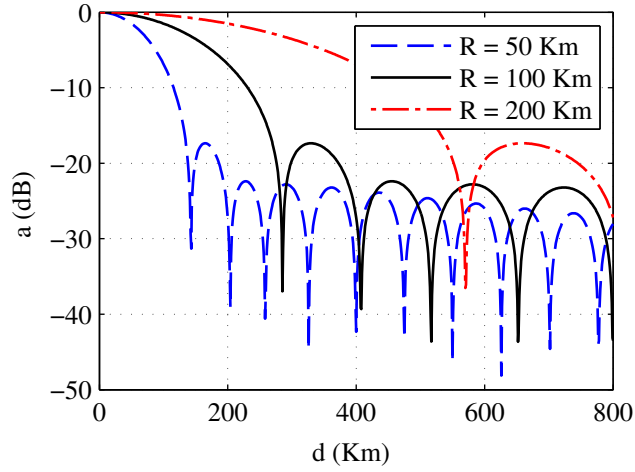
Figure 3.2: Evolution of the antenna pattern with respect to the beam center for different beam radius. Different link budgets have been used to enforce the same maximum gain.

3.2.2   Rain attenuation

The elements δ_i of the diagonal rain attenuation matrix $\mathbf{D} = \text{diag}(\boldsymbol{\delta})$ follow [26]

$$- 20 \log_{10} (\delta_i) \sim \mathcal{LN} (\mu_i, \sigma_i) \tag{3.7}$$

where μ_i and σ_i are the log-normal location and scale parameters, respectively, expressed in dB. Under this assumption, both δ and δ^2 follow a double log-normal distribution, as defined below.

Definition 1. *Let Y be a log-normally distributed random variable with location parameter μ and scale parameter σ, $Y \sim \mathcal{LN} (\mu, \sigma)$. Then, a random variable $X = A^{-\alpha Y}$ is said to be a double log-normal or log-lognormal random variable with scale parameter σ and location parameter $\beta \doteq \mu + \log \log A + \log \alpha$, $X \sim \mathcal{L}^2\mathcal{N} (\beta, \sigma)$.*

Lemma 1 (PDF and CDF of a $\mathcal{L}^2\mathcal{N}$ random variable). *The PDF of a double log-normal random variable is given by [78]*

$$f_X(x) = - \frac{1}{\sqrt{2\pi}\sigma x \log x} \cdot e^{-\frac{1}{2\sigma^2}(\log(-\log x) - \beta)^2} \tag{3.8}$$

with $0 < x < 1$, while the CDF is given by

$$F_X(x) = Q \left(\frac{\log (- \log x) - \beta}{\sigma} \right) \tag{3.9}$$

with $Q(x)$ representing the widely used Gaussian Q-function.

Proof. Consider the function $g(x) = A^{-\alpha x}$, and that we wish to obtain the PDF of the transformation $g(Y)$; the result follows immediately after considering that the inverse transformation is given by $-1/(\alpha \log A) \log x$, while its derivative equals $1/(\alpha x \log A)$. $\qquad\square$

From the definitions above, it is clear that δ corresponds to $\alpha = 1/20$ and δ^2 to $\alpha = 1/10$.

The moments of δ^2 are not available in closed-form, since this would require solving

$$\int_0^1 \frac{x^{k-1}}{\log x} \cdot e^{-\frac{1}{2\sigma^2}(\log(-\log x) - \beta)^2} \, \mathrm{d}x. \tag{3.10}$$

However, a closed-form expression can be obtained after resorting to a sigmoid approximation of the error function, as shown below.

Lemma 2. *Assume a tight approximation of the error function in the interval $[0, u/\sqrt{2})$, $u > 0$, given by $\text{erf}(x) \approx \sum_{i=1}^{N_c} a_i e^{-b_i x}$ and the sets of coefficients $\{a_i\}_{i=1}^{N_c}$, $\{b_i\}_{i=1}^{N_c}$. Then, the k-th moment of the random variable $X \sim \mathcal{L}^2\mathcal{N} (\beta, \sigma)$, $M_X(k)$, can be approximated as shown in (3.15), where $\varsigma_j \doteq \frac{b_j}{\sqrt{2}\sigma}$ and $\Gamma (a, x)$ is the upper incomplete Gamma function $\Gamma(c, x) = \int_x^\infty t^{c-1} e^{-t} \mathrm{d}t$ [79, Eq. 6.5.3].*

$$2M_X(k) \approx e^{-k \cdot e^{\beta - u\sigma}} - e^{-k \cdot e^{\beta + u\sigma}} + \sum_{j=1}^{L} a_j \left(ke^{\beta}\right)^{\varsigma_j}$$

$$\times \left(\Gamma\left(1 - \varsigma_j, ke^{\beta}\right) - \Gamma\left(1 - \varsigma_j, ke^{\beta + u\sigma}\right) + \left(ke^{\beta}\right)^{-2\varsigma_j} \Gamma\left(1 + \varsigma_j, ke^{\beta}\right) - \left(ke^{\beta}\right)^{-2\varsigma_j} \Gamma\left(1 + \varsigma_j, ke^{\beta - u\sigma}\right) \right)$$
(3.15)

Proof. see Appendix 3.A. □

3.2.3   Rain spatial correlation

Once characterized the marginal statistics of the rain coefficients d_i, the main interest is to evaluate their correlation and the corresponding impact on the return link capacity. Taking into account correlation, $\boldsymbol{\delta}$ verifies

$$-20 \log_{10}(\log \boldsymbol{\delta}) \sim \mathcal{N}\left(\boldsymbol{\mu}, \mathrm{diag}(\boldsymbol{\sigma})\mathbf{R}\,\mathrm{diag}(\boldsymbol{\sigma})\right)$$
(3.11)

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the location and scale parameter vectors, respectively, and $\mathbf{R} = \{r'_{ij}\}$ is the matrix of correlation coefficients among the so-called reduced Gaussian variables $u_i = \left(-20 \log_{10}(\log \delta_i) - \mu_i\right)/\sigma_i$.

There are many different models for r', as summarized in [80]. Here we will use the two-exponential model

$$r'_{ij} = p_b(d_{ij}) \doteq 0.94 e^{-\frac{d_{ij}}{30}} + 0.06 e^{-\left(\frac{d_{ij}}{500}\right)^2},$$
(3.12)

which was adopted by ITU in ITU-R P. 618-10 [81], although introducing any other model would be straightforward.

In what refers to the associated log-normal random variables, their correlation is given by [29]

$$\rho_{ij} = \frac{e^{\sigma_i \sigma_j p_b(d_{ij})} - 1}{\sqrt{e^{\sigma_i} - 1}\sqrt{e^{\sigma_j} - 1}}.$$
(3.13)

We will further impose $\mu_i \approx \mu, \sigma_i \approx \sigma \;\forall i$, so that the above equation simplifies to[2]

$$\rho_{ij} = \frac{e^{\sigma^2 p_b(d_{ij})} - 1}{e^{\sigma^2} - 1}.$$
(3.14)

Differently from the antenna pattern, for the rain correlation we will consider the influence of all the cells, and not only the adjacent ones. The distance between the centers of any i-th

---

[2]In reality, the values of μ and σ measured in relatively close areas experience some small differences. However, since we will focus on the case where some degree of correlation exists among the rain attenuation coefficients, it is sensible to assume that the marginal statistics will be very similar.

and j-th cells is given by

$$d_{ij} = 3d_0\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (x_j - x_i)(y_j - y_i)} \qquad (3.16)$$

with

$$x_i = \left\lfloor \frac{i}{M} \right\rfloor + 1, \quad y_i = i - \left\lfloor \frac{i}{M} \right\rfloor M \qquad (3.17)$$

and d_0 the distance between the centers of adjacent hexagons, as shown in Figure 3.1.

Let $\mathbf{P}$ be the matrix of correlation coefficients, $\mathbf{P}_{ij} = \rho_{ij}$. For notational convenience, we will write $\mathbf{P}$ as

$$\mathbf{P} = \frac{1}{e^{\sigma^2} - 1}(\mathbf{A} - 1) \qquad (3.18)$$

with $\mathbf{A}_{ij} \doteq e^{\sigma^2 p_b(d_{ij})}$.

3.3 Performance under correlated rain attenuation

We will start by briefly addressing the ergodic capacity of a multibeam satellite system; as we will see, rain correlation has no effect on this metric. We will then study the outage capacity of the channel, and assess the impact of the system's geometry and of the rain characteristics.

3.3.1 Ergodic capacity

For the channel under discussion, ergodic capacity is obtained by the well-known *logdet* formula $C_{\text{erg}} = \mathbb{E}_{\mathbf{D}}\left[\log_2 \det\left(\mathbf{I} + \gamma\mathbf{D}^2\mathbf{G}^H\mathbf{G}\right)\right]$. The expectation of the logarithm of the determinant is difficult to obtain, since it would require an analytical characterization of the (stochastic) eigenvalues of the product $\mathbf{D}^2\mathbf{G}^H\mathbf{G}$; instead, we will focus on the high and low SNR regimes.

Theorem 6. *The achievable sum rate can be approximated at low SNR by:*

$$C_{\text{low}} = \gamma \cdot M_\delta(2)\,\text{trace}\left(\mathbf{G}^H\mathbf{G}\right)\log_2 e \qquad (3.19)$$

which, for the antenna model proposed in Section 3.2.1, particularizes to

$$C_{\text{low}} = \gamma \cdot M_\delta(2)\log_2 e\left(2a^2(3LM - 2L - 2M + 1) + LM\right) \qquad (3.20)$$

with a^2 from (3.6) ; $M_\delta(2) = \mathbb{E}\left[\delta^2\right]$ can be obtained from (3.15).

On the other hand, at high SNR it would read as

$$C_{\text{high}} = C_{\text{awgn}} - \frac{\log_2 10}{10}K \cdot e^{\mu + \sigma^2/2} \qquad (3.21)$$

$$C_{\mathrm{awgn}} \approx \prod_{l=1}^{L} \prod_{m=1}^{M} \left(1 + \left(1 + 2a \left(\cos \left(\frac{2\pi l}{L} \right) + \cos \left(\frac{2\pi m}{M} \right) + \cos \left(2\pi \left(\frac{l}{L} + \frac{m}{M} \right) \right) \right) \right)^2 \right) \quad (3.22)$$

with C_{awgn} obtained either numerically from $\mathbf{G}$, or by (3.22) for large systems.

Proof. At low SNR, using $\log(1 + x) \approx x$, the achievable sum rate is

$$C_{\mathrm{low}} = \gamma \mathbb{E} \left[\mathrm{trace} \left(\mathbf{D}^2 \mathbf{G}^H \mathbf{G} \right) \right] \log_2 e = \gamma \mathbb{E} \left[\delta^2 \right] \mathrm{trace} \left(\mathbf{G}^H \mathbf{G} \right) \log_2 e \qquad (3.23)$$

where the last equality follows from assuming the same attenuation statistics in all the paths; the extension to different statistics is straightforward and can be found in [82].

In what refers to $\mathrm{trace} \left(\mathbf{G}^H \mathbf{G} \right)$, its value can be computed in closed form and results into (see Appendix 3.B for the proof)

$$\mathrm{trace} \left(\mathbf{G}^H \mathbf{G} \right) = 2a^2 \left(3LM - 2L - 2M + 1 \right) + LM, \qquad (3.24)$$

so that substituting the value of a^2 completes the proof.

$\square$

Remark: The high SNR approximation holds whenever $\mathbf{I} << \gamma \mathbf{D}^2 \mathbf{G}^H \mathbf{G}$, that is, when the system is not noise limited. For future satellite systems using strong frequency reuse, the system becomes strongly interference limited, and as a consequence high SNR analysis tend to be more relevant.

We have just shown how ergodic capacity can be expressed in a rain faded channel: at high SNR, it exhibits a constant loss with respect to the unfaded capacity, while at low SNR it amounts to a scaling of such unfaded capacity which depends on the rain statistics and antenna characteristics; similar problems had been tackled, among others, in [83], in that case by upper bounding capacity. However, we should question the importance of this metric for the channel under discussion.

It is well-known that, as the codeword length approaches infinity, the maximum rate at which reliable communication is possible approaches $\mathbb{E} \left[\log_2 \det \left(\mathbf{I} + \gamma \mathbf{D}^2 \mathbf{G}^H \mathbf{G} \right) \right]$ [38]. But this is not the case of rain-faded satellite links, because codewords usually span only a few realizations of the channel[3]; this is called *slow fading*. In this case, ergodic capacity represents only the average rate at which we can transmit if we have perfect CSI at the transmitter (and without assuming any power allocation over time).

Also, note that the multiuser case has an additional motivation for further analysis: from

---

[3]This could be fixed by introducing an arbitrarily large interleaver, but this would come at the price of an arbitrarily large delay, unaffordable in our scenario.
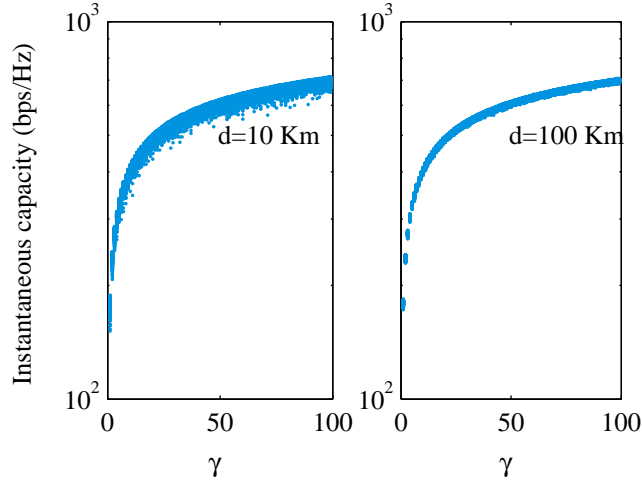
Figure 3.3: Realizations of $\log_2 \det \left(\mathbf{I} + \gamma \mathbf{D}^2 \mathbf{G}^H \mathbf{G} \right)$ for two levels of correlation of the entries in $\mathbf{D}$, and fixing $\mathbf{G}$.

the derived expressions (3.20) and (3.21), the correlation among the users rain attenuation has no effect on the ergodic capacity. As a consequence, this metric does not allow us to assess the induced system degradation. For these reasons, we will focus our analysis on the outage capacity of the link.

Before going further into the outage analysis, Figure 3.3 plots the evolution of the average and instantaneous capacity for different levels of correlation; the mean is always the same, but the dispersion of the dots is much higher for the case with higher correlation (left). In short, the average rate will not be much lower but the availability of the link will suffer the most. We will now analyze all these effects in more detail.

3.3.2 Outage capacity

As stated, correlation among the diagonal elements in $\mathbf{D}$ has no effect on ergodic capacity because it does not affect the expectation

$$C_{\text{erg}} = \sum_i \mathbb{E}_{\mathbf{D}} \left[\log_2(1 + \gamma \lambda_i \left\{ \mathbf{D}^2 \mathbf{G}^H \mathbf{G} \right\}) \right] \tag{3.25}$$

where λ_i denotes the i-th largest eigenvalue of a matrix. However, other metrics are greatly affected by the fluctuations of the random variable $\mathcal{I} = \sum_i \log_2(1 + \gamma \lambda_i \left\{ \mathbf{D}^2 \mathbf{G}^H \mathbf{G} \right\})$.

In a slow fading channel, for a given rate R, the probability that the instantaneous capacity is lower than R is given by the outage probability

$$p_{\text{out}}(R) \doteq \mathrm{P}\left[\mathcal{I} < R \right] \tag{3.26}$$

where $\mathcal{I}$ is the instantaneous capacity.

Also of highly practical significance is the ϵ-*outage capacity*, that is, the largest transmission rate at which the outage probability is less than ϵ [38, 84]. In other words: the maximum rate the channel will allow with probability $1 - \epsilon$.

In our case, from (3.26), we have that C_ϵ is the maximum value satisfying

$$\mathrm{P}\left[\log_2 \det\left(\mathbf{I} + \gamma\mathbf{D}^2\mathbf{G}^H\mathbf{G}\right) < C_\epsilon\right] = \epsilon. \tag{3.27}$$

Let us analyze this expression for the high and low SNR regimes.

Theorem 7. *At high SNR, the outage capacity can be approximated by*

$$C_\epsilon^{\mathrm{H}} = C_{\mathrm{awgn}}^{\mathrm{H}} - e^{Q^{-1}(\epsilon)\Omega + M_{\mathrm{H}}} \tag{3.28}$$

with

$$\Omega^2 = \log\left(e^{\sigma^2} + \frac{2}{K}S\right) - \log K, \tag{3.29}$$

$$M_{\mathrm{H}} = \mu + \frac{\sigma^2}{2} - \frac{1}{2}\log\left(e^{\sigma^2} + \frac{2}{K}S\right) + \frac{3}{2}\log K + \log\left(\frac{\log_2 10}{10}\right), \tag{3.30}$$

$$S \doteq \sum_{\substack{i,j \\ i<j}} \mathbf{A}_{ij}. \tag{3.31}$$

Proof. See Appendix 3.C. $\qquad\qquad\square$

What (3.28) tells us is that, at high SNR, rain induces a constant loss also in terms of outage capacity. This loss depends on S in (3.31), which is a function of the inter-user distance and of the rain geometry model used, but it does not depend on the SNR γ. In particular, the loss grows with S, which is a sum of elements which increase exponentially with the correlation value. As a consequence, decreasing the distance (increasing correlation) makes the loss grow sharply, as we will see in Section 3.4. However, decreasing the distance would also affect $\mathbf{G}$. The overall effect thus depends also on the particular beam pattern under use. Note that these same expressions would hold for any other model, as long as we are able to compute the summation that leads to S.

Following the model we proposed, C_{awgn} also depends on $\mathbf{G}$. Since in (3.28) the contributions of the antenna pattern and of the rain are decoupled, it would be easy to extend the analysis to any other antenna model. We will exploit this to illustrate the influence of rain alone in Section 3.4.

A final remark concerns the case with only one user: noting that $K = 1$ and $S = 0$, we can easily particularize (3.28) and obtain

$$C_\epsilon^{\mathrm{H}} = C_{\mathrm{awgn}}^{\mathrm{H}} - e^{Q^{-1}(\epsilon)\sigma + \mu + \log\left(\frac{\log_2 10}{10}\right)} \tag{3.32}$$

which, interestingly, is the expression of the outage capacity for a single-user channel impaired by rain attenuation at high SNR, as we will show in the next section.

Theorem 8. *At low SNR, and under moderate correlation, the outage capacity can be approximated by*

$$C_\epsilon^{\mathrm{L}} \approx \sqrt{V_{\mathrm{L}}} \cdot Q^{-1}(1 - \epsilon) + C_{\mathrm{low}} \tag{3.33}$$

where C_{low} is given by (3.19), and with

$$V_{\mathrm{L}} = (\log_2 e)^2 \gamma^2 \left(M_\delta(4) \sum_{i=1}^{K} ||\mathbf{g}_k||^4 + 2 \sum_{\substack{i,j \\ i<j}} \varphi_{i,j} \cdot ||\mathbf{g}_i||^2 ||\mathbf{g}_j||^2 \right) - C_{\mathrm{low}}^2, \tag{3.34}$$

$$\varphi_{i,j} \triangleq \mathbb{E}\left[\delta_i \delta_j\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 10^{-\frac{1}{10}\left(e^{\sigma x + \mu} + e^{\sigma y + \mu}\right)} f(x,y) \, \mathrm{d}x \mathrm{d}y, \tag{3.35}$$

and $f(x,y)$ the PDF of a bivariate Gaussian distribution,

$$f(x,y) = \frac{1}{2\pi\sqrt{1 - (r'_{ij})^2}} \cdot e^{-\frac{1}{2(1-(r'_{i,j})^2)}(x^2 + y^2 - 2r'_{i,j}xy)}. \tag{3.36}$$

Proof. See Appendix 3.D. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The value $Q^{-1}(1 - \epsilon)$ is negative for $\epsilon < 0.5$. Thus, increasing V_{L} decreases the outage capacity.

Remark: Differently from the high SNR case, here the effect of the distance cannot be easily separated in antenna effects and correlation effects. Note that decreasing the distance will increase the values $\varphi_{i,j} = \mathbb{E}\left[\delta_i \delta_j\right]$, and will thus tend to decrease the outage capacity; however, decreasing the distance at the same time increases the values of $||\mathbf{g}_i||$ and of C_{low}. Summarizing: Theorem 8 is valid for any antenna pattern, but the shape of C_ϵ^{L} as a function of the distance will depend on its particular values.

Remark: As explained in the appendix, the result above relies on approximating trace $\left(\mathbf{D}^2 \mathbf{G}^H \mathbf{G}\right)$ by a Gaussian random variable, which proves to be accurate whenever the correlation among the elements in $\mathbf{D}$ is not very high. For this latter case, the trivial approximation $\mathbf{D} \approx \delta\mathbf{I}$, with δ a single log-lognormal random variable, should be used instead.

3.3.3 The single-user case

So far, we have tackled the multi-antenna, multiuser case, showing that the outage induced by rain is constant at high SNR, and that it is larger when the rain correlation is stronger. The obtained results are relevant for systems employing full frequency reuse and joint multiuser detection.

In this section, we will explore the simplified case with only one user and one receiving antenna. Apart from yielding very illustrative results, this scenario has great operational significance, as it corresponds to the multibeam scenarios in which partial frequency reuse is employed; in such cases, adjacent beams are assigned different frequencies and it is customary to operate the link without exploiting the residual interference.

At time instant k, the signal model would be

$$y_k = \sqrt{\gamma} \cdot h_k s_k + n_k \tag{3.37}$$

where y_k and s_k are the received and transmitted symbols, respectively, h_k is the channel coefficient, n_k is a complex standard normal noise sample, $n_k \sim \mathcal{CN}(0,1)$, and γ is a variable that denotes the SNR taking into account all the deterministic coefficients of the link budget, including the noise power, the path losses, and any other attenuation in the transmission chain.

Solving (3.26) for $p_{\text{out}} = \epsilon$ yields

$$C_\epsilon = \log_2 \left(1 + F^{-1}\left(\epsilon\right) \cdot \gamma \right) \tag{3.38}$$

where F denotes the *cumulative distribution function* of $|h|^2$, so that substituting (3.9) in (3.38) we obtain

$$C_\epsilon = \log_2 \left(1 + \gamma \cdot e^{-e^{\sigma Q^{-1}(\epsilon)+\beta}} \right). \tag{3.39}$$

An immediate conclusion from (3.38) is that, to obtain the same rate as in an AWGN channel –but with an outage probability of ϵ instead – we will need an extra power of

$$\begin{aligned} \Delta_P &\doteq -10 \log_{10} F^{-1}(\epsilon) \\ &= e^{\sigma Q^{-1}(\epsilon)+\beta} \cdot 10 \log_{10} e \text{ dB.} \end{aligned} \tag{3.40}$$

We can see that an increase in the location parameter μ, present in β, exponentially increases the power loss term. The same behavior holds for the scale parameter σ, but this is *amplified* by the value of $Q^{-1}(\epsilon)$, which grows larger as ϵ decreases; this can be seen on Figure 3.4.

Even though the extra power margin needed is the same regardless of the power regime, the effect of fading on the outage capacity does change with the SNR, as we will see in the following.

Theorem 9. *The outage capacity at low SNR can be approximated by*

$$C_\epsilon^{\text{L}} \approx C_{\text{awgn}} \cdot e^{-e^{\sigma Q^{-1}(\epsilon)+\beta}}, \tag{3.41}$$

while at high SNR it reads as

$$C_\epsilon^{\text{H}} \approx C_{\text{awgn}} - e^{\sigma Q^{-1}(\epsilon)+\beta} \cdot \log_2 e. \tag{3.42}$$

Proof. Both identities can be proven by applying $\log_2(1+x) \approx x/\log 2$ when $0 < x << 1$ (low
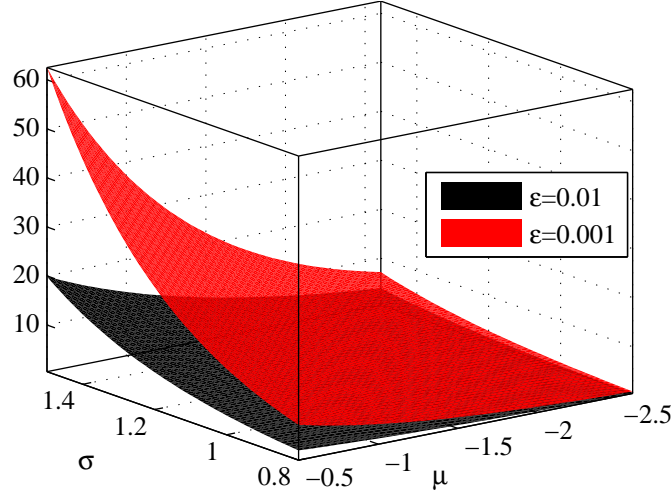
Figure 3.4: Evolution of Δ_P as a function of σ and μ for different values of ϵ.

SNR) and $\log_2(1 + x) \approx \log_2(x)$ (high SNR). $\qquad\square$

From the expressions above, it is easy to check that (3.42) is the same as (3.32) by substituting the expression of β; note that this equality happens even though in the multiuser case we had used an additional approximation, assuming that a sum of log-normally distributed random variables can be fit by a single log-normal random variable. Also, (3.41) tells us that the rain effect is quite different at low SNR: it scales capacity by a double exponential of $\sigma Q^{-1}(\epsilon) + \beta$.

3.4   Numerical results

In this section, we report some numerical results illustrating the behavior of a correlated rain-faded satellite return link. Simulations have been carried out with $\mu = -1.013$, $\sigma = 1.076$, as obtained in [28] for the city of Aarhus; $R = 100\,\mathrm{Km}$; and $L = 10$, $M = 10$, so that $K = 100$.

Figure 3.5 shows the evolution of capacity as a function of the inter-user distance for different values of SNR (γ), in a setup with no rain attenuation; the aim of this figure is to depict only the influence of the deterministic antenna pattern, which resembles that shown in [6].

Now we wish to assess the outage capacity in the presence of spatially correlated rain attenuation; to start with, Figure 3.6 depicts the outage capacity as a function of distance for the high and low SNR cases. Judging from Figures 3.6 and 3.5, it would seem that outage capacity as a function of inter-user distance is affected mostly by changes in the antenna pattern, which lead to changes in C_{awgn}^H. This, however, does not have to be true for every antenna pattern: other patterns could exist for which C_{awgn}^H had a value comparable to $-e^{\Omega Q^{-1}(1-\epsilon) + M_{\mathrm{H}}}$.

Figure 3.7 shows the CDF of the instantaneous capacity at high and low SNR for different distance values. We can see the effect of inter-user distance in the way the curves are shifted, and also that the derived analytical approximations tightly fit the Monte Carlo simulation.
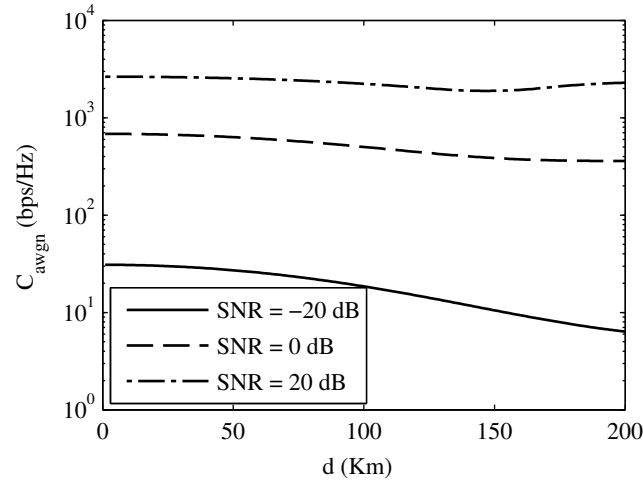
Figure 3.5: Capacity as a function of distance for the unfaded channel.
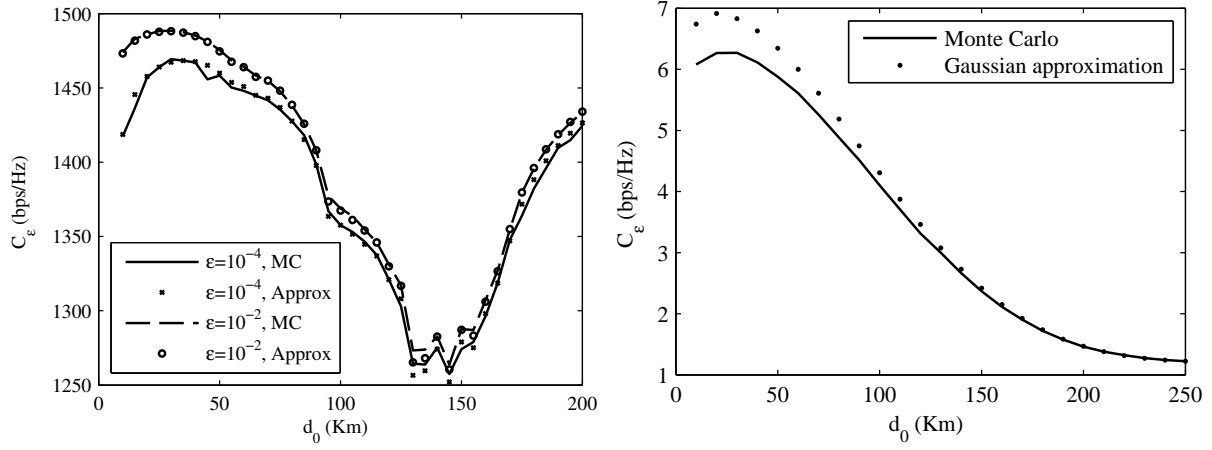
Figure 3.6: C_ϵ as a function of d_0 at high SNR (left, 45 dB) and low SNR (right, -20 dB).
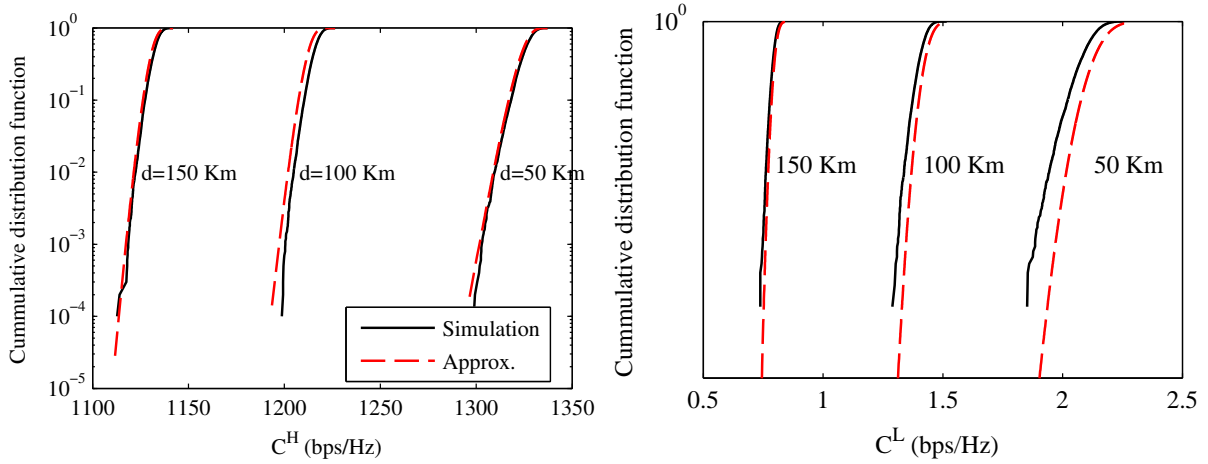
Figure 3.7: CDF of the instantaneous capacity at high (left, 40 dB) and low (right, -20 dB) SNR for different values of d.
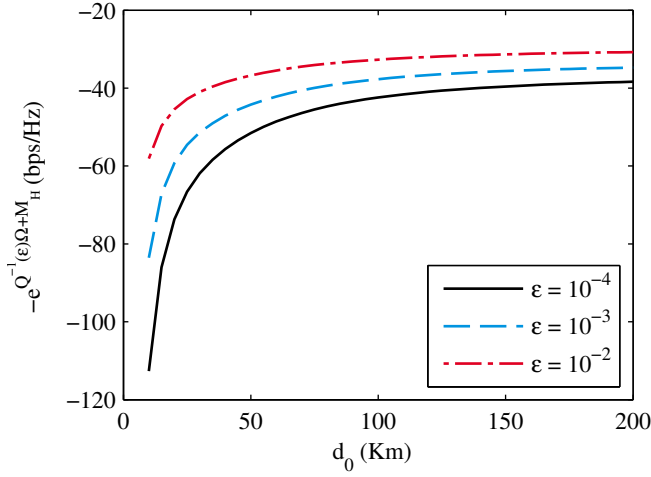
Figure 3.8: Outage capacity loss induced by rain as a function of d_0 for different values of ϵ.
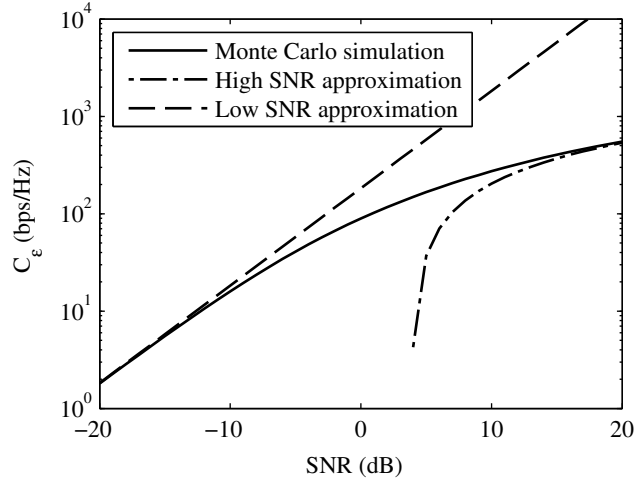
Figure 3.9: Outage capacity versus SNR, $\epsilon = 10^{-3}$.

To illustrate the influence of rain alone at high SNR, Figure 3.8 shows the evolution of the term $-e^{Q^{-1}(\epsilon)\Omega + M_{\mathrm{H}}}$, which is the ϵ-capacity loss induced by rain attenuation, for different values of ϵ; we can see that, for example, with $\epsilon = 10^{-4}$ and $d = 100\,\mathrm{Km}$, the total loss equals about $42\,\mathrm{bps/Hz}$, which roughly means $0.42\,\mathrm{bps/Hz}$ on average per user. Note that these losses are independent of the antenna pattern.

We can also get some insights on the behavior of outage capacity for the whole SNR range. Figure 3.9 shows its behavior for $\epsilon = 10^{-3}$, illustrating also the ranges of validity of the high and low SNR approximations. To this end, recall that the high SNR range is probably the most relevant, since it corresponds also to an interference limited case.

Before reporting the single user results, we will briefly analyze the effect of correlation over the rate of specific users, rather than on the overall sum rate. For this purpose, Figure 3.10 shows the outage probability of the minimum and maximum rate –that is, of the rate of the user
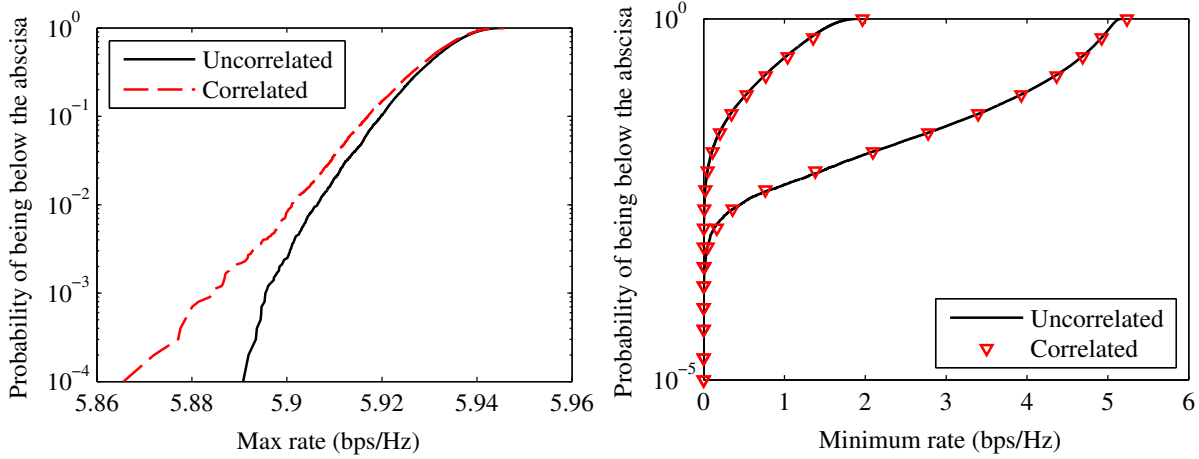
Figure 3.10: Outage probability of the maximum rate with SNR = 10 dB (left), and of the minimum rate on the right; for the latter, upper lines show SNR = 40 dB, lower lines SNR = 10 dB.

in the worst and best conditions, respectively– obtained after simulating the successive decoding of the users. For this simulation, we have selected in each iteration the user j with the largest value of $||\mathbf{h}_j||^2$, obtained its rate, and then removed the j-th column from the channel matrix. Correlation can be seen to have almost no effect on the minimum rate, but rather a significant impact on the maximum rate.

Finally, we present some results regarding the single-user scenario; we will rather depict $C_\epsilon/C_{\mathrm{awgn}}$. Figure 3.11 shows the derived approximations together with the original curve. We can see that, with the same power as in clear sky conditions, ensuring an availability of 99.999 % would imply reducing the rate down to a 10 % at low SNR, and even as low as 30 % for moderate SNR values like 10 dB.

3.5   Conclusions

We have studied the effect of spatially correlated rain attenuation on a multibeam satellite return link from the point of view of outage capacity.

Results have shown that inter-user distance affects both the antenna pattern and the correlation among the beams. Focusing only on the latter, we have seen that correlation induces noticeable losses on the outage capacity of the system. For example, with $\epsilon = 10^{-4}$, an inter-user distance of 100 Km and a beam radius of 100 Km, the total loss equals about 42 bps/Hz, which roughly means 0.42 bps/Hz in average per user.

For the particular case of a single user, single antenna link, results have shown that ensuring an outage probability of 10^{-3} requires an extra power offset of about 15 dB for common rain profiles. In terms of outage capacity, this means reaching only 10 % of the unfaded capacity at low SNR and 65 % at high SNR if we do not increase the power margin.
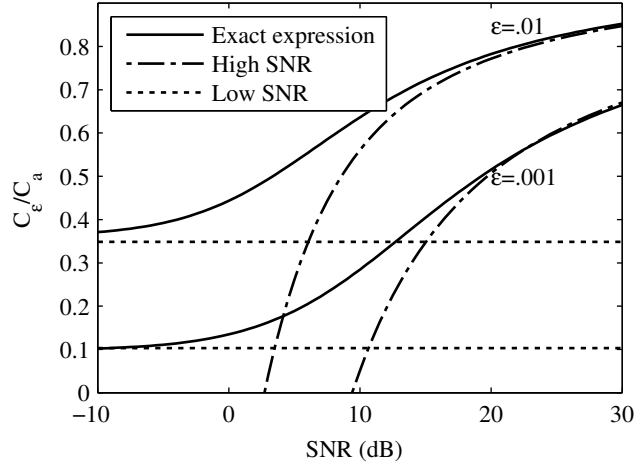
Figure 3.11: Accuracy of the high and low SNR approximations for $\epsilon = 0.01$ and $\epsilon = 0.001$. $\mu = -2.6$ dB and $\sigma = 1.3$ dB.

