UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament de Teoria del Senyal
i Comunicacions

Ph.D. Dissertation

Resource Management Techniques for Sustainable Networks with Energy Harvesting Nodes

Author: Javier Rubio López
Advisor: Dr. Antonio Pascual Iserte

Signal Processing and Communications Group
Department of Signal Theory and Communications
Universitat Politècnica de Catalunya

Barcelona, May 2016

A mis padres y a Roser,

Abstract

In the last years there has been a considerable expansion of the wireless networks jointly with a continuous increase of the number of users. This expansion and the fact that new applications require higher data rates involve a need for a substantial increase of system capacity. In wireless networks, this capacity increase is technically challenging since the resources to be shared among the users are limited. At the same time, in order to be more spectrally efficient, the general trend is to reduce the coverage radius of the access networks. Due to such short distances between transmitters and receivers, the radiated powers can be comparable with or even lower than the powers consumed by the front-end and the baseband stages. In this context, novel strategies for assigning the network resources that take into account all the power consumption sinks should be developed.

On the other hand, it is important to emphasize that battery-powered terminals are becoming broadly used as they provide high mobility to the end users. Additionally, transmitters (base stations in the context of cellular communications) or sensors are sometimes deployed in places where there is no access to the power grid, such as rural areas, making batteries the only mean to be able to offer communications service in those areas. Unfortunately, one of the limiting factors of current technology is the short lifetime of the batteries; the battery technology has not evolved fast enough to cope with the increase in energy consumption associated with the growth of the processing capability of the devices. In wireless sensors networks this can be a serious issue since sensors are usually placed in positions that are difficult or impossible to be accessed, thus, making the process of replacing the batteries very costly. In cellular environments, the telecommunication providers have put a lot of attention on providing good services with enhanced coverage, but this will not be translated into a really perceived added value if the users cannot make use of them due to the mentioned battery limitations. In this regard, it is not enough to become more efficient in terms of energy consumption; instead, a solution that provides sustainable networks is required. Energy harvesting is a technological solution that enables the network devices to recharge the batteries by collecting energy from the environment. Hence, it is a potential technology both to increase the lifetime of battery-powered devices and to reduce the overall carbon footprint, which is also a global major concern nowadays.

This dissertation proposes novel techniques for assigning network resources among the users by considering that the coverage radii are small, implying that some power consumption sinks not considered so far should now be introduced, and by considering that the network devices are battery-powered terminals provided with energy harvesting capabilities. We develop resource

allocation techniques for different scenarios, from the classical single-input single-output settings to more advanced multiple-input multiple-output configurations, and assuming that the channel as well as the battery information at the transmitter are fully or partially known. In this framework, two different configurations in terms of harvesting capabilities are considered. First, we assume that the energy source is external and not controllable, such as solar energy. In this context, the proposed design should adapt to the energy that is currently being harvested. We also study in this context the effect of having a finite backhaul connection that links the wireless access network with the core network, optimizing the resources according to some backhaul capacity limitation. On the other hand, we propose a design in which the transmitter feeds actively the receivers with energy by transmitting signals that receivers use for recharging their batteries. In this case, the power transfer design should be carried out jointly with the power control strategy for users that receive information as both procedures, transfer of information and transfer of power, are implemented at the transmitter and make use of a common resource, i.e., power.

Apart from techniques for assigning the radio resources, this dissertation develops a procedure for switching on and off base stations. Concerning this, it is important to notice that one characteristic of the traffic profile is that it is not constant throughout the day; usually larger traffic demands are required during day hours and smaller ones during night periods. This is precisely the feature that can be exploited to define a strategy based on a dynamic selection of the base stations to be switched off when the traffic load is low, without affecting the quality experienced by the users. Thanks to this procedure, we are able to deploy smaller energy harvesting sources, e.g., solar panels, and smaller batteries and, thus, to reduce the cost of the equipment which directly translates into a reduction of the capital expenditures of the network deployment.

Finally, we derive some procedures to optimize high level decisions of the network operation in which variables from several layers of the protocol stack are involved. In this context, admission control procedures for deciding which user should be connected to which base station are studied and evaluated, based on the expected network aggregate throughput, the average channel information, and the current battery levels, among others. A multi-tier multi-cell scenario is assumed in which base stations belonging to different tiers have different capabilities, e.g., transmission power, battery size, end energy harvesting source size. A set of strategies that require different computational complexity are derived for scenarios with different user mobility requirements.

Resumen

En los últimos años ha habido una expansión de las redes sin cables (en inglés *wireless*) a la vez que el número de usuarios ha incrementado. Esta expansión y el hecho de que las nuevas aplicaciones necesitan una mayor velocidad de transmisión, requieren incrementar la capacidad de las redes *wireless* de forma sustancial. En las redes *wireless*, este incremento de capacidad plantea un reto importante ya que los recursos disponibles de éstas son compartidos entre los distintos usuarios y son limitados. Además, la tendencia actual para ser eficientes en el uso del espectro, es la de reducir los radios de coberturas de las redes de acceso *wireless*. Debido a la corta distancia entre los transmisores y los receptores, las potencias radiadas pueden ser comparables o incluso más bajas que las potencias consumidas por las etapas de radio-frecuencia y de procesado de banda base. En este contexto, es necesario desarrollar estrategias innovadoras que asignen los recursos disponibles en las redes *wireless* teniendo en cuenta todas las fuentes de consumo de potencia y no sólo la potencia radiada como se ha hecho hasta ahora.

Por otro lado, es también importante enfatizar que los receptores con fuentes de alimentación finitas, es decir baterías finitas, están siendo utilizados en muchas aplicaciones, como teléfonos móviles, etc. debido a que proporcionan una alta movilidad a los usuarios finales. Además, los transmisores (estaciones base en el contexto de comunicaciones móviles) y los sensores son a menudo desplegados en lugares donde no hay acceso a la red eléctrica, como en zonas rurales, haciendo de las baterías el único medio para proporcionar energía y así poder ofrecer servicio en esas areas. Desafortunadamente, uno de los factores más restrictivos de la tecnología actual es la corta duración de las baterías; la tecnología de las baterías no ha sido desarrollada lo suficientemente rápido como para compensar el incremento de energía consumida en los dispositivos asociado al aumento de capacidad de procesado de los mismos. En redes de sensores esto puede ser un gran problema debido a que los sensores son desplegados en numerosas ocasiones en zonas de difícil acceso, lo que hace que el reemplazo de las baterías sea muy costoso. En las redes celulares, los proveedores de servicios de telecomunicaciones han puesto mucha atención en proporcionar servicios de alta calidad pero esto no se traduce a un valor añadido percibido si los usuarios no pueden hacer uso de ellos debido a los problemas de batería mencionados anteriormente. En este sentido, no es suficiente con ser más eficiente desde un punto de vista de la energía consumida, se necesita una solución que sea sostenible.

Sistemas de captación de energía (en inglés *energy harvesting*) son una solución tecnológica que permiten que los diferentes dispositivos de la red puedan recargar sus baterías recolectando energía del medio ambiente (por ejemplo, mediante placas solares). Esto lo convierte en una

tecnología potencial para incrementar la duración de las baterías de los dispositivos y para reducir el impacto medioambiental de las redes de telecomunicaciones que es otra de las grandes preocupaciones hoy en día.

Esta tesis doctoral propone técnicas innovadoras para asignar los recursos disponibles en las redes *wireless* considerando que los radios de cobertura son pequeños, lo que implica que otras fuentes de consumo de energía no consideradas hasta ahora se tienen que introducir en los diseños, y considerando que los dispositivos están alimentados con baterías finitas y que tienen a su disposición fuentes de *energy harvesting*. En esta tesis se desarrollan técnicas de asignación de recursos para diferentes escenarios, desde el clásico escenario *single-input single-output* hasta configuraciones más avanzadas como *multiple-input multiple-output*, y asumiendo que el canal de comunicaciones y la información de la batería pueden ser conocidos de forma completa o parcial a la hora de hacer el diseño de asignación de recursos. En este contexto, se consideran dos configuraciones diferentes en función de las capacidades de *energy harvesting*. En primer lugar, se asumirá que la fuente de energía es externa e incontrolable, como por ejemplo la energía solar. Los diseños propuestos deben adaptarse a la energía que está siendo recolectada en ese preciso momento. Para este escenario, también se estudia el efecto de tener una conexión de *backhaul* con capacidad finita. En segundo lugar, se propone un diseño en el cual el transmisor es capaz de enviar energía a los receptores mediante señales de radio-frecuencia diseñadas para ese fin, energía que es utilizada para recargar las baterías. En este ámbito, la estrategia de asignación de recursos debe realizarse de forma conjunta entre los usuarios que reciben información y los que reciben energía ya que comparten un recurso de la red común como es la potencia.

Aparte de técnicas de asignación de recursos radio, en esta tesis doctoral se desarrolla un procedimiento dinámico para apagar y encender estaciones base. En este ámbito, es importante notar que una de las características del perfil de tráfico es que no es constante a lo largo del día. Usualmente, la demanda de tráfico es mayor durante las horas diurnas y menor durante las horas nocturnas. Este es precisamente el patrón que se puede explotar para definir una estrategia dinámica con la que decidir qué estaciones base deben ser apagadas cuando la demanda de tráfico es baja, todo ello sin afectar a la calidad experimentada por los usuarios. Gracias a este procedimiento, es posible desplegar fuentes de *energy harvesting* más pequeñas, es decir, paneles solares, y baterías más pequeñas. Gracias a esto, es posible reducir el coste de los equipos, lo que se traduce en una reducción en los costes de inversión del despliegue de la red.

Finalmente, esta tesis doctoral presenta procedimientos para optimizar decisiones de nivel más alto que afectan directamente al funcionamiento global de la red de acceso. Para tomar estas decisiones, se hace uso de diversas variables que pertenecen a diferentes capas de la pila de protocolos. En particular, esta tesis aborda el diseño de técnicas de control de admisión de usuarios a estaciones base en entornos con múltiples estaciones base, basándose en la velocidad de transmisión agregada obtenida en la red, la información estadística de los canales, y el nivel actual de las baterías, entre otros. Como se ha comentado, el escenario considerado está

compuesto de múltiples estaciones base, donde cada estación base pertenece a una familia y donde cada una de estas familias presenta diferentes capacidades, como por ejemplo, potencia de transmisión, tamaño de batería, o tamaño de la fuente de *energy harvesting*. En esta tesis se derivan un conjunto de técnicas con diferentes cargas computacionales que son de utilidad para aplicar a escenarios con diferente movilidad de usuarios, siendo las técnicas que requieren menos carga computacional más idóneas para escenarios donde la movilidad de usuarios es un factor determinante.

Resum

Durant els últim anys hi ha hagut una expansió de les xarxes sense cables (*wireless* pel seu terme en anglès) i, de la mateixa manera, ha augmentat el nombre d'usuaris. Aquesta expansió, justament amb el fet que les noves aplicacions necessiten una velocitat de transmissió més elevada, requereixen incrementar la capacitat de les xarxes *wireless* de forma substancial. En les xarxes *wireless*, aquest increment de la capacitat planteja un repte important ja que els recursos disponibles d'aquestes són compartits entre diferents usuaris i són limitats. A més a més, la tendència actual per poder ser eficients en l'ús de l'espectre, és la de reduir els radis de cobertura de les xarxes d'accés *wireless*. Degut a la curta distancia entre els transmissors i els receptors, les potencies irradiades solen ser comparables, o inclús més baixes, que les potencies consumides per les etapes de radio-freqüència i de processat de banda base. En aquest àmbit, és necessari desenvolupar estratègies innovadores que assignin els recursos disponibles a les xarxes *wireless*, tenint en compte, totes les fonts de consum de potència i, no, tant sols la potència irradiada, com s'ha fet fins ara.

Per altra banda, és també important emfatitzar que els receptors amb fonts d'alimentació finites, és a dir, bateries finites, s'utilitzen en moltes aplicacions, com per exemple telèfons mòbils, degut a que proporcionen una alta mobilitat als usuaris finals. A més a més, els transmissors (estacions base en el context de comunicacions mòbils) i els sensors són molts cops desplegats en llocs on no es té accés a la xarxa elèctrica, per exemple en zones rurals, fent de les bateries l'únic mitjà per proporcionar energia i, així, poder oferir-hi els serveis. Malauradament, un dels factors més restrictius de la tecnologia actual és la curta duració de les bateries; la tecnologia de les bateries no ha estat desenvolupada suficientment ràpid com per poder compensar l'increment d'energia consumida en els dispositius, associada a l'augment de capacitat de processat del senyal. En xarxes de sensors això pot ser un gran problema degut a que els sensors són desplegats en nombroses ocasions en zones de difícil accés, el que dificulta el reemplaçament de les bateries. En les xarxes cel·lulars, els proveïdors de serveis de telecomunicacions han posat molta atenció en proporcionar serveis d'alta qualitat, però això no és tradueix en un valor afegit percebut si els usuaris no poden fer ús dels serveis degut als problemes de bateria mencionats anteriorment. En aquest sentit, no és suficient el fet de ser més eficient des d'un punt de vista d'energia consumida, es necessita una solució que sigui sostenible.

Sistemes de captació d'energia (*energy harvesting* pel seu terme en anglès) són una solució tecnològica que permet que els diferents dispositius de la xarxa puguin recarregar les bateries recol·lectant energia del medi ambient (per exemple, mitjançant plaques solars). Això ho con-

verteix en una tecnologia en potència per incrementar la duració de les bateries dels dispositius, i per a reduir l'impacte mediambiental de les xarxes de telecomunicacions, que és una de les grans preocupacions d'avui en dia.

Aquesta tesis doctoral proposa tècniques innovadores per assignar els recursos disponibles a les xarxes *wireless* considerant que els radis de cobertura són petits, el que implica que altres fonts de consum d'energia no considerades fins al moment s'hagin d'introduir dins els dissenys, i considerant que els dispositius estan alimentats amb bateries finites i que tenen a la seva disposició fonts de *energy harvesting*. En aquesta tesis es desenvolupen tècniques d'assignació de recursos per a diferents escenaris, des del clàssic escenari *single-input single-output*, fins a les configuracions més avançades com *multiple-input multiple-output*, i també assumint que el canal de comunicacions i la informació de la bateria poder ser coneguts de forma completa o parcial a l'hora de fer el disseny d'assignació de recursos. En aquest context, es consideren dues configuracions diferents en funció de les capacitats de l'*energy harvesting*. En primer lloc, s'assumirà que la font d'energia és externa i incontrolable com, per exemple, l'energia solar. Els dissenys proposats han d'adaptar-se a l'energia que s'està recol·lectant en un precís moment. Per aquest escenari, també s'estudia l'efecte de tenir una connexió de *backhaul* amb capacitat finita, el qual enllaça la xarxa d'accés *wireless* amb la xarxa troncal de l'operador. En segon lloc, es proposa un disseny en el qual el transmissor és capaç d'enviar energia als receptors mitjançant senyals de radiofreqüència dissenyats per aquest fi, energia que és utilitzada per recarregar les bateries. En aquest àmbit, l'estratègia d'assignació de recursos s'ha de realitzar de forma conjunta entre els usuaris que reben informació i els que reben energia ja que comparteixen un recurs de la xarxa comú, la potència.

A part de tècniques d'assignació de recursos radio, en aquesta tesis doctoral es desenvolupa un procediment dinàmic per apagar i encendre estacions base. En aquest àmbit, és important notar que una de les característiques del perfil de tràfic és que no és constant al llarg del dia. Usualment, la demanda de tràfic és superior durant els hores diürnes i inferior durant les hores nocturnes. Aquest és precisament el patró que es pot explotar per definir una estratègia dinàmica per poder decidir quines estaciones base han de ser apagades quan la demanda de tràfic és baixa, tot això sense afectar la qualitat experimentada pels usuaris. Gràcies a aquest procediment, es possible desplegar fonts d'*energy harvesting* més petites, és a dir, panells solars, i bateries més petites. Això implica que és possible reduir el cost dels equips, fet que es tradueix en una reducció dels costos d'inversió del desplegament de la xarxa.

Finalment, aquesta tesis doctoral presenta procediments per optimitzar decisions de nivell més alt que afecten directament al funcionament global de la xarxa d'accés. Per prendre aquestes decisions, es fa ús de diverses variables que pertanyen a diferents capes de la pila de protocols. En aquest context, aquesta tesis aborda el disseny de tècniques de control d'admissió d'usuaris a estacions base en entorns amb múltiples estacions base, basant-se amb la velocitat de transmissió agregada obtinguda a la xarxa, la informació estadística dels canals, i el nivell actual de les

bateries, entre altres. Com ja s'ha comentat anteriorment, l'escenari considerat està format per múltiples estacions base, on cada estació base pertany a una família, i on cada una d'aquestes famílies presenta diferents capacitats, per exemple, potència de transmissió, mida de la bateria, o mida de la font d'*energy harvesting*. En aquesta tesis es deriven un conjunt de tècniques amb diferents costos computacionals que són d'utilitat per a poder aplicar a escenaris amb diferents mobilitats d'usuaris; sent les tècniques que requereixen menys cost computacional més idònies per escenaris on la mobilitat d'usuaris és un factor determinant.

Agradecimientos

Y todo llega a su fin. Con estas líneas se pone fin a una aventura que comenzó hace poco más de tres años, tiempo que he necesitado para madurar en mi carrera investigadora y poder desarrollar todo el contenido presente en este documento. La elaboración de una tesis no es un camino fácil, uno se tiene que doctorar en muchos ámbitos, no sólo en el profesional. Aunque siendo sincero, y a pesar de que estoy a punto de doctorarme, escribir estas líneas está siendo la tarea más difícil que he tenido que superar en todo este tiempo. Me gustaría dedicar estas líneas que ponen fin a esta etapa a todos los que han contribuido de alguna manera u otra a que esta aventura haya sido como ha sido ya que no cambiaría absolutamente nada de lo transcurrido en ella. Es por ello que estas líneas van dedicadas a todas y cada una de esas personas.

En primer lugar me gustaría dar las gracias a mi director de tesis, Toni Pascual, por su ayuda incondicional e incesante durante todo este tiempo. Gracias por creer en mi y brindarme esta oportunidad. Es un placer poder decir que tu puerta siempre estuvo abierta para charlar sobre cualquier tema, sin importar la hora ni el día de la semana. Sin duda, el trabajo de este documento también es tuyo y no tendría ni la forma ni el contenido que tiene si no fuese por ti. Gracias. Espero que tú también te lleves un buen recuerdo de mi etapa como doctorando. Estoy convencido de que nuestra amistad durará de por vida, sin importar donde estemos.

Considero que he tenido la gran suerte de pertenecer a un extraordinario grupo de investigación. Compañeros de trabajo con los que he compartido vivencias, desarrollo de proyectos, y buenos momentos. He podido trabajar con ellos y aprender de ellos. Espero que ellos también hayan aprendido algo de mi. Al menos, lo intenté. Gracias a todos: Sandra Lagén, Jaume Del Olmo, Olga Muñoz, Toni Pascual, Josep Vidal, Adrián Agustín, y Juan Fernández. No me puedo olvidar tampoco de los compañeros del proyecto NEWCOM# del CTTC: Maria Gregori, Miquel Payaró, y Jesús Gómez.

Especial agradecimiento para mis compañeros de despacho, que en mayor o menor medida me han ayudado siempre que lo he necesitado. En particular: Sandra Lagén, Adriano Pastore, Màrius Caus, Josep Font, Eva Lagunas, Marc Torrellas, Juanma Castro, Jaume Del Olmo, y Pere Giménez. Espero haber sido útil cuando solicitasteis mis consejos. A los que ya os fuisteis, desearos todo lo mejor, a los que estáis a punto de terminar, mucho ánimo para la recta final (que además hace cuesta arriba), y a los que recién comenzáis, desearos toda la suerte del mundo. Trabajad duro y veréis como llegan los resultados.

Me gustaría agradecer el esfuerzo y la dedicación que han mostrado los profesores externos a mi grupo con los que he podido colaborar en algún momento determinado. Mis agradecimientos van a Andrea Goldsmith, Daniel P. Palomar, y Antonio G. Marqués. Me gustaría dar un agradecimiento especial a la profesora Andrea Goldsmith, por acogerme en su grupo de investigación y tratarme como un miembro más durante mi estancia en la *Stanford University*. Sin duda una experiencia inolvidable. Allí tuve la oportunidad de conocer gente extraordinaria con la que pasé buenísimos momentos. Gracias a todos: Mahnoosh Alizadeh, Nima Soltani, Mainak Chowdhury, Jinyuan Chen, Milind Rao, Alon Kipnis, Rui Song, Stefano Rini, y Marco Mondelli, y a la pequeña familia española Idoia Ochoa, Mikel Hernaez, y Javier García. *What a wonderful time we had exploring the bay area with the bikes! Thanks Mainak and Nima.* Tampoco me puedo olvidar de la maravillosa familia que me acogió durante esos meses. La familia Osborn, compuesta por Elizabeth, Henry, Margaret, y Paul, vivía en una preciosa casa en la ciudad de Menlo Park. Gracias por cuidarme como si fuera uno de los vuestros.

No me puedo olvidar de dar las gracias a mis amigos de toda la vida. Con los que comparto los buenos y malos momentos de mi vida personal. En especial: Carlos Rubio, Israel Díaz, Hugo Hernández, Víctor Moreno, y a mis dos buenos amigos de la carrera Carlos Serra y Arnau Raventós. Algunos de vosotros estáis inmersos en cambios importantes en vuestra vidas. Sólo desearos todo lo mejor. Ya sabéis que podéis contar conmigo al igual que yo también he contado con vosotros durante todos estos años. Gracias por aguantarme (que no es fácil).

Last, but not least, me gustaría dar un especial agradecimiento muy merecido a mi padres Montse y Javier, y a mi pareja Roser, y dedicar esta tesis a ellos por su apoyo siempre que lo he necesitado y por hacerme tan feliz durante todos estos años. Estoy convencido de que todos ellos están contentos de que haya llegado este momento. Quiero que sepáis, que sin vosotros, esta aventura no hubiera tenido ni el color ni la forma que tiene. *I love you all.*

Javier Rubio López
Barcelona, Mayo 2016

ANEXO

FINANCIACIÓN RECIBIDA DURANTE EL DESARROLLO DE LA TESIS DOCTORAL

El trabajo de investigación que ha dado lugar a la tesis doctoral "Resource Management Techniques for Sustainable Networks with Energy Harvesting Nodes" ha recibido financiación del Grup de Processament del Senyal i Comunicacions (SPCOM-GPS) de la Universitat Politècnica de Catalunya (UPC), del Ministerio de Economía y Competitividad de España a través de las becas FPI BES-2012-052850 y EEBB-I-15-09380 y a través de los proyectos nacionales GRE3N-LINK-MAC (TEC2011-29006-C03-02) y DISNET (TEC2013-41315-R), de la Comisión Europea a través de los proyectos Network of Excellence in Wireless COMmunications NEWCOM# (no. 318306) y TUCAN3G (no. ICT-2011-601102), del programa Europeo de Cooperación en Ciencia y Tecnología a través del proyecto COST Action IC0902, y del Govern de la Generalitat de Catalunya (AGAUR) a través de las ayudas 2009 SGR 891 y 2014 SGR 60.

Barcelona, mayo de 2016.

Contents

Notation

$\mathbb{R}$, $\mathbb{C}$	The set of real and complex numbers, respectively.
$\mathbb{R}_+$	The set of nonnegative real numbers.
$\mathbb{R}^n$, $\mathbb{R}^n_+$, $\mathbb{C}^n$	The set of vectors of dimension n with entries in $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{C}$, respectively.
$\mathbb{R}^{n\times m}$, $\mathbb{R}^{n\times m}_+$, $\mathbb{C}^{n\times m}$	The set of $n \times m$ matrices with entries in $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{C}$, respectively.
x	Scalar.
$\mathbf{x}$	Column vector.
$\mathbf{X}$	Matrix.
$\mathcal{X}$	Set.
$\lvert\mathcal{X}\rvert$	Cardinality of set $\mathcal{X}$.
$[\mathbf{x}]_n$	The n-th component of vector $\mathbf{x}$.
$[\mathbf{X}]_{pq}$	Element in the p-th row and q-th column of matrix $\mathbf{X}$.
$[\mathbf{X}]_{:,q}$	The q-th column of matrix $\mathbf{X}$.
$\mathbf{I}_n$	Identity matrix of order n. The dimension n might be omitted when it can be deduced from the context.
$\mathbf{1}_n$, $\mathbf{0}_n$	Column vector of n ones or zeros. The dimension n might be omitted when it can be deduced from the context.
$(\cdot)^T$	Transpose operator.
$(\cdot)^*$	Complex conjugate operator.
$(\cdot)^H$	Complex conjugate and transpose (Hermitian) operator.
$\mathrm{Tr}(\mathbf{X})$	Trace of matrix $\mathbf{X}$.
$\lvert\mathbf{X}\rvert$ or $\det(\mathbf{X})$	Determinant of matrix $\mathbf{X}$.
$\mathrm{Diag}(\mathbf{x})$	Diagonal matrix where the diagonal entries are given by the vector $\mathbf{x}$.
$\lvert\lvert x\rvert\rvert$	Vector norm.
$\lvert\lvert\mathbf{X}\rvert\rvert_F$	Frobenius norm of matrix $\mathbf{X}$: $\lvert\lvert\mathbf{X}\rvert\rvert_F = \sqrt{\mathbf{Tr}(\mathbf{X}^H\mathbf{X})}$
$\mathbf{X}^{-1}$	Inverse of matrix $\mathbf{X}$.

$\mathbf{X}^{1/2}$	Hermitian square root of the positive semidefinite matrix $\mathbf{X}$, i.e., $\mathbf{X} = \mathbf{X}^{1/2}\mathbf{X}^{1/2}$.
$\mathbf{S}^n$	The set of Hermitian $n \times n$ matrices $\mathbf{S}^n \triangleq \{\mathbf{X} \in \mathbb{C}^{n\times n} \vert \mathbf{X} = \mathbf{X}^H\}$.
$\mathbf{S}_+^n$	The set of Hermitian positive semidefinite $n \times n$ matrices $\mathbf{S}_+^n \triangleq \{\mathbf{X} \in \mathbb{C}^{n\times n} \vert \mathbf{X} = \mathbf{X}^H \succeq \mathbf{0}\}$.
$\mathbf{S}_{++}^n$	The set of Hermitian positive definite $n \times n$ matrices $\mathbf{S}_{++}^n \triangleq \{\mathbf{X} \in \mathbb{C}^{n\times n} \vert \mathbf{X} = \mathbf{X}^H \succ \mathbf{0}\}$.
$\mathbf{X} \succeq \mathbf{0}$	Matrix $\mathbf{X}$ is positive semidefinite.
$\mathbf{X} \succ \mathbf{0}$	Matrix $\mathbf{X}$ is positive definite.
$\mathbf{X} \succeq \mathbf{Y}$	$\mathbf{X} - \mathbf{Y}$ is positive semidefinite.
$\mathbf{X} \succ \mathbf{Y}$	$\mathbf{X} - \mathbf{Y}$ is positive definite.
$\mathrm{vec}(\mathbf{X})$	Vec operator: if $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, then $\mathrm{vec}(\mathbf{X})$ is the column vector $[\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T]^T$.
$\lambda_{\max}(\mathbf{X})$	Maximum eigenvalue of the positive semidefinite matrix $\mathbf{X}$.
$\mathbf{u}_{\max}(\mathbf{X})$	Normalized eigenvector associated with the maximum eigenvalue of the positive semidefinite matrix $\mathbf{X}$.
$\mathbf{x} \succeq \mathbf{0}$	Componentwise greater, i.e., if $\mathbf{x} = (x_1, \ldots, x_n)$, then $x_i \geq 0$, $i = 1, \ldots, n$.
$\vert x \vert$	Modulus of the complex scalar x.
$(x)^+$	Projection of x onto the nonnegative orthant, i.e., $(x)^+ = \max\{0, x\}$.
$(x)_a^b$	Projection of x onto the interval $[a, b]$, i.e., $(x)_a^b = \min\{\max\{a, x\}, b\}$.
$\sim$	Distributed according to.
$\Pr(\cdot)$	Probability.
$\mathbb{E}[\cdot]$	Mathematical expectation.
$\mathcal{N}(\mathbf{m}, \mathbf{C})$	Real Gaussian vector distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$.
$\mathcal{CN}(\mathbf{m}, \mathbf{C})$	Complex circularly symmetric Gaussian vector distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$.
$\otimes$	Kronecker product.
$\mathrm{Re}\{\cdot\}$	Real part.
$\mathrm{Im}\{\cdot\}$	Imaginary part.
$\triangleq$	Defined as.

$\approx$	Approximately equal.
arg	Argument.
max, min	Maximum and minimum.
sup, inf	Supremum and infimum.
$x^\star$	Optimum value of variable x in an optimization problem.
$\mathrm{dom} f$	Domain of function f.
$f(x)\Big\vert_{x=a}$	Function $f(x)$ evaluated at $x = a$.
$\sup_{\mathbf{x}} f$, $\inf_{\mathbf{x}} f$	Supremum and infimum of f with respect to $\mathbf{x}$.
$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$, $\nabla_{\mathbf{x}} f(\mathbf{x})$	Gradient of function f with respect to $\mathbf{x}$.
$\frac{\partial f(\mathbf{x})}{\partial x_i}$	Partial derivative of function f with respect to x_i.
$\frac{\partial f(x)}{\partial x}$, $\dot{f}(x)$	Derivative of function $f(x)$.
$\nabla^2_{\mathbf{x}} f(\mathbf{x})$	Hessian matrix of function f with respect to $\mathbf{x}$.
$\int_a^b f(x)\,\mathrm{d}x$	Integral of $f(x)$ with respect to x in the interval $[a, b]$.
lim	Limit.
$\exp(\cdot)$	Exponential function.
$\log(\cdot)$	Base-2 logarithm.
$\mathbb{1}_{\{x\}}$	Indicator function defined as $\mathbb{1}_{\{x\}} = 1$ if x is true; otherwise $\mathbb{1}_{\{x\}} = 0$.

Acronyms

ADC	Analog to Digital Converter.
AWGN	Additive White Gaussian Noise.
BD	Block Diagonalization.
BER	Bit Error Rate.
BS	Base Station.
CAPEX	Capital Expenditures.
CDF	Cumulative Distribution Function.
CDMA	Code Division Multiple Access.
CoMP	Coordinate Multi-Point.
CRE	Cell Range Expansion.
CSI	Channel State Information.
DAC	Digital to Analog Converter.
DL	Downlink.
DP	Dynamic Programming.
EVD	Eigenvalue Decomposition.
FDD	Frequency Division Duplexing.
GP	Geometric Programming.
HetNets	Heterogeneous Networks.
HPA	High Power Amplifier.
ICT	Information and Communications Technology.
i.i.d.	independent and identically distributed.
KKT	Karush-Kuhn-Tucker.
LMS	Least Mean Square.
LP	Linear Programming.
LTE	Long Term Evolution.
MAC	Medium Access Control.

MIMO	Multiple-Input Multiple-Output.
MISO	Multiple-Input Single-Output.
ML	Maximum Likelihood.
MM	Majorization-Minimization or Minorization-Maximization.
MMSE	Minimum Mean-Square Error.
MSE	Mean-Square Error.
MUI	Multiuser Interference.
OFDM	Orthogonal Frequency Division Multiplexing.
PDF	Probability Distribution Function.
PF	Proportional Fair.
PHY	Physical.
QoS	Quality of Service.
QP	Quadratic Programming.
RF	Radio Frequency.
RFID	Radio Frequency Identification.
SC	Small Cell.
SCA	Successive Convex Approximation.
SDP	Semidefinite Programming.
SINR	Signal to Interference Noise Ratio.
SISO	Single-Input Single-Output.
SNR	Signal to Noise Ratio.
SVD	Singular Value Decomposition.
SWIPT	Simultaneous Wireless Information and Power Transfer.
TDD	Time Division Duplexing.
UMTS	Universal Mobile Telecommunications System.
UL	Uplink.
WLAN	Wireless Local Area Networks.
w.l.o.g.	without loss of generality.
w.r.t.	with respect to.
ZMCCS	Zero Mean Circularly Complex Symmetric.
ZF	Zero Forcing.

List of Figures

List of Algorithms

Part I

Introduction

Chapter 1

Introduction

1.1 Motivation

The increasing popularity of Internet applications together with the widespread use of new generation smartphones, netbooks, and tablets, will lead to a massive growth in data traffic [Cis]. In order to be able to process all this data, a substantial capacity increase of current wireless networks is needed. Supporting the expected traffic volume with the current state-of-the-art network management strategies could significantly increase both deployment and running costs, driven mainly by power consumption of network nodes [Cis]. The traffic growth has already increased the energetic demand of the information and communications technology (ICT) sector: in 2008, the ICT community expended 7.15% of the overall global electricity bill [Ver10] and the forecasts predict that this demand will double by 2020 [Ver10]. In fact, an analysis of the global carbon footprint of ICT sector prognosticates that this footprint will increase slightly year by year, in total by 70% in 2020 compared to 2007, to about 1000 Mto CO_2e (equivalent carbon dioxide), which equals 1.9% of the estimated total global CO_2e [Mal13]. In particular, an analysis of the global carbon footprint of cellular communications systems predicts that the carbon emissions will increase by a factor of three by 2020 compared to 2007, rising from about 86 to 235 Mto CO_2e [Feh11]. This situation is not sustainable and as a consequence the European Council has defined a target for 2020 of a 20% reduction in emissions compared to 1990 levels and a 20% share of renewables energies in overall European Union energy consumption [EC08].

On the other hand, it is important to emphasize that battery-powered terminals are becoming broadly used as they provide high mobility to the end users and, in terms of network entities, e.g., base stations (BS), mobile terminals, or sensors, they can be deployed in places where there is no access to the power grid, such as rural areas. Unfortunately, one of the limiting factors of current technology is the short lifetime of the batteries [Sud11]. The battery technology has not evolved fast enough to cope with the increase in energy consumption associated to the growth of the processing capability of the devices[1]. In wireless sensors networks, this can be a serious

[1] In this dissertation, the concept of device and node will be used interchangeably to refer to both transmitters and receivers. The text will clarify what communication end we are referring to.

issue, since sensors are usually placed in positions that are difficult or impossible to be accessed, thus, making the process of replacing the batteries very costly. In cellular environments, the telecommunication providers have put a lot of attention on providing good services with enhanced coverage, but this will not be translated into a really perceived added value if the users cannot make use of them due to the mentioned battery limitations.

Within this context, there have been several studies that address the issue of how to reduce the energy required by the network terminals (see for example [Gol02], [Li11a], [Bel10], [Has11], [Li11b], and [Eic09]). Nonetheless, it is not enough to become more efficient in terms of energy consumption; instead, a solution that provides sustainable networks is required. In this sense, energy harvesting [Par05], [Sud11], which is a technological solution to recharge the batteries by collecting energy from the environment, is emerging as a potential solution both to increase the lifetime of battery-powered devices and to reduce the carbon footprint and meet the 2020 targets of the European Council. Energy harvesting devices, i.e., devices that are provided with energy harvesting sources to recharge their batteries, have many applications. For example, in rural areas where the access to the power grid is impossible or very expensive, deploying sustainable networks based on energy harvesting nodes is the only viable solution. In this context, BSs can be powered with, for example, solar or wind energy along with a battery that provides enough energy to run the network even during night periods. In the context of Internet of Things [Atz10], [Jio14], [Per14], in which it is envisioned that thousands of wireless sensor nodes will be connected to the current networks, energy harvesting devices will play a major role because access to the electric power grid may be difficult in areas where sensors are usually placed. All in all, energy harvesting devices will be deployed massively in the forthcoming years as it is confirmed in the market study performed by Yole Développement [Dev12a], the results of which are depicted in Figure 1.1.

Traditionally, the research community has concentrated its effort on designing communication systems with the emphasis on providing quality of service (QoS) to the users. However, if the network nodes are energy limited and have energy harvesting sources, a new research paradigm is opened: traditional design strategies are no longer optimal in systems with energy harvesting devices, since the power radiated by the transmitter or the available battery level at the receiver are no longer constant, i.e., they depend on the stochastic harvesting process.

Additionally, some solutions to enhance the system capacity and overcome the issue of massive traffic growth have failed to consider other existing power consumption sinks and, thus, they are not fully representative of the system's behavior. In order to have a network as energy-sustainable as possible, all the energy sinks should be considered simultaneously in the transmission strategy. The goal of this dissertation is to provide a general framework for network resource management with energy harvesting devices in order to enhance the system throughput and the lifetime of the network devices. Such designs will encompass physical (PHY) layer designs where spatial beamforming and power allocation techniques will be developed and also

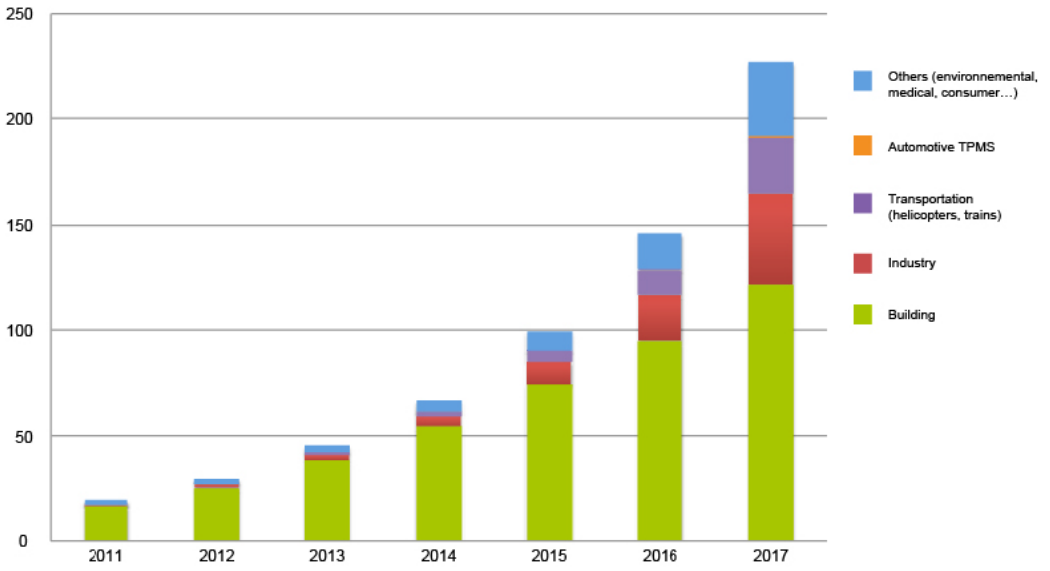


Figure 1.1: Expected sales of energy harvesting modules by application (in M$) (source: [Dev12a]).

medium access control (MAC) strategies in which higher level decisions such as user scheduling and admission control procedures will be implemented. In addition, cross-layer approaches in which both PHY and MAC parameters are jointly processed to enhance the performance of the existing algorithms will be studied.

1.1.1 Power Consumption Sinks

As mentioned in the previous section, system capacity and energy-sustainability of the wireless devices are two main issues that arise in the context of future wireless networks. The first problem is technically challenging as the resources to be shared among users are limited. A general trend that has been proposed widely in the literature is to deploy more crowded networks, i.e., densify the network by deploying more BSs, such as femtocells in cellular environment [Cha08]. By reducing the coverage radius of the access networks, the area spectral efficiency is considerably enhanced. Note that, due to having short distances between transmitters and receivers, the radiated power can be comparable with or even lower than the power consumed by the front-end and the baseband stages [Gro11], [Cui07]. In [Aue11] and [Deb11], authors showed that, in microcells, the radiated energy is practically the same as the energy consumed by the baseband signal processing stages, and for picocells and femtocells the power consumed by the baseband stage is even higher than the actual power radiated by the high power amplifier (HPA). This warrants the need to incorporate such power consumption sinks in the models. In those papers, authors provided a detailed model for the power consumption of a traditional BS and concluded that, if no energy harvesting source is available at the BS, the best way to reduce the overall energy consumption is to turn off the radio frequency (RF) chains whenever the BS is not needed, i.e., shut down the BS [Soh13], [Oh10].

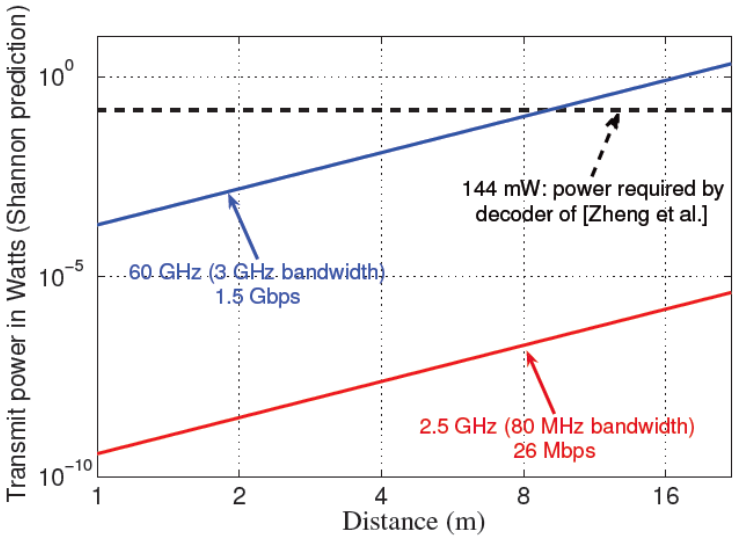


Figure 1.2: The required transmit power for bands of interest assuming short distances. For the ISM band, centered at 2.5 GHz, the figure shows the required power for Bluetooth applications (80 MHz bandwidth) for a data-rate of 26 Mbps. The 60 GHz band presents the upcoming high-bandwidth high-throughput wireless paradigm. The bandwidth is large (3 GHz) and the throughput is 1.5 Gbps. Path-loss exponents are assumed to be 3 (indoor environment) and the noise figure is 3 dB. Most applications today lie somewhere between the two curves. Observe that even for 1.5 Gbps link, the transmit power is not more than a few hundred milliwatts for a distance of 3 m. Many of these applications are designed for even smaller distances, where the transmit power is only a few tens of milliwatts. The source for this figure is [Gro11].

However, it is even more important to focus the attention on the terminal side as it is usually the one experiencing short battery lifetime. In this regard, the RF power consumption as well as the decoding consumption must be included in the models since, as shown in [Gro11], such energy consumption sinks are not negligible and can affect importantly the lifetime of the mobile terminals. In fact, as mentioned before, the energy consumed by the decoding process is higher than the radiated power for some network configurations and distances, as it is shown in Figure 1.2. Authors in [Jen12] provided mathematical models for the consumption of the RF chains and decoder stage of a Long Term Evolution (LTE) mobile terminals based on power consumption measurements of a real LTE device. Also in this case, if no energy harvesting source is provided to the terminal, the best way to reduce the overall energy consumption is by means of disconnecting the RF chains when they are not used (see for example [Zho08], [Wig09], and [Jha13]).

All this suggests that, in network scenarios with small cells[2] (SCs), it is crucial that all power consumption sinks be considered and not just the power radiated by the HPA. As it will be shown in further chapters of this thesis, by considering all these aspects, it is possible to provide higher capacity in terms of network throughput.

[2]Small cells are considered a type of BS in which the coverage area of the cell is considerably smaller than that of a traditional macro BS. In this context, picocells and femtocells are examples of SCs.

1.1.2 Energy Harvesting Systems

If the network must be as sustainable as possible, then energy harvesting nodes need to be deployed. Traditionally, energy harvesting techniques have been developed based on energy sources such as wind or solar energy. Nevertheless, there are other techniques that could be applied to moving terminals (this may be the case of cellular phones) based on piezoelectric technologies. Additionally, ambient RF signals can be used as a source for energy scavenging. In the following, we present a brief summary of the main techniques that are being currently researched and developed. Table 1.1 presents numerical examples of the amount of energy that can be harvested with different technologies. Additionally, the table also presents the level of predictability of each harvesting technology and whether the technology can be controlled or not. Note, however, that the harvesting process is ultimately a random process and, thus, most of the times predictability can only be guaranteed in terms of expected values.

Energy source	**Characteristics**	**Harvested power**
Solar energy [Vul10]		
Outdoor	Uncontrollable (predictable)	10 mW/cm^2
Indoor	Uncontrollable (predictable)	10μW/cm^2
Thermal energy [Vul10]		
Human	Uncontrollable (predictable)	30 μW/cm^2
Industrial	Uncontrollable (predictable)	$1 - 10$ μW/cm^2
Wind energy [Ceb]	Uncontrollable (unpredictable)	1 W at 2000 rpm
Kinetic energy [Sud11]		
Finger motion	Controllable	2.1 mW
Footfalls	Uncontrollable (predictable)	5 mW
RF energy [Vul10], [Vis08]		
GSM	Uncontrollable (unpredictable)	0.3 μW/cm^2 at 25 m
TV broadcasting	Uncontrollable (unpredictable)	0.1 μW/cm^2 at 4 km

Table 1.1: Characteristics and power harvested by different harvesting technologies.

1.1.2.1 Solar Energy

Solar energy is the most commonly exploited source of energy. Solar energy could be replaced by artificial light but the applicability of this strategy is reduced by the amount of energy that can be received and stored. The energy transducer is a solar panel, which generates electricity by converting light into current through the photovoltaic effect. The main problem associated with solar energy is that it is not always available as it depends on weather conditions. In systems running with solar energy, the battery dimensioning is usually carried out considering a worst

case scenario, i.e., the battery is over-dimensioned in order to ensure energy for a certain period of time in case the weather conditions are not good.

1.1.2.2 Wind Energy

The wind is also a quite powerful energy to generate electricity. Usually a wind turbine is used to collect energy from the air flows and the amount of energy depends on the size of the turbine. Usually, this technique provides extraordinary results in terms of power generated, but the main issue with this source of energy is that it is uncontrollable and unpredictable as fast variations in the air flows can easily occur.

1.1.2.3 Thermal Energy

If the wireless device is exposed to temperature gradients, then a thermoelectric generator or a thermogenerator can be implemented in order to harvest such thermal energy and convert the temperature gradients in electrical energy by means of the Seebeck effect [Mat06]. The amount of harvested energy is limited by the temperature gradient the device is exposed to and it is generally incontrollable but predictable.

1.1.2.4 Kinetic Energy

Piezoelectric generators or electrostatic and electromagnetic converters can be used to transform movements or vibrations of objects into electric energy [Rou03], [Bee06], [Gor]. For example, by moving a finger or just simply walking, it is possible to generate a certain amount of power (see Table 1.1 for specific values). This energy is usually controllable and predictable.

1.1.2.5 RF Energy

Ambient RF energy, coming for example from wireless communications systems, can be exploited to charge the batteries of the wireless devices. Apart from energy of existing communications systems, the transmitter could be designed to send specific RF energy to the devices. For example, RF energy is used in Radio Frequency IDentification (RFID) technologies [Lan05]. In [Bha06], authors showed that Wireless Local Area Networks (WLAN) transmissions are able to power sensor nodes. The major drawback of this technique is the energy absorption of the wireless channel. For this reason, the distance between the transmitter and the receiver must be short. RF energy is controllable in applications in which the power source is controlled by the system designer and uncontrollable and unpredictable when energy is harvested from spontaneous transmissions of other networks.

1.1.2.6 Energy Harvesting Modeling

In the previous section, a summary of the different harvesting technologies was presented. Regardless of the technology being used, the only thing that is needed from the designer point of view is the *power harvesting profile*, i.e., the characterization of the temporal evolution of the harvested power. Usually, the amount of power that is being harvested at a particular time depends, apart from the technology, on the dimension of the harvesting source and its on efficiency. As the harvesting profile is usually a stochastic process, having knowledge of the exact profile is not always possible. However, there are cases where it is possible to know the entire realization or to predict it in statistical sense. Most of the efforts in the literature have focused on modeling the power harvesting profile of the solar harvester [Gor], [Kan07], [Mio14]. For example, in [Mio14], authors presented a methodology and a tool to derive simple but yet accurate stochastic Markov processes for the description of the energy scavenged by outdoor solar sources. Yet more research work is needed to fully characterize the other harvesting sources.

In any case, in order to develop resource management techniques, an abstract model that comprises any possible harvesting source is required. The parameters of the model can be computed by discretizing the power harvesting profile (taking samples of the stochastic process throughout time) and by designing a statistical model that characterizes such profile. Authors in [Sey08] proposed a stationary Markov model. They obtained the probability of the average time to run out of energy, a measure that can be used to further optimize systems. Later, authors in [Ho10] generalized the previous work by proposing a model that captures the non-stationarity behavior of Markovian models. Based on the empirical measurements and by using a Bayesian criterion, they concluded that the piezoelectric energy can be better described by their model, while a stationary Markov model is sufficient for the solar energy.

Due to the previous reasons, a Markovian model will be extensively used in this dissertation, and specific models will be pointed out if specific technologies are to be employed. This Markovian model allow us to describe the intensity of the energy source and the size and the efficiency of the harvesting source, among others, by adjusting the parameters of the model. The specific statistical representations of the harvesting profiles can be obtained, for example, from databases on the Internet or real campaign measurements.

1.2 State of the Art of Resource Allocation Strategies with Energy Harvesting Nodes

Let us start the review of the state of the art by noting that, in this dissertation, resource management techniques with energy harvesting devices have been considered at different levels of the protocol stack, such as PHY designs and admission control procedures. However, the vast majority of applications of energy harvesting devices in the literature have dealt with resource

allocation in transmission policies, i.e., PHY designs, which will be, therefore, the main focus of this section. Specific review of the literature of each application will be further provided at each individual chapter of this dissertation.

Energy harvesting strategies can be grouped into two main families that we call *passive* energy harvesting systems and *active* energy harvesting systems. In passive energy harvesting systems, the network devices are provided with an external energy harvesting source and the transmission strategy has to adapt and optimize itself according to such harvesting source. This is the case of having, for example, a solar panel or a wind turbine. On the other hand, active harvesting systems consider transmission strategies in which the transmitter actively feeds the receivers with energy via RF signals. The receivers collect such ambient-specific RF signals and convert them into energy. In this case, the transmission strategy has two simultaneous goals: transmit useful information and provide energy to the receivers. In the following, we present a review of the main techniques available in the literature concerning passive and active harvesting systems.

1.2.1 Passive Energy Harvesting Systems

In passive energy harvesting systems, the network device is using an external energy source that cannot be controlled and, thus, the transmission strategy has to adapt to the energy source behavior. It has been shown in the literature that traditional transmission strategies, such as water-filling [Cov06] in classical mutual information maximization or mercury/water-filling [Loz06] in capacity maximization with an arbitrary input distribution, are no longer optimal in systems with energy harvesting devices. In classical transmission strategy designs, the available power limiting the performance is considered as a fixed constraint, but in systems with harvesting devices, the available power changes over time as it depends on the amount of energy that the devices are capturing from the environment. Therefore, constraints on the amount of energy available at the battery must be imposed for each time period, making the overall problem much more challenging than the one encountered in classical transmission strategies. The vast majority of the works in the literature deal with the scenario in which the device provided with the energy harvesting source is the transmitter. In this dissertation, resource allocation for both energy harvesting transmitters and energy harvesting receivers will be studied and evaluated.

In this context, there exist two well established approaches: online and offline. The difference between these two approaches lies on the availability of information in terms of the power harvesting profile at the transmitter. The degree of knowledge of this information, of course, influences the overall strategy and performance of the transmission. *Offline transmission strategies* assume that the transmitter has full knowledge (i.e., from the past, present, and future realizations) of the harvesting process. In other words, the transmitter knows exactly in advance when and how much harvested energy will be available. On the other hand, *online*

transmission strategies consider that the transmitter has only information from the past and present realizations of the harvesting process and, possibly, also some statistical information concerning its future behavior. Offline transmissions can be studied to obtain upper-bounds on the performance of systems with energy harvesting sources; however, the transmission strategies obtained from this approach are difficult to implement in reality as the assumption of knowing exactly when and how much energy will be available at the battery is somewhat unrealistic. This is the main reason why in this dissertation we focus entirely on online solutions that can be implemented in real systems with causal available information.

In terms of offline solutions, the generalization of the classical water-filling was initially derived in [Oze11]. In that paper, authors maximized the mutual information [Cov06] of a single user scenario for a set of the consecutive channel access with energy causality constraints, considering that the transmitter was provided with an energy harvesting source. Then, it was generalized in [Gre13c] by considering an arbitrary input distribution and a multiple-input multiple-output (MIMO) setup. The authors of [Oze10] studied the coding problem from an information theoretic perspective. The transmission policy that maximizes the mutual information considering finite battery capacity was derived in [Tut12], assuming an infinite backlog of data at the transmitter. The effect of battery imperfections in terms of battery leakage was studied in [Dev12b].

The concept of cumulative curves (see, for example, [Zaf07], [Zaf09], and [Gre14a]) has been used to obtain optimal strategies with harvesting devices. These cumulative curves allow the derivation of a graphical and intuitive solution by representing, for instance, the number of bits that have been cumulatively transmitted by a given node in a given time interval. In this regard, authors in [Yan12b] derived the transmission strategy that minimizes the delivery time of all data packets with energy constraints under the assumption of having infinite battery capacity. The case with finite battery capacity was studied in [Tut12]; however, it was considered that all data packets were available from the beginning of the transmission. It was in [Gre13a], where authors considered that both data packets and energy packets arrived at the transmitter dynamically throughout time.

The works mentioned so far considered that the power radiated by the HPA is the only non-negligible source of energy sink of the device. As it was shown before, this is not true in short-distance networks. Nevertheless, some works in the literature also incorporated other energy sinks in their models (see, for example, [Orh12b] and [Gre14b]).

There have been also some attempts to derive optimal policies for the multiuser scenario. The problem was extended to the broadcast channel in [Yan12a], [Ant11], assuming infinite battery capacity, and [Oze12], where authors found the rate maximizing scheduling policy under the assumption of finite battery capacity. In terms of the uplink (UL) channel, [Yan11] proposed the power allocation strategy that minimizes the transmission completion time. The relay and

multi-hop channels have also been studied in the context of energy harvesting devices. For example, two-hop communications were studied in [Gü11], [Gur12], [Orh12a], and [Luo13]. The relay channel (with direct link between transmitter and receiver) was addressed in [Feg13], [Hua13], and [Med10].

In the context of online solutions, the state of the art is much limited. This is basically due to the complexity of the optimal online approaches. The optimal formulation is time-coupled over all time periods (or frames) considered and the solution has to be found jointly for all time instants as the energy not used in a particular frame can be used in any future frame. If the global optimum is to be achieved, then we need to resort to dynamic programming techniques [Ber05]. However, these techniques require a high number of computational resources, making the strategies difficult to implement in reality. For example, in [Ho12b] authors developed an online transmission strategy considering dynamic programming techniques for a single user single-input single-output (SISO) scenario. On the other hand, in [Bla12] and its journal version [Bla13], authors presented an online approach based on learning techniques under the framework of dynamic programming. However, the scenario under consideration is a simple single-user SISO system. The extension to multiuser systems is a difficult task. In order to avoid using such complex techniques, suboptimal techniques have to be studied to solve the online approach (see for example [Gre13c]).

In any case, much attention has been put on solving and modeling the offline counterpart and just very little effort on solving the online problem. For this reason, in this dissertation, we propose online resource allocation strategies that require low computational burden and are easy to manage even in multiuser scenarios. The idea behind the proposed strategies is to consider a *greedy* approach [Cur03] in which the resource allocation procedure is solved at each frame independently. The coupling energy constraints that appear in the offline formulation will be relaxed and, instead, a per-frame energy constraint will be assumed. This per-frame energy constraints will be a function of the available energy at the battery, which already incorporates the harvesting collected in previous frames. The function that models the extraction of energy from the battery on a per-frame basis can be further optimized as it will be shown later in this dissertation.

So far we have reviewed the state of the art considering energy harvesting transmitters. In terms of transmission strategies considering energy harvesting receivers, only a few works can be found in the literature. For example, authors in [MD13] and [MD14] determined the sampling policy for a given code rate, channel capacity, and battery capacity, for static channels and fading channels, respectively. Then, the same authors extended the previous works in [Yat15] and determined the decoding policies at the receiver that maximize the packet throughput. In [Bai13], throughput maximization with respect to (w.r.t.) the analog to digital converter (ADC) and the transmission bandwidth was derived in a single user scenario. In terms of offline strategies considering energy harvesting receivers, the work in [Ara14] presented a throughput

maximization strategy in which the transmitter had full knowledge of the stochastic harvesting process of the receiver. The modeling employed for the energy consumption at the receiver was based on the model proposed in this dissertation.

1.2.2 Active Energy Harvesting Systems

In active energy harvesting systems, the transmitter radiates electromagnetic energy explicitly towards the receivers. Once this energy is captured by each receiver, it is then transformed into electric energy that is stored at the battery for future use. Usually, at the same time, the transmitter also sends useful information that is of interest to the receivers. This is known in the literature as *simultaneous wireless information and power transfer* (SWIPT). Research opportunities and challenges on SWIPT designs can be found in [Bi15]. Note that, active harvesting systems are online approaches by nature.

The concept of SWIPT was first introduced in the work [Var08] by Varshney. He showed that, for the single-antenna additive white Gaussian noise (AWGN) channel, there exists a non-trivial trade-off in maximizing the data rate versus the power transmission. That work was then generalized to multiuser systems in [Fou12], following a similar information-theoretic approach. Later, in [Gro10], authors extended the previous work by considering frequency-selective single-antenna AWGN channels. In [Zha11] (and its journal version [Zha13]), authors considered a MIMO scenario with one transmitter capable of transmitting information and power simultaneously to two receivers. They proposed two receiver architectures, namely time-switching and power-splitting that were able to combine both sources (information and energy) at the same time (see also [Liu13a] for a dynamic power splitting approach). In [Liu13b], authors introduced time scheduling between information and energy transfer and derived the optimal switching policy with time-varying co-channel interference. The receiver, thus, replenished energy opportunistically via wireless power transfer from the unintended interference and/or the intended signal sent by the transmitter. On the other hand, SWIPT scenarios with imperfect channel state information (CSI) have also been considered [Xia12]. In that work, authors proposed a robust beamforming design policy based on the same scenario presented in [Zha11].

In [Zho13b], authors further developed the previous concepts by presenting with much more detail specific receiver architectures and their performance. The concept of SWIPT was then extended to the multiuser multiple-input single-output (MISO) scenario in [Xu13] and [Shi14] and to the MIMO relay scenario in [Cha11] and [Che15]. Finally, authors in [Zen15] addressed the design of the training duration in MIMO systems for channel estimation considering the energy not being harvested due to the training process. They formulated the problem as an optimization problem and obtained the optimum training duration and the power allocated to the training and the power transfer phases. A more complete review of the state of the art will be given later in Chapter 4 of this dissertation.

1.3 Outline of Dissertation

This dissertation proposes some resource allocation and network procedure strategies, considering that the network nodes are provided with energy harvesting capabilities and finite batteries. For structuring purposes, this dissertation is divided into two main parts. In the first part of the document, we consider that the receivers are battery-constrained and that they have harvesting capabilities whereas, in the second part, we assume that the transmitters do not have access to the power grid and, thus, energy harvesting sources enable them to provide service to the final users. In the following, we provide a detail summary of the different scenarios and design problems considered throughout this dissertation.

1.3.1 Energy Harvesting Capabilities at the Receiver Side

As mentioned before, in the first part of the thesis we assume that the transmitter is able to be connected to the power grid and the mobile terminals are provided with a finite battery that can be recharged by means of energy harvesting. For simplicity, we consider a scenario with just one transmitter and without external interference (e.g. inter-cell interference), see Figure 1.3. We assume that the distance between the transmitter and the receivers is short and, thus, the radiated power is comparable to other sinks of power consumption. Hence, we derive power models that include RF chain consumption and signal processing blocks consumption.

This first part of the thesis is structured into two chapters. In Chapter 3, we deal with resource allocation strategies for a broadcast multiuser MIMO network where the energy harvesting source is external, that is, the transmitter is not able to control how much energy the receivers are actually harvesting. The main contributions of Chapter 3 are:

- Description of different decoding and RF power consumption models.
- Proposal of a power and rate allocation strategy taking into account the state of the batteries of the terminals.
- Development of an online precoder design considering imperfect CSI and evaluation of its impact on the evolution of the data rate and battery levels of the terminals.
- Proposal of a robust design based on imperfect knowledge of the battery due to the finite data rate feedback link used to collect the battery information.
- Asymptotic analysis and characterization of the battery and data rate evolution of the users in the proposed system.
- User scheduling strategies considering the current battery levels and the average throughputs achieved by the users.

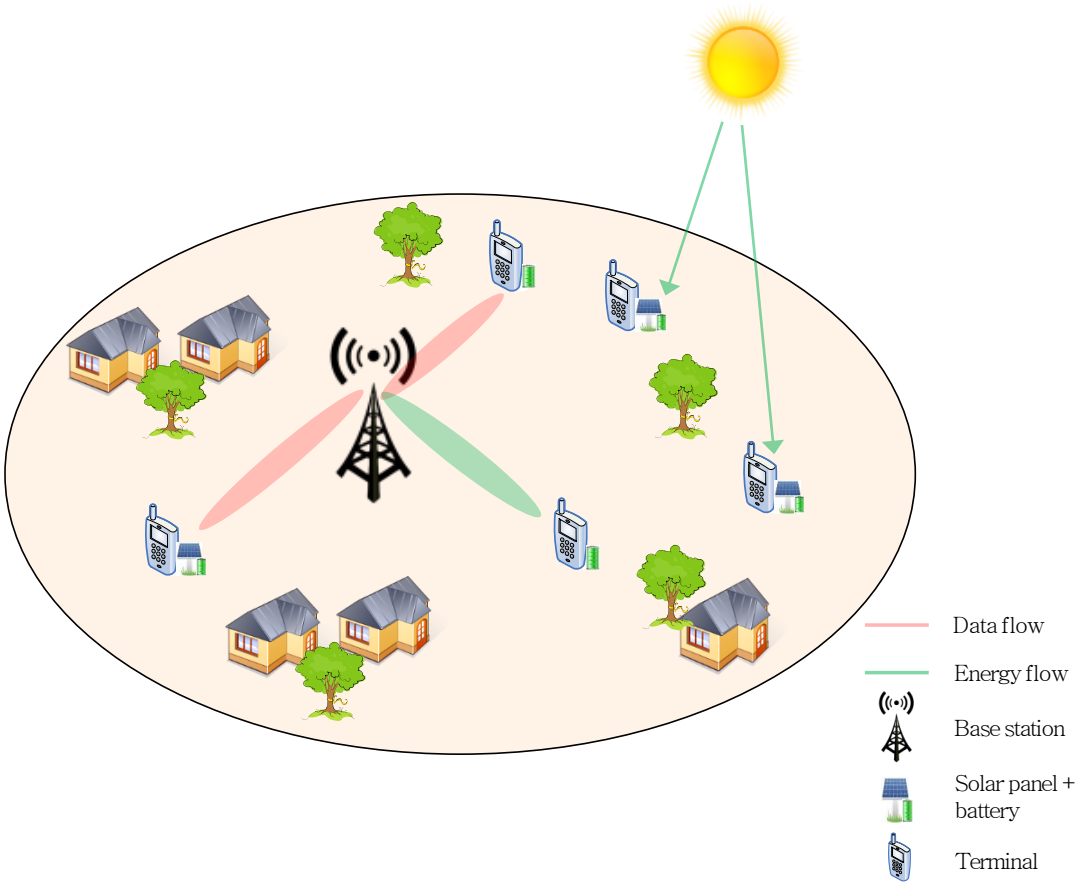


Figure 1.3: Reference scenario for the first part of the dissertation. Receivers are battery-constrained with energy harvesting capabilities.

The results presented in Chapter 3 have been published in one journal paper [J1] and three conference papers, [C5], [C7], and [C9] (see next section). Publications [C1] and [C2] are also related to Chapter 3 but the contents are not presented in this dissertation.

In Chapter 4, we develop some resource allocation strategies but assuming that the transmitter can actively feed the receivers with power through RF signals. Therefore, the energy harvesting process can be controlled and, thus, is included in the optimization design problem. In this context, the main novelty presented in Chapter 4 w.r.t. the current literature is:

- Proposal of a multiuser multi-stream MIMO broadcast transmission strategy. The system weighted sum rate with individual per-user harvesting constraints are considered in the proposed transmission strategy design. We also take into account the state of the batteries of the terminals in the proposed strategy. We study particular cases in which only information and only harvesting users are present in the system.

- Development of an efficient algorithm that computes the optimal precoding matrices for the multiuser MIMO broadcast network setup mentioned previously.

- The fundamental (multidimensional) trade-off between system performance and (per-user) harvested energy is studied and characterized, putting emphasis and giving specific closed-

form expressions for some particular cases of interest.

- Power consumption models at the transmitter and the receivers are incorporated. In particular, the decoding power consumption at the receivers and its impact on the system performance is considered.

- Development of harvesting-constrained user grouping schemes that employ a two-stage user scheduling mechanism, aiming at enhancing the system throughput and/or fairness among users.

- Advanced optimization techniques based on the majorization-minimization method are proposed to solve the non-convex multiuser broadcast MIMO system.

- Proposal of different strategies to manage the power to be harvested by users by controlling how much energy a user should be able to harvest, modeling the impact of that decision in the system performance.

The results presented in Chapter 4 have been submitted for publication in two journal papers [J4] and [J6], in a conference paper [C14], and published in two conference papers [C6] and [C11].

1.3.2 Energy Harvesting Capabilities at the Transmitter Side

In the second part of the dissertation, we focus our attention on the scenario in which the transmitter is the one with energy harvesting capabilities. This may be the case of, for example, BSs placed in remote rural areas where the access to the power grid is not possible or too expensive (see Figure 1.4). In this context, we optimize different network parameters, such as turning on and off transmitters, resource allocation considering backhaul capacity constraints, or admission control procedures in multi-cell scenarios.

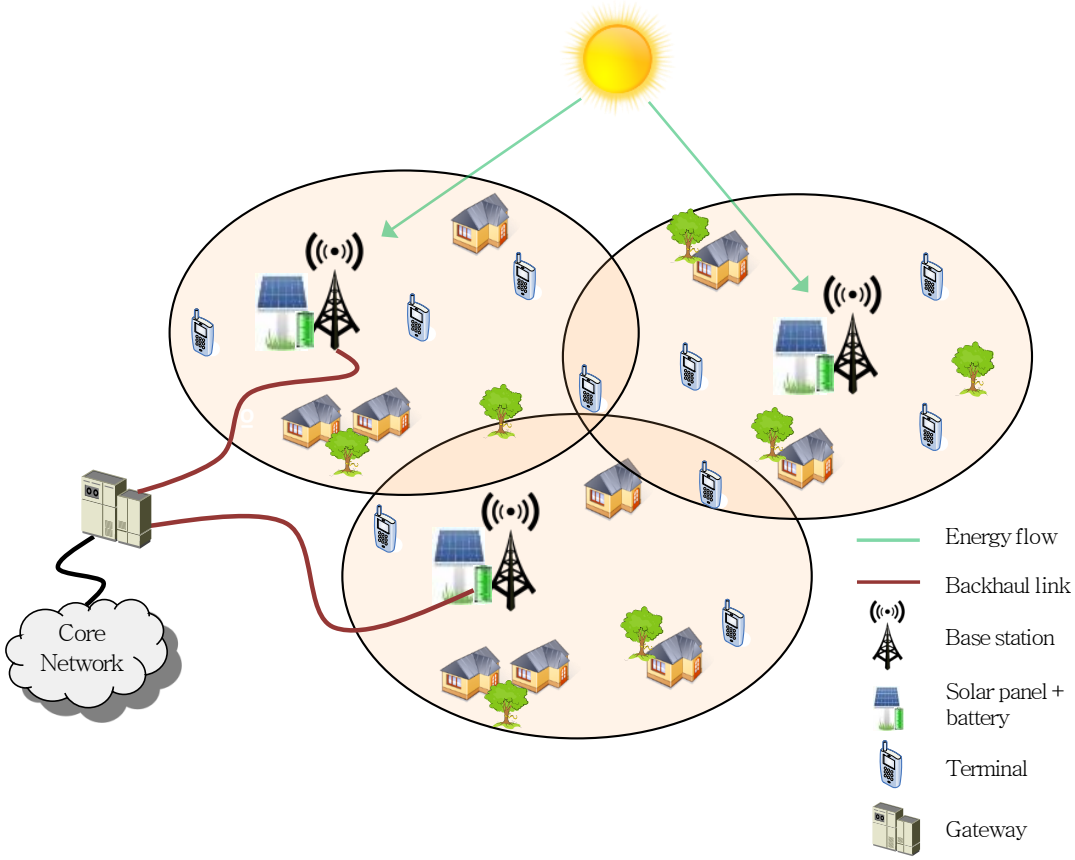


Figure 1.4: Reference scenario for the second part of the dissertation. Transmitters are battery-constrained with energy harvesting capabilities.

This second part of the thesis is structured in three chapters. Chapter 5 deals with strategies for switching on and off transmitters when the current network traffic is under a pre-defined threshold. We propose strategies to design such on/off threshold based on the quality of service experienced by the users and the power needed to run the network. The main contributions are:

- Proposal of a methodology for dimensioning the energy units, i.e., solar panels and batteries. The output of the dimensioning strategy is the required number of units along with their sizes.

- Development of a dynamic switching on/off strategy that provides a size reduction of the energy units. The decision to switch off one of the BSs is based on the required power and not just on the traffic demand.

- Two different approaches to determine the switching on/off threshold are derived: a deterministic approach, where we assume that full knowledge of the traffic profile is available, and a more realistic statistical robust approach, which accounts for possible error traffic estimation and modeling, both for the case of having single type of traffic (voice) and mixed traffic (voice and data).

The results presented in Chapter 5 have been submitted for publication in one journal paper [J3], and published in two conference papers [C8] and [C12].

Later in Chapter 6, we propose some resource allocation strategies, that is, we allocate power, number of codes, and data rates, for the single-cell scenario, in which there exists a backhaul connection that constrains the data rate at the access network. The backhaul constraint is assumed to be long-term and, thus, ergodic optimization techniques are employed to tackle the problem. Resource allocation strategies for the downlink (DL) as well as for the UL are developed. The main contributions are:

- Proposal of a fair scheduling algorithm considering a long-term backhaul constraint, the battery status of the BS, and the energy that it is being harvested.
- An online strategy based on ergodic optimization (also known as stochastic approximation) is developed.
- Two different types of users, voice users and data users, are assumed to coexist in the network. Each type of users demands a different quality of service.
- Resource allocation strategies are proposed for both, the DL and the UL scenarios.

The results presented in Chapter 6 have been published in one journal paper [J2], and submitted for publication in one conference paper [C13].

Finally, in Chapter 7, we propose some user association strategies for a multi-tier heterogeneous network that try to achieve load balancing in terms of overall network throughput among the different BSs. Two different approaches are derived. The first is based on a heuristic approach, the same followed in previous chapters of the dissertation, in which the available power is a function of the current battery, and a second approach in which we make use of ergodic optimization tools to decide how to better use the available energy at the battery. In this context, the main contributions of this chapter are:

- Proposal of a network model for heterogeneous networks in which different types of BS coexist. These BSs have different capabilities, such as smaller/larger solar panels, smaller/larger batteries, coverage area, number of available resources, etc.
- Development of several user association strategies. Some are low complexity solutions that are to be used in scenarios where the computational complexity is a limitation.
- Design of an iterative decentralized strategy in which each BS runs the association algorithm with just local information, i.e., local CSI of the users.
- The asymptotic behavior of the battery is characterized.

- Development of an association strategy with ergodic optimization tools in which the objective function as well as some energy constraints are assumed to be long-term.

The results presented in Chapter 7 have been submitted for publication in two journal papers [J5] and [J7], and published in one conference paper [C10].

1.4 Research Contributions

The research performed during this dissertation has led to the following publications that are either accepted, submitted, or in preparation for submission in the near future:

Journal papers

- [J7] **J. Rubio**, A. Pascual Iserte, and Antonio G. Marqués, "Stochastic Cell User Association with Energy Harvesting HetNets", in preparation.
- [J6] **J. Rubio**, A. Pascual Iserte, D. P. Palomar, and A. Goldsmith, "Majorization-Minimization Strategies for SWIPT in Multiuser MIMO Broadcast Systems", *submitted to IEEE Transactions on Signal Processing*, April 2016.
- [J5] **J. Rubio**, A. Pascual Iserte, J. Del Olmo and J. Vidal, "User Association for Dynamic Rate Balancing in Heterogeneous Networks with Energy Harvesting Constraints", *submitted to IEEE Transactions on Mobile Computing*, January 2016.
- [J4] **J. Rubio** and A. Pascual Iserte, "User Grouping and Resource Allocation in Multiuser MIMO Systems with Simultaneous Wireless Information and Power Transfer", *submitted to IEEE Transactions on Mobile Computing*, February 2016.
- [J3] **J. Rubio**, J. Del Olmo, A. Pascual Iserte, J. Vidal, O. Muñoz, and A. Agustín, "Access Network Dimensioning and On/Off BS Switching Strategies for Sustainable WCDMA Wireless Networks in Isolated Rural Areas", *Submitted to Computer Networks Journal*.
- [J2] **J. Rubio**, O. Muñoz Medina and A. Pascual Iserte, "A Stochastic Approach for Resource Allocation with Backhaul and Energy Harvesting Constraints", *accepted at IEEE Transactions on Vehicular Technology*, August 2015.
- [J1] **J. Rubio** and A. Pascual Iserte, "Energy-Aware Broadcast Multiuser-MIMO Precoder Design with Imperfect Channel and Battery Knowledge", *IEEE Transactions on Wireless Communications*, Vol. 13, pp. 3137 - 3152, June 2014.

Conference papers

- [C14] **J. Rubio**, A. Pascual Iserte, D.P. Palomar, and A. Goldsmith, "SWIPT Techniques for Multiuser MIMO Broadcast Systems", *submitted to IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Valencia (Spain), September 2016.

- [C13] **J. Rubio**, O. Muñoz, A. Pascual Iserte, and J. Vidal, "Stochastic Resource Allocation with a Backhaul Constraint for the Uplink", *submitted to IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Valencia (Spain), September 2016.

- [C12] **J. Rubio**, A. Pascual Iserte and J. Vidal, "BS On/Off Strategies for Wireless Networks Powered with Energy Harvesting Sources", *Joint Newcom/COST Workshop on Wireless Communications*, Barcelona (Spain), October 2015.

- [C11] **J. Rubio** and A. Pascual Iserte, "Harvesting Management in Multiuser MIMO Systems with Simultaneous Wireless Information and Power Transfer", *IEEE Vehicular Technology Conference (VTC) Spring*, Glasgow (UK), May 2015.

- [C10] **J. Rubio**, A. Pascual Iserte, J. Del Olmo and J. Vidal, "User Association for Load Balancing in Heterogeneous Networks Powered with Energy Harvesting Sources", *IEEE Global Communications Conference (GLOBECOM)*, Austin (USA), December 2014.

- [C9] **J. Rubio** and A. Pascual Iserte, "Energy-Aware User Scheduling for Downlink Multiuser-MIMO Systems", *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Washington DC (USA), September 2014.

- [C8] **J. Rubio**, A. Pascual Iserte, J. Del Olmo and J. Vidal, "Dynamic BS Switch On/Off Strategies for Sustainable Wireless Networks", *IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*, Toronto (Canada), June 2014.

- [C7] **J. Rubio**, A. Pascual Iserte, J. J. García Fernández, A. García Armada, O. Font Bach and N. Bartzoudis, "Asymptotic Analysis of Multiuser-MIMO Networks with Battery-Constrained Receivers", *European Wireless*, Barcelona (Spain), May 2014.

- [C6] **J. Rubio** and A. Pascual Iserte, "Simultaneous Wireless Information and Power Transfer in Multiuser MIMO Systems", *IEEE Global Communications Conference (GLOBECOM)*, Atlanta (USA), December 2013.

- [C5] **J. Rubio** and A. Pascual Iserte, "Energy-Aware Broadcast MU-MIMO Precoder Design with Imperfect Battery Knowledge", *IEEE Global Communications Conference (GLOBECOM)*, Atlanta (USA), December 2013.

Other publications not presented in this dissertation

During the elaboration of this dissertation, the author has been involved in different projects and collaborations. From those collaborations, some papers that are not included in this dissertation have been published in international conferences. The topics of those works are: resource allocation with out of cluster interference and energy modeling of signal processing blocks. The list of publications is:

- [C4] N. Bartzoudis, O. Font Bach, M. Payaró, A. Pascual Iserte, **J. Rubio**, J. J. García Fernández and A. García Armada, "Energy Profiling of FPGA-Based PHY-Layer Building Blocks Encountered in Modern Wireless Communication Systems", *IEEE Sensor Array and Multichannel Signal Processing Workshop*, A Coruña (Spain), June 2014.

- [C3] J. J. García Fernández, A. García Armada, **J. Rubio**, A. Pascual Iserte, O. Font Bach and N. Bartzoudis, "Adaptive Block Diagonalization and User Scheduling with Out of Cluster Interference", *European Wireless*, Barcelona (Spain), May 2014.

- [C2] **J. Rubio** and A. Pascual Iserte, "Energy-Efficient Resource Allocation Techniques for Battery Management with Energy Harvesting Nodes: a Theoretical Approach", *IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai (China), April 2013.

- [C1] **J. Rubio**, A. Pascual Iserte and M. Payaró, "Energy-Efficient Resource Allocation Techniques for Battery Management with Energy Harvesting Nodes: a Practical Approach", *European Wireless*, Guildford (UK), April 2013.

Chapter 2

Mathematical Preliminaries

In this chapter, we are going to provide an overview of some important mathematical tools that are constantly used throughout the dissertation. These tools are convex optimization, ergodic stochastic optimization, multi-objective optimization, and majorization-minimization theory. They are basically used to solve the optimization problems that we derive in this thesis.

2.1 Convex Optimization Theory

In this section, a mathematical framework based on *convex optimization* theory is presented. This section is a brief summary of [Boy04] that aims to provide a self-contained document which will give a broad understanding of the basic ideas to the reader that is not familiar with this theory. These mathematical preliminaries will be useful in some chapters of this thesis where nonlinear constrained convex optimization problems must be solved. For a more in depth theory of convex analysis, the interest reader is referred to [Boy04], [Ber99], and [Roc97].

Convex optimization theory provides a framework for solving a variety of constrained optimization problems. There is, in general, no analytical formula for the solution of convex optimization problems. However, in some cases, it is possible to obtain a closed-form solution, or at least a semi-analytical solution[1], based on the application of the Karush-Kuhn-Tucker (KKT) conditions under some mild conditions, as it will be shown later on in this section. Besides, there exists a great variety of very effective numerical methods for solving the problems with no analytical solution, such as for example *interior point methods*. It is out of the scope of this thesis to present such methods. For a formal description the reader is referred to [Boy04].

[1] A semi-analytical solution means that one part of the solution is given by an analytical expression but the other part of the solution has to be found numerically by means of, for example, iterative algorithms.

2.1.1 Convex Sets and Convex Functions

Suppose $\mathbf{x}_1 \neq \mathbf{x}_2$ are two points in $\mathbb{R}^N$. Points of the form

$$\mathbf{y} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, \tag{2.1}$$

where $\theta \in \mathbb{R}$, form the *line* passing through $\mathbf{x}_1$ and $\mathbf{x}_2$. The parameter value $\theta = 0$ corresponds to $\mathbf{y} = \mathbf{x}_2$, and the parameter value $\theta = 1$ corresponds to $\mathbf{y} = \mathbf{x}_1$. Values of the parameter θ between 0 and 1 correspond to the (closed) *line segment* between $\mathbf{x}_1$ and $\mathbf{x}_2$.

A set $\mathcal{A}$ is a *convex set* if the line segment between any two points in $\mathcal{A}$ lies in $\mathcal{A}$. This can be expressed mathematically as

$$\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{A}, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}, \quad \forall \theta \in [0,1]. \tag{2.2}$$

Figure 2.1 depicts a simple example of a convex and a nonconvex set in $\mathbb{R}^2$. Basically, a set is convex if every point in the set can be reached by any other point in the set through a straight segment that must also lie in the set. There are many examples of convex sets. The most important ones are cones, hyperplanes and halfspaces, Euclidean balls and ellipsoids, among others. There exist a variety of properties and mathematical operations that preserve convexity concerning convex sets. For instance, the intersection of an infinite number of convex sets is convex, i.e., if S_α is convex for every $\alpha \in \mathcal{A}$, then $\bigcap_{\alpha\in\mathcal{A}} S_\alpha$ is convex. See [Boy04] for a more complete list of properties and examples of convex sets.

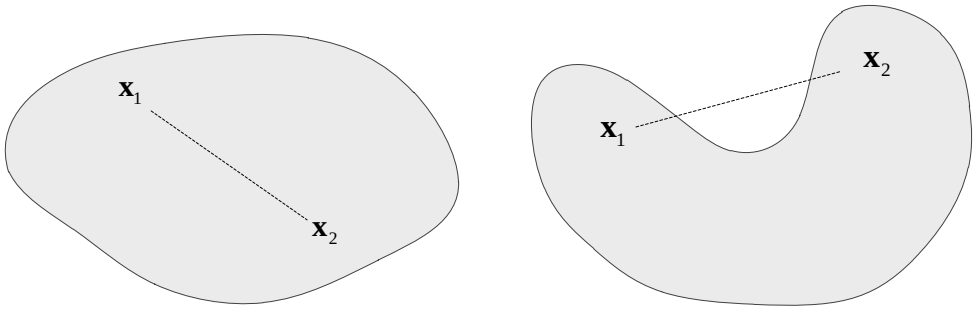
A function $f : \mathbb{R}^N \to \mathbb{R}$ is a *convex function* if **dom** f is a convex set and if $\forall \mathbf{x}_1, \mathbf{x}_2 \in$ **dom** f and $\forall \theta \in [0,1]$, we have

$$f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2). \tag{2.3}$$

Geometrically, the previous inequality means that the line segment between $(\mathbf{x}_1, f(\mathbf{x}_1))$ and $(\mathbf{x}_2, f(\mathbf{x}_2))$ lies above the graph of f. The simplest example of a convex function is the set of affine functions which have the form $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$. A function f is *concave* if $-f$ is convex. Examples of convex and concave functions are exponentials functions, powers, logarithms, etc. We say f is *strictly convex* if strict inequality holds in (2.3) whenever $\mathbf{x}_1 \neq \mathbf{x}_2$ and $0 < \theta < 1$ (strictly concave if $-f$ is strictly convex). See Figure 2.2 for a representation of a convex and a concave function.

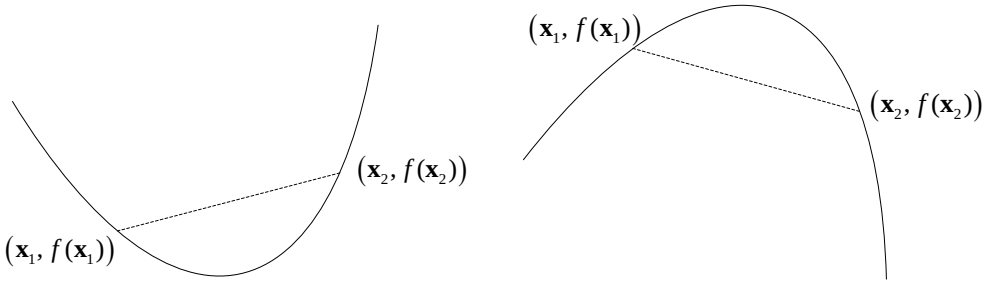
In order to verify the convexity of a function, either of the two following conditions must hold. Let us start with the so-called *first-order condition*. Suppose f is differentiable and its domain convex. Then, f is convex if, and only if,

$$f(\mathbf{x}_2) \geq f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)^T(\mathbf{x}_2 - \mathbf{x}_1), \tag{2.4}$$



(a) Example of a convex set. (b) Example of a non-convex set.

Figure 2.1: Convex sets and non-convex sets.



(a) Example of a convex function. (b) Example of a concave function.

Figure 2.2: Example of a convex and a concave function.

holds for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{dom}\ f$. On the other hand, the *second-order condition* states that if f is twice differentiable, i.e., if its Hessian $\nabla^2 f$ exists at each point in $\mathbf{dom}\ f$, then f is convex if, and only if, its domain is convex and

$$\nabla^2 f(\mathbf{x}) \succeq 0, \tag{2.5}$$

that is, its Hessian matrix is positive semidefinite.

A important property of a convex function is that the associated *sublevel set* is convex, where the $\alpha-$sublevel set $S_\alpha f$ is defined as

$$S_\alpha f \triangleq \{\mathbf{x} \in \mathrm{dom} f : f(\mathbf{x}) \leq \alpha\}. \tag{2.6}$$

2.1.2 Definition of Convex Problems

The general notation to describe a constrained optimization problem is as follows [Boy04], [Ber99]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, M, \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, P. \end{aligned} \tag{2.7}$$

This notation represents the problem of finding the value of $\mathbf{x}$ that minimizes $f_0(\mathbf{x})$ (called the *objective* or *cost function*) among all $\mathbf{x}$ that satisfy the conditions $f_i(\mathbf{x}) \leq 0$, $i = 1, \ldots, M$ and $h_i(\mathbf{x}) = 0$, $i = 1, \ldots, P$. The variable $\mathbf{x}$ is called the *primal optimization variable*. The inequalities $f_i(\mathbf{x}) \leq 0$ and their corresponding functions $f_i : \mathbb{R}^N \rightarrow \mathbb{R}$ are called the *inequality constraints* and *inequality constraint functions*, respectively. Conversely, the equations $h_i(\mathbf{x}) = 0$ and their corresponding functions $h_i : \mathbb{R}^N \rightarrow \mathbb{R}$ are called the *equality constraints* and *equality constraint functions*, respectively. If there are no constraints, the problem is called *unconstrained optimization problem*.

The set of points for which the objective and all constraint functions are defined is called the *domain* of the optimization problem (2.7) and it is expressed as

$$\mathcal{D} = \bigcap_{i=0}^{M} \mathbf{dom}\, f_i \ \cap \ \bigcap_{i=1}^{P} \mathbf{dom}\, h_i. \tag{2.8}$$

A point $\mathbf{x} \in \mathcal{D}$ is *feasible* if it satisfies the constraints $f_i(\mathbf{x}) \leq 0$, $i = 1, \ldots, M$ and $h_i(\mathbf{x}) = 0$, $i = 1, \ldots, P$, simultaneously. If there exists at least one feasible point in the problem (2.7), then the problem is said to be feasible, and *infeasible* otherwise. The set of all feasible points is called the *feasible set* or the *constraint set*.

The *optimal value* $p^\star$ of the problem (2.7) is defined as

$$p^\star = \inf \{ f_0(\mathbf{x}) \mid f_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, M,\ h_i(\mathbf{x}) = 0,\ i = 1, \ldots, P \}. \tag{2.9}$$

The optimal value $p^\star$ is allowed to have values within the range $\pm\infty$. If the optimal value is $p^\star = \infty$, then the problem is infeasible, whereas the problem is said to be unbounded below if $p^\star = -\infty$.

If $\mathbf{x}^\star$ is feasible and $f_0(\mathbf{x}^\star) = p^\star$, then we say $\mathbf{x}^\star$ is an *optimal point* or *optimal solution*. The set of all optimal points is called the *optimal set*, which is expressed as

$$\mathcal{X} = \{ \mathbf{x} \mid f_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, M,\ h_i(\mathbf{x}) = 0,\ i = 1, \ldots, P,\ f_0(\mathbf{x}) = p^\star \}. \tag{2.10}$$

It is said that the optimum value is *attained* or *achieved* if there exists at least one optimal point for the problem (2.7). Accordingly, if $\mathcal{X}$ is empty, it is said that the optimal value is not attained or not achieved.

A feasible point $\mathbf{x}$ is *locally optimal* if there exists $R > 0$ such that

$$f_0(\mathbf{x}) = \inf \{ f_0(\mathbf{z}) \mid f_i(\mathbf{z}) \leq 0,\ i = 1, \ldots, M,\ h_i(\mathbf{z}) = 0,\ i = 1, \ldots, P,\ ||\mathbf{z} - \mathbf{x}||_2 \leq R \}. \tag{2.11}$$

All optimal points are also locally optimal, but the converse is not always true. There may exist some locally optimal points that do not yield optimal solutions (i.e., globally optimal).

The problem (2.7) is said to be a *convex optimization problem* if the objective function is convex, if the inequality constraint functions $f_i(\mathbf{x})$, $i = 1, \ldots, M$ are all convex, and if the equality constraint functions $h_i(\mathbf{x}), i = 1, \ldots, P$ are all affine, i.e., $h_i(\mathbf{x}) = \mathbf{a}_i^T\mathbf{x} - b_i, i = 1, \ldots, P$. Based on the definition of the domain of the optimization problem, it is easy to realize that the domain of a convex optimization problem is also convex, since it is the intersection of the set of domains of all the functions, which are convex. Additionally, the constraint set is also convex since it is the intersection of M convex sublevel sets, and P hyperplanes. A fundamental property of convex optimization problems is that any locally optimal point is at the same time globally optimal [Boy04]. Sometimes it is useful to transform the original problem into an equivalent problem. This is done mostly when the original problem is not convex. These tricks may include a change of variables or the introduction of slack variables. Some of these techniques will be used and commented along this thesis whenever needed.

The most common convex optimization problems are *linear programming* (LP)[2], *quadratic programming* (QP)[3], *semidefinite programming* (SDP)[4], and *geometric programming* (GP)[5] problems. For a large descriptions of convex optimization problems, see [Boy04].

A *concave optimization problem* is a problem where the objective function is to be maximized, the objective function is concave, the inequality constraint functions $f_i(\mathbf{x})$, $i = 1, \ldots, M$ are convex, and the equality constraint functions $h_i(\mathbf{x})$, $i = 1, \ldots, P$ are affine. This problem can be easily transformed to a standard convex optimization problem by just minimizing the convex function $-f_0(\mathbf{x})$ subject to the same original constraints.

2.1.3 Duality Theory and KKT Conditions

Given the optimization problem in standard form (2.7) (not necessarily convex), and assuming that its domain is not empty, the *Lagrangian* of the problem can be defined as

$$L(\mathbf{x}; \boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq f_0(\mathbf{x}) + \sum_{i=1}^{M} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{P} \nu_i h_i(\mathbf{x}), \tag{2.12}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)^T$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_P)^T$ are the vector of *Lagrange multipliers* associated to the inequality and equality constraint functions, respectively. The Lagrange multipliers are commonly referred as *dual variables*. We define the *Lagrange dual function* (or just *dual function*)

[2] An optimization problem is said to be LP if the objective function and all the constraints functions are linear.

[3] An optimization problem is said to be QP if the objective function is quadratic and the constraint functions are affine.

[4] An optimization problem is said to be SDP if the objective function is linear and the constraint set is formed by the intersection of the cone of positive semidefinite matrices with an affine space.

[5] An optimization problem is said to be GP if the objective function and the inequality constraint functions are posynomials, and the equality constraint functions are monomials. It is important to note that GP problems are generally not convex in their standard formulation. Nevertheless, they can be transformed to convex problems by a change of variables and a transformation of the objective and constraint functions.

as the infimum of the Lagrangian over $\mathbf{x}$

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}; \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{M} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{P} \nu_i h_i(\mathbf{x}) \right), \tag{2.13}$$

such that $\boldsymbol{\lambda} \in \mathbb{R}^M, \boldsymbol{\nu} \in \mathbb{R}^P$ and $\mathcal{D}$ is the domain of the original problem (2.7). When the Lagrangian is unbounded below, the dual function takes on the value $-\infty$. Since the dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is a point-wise infimum of a family of affine functions of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$, it is concave even when the problem (2.7) is not convex [Boy04]. A point $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is said to be *dual feasible* if $\boldsymbol{\lambda} \succeq 0$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbf{dom}\, g(\cdot)$. An important statement is that the dual function evaluated at any dual feasible point yields a lower bound on the optimal value $p^\star$ of the original problem (2.7). The proof is as follows:

$$\begin{aligned} p^\star = f_0(\mathbf{x}^\star) &\geq f_0(\mathbf{x}^\star) + \sum_{i=1}^{M} \lambda_i f_i(\mathbf{x}^\star) + \sum_{i=1}^{P} \nu_i h_i(\mathbf{x}^\star) \\ &\geq \inf_{\mathbf{z} \in \mathcal{D}} \left(f_0(\mathbf{z}) + \sum_{i=1}^{M} \lambda_i f_i(\mathbf{z}) + \sum_{i=1}^{P} \nu_i h_i(\mathbf{z}) \right) \\ &= g(\boldsymbol{\lambda}, \boldsymbol{\nu}). \end{aligned} \tag{2.14}$$

The best lower bound that can be obtained from the Lagrange dual function may be found by solving the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \succeq 0. \end{aligned} \tag{2.15}$$

This problem is called the *Lagrange dual problem* (or just *dual problem*) associated with the problem (2.7). Let $d^\star$ be the optimal value of problem (2.15) and let $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ be the optimum dual variables at which the optimum value is attained, i.e., $d^\star = g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$. As it was mentioned before, it is the best lower bound on $p^\star$ resulting from the definition of the dual function:

$$d^\star \leq p^\star. \tag{2.16}$$

The previous inequality always holds even if the original problem is not convex and the difference $p^\star - d^\star$ is known as duality gap. This property is known as *weak duality*. On the other hand, if the equality

$$d^\star = p^\star, \tag{2.17}$$

is fulfilled, then it is said that *strong duality* holds, implying zero duality gap. This shows that under certain conditions, the bound obtained from the Lagrange dual problem is tight. However, strong duality does not always hold. In order to assure strong duality in convex problems, it is only needed to prove that some mild technical conditions are satisfied. These conditions are

called *constraint qualifications*. An example of a constraint qualification is *Slater's condition*. It states that strong duality holds if there exists a point $\mathbf{x}$ such that $f_i(\mathbf{x}) < 0$, $i = 1, \ldots, M$ and $h_i(\mathbf{x}) = 0\,, i = 1, \ldots, P$ (this point is commonly referred as *strictly feasible* point).

If we assume that strong duality holds for given primal optimal $\mathbf{x}^\star$ and dual optimal $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ variables, the following expressions hold

$$\begin{aligned} f_0(\mathbf{x}^\star) &= g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) && (2.18)\\ &= \inf_{\mathbf{x}\in\mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{M} \lambda_i^\star f_i(\mathbf{x}) + \sum_{i=1}^{P} \nu_i^\star h_i(\mathbf{x}) \right) && (2.19)\\ &\leq f_0(\mathbf{x}^\star) + \sum_{i=1}^{M} \lambda_i^\star f_i(\mathbf{x}^\star) + \sum_{i=1}^{P} \nu_i^\star h_i(\mathbf{x}^\star) && (2.20)\\ &\leq f_0(\mathbf{x}^\star). && (2.21) \end{aligned}$$

An important conclusion can be drawn from the previous statements. Notice that the last inequality must be an equality, which implies that

$$\sum_{i=1}^{M} \lambda_i^\star f_i(\mathbf{x}^\star) = 0. \tag{2.22}$$

Since each term in this sum is nonpositive, it can be concluded that

$$\lambda_i^\star f_i^\star(\mathbf{x}^\star) = 0, \quad i = 1, \ldots, M. \tag{2.23}$$

This condition is known as *complementary slackness*, and it holds for any primal and any dual optimal variables. As a consequence, the complementary slackness implies that

$$\lambda_i^\star > 0 \implies f_i(\mathbf{x}^\star) = 0, \tag{2.24}$$

or,

$$f_i(\mathbf{x}^\star) < 0 \implies \lambda_i^\star = 0. \tag{2.25}$$

Using all the previous statements and assuming that all functions $(f_0, \ldots, f_M, h_1, \ldots, h_P)$ are differentiable, we are able to conclude that the following equations must be fulfilled for any primal and dual optimal variables:

$$\begin{aligned} f_i(\mathbf{x}^\star) &\leq 0, \qquad i = 1, \ldots, M, && (2.26)\\ h_i(\mathbf{x}^\star) &= 0, \qquad i = 1, \ldots, P, && (2.27)\\ \lambda_i^\star &\geq 0, \qquad i = 1, \ldots, M, && (2.28)\\ \lambda_i^\star f_i^\star(\mathbf{x}^\star) &= 0, \qquad i = 1, \ldots, M, && (2.29)\\ \nabla f_0(\mathbf{x}^\star) + \sum_{i=1}^{M} \lambda_i^\star \nabla f_i(\mathbf{x}^\star) + \sum_{i=1}^{P} \nu_i^\star \nabla h_i(\mathbf{x}^\star) &= 0. && (2.30) \end{aligned}$$

The previous equations are called the KKT conditions. In practice, the KKT conditions are very useful to obtain optimal solutions analytically (whenever this is possible) or to support the application of numerical methods to solve those equations.

To summarize, for *any* optimization problem with differentiable objective and constraint functions for which strong duality holds, any pair of primal and dual optimal points must satisfy the KKT conditions (2.26) - (2.30). This means that KKT conditions are always necessary conditions for optimality. If, moreover, the problem is *convex*, KKT conditions are also sufficient for optimality. In other words, if the points $\mathbf{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ satisfy the KKT conditions, then these points are primal and dual optimal, with zero duality gap.

2.1.4 Solving Convex Problems

As commented before, there is no general analytical formula to solve convex optimization problems. However, in cases where the objective and constraint functions are differentiable (which means that the KKT conditions hold), being able to solve the KKT equations may yield to a closed-form solution of the problem due to the fact that KKT conditions are necessary and sufficient for optimality.

Notice, nevertheless, that even if not closed-form solutions can be obtained, an efficient numerical algorithm can always be applied to achieve the optimal value thanks to the fundamental property that states that any locally optimal point is also a globally solution. This motivates the search for very fast algorithms to solve convex optimization problems. Indeed, almost real-time algorithms for solving convex problems can be found today. Among the most famous ones, the *interior point methods* (also called *barrier methods*) must be emphasized. These algorithms reach the optimal solution by solving a sequence of smooth unconstrained problems, usually using Newton's method [Boy04]. Generally, the interior-point based methods are able to provide not only the optimal primal variables, but also the optimal dual ones (that is, the Lagrange multipliers). These techniques are called *primal-dual interior point methods*.

2.2 Ergodic Stochastic Optimization Theory

In this section, we are going to present an approach to solve optimization problems that involve time varying random variables with ergodic objective and constraint functions. The framework under which these problems can be solved is known as *ergodic stochastic optimization theory* or *stochastic approximation theory*, and it was first introduced by [Wan07]. It has been applied to different scenarios (see the extensive literature [Rib10a], [Rib10c], [Hu10], [Hu11], [Gat09], [Rib08a], [Rib08b], [Hu13], [Hu12], [Hu11], [Rib12], [Rib10d], [Gat10], [Mar12], [Mar09], [Mar13], [Gat14], [FB13], and [Mar11b]). However, it was in [Rib10b] where A. Ribeiro developed an extensive tutorial and a summary of the framework of ergodic stochastic optimization theory.

This theory will be used in some chapters throughout this dissertation as the technique employed to solve some optimization problems. As the main theoretical proofs of this technique rely on the theory presented in [Rib10b], those proofs will not be included in this dissertation. Nevertheless and for the sake of completeness, in this section, we will develop a summary of the main contributions of [Rib10b].

2.2.1 Problem Formulation

Let us consider a problem involving a time varying random state $\mathbf{h}(t) \in \mathcal{H}$ with probability distribution function (PDF) $m_{\mathbf{h}}(\mathbf{h}) = m(\mathbf{h})$, a resource allocation function $\mathbf{p}(\mathbf{h}(t))$ associated with state realization $\mathbf{h}$ and having a PDF $m_{\mathbf{p}(\mathbf{h})}[\mathbf{p}(\mathbf{h})] = m(\mathbf{p}(\mathbf{h}))$, and an ergodic limit variable $\mathbf{x} = \lim_{t\to\infty}(1/t)\sum_{u=1}^{t}\mathbf{x}(u)$, being $\mathbf{x}(t)$ a resource allocation variable. The goal is to design an adaptive algorithm that observes $\mathbf{h}(t)$ to determine $\mathbf{p}(\mathbf{h}(t))$ and $\mathbf{x}(t)$ without knowledge of the state's distribution $m(\mathbf{h})$ in order to satisfy given problem constraints and optimality criteria. In other words, the goal is to determine the resource allocation functions and ergodic limits that are optimal in the sense of solving the following optimization problem:

$$\begin{aligned} p^{\star} = \text{maximize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & C1 : \mathbf{x} \preceq \mathbb{E}_{\mathbf{h}}\left[\mathbb{E}_{m(\mathbf{p}(\mathbf{h}))}\left[\mathbf{f}_1(\mathbf{p}(\mathbf{h});\mathbf{h})\right]\right] \\ & C2 : \mathbf{f}_2(\mathbf{x}) \succeq \mathbf{0} \\ & C3 : \mathbf{x} \in \mathcal{X} \\ & C4 : \{m(\mathbf{p}(\mathbf{h})) : \mathbf{p}(\mathbf{h}) \in \mathcal{P}(\mathbf{h})\}_{\mathbf{h}\in\mathcal{H}}\,. \end{aligned} \tag{2.31}$$

The previous optimization is w.r.t. ergodic limits $\mathbf{x}$ and probability distributions $m(\mathbf{p}(\mathbf{h}))$ for all $\mathbf{h} \in \mathcal{H}$. We emphasize that the expected value is taken w.r.t. the PDF's $m(\mathbf{h})$ of the state $\mathbf{h}$ and $m(\mathbf{p}(\mathbf{h}))$ of the resource allocation $\mathbf{p}(\mathbf{h})$, and while $m(\mathbf{h})$ is fixed the PDFs $m(\mathbf{p}(\mathbf{h}))$ are part of the optimization space.