The contents in this chapter were done in collaboration with Dimitrios Christopoulos, Dr. Symeon Chatzinotas, Prof. Carlos Mosquera and Prof. Björn Ottersten. Parts of this material were published at the ASMS & SPSC 2012 conference [82], and in *IEEE Transactions on Wireless Communications* [27].

Appendix 3.A   Approximation of the moments of the log-lognormal distribution

Here, we derive a semi-analytical formula for the moments of X, which needs of a tight approximation of the error function $\mathrm{erf}(x)$ in the positive interval; once the approximation is available, the expression of the moments follows in closed form.

Before starting, let us state an important property of the log-lognormal distribution: if $X \sim \mathcal{L}^2\mathcal{N}\left(\beta, \sigma\right)$, then for any $k > 0$ it holds that

$$X^k \sim \mathcal{L}^2\mathcal{N}\left(\beta + \log k, \sigma\right). \tag{3.43}$$

This property will be very useful for the computation of the moments $\mathbb{E}\left[X^k\right]$.

To start with, let us reformulate the problem in a more convenient way.

Lemma 3. *Let $M_X(k)$ be the k-th moment of the random variable X. If X follows a log-lognormal distribution, then its computation can be rewritten as*

$$M_X(k) = \frac{1}{2} - \frac{1}{2}\int_0^\infty \mathrm{erf}\left(\frac{\beta + \log k - \log z}{\sqrt{2}\sigma}\right)e^{-z}\,\mathrm{d}z. \tag{3.44}$$

Proof. Recall that the expectation of a non-negative random variable can be written as [85, Eq.

Table 3.2: Coefficients of the approximation (3.47) of $\text{erf}(x)$ from 0 to $5/\sqrt{2}$

N_c	Coefficients
2	$a_1 = -1.295$, $a_2 = 1.24$, $b_1 = 1.346$, $b_2 = 0.06363$
4	$a_1 = 0.0027151$, $a_2 = -1.026$, $a_3 = 2$, $a_4 = -1.02$
	$b_1 = -1.20$, $b_2 = 0.9814$, $b_3 = 0.2306$, $b_4 = 0.9814$

(5-53)]

$$\mathbb{E}\left[X\right] = \int_0^\infty \left(1 - F_X(x)\right) \, \mathrm{d}x. \tag{3.45}$$

Now, noting that $Q(x) = 1/2 - 1/2 \cdot \text{erf}\left(x/\sqrt{2}\right)$ and $\text{erf}(-x) = -\text{erf}(x)$, we arrive at

$$
\begin{aligned}
\mathbb{E}\left[X\right] &= \int_0^1 \left(1 - Q\left(\frac{\log\left(-\log x\right) - \beta}{\sigma}\right)\right) \, \mathrm{d}x \\
&= \frac{1}{2} - \frac{1}{2}\int_0^1 \text{erf}\left(\frac{\beta - \log\left(-\log x\right)}{\sqrt{2}\sigma}\right) \, \mathrm{d}x.
\end{aligned} \tag{3.46}
$$

The proof concludes by using (3.43) and applying the change of variables $\log x = -z$. $\qquad\square$

Unfortunately, the integral in (3.44) still appears to be intractable. At this point, we will look for a good substitute for $\text{erf}(x)$ in an interval of the form $[0, r)$ that allows to compute the integral in closed form and, at the same time, offers good accuracy; the idea of focusing in a reduced interval is precisely to improve the accuracy by dealing with the asymptote $\text{erf}(x) = 1$ separately. Well-known accurate existing approximations of the Q function, or of the error functions, involve exponentials of quadratic argument or polynomials (see [86] and references therein). Inspired in part by [87], we developed an approximation given by a sum of N_c exponentials of x:

$$\text{erf}(x) \approx \sum_{i=1}^{N_c} a_i e^{-b_i x} \quad 0 \leq x < r \tag{3.47}$$

for $N_c > 1$.

Recall that looking for an approximation in a reduced interval $[0, r)$ aims at improving the accuracy in the curvy parts of the function; the almost constant values of $\text{erf}(x)$ when x grows large could be handled separately by setting r to a small value, as we will show. Table 3.2 shows the optimized coefficients, which were obtained by applying non-linear mean-squared optimization in Matlab$^{\copyright}$, for different values of N_c and $r = 5/\sqrt{2}$.

Once an approximation of the error function of the form of (3.47) is available, the moments of X can be computed in closed form: Let us focus on the case $k = 1$, since any other case would follow by a simple change of variables. We start by splitting the definite integral into two parts,

so that the argument of $\operatorname{erf}(x)$ is always positive, thus resulting into

$$
\begin{aligned}
2M_X(1) = 1 &- \int_0^{e^\beta} \operatorname{erf}\left(\frac{\beta - \log z}{\sqrt{2}\sigma}\right) e^{-z} \, \mathrm{d}z \\
&+ \int_{e^\beta}^\infty \operatorname{erf}\left(-\frac{\beta - \log z}{\sqrt{2}\sigma}\right) e^{-z} \, \mathrm{d}z \\
&= 1 - I_1 + I_2.
\end{aligned}
\tag{3.48}
$$

We further split each integral to separate the part in which $\operatorname{erf}(x) \approx 1$:

$$
\begin{aligned}
I_1 &\doteq \int_0^{e^\beta} \operatorname{erf}\left(\frac{\beta - \log z}{\sqrt{2}\sigma}\right) e^{-z} \, \mathrm{d}z \\
&= 1 - e^{-e^{\beta - u\sigma}} + \int_{e^{\beta - u\sigma}}^{e^\beta} \operatorname{erf}\left(\frac{\beta - \log z}{\sqrt{2}\sigma}\right) e^{-z} \, \mathrm{d}z
\end{aligned}
\tag{3.49}
$$

The same operation applied on the second integral yields

$$
\begin{aligned}
I_2 &\doteq \int_{e^\beta}^\infty \operatorname{erf}\left(-\frac{\beta - \log z}{\sqrt{2}\sigma}\right) e^{-z} \, \mathrm{d}z \\
&= -e^{-e^{\beta + u\sigma}} + \int_{e^\beta}^{e^{\beta + u\sigma}} \operatorname{erf}\left(-\frac{\beta - \log z}{\sqrt{2}\sigma}\right) e^{-z} \, \mathrm{d}z.
\end{aligned}
\tag{3.50}
$$

The two unsolved integrals can be worked out in the same way. Picking the first one, we can apply the exponential sum approximation (3.47) to obtain

$$
\begin{aligned}
&\int_{e^{\beta - u\sigma}}^{e^\beta} \operatorname{erf}\left(\frac{\beta - \log z}{\sqrt{2}\sigma}\right) e^{-z} \, \mathrm{d}z \\
&\approx \sum_{j=1}^{N_c} a_j \int_{e^{\beta - u\sigma}}^{e^\beta} e^{-b_j \frac{\beta - \log z}{\sqrt{2}\sigma} - z} \, \mathrm{d}z \\
&= \sum_{j=1}^{N_c} a_j e^{-\varsigma_j \beta} \int_{e^{\beta - u\sigma}}^{e^\beta} z^{\varsigma_j} e^{-z} \, \mathrm{d}z \\
&= \sum_{j=1}^{N_c} a_j e^{-\varsigma_j \beta} \left(-\Gamma\left(1 + \varsigma_j, e^\beta\right) + \Gamma\left(1 + \varsigma_j, e^{\beta - u\sigma}\right)\right)
\end{aligned}
\tag{3.51}
$$

where the last equality follows from the definition of the upper incomplete Gamma function.

Applying the same procedure to the integral in I_2 and substituting in (3.48) concludes the proof. $\qquad\square$

Appendix 3.B   Computation of $\text{trace}\left(\mathbf{G}^H\mathbf{G}\right)$

It is straightforward to check that the block-diagonal of $\mathbf{G}^H\mathbf{G}$ is formed by $L-2$ matrices $\mathbf{S}^H\mathbf{S} + \mathbf{T}^2 + \mathbf{S}\mathbf{S}^H$ and 2 matrices $\mathbf{S}^H\mathbf{S} + \mathbf{T}^2$, so that

$$\text{trace}\left(\mathbf{G}^H\mathbf{G}\right) = L\,\text{trace}\left(\mathbf{T}^2\right) + 2(L-1)\,\text{trace}\left(\mathbf{S}^H\mathbf{S}\right) \tag{3.52}$$

since $\text{trace}(\mathbf{S}^H\mathbf{S}) = \text{trace}(\mathbf{S}\mathbf{S}^H)$. The traces involved represent the Fröbenius norm of matrices $\mathbf{T}$ and $\mathbf{S}$, respectively. In other words, they represent the sum of the power of their columns. From this premise, they can be readily found to be

$$\text{trace}\left(\mathbf{S}^H\mathbf{S}\right) = a^2 + (M-1)2a^2 = a^2(2M-1) \tag{3.53}$$

and

$$\begin{aligned}
\text{trace}\left(\mathbf{T}^2\right) &= 2(1+a^2) + (M-2)(1+2a^2) \\
&= 2a^2(M-1) + M.
\end{aligned} \tag{3.54}$$

The proof finishes immediately by combining the results above, yielding

$$\text{trace}\left(\mathbf{G}^H\mathbf{G}\right) = 2a^2(3LM - 2L - 2M + 1) + LM. \tag{3.55}$$

$\square$

Appendix 3.C   Proof of Theorem 7

The usual approximation results into

$$\begin{aligned}
C &\approx \log_2 \det\left(\gamma \mathbf{D}^2 \mathbf{G}^H \mathbf{G}\right) \\
&= \log_2 \det\left(\gamma \mathbf{G}^H \mathbf{G}\right) - \left(-\log_2 \det \mathbf{D}^2\right) \\
&= C_{\text{awgn}}^{\text{H}} - \Delta_c
\end{aligned} \tag{3.56}$$

where we have defined $\Delta_c \doteq -\log_2 \det \mathbf{D}^2$, the loss in spectral efficiency induced by rain attenuation

$$\begin{aligned}
\Delta_c &= -\sum_{i=1}^{K} \log_2 \delta_i^2 \\
&= \frac{\log_2 10}{10} \sum_{i=1}^{K} \xi_i \qquad \xi_i \sim \mathcal{LN}\left(\mu, \sigma\right).
\end{aligned} \tag{3.57}$$

From (3.56) we have that the outage probability for a certain *overall rate R* is given by

$$p_{\text{out}}(R) = \text{P}\left[C_{\text{awgn}}^{\text{H}} - \Delta_c < C_\epsilon\right] \tag{3.58}$$

so that, solving $p_{\text{out}}(C_\epsilon^{\text{H}}) = \epsilon$ we obtain

$$C_\epsilon^{\text{H}} = C_{\text{awgn}} - F_{\Delta_c}^{-1}(1 - \epsilon) \tag{3.59}$$

where F_{Δ_c} is the cumulative distribution function (CDF) of Δ_c.

Let us see how can we obtain F_{Δ_c}. The sum of (correlated or uncorrelated) log-normal random variables has been extensively studied in the literature; a summary of the most relevant alternatives can be found in [88]. Here, and for simplicity, we will make use of Fenton-Wilkinson's approximation, which states that a sum of log-normal variables can be approximated by another log-normal variable by matching the first and second order moments.

Let us derive the expressions of these two moments, expressing them as a function of the correlation coefficients. The first moment, $\mu_1 \doteq \mathbb{E}\left[\Delta_c\right]$ trivially reads as

$$\mu_1 = \frac{\log_2 10}{10}\sum_{i=1}^{K}\mathbb{E}\left[\xi_i\right] = \frac{\log_2 10}{10}\cdot Ke^{\mu + \frac{1}{2}\sigma^2} \tag{3.60}$$

where we have used the fact that, if $x \sim \mathcal{LN}(\mu, \sigma)$, then $\mathbb{E}\left[x\right] = e^{\mu + 1/2\sigma^2}$.

The second order moment is more involved; using the binomial expansion

$$\left(\sum_{i=1}^{N}x_j\right)^2 = \sum_{i=1}^{N}x_i^2 + 2\sum_{\substack{i,j\\i<j}}^{N}x_ix_j \tag{3.61}$$

we arrive at

$$\begin{aligned}
\mu_2 &= \frac{(\log_2 10)^2}{100}\mathbb{E}\left[\sum_{i=1}^{K}\xi_i^2 + 2\sum_{\substack{i,j\\i<j}}^{K}\xi_i\xi_j\right]\\
&= \frac{(\log_2 10)^2}{100}\left(K\mathbb{E}\left[\xi^2\right] + 2\sum_{\substack{i,j\\i<j}}^{K}\left(\text{var}\left(\xi\right)\rho_{ij} + \mathbb{E}\left[\xi^2\right]\right)\right)\\
&= \frac{(\log_2 10)^2}{100}\left(K\mathbb{E}\left[\xi^2\right] + 2\,\text{var}\left(\xi\right)\sum_{\substack{i,j\\i<j}}^{K}\rho_{ij} + 2\frac{K(K-1)}{2}\mathbb{E}\left[\xi\right]^2\right).
\end{aligned} \tag{3.62}$$

were we have used the fact that $\sum_{\substack{i,j\\i<j}}^{K}1 = K(K-1)/2$. We further use the following set of identities: $\mathbb{E}\left[\xi^2\right] = e^{2\mu + 2\sigma^2}$, $\mathbb{E}\left[\xi\right]^2 = e^{2\mu + \sigma^2}$, $\text{var}\left(\xi\right) = \left(e^{\sigma^2} - 1\right)e^{2\mu + \sigma^2}$; plugging these

expressions into (3.62), we get

$$
\mu_2 = \frac{(\log_2 10)^2}{100} K e^{2\mu+\sigma^2} \left(e^{\sigma^2} + K - 1 + 2\frac{e^{\sigma^2}-1}{K} \sum_{\substack{i,j \\ i<j}}^{K} \rho_{ij} \right). \tag{3.63}
$$

Note that, as the correlation among the fading variables increases, the value of μ_2 also increases, thus resulting into an increase in variance.

Computing the pending summation in (3.63), we obtain

$$
\begin{aligned}
\sum_{\substack{i,j \\ i<j}} \rho_{ij} &= \sum_{\substack{i,j \\ i<j}} \mathbf{P}_{ij} \\
&= \frac{1}{e^{\sigma^2}-1} \sum_{\substack{i,j \\ i<j}} (\mathbf{A}_{ij}-1) \\
&= \frac{1}{e^{\sigma^2}-1} \left(S - \frac{K(K-1)}{2} \right)
\end{aligned} \tag{3.64}
$$

where, for notational convenience, we have defined $S \doteq \sum_{\substack{i,j \\ i<j}} \mathbf{A}_{ij}$. which in this case would be given by

$$
S = \sum_{\substack{i,j \\ i<j}} e^{\sigma^2 p_b(d_{ij})} \tag{3.65}
$$

Finally, plugging (3.64) into (3.63) we arrive at

$$
\mu_2 = \frac{(\log_2 10)^2}{100} K e^{2\mu+\sigma^2} \left(e^{\sigma^2} + \frac{2}{K} S \right) \tag{3.66}
$$

Now that the moments are available, the sum can be approximated by a log-normal random variable with location parameter M_{H} and scale parameter Ω given by

$$
\begin{aligned}
\Omega^2 &= \log \left(\frac{\mu_2}{\mu_1^2} \right) \\
&= \log \left(e^{\sigma^2} + \frac{2}{K} S \right) - \log K
\end{aligned} \tag{3.67}
$$

$$
\begin{aligned}
M_{\mathrm{H}} &= \log \mu_1 - \frac{1}{2}\Omega^2 \\
&= \mu + \frac{\sigma^2}{2} - \frac{1}{2} \log \left(e^{\sigma^2} + \frac{2}{K} S \right) \\
&\quad + \frac{3}{2} \log K + \log \left(\frac{\log_2 10}{10} \right)
\end{aligned} \tag{3.68}
$$

Finally, using $-Q^{-1}(x) = Q^{-1}(1-x)$, the outage capacity reads as

$$C_\epsilon^{\mathrm{H}} = C_{\mathrm{awgn}}^{\mathrm{H}} - e^{Q^{-1}(\epsilon)\Omega + M_{\mathrm{H}}} \tag{3.69}$$

with $C_{\mathrm{awgn}}^{\mathrm{H}}$ given by (3.22) when K is large.

Appendix 3.D Proof of Theorem 8

The approximation $\log_2(1+x) \approx x/\log 2$ applied to (3.27) yields $\mathrm{P}\left[\gamma \operatorname{trace}\left(\mathbf{D}^2\mathbf{G}^H\mathbf{G}\right)\log_2 e < C_\epsilon^{\mathrm{L}}\right] = \epsilon$. Let us define the random variable $\mathcal{I}_{\mathrm{L}} \triangleq \gamma \operatorname{trace}\left(\mathbf{D}^2\mathbf{G}^H\mathbf{G}\right)\log_2 e$, which is a sum of correlated random variables; if we approximate it with a Gaussian random variable, then the outage capacity reads as $C_\epsilon^{\mathrm{L}} = \sqrt{V_{\mathrm{L}}}Q^{-1}(1-\epsilon) + M_{\mathrm{L}}$, where M_{L} and V_{L} are the mean and variance of $\mathcal{I}_{\mathrm{L}}$, respectively.

The first value is easy to obtain, since $\mathbb{E}\left[\mathcal{I}_{\mathrm{L}}\right] = C_{\mathrm{low}}$ in (3.20) by definition. The second one is more involved; using the identity $\operatorname{var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, the binomial expansion (3.61), and the equality $\operatorname{trace}(\mathbf{D}^2\mathbf{G}^H\mathbf{G}) = \sum_{i=1}^K \delta_i^2 ||\mathbf{g}_i||^2$, we can write

$$\operatorname{var}[\mathcal{I}_{\mathrm{L}}] = (\log_2 e)^2 \gamma^2 \left(\mathbb{E}[\delta^4] \sum_{i=1}^K ||\mathbf{g}_k||^4 + 2 \sum_{\substack{i,j \\ i<j}} \mathbb{E}[\delta_i \delta_j] \cdot ||\mathbf{g}_i||^2 ||\mathbf{g}_j||^2 \right) - C_{\mathrm{low}}^2. \tag{3.70}$$

The value of $\mathbb{E}[\delta^4] = M_\delta(4)$ can be obtained from the approximation of the moments; on the other hand, the value of $\mathbb{E}[\delta_i \delta_j]$ has to be obtained by numerically solving the corresponding integral (3.35).

Chapter 4

Multiuser detection with imperfect channel state information

Contents

4.1 Introduction

So far, we have assumed that the receiver has perfect channel state information. This, however, is never true, and CSI must be acquired at the receiving end. Before finishing this first part of the thesis, we will devote this chapter to analyzing the effect of this impairment on the performance of multi-user detection schemes, with special emphasis on multibeam satellite systems.

We focus on data-aided estimation, where a distinct sequence of length L symbols, called unique word (UW), is transmitted by each terminal. Upon reception of them, the receiver performs some processing, which we assume linear, and obtains an estimate of the channel matrix, $\hat{\mathbf{H}}$. This estimate, and not the actual channel matrix (which remains unknown), will be the one used to perform multi-user detection.

To start with, we analyze the case where the estimate has the form $\mathbf{H} + \boldsymbol{\Sigma}^{\frac{H}{2}}\mathbf{E}$, where $\mathbf{H}$ is the actual channel matrix, $\boldsymbol{\Sigma}$ is the noise covariance matrix, and $\mathbf{E}$ is a random matrix modeling the estimation errors. This scenario is tough to analyze, because it requires manipulating the eigenvalue distribution of complicated matrices. The available literature reports solutions for some cases; for instance, [7] studies the performance of both ZF and MMSE receivers under imperfect CSI, but under the assumptions of $\mathbf{H}$ consisting of i.i.d. entries (which is common in terrestrial scenarios).

Here, we propose to study the mean squared error (MSE) obtained by the LMMSE receiver. Differently from other studies, we will allow $\mathbf{H}$ to be any fixed matrix with finite entries; this allows to accommodate any antenna radiation pattern, provided in numerical or analytical form, at the cost of focusing over a time interval during which $\mathbf{H}$ can be assumed not to change (that is, we focus on a coherence interval of the channel). Also, we will be assuming that the user terminals are fixed.

The proposed method relies on a tight approximation for the error covariance matrix for medium and high SNR, and then exploits the large dimensionality of the system by resorting to asymptotic random matrix theory. As a result, the MSE will be proven to converge to a non-random value as the matrix dimensions grow large while preserving their proportion, and this value will be computed efficiently by solving a system of equations with a unique solution. This procedure will be valid for any error matrix $\mathbf{E}$ with zero-mean and independent entries of finite variance, which leaves room for a number of estimation error profiles.

Before finishing the chapter, we will briefly look at a different scenario, in which the estimate is of the form $\mathbf{H}\mathbf{E}_1 + \mathbf{E}_2$, that is, there is also a multiplicative error. This model, which has been proposed in [5] for the case of imperfect synchronization, will also be analyzed through a high SNR approximation.

4.2   System model

4.2.1   Channel model

The signal model for a given time instant will read as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{4.1}$$

where $\mathbf{y} \in \mathcal{C}^{N \times 1}$ is the set of received symbols, $\mathbf{s} \in \mathcal{C}^{K \times 1}$ is the set of unit-power transmitted symbols, $\mathbf{H} \in \mathcal{C}^{N \times K}$ is the channel matrix and $\mathbf{n} \in \mathcal{C}^{N \times 1}$ contains zero-mean Gaussian samples with covariance matrix $\mathbb{E}\left[\mathbf{nn}^H\right] \doteq \boldsymbol{\Sigma}$; for notational convenience, we will also define $\tilde{\boldsymbol{\Sigma}}$ such that

$$\boldsymbol{\Sigma} = N_0 \tilde{\boldsymbol{\Sigma}} \tag{4.2}$$

with N_0 the noise power; note that both $\boldsymbol{\Sigma}$ and $\tilde{\boldsymbol{\Sigma}}$ are positive semidefinite by definition, and thus admit a factorization of the form $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{H/2}\boldsymbol{\Sigma}^{1/2}$.

Upon reception of $\mathbf{y}$, the gateway will use a LMMSE receiver to recover $\mathbf{s}$. However, we will assume that it has perfect knowledge of the noise covariance, but only an estimate of the channel matrix, $\hat{\mathbf{H}}$, so that the receiver operation will read as $\hat{\mathbf{s}} = \bar{\mathbf{W}}^H \mathbf{y}$, where $\bar{\mathbf{W}}^H$ is the LMMSE combining matrix built from channel estimates

$$\bar{\mathbf{W}}^H = \left(\mathbf{I} + \hat{\mathbf{H}}^H \boldsymbol{\Sigma}^{-1} \hat{\mathbf{H}}\right)^{-1} \hat{\mathbf{H}}^H \boldsymbol{\Sigma}^{-1}. \tag{4.3}$$

4.2.2 Channel estimation techniques

Synchronous transmissions

Recall that we focus on data-aided estimation. Provided that the transmission among all user terminals is synchronized, the set of signals received at the gateway would be

$$\mathbf{Y} = \mathbf{HC} + \mathbf{N} \tag{4.4}$$

where $\mathbf{H} \in \mathcal{C}^{K \times K}$ is the channel matrix, $\mathbf{Y} \in \mathcal{C}^{K \times L}$ is the stack of received sequences (one at each antenna element), $\mathbf{C} \in \mathcal{C}^{K \times L}$ is the stack of transmitted sequences and $\mathbf{N} \in \mathcal{C}^{K \times K}$ contains the additive noise samples.

Because of their limited computational complexity, linear estimation techniques are often preferred. Such techniques try to estimate the channel as

$$\hat{\mathbf{H}} = \mathbf{YA} \tag{4.5}$$

where $\mathbf{A}$ is a matrix that can be obtained following different criteria, as summarized in [89]. For example, the MMSE solution is given by

$$\mathbf{A} = \arg\min_{\mathbf{A}} \mathbb{E}\left[||\mathbf{H} - \mathbf{YA}||^2\right]. \tag{4.6}$$

On the other hand, the least squares (LS) estimate coincides with the maximum likelihood solution if the noise is Gaussian [40], and is given by

$$\hat{\mathbf{H}} = \mathbf{YC}^\dagger = \mathbf{H} + \mathbf{NC}^\dagger \tag{4.7}$$

where $\mathbf{C}^{\dagger}$ is the (right-hand side) pseudoinverse [90] of $\mathbf{C}$; this pseudoinverse will exist if $\mathbf{C}$ has full rank, and thus $L \geq K$ is a necessary condition. As proven among others in [89], orthogonal sequences minimize the mean squared estimation error, and thus Walsh-Hadamard sequences constitute a popular choice.

LS estimation has been a popular choice for satellite communications [5, 21, 91]; note that it does not require knowledge of the noise statistics or any channel properties.

Asynchronous transmission

Despite the existence of synchronization mechanisms at the system level, signals coming from different terminals present a degree of asynchronism at the gateway which may affect, among other things, the quality of the channel estimates. Starting from a vector of length L, $\mathbf{c}_k = (c_{k1} \ c_{k2} \ \cdots \ c_{kL})$, we have that

$$c_k(t) = \sum_{n=1}^{L} c_{kn} g(t - nT), \quad k = 1, \ldots, K \tag{4.8}$$

where $g(t)$ is the transmit pulse and T the symbol period. At a given time instant t, the received baseband waveform corresponding to the k-th training sequence can be written as [5]

$$c_k(t + \tau_k) e^{j\omega_k t + \theta_k}, \quad k = 1, \ldots, K \tag{4.9}$$

with $\{\tau_k, \omega_l, \theta_k\}$ the specific timing, frequency and phase parameters of the k-th received waveform.

Sampling the above sequences at symbol period would produce a matrix $\mathbf{C}$ in (4.4) with imperfectly known entries unless $\{\tau_k, \omega_l, \theta_k\}$ are precisely known for all k. Different timing and frequency recovery techniques have been tested specifically for multibeam return links using multiuser detection [92]; the latter reference proposed a hybrid algorithm, where a non-data aided estimate is refined by exploiting the pilot pattern, and concluded that good performance results can be obtained if the start of the frames can be correctly decoded, even though better pilot sequences would probably be required.

Even in the case that the estimation yields a perfectly known matrix $\mathbf{C}$ of received training sequences, their properties may have been lost with respect to those of the transmitted sequences. By dropping orthogonality some robustness can be obtained. To accomplish this, [5] proposes the use of pseudo-random sequences instead of Walsh-Hadamard sequences, trying to preserve performance even without orthogonality. Using correlation, and not the pseudo-inverse, is the proposal in this case, and so the estimated sequence reads as

$$\hat{\mathbf{H}} = \frac{1}{L} \mathbf{Y} \mathbf{C}^H = \frac{1}{L} \mathbf{H} \mathbf{C} \mathbf{C}^H + \frac{1}{L} \mathbf{N} \mathbf{C}^H \tag{4.10}$$

where we introduce the factor $1/L$ to normalize the output correlation.

In the next sections we will try to quantify the performance degradation introduced by these two channel estimation alternatives.

4.3 Performance with synchronous transmission

4.3.1 Estimation errors

In this section, and as in [21] and [5], we will neglect the impairments caused by imperfect synchronization, and assume that the different unique words can be considered orthogonal. Following these considerations, we come up with the following model for the estimation errors

$$
\begin{aligned}
\hat{\mathbf{H}} &= \mathbf{H} + \sqrt{\frac{N_0}{L}} \tilde{\boldsymbol{\Sigma}}^{H/2} \mathbf{E} \\
&= \mathbf{H} + \alpha \tilde{\boldsymbol{\Sigma}}^{H/2} \mathbf{E}
\end{aligned}
\tag{4.11}
$$

where $\mathbf{H}$ is a perfectly known deterministic matrix, $\mathbf{E}$ is assumed to be a matrix of independent zero-mean Gaussian entries, $\tilde{\boldsymbol{\Sigma}}$ has already been defined in Eq. 4.2, and we have defined $\alpha \doteq \sqrt{N_0/L}$ for simplicity.

From (4.11), we can see that the variance of the error at each entry will be proportional to the noise power and inversely proportional to the training sequence length. However, note that so far we have imposed no further constraints on the variance of each element in $\mathbf{E}$. As we will discuss in Section 4.3.3, we will allow each element to have a different variance as long as some mild constraints are satisfied.

4.3.2 Problem statement

The analytical prediction of the system performance in the presence of channel estimation errors is in general a difficult task. For example, the error covariance matrix, denoted as $\mathbf{Q}$, would read as

$$
\mathbf{Q} = \mathbb{E}\left[(\mathbf{s} - \hat{\mathbf{s}})(\mathbf{s} - \hat{\mathbf{s}})^H \right]
\tag{4.12}
$$

so that, averaging the error over all the users, we can define:

$$
\epsilon^2 \doteq \frac{1}{K} \operatorname{trace} \mathbf{Q}.
\tag{4.13}
$$

Recall that it can be analytically tackled for a number of cases of interest. For example, under the assumption of perfect CSI, the matrix $\mathbf{Q}$ would have the following expression:

$$
\mathbf{Q} = \left(\mathbf{I} + \mathbf{H}^H \boldsymbol{\Sigma}^{-1} \mathbf{H} \right)^{-1}.
\tag{4.14}
$$

Unfortunately, obtaining $\mathbf{Q}$ in our case would lead to a much more complex expression, so

that deriving a closed-form formula seems analytically intractable. To overcome this problem, in the next section we will explore the use of an approximation of $\mathbf{Q}$ which holds whenever the SINR is not very low. From this expression, we will make use of an existing result in the field of RMT to compute ϵ^2; we will also discuss on the conditions the system must meet for the validity of this result.

4.3.3 Approximation of ϵ^2

Approximation of Q

One of the problems for computing ϵ^2 is that, with imperfect CSI, the covariance matrix has a much more involved expression. As a consequence, we will firstly derive a much more tractable approximation which will hold for medium and high SINR values.

Lemma 4. *Assume that the receiver uses a LMMSE receiver, but has only an imperfect estimation of the channel matrix* $\mathbf{H}$ *given by* $\hat{\mathbf{H}} = \mathbf{H} + \alpha\tilde{\boldsymbol{\Sigma}}^{H/2}\mathbf{E}$. *Then, the error covariance matrix can be approximated by*

$$\mathbf{Q} \approx \left(1 + \frac{K}{L}\right)\left(\mathbf{I} + \hat{\mathbf{H}}^H\boldsymbol{\Sigma}^{-1}\hat{\mathbf{H}}\right)^{-1}. \tag{4.15}$$

Proof. Expanding the channel model, we have

$$\mathbf{y} = \hat{\mathbf{s}} - \alpha\tilde{\boldsymbol{\Sigma}}^{H/2}\mathbf{E}\mathbf{s} + \mathbf{n}. \tag{4.16}$$

Let us define $\mathbf{u} \doteq \alpha\tilde{\boldsymbol{\Sigma}}^{H/2}\mathbf{E}\mathbf{s} + \mathbf{n}$ with covariance matrix

$$\begin{aligned}
\mathbb{E}\left[\mathbf{u}\mathbf{u}^H\right] &= \alpha^2\tilde{\boldsymbol{\Sigma}}^{H/2}\mathbb{E}\left[\mathbf{E}\mathbf{E}^H\right]\tilde{\boldsymbol{\Sigma}}^{1/2} + \boldsymbol{\Sigma} \\
&= \alpha^2 K\tilde{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma} \\
&= \left(1 + \frac{K}{L}\right)\boldsymbol{\Sigma}.
\end{aligned} \tag{4.17}$$

After applying the LMMSE receiver, we will have

$$\hat{\mathbf{s}} = \bar{\mathbf{W}}^H\mathbf{H}\mathbf{s} - \bar{\mathbf{W}}^H\mathbf{u}. \tag{4.18}$$

To obtain an approximation, we will start by focusing on the SINR interval in which the contribution to the error is much greater in $\mathbf{W}^H\mathbf{u}$ than in $\mathbf{W}^H\mathbf{H}\mathbf{s}$; to this end, recall that the transmit power is included into $\mathbf{H}$. Assuming this is so, we can approximate the covariance matrix by

$$\mathbf{Q} \approx \mathbb{E}\left[\bar{\mathbf{W}}^H\mathbf{u}\mathbf{u}^H\bar{\mathbf{W}}\right] = \left(1 + \frac{K}{L}\right)\bar{\mathbf{W}}^H\boldsymbol{\Sigma}\bar{\mathbf{W}}. \tag{4.19}$$

Plugging the expression of $\bar{\mathbf{W}}$ we get

$$
\begin{aligned}
\mathbf{Q} &\approx \left(1 + \frac{K}{L}\right) \left(\mathbf{I} + \hat{\mathbf{H}}^H \mathbf{\Sigma}^{-1} \hat{\mathbf{H}}\right)^{-1} \hat{\mathbf{H}}^H \mathbf{\Sigma}^{-1} \hat{\mathbf{H}} \left(\mathbf{I} + \hat{\mathbf{H}}^H \mathbf{\Sigma}^{-1} \hat{\mathbf{H}}\right)^{-1} \\
&\approx \left(1 + \frac{K}{L}\right) \left(\mathbf{I} + \hat{\mathbf{H}}^H \mathbf{\Sigma}^{-1} \hat{\mathbf{H}}\right)^{-1}
\end{aligned}
\tag{4.20}
$$

where in the last equality we have used the approximation $\left(\mathbf{I} + \hat{\mathbf{H}}^H \mathbf{\Sigma}^{-1} \mathbf{H}\right)^{-1} \hat{\mathbf{H}}^H \mathbf{\Sigma}^{-1} \hat{\mathbf{H}} \approx \mathbf{I}$ for simplifying the first two factors. $\qquad\square$

The main advantage of the expression above is that it allows a simple computation of ϵ^2. Recalling that $\mathbf{\Sigma} = N_0 \tilde{\mathbf{\Sigma}}$, then

$$
\begin{aligned}
\epsilon^2 &= \frac{1}{K} \operatorname{trace}\left[\left(\mathbf{I} + \hat{\mathbf{H}}^H \mathbf{\Sigma}^{-1} \hat{\mathbf{H}}\right)^{-1}\right] \\
&= \frac{1}{K} \sum_{i=1}^{K} \frac{1}{1 + \lambda_i\left\{\hat{\mathbf{H}}^H \mathbf{\Sigma}^{-1} \hat{\mathbf{H}}\right\}} \\
&= \frac{N_0}{K} \sum_{i=1}^{K} \frac{1}{N_0 + \lambda_i\left\{\hat{\mathbf{H}}^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}}\right\}}
\end{aligned}
\tag{4.21}
$$

where $\lambda_i\{\}$ denotes the i-th largest eigenvalue of the matrix between brackets. In the limit, we have that

$$
\begin{aligned}
\lim_{N,K \to \infty} \epsilon^2 &= \lim_{N,K \to \infty} \frac{N_0}{K} \sum_{i=1}^{K} \frac{1}{N_0 + \lambda_i\left\{\hat{\mathbf{H}}^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}}\right\}} \\
&= N_0 \mathbb{E}\left[\frac{1}{N_0 + \lambda\left\{\hat{\mathbf{H}}^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}}\right\}}\right] \\
&= N_0 \cdot \mathcal{S}_{\hat{\mathbf{H}}^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}}}(-N_0).
\end{aligned}
\tag{4.22}
$$

Here, $\mathcal{S}_{\mathbf{A}}(z)$ denotes the *Stieltjes transform* [56, 57] of the (empirical eigenvalue distribution of) matrix $\mathbf{A}$, hence the last equality follows directly from its definition

$$
\mathcal{S}_X(z) = \mathbb{E}\left[\frac{1}{X - z}\right].
\tag{4.23}
$$

Computing the Stieltjes transform

Linking ϵ^2 with the Stieltjes transform of $\hat{\mathbf{H}}^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}}$ will prove very useful for our purposes, since this transform has been derived for a wide range of matrices, many times without requiring the actual empirical eigenvalue distribution. In our case, we will make use of the following result from [93]:

Lemma 5. *Let $\mathbf{\Delta}$ be a matrix of the form*

$$\mathbf{\Delta} = \frac{1}{\sqrt{K}}\mathbf{\Xi} \odot \mathbf{X} \tag{4.24}$$

where $\odot$ denotes the Hadamard (or entry-wise) product, $\mathbf{X}$ has zero-mean i.i.d elements with some finite moment of order higher than four and $\mathbf{\Xi}$, the so-called variance profile, has only real finite entries; consider also a deterministic matrix $\mathbf{A}$ whose columns have finite Euclidean norm, and denote $\mathbf{\Theta} = \mathbf{A} + \mathbf{\Delta}$.

Then, there exists a deterministic matrix $\tilde{\mathbf{T}}(N_0)$ such that, while the ratio K/N is kept constant,

$$\lim_{N,K\to\infty} \frac{1}{K}\operatorname{trace}\tilde{\mathbf{T}}(N_0) = N_0 \cdot \mathcal{S}_{\mathbf{\Theta}^H\mathbf{\Theta}}(-N_0). \tag{4.25}$$

In [93] they also prove that $1/K$ trace $\tilde{\mathbf{T}}(N_0)$ can be computed by solving a system of $N+K$ equations which has a unique solution.

Applying this result to our problem, we immediately have that

$$\mathbf{A} = \tilde{\mathbf{\Sigma}}^{-H/2}\hat{\mathbf{H}}, \qquad \mathbf{\Delta} = \alpha\mathbf{E}. \tag{4.26}$$

The particular case $\mathbf{\Sigma} = N_0\mathbf{I}$, $\mathbf{\Xi}_{ij} = 1$ Assume now that $\mathbf{\Sigma} = N_0\mathbf{I}$, and that the elements of $\mathbf{E}$ have unit variance. These conditions lead to a very simple, illustrative solution since, in this case, the system of equations $N + K$ to be solved reads as

$$
\begin{aligned}
x &= \operatorname{trace}(\mathbf{T}) \\
y &= \operatorname{trace}(\tilde{\mathbf{T}}) \\
\mathbf{T} &= \left(N_0(1+\alpha x)\mathbf{I} + \frac{1}{1+\alpha y}\hat{\mathbf{H}}\hat{\mathbf{H}}^H \right)^{-1} \\
\tilde{\mathbf{T}} &= \left(N_0(1+\alpha y)\mathbf{I} + \frac{1}{1+\alpha x}\hat{\mathbf{H}}^H\hat{\mathbf{H}} \right)^{-1}
\end{aligned}
\tag{4.27}
$$

with $\alpha = N_0/L$; as proven in [93], the solution is unique.

The above system can be straightforwardly extended to the case $\mathbf{\Xi}_{ij} = C$, with C some real, finite constant different from 1, by a simple scaling of α. However, when the assumptions above are not met, we must resort to a more involved formulation of the system which requires the definition of some auxiliary matrices; the expression for the general case can be checked on Appendix 4.A.

Before going further and showing the simulation results, let us summarize the procedure we have described so far: in order to compute ϵ^2, we start by obtaining an approximation of the error covariance matrix $\mathbf{Q}$ (Lemma 1); then, we use a result from RMT (Lemma 2) to compute its trace –and thus compute ϵ^2– by solving a system of $N + K$ equations.

4.3.4 Simulation results

In this section, we illustrate the performance of the derived approximation by computer simulations. The simulator setup was summarized in Table 2.1, only that in this case we used 10000 Monte Carlo realizations, clear sky conditions and fixed UT location.