The optimization problem formulated in (2.31) appears usually in optimal resource allocation problems where infinite time horizons allowing performance characterization through ergodic limit is considered. The system is affected by a random state with realizations $\mathbf{h}(t)$. In response to the observed state $\mathbf{h}(t)$, a resource allocation variable $\mathbf{p}(t) \in \mathcal{P}(\mathbf{H}(t))$ measuring how many units of a certain resource are allocated at time t is determined. Allocation of $\mathbf{p}(t)$ units of resource when the random state is $\mathbf{h}(t)$, results in the production of $\mathbf{f}_1(\mathbf{p}(t;\mathbf{h}(t)))$ units of goods. In the same time slot t, consumption is determined by $\mathbf{x}(t) \in \mathcal{X}$ variables. Consumption cannot exceed production, but if long time horizons are of interest, instead of imposing such restriction for every t, it suffices to constraint the limits of the time averages, i.e.,

$$\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}\mathbf{x}(u) \preceq \lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}\mathbf{f}_1(\mathbf{p}(u);\mathbf{h}(u)). \tag{2.32}$$

The first constraint in (2.31) follows upon defining the ergodic limit $\mathbf{x} = \lim_{t\to\infty}(1/t)\sum_{u=1}^{t}\mathbf{x}(u)$ and assuming ergodicity in the limit on the right hand side of (2.32). The constraint $\mathbf{f}_2(\mathbf{x}) \geq \mathbf{0}$ imposes further restrictions on the ergodic average $\mathbf{x}$.

Functions $f_0(\mathbf{x})$ and $f_2(\mathbf{x})$ in (2.31) are assumed to be concave w.r.t. their argument $\mathbf{x}$. The family of functions $\mathbf{f}_1(\mathbf{p}(\mathbf{h}); \mathbf{h})$ is parametrized by the random state $\mathbf{h}$ and, different from $f_0(\mathbf{x})$ and $f_2(\mathbf{x})$, is not necessarily concave w.r.t. the resource allocation $\mathbf{p}(\mathbf{h})$. The sole requirement for the functions $\mathbf{f}_1(\mathbf{p}(\mathbf{h}); \mathbf{h})$ is that they be finite for finite arguments, i.e., for every bounded vector of resources $\mathbf{p}(\mathbf{h}) \prec \infty$, the vector function $\mathbf{f}_1(\mathbf{p}(\mathbf{h}); \mathbf{h})$ is also bounded (i.e., $\mathbf{f}_1(\mathbf{p}(\mathbf{h}); \mathbf{h}) \prec \infty$). The set $\mathcal{X}$ is compact and convex while the set $\mathcal{P}(\mathbf{h})$ is compact but not necessarily convex.

2.2.2 Problem Resolution

Because of the distributions $m(\mathbf{p}(\mathbf{h}))$, there is an infinite number of variables in the primal domain. However, observe that there is a finite number of constraints, which means that the dual problem contains a finite number of variables hinting that the problem is likely more tractable in the dual domain. Let $\boldsymbol{\lambda}_1 \succeq \mathbf{0}$ and $\boldsymbol{\lambda}_2 \succeq \mathbf{0}$ be the vector of dual variables associated with constraints $C1$ and $C2$ in problem (2.31), respectively. Using these definitions, the Lagrangian for the optimization problem (2.31) can be defined as

$$\begin{aligned}\mathcal{L}\left(\mathbf{x}, m(\mathbf{p}(\mathbf{h}))\,; \boldsymbol{\lambda}\right) &= f_0(\mathbf{x}) + \boldsymbol{\lambda}_1^T\left(\mathbb{E}_{\mathbf{h}}\left[\mathbb{E}_{m(\mathbf{p}(\mathbf{h}))}\left[\mathbf{f}_1(\mathbf{p}(\mathbf{h}); \mathbf{h})\right]\right] - \mathbf{x}\right) + \boldsymbol{\lambda}_2^T\mathbf{f}_2(\mathbf{x}) \qquad (2.33)\\ &= f_0(\mathbf{x}) - \boldsymbol{\lambda}_1^T\mathbf{x} + \boldsymbol{\lambda}_2^T\mathbf{f}_2(\mathbf{x}) + \mathbb{E}_{\mathbf{h}}\left[\mathbb{E}_{m(\mathbf{p}(\mathbf{h}))}\left[\boldsymbol{\lambda}_1^T\mathbf{f}_1(\mathbf{p}(\mathbf{h}); \mathbf{h})\right]\right], \qquad (2.34)\end{aligned}$$

where we have defined the aggregate dual variable $\boldsymbol{\lambda} \triangleq [\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T]^T$. The dual function is defined as the maximum of the Lagrangian over the set of feasible ergodic limits $\mathbf{x} \in \mathcal{X}$ and probability distributions $m(\mathbf{p}(\mathbf{h})$ in the set of feasible powers $\mathbf{p}(\mathbf{h}) \in \mathcal{P}(\mathbf{h})$, i.e.,

$$\begin{aligned} g(\boldsymbol{\lambda}) = \sup \quad & \mathcal{L}\left(\mathbf{x}, m(\mathbf{p}(\mathbf{h}))\,; \boldsymbol{\lambda}\right) \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \quad \left\{m(\mathbf{p}(\mathbf{h})) : \mathbf{p}(\mathbf{h}) \in \mathcal{P}(\mathbf{h})\right\}_{\mathbf{h}\in\mathcal{H}}. \end{aligned} \qquad (2.35)$$

The dual problem is defined as the minimization of the dual function over all positive dual variables, i.e.,

$$d^\star = \min_{\boldsymbol{\lambda}\succeq\mathbf{0}}\ g(\boldsymbol{\lambda}). \qquad (2.36)$$

Let us now introduce a discrete time index t and consider a state stochastic process denoted by $\mathbf{H}(\mathbb{N})$ with realizations $\mathbf{h}(\mathbb{N})$ having values $\mathbf{h}(t)$ independent and identically distributed (i.i.d.) according to $m(\mathbf{h})$. The proposed ergodic algorithm consists of the iterative application of the following steps:

1. *Primal Iteration:* given multipliers $\boldsymbol{\lambda}(t)$, find primal variables $\mathbf{x}(t)$ and $\mathbf{p}(\mathbf{h}(t)) \in \mathcal{P}(\mathbf{h}(t))$ such that

$$\mathbf{x}(t) = \mathbf{x}(\boldsymbol{\lambda}(t)) = \arg\max_{\mathbf{x}\in\mathcal{X}} \; f_0(\mathbf{x}) - \boldsymbol{\lambda}_1^T(t)\mathbf{x} + \boldsymbol{\lambda}_2^T(t)\mathbf{f}_2(\mathbf{x}), \tag{2.37}$$

$$\mathbf{p}(t) = \mathbf{p}(\mathbf{h}(t), \boldsymbol{\lambda}(t)) = \arg\max_{\mathbf{p}(\mathbf{h}(t))\in\mathcal{P}(\mathbf{h}(t))} \; \boldsymbol{\lambda}_1^T(t)\mathbf{f}_1(\mathbf{p}(\mathbf{h}(t)); \mathbf{h}(t)). \tag{2.38}$$

2. *Dual Stochastic Subgradients:* define the stochastic subgradients $\hat{\mathbf{s}}(t) = \hat{\mathbf{s}}(\mathbf{h}(t), \boldsymbol{\lambda}(t)) = [\hat{\mathbf{s}}_1^T(t), \hat{\mathbf{s}}_2^T(t)]$ of the dual functions with components:

$$\hat{\mathbf{s}}_1(t) \triangleq \mathbf{f}_1(\mathbf{p}(t); \mathbf{h}(t)) - \mathbf{x}(t), \tag{2.39}$$

$$\hat{\mathbf{s}}_2(t) \triangleq \mathbf{f}_2(\mathbf{x}(t)). \tag{2.40}$$

3. *Dual Iteration:* the iteration is completed with an update in the dual domain with a predetermined step size ϵ along the direction $-\hat{\mathbf{s}}(t)$:

$$\boldsymbol{\lambda}(t+1) = (\boldsymbol{\lambda}(t) - \epsilon\hat{\mathbf{s}}(t))_0^\infty \tag{2.41}$$

$$= \begin{pmatrix} \boldsymbol{\lambda}_1(t) - \epsilon(\mathbf{f}_1(\mathbf{p}(t); \mathbf{h}(t)) - \mathbf{x}(t)) \\ \boldsymbol{\lambda}_2(t) - \epsilon\mathbf{f}_2(\mathbf{x}(t)) \end{pmatrix}_0^\infty, \tag{2.42}$$

where $(x)_a^b$ is the projection of x onto the interval $[a, b]$, i.e., $(x)_a^b = \min\{\max\{a, x\}, b\}$.

Note that, in order to compute the update of the dual variables we used a stochastic unbiased subgradient instead of the actual true subgradient. Note that, the true subgradient would involve the expectations, i.e., $\mathbf{s}_1 \triangleq \mathbb{E}_{\mathbf{h}}\left[\mathbb{E}_{m(\mathbf{p}(\mathbf{h}))}\left[\mathbf{f}_1(\mathbf{p}(\mathbf{h}); \mathbf{h})\right]\right] - \mathbf{x}$ and $\mathbf{s}_2 \triangleq \mathbf{f}_2(\mathbf{x})$. The use of a stochastic subgradient follows the same philosophy as the one considered in the derivation of the well-known gradient descent algorithm to minimize mean squared errors with the least mean square (LMS) algorithm [Hay02]. In fact, in [Rib10b], the author shows the following result:

Proposition 2.1. *Let $\boldsymbol{\lambda}(t)$ be given and define $g(t) \triangleq g(\boldsymbol{\lambda}(t))$. Then, the conditional expected value $\mathbb{E}[\hat{\boldsymbol{s}}(t) \,|\, \boldsymbol{\lambda}(t)]$ of the stochastic subgradient $\hat{\mathbf{s}}(t)$ in (2.39) is a subgradient of the dual function. Specifically, for arbitrary $\boldsymbol{\lambda}$*

$$\mathbb{E}[\hat{\boldsymbol{s}}^T(t) \,|\, \boldsymbol{\lambda}(t)](\boldsymbol{\lambda}(t) - \boldsymbol{\lambda}) \geq g(t) - g(\boldsymbol{\lambda}). \tag{2.43}$$

Proof. See [Rib10b]. ■

The previous proposition establishes that the expected value of the stochastic subgradient is a descent direction of the dual function. Since the dual function is always convex, we expect that a descend algorithm constructed by replacing $\mathbb{E}[\hat{\mathbf{s}}(t) \,|\, \boldsymbol{\lambda}(t)]$ for $\hat{\mathbf{s}}(t)$ in (2.42) would eventually

approach an optimal dual variable $\boldsymbol{\lambda}^\star$. Since the stochastic subgradient $\hat{\mathbf{s}}(t)$ varies randombly around its mean $\mathbb{E}[\hat{\mathbf{s}}(t) \,|\, \boldsymbol{\lambda}(t)]$, it is reasonable to expect that iterates $\boldsymbol{\lambda}(t)$ of (2.42) will also come close to $\boldsymbol{\lambda}^\star$. In fact, this reasoning is formalized with the following result:

Theorem 2.1. *Consider the ergodic algorithm defined in* (2.37)-(2.42). *Let* $\hat{S}^2 \geq \mathbb{E}[||\hat{\boldsymbol{s}}(t)||^2 \,|\, \boldsymbol{\lambda}(t)]$ *be a bound on the second moment of the norm of the stochastic subgradients* $\hat{\boldsymbol{s}}(t)$ *and assume that there exist strictly feasible* $\boldsymbol{x}_0 \in \mathcal{X}$ *and* $m_0(\boldsymbol{p}(\boldsymbol{h}))$ *such that* $\mathbb{E}_{\boldsymbol{h}}\left[\mathbb{E}_{m_0(\boldsymbol{p}(\boldsymbol{h}))}\left[\boldsymbol{f}_1(\boldsymbol{p}(\boldsymbol{h}); \boldsymbol{h})\right]\right] - \boldsymbol{x}_0 > C$ *and* $\boldsymbol{f}_2(\boldsymbol{x}_0) > C$ *for some strictly positive constant* $C > 0$. *Assume that dual variable* $\boldsymbol{\lambda}(T_0)$ *at given time* T_0 *is given and define the best value of the dual function at time* t *as* $g_{best}[t \,|\, \boldsymbol{\lambda}(T_0)] \triangleq \min_{u \in [T_0, t]} g(u)$. *Such best value almost surely converges to within* $(\epsilon \hat{S}^2)/2$ *of* $d^\star$, *i.e.,*

$$\lim_{t \to \infty} g_{best}[t \,|\, \boldsymbol{\lambda}(T_0)] - d^\star \leq \frac{\epsilon \hat{S}^2}{2} \quad a.s. \tag{2.44}$$

Proof. See [Rib10b]. ■

Let us now come back to the original ergodic algorithm. Resource allocation iterates $\mathbf{p}(t)$ are functions of dual iterates $\boldsymbol{\lambda}(t)$ and state realizations $\mathbf{h}(t)$. Since $\mathbf{h}(t)$ is random, so is $\mathbf{p}(t)$, implying that the sequence $\mathbf{p}(\mathbb{N})$ is a realization of a stochastic process $\mathbf{P}(\mathbb{N})$. In fact, all other sequences $\mathbf{x}(\mathbb{N})$, $\boldsymbol{\lambda}(\mathbb{N})$, and $\hat{\mathbf{s}}(\mathbb{N})$ are also realizations of random processes.

Solving an optimization problem such as (2.31) entails finding the optimal value $p^\star$ and the optimal arguments $\mathbf{x}^\star$ and $m^\star(\mathbf{p}(\mathbf{h}))$ such that the constraints in (2.31) are satisfied and $p^\star = f_0(\mathbf{x}^\star)$. As we have ergodic constraints, we adopt a different definition of solution. The goal is not to find $\mathbf{x}^\star$ and $m^\star(\mathbf{p}(\mathbf{h}))$ but to show that sequences $\mathbf{x}(\mathbb{N})$ and $\mathbf{p}(\mathbb{N})$ generated by the ergodic algorithm in (2.37)-(2.42) satisfy (2.32) with the ergodic limit $\mathbf{x}$ of the sequence $\mathbf{x}(\mathbb{N})$ further satisfying $\mathbf{f}_2(\mathbf{x}) \succeq \mathbf{0}$ and $p^\star \approx f_0(\mathbf{x})$. Because the algorithm is stochastic, these results will be established in probability. Specifically, consider the following results:

Theorem 2.2. *Consider the optimization problem in* (2.31) *and sequences* $\boldsymbol{x}(\mathbb{N})$ *and* $\boldsymbol{p}(\mathbb{N})$ *generated by the ergodic algorithm defined in* (2.37)-(2.42). *Let* $\hat{S}^2 \geq \mathbb{E}[||\hat{\boldsymbol{s}}(t)||^2 \,|\, \boldsymbol{\lambda}(t)]$ *be a bound on the second moment of the norm of the stochastic subgradients* $\hat{\boldsymbol{s}}(t)$ *and assume that there exist strictly feasible* $\boldsymbol{x}_0 \in \mathcal{X}$ *and* $m_0(\boldsymbol{p}(\boldsymbol{h}))$ *such that* $\mathbb{E}_{\boldsymbol{h}}\left[\mathbb{E}_{m_0(\boldsymbol{p}(\boldsymbol{h}))}\left[\boldsymbol{f}_1(\boldsymbol{p}(\boldsymbol{h}); \boldsymbol{h})\right]\right] - \boldsymbol{x}_0 > C$ *and* $\boldsymbol{f}_2(\boldsymbol{x}_0) > C$ *for some strictly positive constant* $C > 0$. *Then*

i) Almost sure feasibility. Sequences $\boldsymbol{x}(\mathbb{N})$ *and* $\boldsymbol{p}(\mathbb{N})$ *are feasible with probability 1, i.e.,*

$$\lim_{t \to \infty} \frac{1}{t} \sum_{u=1}^{t} \boldsymbol{x}(u) \preceq \lim_{t \to \infty} \frac{1}{t} \sum_{u=1}^{t} \boldsymbol{f}_1(\boldsymbol{p}(u); \boldsymbol{h}(u)), \quad a.s., \tag{2.45}$$

$$\boldsymbol{f}_2\left(\lim_{t \to \infty} \frac{1}{t} \sum_{u=1}^{t} \boldsymbol{x}(u)\right) \succeq \boldsymbol{0}, \quad a.s. \tag{2.46}$$

ii) *Almost sure near optimality. The ergodic average of $f_0(\boldsymbol{x})$ converges almost surely to a value with optimality gap smaller than $\epsilon\hat{S}^2/2$, i.e.,*

$$p^\star - f_0\left(\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}\boldsymbol{x}(u)\right) \leq \frac{\epsilon\hat{S}^2}{2}, \quad a.s. \tag{2.47}$$

Proof. See [Rib10b]. ■

It is important to elaborate on what (2.45)-(2.47) imply in terms of finding the solution of (2.31). The ergodic limit $\mathbf{x} \triangleq \frac{1}{t}\sum_{u=1}^{t}\mathbf{x}(u)$ satisfies the constraints in (2.31) and the objective function evaluated at $\mathbf{x}$ is within $\epsilon\hat{S}^2/2$ of optimal. Since $\mathcal{X}$ and $\mathcal{P}(\mathbf{h})$ are compact sets, it follows that the bound $\hat{S}^2$ is finite. Therefore, by reducing ϵ it is possible to make $f_0(\mathbf{x})$ arbitrarily close to $p^\star$ and, as a consequence, $\mathbf{x}$ arbitrarily close to some optimal argument $\mathbf{x}^\star$. The optimal resource allocation distribution $m^\star(\mathbf{p}(\mathbf{h}))$, however, is not computed by the ergodic algorithm. Rather, (2.45) implies that, asymptotically, the ergodic algorithm is drawing resource allocation realizations $\mathbf{p}(t)$ from a resource allocation distribution that is close to the optimal $m^\star(\mathbf{p}(\mathbf{h}))$. This is not a drawback in practice because the realizations $\mathbf{p}(t)$ are sufficient for implementation. In that sense, (2.37)-(2.42) yield an optimal resource allocation policy (allocation of $\mathbf{p}(t)$ units at time t) that supports the optimal consumption $\mathbf{x}$ in an ergodic sense.

Note that the problem formulation in (2.31) makes a distinction between constraints $\mathbf{f}_2(\mathbf{x}) \succeq \mathbf{0}$ and $\mathbf{x} \in \mathcal{X}$. While one is expressed as an inequality and the other as a set inclusion, both are convex constraints in the ergodic limit $\mathbf{x}$. However, they are intended to model different constraint modalities. The constraint $\mathbf{f}_2(\mathbf{x}) \succeq \mathbf{0}$ is dualized and incorporated into the Lagrangian in (2.33) and becomes a maximization objective in the primal ergodic iteration in (2.37). As a consequence, it is satisfied in an ergodic sense. The ergodic limits of sequences $\mathbf{x}(\mathbb{N})$ satisfy $\mathbf{f}_2(\mathbf{x}) \succeq \mathbf{0}$ but individual variables $\mathbf{x}(t)$ might or might not satisfy $\mathbf{f}_2(\mathbf{x}) \succeq \mathbf{0}$. On the other hand, the constraint $\mathbf{x} \in \mathcal{X}$ is not incorporated into the Lagrangian in (2.33) and is an implicit constraint in the primal ergodic iteration in (2.37). It is, thus, satisfied for all times, i.e., $\mathbf{x}(t) \in \mathcal{X}, \forall t$. This is an important distinction in applications (e.g., transmitted power in wireless communications must comply with both, ergodic and instantaneous power constraints).

2.3 Multi-Objective Optimization

In this section, we are going to generalize the theory presented before and consider that multiple objective functions must be optimized at the same time. In this sense, multi-objective optimization (also known as multicriteria optimization, vector optimization, etc.) is a type of mathematical optimization problem that involves more than one objective function to be optimized simultaneously [Ehr05]. For a nontrivial multi-objective optimization problem, there

does not exist a single solution that simultaneously optimizes each objective. In that case, the objective functions are said to be conflicting, and there exists a (possibly infinite) number of Pareto optimal solutions. A solution is called Pareto optimal if none of the objective functions can be improved in value without degrading some of the other objective values.

2.3.1 Definitions

Definition 2.1 ([Ehr05]). *A multi-objective problem can be formally expressed as*

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{maximize}} \quad \boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_K(\boldsymbol{x})) \\ &\text{subject to} \quad \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2.48}$$

where $f_k : \mathbb{C}^N \to \mathbb{R}$ for $k = 1, \ldots, K$ and $\mathcal{X}$ is the feasible set that represents the constraints. Let $\mathcal{Y}$ be the set of all attainable points for all feasible solutions, i.e., $\mathcal{Y} = \boldsymbol{f}(\mathcal{X})$.

2.3.2 Efficient Solutions

Definition 2.2 ([Ehr05], Definition 2.1). *A point $\boldsymbol{x} \in \mathcal{X}$ is called Pareto optimal if there is no other $\boldsymbol{x}' \in \mathcal{X}$ such that $\mathbf{f}(\mathbf{x}') \succeq \mathbf{f}(\mathbf{x})$, where $\succeq$ refers to the component-wise inequality, i.e., $f_i(\boldsymbol{x}') \geq f_i(\boldsymbol{x})$, $i = 1, \ldots, K$.*

Sometimes, ensuring Pareto optimality for some problems is difficult. Due to this, the condition of optimality can be relaxed as follows.

Definition 2.3 ([Ehr05], Definition 2.24). *A point $\boldsymbol{x} \in \mathcal{X}$ is called weakly Pareto optimal (or weakly efficient) if there is no other $\boldsymbol{x}' \in \mathcal{X}$ such that $\mathbf{f}(\mathbf{x}') \succ \mathbf{f}(\mathbf{x})$, where $\succ$ refers to the strict component-wise inequality, i.e., $f_i(\boldsymbol{x}') > f_i(\boldsymbol{x})$, $i = 1, \ldots, K$. All Pareto optimal solutions are also weakly Pareto optimal.*

2.3.3 Finding Pareto Optimal Points

There are several methods for finding the Pareto points of a multi-objective problem. In the sequel, we present three different (scalarization) techniques.

3.1) Weighted sum method: the simplest scalarization technique is the weighted sum method which collapses the vector-objective into a single-objective component sum:

$$\underset{\mathbf{x} \in \mathcal{X}}{\text{maximize}} \quad \sum_{k=1}^{K} \beta_k f_k(\mathbf{x}), \tag{2.49}$$

where β_k are real non-negative weights. The following results present the relation between the optimal solutions of (2.49) and the Pareto optimal points of the original problem (2.48).

Proposition 2.2 ([Ehr05], Proposition 3.9)**.** *Suppose that $\boldsymbol{x}^\star$ is an optimal solution of* (2.49)*. Then, $\boldsymbol{x}^\star$ is weakly efficient.*

Proposition 2.3 ([Ehr05], Proposition 3.10)**.** *Let $\mathcal{X}$ be a convex set, and let f_k be concave functions, $k = 1, \ldots, K$. If $\boldsymbol{x}^\star$ is weakly efficient, there are some $\beta_k \geq 0$ such that $\boldsymbol{x}^\star$ is an optimal solution of* (2.49)*.*

As as result, convexity is apparently required for finding all weakly Pareto optimal points with the weighted sum method, which means that if the original problem is not convex, all the Pareto optimal points may not be found by using the weighted sum method. However, there are other weighted sum techniques in the literature (see, for example, the adaptive weighted sum method [Kim06]) that are able to find all Pareto optimal points for nonconvex problems at the expense of a higher computational complexity.

2.3.3.1 Epsilon-Constraint Method

In this method, only one of the original objectives is maximized while the others are transformed into constraints:

$$\begin{aligned} \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad & f_j(\mathbf{x}) \\ \text{subject to} \quad & f_k(\mathbf{x}) \geq \epsilon_k, \quad k = 1, \ldots, K,\ k \neq j. \end{aligned} \tag{2.50}$$

Let us introduce the following results.

Proposition 2.4 ([Ehr05], Proposition 4.3)**.** *Let $\boldsymbol{x}^\star$ be an optimal solution of* (2.50) *for some j. Then $\boldsymbol{x}^\star$ is weakly Pareto optimal.*

Proposition 2.5 ([Ehr05], Proposition 4.5)**.** *A feasible solution $\boldsymbol{x}^\star \in \mathcal{X}$ is Pareto optimal if, and only if, there exists a set of $\hat{\epsilon}_k, k = 1, \ldots, K$ such that $\boldsymbol{x}^\star$ is an optimal solution of* (2.50) *for all $j = 1, \ldots, K$.*

Contrarily to the weighted sum method, convexity is not needed in the previous two propositions (but convexity is still typically required to solve problems like (2.50)).

2.3.3.2 Hybrid Method

This method combines the previous two methods, i.e., the weighted sum method and the epsilon-constraint method. In this case, the scalarized problem to be solved has a weighted sum objective

and constraints on all (or some) objectives.

$$\begin{aligned} \underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad & \sum_{k\in\mathcal{K}_1} \beta_k f_k(\mathbf{x}) \\ \text{subject to} \quad & f_k(\mathbf{x}) \geq \epsilon_k, \quad k \in \mathcal{K}_2, \end{aligned} \tag{2.51}$$

where $|\mathcal{K}_1| \leq K$ and $|\mathcal{K}_2| \leq K$, being $|\mathcal{A}|$ the cardinality of set $\mathcal{A}$, and β_k are real non-negative weights.

2.4 The Majorization-Minimization Optimization Method

The majorization-minimization or minorization-maximization (MM) method is an approach to solve some optimization problems that are too difficult to be solved directly. The principle behind the MM method is to transform a difficult problem into a series of simple problems. In the following, we present a brief summary of this technique. Interested readers may refer to [Hun04] and references therein for an extensive summary of the theory and applications.

The method works as follows. Suppose that we want to maximize $f_0(\mathbf{x})$ over $\mathcal{X}$. In the MM approach, instead of maximizing the cost function $f_0(\mathbf{x})$ directly, the algorithm optimizes a sequence of approximate objective functions that minorize $f_0(\mathbf{x})$, producing a sequence $\{\mathbf{x}^{(k)}\}$ according to the following update rule

$$\mathbf{x}^{(k+1)} = \underset{\mathbf{x}\in\mathcal{X}}{\arg\max} \ \hat{f}_0(\mathbf{x}, \mathbf{x}^{(k)}), \tag{2.52}$$

where $\mathbf{x}^{(k)}$ is the point generated by the algorithm at iteration k, and $\hat{f}_0(\mathbf{x}, \mathbf{x}^{(k)})$, known as surrogate function, is the minorization function of $f_0(\mathbf{x})$ at $\mathbf{x}^{(k)}$. Formally, the function $\hat{f}_0(\mathbf{x}, \mathbf{x}^{(k)})$ is said to minorize the function $f_0(\mathbf{x})$ at point $\mathbf{x}^{(k)}$ if

$$\hat{f}_0(\mathbf{x}, \mathbf{x}^{(k)}) \leq f_0(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X}, \tag{2.53}$$

$$\hat{f}_0(\mathbf{x}^{(k)}, \mathbf{x}^{(k)}) = f_0(\mathbf{x}^{(k)}). \tag{2.54}$$

That means that the surrogate function $\hat{f}_0(\mathbf{x}, \mathbf{x}^{(k)})$ is a global lower bound for $f_0(\mathbf{x})$. The surrogate function must also be continuous in $\mathbf{x}$ and $\mathbf{x}^{(k)}$. The last condition that the surrogate function must fulfill is that the directional derivatives[6] of itself and of the original objective

[6] Let $f : \mathbb{R}^N \to \mathbb{R}$. Then, the directional derivative of $f(\mathbf{x})$ in the direction of vector $\mathbf{d}$ is given by $f'(\mathbf{x}; \mathbf{d}) \triangleq \lim_{\lambda\to 0} \frac{f(\mathbf{x}+\lambda\mathbf{d}) - f(\mathbf{x})}{\lambda}$.

function $f_0(\mathbf{x})$ must be equal at the point $\mathbf{x}^{(k)}$. All in all, the four conditions are

$$(\text{A1}): \quad \hat{f}_0(\mathbf{y},\mathbf{y}) = f_0(\mathbf{y}), \quad \forall \mathbf{y} \in \mathcal{X}, \tag{2.55}$$

$$(\text{A2}): \quad \hat{f}_0(\mathbf{x},\mathbf{y}) \leq f_0(\mathbf{x}), \quad \forall \mathbf{x},\mathbf{y} \in \mathcal{X}, \tag{2.56}$$

$$(\text{A3}): \quad \hat{f}_0'(\mathbf{x},\mathbf{y};\mathbf{d})|_{\mathbf{x}=\mathbf{y}} = f_0'(\mathbf{y};\mathbf{d}), \quad \forall \mathbf{d} \text{ with } \mathbf{y}+\mathbf{d} \in \mathcal{X}, \tag{2.57}$$

$$(\text{A4}): \quad \hat{f}_0(\mathbf{x},\mathbf{y}) \text{ is continuous in } \mathbf{x} \text{ and } \mathbf{y}. \tag{2.58}$$

It can be checked easily that the sequence of points generated by the MM algorithm decreases monotonously the function to be minimized. However, the key property of this algorithm is that, in addition to the previous comment, under assumptions (A1) – (A4), it can be proved (see [Hun04] for details) that every limit point of the sequence $\{\mathbf{x}^{(k)}\}$ is a locally optimal point of the original problem, even if this problem is not convex. Of course, if the problem is convex, the global optimum solution can be obtained.

2.5 Extension to the Case of Complex Variables

In the previous sections, we have assumed that the involved functions were real-valued functions of real arguments. However, the majority of applications in signal processing and communications involve complex variables. The applicability of the previous mathematical tools is almost identical if complex arguments are considered; the only thing that needs to be revised is the derivation rule. Let $f : \mathbb{C}^N \rightarrow \mathbb{R}$ denote a real-valued function of complex argument and let $\mathbf{x} = (x_1,\ldots,x_N) = \mathbf{x}^{(r)} + j\mathbf{x}^{(i)} \in \mathbb{C}^N$ denote a complex vector, being $\mathbf{x}^{(r)} = \text{Re}\{\mathbf{x}\}$ and $\mathbf{x}^{(i)} = \text{Im}\{\mathbf{x}\}$ the real part and the imaginary part of vector $\mathbf{x}$, respectively, and where j is the imaginary unit. Said that, the derivation rule assumed throughout this dissertation is given by [Hjo11]

$$\nabla_{x_m} f(\mathbf{x})\Big|_{\mathbf{x}=\tilde{\mathbf{x}}} = \frac{1}{2}\left(\frac{\partial f(\mathbf{x})}{\partial x_m^{(r)}}\Big|_{\mathbf{x}=\tilde{\mathbf{x}}} - j\frac{\partial f(\mathbf{x})}{\partial x_m^{(i)}}\Big|_{\mathbf{x}=\tilde{\mathbf{x}}}\right), \tag{2.59}$$

$$\nabla_{x_m^*} f(\mathbf{x})\Big|_{\mathbf{x}=\tilde{\mathbf{x}}} = \frac{1}{2}\left(\frac{\partial f(\mathbf{x})}{\partial x_m^{(r)}}\Big|_{\mathbf{x}=\tilde{\mathbf{x}}} + j\frac{\partial f(\mathbf{x})}{\partial x_m^{(i)}}\Big|_{\mathbf{x}=\tilde{\mathbf{x}}}\right), \tag{2.60}$$

where the operator $(\cdot)^*$ denotes the complex conjugate of $\mathbf{x}$, i.e., $\mathbf{x}^* = \mathbf{x}^{(r)} - j\mathbf{x}^{(i)}$. Following the previous convention, if a real-valued function f is being optimized w.r.t. the variable $\mathbf{x}$ by means of the steepest ascent or descent method, it follows that the updating term must be proportional to $\nabla_{\mathbf{x}^*} f(\mathbf{x})$, and not $\nabla_{\mathbf{x}} f(\mathbf{x})$ [Hjo11]. Hence, the update equation for optimizing the real-valued function by means of the steepest ascent or descent method can be expressed as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu \nabla_{\mathbf{x}^*} f(\mathbf{x}^{(k)}), \tag{2.61}$$

where μ is a real positive constant if it is a maximization problem or a real negative constant if it is a minimization problem.

Part II

Energy Harvesting Techniques at the Receiver Side

Chapter 3

Energy-Aware Resource Allocation for Battery-Constrained Receivers

3.1 Introduction

As it was mentioned in the introduction of this dissertation, recently, there has been a considerable expansion of wireless networks jointly with a continuous increase of the number of users. This expansion involves a need for a substantial increase of system capacity. In wireless networks, this capacity increase is technically challenging since the resources to be shared among users are limited. At the same time, in order to be more spectrally efficient, the general trend is to reduce the cell coverage radius (such as in picocells and femtocells in cellular environments). As it was motivated in the introduction of this thesis (see Section 1.1.1), in networks with short distances between transmitter and receivers, the radiated powers can be comparable with or even lower than the powers consumed by the front-end and the baseband stages [Gro11], [Cui07]. In [Aue11] and [Deb11], authors showed that in microcells the radiated power is practically the same as the power consumed by the baseband signal processing blocks, and for pico and femtocells the power consumed by the baseband stage can be even higher than the actual radiated power coming from the HPA.

On the other hand, one of the important issues mentioned in the introduction of this thesis is the fact that one of the limiting factors of current technology is the short lifetime of the batteries. Due to such short lifetimes, the high data rate demanded by the terminals entails situations where the users run out of battery noteworthy fast. In wireless sensors networks, this can be a serious issue, since such sensors are placed in positions that cannot be accessed to replace their batteries. In cellular environments, the telecommunication providers has put a lot of attention on providing good services with enhanced coverage, but this will not be translated into a really perceived added value if the users cannot make use of them due to the mentioned battery limitations.

Within this framework, the current work on energy harvesting is emerging [Par05], [Sud11].

Traditionally, energy harvesting techniques have been developed based on energy sources such as, for example, wind or solar energy. Nevertheless, as introduced before in Section 1.1.2, there are other techniques that could be applied to moving sensors (this may be the case of cellular phones) based on piezoelectric technologies. All this suggests that new strategies for allocating the radio resources should be developed, considering explicitly such battery-related aspects, i.e., the harvesting capabilities and battery status of the terminals. As an opposing aspect, it is shown in [Par05], [Sud11], and [Lu15] that the levels of energy harvesting that can be achieved by current technology are still low (few micro or mili Watts). Nevertheless, the research community is making a lot of effort to improve these numbers and the results are being quite promising. Despite the numbers, there are already many widely recognized published papers in the literature (that will be presented later in this introduction) that are taking energy harvesting into explicit consideration, showing a clear trust in the potentiality of such technology.

In classical precoder design strategies for MIMO systems, a given objective function is optimized subject to some constraints typically considering only the radiated power. The main goal of this chapter is to consider the online precoder design and resource allocation in a multiuser broadcast system incorporating not only the idea that the nodes are battery-limited devices provided with energy harvesting capabilities, but also considering explicitly the power needed for decoding the received data. Thus, the information concerning the status of the battery levels will play an explicit role and have an impact on the proposed design.

3.1.1 Related Work

Some works can be found in the literature concerning resource allocation and scheduling based on energy efficiency metrics in multiuser systems, but to the best of our knowledge, none of them has taken into account the information about the battery levels of the terminals. In [Cui04], authors developed strategies to optimize the modulation type based on a global energy minimization while satisfying given throughput and delay requirements. In [Cui07], authors presented a joint design of the PHY, MAC, and routing layers to minimize network energy consumption. Note that, most of previous works are based on the quotient metric of bits per Joule (J) that was firstly introduced by Verdú in [Ver90]. For example, Miao et al. developed resource allocation strategies based on this quotient (e.g. [Mia10]). The main problem associated with this approach is that, even if such quotient is maximized, the data rate obtained by the optimization problem can be very small and not enough for some applications. If, however, minimum rate constraints are considered, then the problem reduces to a classical design problem where the goal is the minimization of the transmitted power under rate constraints. In addition, these works do not take into account in the proposed strategies either the current battery status or the energy spent by the receiver.

Resource allocation in point-to-point wireless links where the transmitter has energy har-

vesting capabilities has also been a research topic lately. For example, in [Yan12c], authors analyzed two different situations. First, they considered that the arrival times and the amounts of energy of the energy harvesting packets were known and that the node had all the data to be transmitted from the beginning and, second, they assumed that all data was not available from the beginning but the data packets arrivals were known (as well as the energy packets). In this approach, authors considered that the battery size of the node was infinite. Later in [Oze11], they extended the previous work to the case of fading channels. In [Tut12], a finite battery capacity transmitter was considered and the minimum completion time for the data transmission was found considering that the data was available from the beginning. In [Ho12b], authors introduced a specific model for the harvesting source based on Markov chains and used convex optimization theory and dynamic programming to obtain the best rate scheduling or power allocation strategy. The effect of battery imperfections in terms of battery leakage was studied in [Dev12b]. There have been also some attempts to derive optimal policies for the multiuser scenario, although the area is not mature yet and only some preliminary papers can be found in the literature. The problem was partially extended to the broadcast channel in [Yan12a], [Ant11], to the relay channel in [Gü11], and to the multiple access channel in [Yan11]. Finally, some works considering both the radiated power and the circuitry power consumption can also be found (e.g., in [Gre13b], a resource allocation policy that maximizes asymptotically the data-rate in a discrete fading environment was derived).

3.1.2 Main Contribution

The main difference of our work when compared to the previous works is that, in those works, the authors did not consider the receivers to be battery-limited devices, but the transmitter instead. Taking into account this framework and the state of the art, in the following, we summarize the main contributions of our work:

- Description of different decoding power consumption models.
- Proposal of an online precoder design and data rate allocation strategy taking into account the state of the batteries of the terminals (most of the works in the literature focus on the transmitter side).
- Extension to multiuser MIMO scenarios from previous works.
- Development of a precoder design considering imperfect CSI and evaluation of its impact on the evolution of the data rate and battery levels of the terminals.
- Proposal of a robust design based on imperfect knowledge of the battery due to the finite data rate feedback link used to collect the battery information.

- Asymptotic analysis and characterization of the battery and data rate evolution of the users in the proposed system.
- User scheduling strategies considering the current battery levels of the users and the average throughput achieved by the users.

3.1.3 Organization of the Chapter

The remainder of this chapter is organized as follows. The energy modeling is presented in Section 3.3. Section 3.4 covers the precoder design with perfect CSI, imperfect CSI and a robust design based on imperfect battery knowledge. Section 3.4.5 addresses the analysis of the asymptotic behavior of the proposed strategy. Section 3.5 presents a user scheduling strategy to form groups of users to be served and, finally, conclusions are provided in Section 3.6.

3.2 Signal Model

Let us consider a set of K users indexed by $k \in \mathcal{K} \triangleq \{1, \ldots, K\}$. We focus on a broadcast scenario where the receivers have multiple antennas and are served simultaneously by a multiple-antenna transmitter BS. In this system, the k-th user is provided with n_{R_k} antennas and the BS with n_T antennas, respectively. This is commonly denoted as $\{n_{R_1}, ..., n_{R_K}\} \times n_T$. The total number of receive antennas is $n_R = \sum_{i=1}^{K} n_{R_i}$. We assume frequency flat and slow varying fading channels such that the coherence time of the channel is larger than the codewords length. In that case, the information-theoretic Shannon's formula is a good approximation of the instantaneous maximum achievable rate. In case of frequency selective channels, the same MIMO processing that will be proposed in this thesis could be performed on each subcarrier if orthogonal frequency division multiplexing (OFDM) transmission is applied. Let us denote the number of substreams assigned to user k as n_{S_k}. We consider that $n_{S_k} = \min\{n_{R_k}, n_T - (n_R - n_{R_k})\}\, \forall k$ is fulfilled[1]. We index frames by $t \in \mathcal{T} \triangleq \{1, \ldots, T\}$ with a duration of T_f seconds each. We consider that the channels remain constant within a frame and change between consecutive frames. The signal model for the baseband samples of the received signals at the receive antennas for the k-th user at the n-th time instant within the t-th frame is

$$\mathbf{y}_k(n,t) = \mathbf{H}_k(t) \sum_{j=1}^{K} \mathbf{B}_j(t)\mathbf{x}_j(n,t) + \mathbf{n}_k(n,t), \tag{3.1}$$

[1]This condition is imposed by the dimensions of the equivalent channel matrices that will be obtained from the application of the BD technique, as will be explained in detail in Section 3.4.1. As it will be shown later also in Section 3.4.1, finally, the number of substreams to be sent will depend not only on the number of antennas, but also on the eigenvalues of the involved channel matrices. As a result from the optimum power allocation, it may happen that some substreams will be assigned zero power, so that they will not be used to transmit information.

where $\mathbf{y}_k(n,t) \in \mathbb{C}^{n_{R_k}\times 1}$ is the received signal vector, $\mathbf{H}_k(t) \in \mathbb{C}^{n_{R_k}\times n_T}$ is the MIMO channel matrix from the BS to the k-th user, where the (p,q)-th entry represents the channel gain from the BS antenna q to antenna p of user k, and $\mathbf{B}_k(t) \in \mathbb{C}^{n_T\times n_{S_k}}$ is the precoder matrix of user k. The transmitted signal for user k is $\mathbf{B}_k(t)\mathbf{x}_k(n,t)$ and its covariance matrix is $\mathbf{Q}_k(t) = \mathbf{B}_k(t)\mathbf{B}_k^H(t)$ if we assume, without loss of generality (w.l.o.g.), that $\mathbb{E}\left[\mathbf{x}_k(n,t)\mathbf{x}_k^H(n,t)\right] = \mathbf{I}_{n_{S_k}}$, where $\mathbf{x}_k(n,t) \in \mathbb{C}^{n_{S_k}\times 1}$ represents the zero-mean data vector for user k, being $x_{k_i}(n,t)$ the i-th symbol to be communicated. Finally, $\mathbf{n}_k(n,t) \in \mathbb{C}^{n_{R_k}\times 1}$ is the additive zero mean circularly complex symmetric (ZMCCS) Gaussian noise with $\mathbb{E}\left[\mathbf{n}_k(n,t)\mathbf{n}_k^H(n,t)\right] = \sigma^2\mathbf{I}_{n_{R_k}}$. For the sake of clarity, we will drop the frame and time dependence whenever possible.

Let $\tilde{\mathbf{x}} = \mathbf{B}\mathbf{x} \in \mathbb{C}^{n_T\times 1}$ denote the signal vector transmitted by the BS, where the joint precoding matrix $\mathbf{B}$ is defined as $\mathbf{B} = [\mathbf{B}_1,\ldots,\mathbf{B}_K] \in \mathbb{C}^{n_T\times n_S}$, being $n_S = \sum_{i=1}^{K} n_{S_i}$ the total number of substreams, and the data vector as $\mathbf{x} = \left[\mathbf{x}_1^T,\ldots,\mathbf{x}_K^T\right]^T \in \mathbb{C}^{n_S\times 1}$. According to the previous notation, the power constraint can be formulated as $\mathbb{E}[\|\tilde{\mathbf{x}}\|^2] = \sum_{k=1}^{K}\mathrm{Tr}(\mathbf{Q}_k) \leq P_T$, where P_T represents the total power radiated at the BS and it has been assumed that the information symbols of different users are independent.

The receivers must inform the BS about their current battery level status in order to incorporate this knowledge in the design of the precoders. This will be based on a causal procedure, that is, the BS will allocate the resources with the current knowledge of the available energy in the users' batteries.

3.3 Energy Modeling

3.3.1 Power Consumption Models

The energy consumed by the transceiver can be modeled as the energy consumed by the front-end plus the energy consumed by the coding/decoding stages (omitting for the moment the power radiated at the transmitter). Although other works consider battery imperfections in their models [Dev12b], we do not consider this in our work for the sake of simplicity. Note, however, that the strategy and formulation presented in this chapter could be extended easily to incorporate those imperfections. In general terms, it should be emphasized that there do not exist precise general models widely accepted by the community for the energy consumption for a transceiver [MD13]. In the following we will comment briefly which is the generic abstract approach followed in this thesis in order to make the proposed strategies independent of the concrete model.

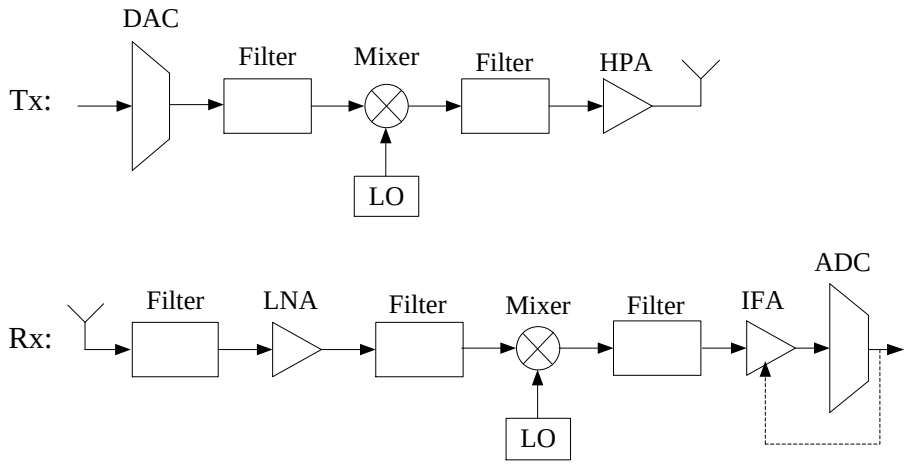


Figure 3.1: RF chain model for the transmitter and the receiver considered in this thesis.

3.3.1.1 Front-end Consumption

As far as the transmitter is concerned, the components that consumes energy are the HPA, the mixers, the filters, and other elements of the RF chain. The RF chain model considered in this thesis is depicted in Figure 3.1. It is important to point out that the energy consumed by the digital to analog converter (DAC) depends on the sampling frequency which depends on the communication bandwidth and is considered as fixed in this thesis. Concerning the receiver, the front-end consumption usually depends on the condition on the channel, i.e., the signal to noise ratio (SNR) (in practice, the receiver should adapt the front-end according to the received power [Jen12], an operation that requires some additional power). As mentioned for the transmitter, the power consumed by the ADC placed at the receiver strongly depends on the sampling frequency. However, in this thesis, we consider the bandwidth to be fixed and so is the sampling rate. In the following, we assume that the component of the receiver front-end consumption that depends on the SNR is negligible [Jen12].

We denote the power consumed by the front-end at the transmitter and the receiver by P_c^{tx} and P_c^{rx}, respectively, following the specific consumption model of each front-end component based on the work [Cui05] (without considering the specific power that is to be radiated by the HPA).

3.3.1.2 Coding/Decoding Stages Consumption

It is reasonable to consider the energy consumed by the coding stage at the transmitter negligible compared to the energy consumed by the front-end (the HPA is the element that consumes most). For this reason, we will not include coding consumption in our models.

On the other hand, the decoding consumption must be included in the models since, as shown in [Gro11], such energy consumption is not negligible and can affect importantly the

lifetime of the mobile terminal. Yet, there has been little work on the modeling of decoding consumption, and there is not any model extensively recognized and accepted by the research community, as Doost and Yates mention explicitly in the introduction of their work [MD13]. However, there is a consensus about the fact that the decoding consumption increases with the data rate (corresponding to the coding rate) R (in fact, it increases with the number of iterations required by the decoding process) for a given target bit error rate (BER). Besides, such coding complexity increases (and therefore, the decoding consumption power) if the target BER decreases, as described in [Gro11]. In order to make our work as general as possible, in the rest of the chapter we will assume an abstract generic function relating the rate and the power consumed by the decoder, denoted as $P_{\text{dec},k}(R_k)$. In the following, we present two simple examples of such decoding consumption function extracted from the literature[2]:

- in [Cui03], authors claim that $P_{\text{dec},k}(R_k)$ grows linearly with the data rate of the k-th user R_k as

$$P_{\text{dec},k}(R_k) = \nu_k R_k, \tag{3.2}$$

  where ν_k is decoder and user specific and models the decoder efficiency.

- in [Ros10], authors propose an exponential model for the decoding complexity. They assume that the state space and the number of possible state transitions in the decoder-trellis expands exponentially with the data rate. According to this, they state that the computation power can be expressed as

$$P_{\text{dec},k}(R_k) = c_{1k} e^{c_{2k} R_k}, \tag{3.3}$$

  where the constants c_{jk} are decoder and user specific and model the decoder efficiency. The previous model may be the object of some criticism due to the fact that the computation power is not zero when the rate assigned to the user is zero. A possible interpretation could be that it corresponds to the power consumption of the electronics of the decoder when it is in idle state.

The two models presented before are rather simple and provide just two illustrative examples. The community should come up with more elaborated models as well as with real measured values for the constants (ν and c_j), which are still open research topics. Due to these reasons and for the sake of generality, in this chapter we consider a generic model denoted by $P_{\text{dec},k}(R_k)$, as already commented previously.

[2] In the following, and in order not to complicate the notation, we assume that a given target BER is given and, therefore, the expressions of the power consumption models depend only on the rate. Nevertheless, if the target BER decreases, the power consumption should increase accordingly.

3.3.1.3 Total Transceiver Consumption

In this section we present the final total power consumption of the transmitter and the receiver, including front-end and decoding stages.

Total Transmitter Consumption: the total consumption at the transmitter is modeled just by considering the front-end consumption as mentioned previously and, thus, it is denoted as

$$P_{\text{tot}}^{t_x} = P_c^{t_x}. \tag{3.4}$$

Total Receiver Consumption: the total power consumption at the receiver is modeled as

$$P_{\text{tot},k}^{r_x}(R_k) = P_{\text{dec},k}(R_k) + P_c^{r_x}. \tag{3.5}$$

Let us add a comment stating that, in case that we want to use a more general model for the front-end power consumption that depends on the channel quality, the only thing that we should do is to generalize the previous expression as $P_{\text{tot},k}^{r_x}(R_k, \mathbf{H}_k) = P_{\text{dec},k}(R_k) + P_c^{r_x}(\mathbf{H}_k)$, which would require the knowledge of the function $P_c^{r_x}(\mathbf{H}_k)$. In any case, if CSI is available, then, for a given concrete channel, $P_{\text{tot},k}^{r_x}(R_k, \mathbf{H}_k)$ would be a function of the rate, and the techniques and strategies proposed in this thesis would still be valid and could be used.

3.3.2 Energy Harvesting Model

We assume a discretized model for the energy arrivals [Yan12a], [Gre13b] where $H_k(t)$ is modeled as an ergodic Bernoulli stochastic process. As a result, only two values of harvested energy are possible, i.e., $H_k(t) \in \{0, e_k\}$, where e_k is the amount of energy, in Joules, contained in an energy packet. The probability of receiving an energy harvesting packet during one frame depends on the actual harvesting intensity (in the case of solar energy, it depends on the particular hour of the day) and is denoted by $p_k(t)$. Note that a higher value of $p_k(t)$ will be obtained in frames where the harvesting intensity is higher, e.g., during the day, and a lower value of $p_k(t)$ during the night or during cloudy days. Note that, in some cases, it is possible to predict or estimate partially the expected energy available to be harvested. This situation can be included in our formulation by adjusting our model parameters by knowing that $\mathbb{E}[H_k(t)] = p_k(t) \cdot e_k$. As a result, the majority of harvesting sources can be modeled with the previous stochastic model.

3.3.3 Battery Dynamics

We consider that each user terminal is provided with a battery with finite capacity and an energy harvesting source that allows to collect energy dynamically from the environment. As

we mentioned before, $H_k(t)$ is the energy harvested in Joules by the k-th user that is available just after the t-th frame. In order to describe the model in an accurate way, we provide the following definitions:

Definition 3.1. (Battery with harvesting). *The battery of the k-th user at the beginning of the t-th frame after adding the energy collected by the harvesting source and subtracting the energy used for decoding is denoted as $C_k(t) = \min\{C^k_{\max}, C_k(t-1) - f(R_k(t-1)) + H_k(t-1)\}$, where $f(R_k(t-1))$ is a function described later in the chapter and $R_k(t-1)$ is the data rate attained by user k during $(t-1)$-th frame.*

3.3.4 Energy Assignment for Decoding

As the users have a limited battery capacity, they must use a finite amount of energy to decode the received message. Based on this, let us introduce the following definition:

Definition 3.2. (Energy available for decoding). *The energy allowed to be spent by the k-th user terminal during the t-th frame for decoding the transmitted information is denoted by $E^k_d(t)$. This energy quantity is taken into account by the resource allocation process at the BS. It must fulfill the following condition: $E^k_d(t) \leq C_k(t)$.*

Notice that $E^k_d(t)$ is different from the current battery level $C_k(t)$ since, as it will be shown later, generally the resource allocation strategy will not allow to use all the available battery level; instead only a fraction of $C_k(t)$ will be used in order to constrain the energy spent by the receiver when decoding the transmitted information and, therefore, increase the battery lifetime. In any case, let us assume for the moment that $E^k_d(t)$ is given (later in Section 3.4.4, we will show that, in fact, $E^k_d(t)$ can be optimized (offline) so that the aggregate sum rate is maximized).

According to (3.5) and Definition 3.2, the energy that is allowed to be consumed by the k-th receiver within the t-th frame is constrained as

$$T_f \cdot P^{rx}_{\text{tot},k}(R_k(t)) = T_f \cdot \left(P_{\text{dec},k}(R_k(t)) + P^{rx}_c\right) \leq E^k_d(t). \tag{3.6}$$

As it is expressed, the energy consumed by the receiver must be lower than the maximum allowed by the resource allocation algorithm, $E^k_d(t)$, that will be calculated as a fraction of the current battery level of the user, as it will be proposed later. This is, in fact, a constraint to be added in the resource allocation algorithm. The previous constraint can also be written in terms of an equivalent upper bound on the maximum data-rate to be supported by the k-th user at the t-th frame,

$$R_k(t) \leq R_{\max,k}(C_k(t)), \tag{3.7}$$

which provides a general framework regardless of the model of the decoding power consumption.

The relation between $R_{\max,k}(\tilde{C}_k(t))$ and the energy allowed by the resource allocation is

$$R_{\max,k}(C_k(t)) = {P_{\text{tot},k}^{r_x}}^{-1}\left(\frac{E_d^k(t)}{T_f}\right). \tag{3.8}$$

If the exponential model is considered, the relation between the maximum data-rate and the energy is

$$R_{\max,k}(C_k(t)) = \frac{1}{c_{2k}}\ln\left(\frac{E_d^k(t) - P_c^{r_x}T_f}{T_f c_{1k}}\right). \tag{3.9}$$

For the linear power consumption model, the relation between the maximum data-rate and the energy is

$$R_{\max,k}(C_k(t)) = \frac{1}{\nu_k}\left(\frac{E_d^k(t) - P_c^{r_x}T_f}{T_f}\right). \tag{3.10}$$

The resource allocator calculates the value of $E_d^k(t)$ for each user at the beginning of each frame. This value represents a given portion of the current battery level $C_k(t)$ of the receiver terminal and represents the energy that is allowed to be spent by the k-th user during the following frame (see Definition 3.2). If the harvesting source has a constant mean intensity, and the BS has only causal information of the incoming energy packets, a proper solution is to assign a fixed percentage of $C_k(t)$:

$$E_d^k(t) = \alpha_k C_k(t), \qquad 0 \leq \alpha_k \leq 1. \tag{3.11}$$

The impact of the value of α_k on the design lies on the fact that high values of α_k would allow to achieve high instantaneous data-rates but the battery level would decrease fast. Nevertheless, if the battery decreases considerably fast, the data-rate obtained in future frames may be low, making the average data-rate be potentially relatively lower. On the other hand, a relatively small α_k value would constrain the maximum instantaneous data-rate, increasing the lifetime of the user notably. Notice, then, that α_k controls the trade-off between the instantaneous data-rate and the lifetime of the users' batteries. The main goal of the thesis is to show that, by considering explicitly the energy constraints formulated in (3.6), it is possible to enhance both the lifetime and the average data-rate simultaneously by taking a proper value for the parameter α_k. A discussion on how to optimally select the numerical value of α_k is addressed in Section 3.4.4.

Note that the real-time optimization and accurate adjustment of the value of α_k for each frame would require the knowledge of the exact statistical properties of the harvesting source (i.e., the current mean intensity of the harvesting) or, even a stronger assumption, knowing the energy that would be harvested in all frames in advance, i.e., in a non-causal way. This assumption is sometimes considered in the literature [Yan12a], however, it is obviously not possible to implement such an approach in practice. In some cases, such as solar harvesting, if the mean intensity of the harvesting changes throughout the day and we know such intensity

Algorithm 3.1 Battery update and energy allocation

initialization: Set $t = 1$, $\alpha_k > 0$, and maximum rates:

$$E_d^k(t) = \alpha_k C_k(t),\ \forall k$$

$$R_{\max,k}(t) = {P_{\text{tot},k}^{r_x}}^{-1}\left(\frac{E_d^k(t)}{T_f}\right),\ \forall k$$

1: **solve** optimization problem and obtain: (see Sections 3.4.1, 3.4.2, and 3.4.3)

$$R_k^\star(t), \mathbf{B}_k^\star(t),\ \forall k$$

2: **update** battery level according to data-rate allocated and harvesting:

$$C_k(t+1) = \min\{C_{\max}^k, C_k(t) - T_f \cdot P_{\text{tot},k}^{r_x}\left(R_k^\star(t)\right) + H_k(t)\},\ \forall k$$

3: **update** energies and maximum rates of users:

$$E_d^k(t+1) = \alpha_k C_k(t+1),\ \forall k$$

$$R_{\max,k}(t+1) = {P_{\text{tot},k}^{r_x}}^{-1}\left(\frac{E_d^k(t+1)}{T_f}\right),\ \forall k$$

4: **set** $t \longleftarrow t+1$ and go to step 1

variation, we could adjust the value of α_k over the day accordingly to match the quantity that is allowed to be spent with the energy that is being harvested. This will be commented with more detail in the simulations section.

The procedure for the assignments of the maximum energies allowed for decoding the information and the battery updates according to the energy spendings and the harvesting capabilities following the previous definitions are presented in Algorithm 3.1.

3.4 Energy-Aware Multiuser MIMO Precoder Design

3.4.1 Precoder Design with Perfect CSI and Energy Constraints

In this section, we present an online precoder design strategy for the already presented multiuser system where users are battery-constrained mobile terminals provided with energy harvesting capabilities. We consider the design of the transmit precoding matrices $\{\mathbf{B}_k\}$ assuming perfect CSI at both ends of the communication link, as depicted in Figure 3.2. We consider, therefore, that the receiver is capable of estimating the channel perfectly and sending it to the transmitter through an ideal feedback channel (although in practical communications such assumption is rather unrealistic). The extension to the case of imperfect CSI will be studied in Section 3.4.2.

The optimum transmission policy in a MIMO broadcast channel is the well-known non-linear dirty paper coding [Gol03]. Nevertheless, such design strategy is highly computational demanding and cannot be computed in real time. Instead, much simpler linear transceiver designs have been shown to achieve almost the same performance with a much lower computational

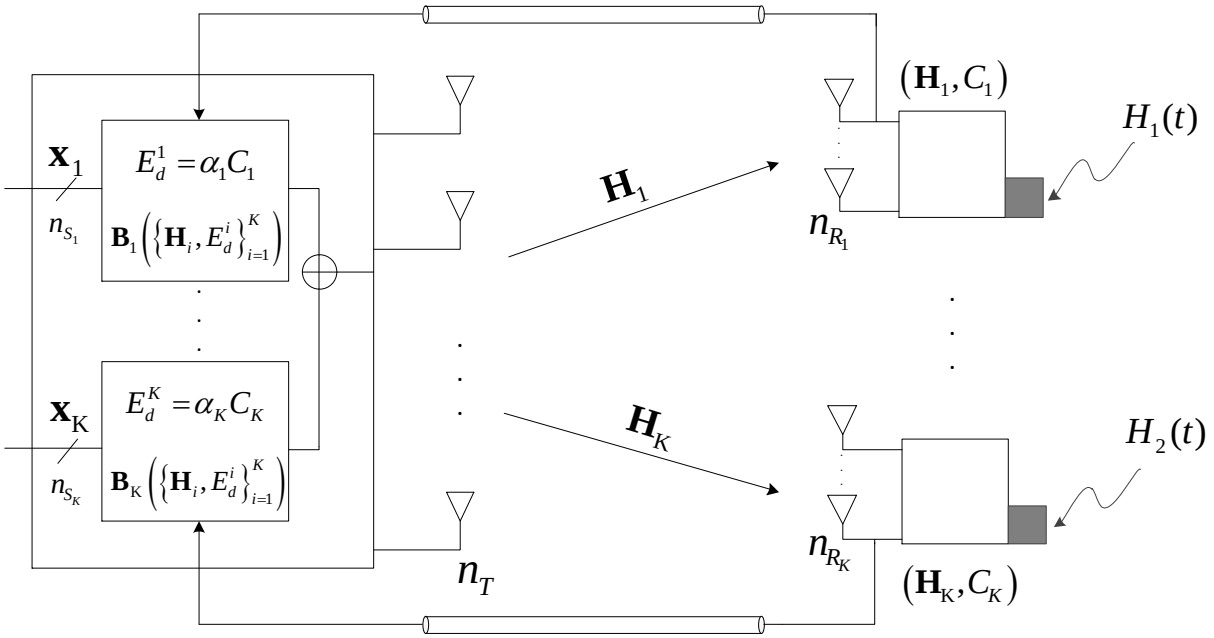


Figure 3.2: Schematic representation of the DL broadcast multiuser communication system. Note that each user feeds back its current battery level and the current CSI. The BS designs the precoder matrices using the feedback information.

complexity [Lee06]. Thus, for simplicity in the transmitter design, in the following we force the precoder matrix to be linear (as already assumed in the signal model (3.1)).

Given that, the rate R_k that can be allocated to the k-th user under linear precoding has to be lower than or equal to the maximum achievable rate, as expressed in the following [Lee06]:

$$R_k = \log \frac{\det\left(\sigma^2\mathbf{I} + \sum_{j=1}^{K} \mathbf{H}_k\mathbf{Q}_j\mathbf{H}_k^H\right)}{\det\left(\sigma^2\mathbf{I} + \sum_{j=1,j\neq k}^{K} \mathbf{H}_k\mathbf{Q}_j\mathbf{H}_k^H\right)}, \tag{3.12}$$

or in a more compact way[3],

$$R_k \leq \log\det\left(\mathbf{I} + \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\mathbf{R}_{\mathrm{ni},k}^{-1}\right), \tag{3.13}$$

where $\mathbf{R}_{\mathrm{ni},k} = \sum_{j=1,j\neq k}^{K} \mathbf{H}_k\mathbf{Q}_j\mathbf{H}_k^H + \sigma^2\mathbf{I} \in \mathbb{C}^{n_{R_k}\times n_{R_k}}$ is the noise plus multiuser interference (MUI) covariance matrix. As will be shown later, the optimization of the resource allocation problem will lead to a solution in which (3.13) holds always with equality (thus, the allocated rate equals the channel capacity).

At this point we want to emphasize, however, that in a practical deployment the rate is usually upper-bounded by a more realistic function related with the practical constraints derived from an actual implementation. In that case, we should just change expression (3.13) by the formulation describing the rate in a realistic way (a possible example corresponds to the introduction of the SNR gap in the capacity expression accounting for a transmission under

[3]All the logarithms considered in this chapter are base-2 logarithms.

a given BER or the fact of using finite size constellations, as explained in [Pal05]). Another issue to take into account is that, as explained in [MD13], [Kha01], and [Ric03], the number of iterations required for the decoding diverges as the rate approaches capacity (the complexity is of the order $\mathcal{O}(1/\text{BER}) + \mathcal{O}(1/\delta \log(1/\delta))$, where $\delta = 1 - R/C$ and C denotes capacity, as long as the decoding is executed in the circuit model specified in that paper), which would imply a divergence in the decoding power consumption. Note that if a prefixed target BER is established, the first term in the decoding complexity is just a constant term. A possibility to solve this divergence problem is to accept a gap in the rate, in the sense that such rate can only achieve a given percentage of the channel capacity, allowing for a finite number of iterations in the decoding. This percentage should, then, be incorporated in the rate formulation.

Note that the aspects mentioned in the previous paragraph depend in general on the concrete technology adopted and the implementation of the decoder, whose detailed analysis are out of the scope of this thesis that is trying to provide an abstract generic approach for the stated problem. This is the reason why in the following we will adopt expression (3.13) for the rate but keeping in mind that the resource allocation strategy developed in this thesis could be easily extended by just adopting the appropriate rate expression corresponding to each specific implementation.

The design of the precoder matrix has received a lot of attention in the past few years. The scientific community has proposed several optimization criteria for the precoder design problem under different constraints and quality of service (QoS) requirements [Pal03]. However, no previous work has taken into account the effect of the battery limitations of the mobile terminals. In this chapter, we propose to consider the sum rate maximization as design objective under a constraint on the maximum energy spent per receiver and a QoS constraint to guarantee a minimum data-rate for every user in order to control the level of fairness. Such allocation problem is solved by the resource allocator at the BS at the beginning of each frame. To simplify the notation, we will drop the time dependence in the variables. Based on the previous considerations, the resource allocation can be formulated in terms of the following optimization problem:

$$
\begin{aligned}
\underset{\{R_k\},\{\mathbf{Q}_k\}}{\text{maximize}} \quad & \sum_{k=1}^{K} R_k && && (3.14)\\
\text{subject to} \quad & C1: \sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_k) + P_c^{tx} \leq P_{\max} && \\
& C2: -\log\det\left(\mathbf{I} + \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^H \mathbf{R}_{\text{ni},k}^{-1}\right) + R_k \leq 0, && 1 \leq k \leq K \\
& C3: R_k \leq R_{\max,k}(C_k), && 1 \leq k \leq K \\
& C4: R_k \geq \text{qos}_k, && 1 \leq k \leq K \\
& C5: \mathbf{Q}_k \succeq 0, && 1 \leq k \leq K,
\end{aligned}
$$

where $P_{\max} = P_T + P_c^{t_x}$ is the total available power at the BS and qos_k is a QoS metric modeled as a minimum data rate to be allocated to the k-th user. Notice that constraint $C3$ corresponds to the maximum energy allowed for decoding per user, where that energy and the maximum data-rate are related through (3.8). We would like to emphasize that although the design of the precoder can be carried out knowing only the maximum rate constraints, such constraints change from frame to frame according to the energy spending of the terminals and the harvested energy (see Algorithm 3.1). Consequently, the energy management policy has a direct impact on the performance throughout time (the evaluation is performed not just based on one frame, but on a number of frames, as will be shown in the simulations section).

Unfortunately, it is easy to check that the problem proposed in (3.14) is not convex since $C2$ is not convex on $\{\mathbf{Q}_k\}$ due to the presence of $\{\mathbf{R}_{\text{ni},k}\}$, which makes it a very difficult problem to be solved.

A possible solution to the previous problem consists in forcing the precoder matrix $\{\mathbf{B}_k\}$ to have a particular structure at the price of making the whole problem solution sub-optimum but tractable. There are different approaches in the literature to do this. The one we propose to be used in this chapter is based on *block diagonalization* (BD) [Spe04]. By means of this approach, it is possible to cancel the MUI completely so that the original problem becomes convex. This technique implies that the following relation concerning the number of antennas has to be fulfilled: $n_T > n_R - \min_k\{n_{R_k}\}$ (see [Spe04] for more details). BD approaches the design of the precoder matrix according to the following decomposition:

$$\mathbf{B}_k = \mathbf{F}_k^a \mathbf{F}_k^b, \tag{3.15}$$

where $\mathbf{F}_k^a \in \mathbb{C}^{n_T \times (n_T - n_R + n_{R_k})}$ is used to eliminate the interference at all antennas of the non-intended receivers by creating parallel single user MIMO channels with no interference among them. Then, $\mathbf{F}_k^b \in \mathbb{C}^{(n_T - n_R + n_{R_k}) \times n_{S_k}}$ is the matrix that processes spatially the user data to optimize the transmission through each equivalent single user MIMO channel $\mathbf{H}_k \mathbf{F}_k^a$ separately. In order to cancel all the MUI, the following constraint must be imposed:

$$\mathbf{H}_k \mathbf{B}_j = \mathbf{0}, \qquad \forall\, k \neq j. \tag{3.16}$$

In order to satisfy (3.16), $\mathbf{B}_k$ should lie in the row null space of $\tilde{\mathbf{H}}_k$, where $\tilde{\mathbf{H}}_k \in \mathbb{C}^{(n_R - n_{R_k}) \times n_T}$ is defined as

$$\tilde{\mathbf{H}}_k = [\mathbf{H}_1^T \quad \dots \quad \mathbf{H}_{k-1}^T \quad \mathbf{H}_{k+1}^T \quad \dots \quad \mathbf{H}_K^T]^T. \tag{3.17}$$

The calculation of such null space is performed as follows. First, let us assume that, since $\mathbf{H}_k$ is random, $\text{rank}(\tilde{\mathbf{H}}_k) = n_R - n_{R_k}$ with probability 1. The singular value decomposition (SVD) of $\tilde{\mathbf{H}}_k$ can be written as

$$\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k \tilde{\mathbf{\Lambda}}_k [\tilde{\mathbf{V}}_k^{(1)} \quad \tilde{\mathbf{V}}_k^{(0)}]^H, \tag{3.18}$$

where the columns of the unitary matrix $\tilde{\mathbf{U}}_k \in \mathbb{C}^{(n_R-n_{R_k})\times(n_R-n_{R_k})}$ are the left singular vectors, the diagonal matrix $\tilde{\mathbf{\Lambda}}_k \in \mathbb{C}^{(n_R-n_{R_k})\times n_T}$ contains the $n_R-n_{R_k}$ singular values of $\tilde{\mathbf{H}}_k$ in decreasing order, matrix $\tilde{\mathbf{V}}_k^{(1)} \in \mathbb{C}^{n_T\times(n_R-n_{R_k})}$ is composed of the first $n_R - n_{R_k}$ right singular vectors, and matrix $\tilde{\mathbf{V}}_k^{(0)} \in \mathbb{C}^{n_T\times(n_T-n_R+n_{R_k})}$ holds the last $n_T - n_R + n_{R_k}$ right singular vectors with associated singular values equal to 0. Since $\tilde{\mathbf{V}}_k^{(0)}$ forms an orthogonal basis for the row null space of $\tilde{\mathbf{H}}_k$, by setting $\mathbf{F}_k^a = \tilde{\mathbf{V}}_k^{(0)}$, we can guarantee that (3.16) is always fulfilled.

According to this notation, the maximum achievable rate of the k-th user can be rewritten as

$$R_k \leq \log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)}\mathbf{F}_k^b\mathbf{F}_k^{bH}\tilde{\mathbf{V}}_k^{(0)H}\mathbf{H}_k^H\right). \tag{3.19}$$

If we consider $\mathbf{G}_k = \mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)} \in \mathbb{C}^{n_{R_k}\times(n_T-n_R+n_{R_k})}$ to be the new equivalent single user MIMO channel and denote $\mathbf{T}_k = \mathbf{F}_k^b\mathbf{F}_k^{bH} \in \mathbb{C}^{(n_T-n_R+n_{R_k})\times(n_T-n_R+n_{R_k})}$, the previous expression for the achievable rate can be written as

$$R_k \leq \log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{G}_k\mathbf{T}_k\mathbf{G}_k^H\right). \tag{3.20}$$

Notice that the interference has been cancelled and only the noise term remains. Now, the original optimization problem (3.14) can be transformed into the following convex optimization problem:

$$\begin{aligned}
\underset{\{R_k\},\{\mathbf{T}_k\}}{\text{maximize}} \quad & \sum_{k=1}^{K} R_k && && (3.21)\\
\text{subject to} \quad & C1: \sum_{k=1}^{K}\mathrm{Tr}(\mathbf{T}_k) \leq P_T && \\
& C2: -\log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{G}_k\mathbf{T}_k\mathbf{G}_k^H\right) + R_k \leq 0, && 1\leq k\leq K\\
& C3: R_k \leq R_{\max,k}(C_k), && 1\leq k\leq K\\
& C4: R_k \geq \mathrm{qos}_k, && 1\leq k\leq K\\
& C5: \mathbf{T}_k \succeq 0, \quad \mathbf{T}_k = \mathbf{T}_k^H, && 1\leq k\leq K.
\end{aligned}$$

We claim that constraint $C2$ is tight at the optimum assuming that data vectors $\mathbf{x}_k$ are Gaussian distributed, i.e., $R_k^\star = \log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{G}_k\mathbf{T}_k^\star\mathbf{G}_k^H\right)$, since if it was not, we could reduce the power (i.e., $\mathrm{Tr}(\mathbf{T}_k)$), which would relax constraint $C1$ until $C2$ was tight and the objective function would not be affected. That means that the objective of the problem is equivalent to maximizing $\sum_{k=1}^{K}\log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{G}_k\mathbf{T}_k\mathbf{G}_k^H\right)$. It is well-known that the previous expression is maximized when the precoder covariance matrix diagonalizes the channel matrix, i.e., $\mathbf{G}_k\mathbf{T}_k\mathbf{G}_k^H$ must be

diagonal [Tel95]. If we compute the SVD of $\mathbf{G}_k$ as

$$\mathbf{G}_k = [\mathbf{U}_k^{(1)} \quad \mathbf{U}_k^{(0)}] \begin{bmatrix} \mathbf{\Lambda}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{V}_k^{(1)} \quad \mathbf{V}_k^{(0)}]^H, \tag{3.22}$$

where $\mathbf{V}_k^{(1)} \in \mathbb{C}^{(n_T - n_R + n_{R_k}) \times n_{S_k}}$ and $\mathbf{\Lambda}_k = \text{diag}(\lambda_{1,k}, ..., \lambda_{n_{S_k},k})$ denotes the non-zero singular values of $\mathbf{G}_k$ in decreasing order (remember that, as indicated at the beginning of Section 3.2, $n_{S_k} = \min\{n_{R_k}, n_T - n_R + n_{R_k}\}$, which is equal to the number of non-zero singular values of $\mathbf{G}_k$ with probability 1 thanks to the randomness of the channel). By applying the identity $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$, the rate for the k-th user becomes $\log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{V}_k^{(1)H}\mathbf{T}_k\mathbf{V}_k^{(1)}\mathbf{\Lambda}_k^2\right)$. It is concluded from [Tel95] that $\mathbf{T}_k$ must be $\mathbf{T}_k = \mathbf{V}_k^{(1)}\mathbf{P}_k\mathbf{V}_k^{(1)H}$, where $\mathbf{P}_k = \text{diag}(p_{1,k}, ..., p_{n_{S_k},k})$ contains the powers to be allocated to the spatial modes of the equivalent channel. Given that, constraint $C2$ can be finally expressed as $-\sum_{i=1}^{n_{S_k}} \log(1 + \frac{1}{\sigma^2}p_{i,k}\lambda_{i,k}^2) + R_k \leq 0$, $\forall k$, matrix $\mathbf{F}_k^b$ can be written as $\mathbf{F}_k^b = \mathbf{V}_k^{(1)}\mathbf{P}_k^{1/2}$, and the precoder matrix $\mathbf{B}_k$ as

$$\mathbf{B}_k = \tilde{\mathbf{V}}_k^{(0)}\mathbf{V}_k^{(1)}\mathbf{P}_k^{1/2} \in \mathbb{C}^{n_T \times n_{S_k}}. \tag{3.23}$$

Finally, the optimization problem can be formulated as

$$\begin{aligned}
\underset{\{R_k\},\{p_{i,k}\}}{\text{maximize}} \quad & \sum_{k=1}^{K} R_k \\
\text{subject to} \quad & C1: \sum_{k=1}^{K}\sum_{i=1}^{n_{S_k}} p_{i,k} \leq P_T \\
& C2: -\sum_{i=1}^{n_{S_k}} \log\left(1 + \frac{1}{\sigma^2}p_{i,k}\lambda_{i,k}^2\right) + R_k \leq 0, \quad 1 \leq k \leq K \\
& C3: R_k \leq R_{\max,k}(C_k), \qquad\qquad 1 \leq k \leq K \\
& C4: R_k \geq \text{qos}_k, \qquad\qquad 1 \leq k \leq K \\
& C5: p_{i,k} \geq 0, \qquad 1 \leq i \leq n_{S_k}, \qquad 1 \leq k \leq K.
\end{aligned} \tag{3.24}$$

The previous problem is a convex optimization problem where now the optimization variables are $\{R_k\}$, $\{p_{i,k}\}$. In order to solve (3.24), effective numerical methods, such as interior point methods, could be applied [Boy04]. However, by means of Lagrange duality and KKT conditions [Boy04], we are able to obtain, in this case, a much more efficient iterative algorithm and to get some insights into the problem that are lost if a generic numerical algorithm is used. Slater's constraint qualification holds for this problem and, thus, KKT conditions are necessary and sufficient for optimality. Given that, let us present the following result:

Proposition 3.1. *According to the KKT conditions, the optimum power allocation of problem*

(3.24) *is given by*

$$p_{i,k}^{\star} = \left(\frac{1 + \nu_k^{\star} - \beta_k^{\star}}{\mu^{\star} \ln(2)} - \frac{\sigma^2}{\lambda_{i,k}^2} \right)^+, \quad 1 \leq i \leq n_{S_k}, \quad 1 \leq k \leq K, \tag{3.25}$$

where $(x)^+ = \max(0, x)$*, and* μ*,* $\{\beta_k\}$*, and* $\{\nu_k\}$ *are the dual variables or Lagrange multipliers corresponding to constraints* $C1$*,* $C3$*, and* $C4$*, respectively.*

Proof. See Appendix 3.A. ■

Finally, the optimum data rate is given by

$$R_k^{\star} = \sum_{i=1}^{n_{S_k}} \log \left(1 + \frac{1}{\sigma^2} p_{i,k}^{\star} \lambda_{i,k}^2 \right). \tag{3.26}$$

Proposition 3.2. *The optimization problem* (3.24) *is optimally solved by the Algorithm 3.2 in a finite number of iterations. The optimum power allocation attained is the multi-level water-filling solution formulated in* (3.25)*.*

Proof. It can be checked that (3.25) satisfies all the KKT conditions and is, therefore, optimal. Algorithm 3.2 is based on hypothesis testing. It first makes the assumption that all users are not energy saturated or constraints $C3$ are inactive, i.e., $R_k^{\star} < R_{\max,k}$, $\forall k$ or $\beta_k^{\star} = 0$, $\forall k$. It assigns power using the classical water-filling policy, and then checks whether the first user to be saturated by energy is indeed saturated, in which case the hypothesis is rejected. The maximum power to fulfill $C3$ is assigned to such user, a new hypothesis is made with one less user, and so forth. Notice that the algorithm computes the corresponding values of the Lagrange multipliers $\{\nu_k\}$, $\{\beta_k\}$, and μ for all users implicitly. The optimum power allocation found in step 21 coincides with the solution in (3.25). In the algorithm, variable c plays the role of $\frac{1}{\mu \ln(2)}$. An upper bound of the required number of iterations is $n_S \times (K + 2)$ since steps 2 and 9 can be computed in no more than n_S iterations each and steps 12-20 require less than $K \times n_S$ iterations. ■

Figure 3.3 depicts an illustrative example of the water levels and power allocations resulting from the proposed algorithm. The first stage computes the thresholds $c_{\min,k}$ and $c_{\max,k}$, $\forall k$, associated with the minimum QoS and maximum rate constraints ($C4$ and $C3$, respectively) and assigns the power $P_k^{\min}$ needed to fulfill the QoS constraints. The second stage assigns the remaining power maximizing the sum rate.

Algorithm 3.2 Multi-level water-filling algorithm with battery and QoS constraints

order substreams increasingly for each user k according to $\frac{\sigma^2}{\lambda_{i,k}^2}$

solve $\text{qos}_k = \sum_{i=1}^{n_{S_k}} \log\left(1 + \left(c_{\min,k} - \frac{\sigma^2}{\lambda_{i,k}^2}\right)^+ \frac{\lambda_{i,k}^2}{\sigma^2}\right) \;\forall\, k \longrightarrow \{c_{\min,k}\}$

calculate $P_k^{\min} = \sum_{i=1}^{n_{S_k}} \left(c_{\min,k} - \frac{\sigma^2}{\lambda_{i,k}^2}\right)^+$

if $\sum_{k=1}^{K} P_k^{\min} > P_T$

    problem is infeasible: relax $\{\text{qos}_k\}$ or drop users from the system $\longrightarrow$ go to 2

else if $\sum_{k=1}^{K} P_k^{\min} = P_T$

    $R_k^\star = \text{qos}_k \quad \forall\, k, \quad p_{i,k}^\star = \left(c_{\min,k} - \frac{\sigma^2}{\lambda_{i,k}^2}\right)^+ \quad \forall\, i,k \longrightarrow$ go to 23

else

    solve $R_{\max,k} = \sum_{i=1}^{n_{S_k}} \log\left(1 + \left(c_{\max,k} - \frac{\sigma^2}{\lambda_{i,k}^2}\right)^+ \frac{\lambda_{i,k}^2}{\sigma^2}\right) \;\forall\, k \longrightarrow \{c_{\max,k}\}$

    order $c_{\max,k}$ increasingly and re-number such that $c_{\max,1} \le c_{\max,2} \le ... \le c_{\max,K}$

    set $\mathcal{A} = \{1, 2, \ldots, K\}$; $q = 1$

    solve $\sum_{k\in\mathcal{A}} \sum_i \left(c - \max\left(c_{\min,k}, \frac{\sigma^2}{\lambda_{i,k}^2}\right)\right)^+ = P_T - \sum_{k\in\mathcal{A}} P_k^{\min} \longrightarrow c$

    if $c > c_{\max,q}$

        $\tilde{P}_q = \sum_i \left(c_{\max,q} - \frac{\sigma^2}{\lambda_{i,q}^2}\right)^+$

        $P_T \longleftarrow P_T - \tilde{P}_q$

        $\mathcal{A} \leftarrow \mathcal{A} - \{q\}$

        $q \leftarrow q + 1$

        go to 12

    end if

end if

assign $p_{i,k}^\star = \left(\min(\max(c, c_{\min,k}), c_{\max,k}) - \frac{\sigma^2}{\lambda_{i,k}^2}\right)^+ \quad \forall\, i,k$

assign $R_k^\star = \sum_{i=1}^{n_{S_k}} \log\left(1 + \frac{p_{i,k}^\star \lambda_{i,k}^2}{\sigma^2}\right) \;\forall\, k$

end algorithm

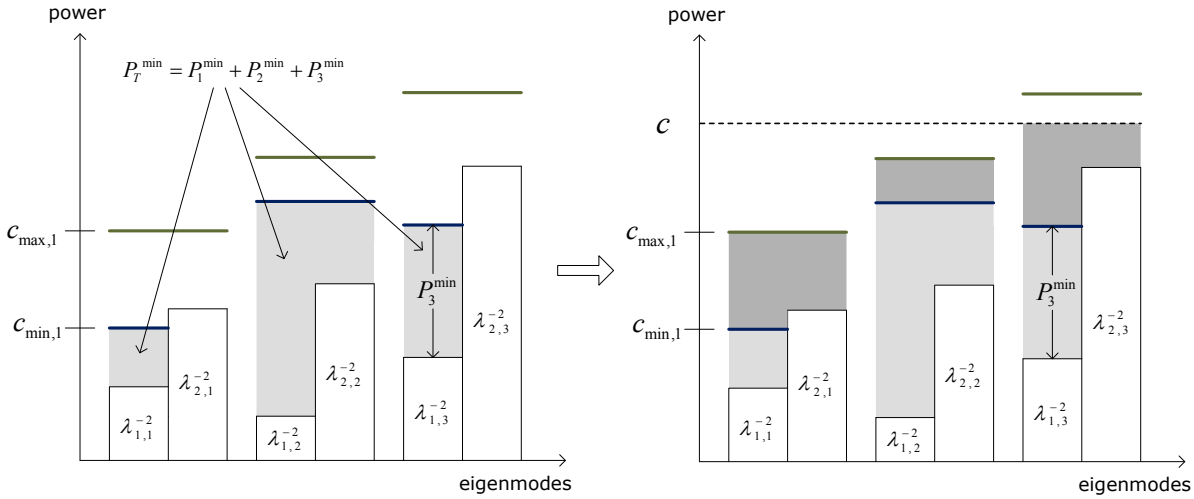


Figure 3.3: Example of the optimum multiuser water-filling in a scenario with 3 users and 2 streams per user. Users 1 and 2 are energy saturated (i.e., $R_k^\star = R_{\max,k}$) whereas user 3 achieves the common water level. It is assumed that $\sigma^2 = 1$.

3.4.1.1 Feasibility

The constraints that must be checked in order to study the feasibility of the problem are $C3$ and $C4$. If $R_{\max,k} < \text{qos}_k$ for some k, then constraints $C3$ and $C4$ cannot be fulfilled simultaneously for such user k. In that situation, we can either relax the QoS constraints or allow some more energy to be used for decoding by increasing $R_{\max,k}$. Then, if $R_{\max,k} > \text{qos}_k$, $\forall k$, the problem will be feasible if there is enough power to guarantee all the QoS constraints $C4$, i.e., $\sum_{k=1}^{K} P_k^{\min} \leq P_T$, where $P_k^{\min} = \sum_{i=1}^{n_{S_k}} \left(c_{\min,k} - \frac{\sigma^2}{\lambda_{i,k}^2} \right)^+$ and $c_{\min,k}$ is obtained as detailed in step 2 of the Algorithm 3.2. If the problem is not feasible, then some users should be dropped from the system or the QoS requirements should be relaxed.

3.4.2 Robust Precoder Design with Imperfect CSI

As we commented before, the assumption of having perfect CSI at the receiver and transmitter sides is rather unrealistic. At the receiver side, the channel is usually estimated through training sequences. On the other hand, the transmitter can acquire the CSI through a feedback channel in frequency division duplexing (FDD) systems or by exploiting channel reciprocity in time division duplexing (TDD) systems[4]. We consider the former case, in which the receivers estimate the channel (obtaining a noisy estimate) and inform the transmitter with an imperfect CSI through a feedback channel. All these facts motivate the introduction of imperfect CSI in the resource allocation strategy. For the sake of simplicity, we consider that this feedback channel is ideal, i.e., neither errors in the feedback communication nor quantization effects are considered. As mentioned in the introduction, we assume slow-varying fading channels so that the current

[4]Even in TDD systems, a feedback channel may be needed since, in general, radio front-ends are not reciprocal. In that case, a proper calibration of the devices is required if feedback is to be avoided.

estimate remains up-to-date for the whole frame.

The receiver estimates the channel by means of known training sequences. The received signal model for user k during the training period is represented as (we assume that the training is carried out using orthogonal training sequences, so that there is no interference among users when estimating the channel):

$$\mathbf{Y}'_k = \mathbf{H}_k\mathbf{M}_k + \mathbf{N}_k, \tag{3.27}$$

where $\mathbf{Y}'_k \in \mathbb{C}^{n_{R_k}\times L_k}$, being L_k the length of the training sequence, $\mathbf{M}_k \in \mathbb{C}^{n_T\times L_k}$ is the matrix containing the training symbols, and $\mathbf{N}_k \in \mathbb{C}^{n_{R_k}\times L_k}$ is the Gaussian noise matrix with each element having a power σ^2. From now on, we will model the entries of matrices $\{\mathbf{H}_k\}$ as i.i.d. random complex Gaussian variables distributed as $\mathcal{CN}(0,\sigma^2_{h_k})$, although the results could be easily extended to other channel statistics.

We denote the channel estimated by user k as $\hat{\mathbf{H}}_k$ and the estimation error matrix as $\mathbf{H}^e_k$ such that:

$$\mathbf{H}_k = \hat{\mathbf{H}}_k + \mathbf{H}^e_k. \tag{3.28}$$

We assume the channel estimate to be obtained under the minimum mean square error (MMSE) criterion, i.e., $\hat{\mathbf{H}}_k = \mathbb{E}_{\mathbf{H}_k|\mathbf{Y}'_k}[\mathbf{H}_k|\mathbf{Y}'_k]$ [Kay93]. If we stack the columns of $\mathbf{Y}'_k$ to produce $\mathbf{y}'_k = \text{vec}(\mathbf{Y}'_k)$, we obtain the channel estimate as $\hat{\mathbf{h}}_k = \boldsymbol{\Upsilon}_k\mathbf{y}'_k$, where $\boldsymbol{\Upsilon}_k = \mathbf{S}^H_k(\mathbf{S}_k\mathbf{S}^H_k + \frac{1}{\sigma^2}\mathbf{I})$, and $\mathbf{S}_k = \mathbf{M}^T_k \otimes \mathbf{I}$. The final estimate $\hat{\mathbf{H}}_k$ is obtained by rearranging $\hat{\mathbf{h}}_k$ into a $n_{R_k} \times n_T$ matrix. By well known properties of the MMSE estimator, we know that the estimated channel matrix and the estimated error matrix are uncorrelated, i.e., $\mathbb{E}[\text{vec}(\hat{\mathbf{H}}_k)\text{vec}(\mathbf{H}^e_k)^H] = \mathbf{0} \ \ \forall k$. The entries of $\hat{\mathbf{H}}_k$ and $\mathbf{H}^e_k$ can be proved to be i.i.d. complex Gaussian with zero mean and with variances $\sigma^2_{h_k} - \sigma^2_{E_k}$ and $\sigma^2_{E_k}$, respectively. The value of $\sigma^2_{E_k}$ represents the quality of the channel estimation and we assume that it is known by the transmitter (if the reader is interested in how this parameter can be estimated, see [Yoo06], and references therein).

Now, if we include (3.28) into the original signal model introduced in (3.1), we obtain

$$\mathbf{y}_k = \hat{\mathbf{H}}_k\mathbf{B}_k\mathbf{x}_k + \mathbf{H}^e_k\mathbf{B}_k\mathbf{x}_k + \hat{\mathbf{H}}_k\sum_{j=1,j\neq k}^{K}\mathbf{B}_j\mathbf{x}_j + \mathbf{H}^e_k\sum_{j=1,j\neq k}^{K}\mathbf{B}_j\mathbf{x}_j + \mathbf{n}_k.$$

Since the transmitter has only information about the estimated matrices $\{\hat{\mathbf{H}}_k\}$, the precoders can be designed only as a function of them. To that end, if we use the BD technique that was presented in last section, the transmit precoder must be designed so that

$$\hat{\mathbf{H}}_k\mathbf{B}_j = \mathbf{0}, \quad \forall\, k \neq j, \tag{3.29}$$

and, therefore, (3.29) reduces to

$$\mathbf{y}_k = \hat{\mathbf{H}}_k\mathbf{B}_k\mathbf{x}_k + \mathbf{H}_k^e\sum_{j=1}^{K}\mathbf{B}_j\mathbf{x}_j + \mathbf{n}_k = \hat{\mathbf{H}}_k\mathbf{B}_k\mathbf{x}_k + \mathbf{u}_k, \tag{3.30}$$

where $\mathbf{u}_k = \mathbf{H}_k^e\sum_{j=1}^{K}\mathbf{B}_j\mathbf{x}_j + \mathbf{n}_k$ is known as the effective noise [Has03] and its covariance matrix is $\mathbf{R}_k = \mathbb{E}_{\mathbf{n}_k,\mathbf{H}_k^e|\hat{\mathbf{H}}_k}[\mathbf{u}_k\mathbf{u}_k^H] = \mathbb{E}_{\mathbf{H}_k^e|\hat{\mathbf{H}}_k}\left[\mathbf{H}_k^e\sum_{j=1}^{K}\mathbf{Q}_j\mathbf{H}_k^{eH}\right] + \sigma^2\mathbf{I}$ with $\mathbf{Q}_j = \mathbf{B}_j\mathbf{B}_j^H$.

Note that although the receiver does not know the estimated error matrix $\mathbf{H}_k^e$, it knows that it is Gaussian distributed, as the data vectors $\mathbf{x}_k$. According to this, it turns out that, the effective noise vector will not be Gaussian distributed since the product of two Gaussian random variables is not Gaussian.

The mutual information for the signal model (3.30) can be expressed as [Med00]

$$\mathcal{I}(\mathbf{x}_k;\mathbf{y}_k|\hat{\mathbf{H}}_k) = \mathcal{H}(\mathbf{x}_k|\hat{\mathbf{H}}_k) - \mathcal{H}(\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k) = \mathcal{H}(\mathbf{x}_k) - \mathcal{H}(\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k). \tag{3.31}$$

In the previous expression, $\hat{\mathbf{H}}_k$ corresponds to the given estimate of the channel. If we assume Gaussian codebooks, then the first entropy is simply $\mathcal{H}(\mathbf{x}_k) = \log\det(\pi\mathrm{e}\mathbf{Q}_k)$. However, the term $\mathcal{H}(\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k)$ cannot be easily computed since $\mathbf{y}_k$ is not Gaussian distributed, as commented before. In this case, we take the procedure followed in references [Has03], [Med00] and upper bound such conditional entropy by considering $\mathbf{y}_k$ to be Gaussian. Finally, according to this procedure, we find the following result:

Proposition 3.3. *The mutual information* (3.31) *can be lower bounded as*

$$\mathcal{I}(\mathbf{x}_k;\mathbf{y}_k|\hat{\mathbf{H}}_k) \geq \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_k\mathbf{Q}_k\hat{\mathbf{H}}_k^H\mathbf{R}_k^{-1}\right) \triangleq \mathcal{R}_k. \tag{3.32}$$

Proof. See Appendix 3.B. ■

We can still further simplify the previous result by observing that the covariance matrix $\mathbf{R}_k$ is indeed diagonal if the elements of $\mathbf{H}_k^e$ are i.i.d., as stated previously. If we take the (p,q)-th entry of $\mathbb{E}_{\mathbf{H}_k^e|\hat{\mathbf{H}}_k}\left[\mathbf{H}_k^e\sum_{j=1}^{K}\mathbf{Q}_j\mathbf{H}_k^{eH}\right]$ and denote $\mathbf{h}_k^{e(p)T}$ as the p-th row of $\mathbf{H}_k^e$, we have that $\mathbb{E}_{\mathbf{H}_k^e|\hat{\mathbf{H}}_k}\left[\mathbf{H}_k^e\mathbf{Q}_j\mathbf{H}_k^{eH}\right]_{pq} = \mathbb{E}_{\mathbf{H}_k^e|\hat{\mathbf{H}}_k}\left[\mathbf{h}_k^{e(p)T}\mathbf{Q}_j\mathbf{h}_k^{e(q)*}\right] = \mathbb{E}_{\mathbf{H}_k^e|\hat{\mathbf{H}}_k}\left[\mathrm{Tr}\left(\mathbf{Q}_j\mathbf{h}_k^{e(q)*}\mathbf{h}_k^{e(p)T}\right)\right]$, and invoking the i.i.d. assumption, finally we have that $\mathbb{E}_{\mathbf{H}_k^e|\hat{\mathbf{H}}_k}\left[\mathbf{H}_k^e\mathbf{Q}_j\mathbf{H}_k^{eH}\right]_{pq} = \sigma_{E_k}^2\,\mathrm{Tr}(\mathbf{Q}_j)\,\delta_{pq}$.

According to the previous results, the optimization problem considered in this section to allocate the resources is the same as that proposed in (3.14), but now replacing the rate expression in constraint $C2$ by the lower bound on the data rate that we have obtained in (3.32). Before presenting the final achievable rate, let us provide the following proposition:

Proposition 3.4. *If there is at least one user which is not saturated (i.e., at least one k such that $R_k^\star < R_{\max,k}$), then the optimum solution fulfills $\sum_{j=1}^{K}\mathrm{Tr}(\mathbf{Q}_j^\star) = P_T$.*

Proof. See Appendix 3.C. ■

According to Proposition 3.4, if at least one user is not saturated in the system, then we know that the achievable rate for user k is $\mathcal{R}_k = \log\det\left(\mathbf{I} + \frac{\hat{\mathbf{H}}_k\mathbf{Q}_k\hat{\mathbf{H}}_k^H}{\sigma_{E_k}^2 P_T+\sigma^2}\right)$. By designing the set of matrices $\{\mathbf{Q}_k\}$ under the BD procedure and taking the previous assumption, the achievable rate is eventually expressed as

$$\mathcal{R}_k = \sum_{i=1}^{n_{S_k}} \log\left(1 + \frac{p_{i,k}\hat{\lambda}_{i,k}^2}{\sigma_{E_k}^2 P_T + \sigma^2}\right), \tag{3.33}$$

where $\hat{\lambda}_{i,k}$ are the eigenvalues computed as in (3.22), simply by substituting the exact channels by the estimated ones.

It turns out that the error due to estimation adds as an equivalent noise with power equal to the variance of the estimation error matrix times the total available power for transmission. The worse the estimation is, the more equivalent noise we have and, as a result, the less power the algorithm will assign to that particular user. As far as the term P_T is concerned, this term appears since now there is interference coming from signals transmitted to all the receivers due to the non-matched BD design at the transmitter side (this happens because only an imperfect channel estimate is available when computing the BD matrices). As a consequence, a given user may have a good channel, but if the receiver performs a noisy estimation, the transmitter will not be able to erase the contribution of this error, which can affect extremely the data rate.

To obtain the optimum power allocations $p_{i,k}^\star$, we only need to change constraint $C2$ in problem (3.24) by (3.33). Since the structure of the optimization problem remains the same, the optimum solution is given by

$$p_{i,k}^\star = \left(\frac{1+\nu_k^\star-\beta_k^\star}{\mu^\star\ln(2)} - \frac{\sigma_{E_k}^2 P_T+\sigma^2}{\hat{\lambda}_{i,k}^2}\right)^+, \quad 1\le i\le n_{S_k}, \quad 1\le k\le K, \tag{3.34}$$

and the Algorithm 3.2 produces the optimum solution by simply substituting $\frac{\sigma^2}{\lambda_{i,k}^2}$ by $\frac{\sigma_{E_k}^2 P_T+\sigma^2}{\hat{\lambda}_{i,k}^2}$.

Notice that, however, if all users in the system are saturated (even though this is a rather unusual situation), the expression $\sum_{j=1}^K \mathrm{Tr}(\mathbf{Q}_j^\star) = P_T$ may no longer be true. In that situation, the achievable rate is expressed as $\mathcal{R}_k = \log\det\left(\mathbf{I} + \frac{\hat{\mathbf{H}}_k\mathbf{Q}_k\hat{\mathbf{H}}_k^H}{\sigma_{E_k}^2\sum_{j=1}^K \mathrm{Tr}(\mathbf{Q}_j)+\sigma^2}\right)$. The optimal rates are $R_k^\star = R_{\max,k}$ but the optimal covariance matrices remain unknown since we do not know the optimal power allocations. In this particular case, Algorithm 3.2 would not produce the optimal solution. At this point we propose an iterative algorithm based on nested intervals through which we can find a lower bound (with tolerance ϵ) on the minimum value of P_T for which all users are saturated. In this situation, we know that Algorithm 3.2 finds the optimum allocation. The proposed nested intervals algorithm is presented in Algorithm 3.3. It is based on Algorithm

Algorithm 3.3 Precoder design with imperfect CSI: all users are energy saturated

solve Algorithm 3.2 substituting $\frac{\sigma^2}{\lambda_{i,k}^2}$ by $\frac{\sigma_{E_k}^2 P_T+\sigma^2}{\hat{\lambda}_{i,k}^2} \longrightarrow \tilde{R}_k, \tilde{\mathbf{Q}}_k, c_{\max,k}$
if $\tilde{R}_k = R_{\max,k}\ \forall k$
    $R = P_T, \quad L = 0, \quad P_T^{(1)} = P_T/2, \quad n = 1$
    repeat
        solve Algorithm 3.2 substituting $\frac{\sigma^2}{\lambda_{i,k}^2}$ by $\frac{\sigma_{E_k}^2 P_T^{(n)}+\sigma^2}{\hat{\lambda}_{i,k}^2} \longrightarrow \tilde{R}_k, \tilde{\mathbf{Q}}_k, c$
        if $\tilde{R}_k = R_{\max,k}, \forall k$
            $R = P_T^{(n)}$
        else
            $L = P_T^{(n)}$
        end if
        $P_T^{(n+1)} = \frac{R+L}{2}$
        $n \longleftarrow n+1$
    until $|R-L| < \epsilon$ and $c < c_{\max,K}$
end if
assign $R_k^\star = \tilde{R}_k, \quad \mathbf{Q}_k^\star = \tilde{\mathbf{Q}}_k, \forall k$
end algorithm

3.2 and produces an ϵ-suboptimal solution, where ϵ can be made as small as possible and trades-off the speed of convergence and the accuracy of the solution (note that, since the algorithm is based on nested intervals, so that in each interval the length of the interval is divided by 2, convergence can always be assured).