Additionally, we will also test the case in which $\boldsymbol{\Sigma}$ is not diagonal, but has small off-diagonal elements. We will do so by assuming a fixed beamforming matrix $\mathbf{B} \in \mathcal{C}^{100 \times 155}$, with $\mathbf{BB}^H$ a matrix with small off-diagonal elements[1], to be applied on the received vector $\mathbf{y}$ before any processing; this is the same matrix used in the previous chapter, and is aimed at minimizing the average interference for a uniform distribution of the users within their beam spots. With this fixed beamforming, the deterministic matrix $\mathbf{A}$ would be given by $\mathbf{BH}\tilde{\boldsymbol{\Sigma}}^{-H/2}$, whereas the random matrix is $\alpha\tilde{\boldsymbol{\Sigma}}^{1/2}\mathbf{E}$ and $\boldsymbol{\Sigma} = N_0\mathbf{BB}^H$.

The system of equations (4.27) was solved by Matlab's function *fminsearch*. With uncorrelated noise, the convergence of the algorithm is very fast –almost instantaneous– but, with correlated noise, the new variance profile of matrix $\boldsymbol{\Delta}$ and the existence of some non-zero off-diagonal elements will affect the accuracy and convergence speed; still, results can be obtained fast by feeding the algorithm with a good starting point.

Focusing on the results, Figure 4.1 compares the derived approximation with the actual Monte Carlo results as a function of the terminals' equivalent isotropic radiated power (EIRP); the accuracy is quite remarkable whenever the EIRP is not very low, since that would be the area where the approximation of $\mathbf{Q}$ in (4.19) does not hold. Additionally, Figure 4.2 depicts the accuracy by means of the *magnitude of the relative error*. We can see that that the proposed method is more accurate when $\tilde{\boldsymbol{\Sigma}} = \mathbf{I}$, and that its accuracy increases as the SINR increases.

4.4 Performance with asynchronous transmission

Here we will tackle the case where pseudo-random sequences and the correlation procedure are used. After obtaining an approximation for the error covariance matrix, we will see that an error floor appears.

4.4.1 Estimation errors

Recall that in this case the estimation errors also appear multiplying the channel matrix (Eq. 4.10)

$$\hat{\mathbf{H}} = \mathbf{H}\frac{\mathbf{CC}^H}{L} + \frac{1}{L}\mathbf{NC}^H \qquad (4.28)$$

so that it is not only additive, but also multiplicative; for convenience let us denote $\mathbf{R} = \mathbf{CC}^H/L$.

---

[1]This is a consequence of $\mathbf{B}$ having to force zeros on off-diagonal elements of the product $\mathbf{BH}$ in order to minimize interference.
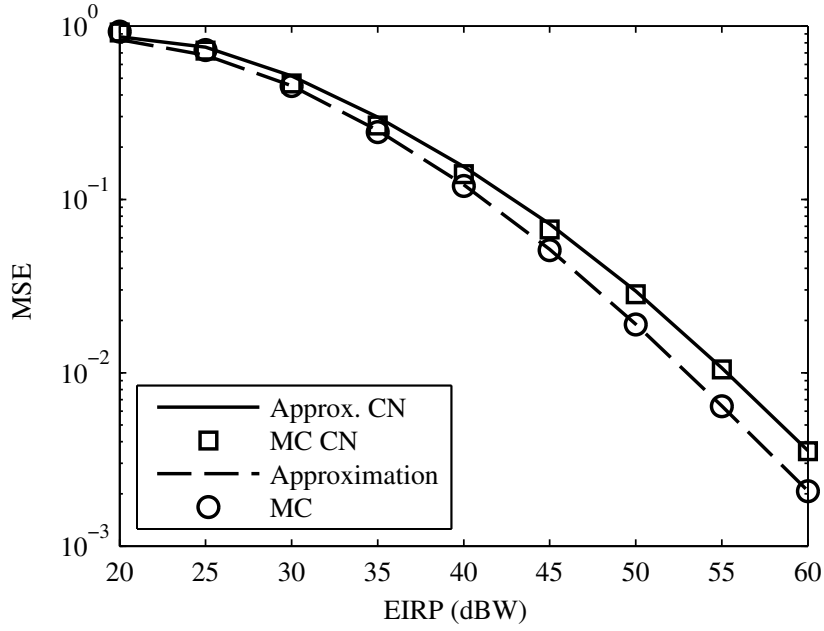
Figure 4.1: Comparison between the MSE and its approximation, both for uncorrelated noise and correlated noise (CN). MC stands for Monte Carlo, and Approx. is the derived approximation.
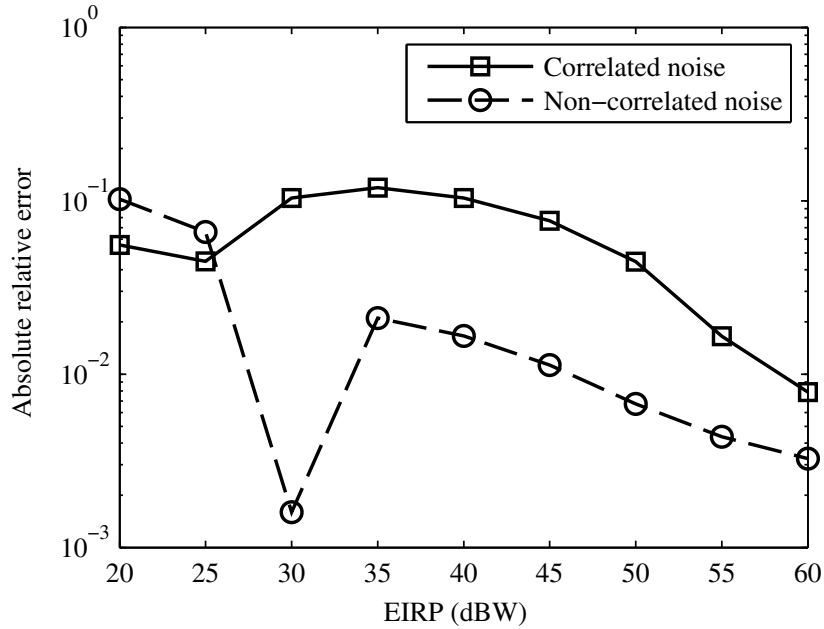
Figure 4.2: Accuracy of the approximation measured by the magnitude of the relative error.

4.4.2 Approximation of the error covariance matrix

As in the case of LS estimation with orthogonal sequences (4.3), obtaining a closed form expression for $\mathbf{Q}$ is a hard task. In the following, we will resort to a high SNR analysis that will show the ultimate limits imposed by this estimation method.

When the SNR is high, that is, when N_0 can be considered low, then

$$\hat{\mathbf{H}} \approx \mathbf{HR} \tag{4.29}$$

and the MMSE receiver can be approximated by the ZF one

$$
\begin{aligned}
\hat{\mathbf{W}}^H &\approx (\hat{\mathbf{H}}^H\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^H = (\mathbf{R}^H\mathbf{H}^H\mathbf{HR})^{-1}\mathbf{R}^H\mathbf{H}^H \\
&= \mathbf{R}^{-1}\mathbf{W}^H
\end{aligned}
\tag{4.30}
$$

where $\mathbf{W}^H$ represents the MMSE combining matrix obtained from the actual channel matrix, $\mathbf{H}$.

We can now obtain the error covariance matrix as

$$
\begin{aligned}
\mathbf{Q} &= \mathbb{E}\left[(\mathbf{x} - \hat{\mathbf{W}}^H\mathbf{y})(\mathbf{x} - \hat{\mathbf{W}}^H\mathbf{y})^H\right] = \\
&= \mathbb{E}[\mathbf{xx}^H] - \mathbb{E}[\mathbf{xy}^H]\hat{\mathbf{W}} - \hat{\mathbf{W}}^H\mathbb{E}[\mathbf{yx}^H] + \hat{\mathbf{W}}^H\mathbb{E}[\mathbf{yy}^H]\hat{\mathbf{W}} \\
&= \mathbf{I} - \mathbf{H}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{R}^{-1} - \mathbf{R}^{-1}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{H} \\
&\quad + \mathbf{R}^{-1}\left(\mathbf{HH}\right)^{-1}\mathbf{H}^H\mathbf{H}\left(\mathbf{HH}\right)\left(\mathbf{HH}\right)^{-1}\mathbf{R}^{-1} \\
&= \mathbf{R}^{-2} - 2\mathbf{R}^{-1} + \mathbf{I}.
\end{aligned}
\tag{4.31}
$$

The expression above states that, when the SNR is high enough for the assumptions to hold, then the cross-correlation properties of the training sequences (present in $\mathbf{R}$) are the only factor limiting performance. Interestingly, the obtained SINR value is constant, independent of the SNR; this is consistent with the observation of a *jitter floor* in this estimation process from [91].

4.4.3 Simulation results

To test the derived expression, we have simulated the channel estimation using a set of 155 pseudo-random, linearly independent sequences of length $L = 1000$ symbols, the same used in [91]; to illustrate their cross-correlation properties, Figure 4.3 shows the entries of matrix $\mathbf{R}$ for this case.

The obtained results are shown in Figure 4.4; the use of correlation produces an error floor of about $-9\,\mathrm{dB}$, which roughly means that the maximum SINR achievable after LMMSE combining is of $9\,\mathrm{dB}$, irrespectively of how large the terminals' EIRP is.
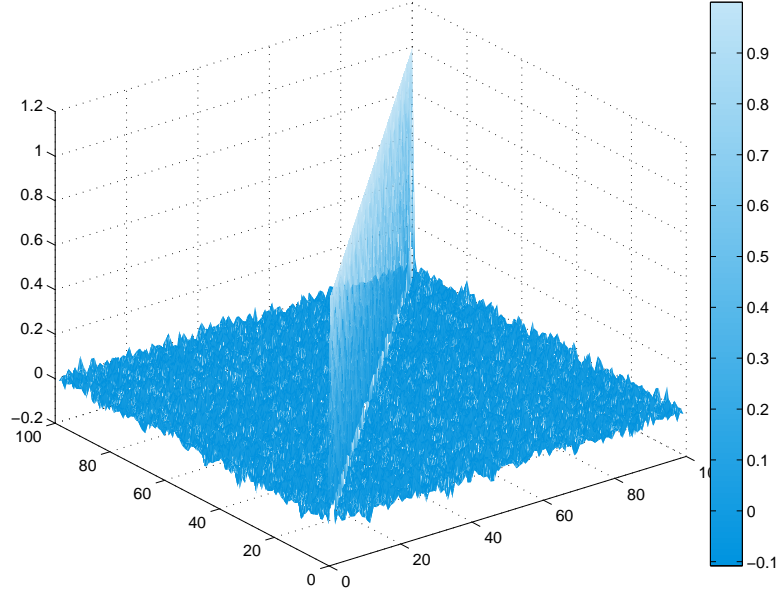
Figure 4.3: Entries of matrix **R** for $L = 1000$.
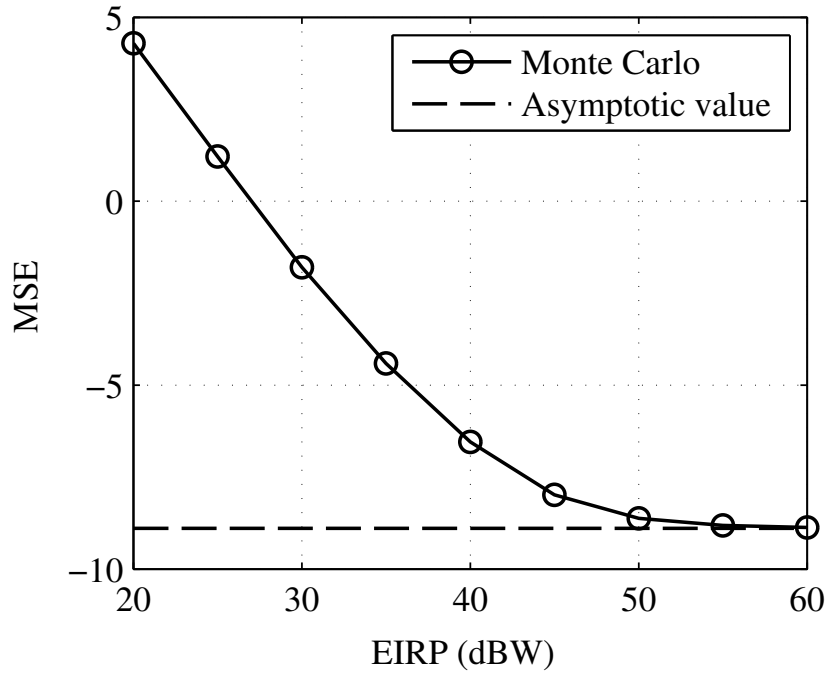
Figure 4.4: Mean squared error, in dB, as a function of the terminals' EIRP when estimation is performed by correlation. $L = 1000$.

4.5 Conclusions

In this chapter, we studied the impact of non-perfect CSI on the performance of the LMMSE receiver. Two data-aided solutions were analyzed: the least-squares solution with orthogonal (Walsh-Hadamard) sequences, and estimation by correlation for pseudo-random, linearly independent sequences.

For the first case, a tight approximation for the mean-squared error has been found. Our method relies on an approximation of the error covariance matrix, for which existing results in the field of random matrix theory can be applied. The derived approximation has been proven to work both with uncorrelated and correlated channel estimation errors, and simulation results have shown a remarkable tightness at medium and high SINR.

Regarding the second case, a high SNR approximation was also obtained. In this case, an error floor has been shown to appear because of the use of the correlation procedure, bounding the achievable output SINR.

The content in this chapter was done in collaboration with Prof. Carlos Mosquera, and partially published in Globecom 2011 [17] and Asilomar 2012 [94].

Appendix 4.A   Finding T in the general case

When the variance profile of matrix $\boldsymbol{\Delta}$ is not constant, then we must resort to a more general formulation of the system, as described in [93]. In this case, before describing the expressions of $\mathbf{T}$ and $\tilde{\mathbf{T}}$, we need to define

$$
\begin{aligned}
\psi_i(z) &= -\frac{1}{z\left(1 + \frac{1}{K}\operatorname{trace}(\tilde{\mathbf{D}}_i\tilde{\mathbf{T}}(z))\right)} & 1 \le i \le N \\
\tilde{\psi}_j(z) &= -\frac{1}{z\left(1 + \frac{1}{K}\operatorname{trace}(\mathbf{D}_j\mathbf{T}(z))\right)} & 1 \le j \le K
\end{aligned}
\tag{4.32}
$$

where

$$
\mathbf{D}_j = \operatorname{diag}\left(\operatorname{col}_j(\boldsymbol{\Xi})\right), \qquad \mathbf{D}_i = \operatorname{diag}\left(\operatorname{row}_i(\boldsymbol{\Xi})\right).
\tag{4.33}
$$

Here, $\operatorname{col}_i(\mathbf{M})$ and $\operatorname{row}_i(\mathbf{M})$ denote the i-th column and the i-th row of matrix $\mathbf{M}$, respectively, and $\operatorname{diag}(\mathbf{m})$ denotes a diagonal matrix whose diagonal is constituted by the elements in vector $\mathbf{m}$.

From (4.32), we also build

$$
\Psi(z) = \operatorname{diag}(\psi_i(z)), \qquad \tilde{\Psi}(z) = \operatorname{diag}(\tilde{\psi}_j(z))
\tag{4.34}
$$

and finally

$$\mathbf{T}(z) = \left(\Psi^{-1}(z) - z\mathbf{A}\tilde{\Psi}(z)\mathbf{A}^H \right)^{-1}$$
$$\tilde{\mathbf{T}}(z) = \left(\tilde{\Psi}^{-1}(z) - z\mathbf{A}^H\Psi(z)\mathbf{A} \right)^{-1}.$$

(4.35)

Part II

Link adaptation in mobile satellite communications

Chapter 5

Introduction to mobile satellite communications

Contents

5.1 Summary

In mobile satellite communications, there is an increasing need for more efficient transmission techniques that enable higher bit-rates at an affordable cost. To this extent, adaptive coding and modulation (ACM) allows the provision of broadband services to large user populations at lower costs, since it makes it possible to operate the links more efficiently by selecting the most suitable modulation and coding scheme (MCS) at each time [13]. However, the use of ACM for mobile links is hindered by the behavior of the land mobile satellite (LMS) channel [95]. This channel is usually modeled by a fast fading component, whose spectrum is related to the mobile speed by the Doppler effect, superimposed on a slow shadowing component; the parameters of both fading and shadowing depend on the environment in which the receiver happens to be. In short, the mobility of the user terminal will cause fast, difficult to predict channel variations, which will pose additional difficulties on the design of both forward and return link strategies.
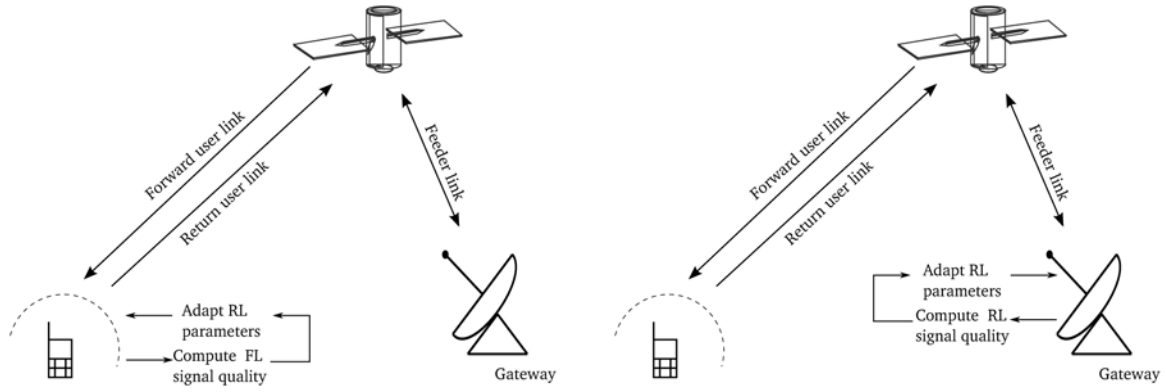
Figure 5.1: Open-loop (left) against closed-loop (right) adaptation.

Elaborating more on this issue, adaptation can be performed in *open-loop* or *closed loop* mode. In an open-loop scheme, the transmitter directly measures the signal quality from the reverse link and changes the parameters accordingly; it usually enjoys negligible delays, although at the price of having only partial information when both links are not perfectly correlated. On the other hand, closed-loop strategies wait for the other end to process their data and operate upon receiving some information about its reception. This makes them more accurate, although the experienced delay is much higher. In Figure 5.1 we show a diagram with the difference between closed loop and open loop adaptation.

Another key problem is that it is difficult to relate the channel statistics -even if we knew them- to the performance of the link because the channel is time varying. For example, it is difficult to tell what is best, a fixed channel with low quality or a fast varying channel with sharp transitions and a higher average quality. In an LMS channel, where transmission usually entails the use of complicated channel codes, if is difficult to do this without resorting to extensive simulations. This, however, is highly undesired for the design of an ACM strategy. In order to overcome this problem, we propose to use physical layer abstraction techniques. Thus, we will use effective SNR metrics (ESM) that have been reported to map complicated channel profiles with their actual performance in a one-to-one manner [96]. Therefore, our key assumption will be the following: a codeword will be correctly transmitted if the ESM of the channel that it undergoes is higher than or equal to the decoding threshold specified for the MCS in use in a static channel.

Focusing firstly on the forward link, the fading is attached to the receiver and, as a consequence, the delay experienced by the CSI will be much longer than the channel coherence time for most speeds of practical interest. Therefore, the received CSI will be completely outdated -even if considering an open-loop scheme [97]- and rate adaptation will be of no use. However, if frames can be retransmitted or, at least, some additional redundancy or parity bits can be sent, then higher throughputs can be achieved at the cost of some latency. The underlying idea would be to transmit at a higher rate during good states of the channel, while somehow keeping

the outage probability low. To achieve this, we propose the use of different MCS for different retransmission indexes. A high throughput is achieved by using high rate MCS in the first transmissions, and an outage constraint is met by reducing the rate when the retransmission index is high. A similar idea was presented in [98], where multi-layer codes (MLC) and MCS that change with the retransmissions are used.

On the other hand, the return link enjoys timely channel information when operating in open-loop mode, but at the cost of having only partial CSI. This partial CSI knowledge is caused by the operation of forward and return links on different frequencies, so the multipath fading is uncorrelated [99]. The shadowing component, however, is the same for both channels. In some cases, open loop adaptation might offer a better performance than closed loop, but selecting the optimum operating regime is not a simple task. We propose to adaptively weight the open loop and closed loop CSI using only the ACK/NAK interchange.

5.2 Signal and system model

The signal model at a given time instant i, for both the forward and return link, is

$$y_i = \sqrt{\text{snr}} \cdot h_i^{\text{xl}} x_i + w_i \tag{5.1}$$

with y_i the received symbol, s_i the transmitted symbol, h_i^{xl} the channel coefficient, with xl $\in$ {fl, rl} the channel in the forward or return link[1], and snr the signal to noise ratio; we also define $\sigma^2 \triangleq \text{snr}^{-1}$. Accordingly, w_i is the unit-power noise plus interference contribution.

In the following, we describe the channel model and present some assumptions on the coding of the system under study. The methods developed in this part of the thesis do not depend on the specific channel model, that will be abstracted by the use of ESM.

5.2.1 Channel model

We will assume a frequency flat channel which, in turn, will be selective over time, with faster or slower fluctuations depending on the speed of the terminal. In particular, we assume h_i^{xl} follows a Loo distribution [100]: slow variations in the LOS component (*shadowing*) are described by a log-normal distribution, whereas fast fluctuations of the signal amplitude (*fading*) are given by a Rician distribution. The PDF of the signal amplitude at a given time instant would be given by

$$f_r(x|t = t_0) = \frac{x}{b_0} \exp\left(-\frac{x^2 + m(t_0)^2}{2b_0}\right) I_0\left(\frac{x \cdot m(t_0)}{b_0}\right) \tag{5.2}$$

where $m(t_0)$ is the offset of the Rician distribution at time instant t_0 (value of the LOS component), b_0 is the scale parameter (power of the non-line-of-sight, NLOS, component) and I_0 is

---

[1]Throughout this part of the thesis, the subscript fl or rl might be removed if the context is clear.

the modified Bessel function of the first kind [79]. Now, given that the LOS component changes over time, we can express the PDF of the signal amplitude as

$$f_r(x) = \mathbb{E}_m \left[\frac{x}{b_0} \exp\left(-\frac{x^2 + m^2}{2b_0} \right) I_0 \left(\frac{x \cdot m}{b_0} \right) \right];$$ (5.3)

finally, assuming that the LOS component follows a log-normal distribution, the above PDF can be written as

$$f_r(x) = \frac{x}{b_0 \sqrt{2\pi d_0}} \int_0^\infty \frac{1}{z} \exp\left(-\frac{(\log z - \mu)^2}{2d_0} - \frac{x^2 + z^2}{2b_0} \right) I_0 \left(\frac{x \cdot z}{b_0} \right) \, \mathrm{d}z$$ (5.4)

where d_0 and μ are the scale parameter and the location parameter of the log-normal distribution, respectively, and b_0 is the variance of the Rician distribution; to determine these parameters we follow the Fontan 3-state model [95]. If we average over the different states of the Fontan model, the resulting channel amplitude has the form of a Loo mixture, with PDF

$$f_r(x) = \sum_{i=1}^N p_i \frac{x}{b_0 \sqrt{2\pi d_{0,i}}} \int_0^\infty \frac{1}{z} \exp\left(-\frac{(\log z - \mu_i)^2}{2d_{0,i}} - \frac{x^2 + z^2}{2b_{0,i}} \right) I_0 \left(\frac{x \cdot z}{b_{0,i}} \right) \, \partial z$$ (5.5)

From an implementation point of view, [95] advocates for generating the LOS component by first obtaining independent Gaussian samples n, exponentiating them $10^{(n/20)}$ to obtain log-normally distributed numbers, and finally interpolating them to obtain the correlation properties specified by the model. This procedure, followed by many more references afterwards, has two main drawbacks: the resulting sequence is not log-normal (interpolating a log-normal sequence does not preserve the original distribution), and the resulting process is not stationary.

A different procedure that preserves log-normality and ensures stationarity is proposed here. Starting from i.i.d. Gaussian samples, the correlation properties are introduced before the exponentiation [101] by a low pass filter whose cutoff frequency is given by $f_{\mathrm{LOS}} = v \cdot T_{\mathrm{symb}}/d_{\mathrm{c}}$ where v is the terminal speed, T_{symb} is the symbol period and d_{c} is the measured correlation distance of the LOS component, which we have obtained from [102].

The NLOS component, on the other hand, is obtained by filtering complex Gaussian samples with a low-pass filter whose cutoff frequency is given by the Doppler spread [103, 104]. Through this thesis we will assume that forward and return link channels experience the exact same LOS realization, but independent NLOS components with the same b_0 parameter and equal Doppler frequencies[2].

To sum up, the forward and return link channels are generated by

$$h_i^{\mathrm{rl}} = h_i^{\mathrm{LOS}} + h_i^{\mathrm{NLOS,rl}}, \;\; h_i^{\mathrm{fl}} = h_i^{\mathrm{LOS}} + h_i^{\mathrm{NLOS,fl}}.$$ (5.6)

---

[2]This is a simplification, as return and forward link operate in different frequencies. The difference between them, however, will always be small in our work.
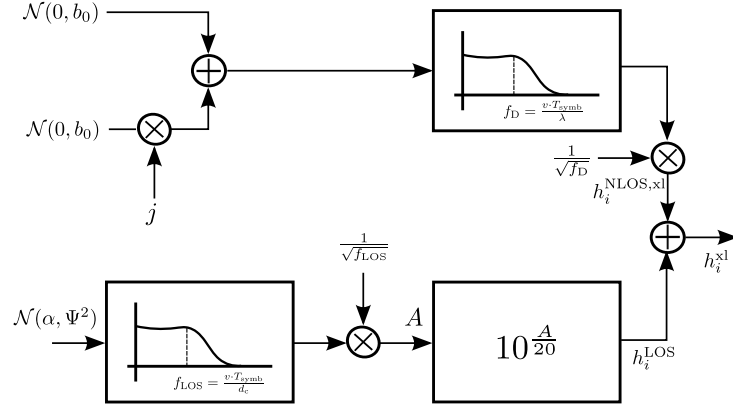
Figure 5.2: Diagram of the channel generation process. α and Ψ denote the mean and standard deviation of the log-normal shadowing when expressed in dB.

A block diagram of the channel generation process is shown in Figure 5.2.

5.3 Physical layer abstraction

5.3.1 Overview

Obtaining CSI in an LMS channel is a compelling task, among other reasons because it is difficult to capture the channel's behavior with a single parameter. As an example, think of how the average SINR would perform in such an environment, with both LOS and NLOS components possibly changing within the time span of a codeword: the same average SINR value would appear for very different channel realizations, with different end-to-end performance also.

If the channel varies within the time span of a codeword, then we need tools to tell whether each codeword will be correctly decoded or not without simulating the whole transmission chain. Such tools would be very interesting for the design of an ACM control strategy if they allowed us to predict whether a codeword will be decoded or not, in a simple way, only from the samples of the channel it will experiment.

Therefore, our goal is quite similar to the one pursued by Physical Layer (PHY) Abstraction techniques: to obtain a metric -that we can compute from estimations of some parameters- which is related in a one-to-one manner to the PER. Such metrics, often called Effective SNR Mappings (ESM), have been studied in [96, 105, 106] among others. Next we summarize the general procedure for PHY abstraction and describe the metric that we have used in this thesis.

5.3.2 Definition

The general procedure for PHY abstraction, as summarized on Fig. 5.3, is the following: for γ_i the SINR experienced by the i-th codeword, the expression for the effective SNR is given by
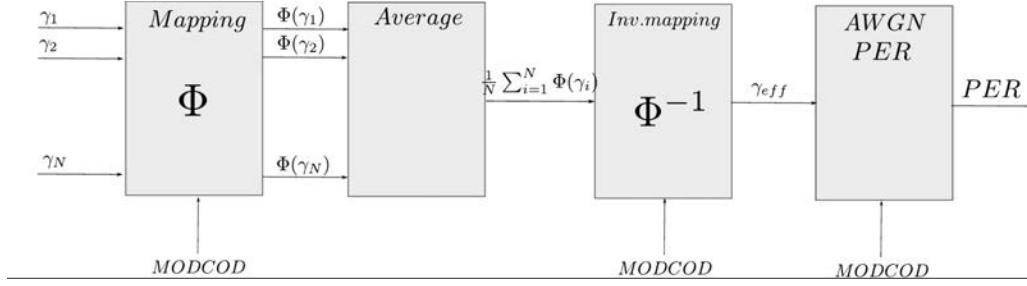
Figure 5.3: General procedure for effective SNR mapping.

$$\gamma_{eff} = \alpha \Phi^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \Phi \left(\frac{\gamma_i}{\beta} \right) \right) \tag{5.7}$$

and the ensuing steps are:

1. Choose a quality metric Φ.

2. Adjust parameters α and β according to measurements/simulations. These parameters are usually dependent on the MCS.

3. Perform simulations to obtain the PER curve in AWGN channel for each MCS.

Apart from in the previously cited references, the accuracy of this abstraction procedure has been tested specifically for mobile satellite communications in [107], among others; for completeness, in Appendix 5.A we report the results of a simulation campaign performed with the aid of a DVB-S software receiver.

A number of metrics for PHY abstraction exist, as reported e.g. in [108]. We will next describe the most relevant metric for our work, along with some of their properties.

- We will call Average SNR (ASNR) to the average of the instantaneous SNR values seen by the symbols within a codeword. Technically, this implies $\alpha = \beta = \Phi = 1$.

- The simplest ESM will probably be the exponential one (EESM), given by

$$\Phi(x) = e^{-\beta x}. \tag{5.8}$$

Although not the most accurate for some applications, its importance is in part due to the fact that it can be used to approximate other, more elaborated metrics, as we will show.

- The ESM metric in which we will focus is the Mutual Information Effective SNR Mapping (MIESM), which must be parametrized in terms of just the constellation used, not the code. It reads as

$$\gamma_{eff} = \frac{1}{M} \left(\frac{1}{N} \sum_{n=1}^{N} \Phi \left(\gamma_n \right) \right) \tag{5.9}$$

where M is the number of points in the constellation. If M changes between different symbols, then the normalization term $1/M$ must be included into the summation. The metric Φ is usually called SI and represents the mutual information associated to a symbol [109], that is

$$\text{SI} = \log_2 M - \frac{1}{M} \sum_{j=1}^{M} \mathbb{E}_w \left\{ \log_2 \left(1 + \sum_{\substack{k=1 \\ k \neq i}}^{M} e^{-\frac{|x_i - x_k + w|^2 - |w|^2}{\sigma^2}} \right) \right\}. \tag{5.10}$$

Here, x_i represents a point in the constellation and w is a zero-mean complex Gaussian random variable with variance $\sigma^2 = 1/(\gamma_n)$. This expression can be easily computed offline for an interval of instantaneous SINRs, where the expected value would be obtained via Monte Carlo. If a closed-form expression is required to perform further analytical studies, then it is possible to approximate SI by a sigmoid function, as explained in [106]. The simplest possible approximation would be

$$\Phi(x) \approx 1 - e^{-\beta x}, \tag{5.11}$$

which can be seen to be equivalent to the exponential effective SNR metric (EESM)[3] [108]. Moreover, the following approximation has been reported to show a remarkably good fit to the original function:

$$\Phi(\gamma) \approx 1 - \alpha_1 e^{-\beta_1 \gamma} + (1 - \alpha_1) e^{-\beta_2 \gamma}. \tag{5.12}$$

In either case, parameters α and β must be tuned for each constellation. The accuracy of each approximation has been illustrated on Fig. 5.4.

5.3.3 Modeling of the ESM in an LMS channel

Despite its advantages, the ESM is difficult to model analytically for an LMS channel, even when resorting to a simple exponential approximation. Before finishing this chapter we will summarize one of the modeling alternatives for a simplified case. A larger version was presented at the ICSSC 2012 conference [110], where it was exploited for adaptation; this work was performed in collaboration with Dr. Alberto Rico-Alvariño and Prof. Carlos Mosquera.

Let us start by assuming that the line of sight component does not change noticeably within the time span of a codeword. If this is the case, then the channel seen by a codeword follows a correlated multivariate Rician distribution. From this, it seems not possible to obtain the distribution of γ_{eff} in closed form; however, some approximations have been proposed in the literature.

In particular, [111] proposes modeling γ_{eff} by a log-normally distributed random variable

---

[3]Note that mapping one minus the mutual information is the same as mapping the mutual information as long as we select the appropriate inverse transformation.
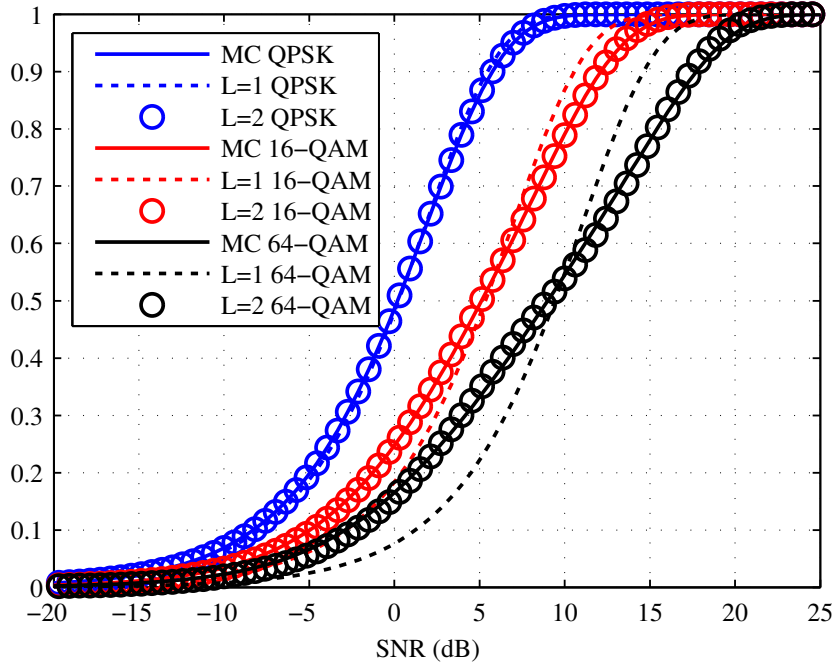
Figure 5.4: Comparison among different approximations to the MI.

for a frequency selective wireless channel. This approximation can be seen to suit also a satellite scenario. Figure 5.5 shows simulation results for a mobile satellite link operated at a center frequency of $2.2\,\text{GHz}$, with 8100 symbol codewords, QPSK modulation and $T_{\text{symb}} = 2.7\,\mu\text{s}$.

Even though the approximation is good, specially as the speed increases (because the samples within a codeword become less correlated), it is still difficult to exploit this. An approximation of these characteristics would require the knowledge of the first two moments of γ_{eff}, which are difficult to estimate and require the knowledge of the parameters of the LMS channel, including the speed at which the terminal is moving.

Appendix 5.A Accuracy of ESM in mobile satellite communications

In this section, we will try to assess how accurate the ESM is as a performance indicator in the LMS channel. To do so, we have simulated a simplified version of a satellite transmission chain that consists of a bit interleaver, a DVB-S2 LDPC codec (that allows up to fifty decoding iterations) and a 4-QAM mapper. The log-likelihood ratios for the decoder have been built as

$$\text{LLR}_i = 2\gamma_i, \tag{5.13}$$

so that we are implicitly assuming perfect knowledge of each symbol's SNR.

The channel has been simulated using Fontan's model, focusing on the so-called *intermediate*
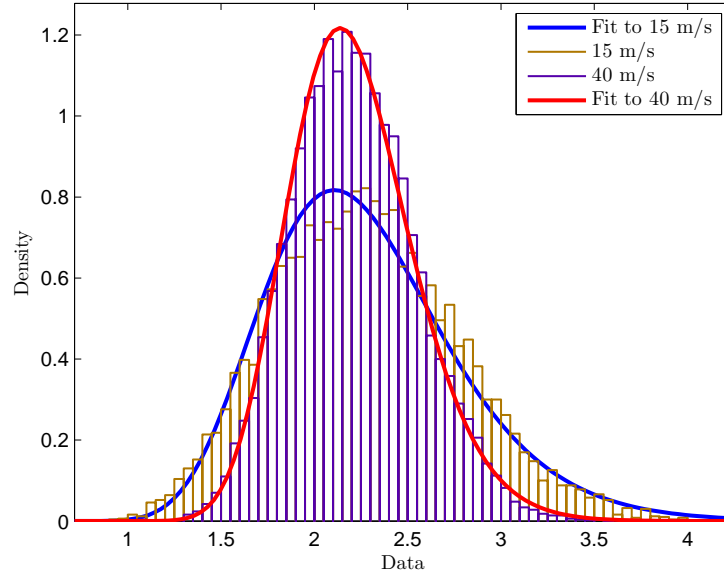
Figure 5.5: Fitting the EESM with by a log-normally distributed random variable.

tree shadowed area (ITS). The link is assumed to be GEO and operate at the S-band with $f_c = 2.2$ GHz and the symbol period is $T_{symb} = 2.7$ μs; the terminal has been assumed to move at a speed of $v = 10$ m/s (36 km/h).

The simulation procedure is the following: for each new codeword, we start by obtaining the corresponding channel samples; then, we apply a scaling so that the ESM of those samples would equal a certain target ESM. We repeat this process for a range of ESM values, and then plot the BER results in the same plot as those from the usual AWGN curve.

The obtained results can be seen on Figures 5.6 and 5.7 for code rates $\eta = 1/2$ and $\eta = 3/4$, respectively. Here, we can see that AWGN and ESM curves are almost indistinguishable. On the other hand, we have illustrated the BER as a function of the average SNR experienced by each codeword; results are not very tight for State 1 –the one with less influence of the non-line-of-sight component– and completely chaotic for the rest of the states.

Let us further investigate these two phenomena. Starting with the accuracy of the ESM as an indicator, Figures 5.8 and 5.9 offer magnified versions of the previous plots in the region of interest. There, we can see that the maximum difference with respect to the AWGN curve is of about 0.1 dB; note, however, that the difference is positive in one case and negative in the other, which suggests that each coding scheme may exhibit a different behavior.

On the other hand, Figure 5.10 shows the average SNR obtained after enforcing an effective SNR of 0.8 dB. We can see that the same ESM in different codewords has led to much different results in terms of average SNR, and, the greater the influence of the NLOS component, the greater the difference; this explains why the average SNR performs such a poor mapping in some of the states.
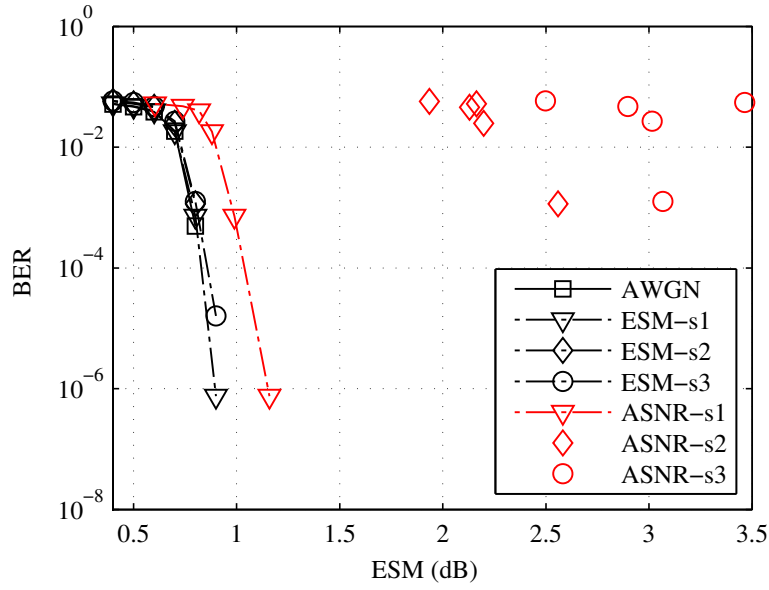
Figure 5.6: BER performance for a code rate $\eta = 1/2$, in the three states of the ITS area. Speed $v = 10$ m/s, $f_c = 2.2$ GHz, $T_{\text{symb}} = 2.7$ μs, 1,500 words of 64,800 bits per point.
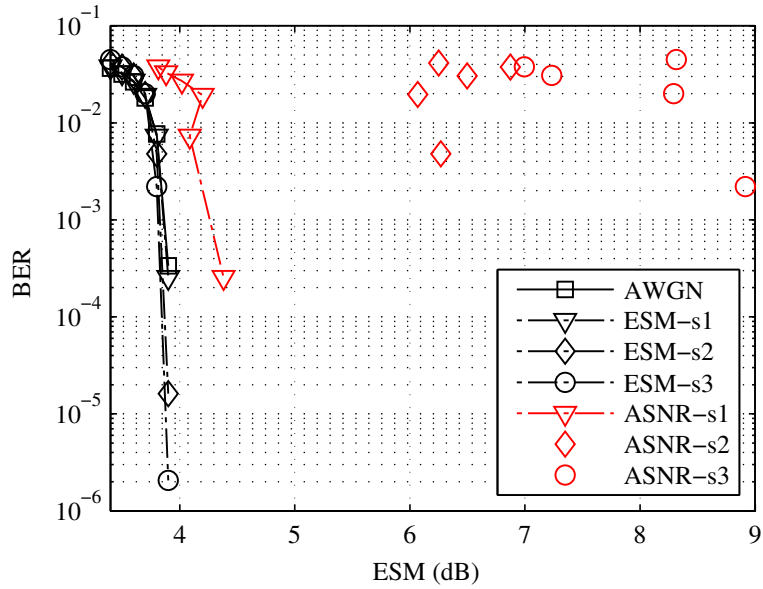
Figure 5.7: BER performance for a code rate $\eta = 3/4$, in the three states of the ITS area. Speed $v = 10$ m/s, $f_c = 2.2$ GHz, $T_{\text{symb}} = 2.7$ μs, 1,500 words of 64,800 bits per point.
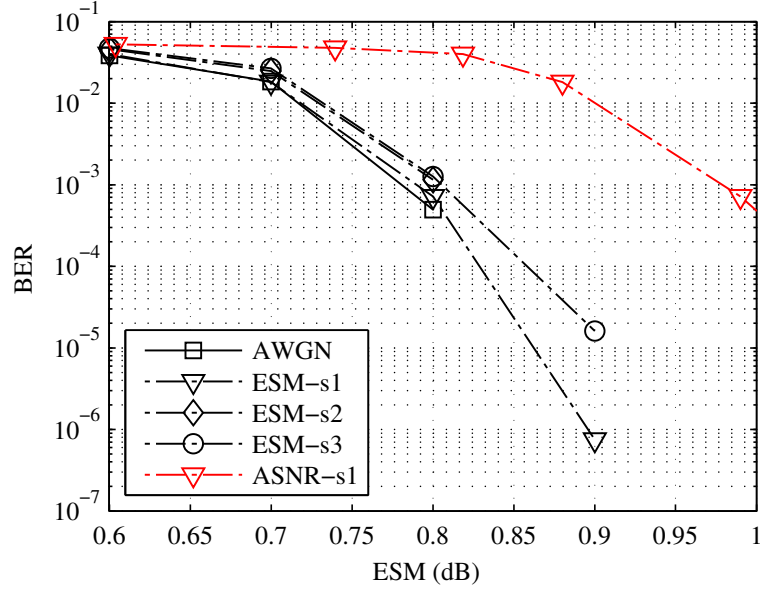
Figure 5.8: Detail of Figure 5.6, showing only the curves close to the AWGN benchmark. Speed $v = 10$ m/s, $f_c = 2.2$ GHz, $T_{symb} = 2.7$ μs, 1,500 words of 64,800 bits per point.
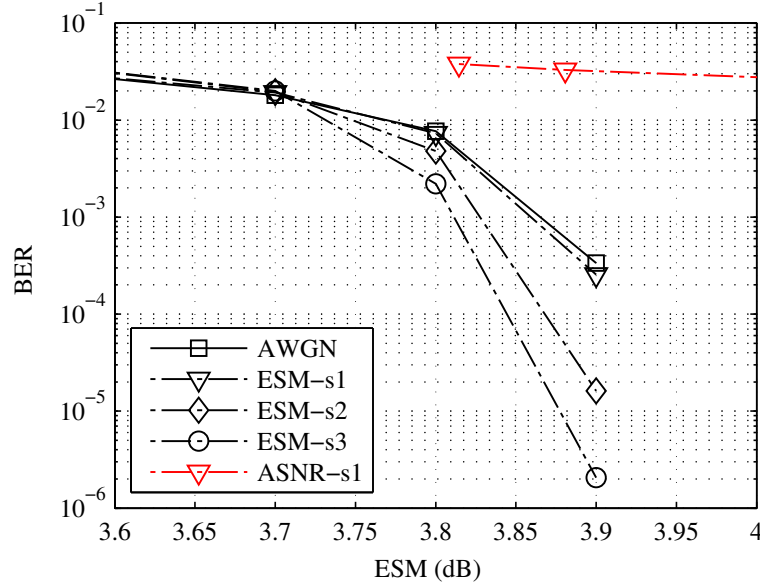
Figure 5.9: Detail of Figure 5.7, showing only the curves close to the AWGN benchmark. Speed $v = 10$ m/s, $f_c = 2.2$ GHz, $T_{symb} = 2.7$ μs, 1,500 words of 64,800 bits per point.

Figure 5.10: Histogram of the average SNR (ASNR) equivalent to $\gamma_{\text{eff}} = 0.8$ dB in the three states of the ITS area. Speed $v = 10$ m/s, $f_{\text{c}} = 2.2$ GHz, $T_{\text{symb}} = 2.7$ μs, 1,500 words of 64,800 bits per point.
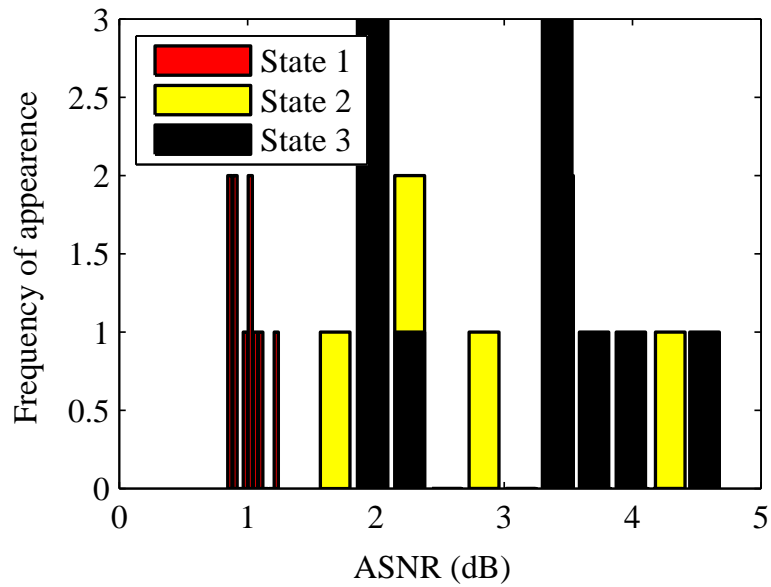
Chapter 6

Physical layer adaptation in the return link I: feasibility of open-loop CSI

Contents

6.1 Introduction

In this chapter we have tested the applicability of open-loop adaptation in the return link of a mobile satellite link using, for improved performance monitoring, effective SNR metrics instead of conventional average SINR as CSI.

We started by testing the applicability of open-loop CSI, simulating it in different environments and at different speeds. Results, given in terms of root mean squared error (RMS) of the difference between the estimated CSI and the actual one, showed apparently acceptable values for ACM. However, this study did not take into account the existence of very different interference levels in both links.

Then, we carried out a performance test focusing on specific working conditions. We used the simplest way to exploit the obtained CSI: the last CSI sample is used, after applying an appropriate margin, to select the most suitable MCS by comparing this value with its working threshold. We used three different interference profiles due to co-channel beams (no interference, mild interference and strong interference), obtained from the simulation of a realistic scenario, to test the design robustness. These simulations were carried out in STYLIST [112], a system simulator developed under an ESA Artes 5.1 activity.

Results have shown that, for the scenario under study, the best performance is obtained in an ITS environment, reaching an improvement of up to 92 % in terms of ASE, and 12 % more availability, even for the most compelling interference pattern simulated. On the other end, the suburban environment has been seen to be the least suitable: performance with open-loop adaptation is worse even with no interference.

The remainder of the chapter is structured as follows: Section 6.2 describes the system and signal model; Section 6.3 studies the properties of open-loop CSI; Section 6.4 illustrates the potential of open-loop techniques in terms of availability and spectral efficiency; finally, Section 6.5 summarizes the conclusions of the chapter.

6.2 System model

We focus on the return link of a BGAN-like [113] mobile satellite link operating at L-band; forward and return links are allocated to different frequencies (1550 MHz for the forward link and 16650 MHz for the return link). We assume the existence of pilots, scattered through the payload, that allow accurate channel estimation. The remainder of the section describes the signal model and offers a brief explanation of physical layer abstraction. The signal model is given by

$$y_k = \sqrt{\text{snr}} \cdot h_k s_k + w_k \tag{6.1}$$

as explained in Section 5.2, with y_k the symbol received at the k-th time instant, s_k the transmitted symbol, h_k the channel coefficient and snr representing the transmitted over noise power.
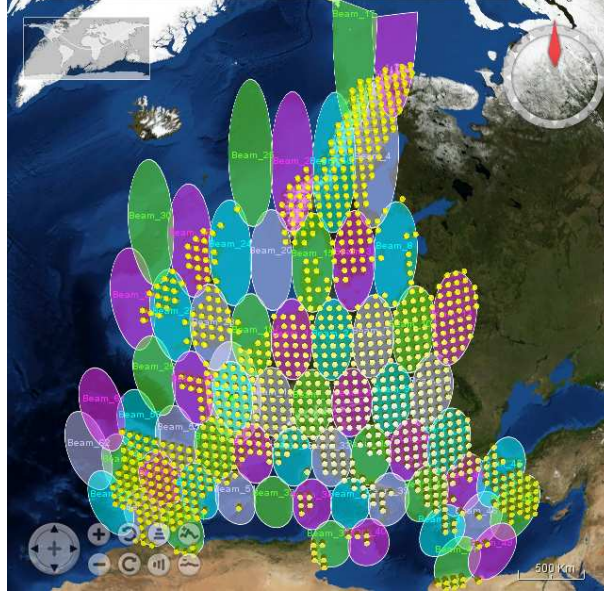
Figure 6.1: Simulated coverage over Europe.

The sample w_k comprises the effect of noise and interference, so that we can write

$$w_k = n_k + i_k \tag{6.2}$$

where n_k is a unit-power noise sample (we are already accounting for the noise power in $\sqrt{\text{snr}}$) and i_k is the interference sample.

After simulating a 4-colors, 62-beam coverage over Europe in STYLIST [112] (see Figure 6.1), we obtained two return link interference profiles, as shown on Figure 6.2 and Figure 6.3; these profiles were later on used for the simulations testing the designed ACM techniques.

6.3 Suitability of open-loop CSI

As we said, the channels experienced by FL and RL will be partially uncorrelated. The common assumption is that *the LOS component will be the same, while the NLOS will be completely uncorrelated* (but with the same statistical parameters whenever the frequency separation is not very high). Therefore, before designing an ACM strategy that relies on open-loop CSI, we will need to test how accurate it is, taking into account this partial uncorrelation among the channels.

However, we will show that forward-return uncorrelation is not the only source of impairments: the channel decorrelates after some time –which depends on terminal's speed–, and we need to make use of the obtained CSI before it runs completely stale.

To test the suitability of open-loop CSI, we performed simulations aimed at assessing the impact of the two impairments mentioned above. In this section, we describe the procedure we
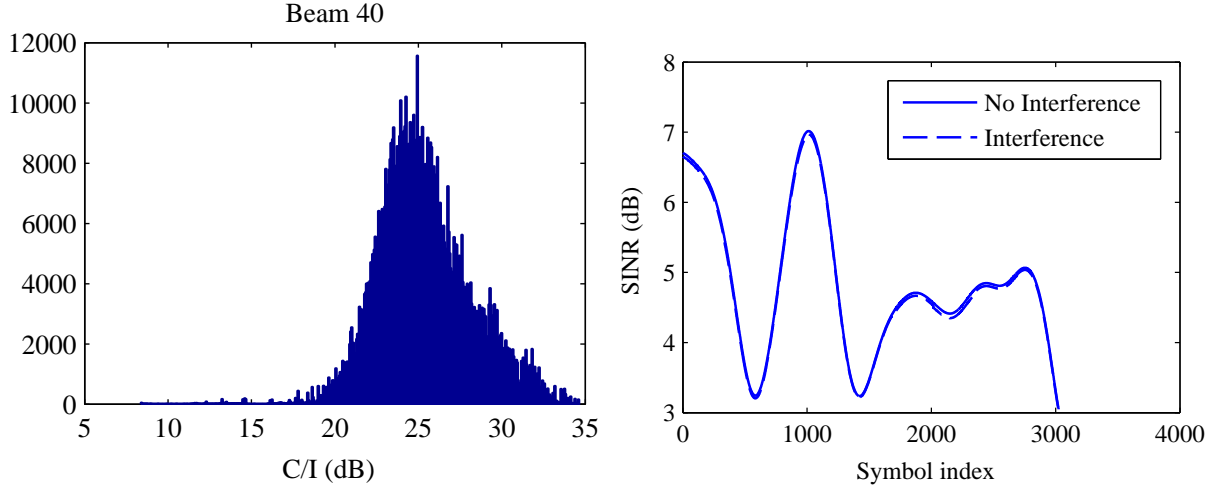
Figure 6.2: Mild interference profile, named SI40. Left is a histogram of the simulated C/I, and right an example of the SINR evolution in an LMS channel including this profile.
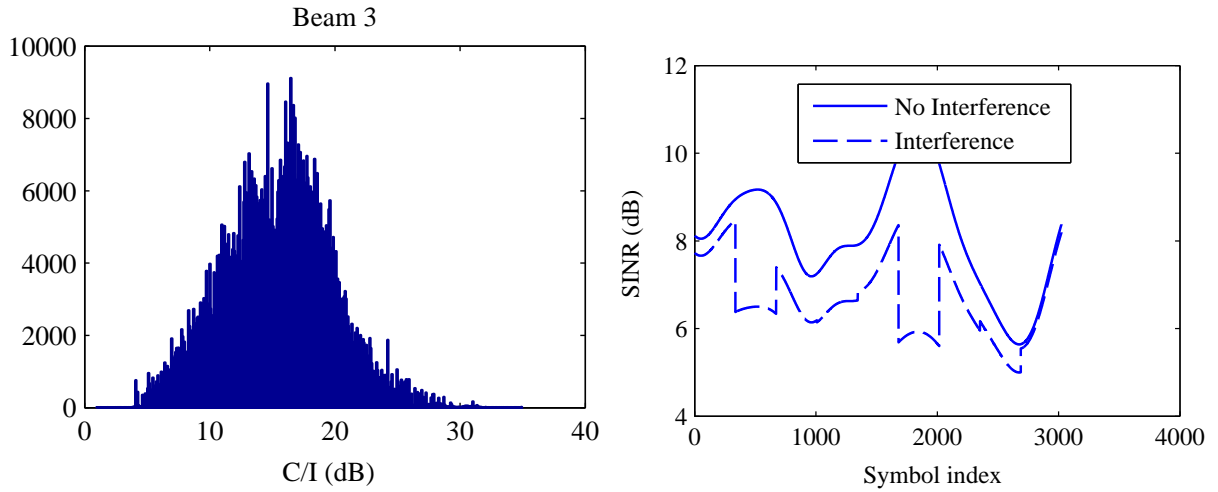
Figure 6.3: Strong interference profile, named SI3. Left is a histogram of the simulated C/I, and right an example of the SINR evolution in an LMS channel including this profile.

followed, and report the results obtained.

6.3.1 BGAN RL timing

After measuring the signal from the FL, we intend to use it for adapting the RL parameters as soon as possible; the more we wait, the less useful the CSI will be. Unfortunately, compliance with the BGAN standard will enforce a minimum waiting time, given by [113]

$$n \cdot 80 + \text{SID} \quad \text{ms} \tag{6.3}$$

after reception of the first symbol from the FL, where n equals 1 when interleaving is done every 10 ms or 20 ms and n equals 2 when it is done every 80 ms; on the other hand, SID = (BeamMaxDelay − Delay), which is usually low and therefore we will neglect it in the sequel. See Figure 6.4 for a graphical explanation from the standard.
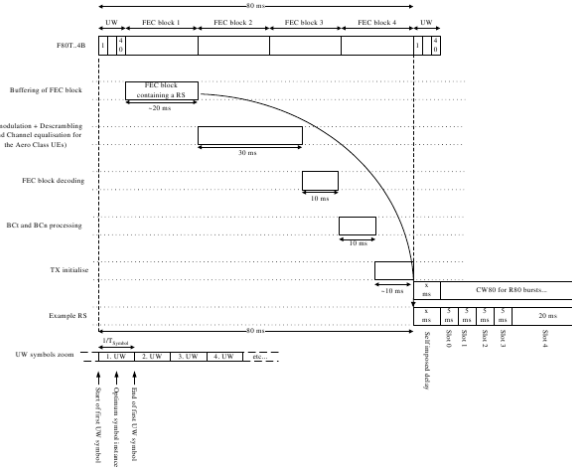
Figure 6.4: Timing requirements for the RL of a BGAN system.

6.3.2 Simulator description

To test the validity of open-loop CSI, we performed several simulations. The procedure, summarized in Figure 6.5, is the following:

1. The parameters are $f_{\text{RL}} = 1550\,\text{MHz}$, $f_{\text{FL}} = 1650\,\text{MHz}$ and $T_{\text{symb}} = 1/33600\,\text{s}$.

2. We generate FL and RL channels with the same LOS but uncorrelated NLOS.

3. From the channel samples we obtain a sequence of instantaneous SINR (γ_{FL} and γ_{RL}), and apply interleaving on them.

4. We compute the effective SNR mapping (ESM) as detailed in Section 5.3.
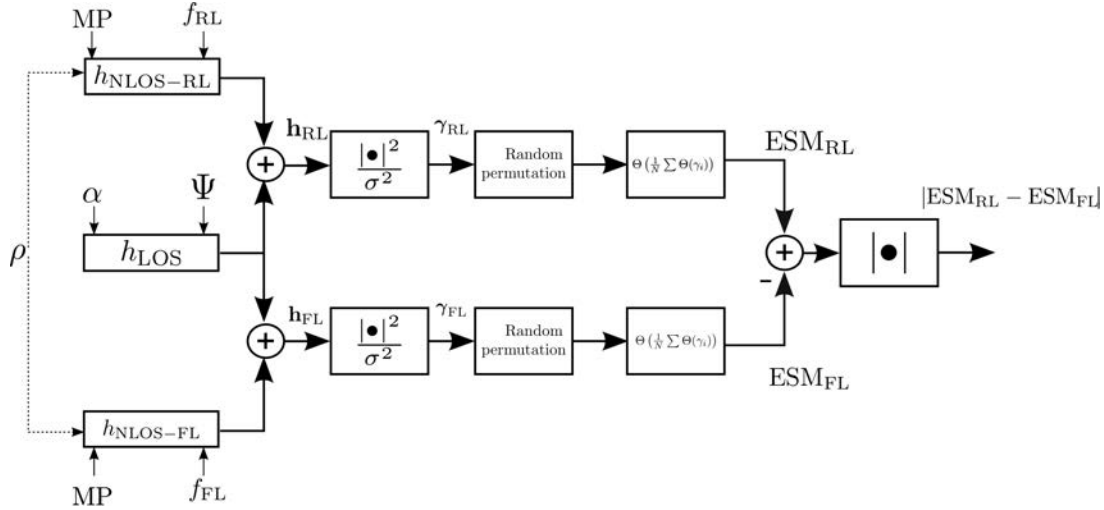
Figure 6.5: Simulation setup for open-loop CSI validation.

5. The ESM values, in dB, are compared, and the absolute value of the difference is the output of the simulator.

Also, we added the possibility of generating partially correlated NLOS components by changing the correlation factor ρ. Although it is customary to assume totally uncorrelated NLOS, this aided in assessing which are the main causes of difference between open-loop and closed-loop CSI.

Additionally, before comparing the values in dB, we may apply a *shift* in order to account for the $n \cdot 80\,\mathrm{ms}$ delay introduced by the standard.

6.3.3   Results

We will now report the most relevant simulation results obtained; additional supporting results can be found in Appendix 6.A. We focus on a scheme with four codewords per frame, which means that the codeword duration is $20\,\mathrm{ms}$; at this stage, however, we always perform interleaving on an $80\,\mathrm{ms}$ basis, for simplicity. All the simulations in this section have emulated an Intermediate Tree Shadowed (ITS) area.

We will plot the root mean squared error (RMS) of the ESM difference $\mathbf{d}$, defined as

$$\mathrm{RMS} \doteq \sqrt{\frac{1}{N}\sum_{i=1}^{N} d_i^2}, \tag{6.4}$$

as a function of the terminal speed.

Figure 6.6: RMS of the ESM difference as a function of speed, $n = 1$, snr = 1; the instantaneous comparison is shown in black, the delayed comparison in red. State by state evolution of the channel on the left, joint time series on the right.
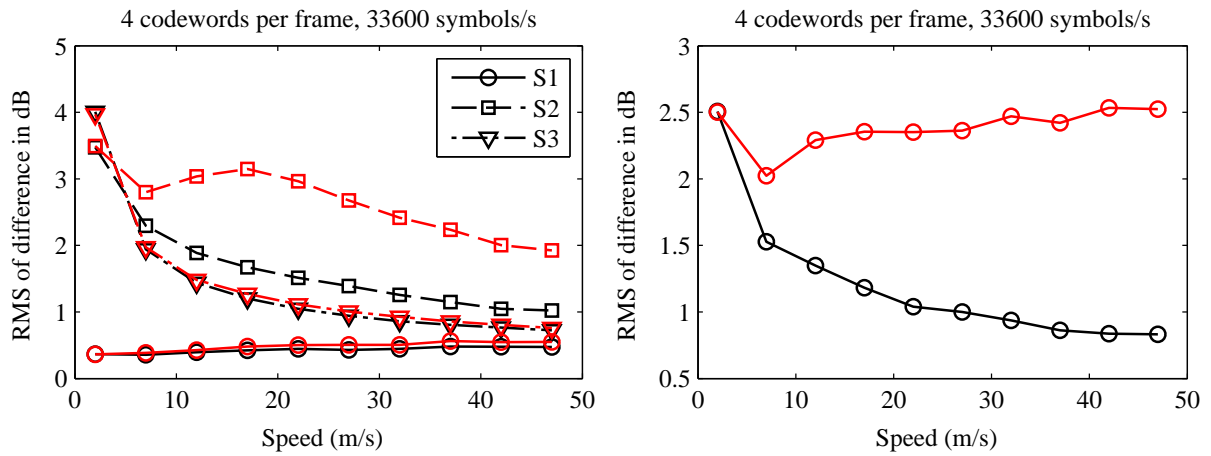
Figure 6.7: RMS of the ESM difference as a function of speed, $n = 1$, snr = 100; the instantaneous comparison is shown in black, the delayed comparison in red. State by state evolution of the channel on the left, joint time series on the right.
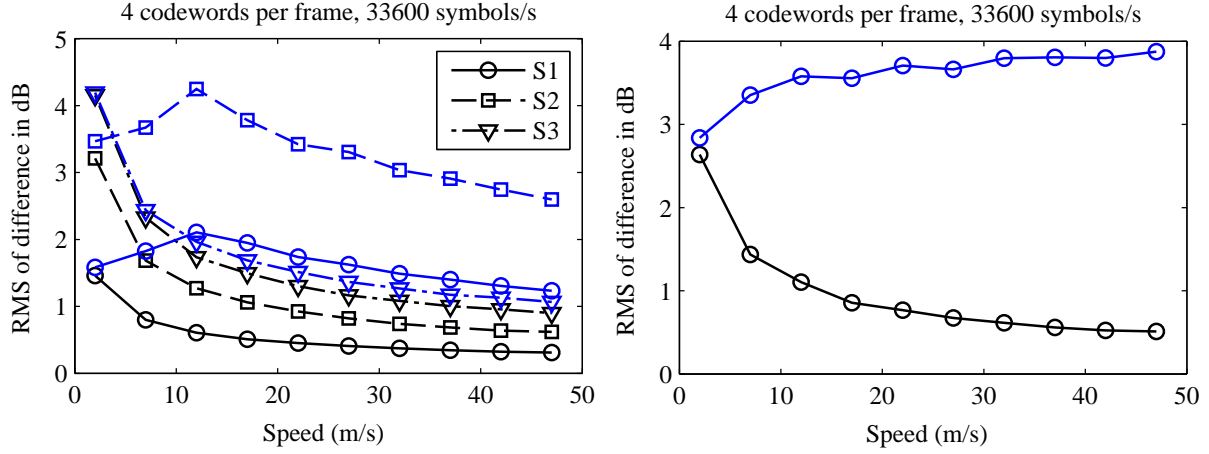
Figure 6.8: RMS of the ESM difference as a function of speed, $n = 2$, snr $= 1$. The instantaneous comparison is shown in black, the delayed comparison in blue. State by state evolution of the channel on the left, joint time series on the right.

Results with an $80\,$ms delay

Figure 6.6 and Figure 6.7 show the comparison result when $n = 1$ –that is, when there is a delay of $80\,$ms before transmitting– for snr $= 1$ and snr $= 100$, respectively.

We can see that differences decrease with speed for the instantaneous comparison (black curves), and that they are larger in the states with larger NLOS power. In what refers to the delayed comparison, it is worth noticing the huge differences induced by state 2: they are due to its fast variations in the LOS component, that render a very different channel after waiting for $80\,$ms.

Results with a $160\,$ms delay

Figure 6.8 and Figure 6.9 show the comparison result when $n = 2$ –that is, when there is a delay of $160\,$ms before transmitting– for snr $= 1$ and snr $= 100$. The same conclusions as for the previous plots hold; only that, as expected, differences in the delayed case are now larger.

Comparison between the two delays

Figure 6.10 summarizes the results above into a single box plot. For each delay, the line within a rectangle box shows the median difference; the upper side of the rectangle shows the $75\,\%$ percentile, and the lower the $25\,\%$ percentile. The remaining points are contained within the whiskers, and those considered outliers are marked separately as dots.

What these results highlight is how large instantaneous differences can be, irrespectively of the RMS obtained; as an example, note that differences of up to $16\,$dB have been recorded at medium speeds. This suggests that a statistical characterization of the difference would be
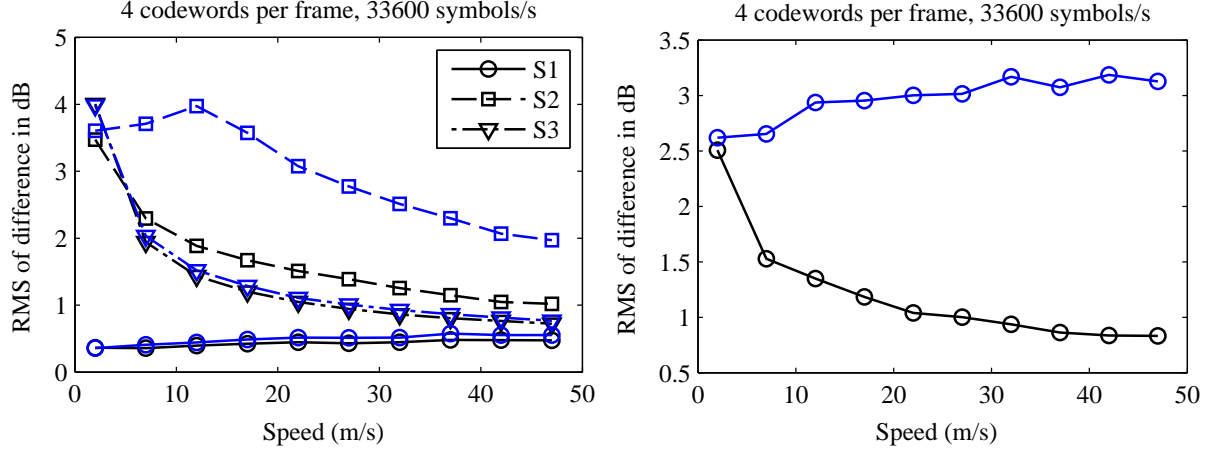
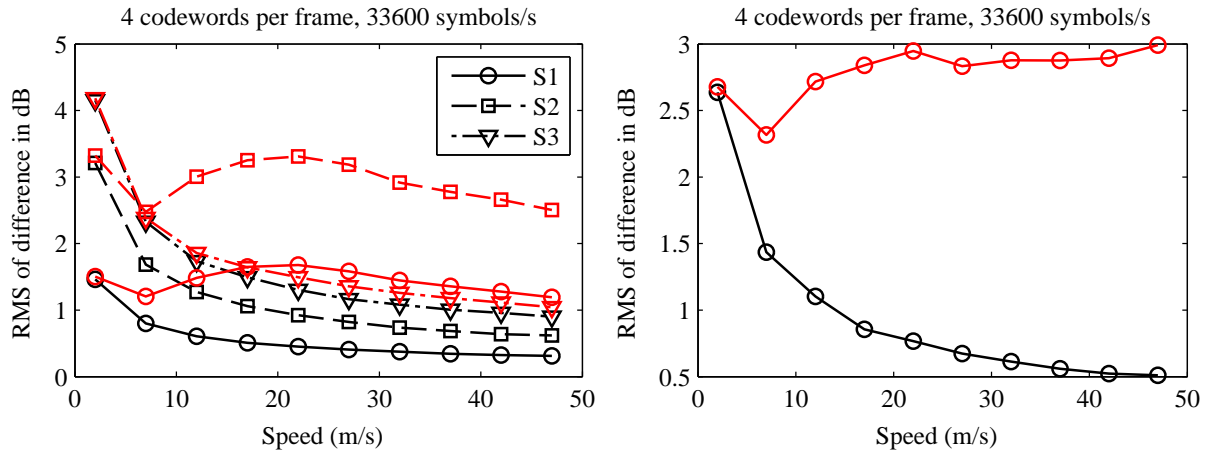Figure 6.9: RMS of the ESM difference as a function of speed, $n = 2$, snr = 100. The instantaneous comparison is shown in black, the delayed comparison in blue. State by state evolution of the channel on the left, joint time series on the right.
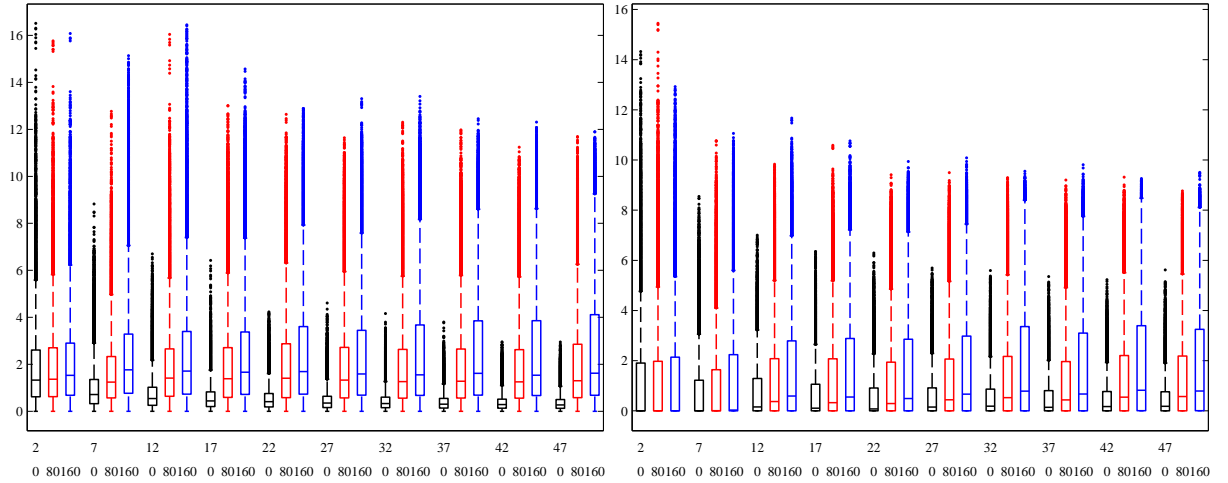

Figure 6.10: Box plot of the ESM differences with no delay and with 80 ms and 160 ms delay, snr = 1 (left) and snr = 100 (right).

necessary for the best exploitation of the channel's capacity. In this work, however, we have disregarded this possibility and focused on a fixed margin countermeasure.

6.3.4 Summary

We have studied the suitability of the CSI obtained from direct measurements on the FL signal. This is a summary of the achieved conclusions.

- Instantaneous comparison.
    - Open-loop CSI is more accurate as the speed increases, and moderately inaccurate for low speeds.
    - Simulations show that the power of the NLOS component is the main driver of inaccuracies.
- Delayed comparison.
    - Now open-loop CSI can still be imprecise at low speeds, depending on the power regime and on the particular state of the Fontan channel.
    - This is due to the differences in the NLOS component, which gain importance because of the delay introduced.

In view of the numerical results, open-loop CSI will be applicable for many speeds, as the most frequent differences are quite low. However, this does not mean that it will outperform traditional closed-loop CSI in each and every case; in particular, note that here we are not taking into account the different interference levels in both links (nor the imperfect estimation of the forward link sample). In the next section, we will illustrate the performance of an ACM link using this kind of open-loop CSI.

6.4 Performance of open-loop ACM

In the previous section, we tested the applicability of open-loop CSI, concluding that its accuracy seems to be enough to be applied in practice. We still have to show which performance does it provide in terms of throughput and availability of the link, and whether it brings any advantage over traditional closed-loop CSI or not. In this section, we report the results of the simulations we performed, aimed at comparing a BGAN-type system with one based on open-loop CSI.

6.4.1 Simulator description

Even though the suitability test in Section 6.3 covered a wide range of cases, here we will focus on a much more specific scenario, as described in Table 6.1. Note that, in assuming a fixed

fading bandwidth, we are implicitly assuming that the terminal speed is fixed (and, in this case, equal to 4 m/s). Also, we focused on codewords spanning 20 ms, since this is the most common case for Demand Assigned Multiple Access (DAMA).

Terminal type	Pocket	T_{symb}	$1/16800$ s
Environment	ITS	Fading bandwidth	20 Hz
C/N_0	50.9 dBHz	f_{FL}	1550 MHz
Bearer bandwidth	21 KHz	f_{RL}	1650 MHz

Table 6.1: Link budget and other parameters used during the simulations.

Figure 6.11 summarizes the simulation procedure, which we describe in the following items. For each case, we simulated four totally independent channel realizations, with 10^5 codewords in each one.

Channel generation

We start by generating both channels with the parameters from Table 6.1; this returns the channel coefficients h_k. We compute the SNR as SNR $= 50.9 - 10\log_{10}(21000) = 7.7$ dB, and then use it in natural units, snr $= 10^{\text{SNR}/10}$.

Adding interference

With the above, we would have the sequence of instantaneous SNR, given by snr$|h_k|^2$. But what we will need for the simulations is actually the sequence of SINR. To obtain it, we used STYLIST and simulated a 62-beam European coverage with the same link budget and parameters as reported above. As an output of the simulation we obtained the sequence of C/I values, which we post-processed to obtain the empirical cumulative distribution function (CDF) of the interference for two different beams, as already shown in Section 6.2. Dragging *one sample per codeword* from the empirical CDF, and using the fact that

$$\left(\frac{C}{N+I}\right)^{-1} = \left(\frac{C}{N}\right)^{-1} + \left(\frac{C}{I}\right)^{-1} \tag{6.5}$$

we could link both outputs and obtain the SINR values.

Computing CSI values

To compare our system with a BGAN-like one, we need to obtain both types of CSI and then select the coding scheme based on it.
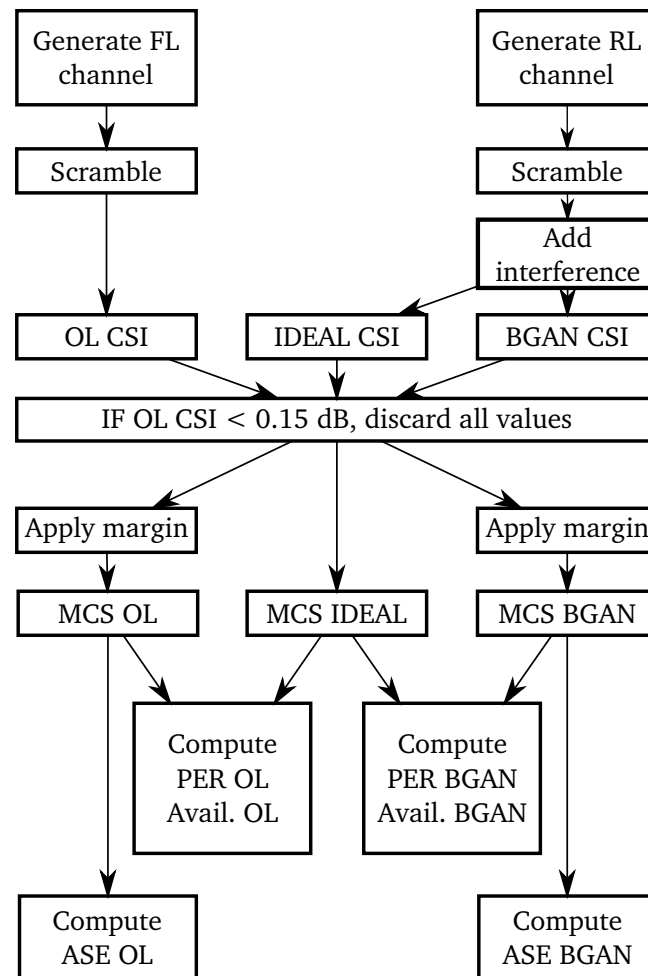
Figure 6.11: ACM simulation description.

	L8	L7	L6	L5	L4	L3	L2	L1	R
γ_{th} (dB)	0.15	0.91	1.87	2.79	3.90	5.07	5.71	6.48	7.67
Coding rate	0.34	0.39	0.46	0.53	0.61	0.69	0.73	0.77	0.81
Info bits	168	200	240	280	328	376	400	424	448

Table 6.2: Coding rate options for the R20T0.5Q-1B bearer.

- **BGAN CSI:** We obtained it by averaging the instantaneous SINR of the return link (closed-loop) during 0.96 s; we then wait for 0.5 s (an RTT) before using it.

- **Open-loop CSI:** This is the last ESM sample computed on the forward link signal, with a delay of 80 ms (since we are assuming 20 ms codewords).

- **Ideal CSI:** This is the name we give to the *actual, totally accurate* CSI of the channel obtained with no delay.

Transmission control protocol

We implement a transmission control policy so that, when the measured quality of the channel is very low, we do nothing. This was done by applying a threshold to the open-loop CSI: if it was below 0.15 dB, the threshold for the most protected coding scheme of the bearer in use, we did not test anything.

Transmitted symbols

In this work, we focused on the bearer R20T0.5Q-1B, which uses only QPSK modulation; the coding schemes available for this bearer can be seen on Table 6.2, along with some details on their efficiency.