3.4.3 Robust Precoder Design with Quantized Battery Knowledge

In the previous sections, we assumed that the feedback channel was ideal and rate unlimited. However, in practical communication scenarios that is not true; the feedback channels are rate limited. This implies that the information to be sent through it should be quantized following a given quantization rule, and then sent back from the receivers to the transmitter. In this section, we assume that the feedback channel is rate limited but we assume that most of the available throughput will be assigned to the feedback of the channel matrix so that the errors due to quantization are negligible. Nonetheless, the errors due to battery quantization are notable and a robust design considering the effect of these errors is studied in this section. For simplicity in the modeling, we assume that no communication errors are produced in the transmission through the feedback link.

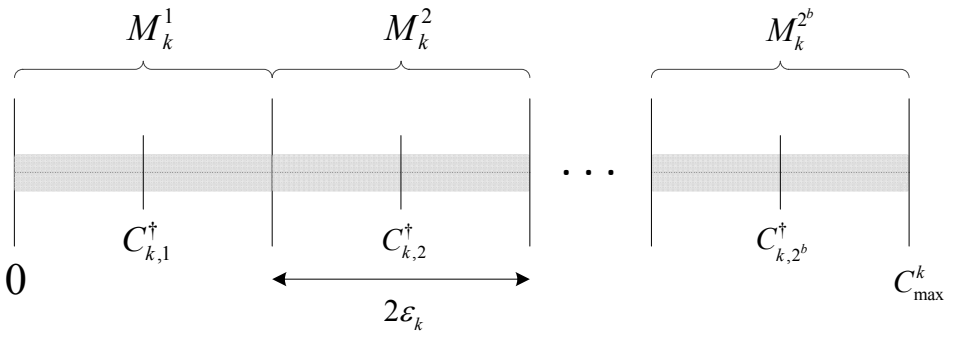


Figure 3.4: Uniform quantization regions of the battery level of the k-th user. As it can be observed, there are 2^b regions corresponding to the use of b quantization bits.

We can model the actual battery level of the k-th user as

$$C_k = C_{k,u}^\dagger + \delta_k, \quad 1 \leq k \leq K, \tag{3.35}$$

where $C_{k,u}^\dagger$ is the quantized version of C_k, being u the quantization index (which is, in fact, what is sent through the feedback channel), and δ_k models the quantization error. We assume that for a given quantized feedback $C_{k,u}^\dagger$, the actual level C_k is inside an uncertainty region denoted as $\mathcal{M}_k^u \in \mathbb{R}$ (i.e., $C_k \in \mathcal{M}_k^u$) that is defined as $\mathcal{M}_k^u = \{C_{k,u}^\dagger + \delta_k \,:\, |\delta_k| \leq \varepsilon_k\}$ where $2\varepsilon_k = q_k$, being $q_k = \frac{C_{\max}^k}{2^b}$ the quantization step in the case of a uniform quantization (as depicted in Figure 3.4), and b the number of bits used for quantization. In case that each user quantizes with different number of bits the BS must also be aware of that as well as of the corresponding value of $C_{\max}^k$. When we introduce this model into the energy equation (3.11), we obtain $E_d^k = \alpha_k C_{k,u}^\dagger + \alpha_k \delta_k$. By using the relationship between data rate and energy (3.8), we end up with $R_{\max,k}(\delta_k)$. Now, constraint $C3$ should be generalized as $C3 : R_k \leq R_{\max,k}(\delta_k),\ \forall\, k, |\delta_k| \leq \varepsilon_k$ to account for the imperfection in the knowledge of the battery level. That means that the optimization problem must be solved guaranteeing that the reformulated constraint $C3$ is fulfilled for all possible values of the quantization error δ_k. That yields directly to a robust design based on the worst-case formulation. Thus, constraint $C3$ can be equivalently rewritten as

$$C3 : R_k \leq \inf_{\delta_k \leq |\epsilon_k|} R_{\max,k}(\delta_k),\ \forall\, k. \tag{3.36}$$

This minimization is straightforward since $R_{\max,k}(\delta_k)$ is a monotonic increasing function of δ_k and, thus, the minimum is attained for the minimum value of δ_k,

$$\inf_{\delta_k \leq |\epsilon_k|} R_{\max,k}(\delta_k) = R_{\max,k}(-\epsilon_k). \tag{3.37}$$

Notice that if $R_{\max,k}(-\varepsilon_k) < \text{qos}_k$, for some user k, the problem will be infeasible. In that case, either the set of $\{\text{qos}_k\}$ constraints should be relaxed or more energy should be allowed to be used for decoding. If, on the other hand, $R_{\max,k}(-\varepsilon_k) = \text{qos}_k$, then we know that $R_k^\star = \text{qos}_k$ if the problem is feasible.

3.4.4 Numerical Simulations

This section evaluates numerically the performance of the proposed multiuser MIMO precoder design detailed in the previous sections. For the simulations, we consider a scenario with a BS with $n_T = 6$ antennas and $K = 3$ receivers with $n_{R_k} = 2$ antennas each. Front-end power consumption at the transmitter and receiver is $P_c^{tx} = 1$ W and $P_c^{rx} = 0.2$ W, respectively. The maximum available power at the BS is $P_{\max} = 5$ W. We assume a normalized noise power given by $\sigma^2 = 1$ W and a frame duration equal to $T_f = 1$ ms. The channel matrices are normalized and generated randomly with i.i.d. entries distributed according to $\mathcal{CN}(0, 1)$. The QoS is set to $\text{qos}_k = 1$ bit/s/Hz $\forall k$. The maximum battery sizes $C_{\max}^k$ are $3,000$ J, $6,000$ J, and $12,000$ J, respectively. The initial battery levels before communication starts are half the maximum battery capacities. The model of decoder consumption is exponential (3.3), unless stated otherwise, where constants $c_1 = 1,000$ W and $c_2 = 8/11$ 1/(bits/s/Hz) (for simplicity, we assume that all users are provided with the same kind of decoder). We consider the energy harvesting packet size to be $e_k = 100$ J, $\forall k$. The probability of energy packet arrival is $p_k = 0.5$, $\forall k$, unless stated otherwise. All figures are averaged over 1,000 channel and 1,000 harvesting realizations.

The values proposed for the different parameters have been chosen just as normalized illustrative values to show an evaluation of the proposed theoretical technique. A time window of 50 frames has been taken in the simulations as illustrative to be able to show the inherent tradeoffs motivating the strategy developed in this chapter[5].

Figure 3.5 shows the average data rate evolution of the system. The resultant rate is averaged over the three users in the system. $\alpha = 1$ corresponds to the traditional max-sum rate without considering energy constraints (i.e., all the energy available at the battery is allowed to be spent in the decoding process). As it can be seen, at the beginning, the traditional approach outperforms the proposed ones since the users are allowed to spend all the battery for decoding, but as time evolves, users start running out of battery and the average rate drops. It is clear that the fewer the bits used for quantization, the lower the data rate obtained. This is because the robust design is battery conservative and, thus, the rate is limited. The saturation experienced in the curves for 2 and 3 bits of battery quantization for the case of $\alpha = 0.1$ is due to the fact that users achieve the maximum data rate constraint since the robust design considers that there is lower battery than there actually is and, therefore, $R_{\max,k}$ is small. This cannot be noticed for the case of $\alpha = 1$ because the whole battery is allowed to be used and, then, the corresponding $R_{\max,k}$ values are larger, making the users not saturate.

[5]Note that, in any system, the evaluation has to be performed for a given fixed length of the time window. The selection of 50 frames has been proposed in this chapter just as an illustrative example. As will be shown in the following figures, this time window is long enough so that the mean rates and the battery levels have almost converged whereas at the same time, the transient period has still an impact on the performance, which is an effect that we also want to take into account in the evaluation.

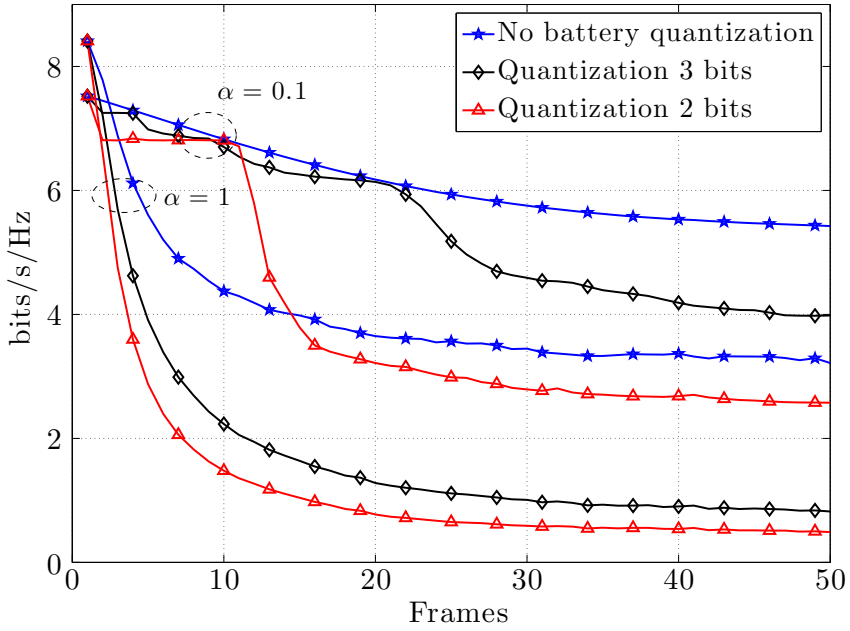


Figure 3.5: Comparison of the average data rate evolution for the given scenario with and without battery quantization for two different values of α.

Figure 3.6 depicts the average battery evolution (averaged over the three users). As it can be observed, the proposed algorithm with $\alpha = 0.1$ shows more than 15% of energy savings at the end of the transmission (50 frames) compared to the traditional approach, where no energy constraints are considered. We would like to emphasize that taking $\alpha = 0.1$ not only provides a solution that saves more energy, but also more rate can be achieved at the same time (see Figure 3.5). Obviously, by using smaller values of α, it is possible to increase the final battery level, but the obtained rate may not be enough (in Figure 3.8, we will present the optimum sum rate as a function of the value of α). Comparing the non-quantized with the quantized scheme we observe that, since the latter is more rate conservative due to the worst-case robust approach, the energy consumption is reduced.

Figure 3.7 shows the average date rate evolution considering imperfect CSI for different values of channel estimation error power $\sigma^2_{E_k}$, where we have assumed that it is the same for all users. The value of α considered is 0.1. As it can be seen and expected, the robust design that accounts for the errors in the channel knowledge produces a much more conservative solution in terms of rate than the perfect CSI scenario. This is translated into a decrease of the system sum rate which is also dependent on the error power. The more error the estimation produces, the lower the sum rate is. Note, however, that at some point the curve corresponding to the perfect CSI case and the curves corresponding to very small values of the estimation error power cross. This is because having imperfect CSI produces low rates at the beginning but users also spend less energy. As time evolves, the imperfect CSI scenario is able to produce a higher data rate eventually. However, the overall sum rate along the 50 frames is still higher for the perfect CSI scenario.

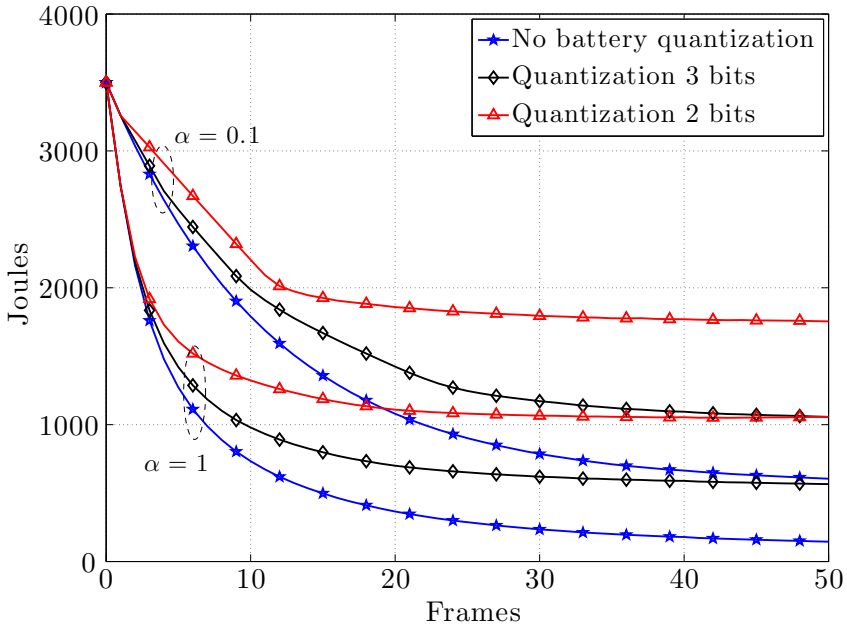


Figure 3.6: Evolution of the average battery levels with and without battery quantization for two different values of α.

Figure 3.8 depicts the achieved average sum rate and the aggregated residual battery level (in percentage) after 50 frames as a function of α for different numbers of users (1, 2, 3, and 4 users). The aggregated residual battery is computed as the ratio of the sum of the final battery levels over the sum of the initial ones. The plotted sum rate corresponds to the average over the 50 frames considered in the simulations. The parameters related to the battery sizes and numbers of antennas are provided in the caption of the figure. We also plot the sum rate obtained for the 4-user case of the offline approach. In that approach, we assume that the whole channel and harvesting realization is known in advance. Thus, that is the performance upper-bound, i.e., the maximum sum rate that could be achieved if future information was available. As it can be observed, if there are more users in the system, the achieved sum rate is higher since the user diversity is indeed larger. Another conclusion from the simulation is that the optimum value of α depends on the number of users (by optimum α we mean the one providing the highest sum rate).

As mentioned before, we present in the same figure the residual aggregated battery level. As expected, this function decays as the value of α increases, since more energy is being allowed to be used for decoding. As it can be seen, the scenario with 4 users corresponds to the case with the highest residual battery level. This phenomenon can be explained by noting that if more users are to be served by the transmitter, the resource allocation strategy will assign low rates to some users, which will allow them to recharge their batteries. In particular, in the scenario with 4 users, the residual battery level at the optimum value of α is 20.46 % and the sum rate is 25.7 bits/s/Hz. If, on the other hand, we would like to end up the whole transmission (at the end of the 50 frames) with the same aggregated battery as at the beginning of the transmission (maybe with individual battery values different from the initial ones but with the same total energy sum

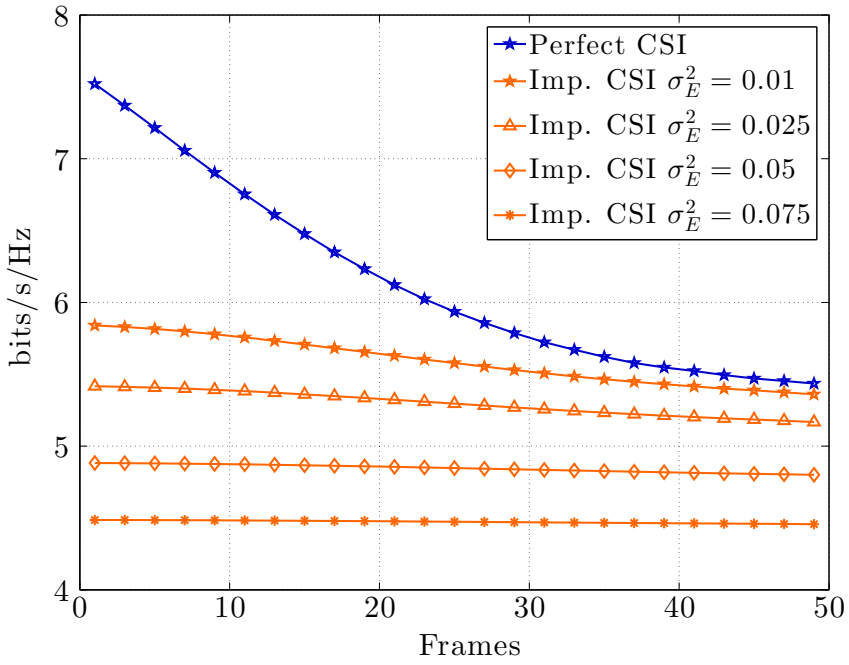


Figure 3.7: Comparison of the average rate with perfect and imperfect CSI with $\alpha = 0.1$.

among users), the optimum value of α obtained from the simulations would be $\alpha = 0.015$ and the sum rate obtained for that particular case would be 21.2 bits/s/Hz.

As commented before, the optimum value of α depends on the particular scenario. In fact, it depends on the mean intensity of the harvesting, the efficiency in the harvesting process, and the number of users in the system, among other parameters. The different kinds of energy sources (e.g. solar harvesting) and abilities to harvest can be modeled in our simulation framework by adjusting properly the probability of energy packet arrival p. Figure 3.9 depicts the optimum value of α as a function of the energy packet arrival probability p considering perfect and imperfect CSI at the transmitter and battery quantization. If we consider a scenario where p changes slowly compared to the time window of the simulations (i.e., 50 frames), then the optimum value of α should be set as according to Figure 3.9. A typical example is the case of solar energy, where in that case p would change slowly throughout the day (taking high values at midday and very low values at night). In that case, α_k could be adjusted according to the current value of p at each time. We also plot the corresponding sum rate achieved for that particular value of p and the sum rate obtained for the offline strategy. As it can be observed from the figure, the optimum value of α increases for higher energy packet arrival probability. It is interesting to note that, even if the harvesting source is always scavenging energy from the environment, the optimum value of α is far from the traditional approach where α is set to 1. Of course, this depends on the size of the energy packet (e), the decoder efficiency (constants c_1 and c_2), and the time window length.

In the last figure, Figure 3.10, we represent the data rate averaged among the users in the system throughout time for different decoder power consumption models and different decoder

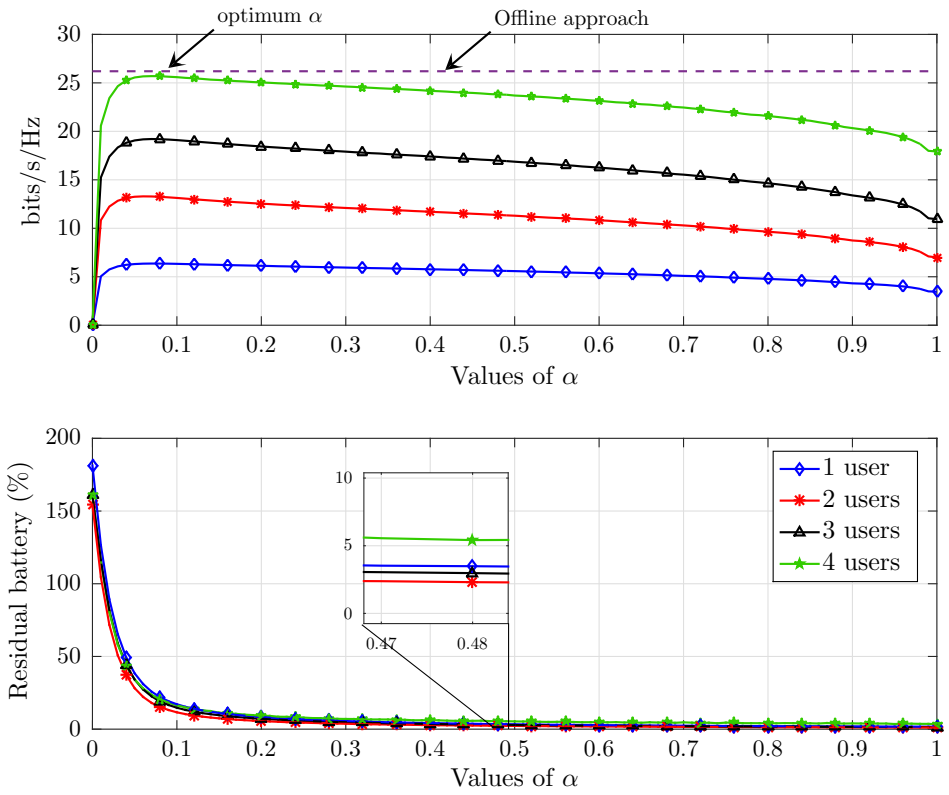


Figure 3.8: Average sum rate (over the 50 frames considered in the simulations) and aggregated residual battery in percentage after 50 frames as a function of α for different numbers of users. A system with up to 4 users is considered whose battery sizes $C^k_{\max}$ are 3,000 J, 6,000 J, 9,000 J, and 12,000 J. The transmitter has 10 antennas and the terminals are provided with 2 antennas each.

efficiencies. In this simulation we have considered that $\alpha_k = 0.5$, $\forall k$. We have extended the simulation window to 200 frames in order to show that the data rates converge to a specific value. Although it is not shown, it could also be checked by simulations that the average battery levels also converge. The specific value of the convergence depends on the statistics of the harvesting (p), the value of α_k, and the type of decoder, among other parameters. A more formal study of such asymptotic behavior will be given in Section 3.4.5. Notice that, if the time window was very large, then the transient period would become negligible. As commented before, a shorter time window was selected for the previous figures since we also want to evaluate the impact of the transient period on the performance of the system. When comparing the behavior of the exponential model with the linear model, we can see that the exponential model has a more abrupt slope than the linear one, since the battery consumption is faster. This phenomenon and the residual aggregated data rate (at the end of the transmission) depend, of course, on the specific efficiencies of the decoders. In this particular example, we see that the linear decoder with $\nu = 15{,}000$ W/(bits/s/Hz) ends up having less residual data rate than the exponential decoder. However, this statement cannot be generalized.

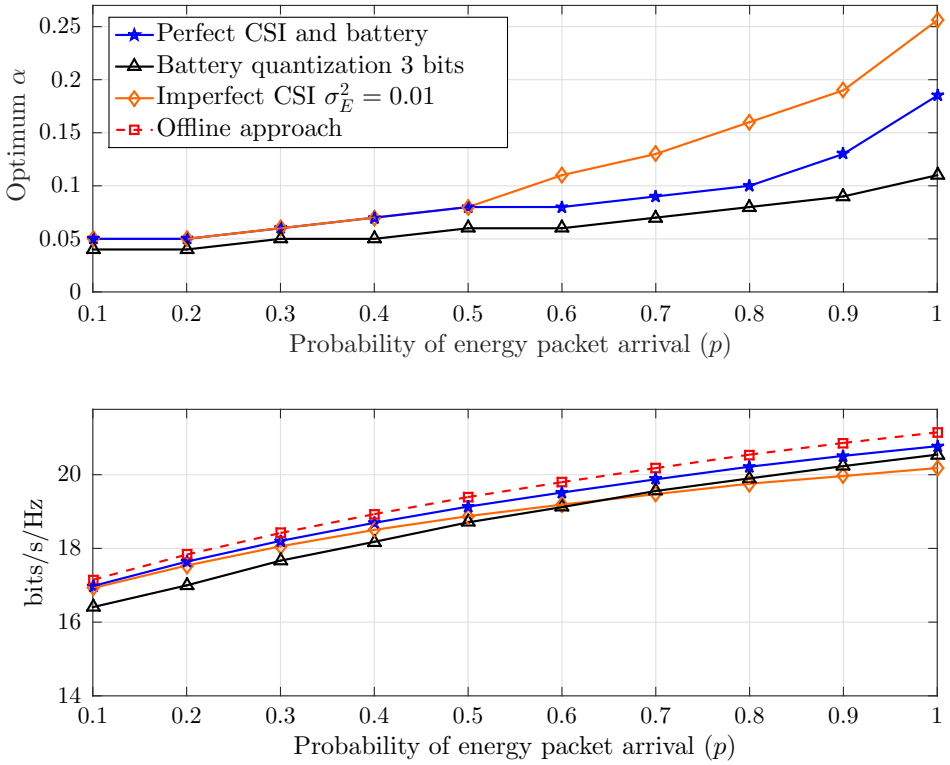


Figure 3.9: Optimum value of α and corresponding sum rate as a function of the probability of energy packet arrival.

3.4.5 Asymptotic Results

In this section we study the asymptotic behavior of the rates and the batteries of the users considering that the transmission is sufficiently long to attain convergence. The values of the data rates and the battery levels at each particular frame are obtained from Algorithm 3.1. Notice that, since the battery levels depend upon the harvesting process, which is a stochastic process, and the rates are obtained as the solution of problem (3.21), where some of the constraints are given by $R_{\max,k}$, (and, therefore, depend on the battery levels), then, both the battery levels and the rates are also stochastic processes. For simplicity, in this section we will consider that the battery sizes are infinite for all users, i.e., $C_{\max}^k = \infty, \forall k$, which means that no battery overflows can occur. Therefore, the battery dynamics are described by the following equation:

$$C_k(t+1) = C_k(t) - T_f \cdot P_{\text{tot},k}^{rx}\left(R_k(t)\right) + H_k(t). \tag{3.38}$$

Throughout the chapter, we have considered the harvesting to be stationary and ergodic. If fact, we modeled the harvesting as a discretized energy packet arrival process, where each energy packet contains a finite amount of Joules. However, as we will show later, even if the harvesting is stationary, that does not imply that the battery levels are stationary. In any case, we are not

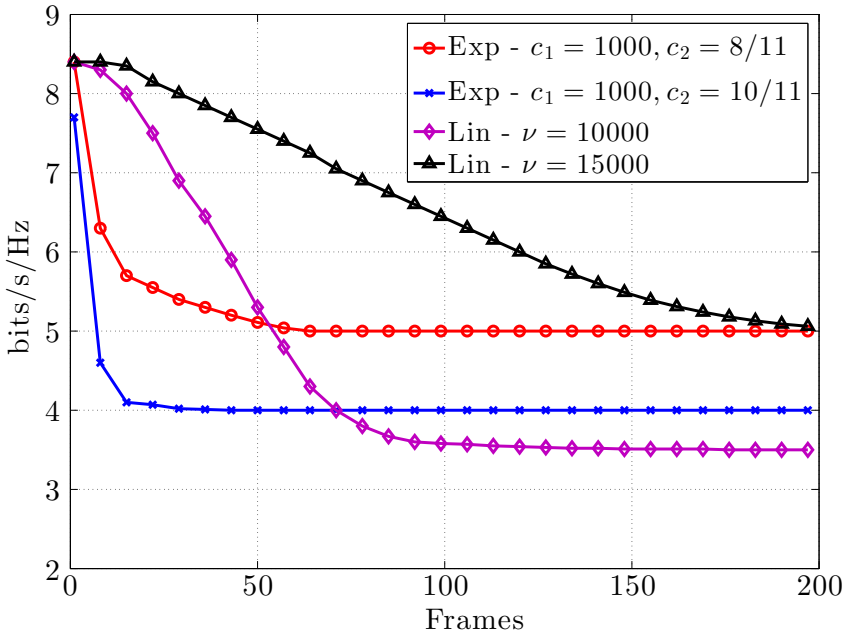


Figure 3.10: Average data rate evolution for different decoder consumption models (exponential and linear) and different decoder efficiencies. Note that, due to space limitations, we have not included the units of the decoder constants c_1, c_2, and ν.

interested in the stationarity of the process, but we seek to characterize the mean convergence[6] of the rates and the batteries. Given that, let us present the following result:

Conjecture 3.1. *In the steady state regime (i.e., as $t \to \infty$), we have that $\lim_{t\to\infty} \mathbb{E}[R_k(t)] = \varphi_k$, where φ_k is a constant such that $0 \leq \varphi_k < \infty$.*

Intuition behind the proof. The idea behind the proof is to note that the rates $R_k(t)$ are the solution of the optimization problem (3.21) where there is a maximum power constraint, $C1$, that implies that $R_k(t) < \infty$ and, therefore, $\mathbb{E}[R_k(t)] < \infty$. Moreover, as the rates depend on the harvesting process, which is stationary, $\mathbb{E}[R_k(t)]$ will not oscillate with time (see Section 3.4.5.2). A more formal proof of the convergence is out of the scope of this thesis. ■

Thus, from previous conjecture, we will assume throughout the section that the rates converge in mean. Of course, the instantaneous value of $R_k(t)$ depends on the current channel and battery level and, thus, $R_k(t)$ will show some random fluctuations throughout time. Unfortunately, the mean convergence of the battery depends upon several parameters, not just the radiated power and the harvesting. As a consequence, there is not a simple relation between the convergence of the rates and the batteries. Before analyzing the battery convergence, let us present another partial result:

Lemma 3.1. *The stochastic variable $R_{\max,k}(t)$ can only converge to the allocated user data rates, $\lim_{t\to\infty} \mathbb{E}[R_{\max,k}(t)] = \lim_{t\to\infty} \mathbb{E}[R_k(t)]$, or diverge $\lim_{t\to\infty} \mathbb{E}[R_{\max,k}(t)] = \infty$.*

[6] Throughout the section, we will say indistinctly mean convergence and convergence but, formally speaking, we refer to mean convergence if not stated otherwise.

Proof. We develop the proof for one particular user k, but it can be extended to the rest of the users. Let us consider that the rate has converged to a given constant, $\mathbb{E}[R_k(t)] = \varphi$. Let us assume that there is a gap between $\mathbb{E}[R_{\max,k}(t)]$ and $\mathbb{E}[R_k(t)]$, i.e., $\mathbb{E}[R_{\max,k}(t)] - \mathbb{E}[R_k(t)] = \kappa$, where $\kappa > 0$ (notice that a negative value of κ is not possible due to constraint $C3$). Then, since $\mathbb{E}[R_{\max,k}(t)]$ is constant, it implies that the battery has also converged, i.e., $\mathbb{E}[C_k(t+1)] = \mathbb{E}[C_k(t)]$. Hence, from (3.38), we have that $\mathbb{E}[T_f P_{\text{tot},k}(R_k(t))] = \mathbb{E}[H_k(t)]$, but since $\mathbb{E}[R_k(t)] < \mathbb{E}[R_{\max,k}(t)]$, constraint $C3$ is not active and the optimum data rates are the ones obtained from the classical water-filling policy. However, as the rates computed from the water-filling have no relation with the harvesting process, the probability of such event is 0, which leads to a contradiction. Thus, $\mathbb{E}[R_{\max,k}(t)]$ may decrease until convergence with $\mathbb{E}[R_k(t)]$ or diverge. ■

Based on the same principle presented in the previous lemma, we are able to define three regions of different asymptotic behaviors that are based on the battery convergence. Let us assume that the algorithm is already in steady state and that the rates have converged, i.e., $t \to \infty$. Then, we define the following three regions:

Definition 3.3. (Region 1 - R1). *This region is defined such that* $\mathbb{E}[T_f P_{\text{tot},k}(R_k(t))] < \mathbb{E}[H_k(t)]$, $\forall k$. *In such case, the batteries increase as t increases, i.e.,* $\mathbb{E}[C_k(t+1)] > \mathbb{E}[C_k(t)]$ *and, as a consequence,* $\lim_{t\to\infty} \mathbb{E}[C_k(t)] = \infty$, $\forall k$. In fact, $\lim_{t\to\infty} \mathbb{E}[R_{\max,k}(t)] = \infty$, *and, as a consequence, the rates* $R_k(t)$ *are only limited by the transmission power* P_T. *This means that the optimum rates are obtained by means of classical water-filling policy given by* $R_k^\star = \sum_{i=1}^{n_{S_k}} \log_2 \left(c \frac{\lambda_{i,k}^2}{\sigma^2} \right)^+$, *where* c *fulfills* $\sum_{i,k} \left(c - \frac{\sigma^2}{\lambda_{i,k}^2} \right)^+ = P_T$. *However, in reality, the batteries have a finite size and, in this case, they would grow until they reach their maximum capacity.*

Definition 3.4. (Region 2 - R2). *This region is defined such that* $\mathbb{E}[T_f P_{\text{tot},k}(R_k(t))] = \mathbb{E}[H_k(t)]$, $\forall k$, *which means that all the batteries have converged to a finite value. As expressed in* Lemma 3.1, *the batteries may increase or decrease until* $\mathbb{E}[R_{\max,k}(t)]$ *converges to* $\mathbb{E}[R_k(t)]$. *This is the most interesting region since it captures the behavior of a battery-limited network.*

Let us present the following result concerning R2:

Lemma 3.2. *Let the harvesting intensity be finite, i.e.,* $\mathbb{E}[H_k(t)] < \infty$ *and let the algorithm be in steady state* ($t \to \infty$). *If* $P_T \to \infty$, *then* $\lim_{t\to\infty} \mathbb{E}[R_k(t)] = \lim_{t\to\infty} \mathbb{E}[R_{\max,k}(t)]$.

Proof. The proof follows directly from Lemma 3.1 and Definition 3.4. ■

Definition 3.5. (Region 3 - R3). *This region is just an intermediate region between R1 and R2 where* $\mathbb{E}[T_f P_{\text{tot},k}(R_k(t))] < \mathbb{E}[H_k(t)]$ *may be true for some users and* $\mathbb{E}[T_f P_{\text{tot},k}(R_k(t))] = \mathbb{E}[H_k(t)]$ *may be true for other users. Hence, some users will experience a battery divergence, while others will experience a battery convergence. This means that, within this region, there are users behaving as being in R1 and others in R2.*

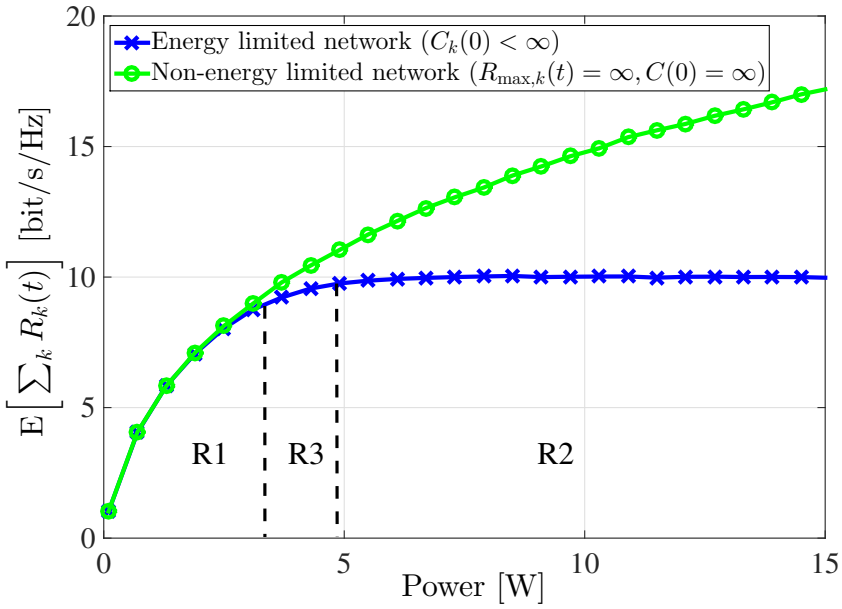


Figure 3.11: Aggregated expected value of the rates $\lim_{t\to\infty} \sum_k \mathbb{E}[R_k(t)]$.

The three regions are represented in Figure 3.11 and Figure 3.12. The setting of the simulation is: $T = 400$ for which we observe that the rates have already converged, the decoder power consumption model is linear with constant $\nu_k = 15,000$ W/(bits/s/Hz) (see (3.2)), $\alpha_k = 0.1$, the harvesting intensity is $p_k = 0.5$, the energy packet is $e_k = 100$ J, and there are three users with initial batteries $C_1(0) = 3,000$ J, $C_2(0) = 6,000$ J, and $C_3(0) = 1,500$ J. The normalized channel matrices are generated randomly with i.i.d. entries distributed according to $\mathcal{CN}(0, 1)$. In Figure 3.11, the green curve represents a system where the initial batteries are infinite, i.e., $R_{\max,k}(t) = \infty$, $\forall t, k$. (classical water-filling policy). The blue curve results from the application of the Algorithm 3.1. As we can see, in R1 the optimum expected rates are the same as the ones obtained from classical water-filling, which means that the network is limited by the radiated power and not by the energy available at the batteries of the receivers. Figure 3.12 depicts the evolution of the expected value of the batteries. Notice that the batteries in R1 should diverge, as stated in Definition 3.3, but as the number of simulated frames is $T = 400$, the obtained battery levels are finite.

From the definitions of the regions, we see that the thresholds between regions depend on the harvesting of the users. In the previous two figures we considered that the users were provided with the same energy harvesting source. Figure 3.13 and Figure 3.14 depict the expected value of the rates and the batteries where users are provided with different energy harvesting sources. By considering different harvesting sources among users (different energy packet sizes e_k), the three regions have been modified accordingly.

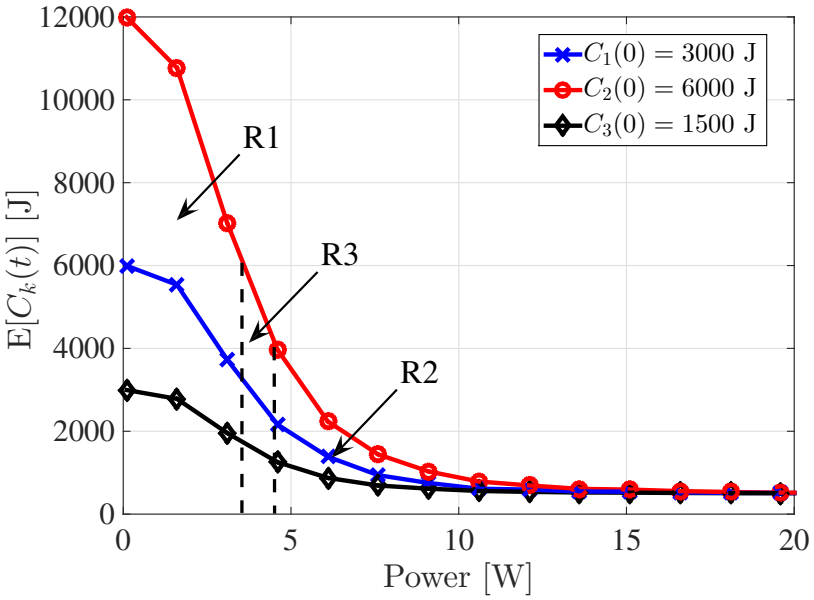


Figure 3.12: Expected value of the individual batteries.

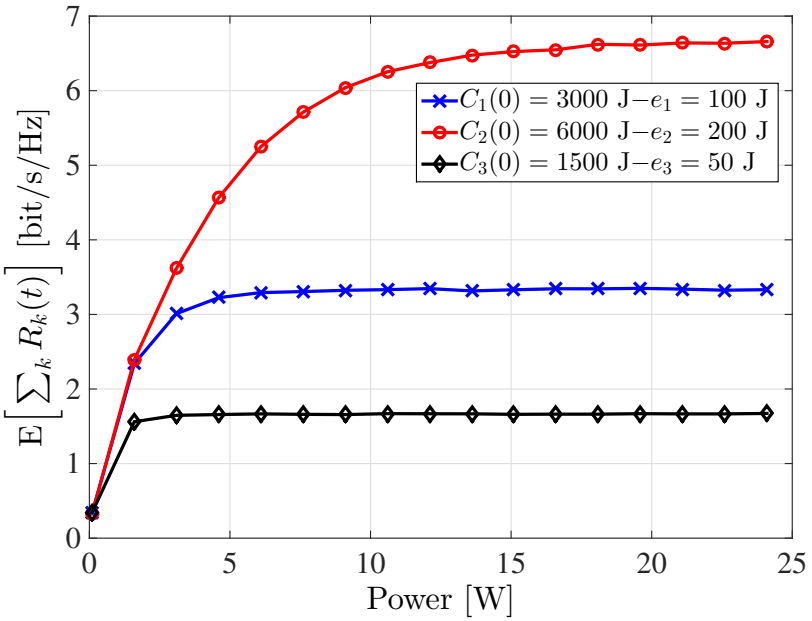


Figure 3.13: Expected value of the rates with users with different harvesting intensities.

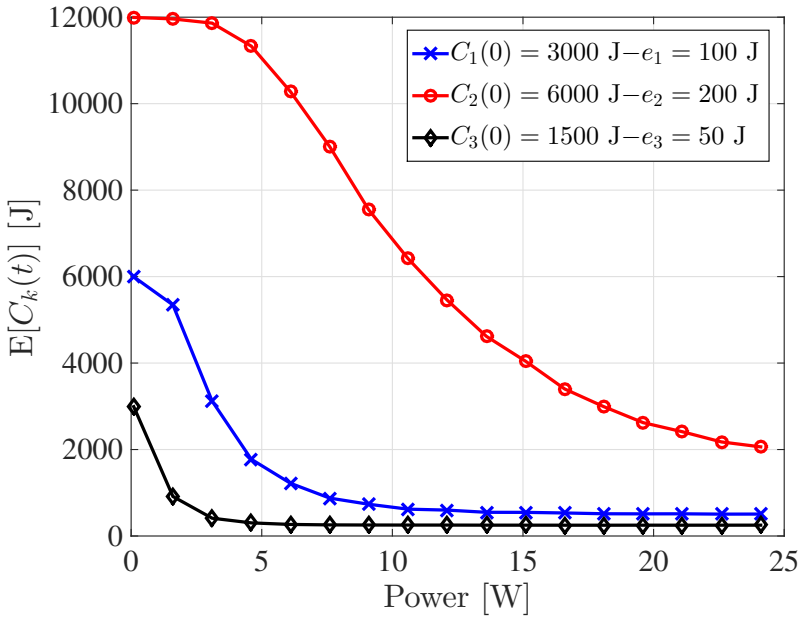


Figure 3.14: Expected value of the batteries with users with different harvesting intensities.

3.4.5.1 Analytic Characterization

In the previous section, we studied the asymptotic characterization of the Algorithm 3.1. We presented under what conditions the rates and the batteries converge and we discussed the different asymptotic behaviors that are possible in these networks. In addition, in this section we characterize the asymptotic behavior analytically for a concrete decoding power consumption model. This would allow us to determine what the values of the rate and the battery levels would be as a function of the initial conditions, the harvesting capabilities, the transmitted power, and the value of α_k.

As we discussed before, in R1 the batteries diverge and the rates are the ones obtained from classical water-filling policy, regardless of the decoder being used. Therefore, we only have to characterize R2, as R3 is a combination of R1 and R2 with an intermediate behavior.

In the previous sections, and throughout the chapter, we have considered two different models for the decoder consumption function $P_{\text{dec},k}(R_k(t))$: a linear model and an exponential model (see Section 3.3.1). In the following, we derive some expressions for the asymptotic values and for each specific decoding model.

Linear decoder consumption model

Recall that the linear decoder consumption is modeled as (3.2)

$$P_{\text{dec},k}(R_k(t)) = \nu_k R_k(t), \tag{3.39}$$

where ν_k models the decoder efficiency. Let us present some asymptotic results concerning this model.

Lemma 3.3. *The expected value of the data rate in convergence,* $\lim_{t\to\infty} \mathbb{E}[R_k(t)]$*, does not depend on* α_k *or on the initial battery level* $C_k(0)$*.*

Proof. Since we are in R2, the battery has converged as $t \to \infty$, i.e., $\mathbb{E}[C_k(t+1)] = \mathbb{E}[C_k(t)]$. Then, $\mathbb{E}[T_f P_{\text{tot},k}(R_k(t))] = \mathbb{E}[H_k(t)]$ holds. By just replacing the model introduced before, we end up with

$$\lim_{t\to\infty} \mathbb{E}[R_k(t)] = P_{\text{dec},k}^{-1}\left(\frac{\mathbb{E}[H_k(t)] - P_c^{rx}T_f}{T_f}\right) = \frac{p_k e_k - P_c^{rx}T_f}{\nu_k T_f}, \tag{3.40}$$

and this concludes the proof. ■

In case that the users are provided with the same harvesting source and the same decoder, they end up with the same expected rate. Then, the expected sum rate in convergence is just $\lim_{t\to\infty} \mathbb{E}[\text{SR}(t)] = \lim_{t\to\infty} \mathbb{E}\left[\sum_{k=1}^{K} R_k(t)\right] = \frac{K(p\cdot e - P_c^{rx}T_f)}{\nu_k T_f}$. Now, let us present a result concerning the value of the battery:

Lemma 3.4. *The expected value of the battery in convergence,* $\lim_{t\to\infty} \mathbb{E}[C_k(t)]$*, does not depend on the initial battery level* $C_k(0)$*.*

Proof. As we are in R2, $\mathbb{E}[R_k(t)] = \mathbb{E}[R_{\max,k}(t)]$ holds. From Lemma 3.3 and (3.10), we obtain the value of the battery in convergence:

$$\lim_{t\to\infty} \mathbb{E}[C_k(t)] = \frac{\mathbb{E}[H_k(t)]}{\alpha_k} = \frac{p_k e_k}{\alpha_k}, \tag{3.41}$$

which concludes de proof. ■

Interestingly, the value of the battery in convergence only depends on the harvesting source and the value of α_k. Thus, the total battery reduction from initial conditions considering the whole transmission would be $\sum_{k=1}^{K} C_k(0) - \sum_{k=1}^{K} \frac{p_k e_k}{\alpha_k}$.

Exponential decoder consumption model

Recall that the exponential decoder consumption is modeled as (3.3)

$$P_{\text{dec},k}(R_k(t)) = c_{1k}\mathrm{e}^{c_{2k}R_k(t)}, \tag{3.42}$$

where c_{1k} and c_{2k} model the decoder efficiency. Let us present some asymptotic results concerning this model.

Lemma 3.5. *The expected value of the data rate in convergence,* $\lim_{t\to\infty} \mathbb{E}[R_k(t)]$*, does not depend on the initial battery level* $C_k(0)$*.*

Proof. Since we are in R2, the battery has converged, i.e., $\mathbb{E}[C_k(t+1)] = \mathbb{E}[C_k(t)]$. Then, $\mathbb{E}[T_f P_{\text{tot},k}(R_k(t))] = \mathbb{E}[H_k(t)]$ holds, which proves that the final rate does not depend on the initial battery value. In addition, we can find an upper-bound for the final average rate. Since $P_{\text{dec},k}(R_k(t))$ is a convex function, we can apply Jensen's inequality[7] [Cov06] and obtain the following upper bound on the expected rate:

$$\mathbb{E}[R_k(t)] \leq P_{\text{dec},k}^{-1}\left(\frac{\mathbb{E}[H_k(t)] - P_c^{rx}T_f}{\nu_k T_f}\right) \leq \frac{1}{c_{2,k}} \ln\left(\frac{p_k e_k - P_c^{rx}T_f}{c_{1,k}T_f}\right), \tag{3.43}$$

which concludes the proof. ■

Unfortunately, we can not provide an exact analytic result for the value of the battery in convergence if the receivers' decoding consumption is modeled with the exponential function.

3.4.5.2 Numerical Simulations

In this section we present some numerical simulations that support the analytic results derived in the previous sections. For the simulations, we consider a scenario with a BS with $n_T = 6$ antennas and $K = 3$ receivers with $n_{R_k} = 2$ antennas each. The front-end power consumption at the receiver is $P_c^{rx} = 0.2$ W. The maximum available power at the BS is $P_T = 5$ W. We assume a normalized noise power given by $\sigma^2 = 1$ W and a frame duration equal to $T_f = 1$ ms. For simplicity, the probability of energy packet arrival is $p_k = 0.5$, $\forall k$, the energy packet size is $e_k = 100$ J (all users are provided with the same energy harvesting source) and $\alpha_k = 0.5$, $\forall k$. This configuration yields to a battery-limited scenario, i.e., all users lie in region 3. The channel matrices are generated randomly with i.i.d. entries distributed according to $\mathcal{CN}(0,1)$. The initial battery levels are 6,000 J, 3,000 J, and 1,500 J. The type of decoder and the decoder efficiencies are specified in the title of the figures. All results are averaged over 1,000 channel and 1,000 harvesting realizations.

Figure 3.15 shows the evolution of the data rates. As we can see, for a given number of frames, the convergence time depends on the specific decoder and its efficiency. If we compute the asymptotic expected rate using equation (3.40) for the linear decoder with the corresponding simulation parameters, we obtain $\lim_{t\to\infty} \mathbb{E}[R_k(t)] = 5$ bits/s/Hz and $\lim_{t\to\infty} \mathbb{E}[R_k(t)] = 3.3$ bits/s/Hz for the efficiencies $\nu_k = 10,000$ W/(bits/s/Hz) and $\nu_k = 15,000$ W/(bits/s/Hz), respectively. We are able to verify that result from the corresponding figures. Focusing on the exponential decoder, if we compute the upper bound given by (3.43), we obtain $\lim_{t\to\infty} \mathbb{E}[R_k(t)] \leq 5.3$ bits/s/Hz

[7] Jensen's inequality states that if X is a random variable and φ is a convex function, then $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$.