MCS selection

From a given CSI value, the corresponding coding rate is selected by applying a threshold. In this work, thresholds have been selected by trial and error to meet a PER constraint of 10^{-3}, using the outputs of Section 6.3 as guidance. If the value obtained after applying the margin is below the threshold of the most protected MCS, we say to select MCS = 0, which means that we do not transmit at all.

Computation of the results

The results of the simulation are offered in terms of average spectral efficiency (ASE), availability and packet error rate (PER), which we define as follows.

	No interference		Beam 40		Beam 3	
	BGAN	Open-loop	BGAN	Open-loop	BGAN	Open-loop
Open	2.15	1.8	2.15	1.9	3.95	4.8
ITS	7.7	6.8	7.6	6.8	6.7	7.1
Suburban	5.9	7				

Table 6.3: Margins used for MCS selection, in dB. Obtained by trial an error to ensure PER = 10^{-3}.

- **ASE:** Average spectral efficiency of the coding rate selected (recall that we modulation is always QPSK).

- **PER:** Let MCS_k^0 be the sequence of coding schemes we would have selected using perfect, ideal CSI of the channel, and let $\mathcal{I}[P]$ be the indicator function, which takes the value 1 when P is true and 0 otherwise; for a sequence of coding schemes $\{\mathrm{MCS}_k\}_{k=1}^K$, we compute the PER as

$$\mathrm{PER} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{I}\left[\mathrm{MCS}_k > \mathrm{MCS}_k^0\right] \tag{6.6}$$

- **Availability:** We consider that a link is available whenever the selected MCS is different from zero, that is

$$\mathrm{Availability} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{I}\left[\mathrm{MCS} > 0\right] \tag{6.7}$$

Also, we call it *normalized availability* when it is normalized by the availability obtained with ideal CSI, which is the maximum performance we could obtain.

6.4.2   Performance results

Here, we will present the results obtained for three different environments: ITS, open and suburban. Using three different interference profiles, we will illustrate the potential of open-loop adaptation. Although the target scenario is ITS, we simulated the other two in order to test how well does the designed system perform in other conditions.

Best performance: ITS

Figure 6.12 depicts a comparison in terms of ASE and availability between the two techniques considered for an ITS environment. The gain in performance is remarkable: close to 92 % in terms of ASE, and 12 % more availability even for the most compelling interference pattern (beam 3).
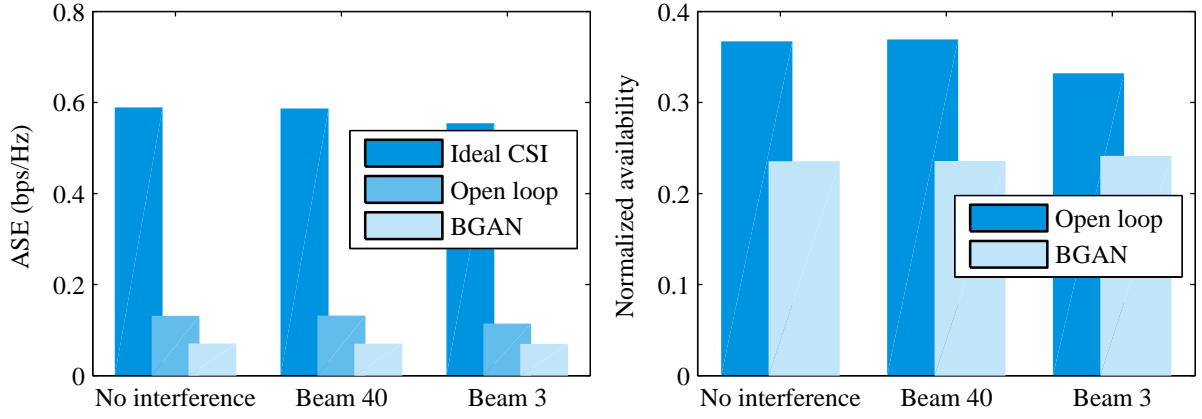
Figure 6.12: ASE (left) and availability (right) comparison for an ITS environment.
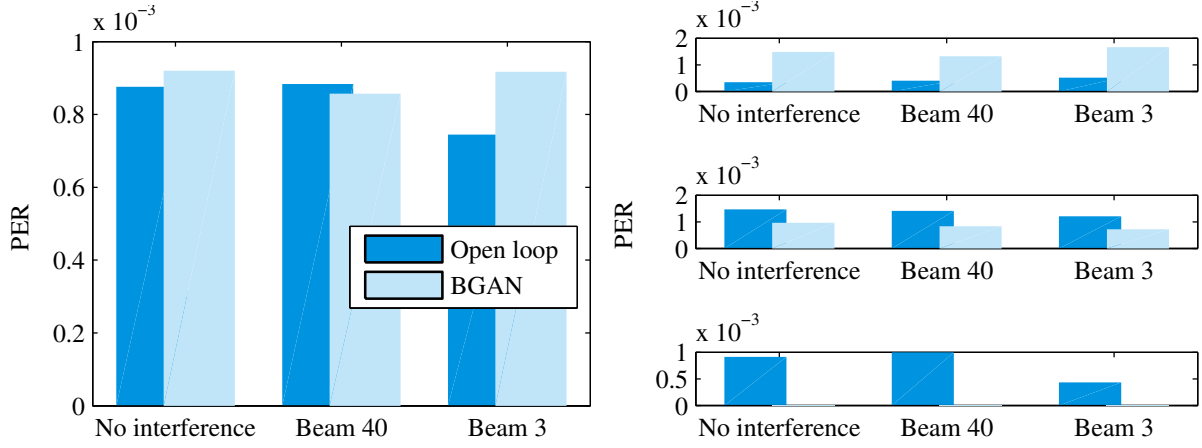
Figure 6.13: PER (left) and state-by-state PER (right) comparison for an ITS environment.

Note that these gains come with no loss in terms of PER: Figure 6.13 proves this fact, showing that a BGAN-like system can be allowed a slightly higher PER without reaching the same ASE as the open-loop alternative. Figure 6.14 shows the same performance results on a state-by state basis.

For a better understanding of the results, Figure 6.15 shows the percentage of usage of each coding scheme with the strongest interference distribution. We can see that, while BGAN uses the most protected mode 72 % of the time, an open-loop strategy explores up to L4 without losing in terms of PER.

Intermediate performance: open environment

Figures 6.16-6.18 show the same results as above, but now for an open environment. As we can see, performance gains here are much smaller, if existent. In fact, with the higher interference levels that beam 30 provides, open-loop performs slightly below BGAN.

The explanation in this case is easy: an open environment is easy to track, even using aver-
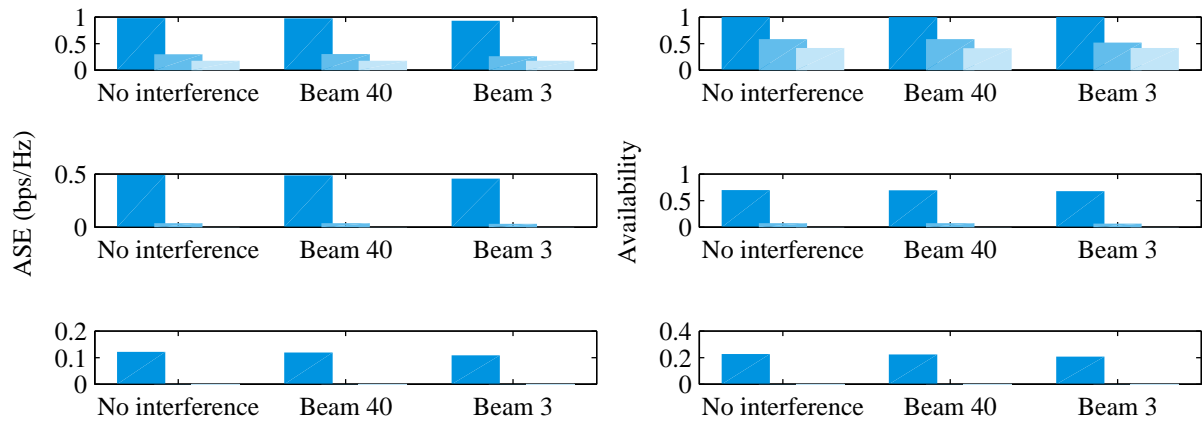
Figure 6.14: State-by-state ASE (left) and state-by-state availability (right) comparison for an ITS environment.
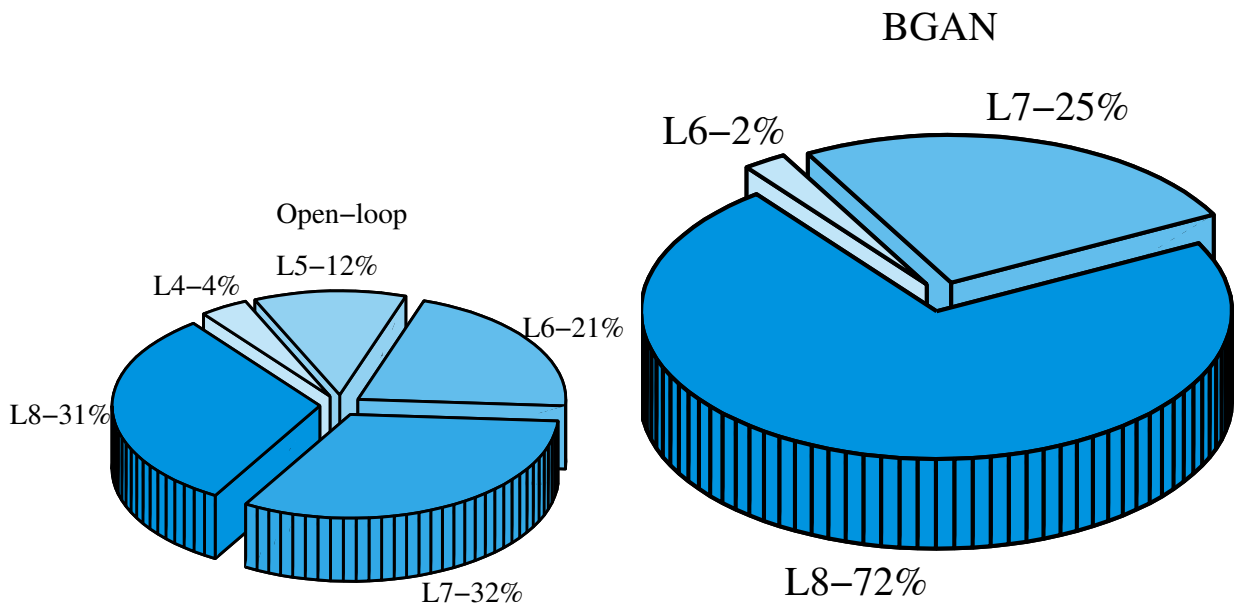
Figure 6.15: Percentage of usage of each MCS for open-loop (left) and BGAN (right), ITS environment, beam 3.
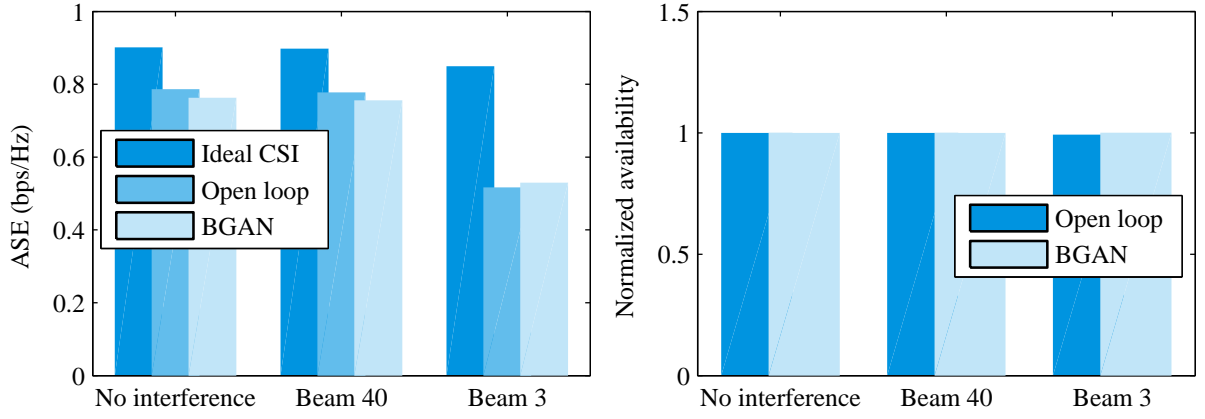
Figure 6.16: ASE (left) and availability (right) comparison for an open environment.
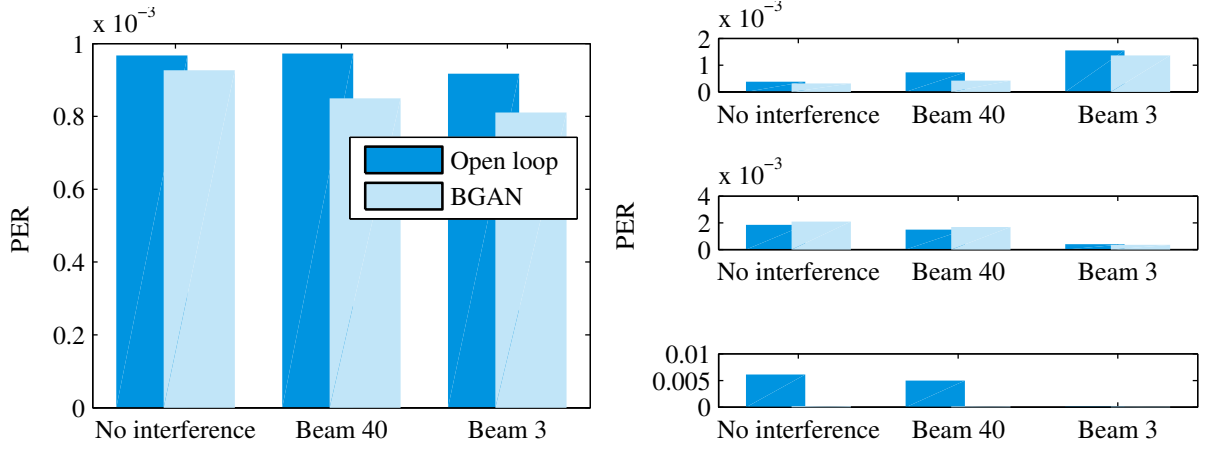
Figure 6.17: PER (left) and state-by-state PER (right) comparison for an open environment.

age SNR and affording long delays. On the other hand, open-loop is unaware of the interference level, which must be counteracted by enlarging the MCS selection margins. As a result, there is no advantage in using it, and it can even perform worse.

Finally, Figure 6.19 shows the MCS distribution, again for the case with the strongest interference.

Worst performance: suburban environment

Following the rationale from the analysis of the open environment, it was expected that open-loop would offer disadvantages in some environments. Figure 6.20 proves this fact, illustrating performance in a suburban environment. In this case, open-loop performance is worse *even without interference.*
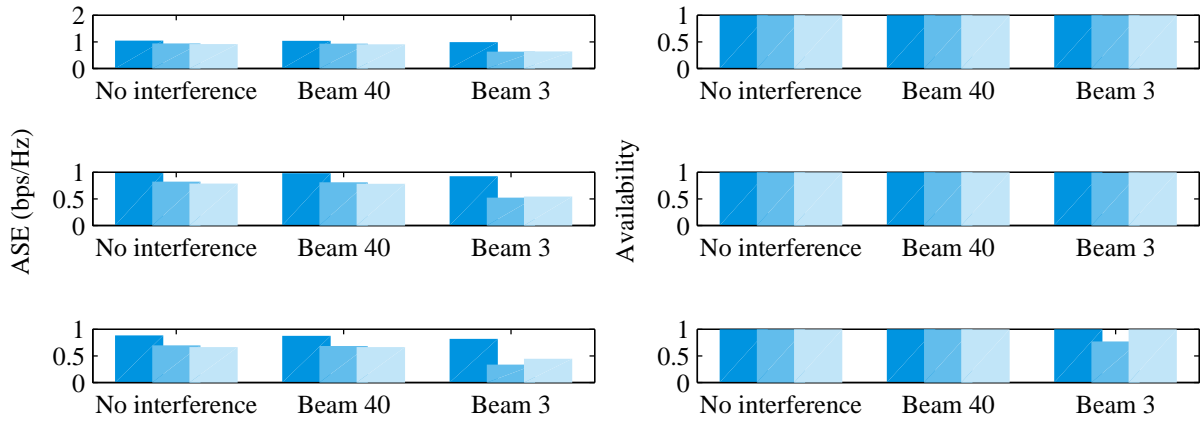
Figure 6.18: State-by-state ASE (left) and state-by-state availability (right) comparison for an open environment.
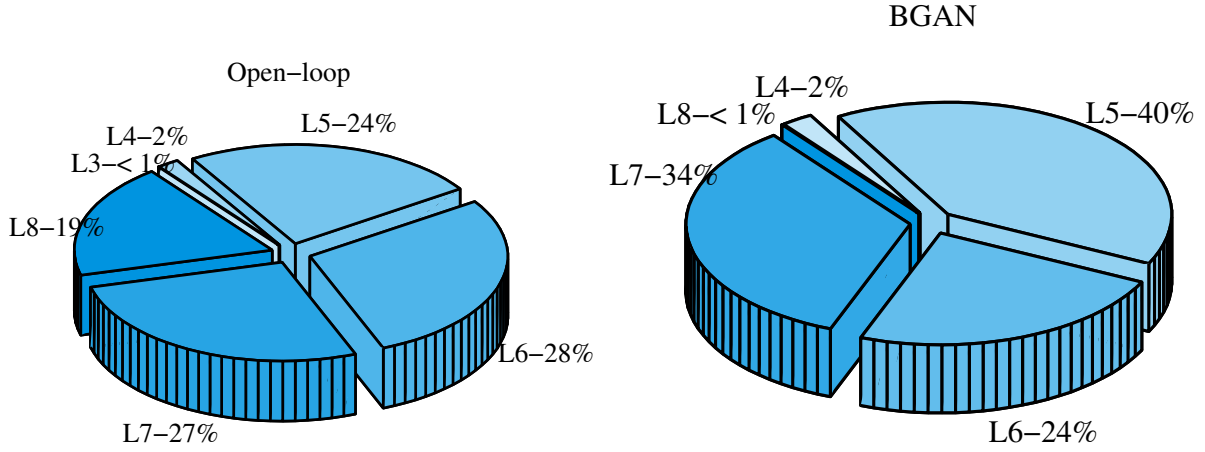
Figure 6.19: Percentage of usage of each MCS for open-loop (left) and BGAN (right), open environment, beam 3.
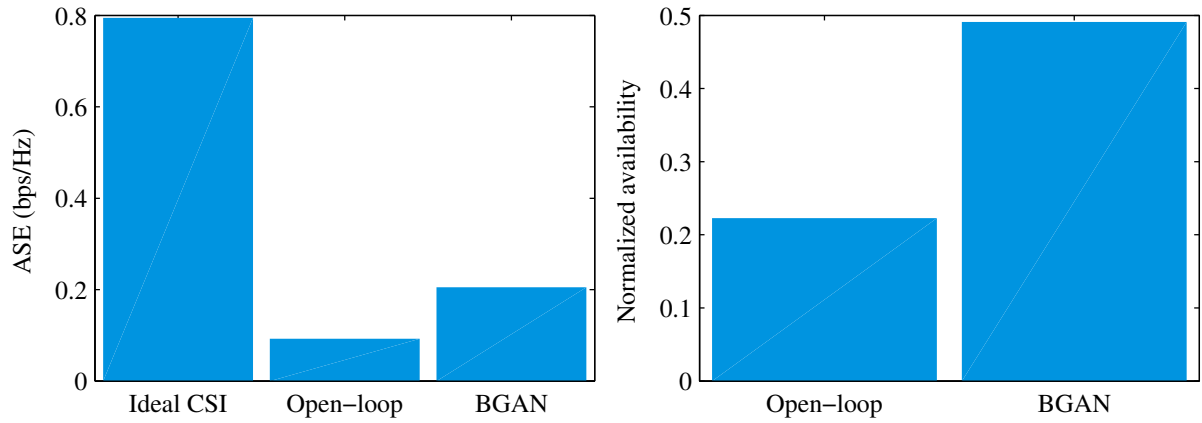
Figure 6.20: ASE (left) and availability (right) comparison for suburban environment.

6.5 Conclusions and further work

6.5.1 Conclusions

In this chapter, we have tested the applicability of open-loop ACM in the return link of a mobile satellite link using, for improved performance monitoring, effective SNR metrics instead of conventional average SINR as CSI.

After a careful study of the two CSI alternatives, we carried out a performance test focusing on specific working conditions. We used the simplest way to exploit the obtained CSI: the last CSI sample is used, after applying a margin, to select the most suitable MCS by comparing this value with its working threshold.

As summarized on Table 6.4, results have shown that, for the scenario under study, the best performance is obtained in an ITS environment, reaching up to 92 % in terms of ASE, and 12 % more availability, even for the most compelling return link interference pattern simulated. On the other end, the suburban environment has been seen to be the least suitable: performance with open-loop adaptation is worse even with no interference.

	No interference		Beam 40		Beam 3	
	BGAN	Open-loop	BGAN	Open-loop	BGAN	Open-loop
Open	0.7627	0.7862	0.7558	0.7776	0.5298	0.5166
ITS	0.0702	0.1314	0.0701	0.1319	0.0694	0.1141
Suburban	0.2050	0.0925				

Table 6.4: Summary of the ASE (bps/Hz) results obtained.

This chapter's contents were done in collaboration with Dr. Pantelis-Daniel Arapoglou, and partially presented at the ICSSC 2013 conference [99].

Appendix 6.A Additional figures on open-loop CSI

Here we will report additional simulation results used for assessing the validity of open-loop CSI.

6.A.1 ITS, 1 codeword per frame

Results with an $80\,\text{ms}$ **delay**

For one codeword per frame, Figure 6.21 and Figure 6.22 show the comparison result when $n = 1$ –that is, when there is a delay of 80 ms before transmitting– for snr = 1 and snr = 100, respectively.

Figure 6.21: RMS of the ESM difference as a function of speed, $n = 1$, snr $= 1$; the instantaneous comparison is shown in black, the delayed comparison in red. State by state evolution of the channel on the left, joint time series on the right.

Results with a $160\,\text{ms}$ delay

Figure 6.23 and Figure 6.24 show the comparison result when $n = 2$ –that is, when there is a delay of $160\,\text{ms}$ before transmitting– for snr $= 1$ and snr $= 100$.

Comparison between the two delays

The box plot comparison between both delays is shown on Figure 6.25.

6.A.2   ITS, 16 codeword per frame

Results with an $80\,\text{ms}$ delay

For one codeword per frame, Figure 6.26 and Figure 6.27 show the comparison result when $n = 1$ –that is, when there is a delay of $80\,\text{ms}$ before transmitting– for snr $= 1$ and snr $= 100$, respectively.

Results with a $160\,\text{ms}$ delay

Figure 6.28 and Figure 6.29 show the comparison result when $n = 2$ –that is, when there is a delay of $160\,\text{ms}$ before transmitting– for snr $= 1$ and snr $= 100$.

Comparison between the two delays

The box plot comparison between both delays is shown on Figure 6.30.
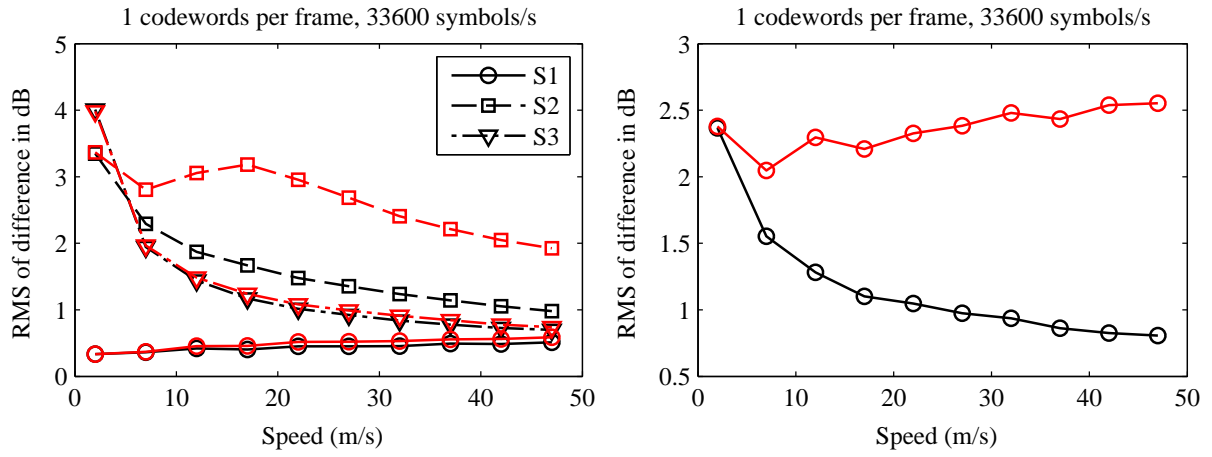
Figure 6.22: RMS of the ESM difference as a function of speed, $n = 1$, snr $= 100$; the instantaneous comparison is shown in black, the delayed comparison in red. State by state evolution of the channel on the left, joint time series on the right.

6.A.3 Suburban, 1 codeword per frame

See Figure 6.31.

6.A.4 Suburban, 4 codeword per frame

See Figure 6.32.

6.A.5 Suburban, 16 codeword per frame

See Figure 6.33.
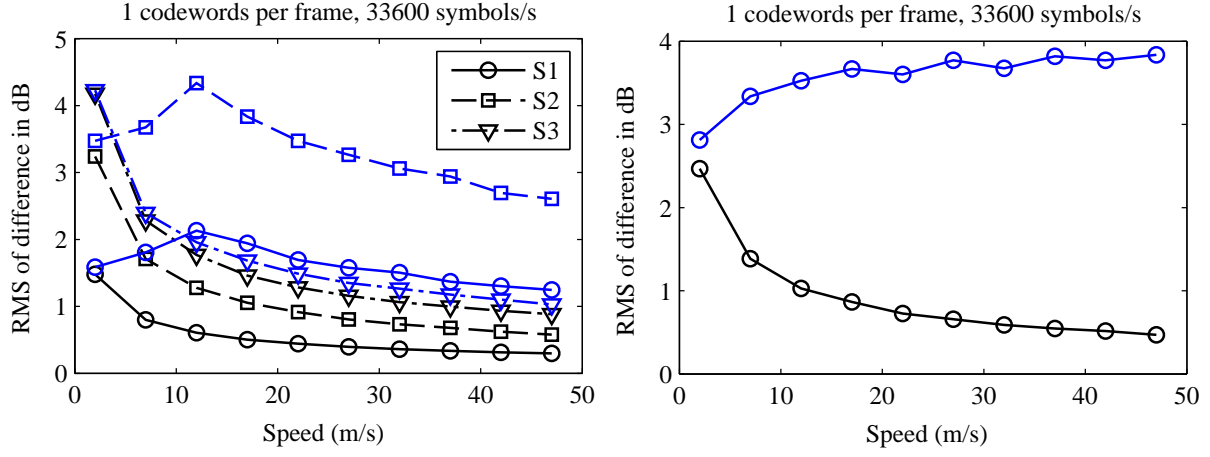
Figure 6.23: RMS of the ESM difference as a function of speed, $n = 2$, snr $= 1$. The instantaneous comparison is shown in black, the delayed comparison in blue. State by state evolution of the channel on the left, joint time series on the right.
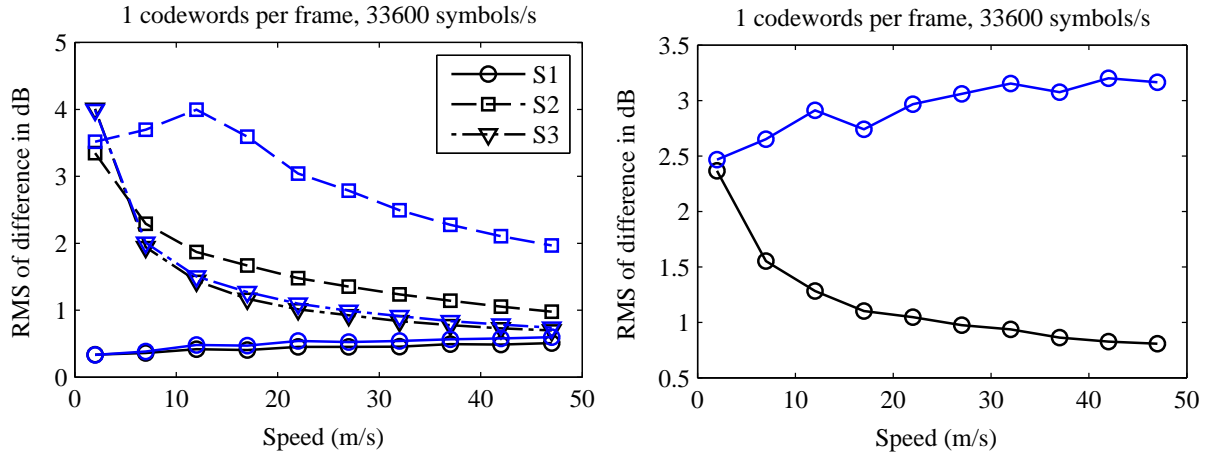
Figure 6.24: RMS of the ESM difference as a function of speed, $n = 2$, snr $= 100$. The instantaneous comparison is shown in black, the delayed comparison in blue. State by state evolution of the channel on the left, joint time series on the right.
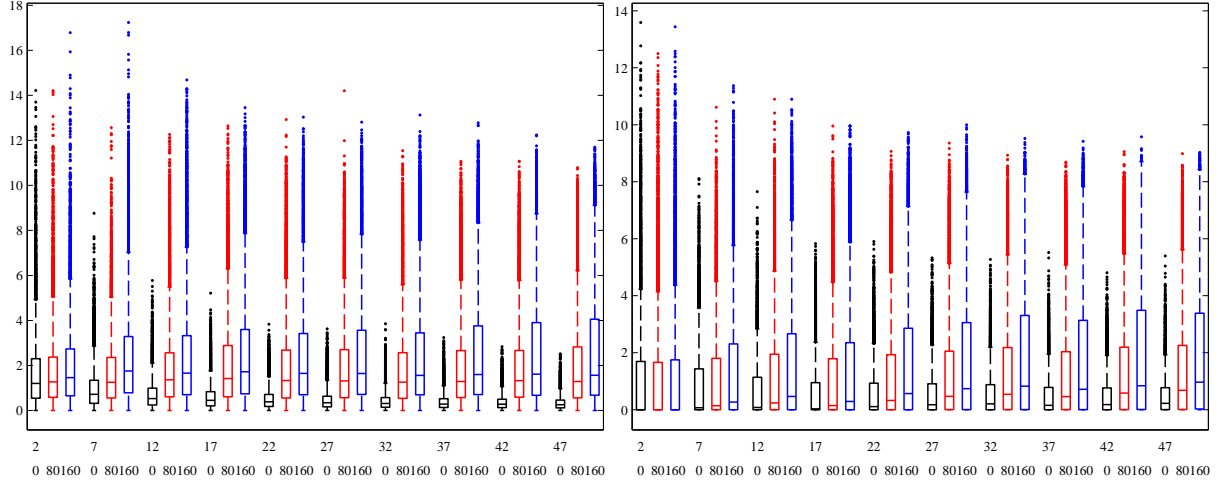
Figure 6.25: Box plot of the ESM differences with no delay and with 80 ms and 160 ms delay, snr = 1 (left) and snr = 100 (right).
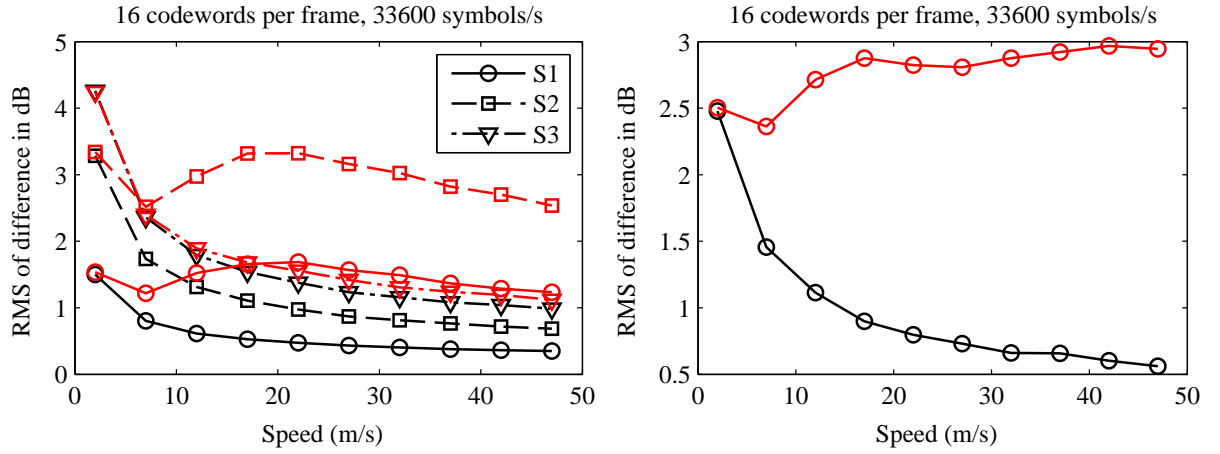
Figure 6.26: RMS of the ESM difference as a function of speed, $n = 1$, snr = 1; the instantaneous comparison is shown in black, the delayed comparison in red. State by state evolution of the channel on the left, joint time series on the right.
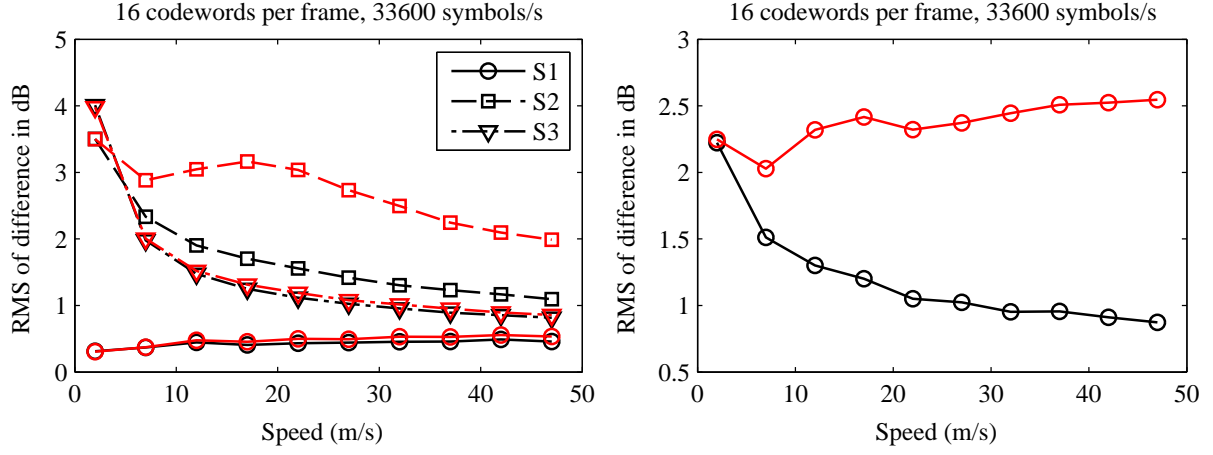
Figure 6.27: RMS of the ESM difference as a function of speed, $n = 1$, snr $= 100$; the instantaneous comparison is shown in black, the delayed comparison in red. State by state evolution of the channel on the left, joint time series on the right.
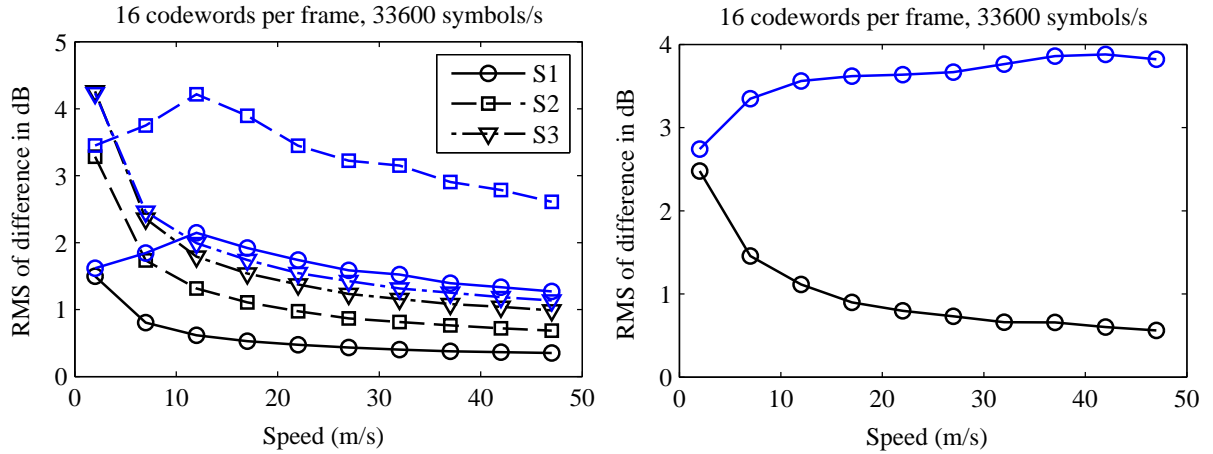
Figure 6.28: RMS of the ESM difference as a function of speed, $n = 2$, snr $= 1$. The instantaneous comparison is shown in black, the delayed comparison in blue. State by state evolution of the channel on the left, joint time series on the right.
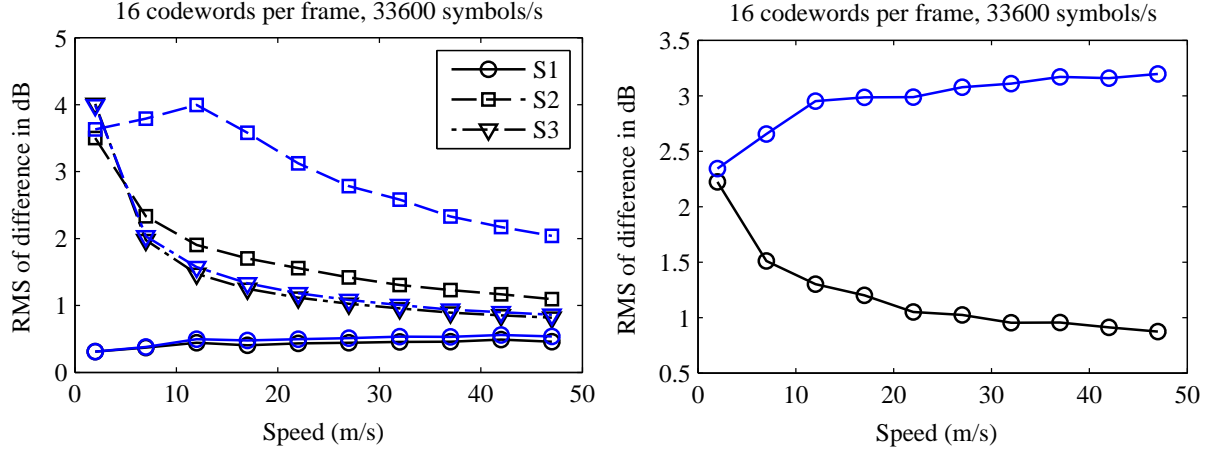
Figure 6.29: RMS of the ESM difference as a function of speed, $n = 2$, snr $= 100$. The instantaneous comparison is shown in black, the delayed comparison in blue. State by state evolution of the channel on the left, joint time series on the right.
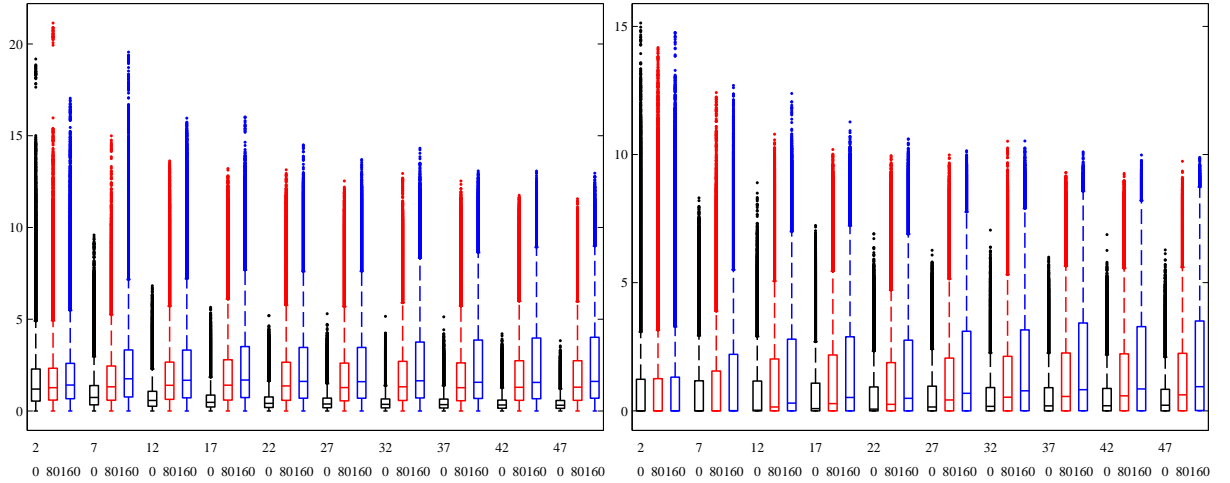
Figure 6.30: Box plot of the ESM differences with no delay and with 80 ms and 160 ms delay, snr $= 1$ (left) and snr $= 100$ (right).
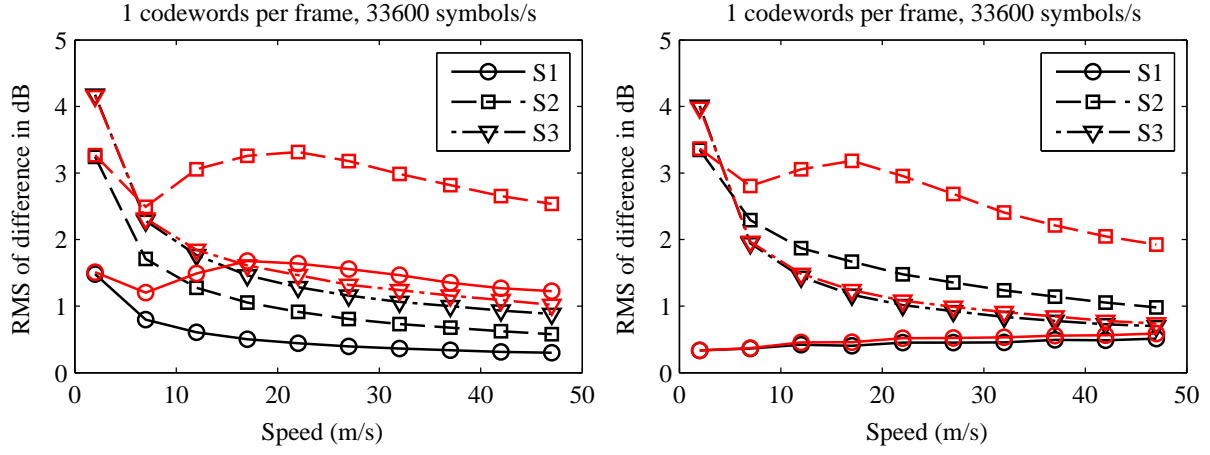
Figure 6.31: RMS of the ESM difference as a function of speed, $n = 1$; the instantaneous comparison is shown in black, the delayed comparison in red. snr $= 1$ on the left, snr $= 316$ on the right.
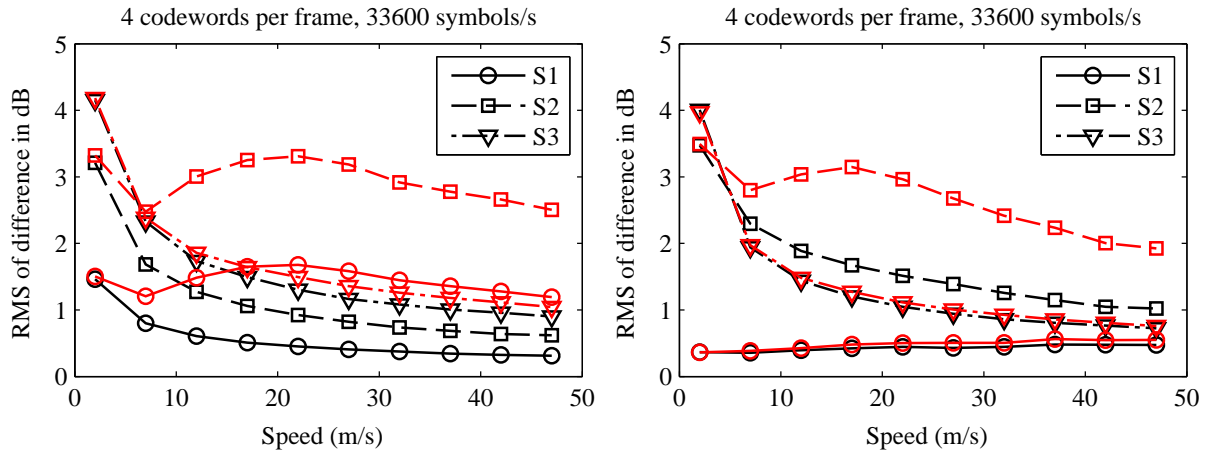
Figure 6.32: RMS of the ESM difference as a function of speed, $n = 1$; the instantaneous comparison is shown in black, the delayed comparison in red. snr $= 1$ on the left, snr $= 316$ on the right.
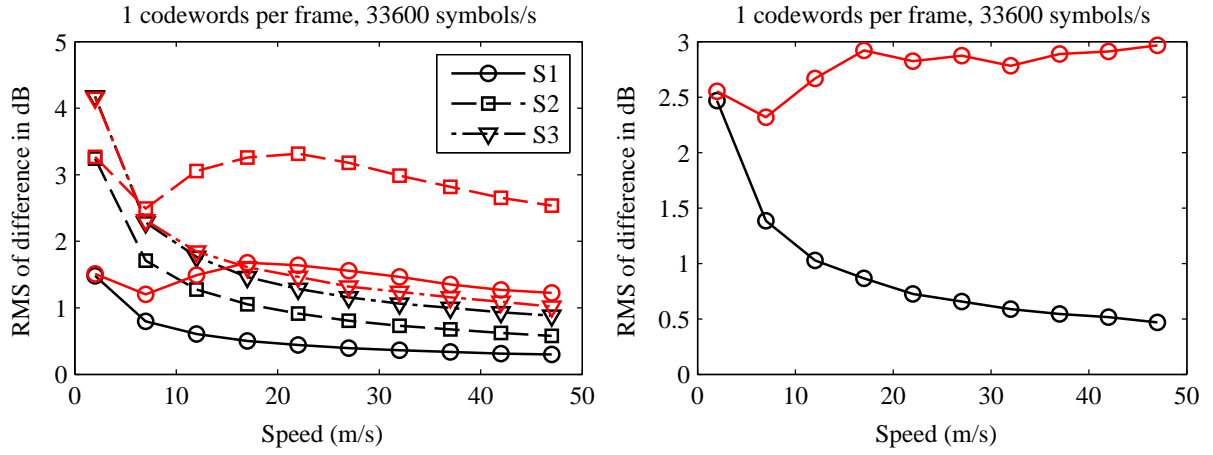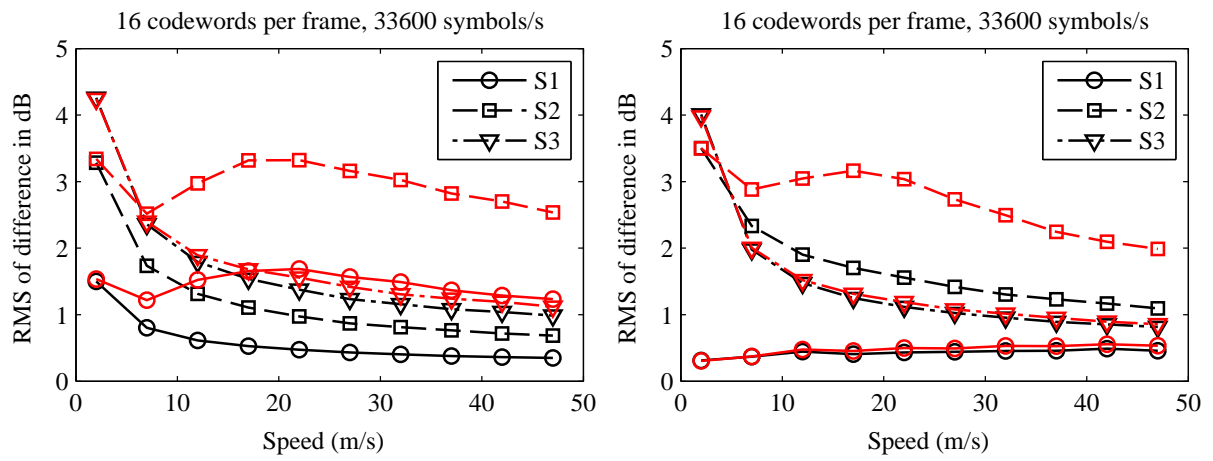
Figure 6.33: RMS of the ESM difference as a function of speed, $n = 1$; the instantaneous comparison is shown in black, the delayed comparison in red. snr $= 1$ on the left, snr $= 316$ on the right.