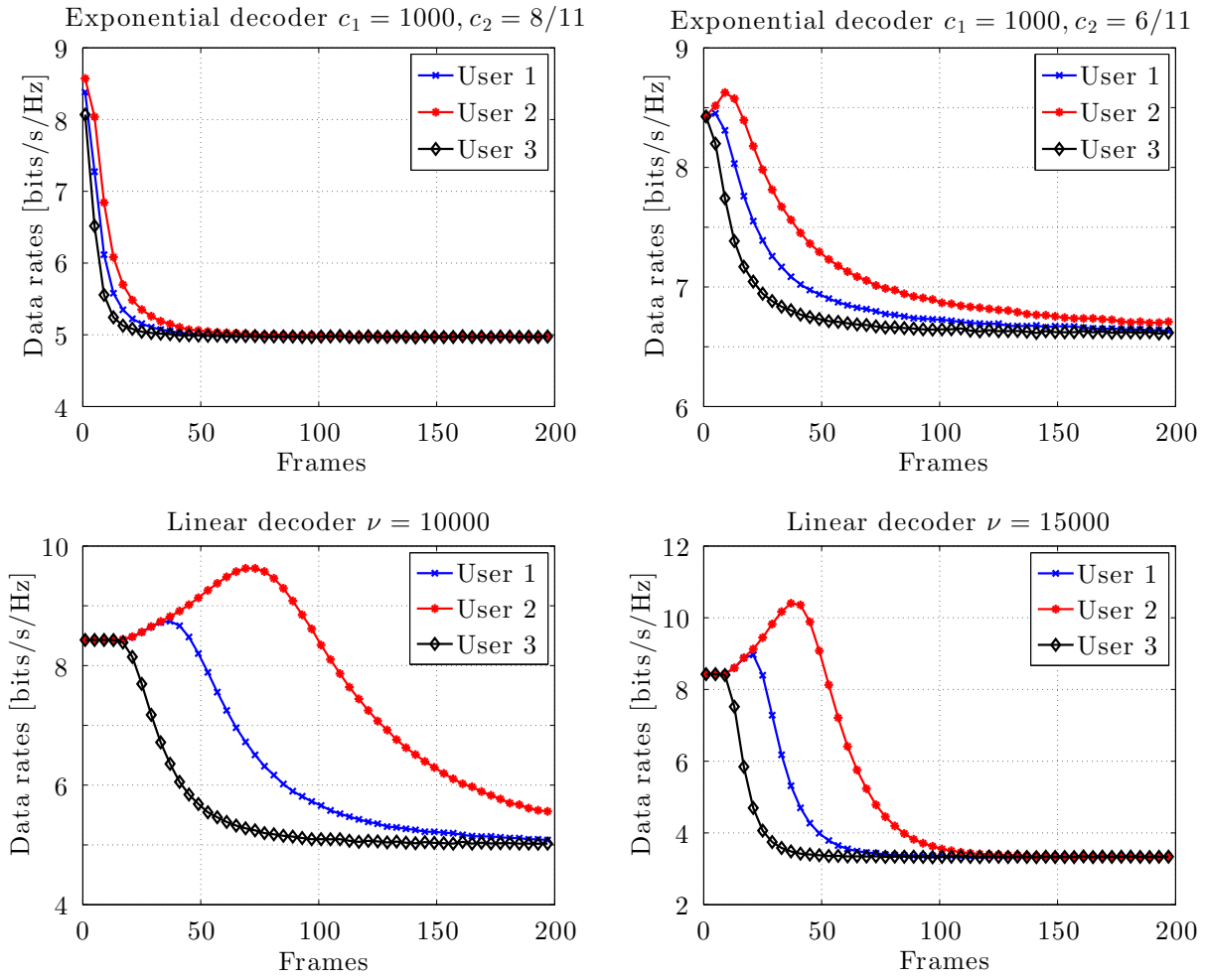


Figure 3.15: Evolution of the data rates for different types of decoder power consumption models. The units of the decoder constants c_1, c_2, and ν are W, 1/(bits/s/Hz), and W/(bits/s/Hz), respectively.

and $\lim_{t\to\infty} \mathbb{E}[R_k(t)] \leq 7$ bits/s/Hz, so there is approximately 0.3 bits/s/Hz of difference with the value obtained in the figures.

Figure 3.16 presents the evolution of the battery levels for the same decoders and decoder efficiencies. We can also verify the analytic result presented in (3.41) for the linear decoder, which is $\lim_{t\to\infty} \mathbb{E}[C_k(t)] = 100$ J.

3.5 Scheduling Procedures for User Selection

3.5.1 Introduction

In the previous sections, we have considered a deployment where MIMO was the physical technology available at both sides of the communication ends. However, practical implementations in multiuser scenarios (such as, for example BD [Spe04]) have the drawback of the limitation on the number of users to be served simultaneously. In the previous sections, we have considered

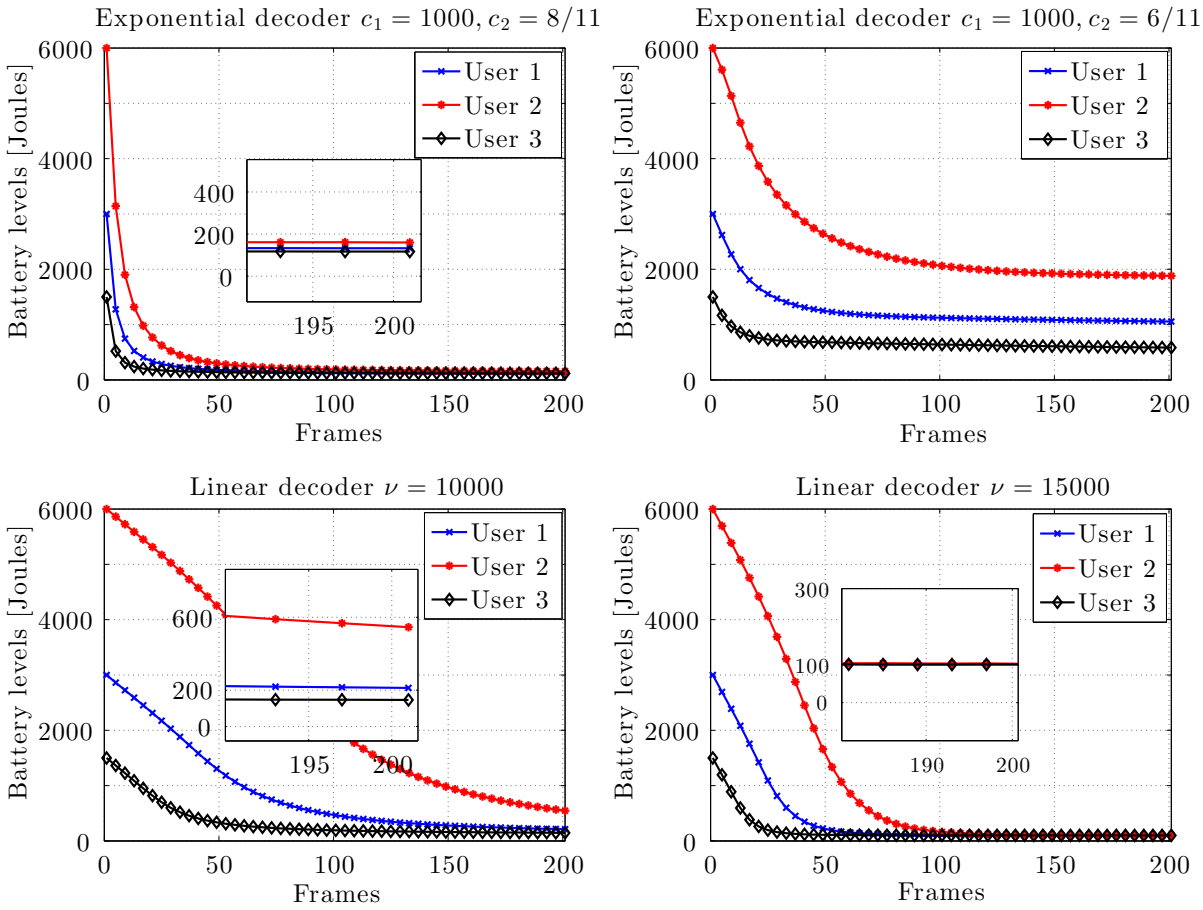


Figure 3.16: Evolution of the battery levels for different types of decoder power consumption models. The units of the decoder constants c_1, c_2, and ν are W, 1/(bits/s/Hz), and W/(bits/s/Hz), respectively.

that the group of users to be served was fixed and known. However, in practical scenarios, users demand access to the network and, thus, a user selection mechanism is needed to group the users under a given criterion. The optimum user selection mechanism usually requires an exhaustive search among all possible subsets, which is very computationally demanding even for a moderate number of users. In this context, a large number of suboptimal yet simplified algorithms have been proposed under different criteria [Liu10], [Yi11], [Dim05], [Sig09]. At this point it is important to remark that none of the previous works has considered the information of the battery status of the terminals in the proposed user selection mechanisms. Note that if users are battery-limited and their battery decrease accordingly to the service being supported, the battery status should be a key parameter to decide which users should be served if an overall metric that evaluates the performance for the whole transmission period is considered. This is the focus and contribution of this section.

In this section, we propose a user selection mechanism for a multiuser MIMO DL scenario based on the proportional fair (PF) scheduling criterion [Jal00] that takes into account the current battery levels of the terminals as well as the instantaneous and average rates. The impact of the harvesting intensities is also evaluated.

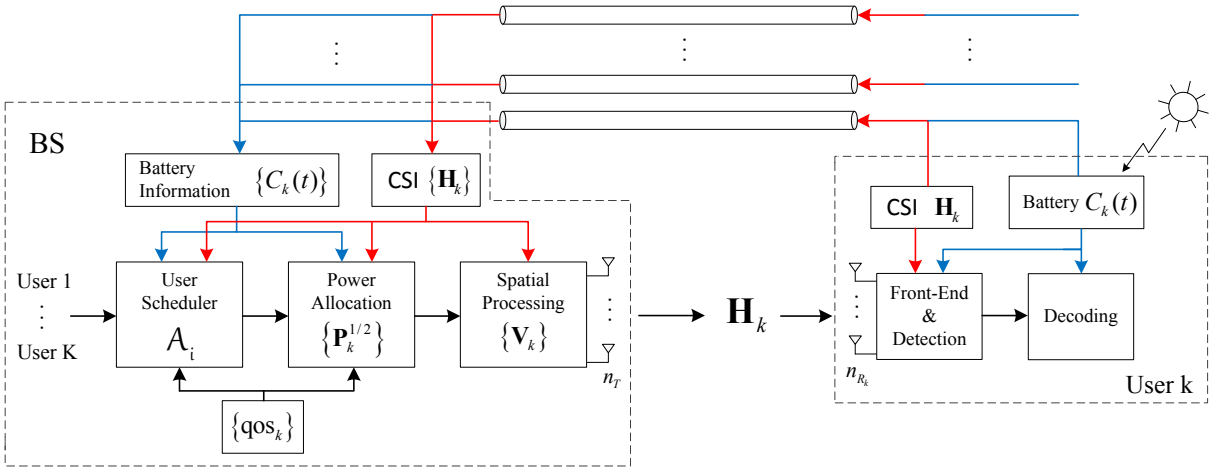


Figure 3.17: General downlink reference scenario.

The proposed scenario is depicted in Figure 3.17, where only one terminal is shown for simplicity. As it is represented, the transmitter needs to allocate a given finite amount of resources among the users. The strategy to be used for spatial precoding is based on the technique developed in Section 3.4 (that was based on BD [Spe04]). With BD, each user's precoding matrix is restricted to be in the null space of all other users' channels. For that reason, the number of users that can be supported simultaneously with BD is limited by the number of transmit and receive antennas. As a consequence, before assigning the resources, the transmitter must select the group of users to be served during each particular frame.

The main idea of this section is to propose a user scheduling mechanism based on PF that considers not only the users' channel dynamics, but also takes into account the battery fluctuations of the terminals. Because users' batteries will be charged and discharged according to the decoding process and harvesting, the scheduler should provide service to users who have enough battery for receiving the messages.

3.5.2 Energy-Aware Proportional Fair Scheduling

Let us first, for simplicity, assume that every user has the same number of antennas, i.e., $n_{R_k} = n_R$, $\forall k \in \mathcal{K}$. The maximum number of simultaneous users to be served following the BD strategy is $\hat{K} = \lceil \frac{n_T}{n_R} \rceil$ [Spe04].

The objective of the user scheduling is to select a set of users at each frame based on the PF criterion [Jal00] and considering energy limitations at the terminals. The PF strikes a good balance between the system throughput and the fair allocation of users' throughput. Let $\mathcal{A}_i$ be a subset of $\mathcal{K}$ and $|\mathcal{A}_i| \leq \hat{K}$. Let $\mathcal{A}$ be the set containing all possible user groupings $\mathcal{A}_i$, i.e., $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \dots\}$.

The PF can be achieved by maximizing the sum of the logarithm of all users' long-term

average throughput [Kel97], i.e.,

$$\max_{\mathcal{A}_i \in \mathcal{A}} \sum_{k \in \mathcal{A}_i} \log(T_k(t)), \tag{3.44}$$

where $T_k(t)$ is the exponentially averaged rate computed as

$$T_k(t) = \left(1 - \frac{1}{T_c}\right) T_k(t-1) + \frac{1}{T_c} R_k(t), \tag{3.45}$$

where $R_k(t)$ is the instantaneous data rate that can be supported by user k within the group $\mathcal{A}_i$ (and that is obtained from the solution of the optimization problem (3.21)) and T_c is the effective length of the exponential impulse response of the averaging filter in terms of the number of frames. If user k is not scheduled in the group, then $R_k(t) = 0$. Notice that since the spatial processing to be employed is based on the BD technique, the instantaneous rate of a given user depends on the particular group under evaluation, i.e., it depends on the spatial geometry of the users' channels, and also on the current battery states of such users. Now, if we substitute (3.45) in (3.44) we get:

$$\begin{aligned}
&\max_{\mathcal{A}_i \in \mathcal{A}} \sum_{k \in \mathcal{A}_i} \log\left(\left(1 - \frac{1}{T_c}\right) T_k(t-1) + \frac{1}{T_c} R_k(t)\right) \\
&= \max_{\mathcal{A}_i \in \mathcal{A}} \sum_{k \in \mathcal{A}_i} \log\left(T_k(t-1) + \frac{1}{T_c}\left(R_k(t) - T_k(t-1)\right)\right) && (3.46) \\
&\approx \max_{\mathcal{A}_i \in \mathcal{A}} \sum_{k \in \mathcal{A}_i} \log\left(T_k(t-1)\right) + \frac{1}{T_c} \sum_{k \in \mathcal{A}_i} \frac{\partial \log\left(T_k(t-1)\right)}{\partial T_k(t-1)} \left(R_k(t) - T_k(t-1)\right) && (3.47) \\
&= \max_{\mathcal{A}_i \in \mathcal{A}} \sum_{k \in \mathcal{A}_i} \log\left(T_k(t-1)\right) + \frac{1}{T_c} \sum_{k \in \mathcal{A}_i} \frac{R_k(t) - T_k(t-1)}{T_k(t-1)} && (3.48) \\
&\equiv \max_{\mathcal{A}_i \in \mathcal{A}} \sum_{k \in \mathcal{A}_i} \frac{R_k(t)}{T_k(t-1)}, && (3.49)
\end{aligned}$$

where in the previos Taylor's approximation[8] we have used the fact that T_c is large, and in the last step we notice that $T_k(t-1)$ is constant at the scheduling period t. The approximation taken in the second step will be an equality only when $T_c \longrightarrow \infty$. Notice that the equation found in the last step coincides with the conventional PF ratio criterion. In any case, as it will be shown later, we prefer the formulation given by the quotient of the instantaneous rate and the average achieved rate rather than the logarithmic formulation, as it is easier to handle. Note that the PF scheduler penalizes the users with large average throughputs and gives priority to users with low average throughputs under the objective of achieving a fair allocation of the resources throughout time.

The optimum solution in terms of maximizing the long-term average rates can only be found by brute force exhaustive search among all possible sets $\mathcal{A}_i \in \mathcal{A}$. The computational complexity of such algorithm is roughly $\mathcal{O}\left(\binom{K}{\hat{K}} \hat{K} n_T^3\right)$ (see Appendix 3.F), which is prohibitive for a decent

[8] Recall that $\log(x + \Delta) \approx \log(x) + \frac{1}{x}\Delta$.

number of users. For that reason, in this section we present a low complexity but suboptimal approach that aims to reduce such computational complexity.

If long-term QoS wants to be considered in the PF criteria, (3.49) needs a modification. A weight function $\eta_k(t)$ is introduced accordingly as follows:

$$\max_{\mathcal{A}_i \in \mathcal{A}} \sum_{k \in \mathcal{A}_i} \eta_k(t) \frac{R_k(t)}{T_k(t-1)}. \tag{3.50}$$

The community has proposed different functions $\eta_k(t)$ [Liu04]. In this chapter, we consider two options:

$$\eta_k(t) = \mathrm{qos}_k, \tag{3.51}$$

whose objective is to assign a proportional average rate corresponding to the user-requested long-term QoS, qos_k, and one based on the barrier function

$$\eta_k(t) = 1 + \nu_k e^{-\xi_k (T_k(t-1) - \mathrm{qos}_k)}, \tag{3.52}$$

where ν_k and ξ_k are constants. In this case, users far below from its qos_k are highly penalized and, thus, served with higher priority. In the simulations, (3.51) will be denoted as PF-QoS and (3.52) as PF-B. Notice that, in order to implement a PF scheduler on top of the resource allocation strategy presented in Section 3.4, we need to incorporate a set of weights in the objective function. These weights, defined as $\omega_k(t) = \frac{\eta_k(t)}{T_k(t-1)}$, will provide the scheduler with information to be able to carry out the decision of which users must be selected at each particular scheduling period.

The proposed algorithm is presented in Algorithm 3.4. It is based on a greedy set selection (users are added one by one at each iteration of step 5). The computational burden of this algorithm is approximately $\mathcal{O}\left(K\hat{K}n_T^3\right)$ (see Appendix 3.E). Notice that as the rates are computed using (3.21), the current battery of the terminals is implicitly considered through constraint $C3$ of (3.21). Users with higher battery level have higher $R_{\max}$ and, thus, they may be served with higher rate.

Algorithm 3.4 Proportional fair-based scheduler with energy constraints

set $\mathcal{K} = \{1, 2, ..., K\}$ and $\mathcal{A} = \emptyset$
find $k_1 = \arg\max_k \max_{\mathbf{Q}_k} \omega_k \log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right)$ s. t. $C1, \ldots, C4$ and $\mathbf{Q}_k \succeq 0$
set $f_{\text{temp}} = \omega_{k_1} R_{k_1}$
set $\mathcal{A} \leftarrow \mathcal{A} \cup \{k_1\},\ \mathcal{K} \leftarrow \mathcal{K} \setminus \{k_1\}$
for $j = 2 : \hat{K}$
    for every $i \in \mathcal{K}$
        let $\mathcal{A}_i = \mathcal{A} \cup \{i\}$
        find precoding $\mathbf{V}_m = \tilde{\mathbf{V}}_m^{(0)}\mathbf{V}_m^{(1)}$ based on SVD of matrices $\{\mathbf{H}_m\}$, $m \in \mathcal{A}_i$ as in (3.23)
        obtain the n_{s_m} singular values $\{\lambda_{n,m}\}$ of $\mathbf{H}_m\tilde{\mathbf{V}}_m^{(0)}$
        water fill over $\{\lambda_{n,m}\}$ based on (3.25), $\forall m \in \mathcal{A}_i$ considering energy constraints
        find the precoder matrices $\mathbf{B}_m = \mathbf{V}_m\mathbf{P}_m^{1/2}$
        find the throughput of the users in the set $\mathcal{A}_i$ as $R_m^\star = \sum_{i=1}^{n_{S_m}} \log_2\left(1 + \frac{1}{\sigma^2}p_{n,m}^\star\lambda_{n,m}^2\right)$
        compute $f_i = \sum_{m\in\mathcal{A}_i} \omega_m R_m$
    end for
    let $k_j = \arg\max_{i\in\mathcal{K}} f_i$
    if $f_{k_j} < f_{\text{temp}} \longrightarrow$ go to 23
    else
        $\mathcal{A} \leftarrow \mathcal{A} \cup \{k_j\},\ \mathcal{K} \leftarrow \mathcal{K} \setminus \{k_j\}$
        let $f_{\text{temp}} = f_{k_j}$
    end if
end for
the selected user set is $\mathcal{A}$
end algorithm

3.5.3 Numerical Simulations of the Scheduling Policy

This section presents numerical results. We consider a scenario with 10 users, i.e., $|\mathcal{K}| = 10$, a BS with $n_T = 6$ antennas and receivers with $n_R = 2$ antennas each. The maximum available power at BS is $P_{\max} = 5$ W and $P_c^{rx} = 0.2$ W. The noise power is[9] $\sigma^2 = -174$ dBm/Hz, the system bandwidth is 10 MHz, and the frame duration is $T_f = 1$ ms. The channel matrices are generated with i.i.d. Gaussian distributed entries $\sim \mathcal{CN}(0, 1)$ and include a model of pathloss $100 + 37.6\log_{10}(\text{d}_{\text{km}})$, where $0 \leq \text{d}_{\text{km}} \leq 0.05$ [km], and log-normal shadowing with a standard

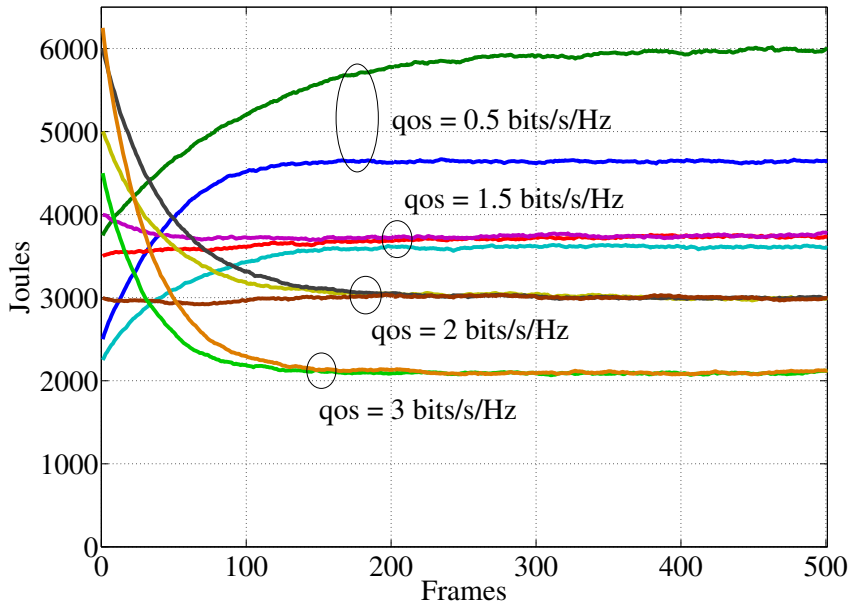
[9]Note that we have assumed so far a normalized noise power $\sigma^2 = 1$ W. However, in this section, in order to evaluate the scheduling of users, we consider a realistic channel model with path loss and, as a consequence, we need to consider a non-normalized noise power so that reasonable throughput values are obtained.

deviation of 8 dB. The maximum battery capacities are generated randomly within the interval $[3,000, 12,000]$ J. The traffic model considered is full buffer for all users. The model of decoder consumption is exponential given by $P_{\text{dec},k}(R_k) = c_1 e^{c_2 R_k}$ with $c_1 = 1,000$ W and $c_2 = 8/10$ 1/(bits/s/Hz). The harvesting is modeled as a stationary Bernoulli process with $p = 0.4$ and with an energy packet size of 200 J for all users. The length of the averaging filter is $T_c = 20$ frames, and the constants of the barrier function are $\nu = \xi = 4$. We consider 4 types of users: 20% demand qos = 0.5 bits/s/Hz, 30% demand qos = 1.5 bits/s/Hz, 30% demand qos = 2 bits/s/Hz, and 20% demand qos = 3 bits/s/Hz. The number of frames for the simulations is 500. All results are averaged over 10,000 channel and harvesting realizations.

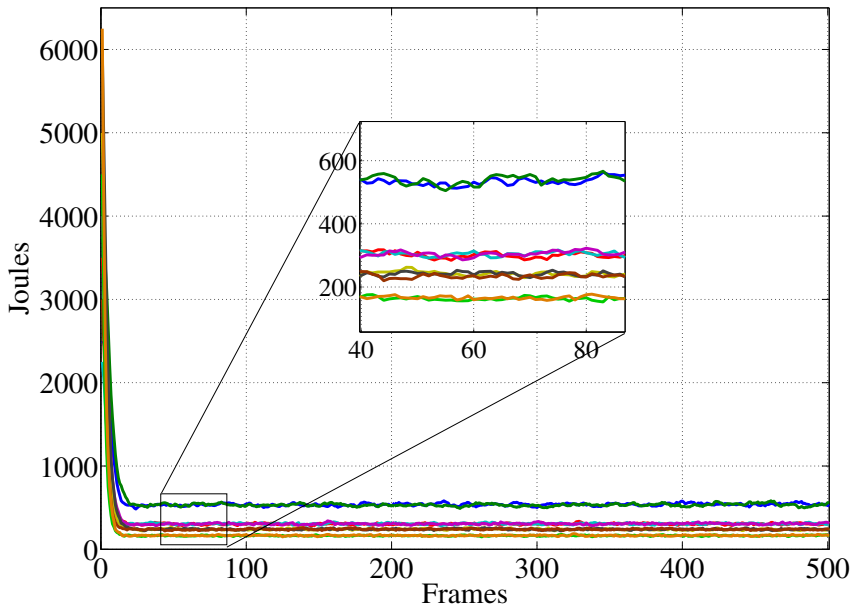
Figure 3.18 shows the evolution of the battery levels for two different values of α for the PF with the barrier function (denoted as PF-B). Recall that $\alpha = 1$ means that the users are able to spend all the battery for decoding, i.e., the users do not have energy limitations but, of course, their batteries are finite. Let us mention that $\alpha = 0.1$ was chosen because, as will be shown later, it was found to be (through simulations) optimal in terms of data rate. In general trends, what we can see is that there is a transient period at the beginning and then the battery levels converge to a steady state. If we focus on the extreme case, Figure 3.18-(b) with $\alpha = 1$, users achieve high data rates at the beginning (since they do not have energy limitations), but their batteries decrease quite fast. After that transient period, the energy collected by the harvesting is able to compensate the battery reduction due to data decoding and the battery levels get stabilized. On the other hand, in Figure 3.18-(a) we see that some batteries even increase at the beginning. This is so because these users are energy-constrained and they achieve lower instantaneous rates (according to their QoS). As a conclusion, for long-term QoS assurance, it is better to control the amount of energy the users are using rather than allowing them to spend all the energy in any frame. Another interesting effect is that users with the same qos end up with similar battery level. This is so because they are forced to achieve the same rate and have the same energy harvesting sources. Note that the initial values of the batteries are irrelevant.

Figure 3.19 depicts the average data rate of the three proposed schedulers for different values of α. As it can be seen and expected, the PF ($\eta_k(t) = 1$) provides approximately the same data rates for all users regardless of their qos values. The PF-B scheme tries to compensate users according to their requested long-term qos but, due to energy limitations, the group with the highest requested QoS cannot achieve it. On the other hand, the scheduler PF-QoS focus only on the proportionality between rates. Thus, users with qos= 1.5 bits/s/Hz should approximately achieve 3 times more rate than users with qos= 0.5 bits/s/Hz.

Note that, although the PF-QoS strategy should provide proportional long-term average rates to users belonging to different QoS groups, from the observation of Figure 3.19 we can deduce that the scheduler is not always able to achieve this behavior. The explanation lies on the fact that users are battery-limited and the scheduler cannot maintain this proportionality factor because the relation between the achieved data rate and the battery consumption is exponential,



(a) $\alpha = 0.1$.



(b) $\alpha = 1$.

Figure 3.18: Evolution of the battery levels for two different values of α and with different users demanding different QoS.

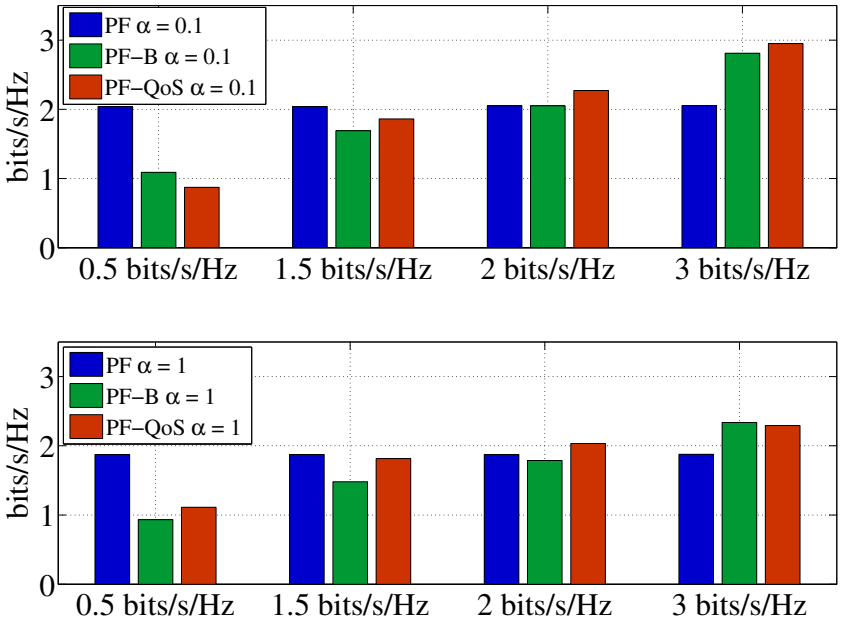


Figure 3.19: Average data rates for users with different types of QoS for the three proposed scheduling strategies and different values of α.

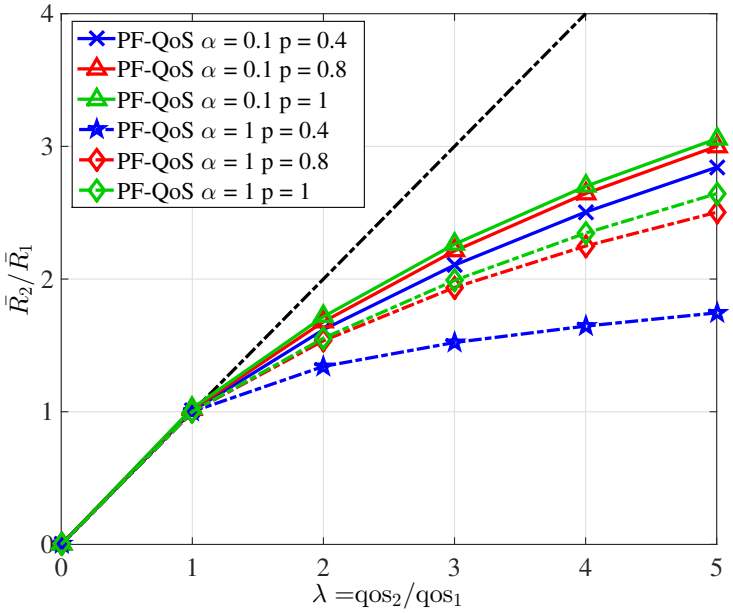


Figure 3.20: Evolution of the ratio of average data rates as a function of the ratio of qos.

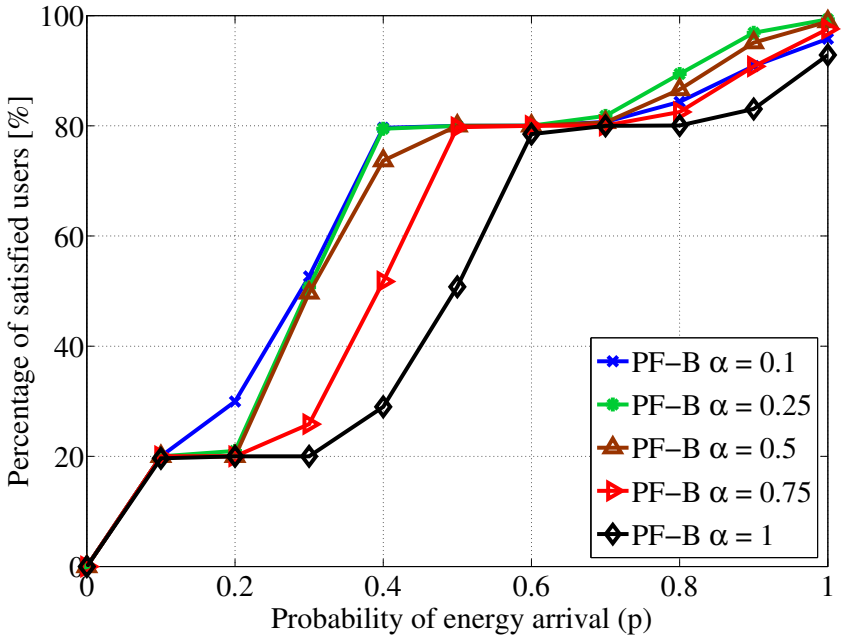


Figure 3.21: Percentage of users achieving their qos.

i.e., a linear increase in the rate would imply an exponential decrease of the battery level. Figure 3.20 shows, for a system with only two types of users with two QoS's, the proportionality between the achieved rates for different proportionalities between the demanded qos. The dotted line is the reference in which the achieved rates maintain the QoS proportionality. As it can be seen, if α increases, so does the saturation. Notice also that, the saturation improves if the probability p of receiving energy harvesting packets increases. Note that even with $p = 1$, the scheduler is not able to provide the requested qos to the users belonging to the group with qos_2 (obviously, this also depends on the energy packet size).

Finally, Figure 3.21 shows the amount of users that achieve their qos constraints for the PF-B scheduler as a function of the harvesting probability p. If we focus on $p = 0.4$, the value considered in the previous simulations, we see that $\alpha = 0.1$ provides the highest amount of satisfied users, 80%, compared only with the 30% of satisfied users when $\alpha = 1$ is considered. Note that, as the intensity of harvesting increases, this is no longer true and the optimum energy allocation for decoding (optimum α) that provides the largest amount of satisfied users is higher, almost 100% of user satisfaction with $\alpha = 0.25$ if $p = 1$.

As general concluding remarks, it is important to point out that it is always better to control the energy to be used for decoding for all users rather than allowing them to spend all the battery freely. The performance improves if the harvesting probability increases.

3.6 Chapter Summary and Conclusions

In this chapter, a multiuser MIMO precoder design for short distance networks has been presented. The mobile terminals have been considered to be battery-powered devices provided with energy harvesting sources. The key point is that the battery status of the users have been taken into account explicitly in the precoder and resource allocation design, increasing the lifetime of the nodes. The optimum solution, based on the application of BD, has been found to diagonalize the equivalent channel but with a maximum allowable decoding power per user due to energy constraints. The resulting algorithm has been a modification of the water-filling policy obtained thanks to duality theory. We have also considered the case of imperfect CSI at the transmitter and a robust precoder design has been derived for that scenario. The effect of imperfect CSI at transmitter is translated into a loss in the system sum rate compared to the perfect CSI case.

Then, we have addressed the problem of battery quantization due to the use of rate-limited feedback channels. The transmitter designs the precoders taking into account the battery knowledge imperfections explicitly, yielding again to a robust approach. The rate obtained with the robust design based on battery knowledge imperfection is lower when compared with the ideal one since it is more rate conservative. Simulations have shown that if the battery status of the users is incorporated in the design, the nodes improve their battery lifetime, whereas the average sum rate is enhanced at the same time. Thus, classical maximum sum rate techniques that do not take into account the status of the batteries of the users that have to be served are inefficient from the point of view of energy efficiency.

Later, we have characterized the asymptotic behavior of the battery levels and the data rates of the proposed resource allocation strategy. In particular, we have obtained in closed form the expressions of the battery levels and the data rates once the system has converged, e.g., for long time windows.

Finally, we have addressed the problem of designing a user selection algorithm. Linear precoding has been adopted throughout the chapter at the transmitter and, thus, the number of users to be served at each particular scheduling period is limited by the number of available transmit and receive antennas. The adopted criterion for the user selection was based on the PF in which the selection strategy exploited not only the channel dynamics but also the battery fluctuations. Simulation results have shown that the average long-term rates approach the requested QoS if the scheduler controls the amount of energy that is being used by the terminals at each scheduling period.

3.A Proof of Proposition 3.1

Let us start the proof by defining the corresponding Lagrangian and the KKT conditions of problem (3.24). For simplicity in the notation, constraint $C5$ in problem (3.24) is omitted in the Lagrangian since we force the positivity of the variables $\{p_{i,k}\}$ in the analytic solution without loss of optimality (approach followed in other works). Let μ, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_K)$ be the Lagrange multipliers associated with constraints $C1$, $C2$, $C3$, and $C4$ of problem (3.24). Collect all multipliers in $\boldsymbol{\Upsilon} = (\mu, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\nu})$. Collect all powers in $\mathbf{p} = (p_{i,k}, \forall i, k)$ and all data rates in $\mathbf{R} = (R_k, \forall k)$. Hence, the Lagrangian of problem (3.24) is defined as

$$\begin{aligned}\mathcal{L}(\mathbf{R}, \mathbf{p}; \boldsymbol{\Upsilon}) = &- \sum_{k=1}^{K} R_k + \mu \left(\sum_{k=1}^{K} \sum_{i=1}^{n_{S_k}} p_{i,k} - P_T \right) \\ &+ \sum_{k=1}^{K} \gamma_k \left(R_k - \sum_{i=1}^{n_{S_k}} \log \left(1 + \frac{1}{\sigma^2} p_{i,k} \lambda_{i,k}^2 \right) \right) \\ &+ \sum_{k=1}^{K} \beta_k \left(R_k - R_{\max,k} \right) + \sum_{k=1}^{K} \nu_k \left(\mathrm{qos}_k - R_k \right). \end{aligned} \tag{3.53}$$

Let $R_k^\star, p_{i,k}^\star$, and $\mu^\star, \gamma_k^\star, \beta_k^\star, \nu_k^\star$ be any set of primal and dual optimal variables. Then, the KKT conditions of problem (3.24) are shown below:

$$-1 + \gamma_k^\star + \beta_k^\star - \nu_k^\star = 0 \, \forall k, \tag{3.54}$$

$$\mu^\star - \frac{1}{\ln(2)} \frac{\gamma_k^\star \lambda_{j,k}^2}{\sigma^2 + p_{j,k}^\star \lambda_{j,k}^2} = 0 \, \forall j, k, \tag{3.55}$$

$$\sum_{k=1}^{K} \sum_{i=1}^{n_{S_k}} p_{i,k}^\star - P_T \leq 0, \tag{3.56}$$

$$R_k^\star - \sum_{i=1}^{n_{S_k}} \log \left(1 + \frac{1}{\sigma^2} p_{i,k}^\star \lambda_{i,k}^2 \right) \leq 0 \, \forall k, \tag{3.57}$$

$$R_k^\star - R_{\max,k} \leq 0, \quad \mathrm{qos}_k - R_k^\star \leq 0 \, \forall k, \tag{3.58}$$

$$\mu^\star \left(\sum_{k=1}^{K} \sum_{i=1}^{n_{S_k}} p_{i,k}^\star - P_T \right) = 0, \tag{3.59}$$

$$\gamma_k^\star \left(R_k^\star - \sum_{i=1}^{n_{S_k}} \log \left(1 + \frac{1}{\sigma^2} p_{i,k}^\star \lambda_{i,k}^2 \right) \right) = 0 \, \forall k, \tag{3.60}$$

$$\beta_k^\star \left(R_k^\star - R_{\max,k} \right) = 0 \, \forall k, \tag{3.61}$$

$$\nu_k^\star \left(\mathrm{qos}_k - R_k^\star \right) = 0 \, \forall k, \tag{3.62}$$

$$\mu^\star \geq 0, \quad \gamma_k^\star \geq 0, \quad \beta_k^\star \geq 0, \quad \nu_k^\star \geq 0 \, \forall k. \tag{3.63}$$

Given the Lagrange multipliers, the optimal power allocation for the j-th substream of the

k-th user is

$$p_{j,k}^{\star} = \left(\frac{1 + \nu_k^{\star} - \beta_k^{\star}}{\mu^{\star} \ln(2)} - \frac{\sigma^2}{\lambda_{j,k}^2} \right)^+ . \tag{3.64}$$

The optimum power allocation policy is obtained through water-filling, where the water-levels are user-dependent $\left(\frac{1+\nu_k^{\star}-\beta_k^{\star}}{\mu^{\star}\ln(2)} \text{ for user } k\right)$. For users for which neither the QoS constraints ($R_k > \text{qos}_k$) nor the energy constraints ($R_k < R_{\max,k}$) are active, we have $\nu_k^{\star} = 0$ and $\beta_k^{\star} = 0$, thus $p_{j,k}^{\star} = \left(\frac{1}{\mu^{\star}\ln(2)} - \frac{\sigma^2}{\lambda_{j,k}^2} \right)^+$, which is the classical water-filling policy where the water level is common to all users.

The proposed optimization algorithm can be described based on three steps:

1. Assign power to all users so that the QoS constraints ($C4$) are fulfilled with equality. This power is referred as $P_k^{\min}$ and is obtained by performing water-filling across the substreams of each user independently. The obtained water-level of each user is denoted by $c_{\min,k}$, i.e., $P_k^{\min} = \sum_{i=1}^{n_{S_k}} \left(c_{\min,k} - \frac{\sigma^2}{\lambda_{i,k}^2} \right)^+$ such that $R_k = \text{qos}_k$. At this stage, it is possible to obtain the feasibility of the problem by checking if $\sum_{k=1}^{K} P_k^{\min} \leq P_T$. If the inequality is not fulfilled, then the problem is not feasible. In that case, two options are considered: users should be dropped out from the allocation or QoS constraints should be relaxed. If $\sum_{k=1}^{K} P_k^{\min} = P_T$, then, the algorithm ends here and the optimum values are $p_{j,k}^{\star} = \left(c_{\min,k} - \frac{\sigma^2}{\lambda_{j,k}^2} \right)^+$ and $R_k^{\star} = \text{qos}_k$.
2. In the second step, we obtain the saturation water levels for all users. These saturation thresholds are the values that the water level of each user $\left(\frac{1-\beta_k^{\star}}{\mu^{\star}\ln(2)} \text{ for user } k\right)$ should reach to have $R_k^{\star} = R_{\max,k}$. These maximum thresholds for the water-levels are denoted as $c_{\max,k}$, and can be obtained by solving a water-filling for each independent user (see step 9 in Algorithm 3.2).
3. In the third step, we solve a water-filling procedure, but now jointly among all substreams of all users using a common water-level c. The minimum powers assigned in step 1 and the corresponding water-levels $c_{\min,k}$ must also be considered as starting points of the algorithm. At this step, whenever a user reaches its maximum data rate according to the energy constraint, i.e., $R_k = R_{\max,k}$, the scheduler drops out this user from the algorithm. The procedure ends when no more power is available or when all users are saturated.

A detailed description of the algorithm is presented in Algorithm 3.2.

3.B Proof of Proposition 3.3

In this appendix, we prove the lower bound on the mutual information when the transmitter has imperfect CSI. We extend the derivation developed in [Med00] that considered a single-user

SISO setup to our multiuser MIMO network. The mutual information for the signal model (3.30) can be expressed as

$$\mathcal{I}(\mathbf{x}_k;\mathbf{y}_k|\hat{\mathbf{H}}_k) = \mathcal{H}(\mathbf{x}_k|\hat{\mathbf{H}}_k) - \mathcal{H}(\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k) = \mathcal{H}(\mathbf{x}_k) - \mathcal{H}(\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k). \tag{3.65}$$

If we assume Gaussian codebooks, then the first entropy is simply $\mathcal{H}(\mathbf{x}_k) = \log\det(\pi\mathrm{e}\mathbf{Q}_{\mathbf{x}_k})$. Then, for the second term, $\mathcal{H}(\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k)$, we can upper-bound it by recalling that the Gaussian distribution is entropy-maximizing and, thus, it results in,

$$\mathcal{H}(\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k) \leq \mathbb{E}_{\mathbf{y}_k|\hat{\mathbf{H}}_k}\left[\log\det\left(\pi\mathrm{e}\mathbf{Q}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k}\right)\right], \tag{3.66}$$

where the covariance $\mathbf{Q}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k}$ is by definition,

$$\mathbf{Q}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k} = \mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k}\left[\left(\mathbf{x}_k - \mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k}[\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k]\right)\left(\mathbf{x}_k - \mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k}[\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k]\right)^H\right]. \tag{3.67}$$

The previous expression can be upper-bounded once more by considering, for example, a linear estimate of $\mathbf{x}_k$, such as the MMSE (see details in [Med00]). Then,

$$\mathbf{Q}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k} \preceq \mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k}\left[\left(\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\right)\left(\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\right)^H\right], \tag{3.68}$$

where $\mathbf{\Psi}_k$ is the matrix obtained by minimizing $\mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k}\left[\|\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\|^2\right]$. The expression of $\mathbf{\Psi}_k$ is given by $\mathbf{\Psi}_k = \mathbf{Q}_{\mathbf{x}_k,\mathbf{y}_k|\hat{\mathbf{H}}_k}\mathbf{Q}^{-1}_{\mathbf{y}_k|\hat{\mathbf{H}}_k}$, wherein the two covariance matrices are

$$\mathbf{Q}_{\mathbf{x}_k,\mathbf{y}_k|\hat{\mathbf{H}}_k} = \mathbf{Q}_{\mathbf{x}_k}\hat{\mathbf{H}}_k^H, \tag{3.69}$$

$$\mathbf{Q}_{\mathbf{y}_k|\hat{\mathbf{H}}_k} = \hat{\mathbf{H}}_k\mathbf{Q}_{\mathbf{x}_k}\hat{\mathbf{H}}_k^H + \mathbf{R}_k, \tag{3.70}$$

and $\mathbf{R}_k = \mathbb{E}_{\mathbf{n}_k}[\mathbf{n}_k\mathbf{n}_k^H] + \mathbb{E}_{\mathbf{y}_k|\hat{\mathbf{H}}_k}\left[\mathbf{H}_k^e\sum_{j=1}^K\mathbf{x}_j\mathbf{x}_j^H\mathbf{H}_k^{eH}\right] = \sigma^2\mathbf{I} + \mathbb{E}_{\mathbf{H}_k^e|\hat{\mathbf{H}}_k}\left[\mathbf{H}_k^e\sum_{j=1}^K\mathbf{Q}_{\mathbf{x}_j}\mathbf{H}_k^{eH}\right]$. Now, if we substitute (3.68) into (3.66), and then further upper-bound the entropy via Jensen's inequality by moving the operator $\mathbb{E}_{\mathbf{y}_k|\hat{\mathbf{H}}_k}$ into the concave $\log\det$ function, we obtain

$$\begin{aligned}\mathcal{H}(\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k) &\leq \mathbb{E}_{\mathbf{y}_k|\hat{\mathbf{H}}_k}\left[\log\det\left(\pi\mathrm{e}\mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k,\hat{\mathbf{H}}_k}\left[\left(\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\right)\left(\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\right)^H\right]\right)\right] \\ &\leq \log\det\left(\pi\mathrm{e}\mathbb{E}_{\mathbf{x}_k,\mathbf{y}_k|\hat{\mathbf{H}}_k}\left[\left(\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\right)\left(\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\right)^H\right]\right).\end{aligned} \tag{3.71}$$

The expectation inside the entropy can be computed as

$$\mathbb{E}_{\mathbf{x}_k,\mathbf{y}_k|\hat{\mathbf{H}}_k}\left[\left(\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\right)\left(\mathbf{x}_k - \mathbf{\Psi}_k\mathbf{y}_k\right)^H\right] = \mathbf{Q}_{\mathbf{x}_k} - \mathbf{Q}_{\mathbf{x}_k,\mathbf{y}_k|\hat{\mathbf{H}}_k}\mathbf{Q}^{-1}_{\mathbf{y}_k|\hat{\mathbf{H}}_k}\mathbf{Q}_{\mathbf{x}_k,\mathbf{y}_k|\hat{\mathbf{H}}_k}. \tag{3.72}$$

After inserting (3.72) into (3.71), the mutual information can be, finally, lower bounded as

$$\mathcal{I}(\mathbf{x}_k;\mathbf{y}_k|\hat{\mathbf{H}}_k) \geq \log\det(\pi\mathrm{e}\mathbf{Q}_{\mathbf{x}_k}) - \log\det\left(\pi\mathrm{e}\left(\mathbf{Q}_{\mathbf{x}_k} - \mathbf{Q}_{\mathbf{x}_k,\mathbf{y}_k|\hat{\mathbf{H}}_k}\mathbf{Q}^{-1}_{\mathbf{y}_k|\hat{\mathbf{H}}_k}\mathbf{Q}_{\mathbf{x}_k,\mathbf{y}_k|\hat{\mathbf{H}}_k}\right)\right) \tag{3.73}$$

$$= \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_k\mathbf{Q}_k\hat{\mathbf{H}}_k^H\mathbf{R}_k^{-1}\right) \triangleq \mathcal{R}_k. \tag{3.74}$$

3.C Proof of Proposition 3.4

Let us denote the set of saturated and non-saturated users by $\mathcal{S}$ and $\mathcal{N}$, respectively. We start the proof by assuming that, for the optimum transmit covariance matrices, we have that $\sum_{j=1}^{K} \mathrm{Tr}(\mathbf{Q}_j) < P_T$ and $\mathcal{N} \neq \emptyset$ (i.e., $R_k^\star < R_{\max,k}$ for some k) and $\mathcal{S} \neq \emptyset$, and check that this assumption leads to a contradiction. Let α be a positive constant, i.e., $\alpha > 1$ such that $\sum_{j=1}^{K} \mathrm{Tr}(\alpha\mathbf{Q}_j) = P_T$. By re-scaling all the transmit covariance matrices using this constant α, it is easy to check that all data-rates increase since

$$\begin{aligned} &\log\,\det\left(\mathbf{I} + \frac{\hat{\mathbf{H}}_k \alpha \mathbf{Q}_k \hat{\mathbf{H}}_k^H}{\sigma_{E_k}^2 \sum_j \mathrm{Tr}(\alpha\mathbf{Q}_j) + \sigma^2}\right) \\ &= \log\det\left(\mathbf{I} + \frac{\hat{\mathbf{H}}_k \mathbf{Q}_k \hat{\mathbf{H}}_k^H}{\sigma_{E_k}^2 \sum_j \mathrm{Tr}(\mathbf{Q}_j) + \sigma^2/\alpha}\right) > \log\det\left(\mathbf{I} + \frac{\hat{\mathbf{H}}_k \mathbf{Q}_k \hat{\mathbf{H}}_k^H}{\sigma_{E_k}^2 \sum_j \mathrm{Tr}(\mathbf{Q}_j) + \sigma^2}\right), \end{aligned} \tag{3.75}$$

is fulfilled. The problem is that by re-scaling all the matrices $\mathbf{Q}_j$ in this way, users belonging to $\mathcal{S}$ will violate the energy constraints, i.e., $R_k^\star > R_{\max,k}$ for $k \in \mathcal{S}$. Now, let us introduce a new set of variables $\{\beta_j\}$ to re-scale again each transmit covariance matrix (i.e., $\alpha\beta_j\mathbf{Q}_j$) such that $\alpha \sum_{j=1}^{K} \beta_j \mathrm{Tr}(\mathbf{Q}_j) = P_T$. Let us denote the achieved rate for user k with the new re-scaled as $f_k(\alpha, \{\beta_j\}) = \log\det\left(\mathbf{I} + \frac{\hat{\mathbf{H}}_k \alpha\beta_k \mathbf{Q}_k \hat{\mathbf{H}}_k^H}{\sigma_{E_k}^2 \alpha \sum_j \beta_j \mathrm{Tr}(\mathbf{Q}_j) + \sigma^2}\right)$. Notice that, in particular, $f_k(1, 1, \ldots, 1)$ corresponds to the initial situation before re-scaling the matrices $\mathbf{Q}_j$. Then, we seek to find a set of $\{\beta_j\}$ such that $f_k(\alpha, \{\beta_j\}) = R_{\max,k}$ for $k \in \mathcal{S}$ and $f_k(1, 1, \ldots, 1) \leq f_k(\alpha, \{\beta_j\}) \leq R_{\max,k}$ for $k \in \mathcal{N}$ (i.e., we increase the rates for the non-saturated users while we keep the maximum rates $R_{\max,k}$, $\forall k \in \mathcal{S}$ for the initially saturated users). Now, the question is whether we will be able to find these values of β_j fulfilling the previous relations. The answer is that, as long as at least one user remains unsaturated (i.e., at least for one $k \in \mathcal{N}$ user $f_k(\alpha, \{\beta_j\}) < R_{\max,k}$), there is available one degree of freedom (its own β) to answer the previous question positively.