Chapter 7

Physical layer adaptation in the return link II: automatic CSI balancing

Contents

7.1 Introduction

As explained in the previous chapter, in some cases open loop operation can offer advantages over closed loop, since the CSI is timely (shadowing events occur next to the mobile terminal). The CSI accuracy, however, is reduced due to uncorrelated multipath fading in return and forward link. Thus, we saw how scenarios with strong multipath and low speed seem more favorable for closed loop operation, whereas open loop would be a better choice for high speed and strong line of sight scenarios.

Switching between open and closed loop modes depending on the scenario is not a trivial task, and different methods could be proposed. One possible option could be the use of a look-up table (LUT), built from simulation results, and containing the optimum operating mode for different parameters (speed, multipath, shadowing coherence distance, etc.). This option,
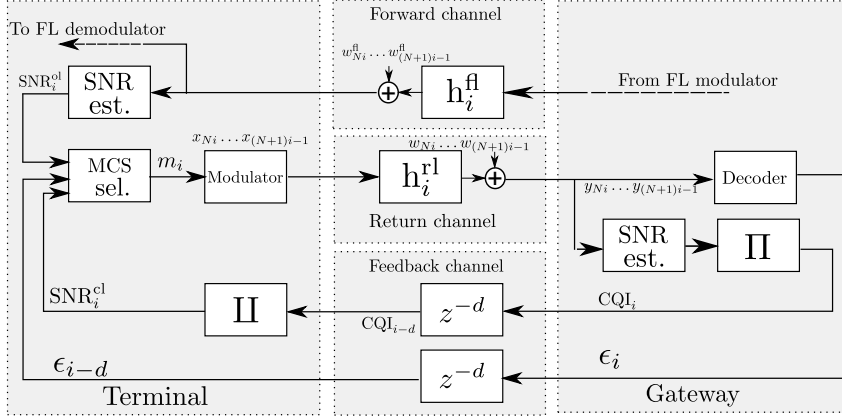
Figure 7.1: Diagram of the information interchange and link adaptation procedure.

however, requires the receiver to estimate these parameters in an accurate way, and is not robust against practical impairments not considered in the simulation step.

In this chapter we present a method to automatically balance the open loop and closed loop CSI. This method relies only on the ACK/NAK interchange between transmitter and receiver, and is able to exploit both CSI values at the same time. An automatic backoff margin is also designed as a byproduct of the proposed method.

7.2   Problem statement

Let us denote by $m_i \in \{1, \ldots, M\}$ the index of the MCS selected in time instant i. We define $\epsilon_i \in \{0, 1\}$, the error event of the i-th codeword, as

$$\epsilon_i = \begin{cases} 1 & \text{if } \gamma_{\text{eff},i} < \gamma_{\text{th}} \\ 0 & \text{otherwise.} \end{cases} \tag{7.1}$$

We also assume that the error probability $\mathbb{P}\left[\epsilon_k = 1\right]$ depends only on m_k and $\mathbf{h}_k$ (the channel vector experienced), and thus is independent of the transmitted message. This assumption is compatible with the proposed error function (7.1), but includes other error functions that characterize coded transmission. We denote the probability of error of the j-th MCS under channel $\mathbf{h}$ as $E\left(j, \mathbf{h}\right)$, and assume that $E\left(j + 1, \mathbf{h}\right) \geq E\left(j, \mathbf{h}\right)$, i.e., higher transmission rates imply higher error probabilities.

We consider two types of feedback: first, the receiver acknowledges the correct decoding of the $(i-d)$-th codeword, so the values $\epsilon_0, \ldots, \epsilon_{i-d}$ are available at the transmitter at time instant i, with d the feedback delay; second, the receiver estimates the channel quality in time instant $i - d$ and includes a channel quality indicator (CQI) in the feedback message. Although there are different ways to calculate the CQI, we consider the index of the highest MCS supported by channel $\mathbf{h}_i^{\text{rl}}$, i.e., $\text{CQI}_i = \arg \max_j \left[j | E\left(j, \mathbf{h}_i^{\text{rl}}\right) \leq p_{\text{CQI}}\right]$. If no MCS is supported, then $\text{CQI}_i = 1$.

p_{CQI} is a value that determines a limit on the error probability. In the case of using Gaussian coding with a sufficiently large code size, for example, p_{CQI} can be set to a number as close to 0 as desired. With the proposed error function (7.1), the error probability can be set to zero if the channel is sufficiently favorable, so we assume $p_{\mathrm{CQI}} = 0$.

The CQI value can be obtained from the effective SNR value by means of a function Π. The function $\Pi(\mathrm{snr})$ is an LUT that maps SNR intervals to values. Throughout this section, we assume that the SNR values are in dB for convenience. For M MCS values, the function Π can be parametrized by $M - 1$ thresholds t_i, $i = 1, \ldots, M - 1$, such that $\Pi(\mathrm{snr}) = j \iff t_{j-1} \leq \mathrm{snr} < t_j$, where the higher and lower thresholds are defined as $t_0 = -\infty$ and $t_M = +\infty$. The function Π is usually referred to as *inner loop* for link adaptation, and we assume the thresholds to be the γ_{th} values for each MCS[1]. If perfect SNR information was available at the transmitter, the link adaptation procedure would be trivial for a calibrated receiver. In practice, however, this information may not be available, so a correction has to be made to the estimated SNR value.

We define the function $\amalg$ as a function that maps MCS indexes to SNR values, such that $\Pi(\amalg(j)) = j$. For example, a function that maps every MCS index to its SNR threshold γ_{th} meets this requirement, and is the one we will use throughout this section. The objective of $\amalg$ is to map CQI to SNR values. We also define $\mathrm{snr}_i^{\mathrm{cl}} \triangleq \amalg(\mathrm{CQI}_{i-d})$ as the SNR that the transmit side gets to know about the received quality d frames earlier in closed loop mode. Note that $\mathrm{snr}_i^{\mathrm{cl}}$ denotes an estimate of the effective SNR performed in time instant $i - d$, but used at the terminal in instant i. In absence of feedback delay ($d = 0$), channel estimation error and other impairments, the optimal MCS selection would be $m_i = \Pi(\amalg(\mathrm{CQI}_i)) = \mathrm{CQI}_i$. A common practice to accommodate these impairments is the application of margins to the received CQI value, so

$$m_i = \Pi\left(\mathrm{snr}_i^{\mathrm{cl}} + c^{\mathrm{cl}}\right) \tag{7.2}$$

with c^{cl} the SNR margin in dB. A possible approach to select c is by means of an LUT that stores values of c for different scenarios, where parameters like channel distribution, Doppler, detection complexity, etc. have to be taken into account. This approach has some drawbacks that limit its application to practical settings. First, filling the LUT requires running exhaustive simulations under many different settings to be applicable to practical scenarios, and its behavior will be unpredictable under conditions that differ from the stored ones. Second, the receiver has to estimate the required parameters, which can be computationally expensive, and errors in the estimation of these parameters might lead to unexpected behavior. Thus, an adaptive method to adjust c^{cl} is required in many cases. An adaptation of c^{cl} based on ACK/NAK reception was proposed in [114], and applied to the satellite scenario in [13].

On top of the feedback information, the terminal is also observing the channel in the forward link to obtain open-loop CSI, as explained in Chapter 6. If the duplexing scheme was TDD, the terminal might gain access to timely and accurate CSI just by measuring the

---

[1]Note that the γ_{th} for the first MCS is not used as a threshold.

forward link channel. In our setting, duplexing is performed by means of frequency separation, so this assumption does not hold. Under our model, however, there is some degree of correlation between the forward and return link, as the LOS component is the same for both links. Therefore, depending on the scenario, the open loop CSI would be more or less accurate. Let us define $\text{snr}^{\text{ol},i}$ as the most recent SNR estimation on the forward link. We assume that this SNR estimation is perfect, and equal to the effective SNR of the previous codeword, i.e., $\text{snr}^{\text{ol},i} = \gamma^{\text{fl}}_{\text{eff},i-1}$. This assumption does not affect the design of the method, and is made for the sole purpose of simplifying the simulations. We might think of performing a similar adaptation as in the closed loop case (7.2)

$$m_i = \Pi\left(\text{snr}^{\text{ol}}_i + c^{\text{ol}}\right). \tag{7.3}$$

Once again, the margin c^{ol} should be obtained adaptively or by means of an LUT. In Figure 7.1 we show a diagram containing the main variables of the system model.

A further question is how to determine the scenarios where (7.3) or (7.2) are more appropriate to be used. It is expected that in scenarios with relatively low speed or strong multipath the closed loop approach would perform better, while strong LOS and high speed scenarios are more suitable for the open loop one. A possible approach is to perform parameter estimation (speed, multipath, etc.) and obtain the optimum strategy from an LUT, which had to be previously filled according to exhaustive simulation results. In the following section we present an adaptive approach that does not require to perform this parameter estimation, and solely relies on the feedback of CQI and ACK/NAK, as well as in the open loop SNR estimation.

7.3 Adaptive CSI balancing

A key observation in (7.2)-(7.3) is that they can be jointly described by

$$m_i = \Pi\left(\xi^{\text{ol}}\text{snr}^{\text{ol}}_i + \xi^{\text{cl}}\text{snr}^{\text{cl}}_i + c\right). \tag{7.4}$$

If we set $\xi^{\text{ol}} = 0$, $\xi^{\text{cl}} = 1$ we arrive to (7.2), and $\xi^{\text{cl}} = 0$, $\xi^{\text{ol}} = 1$ leads to (7.3). Note that (7.4) includes any affine combination of snr^{ol} and snr^{cl}, so it generalizes the open loop and closed loop strategies. We now derive an adaptation method for general values of ξ^{cl} and ξ^{ol}, and in Section 7.4 we introduce an alternative formulation where their value is constrained to sum one, i.e., $\xi^{\text{cl}} + \xi^{\text{ol}} = 1$.

For simplicity, we denote $\mathbf{snr}_i \triangleq \left[\text{snr}^{\text{cl}}_i \ \text{snr}^{\text{ol}}_i\right]^T$ and $\boldsymbol{\xi} \triangleq \left[\xi^{\text{cl}} \ \xi^{\text{ol}}\right]^T$; the derivations from now on could be generalized for vectors $\mathbf{snr}$ and $\boldsymbol{\xi}$ of any size, so we could include channel prediction in this framework, for example.

Following a similar approach as [115], we state the problem of finding the margin c and SNR balancing weights $\boldsymbol{\xi}$ such that the PER converges to a fixed *target PER* p_0. The desired

values can be obtained as the solution to the following optimization problem

$$\min_{c,\boldsymbol{\xi}} J(c,\boldsymbol{\xi}) = |\mathbb{E}\left[\epsilon\right] - p_0|^2. \tag{7.5}$$

Note that (7.5) does not have any optimality properties in terms of throughput, but just sets the mean packet error rate to the desired value p_0. In practice, nevertheless, it is expected that high SNR values will lead to the use of higher rate MCS to meet the target PER p_0, so the throughput is implicitly increased.

Problem (7.5) can be solved by performing a gradient descent on $J(c,\boldsymbol{\xi})$. The gradient of $J(c,\boldsymbol{\xi})$ can be worked out as

$$\nabla J(c,\boldsymbol{\xi}) = 2\left(\mathbb{E}\left[\epsilon\right] - p_0\right)\nabla\mathbb{E}\left[\epsilon\right]. \tag{7.6}$$

A gradient descent iteration reads as

$$\begin{bmatrix} c_{i+1} \\ \boldsymbol{\xi}_{i+1} \end{bmatrix} = \begin{bmatrix} c_i \\ \boldsymbol{\xi}_i \end{bmatrix} - \mu_i \cdot \nabla\left. J(c,\boldsymbol{\xi})\right|_{c_i,\boldsymbol{\xi}_i}. \tag{7.7}$$

Obtaining a numerical expression for the gradient $\left. J(c,\boldsymbol{\xi})\right|_{c_i,\boldsymbol{\xi}_i}$ is not possible: the expectation of ϵ depends on the PDF of the channel, which we assume unknown at the transmitter. On top of this, the PDF of the channel might change over time. Instead, we propose a stochastic gradient approach, where the expectations are substituted by instantaneous observations.

Let us define

$$\Omega \triangleq \boldsymbol{\xi}^T\mathbf{snr} + c, \tag{7.8}$$

as the indicator SNR with which the MCS m_i is selected in (7.4). Ω is a function of c and $\boldsymbol{\xi}$ whose gradient is trivial to compute, so that applying the chain rule of differentiation in (7.6) we arrive at

$$\begin{aligned} \nabla J(c,\boldsymbol{\xi}) &= 2\left(\mathbb{E}\left[\epsilon\right] - p_0\right)\nabla\mathbb{E}\left[\epsilon\right] \\ &= 2\left(\mathbb{E}\left[\epsilon\right] - p_0\right)\mathbb{E}\left[\frac{\partial\epsilon}{\partial\Omega}\nabla\Omega\right] \\ &= 2\mathbb{E}\left[\frac{\partial\epsilon}{\partial\Omega}\right]\left(\mathbb{E}\left[\epsilon\right] - p_0\right)\begin{bmatrix} 1 \\ \mathbf{snr} \end{bmatrix} \end{aligned} \tag{7.9}$$

Following the stochastic gradient approach, we substitute $\mathbb{E}\left[\epsilon\right]$ by the instantaneous value ϵ_{i-d}; also, and since $2\partial\epsilon/\partial\Omega$ is positive[2], we embed $2\mathbb{E}\left[\partial\epsilon/\partial\Omega\right]$ into the positive adaptation constant μ_i. The resulting expression for the update of c and $\boldsymbol{\xi}$ reads as (see Figure 7.2)

---

[2]This can be proved by writing probability of error, averaged over all channel states, in integral form, and using the assumption that for a channel state, the probability of error is higher for higher MCS. Higher values of Ω lead to higher MCS values, which increase the probability of error.
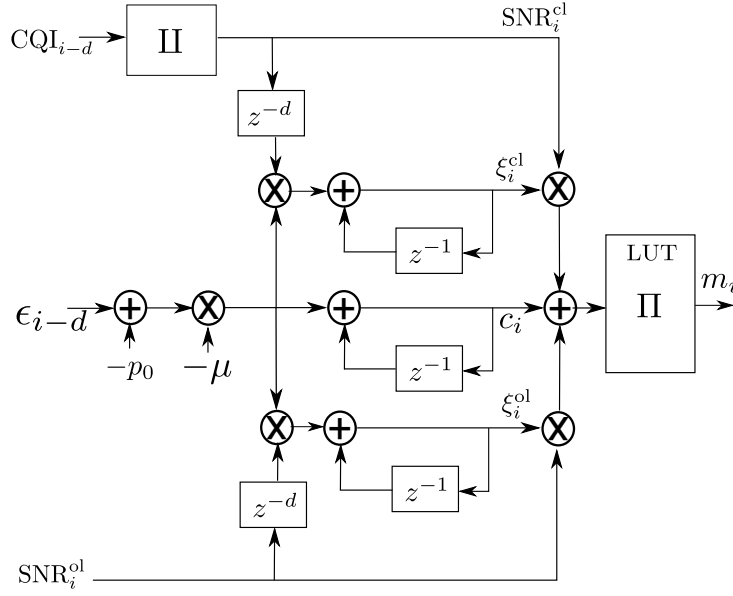
Figure 7.2: Diagram of the adaptation process

$$\begin{bmatrix} c_{i+1} \\ \boldsymbol{\xi}_{i+1} \end{bmatrix} = \begin{bmatrix} c_i \\ \boldsymbol{\xi}_i \end{bmatrix} - \mu\left(\epsilon_{i-d} - p_0\right) \begin{bmatrix} 1 \\ \mathbf{snr}_{i-d} \end{bmatrix}. \tag{7.10}$$

where we removed the dependence of μ with time, so we are using a constant stepsize.

Note that in time instant i the last received feedback is the one corresponding to the information transmitted in time $i - d$. The SNR values used for adaptation in (7.10) have to be the ones used for the MCS selection of the packet the ACK/NAK is referred to. If the transmitter knows the delay introduced by the channel, a delay of z^{-d} has to be introduced in the adaptation algorithm, as shown in Figure 7.2. If the delay value is not known or is variable (in case of ACK/NAK grouping, for example), then the transmitter should store in memory the SNR values used for adaptation of every packet, indexed by a packet ID. When the ACK/NAK for a packet ID is received, the parameter update is performed by recovering the corresponding SNR values from memory. Note that the closed loop SNR value $\mathrm{snr}_i^{\mathrm{cl}}$ used for MCS selection is the one generated by CQI_{i-d}, but the one used for the adaptation of ξ_i^{cl} is $\mathrm{snr}_{i-d}^{\mathrm{cl}}$, which is generated by CQI_{i-2d}.

Remark 1. *Note that the adaptive margin algorithm proposed in [13, 114] is equivalent to the adaptation of c in (7.10). This adaptation is also the one described in [115]. More precisely, the algorithm for update described in [13, 114] is*

$$c_{i+1} = \begin{cases} c_i + \delta_{up} & \text{if } \epsilon_{i-d} = 0 \\ c_i - \delta_{down} & \text{if } \epsilon_{i-d} = 1 \end{cases} \tag{7.11}$$

with δ_{up} and δ_{down} values such that[3]

$$\delta_{down} = \delta_{up}\frac{p_0}{1 - p_0}. \tag{7.12}$$

It can be seen that (7.10) and (7.11) describe the same adaptation, provided that $\delta_{up} = \mu p_0$ and $\delta_{down} = \mu (1 - p_0)$.

7.4   Convergence enhancements

We observed that the adaptation method described by (7.10) offers a noisy behavior in convergence, thus needing small values of μ. This affects the convergence speed of the algorithm, dramatically decreasing it. Note that (7.10) resembles a least mean squares (LMS) adaptation with input $[1 \ \mathbf{snr}_{i-d}]^T$ and error $\epsilon_{i-d} - p_0$. Normalized LMS (NLMS) [40] is well known to outperform LMS in convergence speed. If the step-size is normalized in (7.10), the NLMS-like version reads as

$$\begin{bmatrix} c_{i+1} \\ \boldsymbol{\xi}_{i+1} \end{bmatrix} = \begin{bmatrix} c_i \\ \boldsymbol{\xi}_i \end{bmatrix} - \frac{\mu}{1 + \|\mathbf{snr}_{i-d}\|^2} \left(\epsilon_{i-d} - \tilde{p}_{0,i}\right) \begin{bmatrix} 1 \\ \mathbf{snr}_{i-d} \end{bmatrix}. \tag{7.13}$$

Note that we substituted p_0 by $\tilde{p}_{0,i}$. In general, (7.13) does not converge to a PER of $\tilde{p}_{0,i}$, since the first component of a stationary point meets

$$\mathbb{E}\left[\frac{\epsilon_i - \tilde{p}_{0,i}}{1 + \|\mathbf{snr}_{i-d}\|^2}\right] = 0 \tag{7.14}$$

which does no necessarily imply $\mathbb{E}[\epsilon_i] = \tilde{p}_{0,i}$. It is expected, however, that an appropriate choice of $\tilde{p}_{0,i}$ (not equal to p_0) led to a PER of p_0. We propose to adjust $\tilde{p}_{0,i}$ following a recursion

$$\tilde{p}_{0,i+1} = \tilde{p}_{0,i} - \lambda\left(\epsilon_{i-d} - p_0\right) \tag{7.15}$$

It is clear that $\mathbb{E}[\epsilon_i] = p_0$ is a stationary point of (7.15), thus leading to the desired PER. A block diagram of this NLMS adaptation is shown in Figure 7.3.

It has been also observed that the NLMS adaptation offers a good convergence performance for the terms ξ, but not for the margin c. This might be caused because the corrections to this term are smaller in absolute value than those for ξ because snr^{ol} and snr^{cl} are usually bigger than 1. To overcome this problem, we propose an alternative formulation that increases the

---

[3]In [13] the steps are selected to meet $\delta_{down} = \delta_{up}p_0$ instead of (7.12). Both formulations, however, are equivalent for low values of p_0.
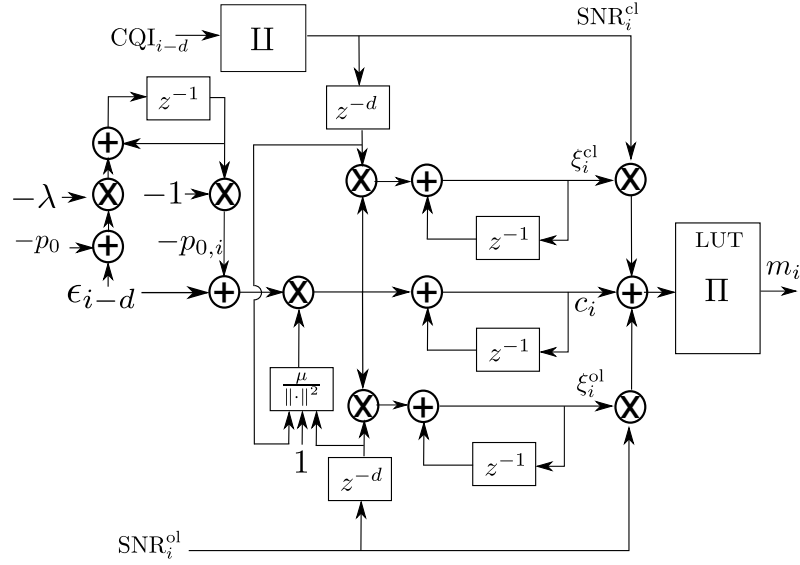
Figure 7.3: Diagram of NLMS adaptation

speed of convergence of c

$$\begin{bmatrix} c_{i+1} \\ \boldsymbol{\xi}_{i+1} \end{bmatrix} = \begin{bmatrix} c_i \\ \boldsymbol{\xi}_i \end{bmatrix} - \frac{\mu}{\theta^2 + \|\mathbf{snr}_{i-d}\|^2} \left(\epsilon_{i-d} - \tilde{p}_{0,i}\right) \begin{bmatrix} \theta \\ \mathbf{snr}_{i-d} \end{bmatrix}. \tag{7.16}$$

We also performed experiments with only one weight ξ instead of two. In this case, we defined the MCS selection rule as

$$m_i = \Pi \left(\left(1 - \xi^{\mathrm{cl}}\right) \mathrm{snr}_i^{\mathrm{ol}} + \xi^{\mathrm{cl}} \mathrm{snr}_i^{\mathrm{cl}} + c \right), \tag{7.17}$$

and the corresponding adaptation rule as

$$\begin{bmatrix} c_{i+1} \\ \xi_{i+1}^{\mathrm{cl}} \end{bmatrix} = \begin{bmatrix} c_i \\ \xi_i^{\mathrm{cl}} \end{bmatrix} - \frac{\mu}{\theta^2 + \left(\mathrm{snr}_{i-d}^{\mathrm{cl}} - \mathrm{snr}_{i-d}^{\mathrm{ol}}\right)^2} \times \tag{7.18}$$

$$\left(\epsilon_{i-d} - \tilde{p}_{0,i}\right) \begin{bmatrix} \theta \\ \mathrm{snr}_{i-d}^{\mathrm{cl}} - \mathrm{snr}_{i-d}^{\mathrm{ol}} \end{bmatrix}.$$

and $\tilde{p}_{o,i}$ following the recursion in (7.15).

The convergence properties of the methods described in this section are object of future research. Nevertheless, their convergence to the desired PER value has been empirically observed, as described the next section.
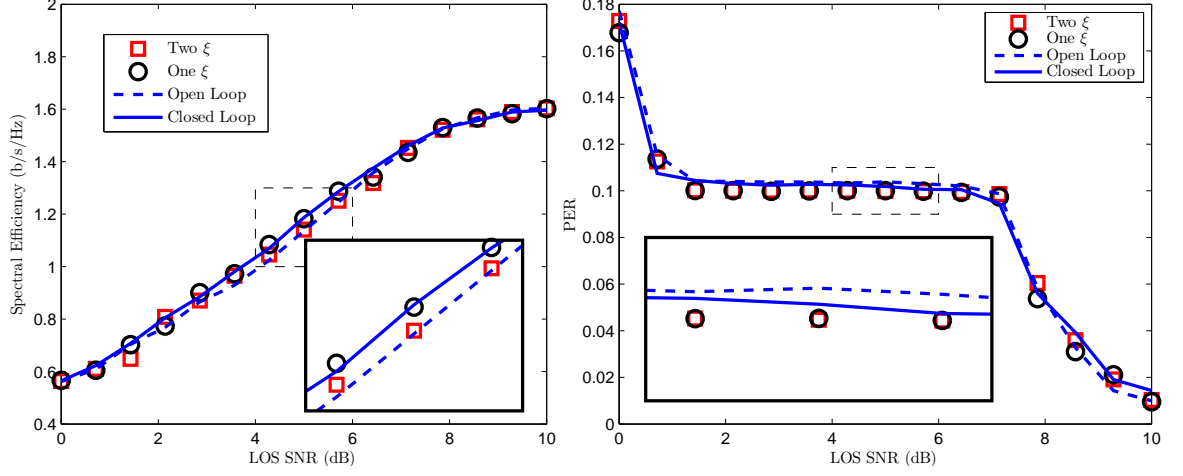
Figure 7.4: Average spectral efficiency (left) and FER (right) for different methods in intermediate tree shadowed environment, state 1, $0.3\,\mathrm{m/s}$, $p_0 = 0.1$
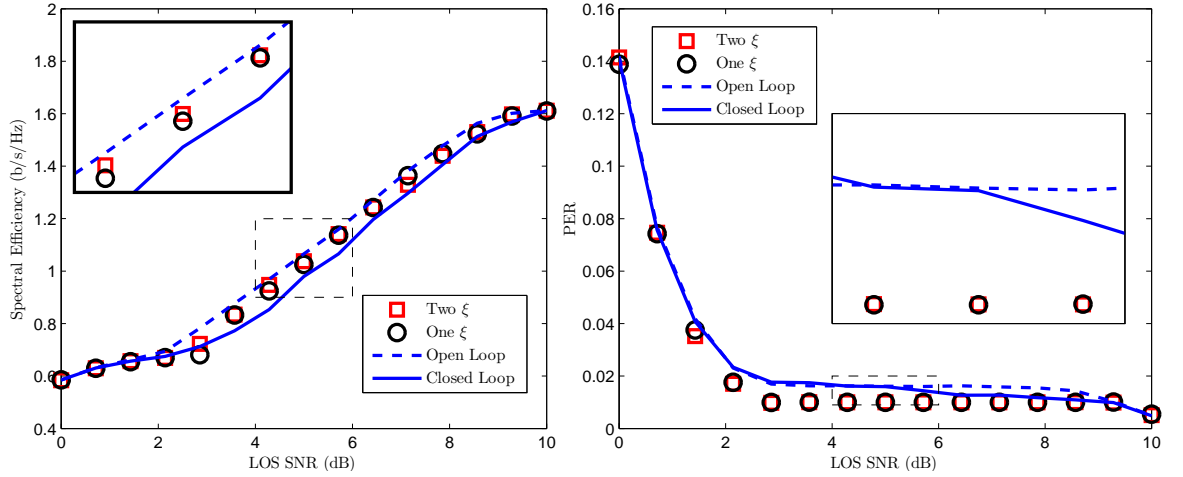
Figure 7.5: Average spectral efficiency (left) and FER (right) for different methods in intermediate tree shadowed environment, state 1, $3\,\mathrm{m/s}$, $p_0 = 0.01$
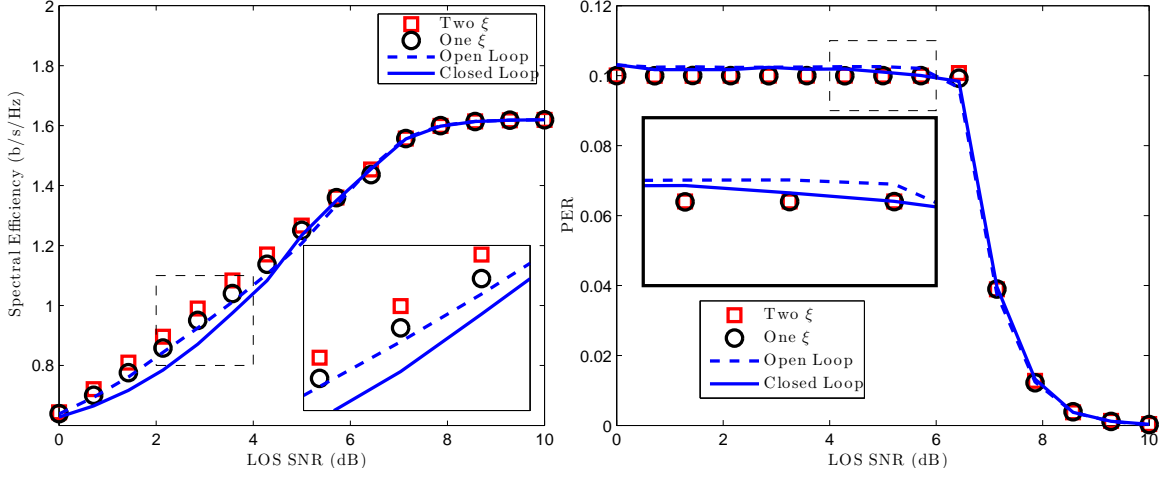
Figure 7.6: Average spectral efficiency (left) and FER (right for different methods in intermediate tree shadowed environment, state 1, $15\,\mathrm{m/s}$, $p_0 = 0.1$

7.5   Simulation results

We performed simulations of the proposed methods, given by equations (7.18) and (7.16), and compare them with open loop and closed loop with automatic margin adaptation, given by (7.11). The adaptation was performed with $\theta = 10$, $\lambda = 10^{-3}$ and $\mu = 1$ for the case of the proposed methods, and $\mu = 10^{-2}$ for the open and closed loop cases, which correspond to $\delta_{\mathrm{up}} = 0.001$ and $\delta_{\mathrm{down}} = 0.009$ for a target PER of 10^{-1}. The parameters were initialized as $c_0 = 0$, $\xi_0^{\mathrm{cl}} = 0.5$, and $\xi_0^{\mathrm{ol}} = 0.5$. We set the feedback delay to $d = 5$ codewords to model the round trip time in a GEO satellite.

We performed simulations over a Loo channel with the parameters for intermediate tree shadowed environment, state 1. Simulations were also carried out for other settings, with similar results observed. We simulated three different terminal speeds, corresponding to $0.3\,\mathrm{m/s}$, $3\,\mathrm{m/s}$ and $15\,\mathrm{m/s}$. The target PER was set to $p_0 = 0.1$ in the first and last speeds, and to $p_0 = 0.01$ for the $3\,\mathrm{m/s}$ case. Average spectral efficiency and throughput results were averaged over the transmission of $6 \cdot 10^4$ packets. Results are shown in Figures 7.4-7.6. Spectral efficiency is defined as $\frac{1}{N}\sum_{i=1}^{N}\epsilon_i r_{m_i}$, with r_j the rate of the j-th MCS.

We see that the lower speed scenario shows a better performance of closed loop with respect to open loop, and that the adaptation with two ξ performs worse than closed loop, and similarly to open loop. Adaptation with one ξ offers approximately the same performance as closed loop. For higher speeds, open loop performs better than closed loop, and so do the proposed methods. For the $15\,\mathrm{m/s}$ case, adaptation with one and two ξ outperform both open loop and closed loop adaptation. It is also noticeable that the proposed methods are more accurate in converging to the desired PER p_0 than the open and closed loop ones. A PER p_0 might not be achievable in very high or very low SNR scenarios. In these cases, the PER converges to the minimum or maximum possible PER values, respectively. The method with one ξ seems to be more robust

than the one with two ξ, and offers a better or at least the same performance than closed loop link adaptation in all the cases, even beating the open loop adaptation for high speeds.

7.6 Implementation aspects

In this chapter we made some simplifications to make the link adaptation problem more tractable. In the following, we make comments about some possible implementation aspects of the proposed algorithm

- **CQI feedback** Throughout the chapter, it was assumed that a CQI value was fed back for every packet. In modern communication standards this is not usually the case, so some packets would have a more outdated CQI than others. Assume that a CQI value is transmitted every K packets. Applying the same weight ξ^{cl} in all K time instants could be suboptimal, since the first packet in every period has much more precise CSI than the last one. A possible approach to overcome this problem is to have a different adaptation loop for each of the possible CQI delays.

- **Use of effective SNR** Although in this section we used effective SNR for convenience, the proposed method is expected to work with other CSI metrics, such as average SNR or RSSI. This can be convenient in case the effective SNR calculation could not be performed by the receiver because of computational complexity.

- **Interference in return link** We neglected the possible interference in the return link, which is impossible to estimate from open loop observations [99]. From the analysis in Chapter 6, the proposed method is in this case expected to converge to weights ξ that reduce the impact of errors in open loop CSI.

- **Estimation errors and uncalibrated receivers** We assumed in the simulations that CQI and SNR estimates in the forward link were perfect, as well as the knowledge of the SNR thresholds for decoding. In practice, there might be some non-negligible errors in the SNR estimation, and the performance of a receiver might not be known a priori. In these cases, the proposed adaptive method should be able to adjust the parameters to meet the PER constraint, although possibly reducing the throughput with respect to the ideal case.

- **Divergence of parameters** In cases of very high or very low SNR, the adaptation parameters ξ and c diverge, as it is not possible to converge to the desired PER p_0. A threshold should be included in the adaptation to prevent this behavior.

7.7 Conclusions

In this chapter we proposed a method for link adaptation in the return link of satellite communications that exploits open loop and closed loop CSI. It works by balancing the importance

given to each CSI value based on the observation of the ACK/NAK exchange between both communication ends. The adaptive algorithm is obtained as a stochastic gradient descent of an unconstrained optimization problem. Interestingly, a baseline algorithm arises as a particular case of this optimization problem. The proposed method is shown to offer a good performance with respect to open loop and closed loop adaptation.

The content of this chapter is the result of a collaboration with Dr. Alberto Rico-Alvariño and Prof. Carlos Mosquera, and was partially published in ASMS 2014 [116].

Chapter 8

Physical layer adaptation in the forward link: ARQ with different MCS

Contents

8.1 Introduction

Differently from the return link, there is no sense in operating the forward link of a satellite communications system with open loop information, since it experiences the same delay as the

closed loop one. As a consequence, the information available in this case will consist of outdated but accurate estimates.

To cope with the large delay, some degree of diversity must be introduced in order to improve the exploitation of the channel capacity. As an example, in [110] it was proposed to use multi-layer codes with different parameters, which were adapted through statistical CSI.

In this chapter, we will look at the case where, instead of different layers, an automatic repeat request (ARQ) mechanism is employed, and more than one transmission per packet is available. In this case, we will try to exploit the use of different MCS in each retransmission attempt. The outage probability will be defined in terms of the post-ARQ packet error probability, and the objective is once again to maximize the throughput subject to a packet error rate constraint.

8.2　System model review

As explained in Section 5.2, the signal model is given by

$$y_k = \sqrt{\mathrm{snr}} \cdot h_k s_k + w_k \tag{8.1}$$

with y_k the symbol received at the k-th time instant, s_k the transmitted symbol, h_k the channel coefficient and snr representing the transmitted over noise power, and w_k a sample of complex Gaussian unit-power noise.

8.2.1　Transmitted and received symbols

The transmitted symbols are the result of applying forward error correction coding and constellation mapping to a stream of bits; we consider a finite set of available codes, as described in Table 8.1. Throughout the chapter we will denote as $\gamma_{\mathrm{th},j}$ the threshold SNR for the i-th MCS, and drop the index i when its meaning is clear. We denote as r_j the spectral efficiency of the j-th MCS, that is obtained by multiplying the code rate with the $\log_2$ of the constellation size (2 for QPSK). In the following, we describe the input-output relationship of the channel. We denote the number of MCS as K.

Symbols form codewords $\mathbf{x}_i = \left[x_{iN}, x_{iN+1}, \ldots, x_{(i+1)N-1} \right]$ of constant length N, such that they see the channel samples

$$\mathbf{h}_i^{\mathrm{fl}} \triangleq \left[h_{iN}^{\mathrm{fl}}, h_{iN+1}^{\mathrm{fl}}, \ldots, h_{(i+1)N-1}^{\mathrm{fl}} \right] \tag{8.2}$$

For each codeword sent, we assume that the other end feeds back an ACK if decoding was possible, and a NAK otherwise. We assume that $N = 2700$, which gives codewords of approximately 80 ms. We neglect the effect of headers or other sort of overhead.

MCS	Coding rate	γ_{th} (dB)	MCS	Coding rate	γ_{th} (dB)
QPSK 1/4	0.357	-1.5	QPSK 2/3	1.261	4
QPSK 1/3	0.616	-0.3	QPSK 3/4	1.390	4.9
QPSK 2/5	0.745	0.6	QPSK 4/5	1.476	5.6
QPSK 1/2	0.831	1.9	QPSK 5/6	1.562	6.1
QPSK 3/5	1.132	3.1	QPSK 8/9	1.691	7.1

Table 8.1: MCS for the forward link, which are the ones used in DVB-S2.
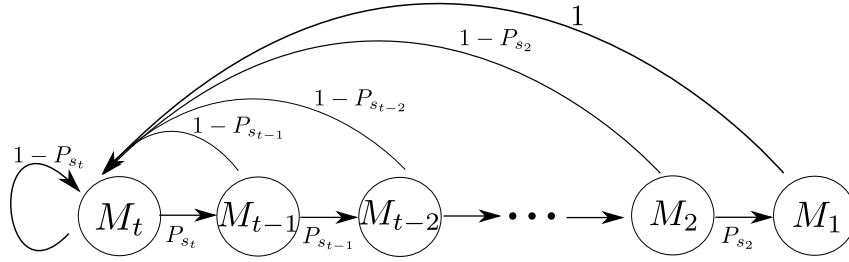
Figure 8.1: Markov model representing the transitions between states.

8.3 Link adaptation

8.3.1 Problem statement

Recall that the objective of link adaptation is to maximize the throughput guaranteeing the data-link layer QoS, expressed in terms of outage or packet error rate and maximum delay. For a maximum delay or, equivalently, for a maximum number of transmissions t, our design variables are the MCS indexes to use on each attempt $\{s_1, \ldots, s_t\} \in [1, \ldots, K]^t$. These values must satisfy an outage constraint given by p_0, so

$$\prod_{i=1}^{t} P_{s_i} \leq p_0. \tag{8.3}$$

Each bit waiting to be transmitted can be in one of t different states denoted by M_i, $i = 1, \ldots, t$, with i the number of maximum attempts before being discarded. In state M_i the data still has i transmissions available. Decoding will fail with a probability P_{s_i}, and data will move onto the next state, M_{i-1}, with $i - 1$ transmissions left; on the other hand, if the transmission is successful -with probability $(1 - P_{s_i})$- the system moves back to the first state. The Markov chain representing this transmission procedure is depicted in Figure 8.1.

With this, we have that the average spectral efficiency, ASE, is given by

$$\eta = \sum_{i=1}^{t} \mathbb{P}\left[M_i\right] r_{s_i} \left(1 - P_{s_i}\right) \tag{8.4}$$

where the probability for a given bit to belong to state i is easily obtained from the diagram in Figure 8.1 as

$$\begin{aligned}
\mathbb{P}\left[M_t\right] &= 1/P \\
\mathbb{P}\left[M_{t-1}\right] &= P_{s_t}/P \\
\mathbb{P}\left[M_{t-2}\right] &= P_{s_t} \cdot P_{s_{t-1}}/P \\
&\cdots \\
\mathbb{P}\left[M_1\right] &= P_{s_t} \cdot P_{s_{t-1}} \cdot \ldots \cdot P_{s_2}/P.
\end{aligned} \tag{8.5}$$

For notational convenience, we have defined $P \triangleq 1 + P_{s_t} + P_{s_{t-1}} \cdot P_{s_{t-1}} + \cdots + P_{s_t} \cdot P_{s_{t-1}} \cdot \ldots \cdot P_{s_2}$. Note that the outage probability of the last state P_{s_1} does not affect the probability of the different states, but only the average spectral efficiency of the last retransmission and the overall outage probability. The input bit rate is supposed to be such that these probabilities remain stationary with time.

Based on the previous considerations, the choice of the MCS indexes can be recast as the following optimization problem:

$$\begin{aligned}
\text{maximize} \quad & \sum_{i=1}^{t} \mathbb{P}\left[M_i\right] r_{s_i}(1 - P_{s_i}) \\
\text{subject to} \quad & \prod_{i=1}^{t} P_{s_i} \leq p_0.
\end{aligned} \tag{8.6}$$

This problem seems hard to solve, at least for the general case of different rates being assigned to subsequent retransmissions. We will now present some simplifications to this problem.

8.3.2 Simplifying the optimization problem

There are some optimality properties that can be exploited to avoid the need for an exhaustive search over the whole set of codes. Intuitively, for example, it seems plausible to expect higher rates at earlier attempts and more robust MCS for the subsequent retransmissions. Next we will prove this for $t = 2$, with the conjecture, supported by practical optimizations, that this is also the case for any t.

Proposition 1. *If t=2, then the solution to (8.6) meets the following two propositions:*

    1. if $\exists j \neq s_i$ such that $r_{s_i}(1 - P_{s_i}) < r_j(1 - P_j)$ then $P_{s_i} \leq P_j \; \forall i=1, 2$

    2. $r_{s_2} \geq r_{s_1}$ (or, equivalently, $P_{s_2} \geq P_{s_1}$).

Proof. The first proposition is easy to prove for $i = 1$. Let us assume (s_2^*, s_1^*) is optimum, and denote $T_x \triangleq r_x(1 - P_x)$, so we can write the objective function as

$$r = \frac{T_{s_2^*} + P_{s_2^*} T_{s_1^*}}{1 + P_{s_2^*}}. \tag{8.7}$$

It is easy to see that if there exists j such that $T_j > T_{s_1^*}$ and $P_j < P_{s_1^*}$ the outage constraint will be met if it was met for (s_2^*, s_1^*) and the objective function will increase, so (s_2^*, s_1^*) cannot be optimum.

Next, assume that (s_2^*, s_1^*) is optimum with $P_{s_1^*} > P_{s_2^*}$. This implies that $T_{s_1^*} > T_{s_2^*}$ from the previous property. Also, we have that the outage probability of (s_2^*, s_1^*) is the same as that of (s_1^*, s_2^*). Therefore, to prove that (s_2^*, s_1^*) is not optimum it suffices to prove that

$$T_{s_1^*}\theta_{s_1^*} + T_{s_2^*}(1 - \theta_{s_1^*}) \geq T_{s_2^*}\theta_{s_2^*} + T_{s_1^*}(1 - \theta_{s_2^*}) \tag{8.8}$$

where $\theta_{s_2^*} \triangleq 1/(1 + P_{s_2^*})$ is the probability of a bit being transmitted in the first state if the order is (s_2^*, s_1^*), and $\theta_{s_1^*} \triangleq 1/(1 + P_{s_1^*})$ is the probability of a bit being transmitted in the first state if the order is (s_1^*, s_2^*). Note also that $\theta_i \geq (1 - \theta_i)$, and $\theta_{s_2^*} \geq \theta_{s_1^*}$. With this, we have that

$$
\begin{aligned}
T_{s_1^*}\theta_{s_1^*} + T_{s_2^*}(1 - \theta_{s_1^*}) &= \theta_{s_1^*}\left(T_{s_1^*} - T_{s_2^*}\right) + T_{s_2^*} \\
&\overset{(i)}{\geq} \left(1 - \theta_{s_1^*}\right)\left(T_{s_1^*} - T_{s_2^*}\right) + T_{s_2^*} \\
&= T_{s_1^*} + \theta_{s_1^*}\left(T_{s_2^*} - T_{s_1^*}\right) \\
&\overset{(ii)}{\geq} T_{s_1^*} + \theta_{s_2^*}\left(T_{s_2^*} - T_{s_1^*}\right) \\
&= T_{s_1^*}\left(1 - \theta_{s_2^*}\right) + T_{s_2^*}\theta_{s_2^*}
\end{aligned}
\tag{8.9}
$$

where (i) is due to $T_{s_1^*} > T_{s_2^*}$ and $\theta_{s_1^*} > (1 - \theta_{s_1^*})$, and (ii) is due to $T_{s_1^*} > T_{s_2^*}$ and $\theta_{s_1^*} > \theta_{s_2^*}$. With this, we prove that it is better to transmit first with the higher rate code, so (s_2^*, s_1^*) is not optimum if $P_{s_1^*} > P_{s_2^*}$. Now, we have to prove proposition I for $i = 2$. This is easy to prove since

$$\frac{\mathrm{d}r}{\mathrm{d}P_{s_2}} = \frac{T_{s_1} - T_{s_2}}{(1 + P_{s_2})} \leq 0 \tag{8.10}$$

as $T_{s_1} \leq T_{s_2}$, and

$$\frac{\mathrm{d}r}{\mathrm{d}T_{s_2}} = \frac{1}{(1 + P_{s_2})} \geq 0 \tag{8.11}$$

so decreasing the error probability of s_2 while increasing T_{s_2} will increase the objective function.

This lemma proves that for $t = 2$ we only have to take into account those MCS where an increment in the outage probability implies an increment in the throughput, and that the MCS are of non-increasing rate. Note that this result is quite general, since the only assumption for the relationship between code rate and error probability is that higher rates imply higher error probabilities. $\qquad\square$

Conjecture 8.1. *For all t, the solution to (8.6) meets the following two propositions:*

1. *if $\exists j \neq s_i$ such that $r_{s_i}\left(1 - P_{s_i}\right) < r_j\left(1 - P_j\right)$ then $P_{s_i} \leq P_j \ \forall i = 1, \ldots, t$*

2. *$r_{s_{i+1}} \geq r_{s_i}$ (or, equivalently, $P_{s_{i+1}} \geq P_{s_i}$) $\forall i = 1, \ldots, t - 1$.*

8.3.3   Optimization algorithm

With the above conjecture, the optimization algorithm could be written as follows: in a first step, we will obtain the subset of K' MCS that meet Condition 1; then, we will perform the optimization over them using Condition 2.

Require: MCS set sorted by descending rate, $i < j \Leftrightarrow r_j \leq r_i \quad \forall r_i, r_j \in \{r_i\}_{i=1}^M$

> $T_0 \leftarrow 0$
> **for** $i = 1$ to K **do**
>> **if** $r_i(1 - P_i) < T_{i-1}$ **then**
>>> Delete MCS i
>> **else**
>>> $T_i = r_i (1 - P_i)$
>> **end if**
> **end for**

We now compute all the valid combinations of MCS:

> $i \leftarrow 0$
> **for** $a_1 = 1$ **to** K' **do**
>> **for** $a_2 = 1$ **to** a_1 **do**
>>> $\vdots$
>>> **for** $a_t = 1$ **to** a_{t-1} **do**
>>>> $F_i = \{a_1, a_2, \cdots, a_t\}$
>>>> **if** $\prod_{j=1}^{t} P_{a_j} \leq p_0$ **then**
>>>>> $G_i = \sum_{j=1}^{t} \mathbb{P}\left[M_{a_j}\right] r_{a_j}(1 - P_{a_j})$
>>>> **else**
>>>>> $G_i = -\infty$
>>>> **end if**
>>> **end for**
>> **end for**
> **end for**
> **return** $F_{\text{opt}} = \arg\max_{F_i} G_i$

The complexity of this algorithm is less than that of a brute-force search over all the possible combinations, which would result into K^t combinations; the complexity is now given by the following proposition.

Lemma 6. *The modified algorithm only needs to check $\binom{K'-1+t}{t}$ combinations.*

Proof. From the algorithm description, we have that the number of combinations will be given

by

$$
\begin{aligned}
N &= \sum_{a_1=0}^{K'-1} \sum_{a_2=0}^{a_1} \cdots \sum_{a_{t-2}=0}^{a_{t-3}} \sum_{a_{t-1}=0}^{a_{t-2}} \sum_{a_t=0}^{a_{t-1}} 1 \\
&= \sum_{a_1=0}^{K'-1} \sum_{a_2=0}^{a_1} \cdots \sum_{a_{t-2}=0}^{a_{t-3}} \sum_{a_{t-1}=0}^{a_{t-2}} \binom{a_{t-1}+1}{1} \\
&= \sum_{a_1=0}^{K'-1} \sum_{a_2=0}^{a_1} \cdots \sum_{a_{t-2}=0}^{a_{t-3}} \binom{a_{t-2}+2}{2} \\
&= \sum_{a_1=0}^{K'-1} \binom{a_1+(t-1)}{t-1} = \binom{K'-1+t}{t}
\end{aligned}
\tag{8.12}
$$

where we have used $\sum_{j=0}^{k} \binom{n+j}{n} = \binom{k+n+1}{n+1}$. $\qquad\square$

The reduction in complexity with respect to the brute force solution is quite remarkable, reaching almost an order of magnitude for $t = 3$ and $K' = 20$.

8.3.4 Modified optimization: alert mode

Assuming perfect knowledge of the long-term PER of each MCS, the solution described above is optimum. However, and as we will show in the following sections, having this kind of CSI is rarely the case; often, we will deal with information obtained during a short period or time, or we will wish to meet the PER constraint also in short time windows and not only in the long-term average.

Finding the optimum solution for the above situations may be very difficult. Here, we will use a simple heuristic and test its performance: if more than a given number of packets are lost during a given time frame, we will always enforce the most protected MCS in the last transmission. Although more conservative than the previous solutions, the latter should avoid an excessive loss of packets.

8.3.5 Knowledge of P_i

Offline adaptation

The simplest possibility to obtain P_i is to focus on the long-term statistics of the channel. If the channel is ergodic, then we can simply obtain them offline, either simulating or performing measurements on a real setup. However, note that solving the optimization problem based on these values would meet the PER constraint in the long term, while short transmission bursts may experience higher outage values.
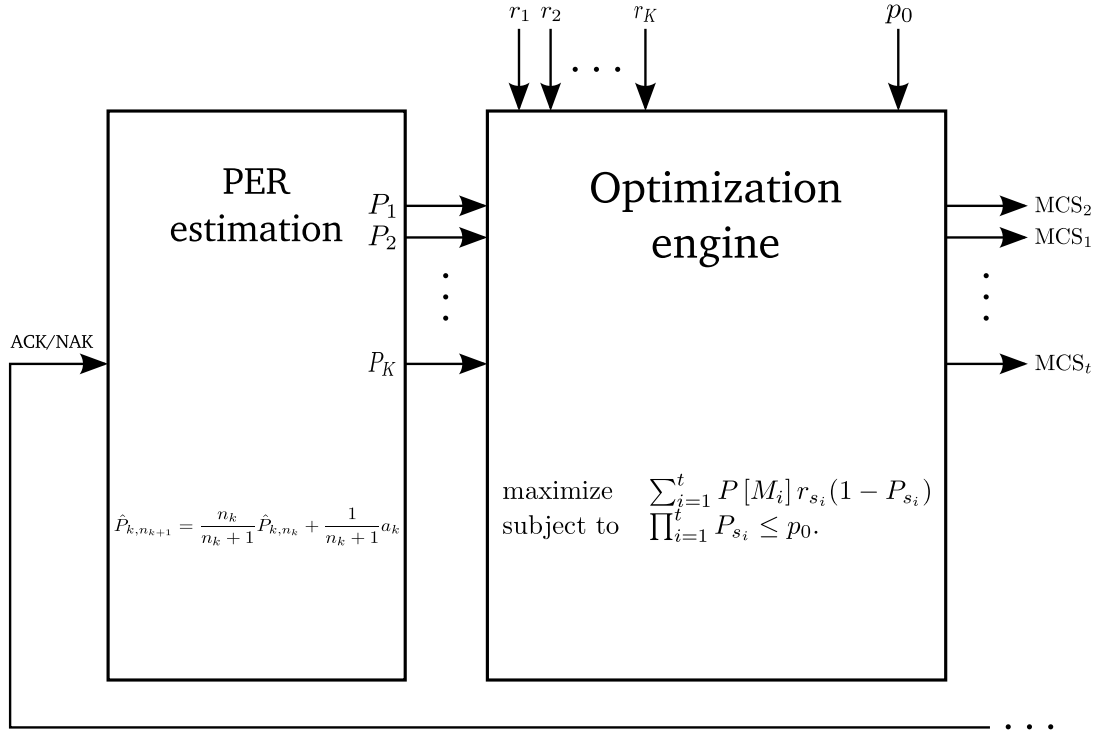
Figure 8.2: Graphic description of the solution for uncorrelated transmissions with online adaptation.

Online adaptation

So far we have been implicitly assuming that the transmitter will know the pdf of γ_{eff} in order to obtain P_k perfectly. But, indeed, the proposed link adaptation algorithm can be easily implemented without this knowledge, in an online fashion. A similar idea was introduced in [117] for link adaptation in MIMO-OFDM systems. Here, the transmit side would estimate directly P_k by just observing the ACK/NAK of the ARQ protocol, as shown in Figure 8.2:

$$\hat{P}_{k,n_{k+1}} = \frac{n_k}{n_k + 1}\hat{P}_{k,n_k} + \frac{1}{n_k + 1}a_k \tag{8.13}$$

with n_k the number of ACK/NAK obsevations for MCS c_k, $a_k = 1(0)$ if an ACK (NAK) is received, and $\hat{P}_{k,n_k}$ the estimation of P_k after n_k packets are received.

A transmitter selecting the MCS following the adaptation algorithm and updating the PER estimations following (8.13) can be easily stuck in a suboptimal solution. If some of the optimum MCS have not been *explored*, the corresponding PER estimation will have a large variance and it can cause the adaptation algorithm not to select that MCS, thus disabling the possibility of improving the estimation. To overcome this issue, we set a fraction ϵ of the packets to be scheduled with a random MCS, so we can improve the estimation of the PER in case an MCS is seldom selected. In our simulations we selected $\epsilon = 0.01$.

Figure 8.3: Graphic description of the solution for uncorrelated transmissions with online adaptation and forgetting factor.

Online adaptation with forgetting factor

In the previous description we assumed that the environment was fixed for the whole operation time. In a realistic scenario, however, infrequent environment changes can be expected. The above described online adaptation algorithm is not suitable for changes, since the update equation (8.13) remains stationary as the number of samples goes to infinity, regardless of any changes in the environment. To overcome this problem, we rewrite (8.13) as (see Figure 8.3)

$$\hat{P}_{k,n_{k+1}} = (1 - \alpha)\hat{P}_{k,n_k} + \alpha a_k \qquad (8.14)$$

with α the forgetting factor. The problem with (8.13) is that as $n_k \to \infty$, $\alpha \to 0$, so the probability estimates are not able to adapt to the new environment. We propose to modify α as

$$\alpha = \max \left\{ \frac{1}{n_k + 1}, \alpha_{min} \right\}. \qquad (8.15)$$

With this definition of α the first steps of the adaptation are the same as in (8.13), but as n_k grows α does not tend to 0, but to α_{min} .

   Remark: For both this option and the former, we are assuming the exchange of ACK/NAK packets between both ends. If this is not true, and package losses are handled with other mechanisms (like timout events), these alternatives would see its performance worsened.
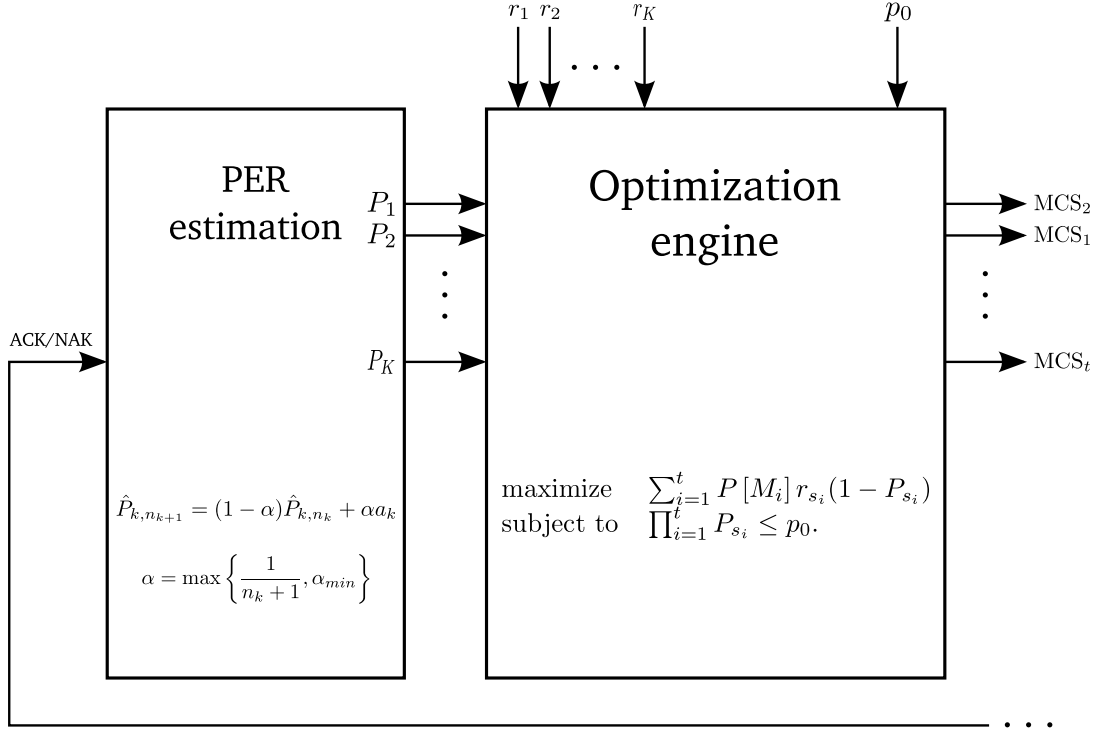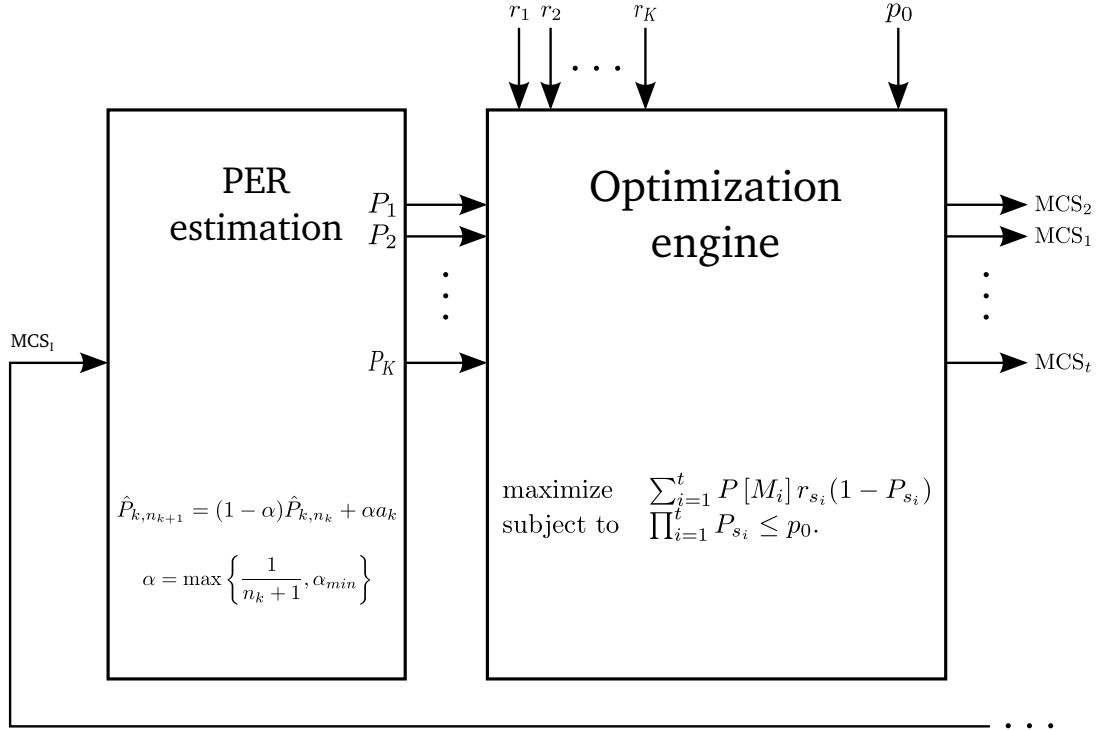
Figure 8.4: Graphic description of the solution for uncorrelated transmissions with online adaptation and MCS feedback.

Online adaptation with MCS feedback

So far, we assumed no processing at the terminal side. But allowing even some minor processing can bring great advantages. As an example, we consider also the case in which the other end feedbacks the best MCS that would have succeeded, that is, the optimum MCS in the case of a single transmission.

The idea, summarized on Figure 8.4, is not to use this MCS directly, but instead exploiting it to update the information of all the MCS: for the ones with a lower SINR threshold, we consider that they would have succeeded and update their PER accordingly; conversely, the MCS with higher thresholds are assumed to have failed.

Remark: Here, we are assuming basic estimation capabilities at the receiver side, which will need to estimate the ESM (based on the pilots sent by the transmitter), infer the corresponding MCS and feed its index back.

8.4 Simulation results

We have simulated the performance of the proposed ARQ scheme for a 3-transmission strategy (2 retransmissions) with an outage probability of 10^{-3}. The chosen central frequency is $f_c = 1550\,\text{MHz}$ with a symbol period $T_{\text{symb}} = 1/(33.6)\,\text{ms}$ and a codeword length of 2688 symbols.

8.4.1   Offline adaptation

In this first approximation, we assume perfect empirical knowledge of the probability density function of γ_{eff}. We compare the results with the case of no retransmissions at all ($t = 1$) under the same outage probability constraint, that could be the constraint in interactive applications, for example. We also compare the proposed approach to a so-called *baseline* method; there, the mobile receiver estimates the SNR by averaging over $10.24\,\text{s}$ (128 frames) and feeds it back to the transmit side. The transmit side selects the MCS by applying a **margin** (usually between $3\,\text{dB}$ and $5\,\text{dB}$) to the received SNR.

Here, we have dragged an independent channel realization for each packet, i.e., removing the channel correlation between different transmissions. This would be the case of the RTT being significantly larger than the channel coherence time. To make a fair comparison, we assume that the baseline knows perfectly the average SNR and that two retransmissions are allowed; then, it selects the MCS by applying a margin of $3\,\text{dB}$ or $5\,\text{dB}$ to that average SNR. In Figures 8.5 and 8.6 we show the evolution of throughput and PER with respect to the LOS SNR. We can see that the introduction of the margin does not suffice to achieve the desired PER, which lays almost an order of magnitude below the obtained results.

Figure 8.5: Comparison with baseline, $1\,\text{m/s}$, independent realizations for each codeword. We denote with B. the baseline mode with a fixed margin.

It is important to remark that the baseline scheme was designed assuming a Rician channel, and that the LMS channel presents a much higher variation because of the existence of shadowing. A first approach for a more fair comparison could be increasing the margins, in order to take into account these higher variations.

However, our preliminary results show that this does not fix the problem: in Figure 8.7 we apply the same adaptation but with a margin of $10\,\text{dB}$ and $12\,\text{dB}$. Although the performance is better than in the previous case, the use of margins is clearly suboptimal: the $12\,\text{dB}$ margin, for example, meets the PER constraint with a value of 10^{-7} around $14\,\text{dB}$ of LOS SNR, but it does not meet the constraint around $20\,\text{dB}$. These results suggest that the variation of the effective
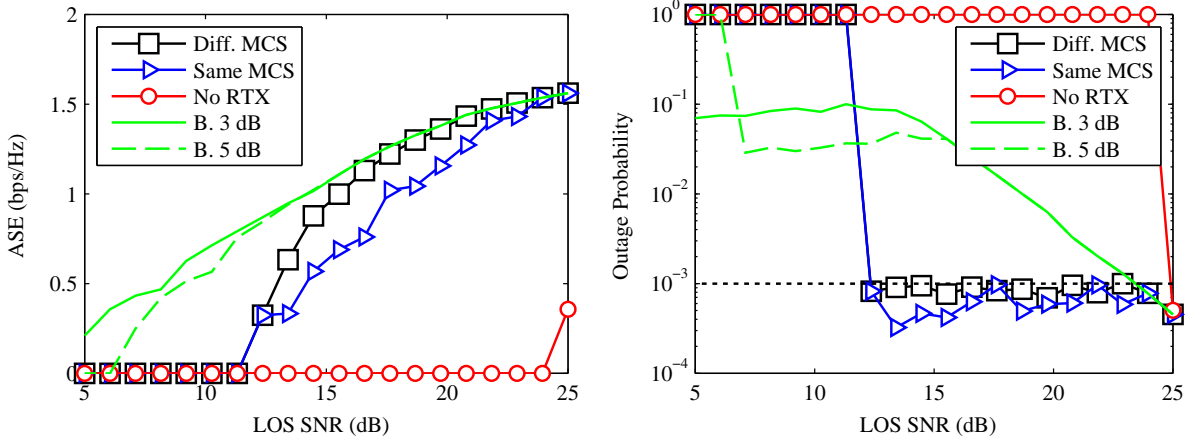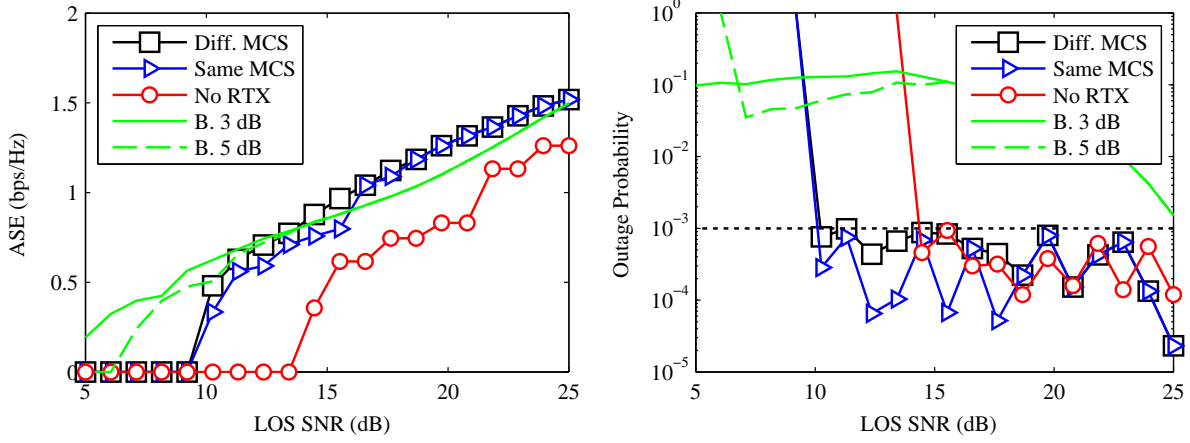
Figure 8.6: Comparison with baseline, 10 m/s, independent realizations for each codeword. We denote with B. the baseline mode with a fixed margin.
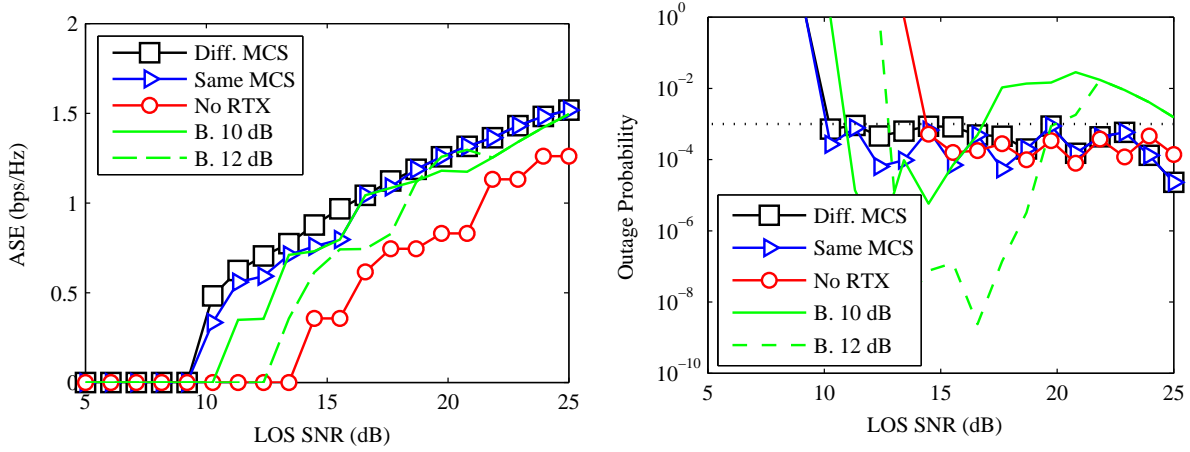
Figure 8.7: Comparison with baseline with large magins, 10 m/s, independent realizations for each codeword. We denote with B. the baseline mode with a fixed margin.

SNR is quite involved to model, and that a margin on the average SNR is not enough to capture its behavior.

It can be seen that the multiple MCS approach outperforms the fixed MCS in all scenarios, specially in the low speed ones due to the lower time diversity. Note that in some cases the fixed MCS is the solution to our problem, leading to the same throughput and outage probability. In the case of not allowing retransmissions the throughput is severely reduced due to the smaller diversity. In fact, for the low speed case, the small diversity introduced by the channel variations causes the system to be unable to meet the outage constraint for almost every SNR.

The solution for 1 m/s for the different points is shown on Table 8.2, where it can be seen that the MCS rate decreases with the transmission index, and that there can be a huge variation between the first and last transmissions. In the simulations we solved the problem by brute force, so these results agree with Conjecture 8.1.