Given that, the construction of the variables α and $\{\beta_j\}$ would be as follows (recall that the term $\alpha \sum_{j=1}^{K} \beta_j \mathrm{Tr}(\mathbf{Q}_j) = P_T$ remains constant by assumption):

1. We compute the value of α as $\alpha = \frac{P_T}{\sum_{j=1}^{K} \mathrm{Tr}(\mathbf{Q}_j)} > 1$.
2. For the saturated users, the β_i values can be computed as $\beta_i = \frac{1}{\alpha} < 1$, $\forall i \in \mathcal{S}$, making $f_i(\alpha, \{\beta_j\}) = R_{\max,i}$.
3. Finally, for the non-saturated users we can choose a set of β_k $\forall k \in \mathcal{N}$ such that $\alpha\beta_k \geq 1$ $\forall k \in \mathcal{N}$ increasing the data rates for these users, i.e., $f_k(1, 1, \ldots, 1) \leq f_k(\alpha, \{\beta_j\})$ $\forall k \in \mathcal{N}$.

Thus, if the original covariance matrices fulfilled $\sum_{j=1}^{K} \mathrm{Tr}(\mathbf{Q}_j) < P_T$, we have shown that it is possible to find another set of covariance matrices $\{\tilde{\mathbf{Q}}_k = \alpha\beta_k\mathbf{Q}_k\}$ such that $\sum_{j=1}^{K} \mathrm{Tr}(\tilde{\mathbf{Q}}_j) = P_T$

and the corresponding sum rate is higher while still fulfilling the constraints. This contradicts the initial assumption.

Using similar reasonings to the previous ones, it can be checked that the case of $\mathcal{S} = \emptyset$, $\mathcal{N} \neq \emptyset$, and $\sum_{j=1}^{K} \mathrm{Tr}(\mathbf{Q}_j) < P_T$ also leads to a contradiction, concluding the proof.

3.D Derivation of the Complexity of Matrix Operations

In this appendix we quantify the computational complexity of the suboptimal greedy algorithm presented in Table 3.4 and compare it with the brute-force approach. The complexity is measured in terms of number of flops. A flop is defined to be a real floating point operation [Gol96]. A real addition, multiplication, or division is counted as one flop. A complex addition and multiplication have two flops and six flops, respectively.

Complexity of Typical Matrix Operations

Let us first provide the flop count of several matrix operations that are used in the proposed algorithms. Let us define $\mathbf{W} \in \mathbb{C}^{m \times n}$, and $\mathbf{Z} \in \mathbb{C}^{n \times p}$. We assume that $m \leq n$ and $n \leq p$.

- The addition of two complex matrices $\mathbf{W} + \mathbf{W}$ takes mn complex additions, hence the flop count is $2mn$.
- The product of two complex matrices $\mathbf{WZ}$ needs $2mp(4n - 1)$ flops.
- The flop count for a SVD of a complex-valued matrix $\mathbf{W}$ is approximated as $24mn^2 + 48m^2n + 54m^3$ by treating every operation as a complex multiplication [Gol96].
- Water-filling over n eigenmodes takes $2n$ additions and $2n$ divisions. Hence, the flop count is[10] $4n$.

3.E Derivation of the Computational Complexity of the Proposed Scheduler

In this appendix, we evaluate the computation complexity of the proposed low-complexity scheduling procedure. For a description of the computational complexity of the matrix op-

[10] Obviously, the total flop count of the water-filling algorithm depends on the algorithm employed. For this computation, we have considered the hypothesis testing algorithm for the computation of the water-filling. Let us explain the details. Consider the power to be assigned to a given eigenmode to be $p_i = [c - \frac{1}{\lambda_i}]^+$ and the total power contraint $\sum_i p_i = P_T$, where $\frac{1}{\lambda_1} < \frac{1}{\lambda_2} < \ldots < \frac{1}{\lambda_n}$. The algorithm starts assuming that there is only one eigenmode active, thus, $c - \frac{1}{\lambda_1} = P_T \rightarrow c = P_T + \frac{1}{\lambda_1}$. Then, If $c - \frac{1}{\lambda_2} > 0$ there is at least two eigenmodes actives. Therefore, $c - \frac{1}{\lambda_1} + c - \frac{1}{\lambda_2} = P_T \rightarrow c = \frac{1}{2}(P_T + \frac{1}{\lambda_1} + \frac{1}{\lambda_2})$ and, then, procced iteratively.

erations involved in the procedure, please see Appendix 3.D. In the sequel, we present the detail of the complexity of the operations involved.

1. $j = 1$ (steps 1-4): we need to compute the SVD of $\mathbf{H}_k$, which has $24n_Rn_T^2 + 48n_R^2n_T + 54n_R^3$ flops. The water-filling algorithm used in step 2 of Algorithm 3.4 needs $6n_S + \frac{n_S^2}{2}$ flops for each user. In total, the first four steps of Algorithm 3.4 require as much as $K\left(24n_Rn_T^2 + 48n_R^2n_T + 54n_R^3 + 6n_S + \frac{n_S^2}{2}\right)$ flops.

2. $j \geq 2$ (steps 5-21):
   For each $i \in \mathcal{K}$ we need to compute an SVD of $\tilde{\mathbf{H}}_i$, so $24(j-1)n_Rn_T^2+48(j-1)^2n_R^2n_T+54(j-1)^3n_R^3$ flops are needed. To compute $\mathbf{G}_i = \mathbf{H}_i\tilde{\mathbf{V}}_i^{(0)} \in \mathbb{C}^{n_R\times(n_T-n_R(j-1))}$, the complexity of this product is $2n_R(n_T - n_R(j-1))(2n_T-1)$. SVD of $\mathbf{G}_i$ introduces $24n_R(n_T - n_R(j-1))^2 + 48n_R^2(n_T - n_R(j-1)) + 54n_R^3$. The computation of water-filling involves $2j^2n_R + j(\frac{n_R^2}{2} + 4n_R)$ flops.

Therefore, the total flop count of the suboptimal user selection algorithm is

$$\begin{aligned}\psi = \sum_{j=2}^{\hat{K}} \Big[&24j(j-1)n_Rn_T^2 + 48j(j-1)^2n_R^2n_T + 54j(j-1)^3n_R^3 + 24jn_R(n_T - n_R(j-1))^2 \\ &+ 48jn_R^2(n_T - n_R(j-1)) + 54jn_R^3 + 4jn_Rn_T^2 - 2jn_Rn_T + 2j^2n_R + j\left(n_R^2/2 + 4n_R\right)\Big] \\ &\times (K-j+1) + K\left(24n_Rn_T^2 + 48n_R^2n_T + 54n_R^3 + 6n_S + n_S^2/2\right) \qquad (3.76)\\ \overset{(a)}{<} \sum_{j=2}^{\hat{K}} \Big[&(24j(j-1)+28j)n_Rn_T^2 - 2n_Rn_T + 2j^2n_R + (48j(j-1)^2 + 48j)n_R^2n_T \qquad (3.77)\\ &+ (54j(j-1)^3 + 54j)n_R^3 + j\left(n_R^2/2 + 4n_R\right)\Big] \cdot (K-j+1) + C \overset{(b)}{\approx} \sum_{j=2}^{\hat{K}} 54j^4n_R^3(K-j) \\ \overset{(c)}{\approx} &\mathcal{O}\left(K\hat{K}n_T^3\right) \qquad (3.78)\end{aligned}$$

where the inequality in (a) is due to the upper bound of $(n_T - n_R(j-1))$ by n_T and C is a constant term, the approximation in (b) only considers the highest computational complexity operations, and finally in (c) we have considered that $K \gg \hat{K}$.

3.F Derivation of the Computational Complexity of the Optimum Scheduler

Recall that $\mathcal{A}$ is the set containing all possible user grouping $\mathcal{A}_i$, i.e., $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \dots\}$. Then, the optimum scheduler must conduct a complete search among all possible user subsets contained in $\mathcal{A}$. The cardinality of $\mathcal{A}$ is $|\mathcal{A}| = \sum_{i=1}^{\hat{K}} \mathcal{C}_i^K = \sum_{i=1}^{\hat{K}} \binom{K}{i}$. For a description of the computational

complexity of the matrix operations involved in the procedure, please see Appendix 3.D. Given that, the computational complexity of this method is given by

$$\begin{aligned}\psi = \sum_{j=1}^{\hat{K}} \binom{K}{j} \Big[& 24j(j-1)n_R n_T^2 + 48j(j-1)^2 n_R^2 n_T + 54j(j-1)^3 n_R^3 + 24jn_R \\ & \times (n_T - n_R(j-1))^2 + 48jn_R^2(n_T - n_R(j-1)) + 54jn_R^3 + 4jn_R n_T^2 - 2jn_R n_T \\ & + 2j^2 n_R + j\left(n_R^2/2 + 4n_R\right) \Big] \qquad (3.79) \\ \overset{(a)}{>} \binom{K}{\hat{K}} \Big[& 24\hat{K}(\hat{K}-1)n_R n_T^2 + 48\hat{K}(\hat{K}-1)^2 n_R^2 n_T + 54\hat{K}(\hat{K}-1)^3 n_R^3 + 24\hat{K}n_R \\ & \times (n_T - n_R(\hat{K}-1))^2 + 48\hat{K}n_R^2(n_T - n_R(\hat{K}-1)) + 54\hat{K}n_R^3 + 4\hat{K}n_R n_T^2 \\ & - 2\hat{K}n_R n_T + 2\hat{K}^2 n_R + \hat{K}\left(n_R^2/2 + 4n_R\right) \Big] \qquad (3.80) \\ \approx \binom{K}{\hat{K}} & 54\hat{K}(\hat{K}-1)^3 n_R^3 \approx \mathcal{O}\left(\binom{K}{\hat{K}} \hat{K} n_T^3\right), \qquad (3.81)\end{aligned}$$

where in (a) we only pick $\hat{K}$ users out of K.

Chapter 4

User Grouping and Resource Allocation Strategies in Multiuser MIMO SWIPT Networks

4.1 Introduction

Traditionally, energy harvesting techniques have been developed based on energy sources such as wind or solar energy [Par05], [Sud11]. Nevertheless, RF signals could also be used as a source for energy scavenging. Unfortunately, some measurements in today's urban landscape show that the actual strength of the received electric field is not high and, thus, the proximity to the transmitter is important [Par05]. In this sense, it is important to emphasize that the newer applications require higher data rates and that this implies that efficient network deployments with more capacity must be considered. In the last years, this increase in capacity efficiency has been shown to be achieved through the deployment of short-distance networks[1] (e.g. femtocells [Cha08]). As we mentioned before in the introduction of this thesis, in Section 1.1.1, the use of shorter distances in this kind of networks allows increasing the received power levels and, consequently, to make mobile terminals able to harvest power from the received radio signals when they are not detecting information data. This is commonly named as *wireless power transfer* (see [Lu15] for an extensive review).

4.1.1 Related Work

The concept of SWIPT was first studied from a theoretical point of view by Varshney [Var08]. He showed that, for the single-antenna AWGN channel, there exists a nontrivial trade-off in maximizing the data rate versus the power transmission. In [Zha13], authors considered a MIMO scenario with one transmitter capable of transmitting information and power simultaneously to one receiver. Later, [Rub13] (the work presented in this chapter) extended the work in [Zha13]

[1] By short-distance network we mean a short distance between the transmitter and the receivers.

by considering that multiple users were present in the broadcast MIMO system. However, the multi-stream transmit covariance optimization that arises in broadcast MIMO systems is a very difficult nonconvex optimization problem. In order to overcome that difficulty, in the first part of this chapter we have considered a BD strategy [Spe04], in which interference is pre-canceled at the transmitter. The BD technique allows for a simple solution but wastes some degrees of freedom and, thus, degrades the overall performance. Works [Par13] and [Par14] considered a MIMO network consisting of multiple transmitter-receiver pairs with co-channel interference. The study in [Par13] focused on the case with two transmitter-receiver pairs whereas in [Par14], the authors generalized [Par13] by considering that k transmitter-receivers pairs were present. The work in [Zon16] considered a MIMO system with single-stream transmission. In contrast to previous works where the system rate was optimized, the objective was to minimize the overall power consumption with per-user signal to interference and noise ratio (SINR) constraints and harvesting constraints. The design of multiuser broadcast networks under the framework of MISO beamformimg optimization has also been addressed in works such as [Xu13] and [Shi14].

In this chapter, we assume that the CSI is known at the transmitter, but some works can be referenced in which techniques for optimizing the training under the SWIPT framework are proposed. This is the case of [Zen15], in which authors studied the design of an efficient channel acquisition method for a point-to-point MIMO SWIPT system by exploiting the channel reciprocity. Additionally, a worst-case robust beamforming design was proposed in [Xia12], in which imperfect CSI at the transmitter was assumed.

Apart from designing the transmit covariance matrices, we also develop some user grouping techniques in which, from frame to frame, it is decided what users will receive information data and what users will harvest energy from RF signals. In this context, there are some works in the literature that deal with user scheduling in the SWIPT framework but they consider a SISO system. Therefore, the scheduling presented in those papers is purely temporal scheduling of users. Among those works, [Liu13b] introduced time scheduling between information and energy transfer and derived the optimal switching policy considering time-varying co-channel interference. The receiver, thus, replenished the battery opportunistically via wireless power transfer from the unintended interference and/or the intended signal sent by the transmitter. Then, in [Mor15] authors studied DL multiuser scheduling for a time-slotted system with SWIPT. In particular, in each time slot, a single user is scheduled to receive information, whereas the remaining users opportunistically harvest energy from ambient signals. Finally, in [Din14] authors considered a multiuser cooperative network where M source-destination SISO pairs communicate with each other via a relay that is energy-constrained and is provided with an energy harvesting source. The key idea is to select a subset of those M pairs to communicate through the relay.

4.1.2 Main Contribution

Within the state of the art described above, the main difference of the work presented in this chapter w.r.t. the previous works is that we assume a broadcast multiuser multi-stream MIMO SWIPT network, which is a scenario not considered before.

In the first part of the chapter, we consider that BD is used and develop the optimal transmit precoding matrices for this setup, considering explicitly some power consumption sources in the problem formulation. We model the receivers as battery-limited nodes and, thus, they need to recharge the batteries to prolong their network lifetimes. In the multiuser MIMO SWIPT framework there are two groups of users to be served: one for power harvesting and another one for information reception. So far in the literature of MIMO SWIPT techniques, authors have considered that these two sets of users were predefined and fixed. In this chapter, we propose some user grouping techniques that work on a per frame basis in order to maximize the system throughput and/or fairness among users. Later, we provide a procedure that manages and configures the minimum energy a given user should harvest considering the impact of imposing those constraints in the system performance.

In the second part of the chapter, we model our transmitter design as a multi-objective problem in which the scenarios studied in [Zha13] and [Rub13] are shown to be just particular solutions. We assume that interference is not pre-canceled (i.e., the BD approach is not applied) and, thus, both larger information transfer and harvested power can be achieved simultaneously. The resulting problem is nonconvex and very difficult to solve. In order to obtain locally optimal solutions, we derive different methods based on the MM approach. By means of this strategy, we are able to reformulate our original nonconvex problem into a series of convex subproblems that are easily solved (i.e., either analytically or through algorithms that have a very low computational complexity) and whose solutions converge to a locally optimal solution of the original nonconvex problem.

Compared to the works mentioned in the previous section devoted to the state of the art, the main contributions of this chapter can be highlighted as:

- We consider a multiuser multi-stream MIMO broadcast transmission strategy in which both the transmitter and the receivers are provided with multiple antennas. The system weighted sum rate with individual per-user harvesting constraints are considered in the proposed transmission strategy design. We also take into account the state of the batteries of the terminals in the proposed strategy. We study particular cases in which only information and only harvesting users are present in the system.

- We develop an efficient algorithm that computes the optimal precoding matrices for the multiuser MIMO broadcast network setup mentioned previously.

- The fundamental (multidimensional) trade-off between system performance and (per-user) harvested energy is studied and characterized, putting emphasis and giving specific closed-form expressions for some particular cases of interest.

- We incorporate power consumption models at the transmitter and the receivers. In particular, we consider the decoding power consumption at the receivers and its impact on the system performance.

- We develop user grouping schemes that employ a two-stage user scheduling mechanism that run at different time scales. In the first stage, a subset of users are grouped to be candidates of information reception and a subset of users are grouped to be candidates of harvesting. Out of these selected users, in the second stage, we perform the final user information and harvesting selection and grouping, with the aim of enhancing the system throughput and/or fairness among users.

- We provide a procedure that manages and configures the minimum amount of energy that a given user should harvest from ambient RF signals, considering the impact of imposing those constraints in the system performance.

- Finally, we generalize the previous transmit covariance design strategy and assume that BD is not applied. In this sense, we develop advanced optimization techniques to solve the nonconvex problem that arises when optimizing the transmit covariance matrices under the SWIPT setup.

4.1.3 Organization of the Chapter

The remainder of this chapter is organized as follows. The system model is presented in Section 4.2. Section 4.3 covers the different proposed techniques considering that BD is applied. In particular, in Section 4.3.1 we present the general problem formulation. Section 4.3.2 covers the transmit covariance matrix design for simultaneous data and power transmission. In Section 4.3.3, we present a scheduling mechanism to decide which users should be scheduled in each particular user set. Later, in Section 4.3.6 we present a methodology to manage the amount of harvested energy and the impact on the system performance. Then, Section 4.4 addresses the second part of the chapter, in which we design the nonconvex transmit covariance matrices based on the MM approach, considering that interference is present in the system. Section 4.4.3 presents some numerical results comparing the performance of the strategies presented in Sections 4.3.2 and 4.4. Finally, conclusions are drawn in Section 4.5.

4.2 System Model

4.2.1 Signal Model

We consider a wireless broadcast system consisting of one BS transmitter equipped with n_T antennas and a set of K receivers, denoted as $\mathcal{U}_T = \{1, 2, \ldots, K\}$, where the k-th receiver is equipped with n_{R_k} antennas as depicted in Figure 4.1.

We index frames by $t \in \mathcal{T} \triangleq \{1, \ldots, T\}$ with a duration of T_f seconds each. We assume block fading channels, that is, the channels remain constant within a frame but change from frame to frame. The equivalent baseband channel from the BS to the k-th receiver is denoted by $\mathbf{H}_k(t) \in \mathbb{C}^{n_{R_k} \times n_T}$. It is also assumed that the set of matrices $\{\mathbf{H}_k(t)\}$ is known to the BS and to the corresponding receivers. The case of imperfect CSI is out of the scope of the chapter.

The set of users is partitioned into two subsets as commented in the introduction. One of the sets contains the users that receive information, denoted as $\mathcal{U}_I(t) \subseteq \mathcal{U}_T$ with $|\mathcal{U}_I(t)| = N$, and the other set, $\mathcal{U}_E(t) \subseteq \mathcal{U}_T$ with $|\mathcal{U}_E(t)| = M$, contains the users that harvest energy from the power radiated by the BS that is used to transmit signals to the information receivers. We assume that a given user is not able to decode information and to harvest energy simultaneously. This forces a user to either receive information or harvest energy during the whole frame, i.e., during the scheduling period, which is a reasonable choice if the scheduling periods are short. That translates into disjoint subsets[2], i.e., $\mathcal{U}_I(t) \cap \mathcal{U}_E(t) = \emptyset$, $|\mathcal{U}_I(t)| + |\mathcal{U}_E(t)| \leq K$. Note that the previous sets depend on t as the specific users in each of them may change from frame to frame. The number of users in each set, N and M, may also change from frame to frame as it will be explained later in the chapter. To simplify the notation when needed, we will assume that the indexing of the users is such that[3] $\mathcal{U}_I(t) = \{1, 2, \ldots, N\}$ and $\mathcal{U}_E(t) = \{N+1, N+2, \ldots, N+M\}$. We will assume that $n_T > n_R - \min_k\{n_{R_k}\}$ is fulfilled, being[4] $n_R = \sum_{k \in \mathcal{U}_I} n_{R_k}$.

As far as the signal model is concerned, the received signal for the i-th information receiver at the n-th time instant within the t-th frame can be modeled as

$$\mathbf{y}_i(n,t) = \mathbf{H}_i(t)\mathbf{B}_i(t)\mathbf{x}_i(n,t) + \mathbf{H}_i(t) \sum_{\substack{k \in \mathcal{U}_I(t) \\ k \neq i}} \mathbf{B}_k(t)\mathbf{x}_k(n,t) + \mathbf{n}_i(n,t) \in \mathbb{C}^{n_{R_i} \times 1}, \qquad \forall i \in \mathcal{U}_I(t). \tag{4.1}$$

In the previous notation, $\mathbf{B}_i(t)\mathbf{x}_i(n,t)$ represents the transmitted signal for user $i \in \mathcal{U}_I(t)$, where $\mathbf{B}_i(t) \in \mathbb{C}^{n_T \times n_{S_i}}$ is the precoder matrix and $\mathbf{x}_i(n,t) \in \mathbb{C}^{n_{S_i} \times 1}$ represents the information sym-

[2] Let us assume for the moment that not all users must be in any group. As will be shown later, some of the users may not be selected for any group in a given scheduling period.

[3] At the beginning of each frame, once the groups have been decided, the users are indexed again in such a way that the first N users are information users and the following M users are harvesting users.

[4] This assumption corresponds to a necessary constraint to be applied when BD is used [Spe04], as will be explained later with more detail in Section 4.3.2.

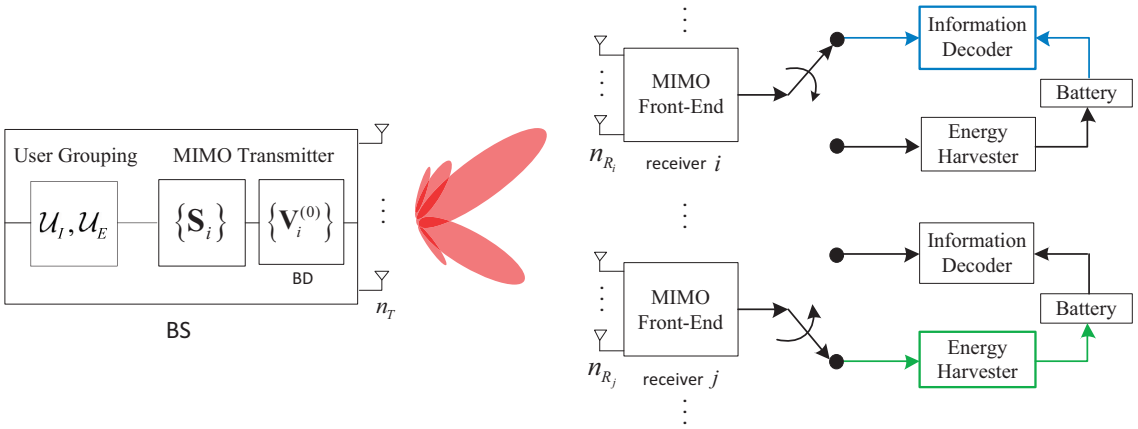


Figure 4.1: Schematic representation of the DL broadcast multiuser communication system. Note that each user can switch from being an information user to being an energy harvester user. Note also that BD is applied at the transmitter. This is the transmitter architecture used in the first part of this chapter. In the second part of the chapter BD will not be applied and, hence, the BS will be composed of user grouping and precoding stages without precancelation matrices $\{\tilde{\mathbf{V}}_i^{(0)}\}$.

bol vector. n_{S_i} denotes the number of streams assigned to user $i \in \mathcal{U}_I(t)$ and, throughout the first part of the chapter, we assume that $n_{S_i} = \min\{n_{R_i}, n_T - (n_R - n_{R_i})\}\, \forall i \in \mathcal{U}_I(t)$ is fulfilled[5]. The transmit covariance matrix is $\mathbf{S}_i(t) = \mathbf{B}_i(t)\mathbf{B}_i^H(t)$ if we assume w.l.o.g. that $\mathbb{E}\left[\mathbf{x}_i(n,t)\mathbf{x}_i^H(n,t)\right] = \mathbf{I}_{n_{S_i}}$. $\mathbf{n}_i(n,t) \in \mathbb{C}^{n_{R_i}\times 1}$ denotes the receiver noise vector, which is considered white and Gaussian with[6] $\mathbb{E}\left[\mathbf{n}_i(n,t)\mathbf{n}_i^H(n,t)\right] = \mathbf{I}_{n_{R_i}}$. Note that the middle term of (4.1) is an interference term that it is usually known as MUI and whose covariance matrix is written as

$$\mathbf{\Omega}_i(\mathbf{S}_{-i}(t))(t) = \mathbf{H}_i(t)\mathbf{S}_{-i}(t)\mathbf{H}_i^H(t) + \mathbf{I}, \quad \forall i \in \mathcal{U}_I(t), \tag{4.2}$$

where $\mathbf{S}_{-i}(t) = \sum_{\substack{k\in\mathcal{U}_I(t)\\k\neq i}} \mathbf{S}_k(t)$. Let $\tilde{\mathbf{x}}(n,t) = \mathbf{B}(t)\mathbf{x}(n,t)$ denote the signal vector transmitted by the BS, where the joint precoding matrix is defined as $\mathbf{B}(t) = [\mathbf{B}_1(t), \ldots, \mathbf{B}_N(t)] \in \mathbb{C}^{n_T\times n_S}$, being $n_S = \sum_{i=1}^N n_{S_i}$ the total number of streams of all information users, and the data vector as $\mathbf{x}(n,t) = \left[\mathbf{x}_1^T(n,t), \ldots, \mathbf{x}_N^T(n,t)\right]^T \in \mathbb{C}^{n_S\times 1}$. $\tilde{\mathbf{x}}(n,t)$ must satisfy the power constraint formulated as $\mathbb{E}[\|\tilde{\mathbf{x}}(n,t)\|^2] = \sum_{i=1}^N \mathrm{Tr}(\mathbf{S}_i(t)) \leq P_T$, where P_T represents the total radiated power at the BS, assuming that the information symbols of different users are independent and zero-mean.

Let us model the total power harvested by the j-th user during the t-th frame, denoted by $\bar{Q}_j(t)$, from all receiving antennas to be proportional to that of the equivalent baseband signal, i.e.,

$$\bar{Q}_j(t) = \zeta_j \mathbb{E}\left[\left\|\mathbf{H}_j(t)\sum_{i\in\mathcal{U}_I(t)}\mathbf{B}_i(t)\mathbf{x}_i(n,t)\right\|^2\right] \tag{4.3}$$

$$= \zeta_j \sum_{i\in\mathcal{U}_I(t)} \mathbb{E}[\|\mathbf{H}_j(t)\mathbf{B}_i(t)\mathbf{x}_i(n,t)\|^2], \quad \forall j \in \mathcal{U}_E(t), \tag{4.4}$$

[5]In fact $\min\{n_{R_i}, n_T - (n_R - n_{R_i})\}$ is an upper bound for the actual number of active streams. Such number will be obtained from the solution of the corresponding optimization problems presented in this chapter.

[6]We assume that noise power $\sigma^2 = 1$ w.l.o.g., otherwise we could simply apply a scale factor at the receiver and re-scale the channels accordingly.

where ζ_j is a constant that accounts for the loss in the energy transducer for converting the harvested power to electrical power to charge the battery. Notice that, for simplicity, in (4.4) we have omitted the harvested power due to the noise term or other external RF sources since they can be assumed negligible. Based on the previous assumptions, (4.4) can be written as

$$\bar{Q}_j(t) = \zeta_j \sum_{i \in \mathcal{U}_I(t)} \mathrm{Tr}(\mathbf{H}_j(t)\mathbf{S}_i(t)\mathbf{H}_j^H(t)), \quad \forall j \in \mathcal{U}_E(t). \tag{4.5}$$

For the sake of clarity, we will drop the time and frame dependence whenever possible.

4.2.2 Power Consumption Models

The power consumption models considered in this chapter are the same as the ones assumed in Chapter 3 but, for the sake of independence of this chapter, we provide a description of them in the sequel. The energy consumed by the transceiver can be modeled as the energy consumed by the front-end plus the energy consumed by the coding/decoding stages (omitting for the moment the power radiated by the transmitter). Although other works consider battery imperfections in their models [Dev12b], we do not consider this in our work for the sake of simplicity. Note, however, that the strategy and formulation presented in this chapter could be extended easily to incorporate those imperfections. In the following we will comment briefly which is the generic abstract approach followed in this chapter in order to make the proposed strategies independent of the concrete model.

1. *Front-end Consumption:* as far as the transmitter is concerned, the components that consume energy are the HPA, the mixers, the filters, and other elements of the RF chain. Concerning the receiver, the front-end consumption usually depends on the condition on the channel, i.e., the SNR (in practice, the receiver should adapt the front-end according to the received power [Jen12], an operation that requires some additional power). In the following, however, we assume that the component of the receiver front-end consumption that depends on the SNR is negligible as it can be concluded from experimental measurements and is adopted in most works [Jen12]. We denote the energy consumed by the front-end at the transmitter and the receiver by P_c^{tx} and P_c^{rx}, respectively.

2. *Coding/Decoding Consumption:* it is reasonable to consider the energy consumed by the coding stage at the transmitter negligible compared to the energy consumed by the front-end. This is illustrated and commented in papers such as [Aue11]. For this reason, we will not include coding consumption in our models. On the other hand, the decoding consumption must be included in the models since, as shown in [Gro11], [Ros10], such energy consumption is not negligible and can affect importantly the lifetime of the mobile terminal. There is a consensus about the fact that the decoding consumption, denoted by $P_{\mathrm{dec},i}(R_i(t))$, increases with the data rate $R_i(t)$. In Chapter 3 of this dissertation, we

presented different models for $P_{\text{dec},i}(R_i(t))$ but, for the sake of generality, we will consider it as a general function in this chapter.

Given the previous models, the total consumption at the transmitter (omitting for the moment the radiated power) just includes the front-end consumption as mentioned previously and, thus, it is denoted as

$$P_{\text{tot}}^{t_x} = P_c^{t_x}. \tag{4.6}$$

On the other hand, the total power consumption at the i-th receiver is expressed as

$$P_{\text{tot},i}^{r_x}(R_i(t)) = P_{\text{dec},i}(R_i(t)) + P_c^{r_x}. \tag{4.7}$$

Notice that the power consumption at the receiver is limited by the current battery level. According to this, the data rate of a given information user (user i) during one frame must be constrained in order not to consume more energy when decoding than the current energy available at the battery $C_i(t)$. Hence,

$$T_f(P_{\text{dec},i}(R_i(t)) + P_c^{r_x}) \leq C_i(t), \tag{4.8}$$

which can be written in terms of a maximum rate constraint as

$$R_i(t) \leq R_{\text{max},i}(C_i(t)), \tag{4.9}$$

where $R_{\text{max},i}(C_i(t)) = P_{\text{dec},i}^{-1}\left(\frac{C_i(t)}{T_f} - P_c^{r_x}\right)$.

4.2.3 Battery Dynamics

We consider that each user terminal is provided with a finite battery capacity whose level decreases accordingly when the user receives and decodes data. The terminals are also able to recharge their batteries by means of collecting power dynamically coming from the BS.

The battery at the beginning of the t-th frame of the i-th information user served with a data rate $R_i(t-1)$ during the previous frame is denoted as

$$C_i(t) = \Big(C_i(t-1) - T_f P_{\text{tot},i}^{r_x}(R_i(t-1))\Big)_0^{C_{\text{max}}^i}, \quad \forall i \in \mathcal{U}_I(t), \tag{4.10}$$

where C_{max}^i is the maximum battery level, and the function $P_{\text{tot,i}}^{r_x}(R_i(t-1))$ was defined in (4.7). Note that $C_i(t)$ has units of Joules.

On the other hand, the battery at the beginning of the t-th frame of the j-th harvesting

user is denoted as

$$C_j(t) = \left(C_j(t-1) + T_f \bar{Q}_j(t-1) - T_f P_c^{rx}\right)_0^{C_{\max}^j}, \quad \forall j \in \mathcal{U}_E(t), \tag{4.11}$$

where $\bar{Q}_j(t-1)$ is the power harvested during the frame $t-1$.

The receivers must inform the BS about their battery level status in order to take decisions whether to serve that user with information or with RF power. In this chapter, we assume that the feedback channel is ideal and not limited in rate.

4.3 Part I: BD-Based Transmit Covariance Optimization Techniques

4.3.1 Joint Resource Allocation and User Grouping Formulation

In this section, we formulate the joint design of the covariance matrices $\mathbf{S}_i(t)$, the data rates $R_i(t)$, and the user grouping $\mathcal{U}_I(t), \mathcal{U}_E(t)$, based on the maximization of the weighted sum rate with individual power harvesting constraints for all time instants $t \in \mathcal{T}$. Given this, the problem is formulated through the following optimization problem:

$$\begin{aligned}
\underset{\substack{\{R_i(t),\mathbf{S}_i(t)\}_{\forall i\in\mathcal{U}_I(t)},\\ \mathcal{U}_I(t),\mathcal{U}_E(t)}}{\text{maximize}} \quad & \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{U}_I(t)} \omega_i(t) R_i(t) && && (4.12)\\
\text{subject to} \quad & C1: \sum_{i\in\mathcal{U}_I(t)} \mathrm{Tr}(\mathbf{H}_j(t)\mathbf{S}_i(t)\mathbf{H}_j^H(t)) \geq Q_j, && \forall j \in \mathcal{U}_E(t),\ \forall t \in \mathcal{T}\\
& C2: \sum_{i\in\mathcal{U}_I(t)} \mathrm{Tr}(\mathbf{S}_i(t)) + P_c^{tx} \leq P_{\max}, && \forall t \in \mathcal{T}\\
& C3: R_i(t) \leq \log\det\left(\mathbf{I} + \mathbf{H}_i(t)\mathbf{S}_i(t)\mathbf{H}_i^H(t)\right), && \forall i \in \mathcal{U}_I(t),\ \forall t \in \mathcal{T}\\
& C4: R_i(t) \leq R_{\max,i}(C_i(t)), && \forall i \in \mathcal{U}_I(t),\ \forall t \in \mathcal{T}\\
& C5: \mathbf{H}_k(t)\mathbf{S}_i(t)\mathbf{H}_k^H(t) = 0, \quad \forall k \neq i,\ k,i \in \mathcal{U}_I(t), && \forall t \in \mathcal{T}\\
& C6: \mathbf{S}_i(t) \succeq 0, && \forall i \in \mathcal{U}_I(t),\ \forall t \in \mathcal{T}\\
& C7: C_i(t) = \left(C_i(t-1) - T_f P_{\mathrm{tot},i}^{rx}(R_i(t-1))\right)_0^{C_{\max}^i}, && \forall i \in \mathcal{U}_I(t),\ \forall t \in \mathcal{T}\\
& C8: C_j(t) = \left(C_j(t-1) + T_f \bar{Q}_j(t-1) - T_f P_c^{rx}\right)_0^{C_{\max}^j}, && \forall j \in \mathcal{U}_E(t),\ \forall t \in \mathcal{T},
\end{aligned}$$

where the weights $\omega_i(t) \geq 0$ can be set to assign priorities to achieve fairness among users[7], $R_i(t) \leq \log\det\left(\mathbf{I} + \mathbf{H}_i(t)\mathbf{S}_i(t)\mathbf{H}_i^H(t)\right)$ denotes the achievable data rate of the i-th user when

[7] A further discussion on how the weights $\omega_i(t) \geq 0$ can be set to provide fairness will be introduced later in Section 4.3.3.

considering linear precoding following a BD strategy[8] [Spe04], $Q_j = \frac{\bar{Q}_j^{\min}}{\zeta_j}$, being $\{\bar{Q}_j^{\min}\}$ the set of minimum power harvesting constraints that can be freely chosen, and $P_{\max}$ is the available power at the BS. In fact, BD is applied through constraint $C5$, which forces the complete cancellation of the MUI making the whole problem more tractable (as it will be shown later in the chapter). Notice that constraint $C1$ is associated with the minimum power to be harvested for a given user. In case that another external energy harvesting source was available and the amount to be harvested could be estimated (or was fully known in advanced), we could subtract such value from Q_j accordingly. Constraint $C4$ assures that the information users do not spent more energy decoding the message than the current energy available at the battery.

Two main difficulties arise when attempting to solve (4.12). First, note that the solution for all time instants has to be found jointly. The reason is that resource allocation decisions at frame t have an impact not only on that frame, but on future frames as well. Some researchers have attempted to solve harvesting (time-coupled) problems by assuming that the whole channel and harvesting realizations are known a-priori, giving rise to off-line approaches that are not implementable in real scenarios [Ho12a], [Yan12a]. As we assume that only causal knowledge of the channel and the harvesting is available, we would have to resort to dynamic programming (DP) techniques [Ber05] in order to find the optimal solution of problem (4.12). However, these techniques usually require the implementation of extremely high complexity algorithms that are impractical in scenarios where the set of variables to be optimized is large and, thus, DP techniques have been applied only in cases where the optimization variables are scalars [Bla13], [LR14]. The second difficulty that we find is that the user grouping must also be optimized jointly with the covariance matrices and the data rates. The user grouping variables are discrete and, hence, the problem becomes combinatorial. The optimum solution has to be found by applying some sort of combinatorial search among all possible user groups, increasing the overall complexity exponentially.

Because we are interested in low-complexity solutions, we have to make some simplifications to problem (4.12) to make it more tractable, with the hope of finding a good suboptimum solution that is close to the global optimum solution of problem (4.12).

The first assumption that we consider is to decouple the problem in time and propose a separate per-frame optimization approach. With this approach, we solve the optimization problem at the beginning of each frame t, taking decisions based on current and past information of the battery levels. The optimization to solve is (we omit the time dependence for the sake of simplicity in the notation even though, all these variables, including the information and

[8]The optimum transmission policy in a MIMO broadcast channel is the well-known non-linear dirty paper coding strategy [Gol03]. Nevertheless, such design strategy is highly computational demanding and cannot be computed in real time. Instead, much simpler linear transceiver designs have shown to achieve almost the same capacity using much lower computational resources [Lee06]. Thus, for simplicity in the transmitter design, in the following we force the precoder to be linear.

harvesting users sets $\mathcal{U}_I$ and $\mathcal{U}_E$, change at each frame):

$$\begin{aligned}
\underset{\substack{\{R_i,\mathbf{S}_i\}_{\forall i\in\mathcal{U}_I},\\ \mathcal{U}_I,\mathcal{U}_E}}{\text{maximize}} \quad & \sum_{i\in\mathcal{U}_I}\omega_i R_i && && (4.13)\\
\text{subject to} \quad & C1: \sum_{i\in\mathcal{U}_I}\text{Tr}(\mathbf{H}_j\mathbf{S}_i\mathbf{H}_j^H)\geq Q_j, && \forall j\in\mathcal{U}_E\\
& C2: \sum_{i\in\mathcal{U}_I}\text{Tr}(\mathbf{S}_i)+P_c^{tx}\leq P_{\max}\\
& C3: R_i\leq \log\det\left(\mathbf{I}+\mathbf{H}_i\mathbf{S}_i\mathbf{H}_i^H\right), && \forall i\in\mathcal{U}_I\\
& C4: R_i\leq R_{\max,i}(C_i), && \forall i\in\mathcal{U}_I\\
& C5: \mathbf{H}_k\mathbf{S}_i\mathbf{H}_k^H=0, && \forall k\neq i,\ \ k,i\in\mathcal{U}_I\\
& C6: \mathbf{S}_i\succeq 0, && \forall i\in\mathcal{U}_I.
\end{aligned}$$

Problem (4.13) is still very difficult to solve as it involves continuous as well as integer variables. Note that for a fixed set of groups, $\mathcal{U}_I$ and $\mathcal{U}_E$, problem (4.13) is convex w.r.t. $\{R_i,\mathbf{S}_i\}$ and can be solved using standard optimization techniques. The optimum solution can be found by solving problem (4.13) for all possible combinations of user groups, that is, an exhaustive search should be implemented. Consider for example that $|\mathcal{U}_I|=4$ and $|\mathcal{U}_E|=4$, and that $K=10$. Then, problem (4.13) (for a fixed $\mathcal{U}_I$ and $\mathcal{U}_E$) should be solved $\frac{K!}{|\mathcal{U}_I|!|\mathcal{U}_E|!(K-|\mathcal{U}_I|-|\mathcal{U}_E|)!}=3,150$ times. Clearly, the optimum solution is impractical even for a system with a small number of users. In that sense, any technique aside from the exhaustive search may be suboptimal.

This fact motivates our second simplification: we decouple the decision of resource allocation and user grouping and propose a two-stage design strategy in which the user grouping is found based on suboptimal but less complex techniques. In other works, at the beginning of each frame, first, we find the user groups $\mathcal{U}_I$ and $\mathcal{U}_E$, and, then, for that fixed user groups, we solve the following convex optimization problem:

$$\begin{aligned}
\underset{\{R_i,\mathbf{S}_i\}_{\forall i\in\mathcal{U}_I}}{\text{maximize}} \quad & \sum_{i\in\mathcal{U}_I}\omega_i R_i && (4.14)\\
\text{subject to} \quad & C1\ldots C6 \text{ of problem (4.13)}.
\end{aligned}$$

Note that, thanks to $C5$ problem (4.14) is convex; otherwise, the objective function, i.e., the weighted sum rate, would not be convex due to the MUI. In the next section, we are going to present a method to solve problem (4.14) for different settings. Later, in Section 4.3.3 we will present the user grouping techniques.

4.3.2 Weighted Sum Rate Maximization with Harvesting Constraints

The optimization problem presented in (4.14) involves harvesting users and information users. In the following subsections, we will study two particular cases of the previous optimization problem that correspond to situations that can be encountered in real scenarios. Such simplified scenarios will yield simpler optimization problems with lower computational complexity in the resolution of the resource allocation algorithm. Finally, the general case presented in (4.14) will be analyzed.

4.3.2.1 Scenario with Only Information Users

Consider first the broadcast scenario with only users to be served with information and no energy harvesting users, i.e., $\mathcal{U}_E = \emptyset$. In this case, problem (4.14) can be expressed as

$$\begin{aligned} \underset{\{R_i, \mathbf{S}_i\}_{\forall i \in \mathcal{U}_I}}{\text{maximize}} \quad & \sum_{i \in \mathcal{U}_I} \omega_i R_i \\ \text{subject to} \quad & C2 \dots C6 \text{ of problem (4.13).} \end{aligned} \tag{4.15}$$

The optimal solution to the above problem was presented in Chapter 3 of this dissertation and is summarized here for the sake of completeness. In this optimization problem, constraint $C3$ is fulfilled with strict equality, i.e., $R_i^\star = \log\det\left(\mathbf{I} + \mathbf{H}_i \mathbf{S}_i^\star \mathbf{H}_i^H\right)$. As a consequence, the objective function is equivalent to $\sum_{i \in \mathcal{U}_I} \omega_i \log\det\left(\mathbf{I} + \mathbf{H}_i \mathbf{S}_i \mathbf{H}_i^H\right)$. In Chapter 3, we showed that under this setup, the optimal transmit covariance matrix has the following structure: $\mathbf{S}_i^\star = \tilde{\mathbf{V}}_i^{(0)} \mathbf{V}_i \mathbf{P}_i \mathbf{V}_i^H \tilde{\mathbf{V}}_k^{(0)H}$, where $\tilde{\mathbf{V}}_i^{(0)} \in \mathbb{C}^{n_T \times (n_T - n_R + n_{R_i})}$ corresponds to the right null space of the extended channel matrix:

$$\tilde{\mathbf{H}}_i = [\mathbf{H}_1^T \quad \dots \quad \mathbf{H}_{i-1}^T \quad \mathbf{H}_{i+1}^T \quad \dots \quad \mathbf{H}_N^T]^T \in \mathbb{C}^{(n_R - n_{R_i}) \times n_T}, \tag{4.16}$$

where $\mathbf{V}_i \in \mathbb{C}^{(n_T - n_R + n_{R_i}) \times n_{S_i}}$ is obtained from the reduced[9] SVD of $\mathbf{H}_i \tilde{\mathbf{V}}_i^{(0)} = \mathbf{U}_i \mathbf{\Gamma}_i^{1/2} \mathbf{V}_i^H$, with $\mathbf{U}_i \in \mathbb{C}^{n_{R_i} \times n_{S_i}}$, $\mathbf{\Gamma}_i = \text{Diag}(\gamma_{1,i}, \dots, \gamma_{n_{S_i},i})$, $\gamma_{1,i} \geq \gamma_{2,i} \geq \cdots \geq \gamma_{n_{S_i},i} > 0$, and $\mathbf{P}_i = \text{Diag}(p_{1,i}, \dots, p_{n_{S_i},i})$ is a diagonal matrix whose elements can be calculated according to the modified multi-level water-filling power allocation policy:

$$p_{k,i}^\star = \left(\frac{\omega_i (1 - \beta_i)}{\mu \ln(2)} - \frac{1}{\gamma_{k,i}} \right)_0^\infty, \quad \forall i, k, \tag{4.17}$$

where μ is a Lagrange multiplier calculated to fulfill $\sum_{i \in \mathcal{U}_I} \text{Tr}(\mathbf{S}_i^\star) = P_{\max} - P_c^{tx}$, and β_i is a multiplier associated with the energy constraint $C4$ (if $C4$ is not active, $\beta_i = 0$). The maximum

[9] Reduced in this context means that we only consider the eigenvalues greater than zero and the corresponding eigenvectors.

sum rate achieved by problem (4.15) is, thus,

$$\mathrm{SR}_{\max}(\{p^{\star}_{k,i}\}) = \sum_{i\in\mathcal{U}_I} \omega_i R^{\star}_i = \sum_{i\in\mathcal{U}_I}\sum_{k=1}^{n_{S_i}} \omega_i \log(1+\gamma_{k,i}p^{\star}_{k,i}). \tag{4.18}$$

4.3.2.2 Scenario with Only Energy Harvesting Users

Let us now consider the case where there are only users who want to harvest energy, i.e., $\mathcal{U}_I = \emptyset$. In this particular case, since there is no objective function, the optimization problem becomes a feasibility problem [Boy04] that can be expressed as

$$\begin{aligned} \text{find} \quad & \mathbf{S} \\ \text{subject to} \quad & C1, C2, C6 \text{ of problem (4.13).} \end{aligned} \tag{4.19}$$

Notice that constraints $C3$, $C4$, and $C5$ from problem (4.14) do not affect since the set $\mathcal{U}_I$ is empty. Notice also that, without loss of optimality, we have changed the optimization variable from a set of precoding matrices $\{\mathbf{S}_i\}$ to a single precoder matrix $\mathbf{S}$. In the following, we will present a necessary condition for feasibility of (4.19).

Proposition 4.1 ([Bha87])**.** *Let $\lambda_{\max}(\boldsymbol{X}) \geq \lambda_2(\boldsymbol{X}) \geq \cdots \geq \lambda_{\min}(\boldsymbol{X})$ be the eigenvalues of the positive semidefinite matrix $\boldsymbol{X}$. Then, for any two semidefinite positive matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we have*

$$\lambda_j(\boldsymbol{AB}) \leq \lambda_{\max}(\boldsymbol{B})\lambda_j(\boldsymbol{A}) \text{ and } \lambda_j(\boldsymbol{BA}) \leq \lambda_{\max}(\boldsymbol{B})\lambda_j(\boldsymbol{A}), \quad \forall j, \tag{4.20}$$

$$\lambda_j(\boldsymbol{AB}) \geq \lambda_{\min}(\boldsymbol{B})\lambda_j(\boldsymbol{A}) \text{ and } \lambda_j(\boldsymbol{BA}) \geq \lambda_{\min}(\boldsymbol{B})\lambda_j(\boldsymbol{A}), \quad \forall j. \tag{4.21}$$

Note that the previous lemma can be generalized as: $\mathrm{Tr}\,(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{B})\,\mathrm{Tr}(\mathbf{A})$ since $\mathrm{Tr}(\mathbf{A}) = \sum_j \lambda_j(\mathbf{A})$. The inequality is attained when $\mathbf{A}$ has rank 1 and is built with the eigenvector associated with the maximum eigenvalue of $\mathbf{B}$, $(\mathbf{e}_{\max}(\mathbf{B}))$, i.e., $\mathbf{A} = k\,\mathbf{e}_{\max}(\mathbf{B})\mathbf{e}_{\max}(\mathbf{B})^H$.

Proposition 4.2. *Let $\boldsymbol{H}_j^H\boldsymbol{H}_j = \boldsymbol{V}_{H,j}\boldsymbol{\Sigma}_{H,j}\boldsymbol{V}_{H,j}^H$ be the reduced eigenvalue decomposition (EVD) of $\boldsymbol{H}_j^H\boldsymbol{H}_j$ with $\boldsymbol{\Sigma}_{H,j} = \mathrm{Diag}(\sigma_{1,j},\ldots,\sigma_{n_{R_j},j})$ and $\sigma_{1,j} \geq \sigma_{2,j} \geq \cdots \geq \sigma_{n_{R_j},j} > 0$. Then, a necessary condition for feasibility of problem (4.19) is $(P_{\max} - P_c^{tx})\sigma_{1,j} - Q_j \geq 0$, $\forall j$.*

Proof. Just as we considered before, if the problem is feasible, at least one solution fulfills $\mathrm{Tr}(\mathbf{S}) = P_{\max} - P_c^{tx}$ and the maximum value that $\mathrm{Tr}(\mathbf{H}_j\mathbf{S}\mathbf{H}_j^H)$ can take, based on Proposition 4.1, is $(P_{\max} - P_c^{tx})\sigma_{1,j}$ with $\mathbf{S} = (P_{\max} - P_c^{tx})\mathbf{v}_{n_{R_j},j}\mathbf{v}_{n_{R_j},j}^H$, where $\mathbf{v}_{n_{R_j},j}$ is the eigenvector associated with the maximum eigenvalue $\sigma_{1,j}$ of $\mathbf{H}_j^H\mathbf{H}_j$. ■

Generally, as we are not able to provide a necessary and sufficient condition, we need to

solve the following convex optimization problem to test the feasibility of problem (4.19):

$$\begin{aligned} \underset{\mathbf{S},\bar{P}_{\max}}{\text{minimize}} \quad & \bar{P}_{\max} \\ \text{subject to} \quad & C2 : \text{Tr}(\mathbf{S}) + P_c^{tx} \leq \bar{P}_{\max} \\ & C1, C6 \text{ of problem (4.13)}. \end{aligned} \tag{4.22}$$

The problem above is categorized as a semidefinite optimization problem. There is no closed-form solution for the above problem, but the optimum solution can be obtained efficiently with the application of interior point methods [Boy04]. Let us denote the optimum solution of the problem above as $\bar{P}^{\star}_{\max}$. Now, it only remains to check whether $\bar{P}^{\star}_{\max} \leq P_{\max}$ (which means that the problem is feasible) or $\bar{P}^{\star}_{\max} > P_{\max}$ (which implies infeasibility). If the problem is feasible, the optimum covariance matrix obtained in (4.22) is the matrix that fulfills all the harvesting power constraints with the minimum transmitted power.