LOS SNR (dB)	$\mathcal{C}_{s_3}$	$\mathcal{C}_{s_2}$	$\mathcal{C}_{s_1}$	$\mathcal{C}_{\text{same MCS}}$
23.9474	QPSK 8/9	QPSK 8/9	QPSK 8/9	QPSK 8/9
21.8421	QPSK 8/9	QPSK 8/9	QPSK 3/4	QPSK 5/6
19.7368	QPSK 8/9	QPSK 8/9	QPSK 2/5	QPSK 2/3
18.6842	QPSK 8/9	QPSK 5/6	QPSK 1/4	QPSK 2/3
17.6316	QPSK 8/9	QPSK 5/6	QPSK 1/4	QPSK 3/5
15.5263	QPSK 3/4	QPSK 3/5	QPSK 1/4	QPSK 2/5
13.4211	QPSK 3/5	QPSK 1/4	QPSK 1/4	QPSK 1/3
12.3684	QPSK 1/4	QPSK 1/4	QPSK 1/4	QPSK 1/4

Table 8.2: Optimum MCS evolution.
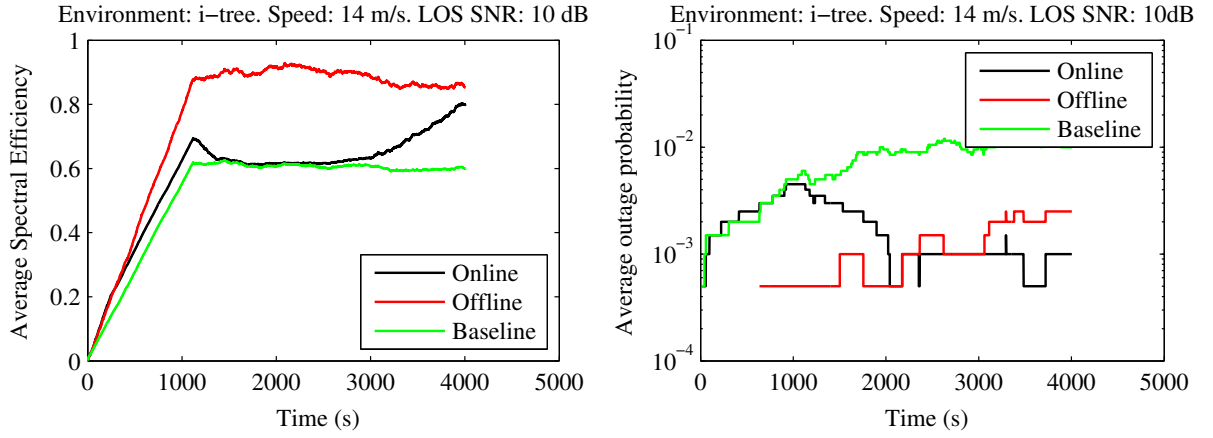
Figure 8.8: Comparison with baseline, online adaptation.

8.4.2 Online adaptation

We show the evolution of the spectral efficiency with time (averaged over a window of 1000 packets) when online adaptation is used following (8.13). As opposed to the results previously shown, here the channel follows its natural time evolution, so that some time correlation will naturally exist at low speeds. For the sake of simplicity, we used *Stop-and-Wait* ARQ for the simulations, so only one packet is transmitted every RTT, although the method can be also applied for more efficient protocols like *Selective Repeat ARQ* [118]. However, if in the latter case ACK packets are sent in groups, the performance of this online scheme could be hindered because it needs to learn the channel performance in the minimum possible time.

Figure 8.8 illustrates the performance in a correlated scenario, compared to the online and offline adaptation procedures. We observe that, once again, the baseline is not able to meet the PER constraint.
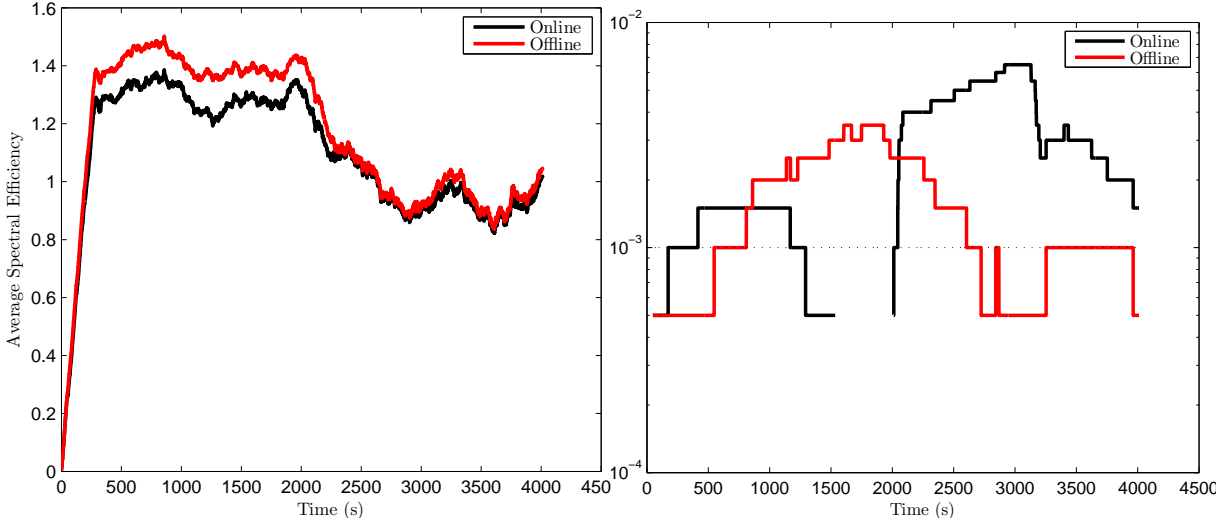
Figure 8.9: Average spectral efficiency (left) and outage probability (right). $\alpha_{\min} : 0$.
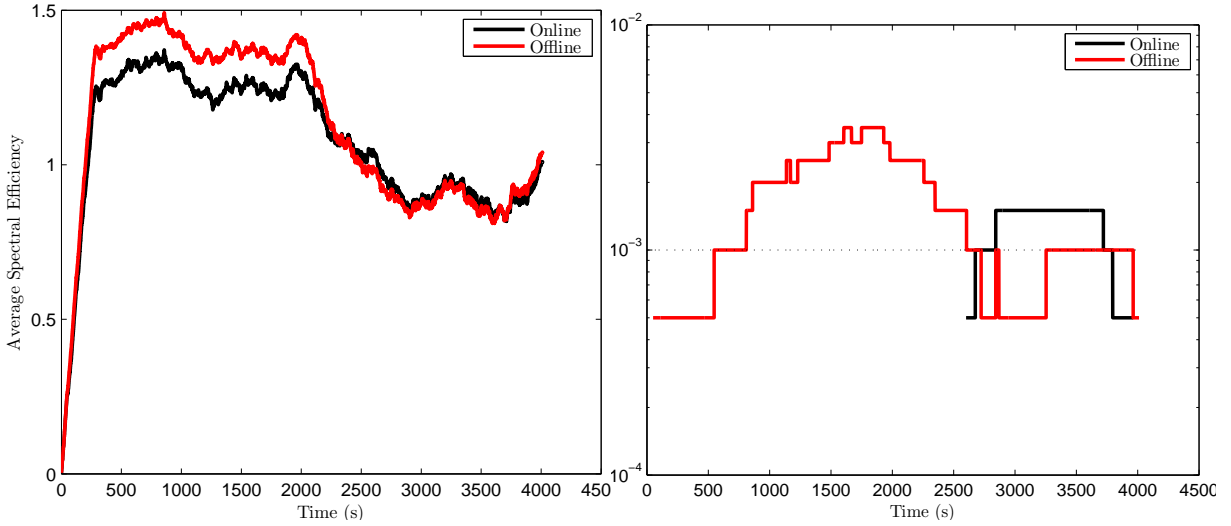
Figure 8.10: Average spectral efficiency (left) and outage probability (right). $\alpha_{\min} : 10^{-2}$.

8.4.3   Online adaptation: change in environment

In Figures 8.9 and 8.10 we show the effect of changing the forget factor α_{min}. We simulate an ITS environment with 20 dB of LOS SNR that changes to 13 dB at $t = 2000\,\text{s}$. In Figure 8.9 we see that the PER dramatically increases when the environment changes, and it remains above the desired threshold for a long time; in Figure 8.10 this effect is no longer observed.

A final example concerns the speed at which the algorithm recovers from a sudden change to a worse state. In Figure 8.11 we simulate the toy example of a constant channel whose power decreases abruptly in 3 dB; right after the change, approximately 30 packets are lost before reaching stability.
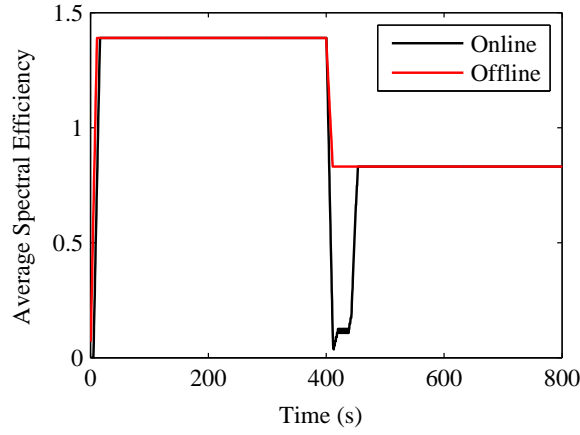
Figure 8.11: ASE over time for the toy example; a sudden drop in SNR occurs at $t = 400$ s.

	SNR drop	Cold start
Conventional	28	28
Alert mode	21	21
MCS feedback	21	7

Table 8.3: Packets lost during the SNR drop and during a cold start assuming 7 packets are transmitted per RTT; codewords span 80 ms.

8.4.4 Online adaptation: alert mode and MCS feedback

For a toy example like the one above, we will now test the two advanced options for online adaptation: alert mode and MCS feedback.

As explained in the previous section, in alert mode the most protected MCS is enforced whenever a burst of packets is lost. Here, we have selected $\alpha_{\min} = 0.99$, and alert mode was activated whenever 3 packets where lost in 30 attempts. Alert mode is also present with MCS feedback but, additionally, the terminal feeds back the best MCS that would have succeeded and not simply the ACK/NAK pair.

Adaptation with MCS feedback is faster, and the packet loss is minimized as shown in Table 8.3 on the left, where we can see the total number of packets lost during the SNR drop. Note that, for computing this table, we assumed that each failure implied the loss of 7 packets, which is the number of packets that fit into an RTT.

Choosing either of these techniques will impact the rate at which we update each MCS information, and this will affect also the convergence of the algorithms and their behavior when the SINR level of the channel improves. As an example, Table 8.3 on the right shows the number of packets loss during a *cold* initialization of the algorithm, in which we assumed no a priori knowledge and all the MCS where assigned the same starting PER; we can see how MCS

feedback clearly outperforms the others. Finally, an interesting observation, from the plots above, is that the obtained PER is slightly over the threshold even in the offline case. This is caused by the fact that a retransmitted packet does not see the same channel distribution as the first transmission due to channel correlation, which is not taken into account by this approach.

8.5   Conclusions

We have derived an optimum strategy to select the MCS to be used in the different retransmissions of a packet. We need the PER performance of each MCS in order to solve the optimization problem, and we have also shown that this PER value can be estimated in a very simple way while still getting good results.

Although this estimation required receiving ACK and NAK packets, any other exchange that led to estimating each MCS's PER in a timely way would do. To this extent, note that many standards work in a *selective acknowledgement mode*: ACKs are received for series of transmitted packets, and retransmissions are ordered indirectly by the expiration of a timer. In such a case, it would be necessary to check if the timings involved still allow a timely packet error rate estimation.

A further improvement on the above solution would require some feedback from the terminal, in the form of a quantized MCS index; this technique has provided the fastest adaptation.

A nice feature of the solutions above is that they do not require knowledge of the terminal speed. However, for speeds below $10\,\text{m/s}$ small discrepancies between the target and obtained PER can occur.

The contents in this chapter were done in collaboration with Dr. Alberto Rico-Alvariño and Prof. Carlos Mosquera. Parts of this work were presented at Globecom 2013 conference [119].

Chapter 9

Conclusions and future work

Contents

9.1 Conclusions

This thesis focuses on techniques to improve the spectral efficiency of next-generation satellite communications. Efforts were directed into two main directions: mitigating the multiuser interference present in the return link of multibeam satellite systems, and adapting to the changing channel conditions in mobile satellite communications.

9.1.1 Interference mitigation in multibeam satellite systems

In Chapter 2 we introduced the system description envisaged for next generation satellite system, along with the most relevant multiuser detectors and their properties. The rest of the chapter focused on analyzing the performance of multiuser detection in a multibeam satellite system, in both a regular scenario of on-ground processing and an alternative one with part of the operations being carried out at the satellite. With respect to the former, an approximation was derived, but the rest of the analysis was left for Chapter 3. With respect to the latter, it was proven that including some processing at the satellite equals, at best, the performance of an architecture with on-ground processing. Moreover, equality is only achieved if the on-board processing is aware of the channel values. The results in this chapter can be useful to help

choosing between one architecture of the other. They also allow to understand the fundamental limitation of on-board processing and hint the line to overcome it.

In Chapter 3 we focused on analyzing the performance, from the outage capacity perspective, of a multibeam satellite system impaired by rain attenuation. The attenuation experienced by the users is correlated, and this correlation depends on their relative distance. Analytical expressions of the outage capacity were obtained at high and low SNR, and then particularized for the single-user case. To the best of our knowledge, these results are new. It was shown that correlation induces noticeable losses on the outage capacity of the system. For example, with $\epsilon = 10^{-4}$, an inter-user distance of $100 \, \mathrm{Km}$ and a beam radius of $100 \, \mathrm{Km}$, the total loss equals about $42 \, \mathrm{bps/Hz}$, which roughly means $0.42 \, \mathrm{bps/Hz}$ in average per user. For the particular case of a single user, single antenna link, results have shown that ensuring an outage probability of 10^{-3} requires an extra power offset of about $15 \, \mathrm{dB}$ for common rain profiles. In terms of outage capacity, this means reaching only $10 \, \%$ of the unfaded capacity at low SNR and $65 \, \%$ at high SNR if we do not increase the power margin. The results in this chapter are useful to properly dimension the typical multibeam satellite system with multiuser detection which is envisaged for the next generation.

Finally, Chapter 4 addresses the problem of having non-perfect channel state information at the receiver, focusing on its effect on the LMMSE receiver. Two error models are studied, the first corresponding to the use of orthogonal training sequences, and the second to non-orthogonal sequences and estimation by correlation. For the first case, a tight approximation for the mean-squared error has been found. Our method relies on an approximation of the error covariance matrix, for which existing results in the field of random matrix theory can be applied. The derived approximation has been proven to work both with uncorrelated and correlated channel estimation errors, and simulation results have shown a remarkable tightness at medium and high SINR. Regarding the second case, a high SNR approximation was also obtained. In this case, an error floor has been shown to appear because of the use of the correlation procedure, bounding the achievable output SINR. The results in this chapter can be used to select an appropriate training sequence length that allows to meet some performance constraint by simply evaluating the obtained expressions.

9.1.2   Link adaptation in mobile satellite communications

Both Chapters 6 and 7 deal with the problem of link adaptation in the return link of mobile satellite systems. In Chapter 6, we tested the applicability of open-loop ACM using, for improved performance monitoring, effective SNR metrics instead of conventional average SINR as CSI. After a careful study of the two alternatives, we carried out a performance test focusing on specific working conditions. Results have shown that, for the scenario under study, the best performance is obtained in an intermediate tree-shadowed environment, achieving remarkable performance gains even for the most compelling interference pattern simulated. On the other end, the suburban environment has been seen to be the least suitable: performance with open-

loop adaptation is worse even with no interference.

In Chapter 7, and building on the results from the previous chapter, we proposed a method for link adaptation in the return link of satellite communications that exploits both open-loop and closed-loop CSI. It works by balancing the importance given to each CSI value based on the observation of the ACK/NAK exchange between both communication ends. The adaptive algorithm is obtained as a stochastic gradient descent of an unconstrained optimization problem. The proposed method is shown to offer a good performance with respect to open loop and closed loop adaptation.

Finally, Chapter 8 focuses on link adaptation in the forward link, where the available CSI is in many cases outdated. Assuming that more than one retransmission is available for each packet, we derived an optimum strategy to select the MCS to be used in the different retransmissions. We need the PER performance of each MCS in order to solve the optimization problem, but we have also shown that this PER value can be estimated in a very simple way while still getting good results. A further improvement on the solution would require some feedback from the terminal, in the form of a quantized MCS index; this technique has been proven to provide the fastest adaptation.

The results in these three chapters are immediately applicable to improve the spectral efficiency of mobile terminals in vehicular conditions, since they require only a minor increase in complexity at the gateway side.

9.2   Future work

9.2.1   Interference mitigation in multibeam satellite systems

- The results in Chapter 2 can be extended to design the optimum (in some sense) on-board beamforming matrix. Since the channel changes over time, and such a beamforming is considered fixed, it would be necessary to design it taking into account the statistics of the channel. Further improvements could be obtained by allowing some form of calibration.

- All the performance measures in Chapter 3 are given at system level. In some applications, however, it is necessary to analyze the performance of individual users. So far, such a problem has been seen as analytically intractable.

- A similar consideration holds for the results in Chapter 4. By focusing on single-user performance metrics, it would be possible to analyze the CSI requirements needed to satisfy, for example, the user experiencing the worse conditions. Another interesting extension would be to drop the assumption of clear-sky conditions.

9.2.2   Link adaptation in mobile satellite communications

- Regarding Chapters 6 and 7, a better characterization of the interference would be useful. With respect to the adaptation algorithm itself, a deeper study of its convergence properties would be useful, too, along with considering imperfect channel information.

- The solution in Chapter 8 is less efficient at low speeds, since in that case the channel experienced by each retransmission can be correlated, violating one of the design assumptions. A possible solution would be to build a Markov model of the channel whose probabilities get updated from estimates, and then obtain the optimum sequence of MCS from the parameters of this Markovian description.

Bibliography

[1] Cisco Systems, "Cisco visual networking index: Global mobile data traffic forecast update, 2013–2018," Tech. Rep., 2014.

[2] J. Pérez-Trufero, S. Watts, G. Peters, B. Evans, T. Frequet, and M. Dervin, "Broadband acess via integrated terrestrial and satellite systems (BATS)," in *Proc. ICSSC*, Florence, Italy, Oct. 2013.

[3] P. Angeletti, F. Coromina, F. Deborgies, R. D. Gaudenzi, A. Ginesi, and A. Vernucci, "SATCOMS 2020 R&D challenges: Part I: Broadband and fixed communications," in *Proc. ICSSC*, Cagliari, Italy, Sep. 2009.

[4] A. Morello, "DVB-Sx: The evolution of the (satellite systems) species," in *Proc. ICSSC*, Florence, Italy, Oct. 2013.

[5] M. Debbah, G. Gallinaro, R. Müller, R. Rinaldo, and A. Vernucci, "Interference mitigation for the reverse-link of interactive satellite networks," in *Proc. SPSC*, Noordwijk, The Netherlands, Sep. 2006.

[6] N. Letzepis and A. Grant, "Capacity of the multiple spot beam satellite channel with rician fading," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5210 –5222, Nov. 2008.

[7] J. Hoydis, M. Kobayashi, and M. Debbah, "Asymptotic performance of linear receivers in network MIMO," in *Proc. ASILOMAR*, Pacific Grove, CA., Nov. 2010, pp. 942 –948.

[8] "Digital video broadcasting (DVB); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications (DVB-S2)," *ETSI EN 302 307 V1.2.1 (2009-08)*.

[9] "Digital video broadcasting (DVB); second generation DVB interactive satellite system (DVB-RCS2); part 1: Overview and system level specification," *ETSI TS 101 545-1 V1.1.1 (2012-05)*.

[10] S. Cioni, R. De Gaudenzi, and R. Rinaldo, "Adaptive coding and modulation for the forward link of broadband satellite networks," in *Proc. IEEE GLOBECOM*, vol. 6, San Francisco, CA, Dec. 2003, pp. 3311–3315 vol.6.

[11] ——, "Channel estimation and physical layer adaptation techniques for satellite networks exploiting adaptive coding and modulation," *International Journal of Satellite Communications and Networking*, vol. 26, no. 2, pp. 157–188, 2008.

[12] ——, "Adaptive coding and modulation for the reverse link of broadband satellite networks," in *Proc. IEEE GLOBECOM*, vol. 2, Dallas, TX, Nov. 2004, pp. 1101–1105 Vol.2.

[13] H. Bischl, H. Brandt, T. de Cola, R. De Gaudenzi, E. Eberlein, N. Girault, E. Alberty, S. Lipp, R. Rinaldo, B. Rislow, J. A. Skard, J. Tousch, and G. Ulbricht, "Adaptive coding and modulation for satellite broadband networks: From theory to practice," *International Journal of Satellite Communications and Networking*, vol. 28, no. 2, pp. 59–111, Mar. 2010.

[14] D. Tarchi, G. E. Corazza, and A. Vanelli-Coralli, "Adaptive coding and modulation techniques for next generation hand-held mobile satellite communications," in *Proc. IEEE ICC*. Budapest, Hungary: IEEE, 2013, pp. 4504–4508.

[15] "The SatNEx III Project." [Online]. Available: http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=30263

[16] S. Verdú, *Multiuser detection*. Cambridge University Press, 1998.

[17] J. Arnau-Yanez, M. Bergmann, E. Candreva, G. Corazza, R. de Gaudenzi, B. Devillers, W. Gappmair, F. Lombardo, C. Mosquera, A. Perez-Neira, I. Thibault, and A. Vanelli-Coralli, "Hybrid space-ground processing for high-capacity multi-beam satellite systems," in *Proc. IEEE GLOBECOM*, Houston, Texas, Dec. 2011, pp. 1–6.

[18] G. Gallinaro, G. Caire, and M. Debbah, "Perspectives of adopting inteference mitigation techniques in the context of broadband multimedia satellite systems," in *Proc. AIAA ICSSC*, Rome, Italy, Sep. 2005, pp. 1–8.

[19] N. Letzepis and A. Grant, "Information capacity of multiple spot beam satellite channels," in *Proc. AusCTW*, Brisbane, Australia, Feb 2005, pp. 168–174.

[20] M. Moher, "Multiuser decoding for multibeam systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1226–1234, Jul. 2000.

[21] J. Arnau, B. Devillers, C. Mosquera, and A. Perez-Neira, "Performance study of multiuser interference mitigation schemes for hybrid broadband multibeam satellite architectures," *EURASIP J. Wirel. Commun. Netw.*, p. 132, 2012.

[22] D. Christopoulos, S. Chatzinotas, M. Matthaiou, and B. Ottersten, "Capacity analysis of multibeam joint decoding over composite satellite channels," in *Proc. ASILOMAR*, Pacific Grove, CA., 2011, pp. 1795–1799.

[23] E. Lutz, M. Werner, and A. Jahn, *Satellite systems for personal and broadband communications*. Springer, 2000.

[24] C. Caini, G. Corazza, G. Falciasecca, M. Ruggieri, and F. Vatalaro, "A spectrum- and power-efficient EHF mobile satellite system to be integrated with terrestrial cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 8, pp. 1315–1325, Oct. 1992.

[25] A. Panagopoulos, P. D. M. Arapoglou, and P. Cottis, "Satellite communications at Ku, Ka, and V bands: Propagation impairments and mitigation techniques," *IEEE Commun. Surveys Tuts.*, vol. 6, no. 3, pp. 2–14, 2004.

[26] ITU-R P.1853, "Tropospheric attenuation time series synthesis," Geneva, 2012.

[27] J. Arnau, D. Christopoulos, S. Chatzinotas, C. Mosquera, and B. Ottersten, "Performance of the multibeam satellite return link with correlated rain attenuation," *IEEE Trans. Wireless Commun.*, 2014.

[28] M. Cheffena, L. Braten, and T. Ekman, "On the space-time variations of rain attenuation," *IEEE Trans. Antennas Propag.*, vol. 57, no. 6, pp. 1771–1782, Jun. 2009.

[29] B. Gremont and M. Filip, "Spatio-temporal rain attenuation model for application to fade mitigation techniques," *IEEE Trans. Antennas Propag.*, vol. 52, no. 5, pp. 1245 – 1256, May 2004.

[30] D. V. B. (DVB), "Second generation DVB interactive satellite system; part 2: Lower layers for satellite standard," 2011.

[31] H. Brandt, O. Lücke, V. Boussemart, C. Párraga-Niebla, T. Flo, C. Kissling, and R. Schweikert, "Resources management using adaptive fade mitigation techniques in DVB-S2/RCS multi-beam systems," in *Proc. ICSSC*, Seoul, Korea, 2007.

[32] C. Studer, A. Burg, and H. Bolcskei, "Soft-output sphere decoding: algorithms and VLSI implementation," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 986–996, Feb. 2008.

[33] B. Shim and I. Kang, "On further reduction of complexity in tree pruning based sphere search," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 417–422, Feb. 2010.

[34] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, Jan. 1999.

[35] A. Rontogiannis, V. Kekatos, and K. Berberidis, "A square-root adaptive V-BLAST algorithm for fast time-varying MIMO channels," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 265–268, May 2006.

[36] M. K. Varanasi, "Decision feedback multiuser detection: a systematic approach," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 219–240, Jan. 1999.

[37] G. Woodward, R. Ratasuk, M. L. Honig, and P. Rapajic, "Minimum mean-squared error multiuser decision-feedback detectors for DS-CDMA," *IEEE Trans. Commun.*, vol. 50, no. 12, Dec. 2002.

[38] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* New York, USA: Cambridge University Press, 2005.

[39] A. Duel-Hallen, "Performance characteristics of cellular systems with different link adaptation strategies," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 4, pp. 630–639, Apr. 1992.

[40] S. Haykin, *Adaptive filter theory.* Prentice Hall, 2002.

[41] J. Choi, *Optimal Combining and Detection: Statistical Signal Processing for Communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2010.

[42] C. Schlegel and A. Grant, *Coordinated multiuser communications.* Springer, 2006.

[43] M. L. Honig, Ed., *Advances in Multiuser Detection.* Wiley, 2009.

[44] M. L. Honig and M. Tsatsanis, "Adaptive techniques for multiuser CDMA receivers," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 49–61, May 2000.

[45] G. Zheng, S. Chatzinotas, and B. Ottersten, "Generic optimization of linear precoding in multibeam satellite systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2308–2320, Jun. 2012.

[46] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE*, Pisa, Italy, Sep 1998, pp. 295–300.

[47] R. C. de Lamare, "Adaptive and iterative multi-branch MMSE decision feedback detection algorithms for multi-antenna systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 5294–5308, Oct. 2013.

[48] F. Rusek, D. Persson, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[49] B. Hassibi, "An efficient square-root algorithm for BLAST," in *Proc. IEEE ICASSP*, vol. 2, Istanbul, Turkey, Jun. 2000, pp. II737–II740.

[50] C. Berrou and A. Glavieux, "Near optimum error-correcting coding and decoding: Turbo codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[51] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.

[52] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[53] J. W. Choi, A. C. Singer, J. Lee, and N. I. Cho, "Improved linear soft-input soft-output detection via soft feedback successive interference cancellation," *IEEE Trans. Commun.*, vol. 58, no. 3, pp. 986–996, Mar. 2010.

[54] J. Wu and H.-N. Lee, "Performance analysis for LDPC-coded modulation in MIMO multiple-access systems," *IEEE Trans. Commun.*, vol. 55, no. 7, pp. 1417–1426, Jul. 2007.

[55] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[56] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends^{TM} Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, Jul. 2004.

[57] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications.* Cambridge University Press, 2011.

[58] N. Letzepis and A. Grant, "Shannon transform of certain matrix products," in *Proc. ISIT*, vol. 1, no. 6, Nice, France, Jun. 2007, pp. 1646–1650.

[59] B. Devillers, A. Pérez-Neira, and C. Mosquera, "Joint linear precoding and beamforming for the forward link of multi-beam broadband satellite systems," in *Proc. IEEE GLOBE-COM*, Houston, Texas, Dec. 2011.

[60] J. Tronc, P. Angeletti, N. Song, M. Haardt, J. Arendt, and G. Gallinaro, "Overview and comparison of on-ground and on-board beamforming techniques in mobile satellite service applications," *Int. J. Satell. Commun. Netw.*, Oct. 2013.

[61] C. Kourogiorgas, A. Panagopoulos, and J. Kanellopoulos, "On the Earth-space site diversity modeling: novel physical-mathematical outage prediction model," *IEEE Trans. Antennas Propag.*, vol. 60, no. 9, pp. 4391–4397, Sep. 2012.

[62] K. Liolis, A. Panagopoulos, P. Cottis, and B. Rao, "On the applicability of MIMO principle to 10-66 GHz BFWA networks: capacity enhancement through spatial multiplexing and interference reduction through selection diversity," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 530–541, Feb. 2009.

[63] F. Lombardo, A. Vanelli-Coralli, E. Candreva, and G. Corazza, "Multi-gateway interference cancellation techniques for the return link of multi-beam broadband satellite systems," in *Proc. IEEE GLOBECOM*, Anaheim, CA, Dec. 2012, pp. 3425–3430.

[64] D. Christopoulos, J. Arnau, S. Chatzinotas, C. Mosquera, and B. Ottersten, "MMSE performance analysis of generalized multibeam satellite channels," *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1332–1335, Jul. 2013.

[65] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261 –1282, Apr. 2005.

[66] O. Oyman, R. Nabar, H. Bölcskei, and A. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2784 – 2795, Nov. 2003.

[67] G. Golub and C. Van Loan, *Matrix Computations*.   Johns Hopkins University Press, 1996.

[68] M. Filip and E. Vilar, "Optimum utilization of the channel capacity of a satellite link in the presence of amplitude scintillations and rain attenuation," *IEEE Trans. Commun.*, vol. 38, no. 11, pp. 1958–1965, Nov. 1990.

[69] D. Christopoulos, S. Chatzinotas, G. Zheng, J. Grotz, and B. Ottersten, "Linear and non-linear techniques for multibeam joint processing in satellite communications," *EURASIP J. Wirel. Commun. Netw.*, p. 162, 2012.

[70] M. Castro and G. Granados, "Cross-layer packet scheduler design of a multibeam broadband satellite system with adaptive coding and modulation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 248–258, jan. 2007.

[71] A. Gharanjik, B. Rao, P.-D. Arapoglou, and B. Ottersten, "Gateway switching in Q/V band satellite feeder links," *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1384–1387, Jul. 2013.

[72] A. Kyrgiazos, B. Evans, P. Thompson, and N. Jeannin, "Gateway diversity scheme for a future broadband satellite system," in *Proc. ASMS & SPSC*, Baiona, Spain, Sep. 2012, pp. 363–370.

[73] K. Liolis, A. D. Panagopoulos, and P. Cottis, "Multi-satellite MIMO communications at Ku-band and above: investigations on spatial multiplexing for capacity improvement and selection diversity for interference mitigation," *EURASIP J. Wirel. Commun. Netw.*, vol. 2007, pp. 16–16, Jan. 2007.

[74] V. Sakarellos, D. Skraparlis, A. Panagopoulos, and J. Kanellopoulos, "Outage performance analysis of a dual-hop radio relay system operating at frequencies above 10 GHz," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3104–3109, Nov. 2010.

[75] A. Panagopoulos, P. D. M. Arapoglou, G. Chatzarakis, J. Kanellopoulos, and P. Cottis, "LMDS diversity systems: a new performance model incorporating stratified rain," *IEEE Commun. Lett.*, vol. 9, no. 2, pp. 145–147, Feb. 2005.

[76] A. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, Nov. 1994.

[77] R. M. Gray, "Toeplitz and circulant matrices: A review," *Found. Trends^{TM} Commun. Inf. Theory*, vol. 2, no. 3, 2005.

[78] B. Holland and M. Ahsanullah, "Further results on a distribution of Meinhold and Singpurwalla," *The American Statistician*, vol. 43, no. 4, pp. 216–219, Nov. 1989.

[79] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*.   New York: Dover, 1964.

[80] L. Luini and C. Capsoni, "The impact of space and time averaging on the spatial correlation of rainfall," *Radio Science*, vol. 47, no. 3, Jun. 2012.

[81] ITU-R P.618-10, "Propagation data and prediction method required for the design of Earth-space telecommunication systems," Geneva, 2009.

[82] J. Arnau and C. Mosquera, "Performance analysis of multiuser detection for multibeam satellites under rain fading," in *Proc. ASMS & SPSC*, Baiona, Spain, Sep. 2012, pp. 197–204.

[83] A. Ishimaru, S. Jaruwatanadilok, J. Ritcey, and Y. Kuga, "A MIMO propagation channel model in a random medium," *IEEE Trans. Antennas Propag.*, vol. 58, no. 1, pp. 178–186, Nov. 2010.

[84] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*, ser. Probability and mathematical statistics. Academic Press, 1981.

[85] A. Papoulis and S. Pillai, *Probability, random variables, and stochastic processes*, 4th ed. McGraw-Hill, 2002.

[86] M. Lopez-Benitez and F. Casadevall, "Versatile, accurate, and analytically tractable approximation for the gaussian Q-function," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 917–922, Feb. 2011.

[87] O. Olabiyi and A. Annamalai, "Invertible exponential-type approximations for the gaussian probability integral $Q(x)$ with applications," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 544 –547, Oct. 2012.

[88] N. Mehta, J. Wu, A. Molisch, and J. Zhang, "Approximating a sum of random variables with a lognormal," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2690 –2699, Jul. 2007.

[89] M. Biguesh and A. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[90] R. Horn and C. Johnson, *Matrix Analysis.* Cambridge University Press, 1990.

[91] M. Bergmann, W. Gappmair, C. Mosquera, and O. Koudelka, "Channel estimation on the forward link of multi-beam satellite systems," in *Personal Satellite Services*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, G. Giambene and C. Sacchi, Eds. Springer Berlin Heidelberg, 2011, vol. 71, pp. 250–259.

[92] Z. Abu-Shaban, B. Shankar, D. Christopoulos, and P.-D. Arapoglou, "Timing and frequency synchronisation for multiuser detection on the return link of interactive mobile satellite networks," in *Proc. AIAA ICSSC*, Firenze, Italy, Oct. 2013.

[93] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *The Annals of Applied Probability*, vol. 17, no. 3, pp. pp. 875–930, Jun. 2007.

[94] J. Arnau and C. Mosquera, "Multiuser detection performance in multibeam satellite links under imperfect CSI," in *Proc. Asilomar*, Pacific Grove, California, Nov. 2012, pp. 468–472.

[95] F. Fontan, M. Vazquez-Castro, C. Cabado, J. Garcia, and E. Kubista, "Statistical modeling of the LMS channel," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1549–1567, Nov. 2001.

[96] "IEEE 802.16m. evaluation methodology for P802.16m-advanced air interface."

[97] A. Monk and L. Milstein, "Open-loop power control error in a land mobile satellite system," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 2, pp. 205 –212, Feb. 1995.

[98] U. Erez, M. Trott, and G. Wornell, "An efficient ARQ scheme with SNR feedback," in *Proc. IZS*, Zurich, Switzerland, Mar. 2008, pp. 88 –91.

[99] J. Arnau and C. Mosquera, "Open loop adaptive coding and modulation for mobile satellite return links," in *Proc. AIAA ICSSC*, Firenze, Italy, Oct. 2013.

[100] C. Loo, "A statistical model for a land mobile satellite link," *IEEE Trans. Veh. Technol.*, vol. 34, no. 3, pp. 122–127, Aug. 1985.

[101] D. Arndt, T. Heyn, J. Konig, A. Ihlow, A. Heuberger, R. Prieto-Cerdeira, and E. Eberlein, "Extended two-state narrowband LMS propagation model for S-band," in *IEEE Int. Symp. Broadband Multimed. Syst. Broadcast.*, Seoul, Korea, Jun. 2012, pp. 1–6.

[102] F. Perez-Fontan, M. Vazquez-Castro, S. Buonomo, J. Poiares-Baptista, and B. Arbesser-Rastburg, "S-band LMS propagation channel behaviour for different environments, degrees of shadowing and elevation angles," *IEEE Trans. Broadcast.*, vol. 44, no. 1, pp. 40–76, Mar. 1998.

[103] F. Gini, M. Luise, and R. Reggiannini, "Cramer-Rao bounds in the parametric estimation of fading radiotransmission channels," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1390–1398, Oct. 1998.

[104] P. Burzigotti, R. Prieto-Cerdeira, A. Bolea-Alamanac, F. Perez-Fontan, and I. Sanchez-Lago, "DVB-SH analysis using a multi-state land mobile satellite channel model," in *Proc. ASMS*, Bologna, Italy, Aug. 2008, pp. 149 –155.

[105] "216715 NEWCOM++ DR.3.2 second report: AMC design for single- and multi-carrier wireless systems."

[106] "216715 NEWCOM++ DR.3.3 final report: AMC design towards next generation wireless systems."

[107] G. Gallinaro, E. Tirro, F. Di Cecca, M. Migliorelli, N. Gatti, and S. Cioni, "Next generation interactive S-band mobile systems. challenges and solutions," in *Proc. ASMS & SPSC*, Baiona, Spain, Sep. 2012, pp. 54–61.

[108] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *Proc. IEEE PIMRC*, vol. 4, Berlin, Germany, Sep. 2005, pp. 2306 –2311.

[109] R. G. Gallager, *Information Theory and Reliable Communication.* New York, NY, USA: John Wiley & Sons, Inc., 1968.

[110] J. Arnau, A. Rico-Alvariño, and C. Mosquera, "Adaptive transmission techniques for mobile satellite links," in *Proc. AIAA ICSSC*, Ottawa, Canada, Oct. 2012.

[111] S. Donthi and N. Mehta, "An accurate model for EESM and its application to analysis of CQI feedback schemes and scheduling in LTE," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3436–3448, Oct. 2011.

[112] SpE, "STYLIST: SW Tool for HYbrid SateLlite / TerrestrIal mobile BroadcaSting SysTem." [Online]. Available: http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=32112

[113] ETSI TS 102 744, "Satellite Component of UMTS (S-UMTS); Family SL satellite radio interface," Oct. 2012, draft.

[114] M. Nakamura, Y. Awad, and S. Vadgama, "Adaptive control of link adaptation for high speed downlink packet access (HSDPA) in W-CDMA," in *Proc. WPMC*, vol. 2, Honolulu, HI, Oct. 2002, pp. 382–386.

[115] T. Cui, F. Lu, V. Sethuraman, A. Goteti, S. P. Rao, and P. Subrahmanya, "Throughput optimization in high speed downlink packet access (HSDPA)," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 474–483, 2011.

[116] A. Rico-Alvariño, J. Arnau, and C. Mosquera, "Balancing closed and open loop CSI in mobile satellite link adaptation," in *Proc. ASMS & SPSC*, Livorno, Italy, Sep. 2014.

[117] S. Yun and C. Caramanis, "Reinforcement learning for link adaptation in MIMO-OFDM wireless systems," in *Proc. IEEE GLOBECOM*, Miami, FL, Dec. 2010, pp. 1–5.

[118] J. F. Kurose and K. W. Ross, *Computer networking.* Pearson Education, 2003, vol. 2.

[119] A. Rico-Alvariño, J. Arnau, and C. Mosquera, "Statistical cross layer adaptation in fast fading mobile satellite channels," in *Proc. Globecom*, Atlanta, Georgia, Dec. 2013, pp. 2822–2827.