If the problem is infeasible, then, different approaches could be taken. One possible solution would be to reduce all the power harvesting constraints $\{Q_j\}$ such that constraints $C1$ become looser until the problem becomes feasible. This reduction can be performed equally or differently among users. The status of the battery level could play an active role concerning on how to reduce such constraints. For example, employing a reduction proportional to the current battery level so that users with higher battery level suffer larger reductions of harvesting demands may be a fair solution. A simpler approach would be to drop out some of the users from the system. Due to space constraints, we do not include this analysis in the chapter.

4.3.2.3 Scenario with Both Information and Harvesting Users

In this section, we consider the scenario where we have both types of users in the system. The problem presented in (4.14) is convex and can be solved using numerical interior point methods [Boy04]. However, those methods usually have high computational complexity and, since we aim at finding a low-complexity solution, a customized algorithm should be developed. In some cases, it is possible to obtain the structure of the transmit covariance matrices in closed form and then develop a simplified and efficient iterative algorithm based on that structure. Unfortunately, it is not possible to find the closed-form expression of the optimal transmit covariances for the previous problem due to the constraint $C4$. However, as we will show later, it is possible to find the transmit covariance structure of problem (4.14) if $C4$ is not active.

In order to guarantee that constraint $C4$ is not active, we will assume that the set of information users are selected by the scheduler in a first stage in a way that they have enough battery such that $R_i^{\star}(t) < R_{\max,i}(C_i(t))$, $\forall i \in \mathcal{U}_I$ can be guaranteed in that particular scheduling period (later, we will comment what to do in the unlikely event of violating the previous requirement). This is a reasonable assumption since users who have very low battery should not be selected

for receiving information but to harvest energy. Due to the previous simplifying assumption, constraint $C4$ will not be active and, thus, we do not consider it in the optimization problem. This assumption simplifies considerably the resolution of the problem.

Given that, notice that constraint $C5$ from the original problem (4.14) forces the precoder matrix $\mathbf{B}_i$ to lie in the right null space of matrix $\tilde{\mathbf{H}}_i = [\mathbf{H}_1^T \quad \dots \quad \mathbf{H}_{i-1}^T \quad \mathbf{H}_{i+1}^T \quad \dots \quad \mathbf{H}_N^T]^T \in \mathbb{C}^{(n_R - n_{R_i}) \times n_T}$ [Spe04]. Computing the SVD of $\tilde{\mathbf{H}}_i$ yields $\tilde{\mathbf{H}}_i = \tilde{\mathbf{U}}_i \tilde{\mathbf{\Lambda}}_i [\tilde{\mathbf{V}}_i^{(1)} \quad \tilde{\mathbf{V}}_i^{(0)}]^H$, where $\tilde{\mathbf{\Lambda}}_i$ is a diagonal matrix containing the singular values and $\tilde{\mathbf{V}}_i^{(0)} \in \mathbb{C}^{n_T \times (n_T - n_R + n_{R_i})}$ contains the right-singular vectors in the null-space of $\tilde{\mathbf{H}}_i$. Thus, $\mathbf{B}_i$ can be written as $\mathbf{B}_i = \tilde{\mathbf{V}}_i^{(0)} \tilde{\mathbf{B}}_i$ (with $\tilde{\mathbf{B}}_i \in \mathbb{C}^{(n_T - n_R + n_{R_i}) \times n_{S_i}}$) and then, $\mathbf{S}_i = \tilde{\mathbf{V}}_i^{(0)} \tilde{\mathbf{S}}_i \tilde{\mathbf{V}}_i^{(0)H}$ where $\tilde{\mathbf{S}}_i = \tilde{\mathbf{B}}_i \tilde{\mathbf{B}}_i^H$. Now, the optimization problem can be rewritten in terms of the new optimization variables $\{\tilde{\mathbf{S}}_i\}$. Let $\hat{\mathbf{H}}_i = \mathbf{H}_i \tilde{\mathbf{V}}_i^{(0)}$ and $\hat{\mathbf{H}}_{ji} = \mathbf{H}_j \tilde{\mathbf{V}}_i^{(0)}$. Note that, if constraint $C4$ is not present in (4.14), constraint $C3$ is tight at the optimum,i.e., $R_i^\star = \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_i \tilde{\mathbf{S}}_i^\star \hat{\mathbf{H}}_i^H\right)$, and, thus, the objective function is directly expressed as $\sum_{i \in \mathcal{U}_I} \omega_i \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_i \tilde{\mathbf{S}}_i \hat{\mathbf{H}}_i^H\right)$. Then, problem (4.14) (without considering $C4$) is reformulated as

$$\begin{aligned} \underset{\{\tilde{\mathbf{S}}_i\}_{\forall i \in \mathcal{U}_I}}{\text{maximize}} \quad & \sum_{i \in \mathcal{U}_I} \omega_i \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_i \tilde{\mathbf{S}}_i \hat{\mathbf{H}}_i^H\right) \\ \text{subject to} \quad & C1: \sum_{i \in \mathcal{U}_I} \mathrm{Tr}(\hat{\mathbf{H}}_{ji} \tilde{\mathbf{S}}_i \hat{\mathbf{H}}_{ji}^H) \geq Q_j, \qquad \forall j \in \mathcal{U}_E \\ & C2: \sum_{i \in \mathcal{U}_I} \mathrm{Tr}(\tilde{\mathbf{S}}_i) + P_c^{tx} \leq P_{\max} \\ & C3: \tilde{\mathbf{S}}_i \succeq 0, \qquad \forall i \in \mathcal{U}_I. \end{aligned} \tag{4.23}$$

The problem above can be checked to be convex since the objective function is concave and the constraints define a convex set. As a consequence, there exists a global optimal solution and can be obtained numerically by means of, for example, interior point methods [Boy04]. However, thanks to the fact that (4.23) is convex and satisfies Slater's conditions [Boy04], the duality gap is zero and, thus, the problem can be solved using tools derived from the Lagrange duality theory and the optimal structure of the transmit covariance matrices $\{\tilde{\mathbf{S}}_i\}$ can be revealed. Let $\boldsymbol{\lambda} = \{\lambda_j\}_{j \in \mathcal{U}_E}$ be the vector of dual variables associated with constraint $C1$ and μ be the dual variable associated with constraint $C2$. The optimal solution of problem (4.23) is given by the following theorem in terms of $\boldsymbol{\lambda}^\star$ and $\mu^\star$.

Theorem 4.1. *The optimal solution of problem* (4.23) *has the following structure:*

$$\tilde{\boldsymbol{S}}_i^\star(\boldsymbol{\lambda}^\star, \mu^\star) = \boldsymbol{A}_i^{-1/2} \hat{\boldsymbol{V}}_i \hat{\boldsymbol{D}}_i \hat{\boldsymbol{V}}_i^H \boldsymbol{A}_i^{-1/2}, \tag{4.24}$$

where matrix $\boldsymbol{A}_i = \mu^\star \boldsymbol{I} - \sum_{j \in \mathcal{U}_E} \lambda_j^\star \hat{\boldsymbol{H}}_{ji}^H \hat{\boldsymbol{H}}_{ji}$, $\hat{\boldsymbol{V}}_i \in \mathbb{C}^{(n_T - n_R + n_{R_i}) \times n_{S_i}}$ *is obtained from the reduced SVD of matrix* $\hat{\boldsymbol{H}}_i^H \boldsymbol{A}_i^{-1/2} = \hat{\boldsymbol{U}}_i \hat{\boldsymbol{\Sigma}}_i^{1/2} \hat{\boldsymbol{V}}_i^H$, *with* $\hat{\boldsymbol{\Sigma}}_i = \mathrm{Diag}(\hat{\sigma}_{1,i}, \dots, \hat{\sigma}_{n_{S_i},i})$, $\hat{\sigma}_{1,i} \geq \hat{\sigma}_{2,i} \geq \cdots \geq \hat{\sigma}_{n_{S_i},i} > 0$, *and* $\hat{\boldsymbol{D}}_i = \mathrm{Diag}(\hat{d}_{1,i}, \dots, \hat{d}_{n_{S_i},i})$, *with* $\hat{d}_{k,i} = (\omega_i / \log(2) - 1/\hat{\sigma}_{k,i})_0^\infty$, $\forall i \in \mathcal{U}_I$ *and*

$k = 1, \ldots, n_{S_i}$.

Proof. See Appendix 4.A. ■

Finally, the optimum data rate achieved by user i is, thus,

$$R_i^\star = \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_i \tilde{\mathbf{S}}_i^\star \hat{\mathbf{H}}_i^H\right) = \sum_{j=1}^{n_{S_i}} \log(1 + \hat{\sigma}_{j,i}\hat{d}_{j,i}), \quad \forall i \in \mathcal{U}_I. \tag{4.25}$$

It is still pending the computation of the optimal dual variables, since in the previous development we assumed the dual variables were given (in Theorem 4.1, matrix $\mathbf{A}_i$ depends on the optimal values of the Lagrange multipliers). As long as we have a closed-formed expression of the covariance matrices $\tilde{\mathbf{S}}_i(\boldsymbol{\lambda}, \mu)$ as a function of the dual variables, we can solve the dual problem of (4.23) by maximizing the dual function $g(\boldsymbol{\lambda}, \mu)$ subject to $\boldsymbol{\lambda} \succeq 0$, $\mu \geq 0$, and $\mathbf{A}_i \succ 0$ $\forall i$. This can be addressed by applying any subgradient-type method, such as for example the ellipsoid method [Bla81]. It can be shown that the subgradient of $g(\boldsymbol{\lambda}, \mu)$ denoted as $\mathbf{t}$ is given by $[\mathbf{t}]_m = Q_{N+m} - \sum_{i \in \mathcal{U}_I} \mathrm{Tr}(\hat{\mathbf{H}}_{(N+m)i} \tilde{\mathbf{S}}_i \hat{\mathbf{H}}_{(N+m)i}^H)$ for $1 \leq m \leq M$ and $[\mathbf{t}]_{M+1} = \mathrm{Tr}(\tilde{\mathbf{S}}_i) - (P_{\max} - P_c^{tx})$ [Ber99], which represents the subgradient of $g(\boldsymbol{\lambda}, \mu)$ w.r.t. λ_m and μ, respectively, ($[\mathbf{t}]_k$ denotes the k-th entry of vector $\mathbf{t}$), and $\tilde{\mathbf{S}}_i$ is computed as in (4.24) for a given $\boldsymbol{\lambda}$ and μ (for each step of the algorithm, we compute $\tilde{\mathbf{S}}_i$ just by replacing expression (4.24) the optimal values of the Lagrange multipliers by their current values). Since the duality gap is zero, when we obtain the optimal dual variables ($\boldsymbol{\lambda}^\star$ and $\mu^\star$) with the ellipsoid method, the optimal solution $\tilde{\mathbf{S}}_i^\star(\boldsymbol{\lambda}^\star, \mu^\star)$ converges to the primal optimal solution of problem (4.23). As a summary, the algorithm that solves problem (4.23) is described in Algorithm 4.1.

4.3.2.4 Trade-Off Analysis Between Weighted Sum Rate and Power Constraints

In this section, we analyze the multidimensional trade-off between the objective function, that is, the weighted sum rate, and the set of power harvesting constraints. For simplicity, let us consider that $C_i(t)$ $\forall i \in \mathcal{U}_I$ is high enough so that it could be assumed that $R_i^\star < R_{\max,i}$ and $R_i^\star = \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_i \tilde{\mathbf{S}}_i^\star \hat{\mathbf{H}}_i^H\right)$. We would like to emphasize that, as the noise and channels are normalized, we will refer to the powers harvested by the receivers in terms of power units instead of Watts. Given that, we propose to use the *rate-power* (R-P) surface to characterize all the achievable sum rates (in bit/s/Hz) and power harvesting (in power units) $M+1$-tuples under a

Algorithm 4.1 Algorithm for solving problem (4.23)

initialize $\boldsymbol{\lambda} \succeq 0$, $\mu \geq 0$ such that $\mu\mathbf{I} - \sum_{j\in\mathcal{U}_E} \lambda_j \hat{\mathbf{H}}_{ji}^H \hat{\mathbf{H}}_{ji} \succ 0$, $\forall i$
repeat
  compute $\tilde{\mathbf{S}}_i(\boldsymbol{\lambda}, \mu)$ $\forall i$ using (4.24)
  compute subgradient of $g(\boldsymbol{\lambda}, \mu)$:
    $[\mathbf{t}]_m = Q_{N+m} - \sum_{i\in\mathcal{U}_I} \mathrm{Tr}(\hat{\mathbf{H}}_{(N+m)i} \tilde{\mathbf{S}}_i \hat{\mathbf{H}}_{(N+m)i}^H)$ for $1 \leq m \leq M$
    $[\mathbf{t}]_{M+1} = \mathrm{Tr}(\tilde{\mathbf{S}}_i) - (P_{\max} - P_c^{tx})$
  update $\boldsymbol{\lambda}$, μ using the ellipsoid method [Bla81] subject to:
    $\boldsymbol{\lambda} \succeq 0$, $\mu \geq 0$ and $\mu\mathbf{I} - \sum_{j\in\mathcal{U}_E} \lambda_j \hat{\mathbf{H}}_{ji}^H \hat{\mathbf{H}}_{ji} \succ 0$, $\forall i$
until dual variables converge
end algorithm

given power constraint as in [Zha13]. The rate-power surface of problem (4.23) is defined as

$$\mathcal{C}_{R-P}((P_{\max} - P_c^{tx}),\, \{\omega_i\}\,) \triangleq \Big\{ (\mathrm{SR}; \{Q_j\}) \mid \exists\, \{\tilde{\mathbf{S}}_i\} \text{ with } \mathrm{SR} \leq \sum_{i\in\mathcal{U}_I} \omega_i \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_i \tilde{\mathbf{S}}_i \hat{\mathbf{H}}_i^H\right),$$
$$\sum_{i\in\mathcal{U}_I} \mathrm{Tr}(\hat{\mathbf{H}}_{ji} \tilde{\mathbf{S}}_i \hat{\mathbf{H}}_{ji}^H) \geq Q_j,\ \sum_{i\in\mathcal{U}_I} \mathrm{Tr}(\tilde{\mathbf{S}}_i) + P_c^{tx} \leq P_{\max},$$
$$\tilde{\mathbf{S}}_i \succeq 0 \quad \forall j \in \mathcal{U}_E,\ \forall i \in \mathcal{U}_I \Big\}. \tag{4.26}$$

In order to be able to show an example of the trade-off graphically, we restrict the cardinality of the set of harvesting users and information users to be two, i.e., $|\mathcal{U}_E| = 2$ and $|\mathcal{U}_I| = 2$, and, for simplicity, we consider that $\omega_i = 1$, $\forall i \in \mathcal{U}_I$. In such a case, the trade-off surface between the sum rate and the two power constraints is a 3-dimensional surface.

The setup taken as an example for this section is a BS with four transmit antennas and where all users have two antennas. The maximum transmission power at the BS is $P_{\max} - P_c^{tx} = 10$ W. The entries of the matrix channels are generated independently from a complex circularly symmetric Gaussian distribution with zero mean and variance equal to one. Only one channel realization is evaluated and depicted.

Figure 4.2 depicts the 3-dimensional rate-power surface for the previous setup. As it can be appreciated, the optimal sum rate solution is jointly concave on Q_1 and Q_2, as expected [Boy04]. The values of Q_1 and Q_2 for which the surface is not defined correspond to situations where problem (4.23) is infeasible. In order to characterize the surface accurately, let us introduce the contour lines of the rate-power surface in Figure 4.3. In the plot, when the lines are close together, the magnitude of the gradient is large. There are also some important boundary points

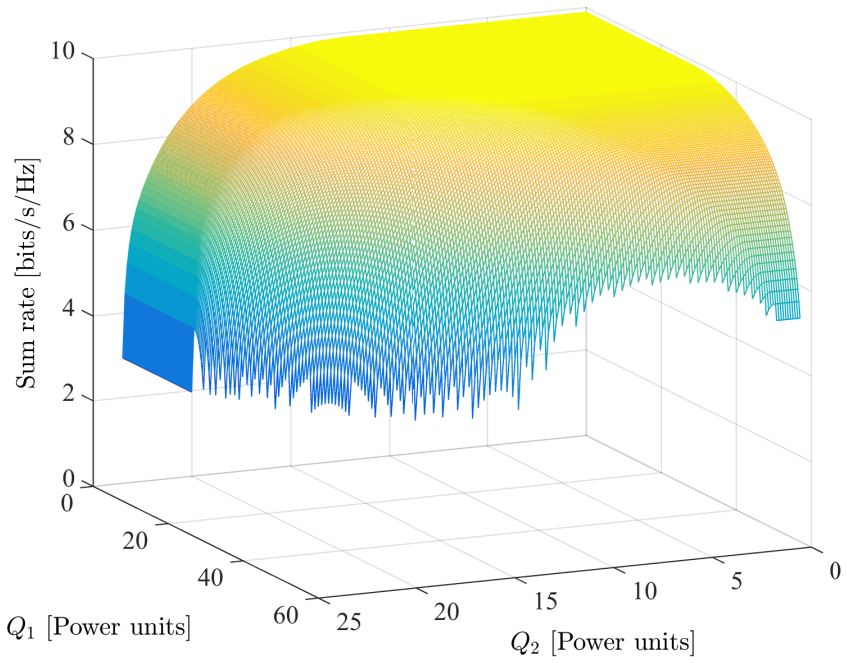


Figure 4.2: Representation of the three-dimensional rate-power surface of problem (4.23). The figure represents the existing trade-off between the optimal solution of the problem, i.e., the weighted sum rate, and the two power harvesting constraints.

marked in the 3-dimensional plot of the surface. Those points can be computed in a simple way and provide us with useful cases that will be commented in what follows.

Let us first start with the boundary point defined by $(\mathrm{SR}_{\max}, 0, 0)$. The power harvesting constraints for users 1 and 2 at this point are set to zero and, thus, the solution of the problem can be obtained from problem (4.15) (or from problem (4.23)) with $Q_1 = Q_2 = 0$. $\mathrm{SR}_{\max}$ represents the maximum sum rate that can be achieved (computed as in (4.18)). The optimum covariance matrices were obtained in Section 4.3.2.1 and are denoted here as $\tilde{\mathbf{S}}^{\star}_{\mathrm{SR}_i}$ for the i-th user. Following that notation, the maximum sum rate can also be expressed as $\mathrm{SR}_{\max} = \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_1\tilde{\mathbf{S}}^{\star}_{\mathrm{SR}_1}\hat{\mathbf{H}}_1^H\right) + \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_2\tilde{\mathbf{S}}^{\star}_{\mathrm{SR}_2}\hat{\mathbf{H}}_2^H\right)$.

Note that, although when computing $\mathrm{SR}_{\max}$ we do not apply power harvesting constraints, this does not mean necessarily that the actual harvested powers are zero. In this context, we have the boundary point $(\mathrm{SR}_{\max}, Q_1^{\mathrm{I}}, 0)$ where Q_1^{I} represents the power harvested by user 1 when the precoder matrices are the ones that maximize the weighted sum rate, i.e., $Q_1^{\mathrm{I}} = \mathrm{Tr}(\hat{\mathbf{H}}_{11}\tilde{\mathbf{S}}^{\star}_{\mathrm{SR}_1}\hat{\mathbf{H}}_{11}^H) + \mathrm{Tr}(\hat{\mathbf{H}}_{12}\tilde{\mathbf{S}}^{\star}_{\mathrm{SR}_2}\hat{\mathbf{H}}_{12}^H)$. The same can be said for the boundary point $(\mathrm{SR}_{\max}, 0, Q_2^{\mathrm{I}})$, where $Q_2^{\mathrm{I}} = \mathrm{Tr}(\hat{\mathbf{H}}_{21}\tilde{\mathbf{S}}^{\star}_{\mathrm{SR}_1}\hat{\mathbf{H}}_{21}^H) + \mathrm{Tr}(\hat{\mathbf{H}}_{22}\tilde{\mathbf{S}}^{\star}_{\mathrm{SR}_2}\hat{\mathbf{H}}_{22}^H)$. Then, there is a fourth point that defines a flat surface (or tableland) of constant sum rate $\mathrm{SR}_{\max}$, which is the combination of the two previous points, $(\mathrm{SR}_{\max}, Q_1^{\mathrm{I}}, Q_2^{\mathrm{I}})$. In other words, the tableland of constant maximum weighted sum rate $\mathrm{SR}_{\max}$ defines all possible values of harvested power constraints for which constraints $C1$ are not active and, thus, do not affect the optimum value of the weighted sum rate objective function.

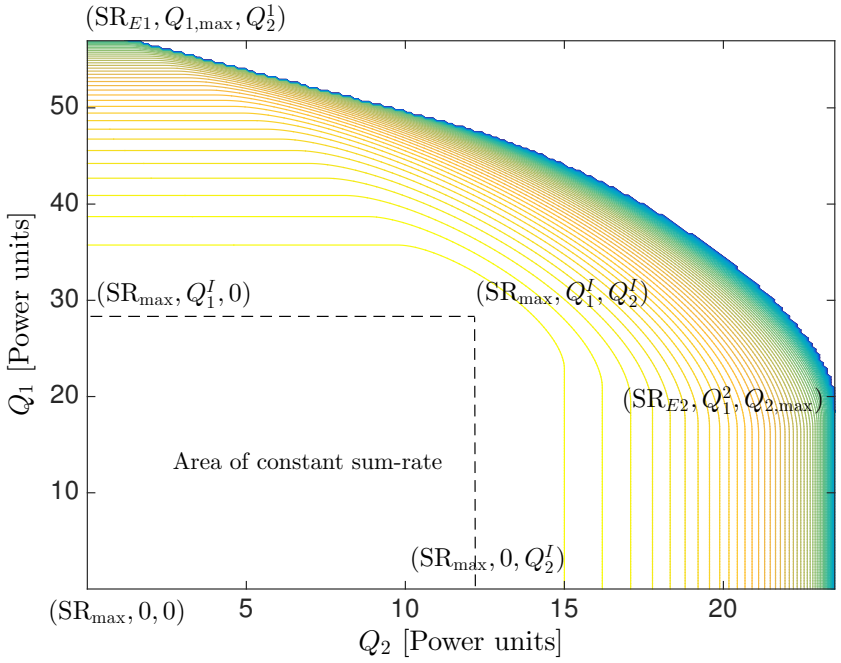


Figure 4.3: Contour lines of the three-dimensional rate-power surface of problem (4.23). Notice that the density of lines increases with the gradient of the surface, and the color indicates the value of such surface. Note also that some important boundary characteristic points have been marked.

Now, let us consider the boundary points in terms of maximum harvested power. On top of the figure, there is the point $(\mathrm{SR}_{\mathrm{E1}}, Q_{1,\max}, Q_2^1)$. This point corresponds to the situation in which the power harvested by user 1 is maximum or, in other words, the maximum value of Q_1 for which problem (4.23) is feasible assuming no constraint on the power to be harvested by user 2. To calculate $Q_{1,\max}$, we solve the following optimization problem:

$$\begin{aligned} \underset{\tilde{\mathbf{S}}_{\mathrm{E1}}}{\text{maximize}} \quad & \mathrm{Tr}(\hat{\mathbf{H}}_{11}\tilde{\mathbf{S}}_{\mathrm{E1}}\hat{\mathbf{H}}_{11}^H) \\ \text{subject to} \quad & C1 : \mathrm{Tr}(\tilde{\mathbf{S}}_{\mathrm{E1}}) + P_c^{tx} \leq P_{\max} \\ & C2 : \tilde{\mathbf{S}}_{\mathrm{E1}} \succeq 0, \end{aligned} \tag{4.27}$$

where $\tilde{\mathbf{S}}_{\mathrm{E1}}$ represents the sum of the two covariance matrices for the information users (note that in this problem the objective function and the constraint depend on such matrices through their sum) and the objective function is the power harvested by user 1. Now, by applying the result from Proposition 4.1, we obtain the solution of problem (4.27) as follows. Let the reduced eigen-decomposition of $\hat{\mathbf{H}}_{11}^H\hat{\mathbf{H}}_{11}$ be $\hat{\mathbf{U}}_{11}\hat{\mathbf{\Lambda}}_{11}\hat{\mathbf{U}}_{11}^H$ such that $\hat{\mathbf{u}}_{11,\max}$ is the eigenvector associated with the maximum eigenvalue $\hat{\lambda}_{11,\max}$. Then, the solution to the previous problem is based on the following inequality: $\mathrm{Tr}(\tilde{\mathbf{S}}_{\mathrm{E1}}\hat{\mathbf{H}}_{11}^H\hat{\mathbf{H}}_{11}) \leq \hat{\lambda}_{11,\max}\,\mathrm{Tr}(\tilde{\mathbf{S}}_{\mathrm{E1}}) = \hat{\lambda}_{11,\max}(P_{\max} - P_c^{tx})$ (since at the optimum $\mathrm{Tr}(\tilde{\mathbf{S}}_{\mathrm{E1}}^\star) = P_{\max} - P_c^{tx}$), where such inequality becomes equality if $\tilde{\mathbf{S}}_{\mathrm{E1}}^\star = (P_{\max} - P_c^{tx})\hat{\mathbf{u}}_{11,\max}\hat{\mathbf{u}}_{11,\max}^H$. In this case, the maximum harvested energy is accomplished by *energy beamforming*[10] (i.e., rank 1) to the best eigenmode of the equivalent channel $\hat{\mathbf{H}}_{11}^H\hat{\mathbf{H}}_{11}$.

[10]The concept of energy beamforming was already introduced in [Zha13].

Then, we obtain $Q_{1,\max} = \mathrm{Tr}(\hat{\mathbf{H}}_{11}\tilde{\mathbf{S}}^{\star}_{\mathrm{E1}}\hat{\mathbf{H}}^{H}_{11}) = (P_{\max} - P_c^{tx})\hat{\lambda}_{11,\max}$. According to this, the weighted sum rate obtained by solving problem (4.23) and $Q_1 = Q_{1,\max}, Q_2 = 0$ (denoted as $\mathrm{SR}_{\mathrm{E1}}$) is $\mathrm{SR}_{\mathrm{E1}} = \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_1\tilde{\mathbf{S}}^{\star}_{\mathrm{E1}}\hat{\mathbf{H}}^{H}_1\right) + \log\det\left(\mathbf{I} + \hat{\mathbf{H}}_2\tilde{\mathbf{S}}^{\star}_{\mathrm{E1}}\hat{\mathbf{H}}^{H}_2\right)$. Notice that, even though we do not apply power harvesting constraint of user 2 when computing $\tilde{\mathbf{S}}_{\mathrm{E1}}$, that does not mean that the actual power harvested by user 2 is zero. In this context, we define the last coordinate of the point, denoted as Q_2^1 which represents the power harvested by user 2 when the covariance matrix is $\tilde{\mathbf{S}}^{\star}_{\mathrm{E2}}$, i.e., $Q_2^1 = \mathrm{Tr}(\hat{\mathbf{H}}_{21}\tilde{\mathbf{S}}^{\star}_{\mathrm{E2}}\hat{\mathbf{H}}^{H}_{21})$. The same reasoning can be applied to obtain the last boundary point $(\mathrm{SR}_{\mathrm{E2}}, Q_1^2, Q_{2,\max})$ by interchanging the roles of users 1 and 2.

The rest of the boundary points in the curve can be obtained by properly varying the values of Q_1 and Q_2 ($0 \leq Q_1 \leq Q_{1,\max}$, $0 \leq Q_2 \leq Q_{2,\max}$) in problem (4.23).

4.3.3 User Selection Policies

So far we have assumed that the two groups of users, i.e., $\mathcal{U}_I$ and $\mathcal{U}_E$ were known. The goal of this section is to propose a grouping strategy to select which users should go into each set in a way that the aggregated throughput over time is maximized. As the channels and batteries fluctuate throughout time, the users in each group may also change from frame to frame. In this section we will assume that the values of $\{Q_j\}$ are known and fixed. A management strategy of these values will be presented later in this chapter.

As commented before, the optimal information and harvesting grouping should be obtained by joint exhaustive search (see Section 4.3.1). This search is prohibitively complex and, thus, suboptimum techniques should be derived. The case of having only information users has been studied in the literature and suboptimal techniques that perform closely to the optimum one have been proposed [Dim05], [Sig09]. In this chapter, in order to keep the overall complexity as low as possible without compromising the performance of the system, we present suboptimal techniques for the user grouping for both kinds of users, i.e., information and harvesting users. The overall user grouping strategy will be divided into two stages. In the first stage (that will be known as super-grouping) we will provide a pre-selection of user candidates to be in each set. This will depend primarily on the current energies available at the batteries. For the second stage, we are going to present two different user grouping strategies. The strategy with the highest complexity provides a better performance than the simpler one.

In the first (simpler) approach, we will split the user grouping further into two stages. The first stage selects the information users, $\mathcal{U}_I$, from the super-grouping set $\mathcal{U}_I^S$ based on a greedy approach, whereas the second stage selects the harvesting users, $\mathcal{U}_E$, based on the already selected information users. In the second approach, we will develop a joint information-harvesting grouping strategy, which constitutes an intermediate approach between the first simple approach and the optimum one based on exhaustive search.

4.3.3.1 User Super-grouping Strategy

Recall that when we derived the optimal precoder matrix in Section 4.3.2.3, we assumed that the optimal rates would fulfill $R_i^\star(t) < R_{\max,i}(t)$, $\forall i \in \mathcal{U}_I$ for any particular frame and, therefore, constraints $C4$ in problem (4.14) were not active. This is achieved by pre-selecting the users that are to be scheduled for data transmission or battery charging. In our proposed approach, we first implement a selection of candidates to be in $\mathcal{U}_I$ and $\mathcal{U}_E$, known as $\mathcal{U}_I^S$ and $\mathcal{U}_E^S$, such that $\mathcal{U}_I \subseteq \mathcal{U}_I^S$, $\mathcal{U}_E \subseteq \mathcal{U}_E^S$, $\mathcal{U}_I^S \cap \mathcal{U}_E^S = \emptyset$, and $|\mathcal{U}_I^S| + |\mathcal{U}_E^S| = K$, and then we select the users that finally go into the sets $\mathcal{U}_I$ and $\mathcal{U}_E$. Basically, we have two layers of grouping, one that is to be run at a longer time scale, every few scheduling periods or frames (denoted as a super-frame) and called *super-grouping*, and one that is to be run at every frame, called simply *grouping*. The proposed super-grouping algorithm is presented in Algorithm 4.2 and works as follows: we set a threshold ϑ, such that $0 \leq \vartheta \leq 1$. Then, we compute the ratio of the current battery level and the battery capacity for all users, and then we order these ratios increasingly. If the middle ratio of the previous list is greater than the value of the threshold ϑ, then we split the overall group by half and put half of the users in $\mathcal{U}_I^S$ and the other half in $\mathcal{U}_E^S$. On the other hand, if the middle ratio of the previous list is lower than the value of ϑ, we find the user with battery ratio closest to the value of ϑ and put all users with lower ratio than the one closest to ϑ in the harvesting set, and the rest of the users in the information set. The larger the value of ϑ, the more users will be included in the harvesting set $\mathcal{U}_E^S$.

4.3.3.2 Disjoint Information and Harvesting User Grouping

This first approach is based on two stages. In the first stage, the selection of the information users follows a greedy approach, in which each user is added at a time and the maximization of the weighted sum rate without harvesting constraints is evaluated for all possible candidate information users with the already selected users. No harvesting users are considered at this stage.

Let us assume, for simplicity, that every information user has the same number of antennas, i.e., $n_{R_i} = N_R$, $\forall i \in \mathcal{U}_T$. The maximum number of simultaneous users to be served following the BD strategy is then $U = \lceil \frac{n_T}{N_R} \rceil$ [Spe04]. The algorithm for selecting the information users is shown in Algorithm 4.3: first, we select the user that can achieve the greatest weighted rate[11]. Then, we incorporate one user at a time into the set only if the accumulated weighted sum

[11] A way to calculate the weights ω_i can be based on the achieved average rate as in the PF scheme [Jal00], [Wan07], [Liu10]. In that case, the weights are computed as $\omega_i(t) = \frac{1}{T_i(t)}$, being $T_i(t)$ the exponentially averaged rate calculated as $T_i(t) = \left(1 - \frac{1}{T_c}\right) T_i(t-2) + \frac{1}{T_c} R_i(t-1)$, where T_c is the effective length of the impulse response of the exponential averaging filter and $R_i(t-1)$ is the rated assigned to the i-th user in the $(t-1)$-th frame. Note that if the i-th user was not selected to be in $\mathcal{U}_I$ during the $(t-1)$-th frame, then $R_i^\star(t-1) = 0$. Otherwise, $R_i(t-1) = R_i^\star(t-1)$, i.e., the rate $R_i(t-1)$ corresponds to the solution of problem (4.23) during the $(t-1)$-th frame. Note that many other fairness criterion could be introduced by adjusting properly the weights.

Algorithm 4.2 Algorithm to obtain the super-frame sets $\mathcal{U}_I^S$ and $\mathcal{U}_E^S$

1: set a threshold $0 \leq \vartheta \leq 1$
2: order the users increasingly with the following rule:
3: $\frac{C_1(t)}{C_{\max}^1} \leq \frac{C_2(t)}{C_{\max}^2} \leq \cdots \leq \frac{C_{K/2}(t)}{C_{\max}^{K/2}} \leq \frac{C_{K/2+1}(t)}{C_{\max}^{K/2+1}} \leq \cdots \leq \frac{C_K(t)}{C_{\max}^K}$
4: **if** $\vartheta < \frac{C_{K/2}(t)}{C_{\max}^{K/2}}$
5: users $\{1, 2, \ldots, K/2\}$ go to $\mathcal{U}_E^S$
6: users $\{K/2+1, K/2+2, \ldots, K\}$ go to $\mathcal{U}_I^S$
7: **else**
8: find the user m such that $m = \arg\min_i \left| \frac{C_i(t)}{C_{\max}^i} - \vartheta \right|$
9: users $\{1, 2, \ldots, m\}$ go to $\mathcal{U}_E^S$
10: users $\{m+1, m+2, \ldots, K\}$ go to $\mathcal{U}_I^S$
11: **end if**
12: **end algorithm**

rate increases thanks to incorporating such user (weighted sum rate evaluated with the already selected users). The algorithm ends when there is no improvement in the weighted sum rate or when the maximum number of users to be scheduled (U) is reached.

Note that the distances from the BS to the users is taken into account implicitly in the algorithm, since in step 2 and step 8 of Algorithm 4.3, we select users according to the rates. These rates depend on the channel matrices $\{\mathbf{H}_i\}$, and the components of these matrices, of course, will be small if the distances are large. Therefore, the distances will have a direct impact on the selection of users.

Once we have selected the information users, we continue with the selection of the harvesting users in the second stage of this grouping strategy. The idea is to select the harvesting users so that when the resource allocation strategy is executed, they affect (reduce) the system performance as least as possible (see Section 4.3.2.4). Let $\bar{\mathbf{S}}^\star = \sum_{i \in \mathcal{U}_I} \mathbf{S}_i^\star$, where $\mathcal{U}_I$ and $\{\mathbf{S}_i^\star\}_{i \in \mathcal{U}_I}$ are the information user set and the optimum covariance matrices obtained from the algorithm detailed in Algorithm 4.3, respectively. The algorithm works as follows. For each harvesting user j, we evaluate and order decreasingly $\mathrm{Tr}(\mathbf{H}_j \mathbf{S}^\star \mathbf{H}_j) - Q_j$, and select the first M harvesting users according to this order (note that M is given, that is, it is an input to the algorithm). Note that in the previous expression, we are evaluating how the optimum covariance covariance matrices of the selected information users transmit power in the geometrical direction of the channels of the harvesting users. We also take into account the minimum required power to be harvested Q_j to ensure feasibility of the solution of the resource allocation problem. The algorithm is presented in Algorithm 4.4.

Algorithm 4.3 Algorithm to obtain the set of information users $\mathcal{U}_I$

set $\mathcal{U}_I = \emptyset$, $Q_i \geq 0$, and $\omega_i > 0$, $\forall i \in \mathcal{U}_T$
find $i_1 = \arg\max_{\forall i \in \mathcal{U}_I^S} \max_{\mathbf{S}_i} \omega_i \log\det\left(\mathbf{I} + \mathbf{H}_i \mathbf{S}_i \mathbf{H}_i^H\right)$ s. t. $\mathrm{Tr}(\mathbf{S}_i) \leq P_T$, $\mathbf{S}_i \succeq 0$
set $f_{\text{temp}} = \omega_{i_1} \log\det(\mathbf{I} + \mathbf{H}_{i_1} \mathbf{S}_{i_1} \mathbf{H}_{i_1}^H)$
set $\mathcal{U}_I \leftarrow \mathcal{U}_I \cup \{i_1\}$, $\mathcal{U}_I^S \leftarrow \mathcal{U}_I^S \setminus \{i_1\}$
for $j = 2$ to U
    for every $i \in \mathcal{U}_I^S$
        let $\mathcal{U}_I^{(i)} = \mathcal{U}_I \cup \{i\}$
        solve (4.23) without $C1$ and obtain $R_m^\star$, $\forall m \in \mathcal{U}_I^{(i)}$
        compute $f_i = \sum_{m \in \mathcal{U}_I^{(i)}} \omega_m R_m^\star$
    end for
    let $i_j = \arg\max_{i \in \mathcal{U}_I^S} f_i$
    if $f_{i_j} < f_{\text{temp}} \longrightarrow$ go to 18 (break for)
    else
        $\mathcal{U}_\mathcal{I} \leftarrow \mathcal{U}_I \cup \{i_j\}$, $\mathcal{U}_I^S \leftarrow \mathcal{U}_I^S \setminus \{i_j\}$
        let $f_{\text{temp}} = f_{i_j}$
    end if
end for
end algorithm

Algorithm 4.4 Algorithm to obtain the set of harvesting users $\mathcal{U}_E$

input: $\mathcal{U}_I$ taken from algorithm in Algorithm 4.3, $\bar{\mathbf{S}}^\star = \sum_{i \in \mathcal{U}_I} \mathbf{S}_i^\star$,
evaluate $m_j = \mathrm{Tr}(\mathbf{H}_j \bar{\mathbf{S}}^\star \mathbf{H}_j) - Q_j$, $\forall j \in \mathcal{U}_E^S$
order m_j decreasingly
construct $\mathcal{U}_E$ with the users corresponding to the first M ordered terms of m_j
end algorithm

4.3.3.3 Joint Information and Harvesting User Grouping

In this second approach, the selection of the information and harvesting users is coupled. Thanks to this joint approach, the system performance will be degraded less by the effect of having harvesting users in the system compared with the previous decoupled approach. However, the computational complexity increases as more combinations need to be evaluated.

Algorithm 4.5 Algorithm to obtain jointly the set of information and harvesting users $\mathcal{U}_I$, $\mathcal{U}_E$

1: set $\mathcal{U}_I = \emptyset$, $Q_i \geq 0$, and $\omega_i > 0$, $\forall i \in \mathcal{U}_T$
2: find $i_1 = \arg\max_{\forall i \in \mathcal{U}_I^S} \max_{\mathbf{S}_i} \omega_i \log\det\left(\mathbf{I} + \mathbf{H}_i \mathbf{S}_i \mathbf{H}_i^H\right)$ s. t. $\mathrm{Tr}(\mathbf{S}_i) \leq P_T$, $\mathbf{S}_i \succeq 0$
3: set $f_{\text{temp}} = \omega_{i_1} \log\det(\mathbf{I} + \mathbf{H}_{i_1} \mathbf{S}_{i_1} \mathbf{H}_{i_1}^H)$
4: set $\mathcal{U}_I \leftarrow \mathcal{U}_I \cup \{i_1\}$, $\mathcal{U}_I^S \leftarrow \mathcal{U}_I^S \setminus \{i_1\}$
5: evaluate $m_j = \mathrm{Tr}(\mathbf{H}_j \mathbf{S}_{i_1}^\star \mathbf{H}_j) - Q_j$, $\forall j \in \mathcal{U}_E^S$
6: find the k users with highest value of m_j. Put them in set $\mathcal{H}$
7: set $\mathcal{U}_E \leftarrow \mathcal{U}_E \cup \mathcal{H}$, $\mathcal{U}_E^S \leftarrow \mathcal{U}_E^S \setminus \mathcal{H}$, $\mathcal{H} = \emptyset$
8: **for** $j = 2$ to U
9: **for** every $i \in \mathcal{U}_I^S$
10: let $\mathcal{U}_I^{(i)} = \mathcal{U}_I \cup \{i\}$
11: solve (4.23) and obtain $R_m^\star$, $\mathbf{S}_m^\star$, $\forall m \in \mathcal{U}_I^{(i)}$
12: compute $f_i = \sum_{m \in \mathcal{U}_I^{(i)}} \omega_m R_m^\star$
13: **end for**
14: let $i_j = \arg\max_{i \in \mathcal{U}_I^S} f_i$
15: **if** $f_{i_j} < f_{\text{temp}} \longrightarrow$ go to 24 (break for)
16: **else**
17: $\mathcal{U}_\mathcal{I} \leftarrow \mathcal{U}_I \cup \{i_j\}$, $\mathcal{U}_I^S \leftarrow \mathcal{U}_I^S \setminus \{i_j\}$
18: let $f_{\text{temp}} = f_{i_j}$
19: **end if**
20: evaluate $m_j = \mathrm{Tr}(\mathbf{H}_j \sum_{i \in \mathcal{U}_I} \mathbf{S}_i^\star \mathbf{H}_j) - Q_j$, $\forall j \in \mathcal{U}_E^S$
21: find the k users with highest value of m_j. Put them in set $\mathcal{H}$
22: set $\mathcal{U}_E \leftarrow \mathcal{U}_E \cup \mathcal{H}$, $\mathcal{U}_E^S \leftarrow \mathcal{U}_E^S \setminus \mathcal{H}$, $\mathcal{H} = \emptyset$
23: **end for**
24: **end algorithm**

The algorithm for selecting the information users is based on the same greedy approach that we presented before. The difference is that, now, instead of selecting the information users and, then, the harvesting users, we select both types of users simultaneously. For simplicity in the formulation, let us consider that M is an integer multiple of U and define $k = \frac{M}{U}$ (we will comment later how we could apply the algorithm if that was not the case). The idea behind the algorithm is as follows. We select one information user q and obtain its optimum covariance matrix $\mathbf{S}_q^\star$. Then, we find the best k harvesting users based on the principle developed in Algorithm 4.4. After that, we select another information user and repeat the same process until there is no improvement in the objective function. Thanks to the fact that the grouping

is coupled, we consider the impact of having selected harvesting users on the future selection of information users. The specific details of the joint algorithm are presented in Algorithm 4.5.

The main difference with algorithm in Algorithm 4.4 is that, now, we solve problem (4.23) with constraints $C1$, that is, with harvesting users, and this increases the complexity of the overall grouping procedure. As addressed before, if M is not an integer multiple of U, we can introduce more harvesting users in step 21 in Algorithm 4.5 in some iterations, e.g., if $M = 7$ and $U = 3$, we first select 3 harvesting users, and then 2 harvesting users in the other 2 iterations.

4.3.4 Overall User Grouping and Resource Allocation Algorithm

In the following, we present a summary of the overall algorithm that consists of the user super-grouping, the user grouping, and the resource allocation stages presented in the previous two sections. Note that the user super-grouping is carried out every few frames whereas the user grouping is executed at each frame. If, for some reason, the super-grouping algorithm fails in fulfilling $R_i^\star(t) < R_{\max,i}(t)$, $\forall i \in \mathcal{U}_I$ (an event that would be unlikely to happen), then for those users for which $R_i^\star(t) \geq R_{\max,i}(t)$ we just transmit information in some channel accesses of the frame until their battery is over. The overall algorithm is detailed in Algorithm 4.6.

4.3.5 Numerical Simulations

In this section, we perform some numerical analysis of the proposed grouping and resource allocation strategies. The system is composed of one transmitter with 8 antennas, and 30 users ($|\mathcal{U}_T| = 30$) with 2 antennas each. The maximum radiated power is $P_{\max} = 11$ W and the transmitter front-end consumption is $P_c^{tx} = 1$ W. Front-end power consumption at the receiver is $P_c^{rx} = 100$ mW and the model used for decoding is exponential, i.e., $P_{\text{dec}}(R) = c_1 \mathrm{e}^{c_2 R}$, where $c_1 = 30$ W and $c_2 = 0.75$ 1/(bits/s/Hz). The frame duration is equal to $T_f = 100$ ms and the super-frame duration is equal to 3 s. The channel matrices are generated randomly with i.i.d. entries distributed according to $\mathcal{CN}(0, 1)$. The noise power is normalized to 1. The effective window length for the PF scheme is $T_c = 5$. The threshold value used for super-grouping is $\vartheta = 0.1$. The battery capacities are generated randomly from 3,000 to 10,000 energy units. As we mentioned before, we assume that all the harvesting constraints are the same for all users and fixed for all periods to $Q_j = 50$ power units, unless stated otherwise.

Algorithm 4.6 Overall user grouping and resource allocation algorithm

beginning of a super-frame:
  run **user super-grouping** algorithm in Algorithm 4.2: obtain sets $\mathcal{U}_I^S$ and $\mathcal{U}_E^S$
beginning of each frame (two options):
option 1:
  run **information user grouping** algorithm in Algorithm 4.3: obtain set $\mathcal{U}_I$
  run **harvesting user grouping** algorithms in Algorithm 4.4: obtain set $\mathcal{U}_E$
  run **resource allocation** algorithm in Algorithm 4.1
option 2:
  run **joint information and harvesting grouping** algorithm in Algorithm 4.5: obtain sets $\mathcal{U}_I$ and $\mathcal{U}_E$
  run **resource allocation** algorithm in Algorithm 4.1
end of each frame:
update batteries:
$$C_i(t) = \Big(C_i(t-1) - T_f P_{\text{tot},i}^{r_x}(R_i^\star(t-1))\Big)_0^{C_{\max}^i}, \quad \forall i \in \mathcal{U}_I$$
$$C_j(t) = \big(C_j(t-1) + T_f \bar{Q}_j(t-1) - T_f P_c^{r_x}\big)_0^{C_{\max}^j}, \quad \forall j \in \mathcal{U}_E$$
update weights (e.g. using a PF approach):
$$w_i(t) = \tfrac{1}{T_i(t)}, \quad T_i(t) = \left(1 - \tfrac{1}{T_c}\right) T_i(t-2) + \tfrac{1}{T_c} R_i^\star(t-1)$$
end algorithm

In the simulations, we compare our proposed two methods with two other schemes. As there are no proposals in the literature for user scheduling in the SWIPT framework, we compare our approaches with traditional schemes. In one of the schemes we assume that the super-grouping and the grouping are implemented with a round robin strategy. We will denote this strategy RR-SF / RR-F. In the other scheme we consider that random selection of users is implemented at both levels as well. This strategy will be denoted by Ra-SF / Ra-F. On the other hand, the proposed super-grouping strategy (Algorithm 4.2) will be denoted by LB and the grouping will be denoted according to the algorithm, DHS for de decoupled approach presented in Section 4.3.3.2 (Algorithms 4.3 and 4.4) and CHS for the coupled approach presented in Section 4.3.3.3 (Algorithm 4.5).

4.3.5.1 Time Evolution Simulations

Figure 4.4 presents the estimate of the expected sum rate of the system (that is computed as $SR(\tau) = \frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{i\in\mathcal{U}_I} R_i(t)$). From the figure, we see that the sum rate of the round robin and the random schemes provide a stable average throughput over time but the magnitude of

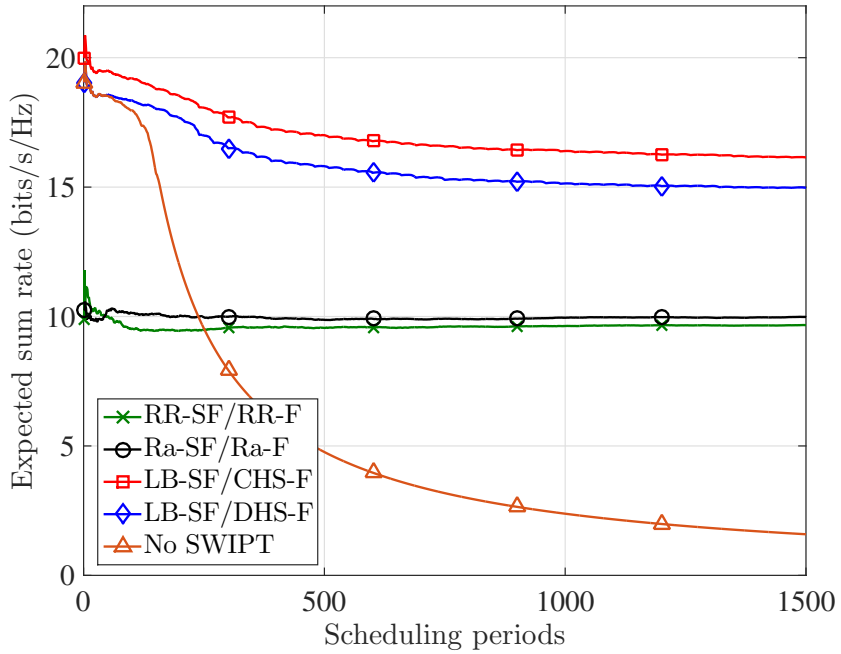


Figure 4.4: Average sum rate of the system for the different approaches.

the throughput is not so high. Then, we see how the proposed schemes outperform notably the previous benchmarking strategies. The simpler approach, DHS, performs similar to the more complex strategy, CHS. We also plot as benchmark the case where no power transfer (no SWIPT) is available. In this case, the users run out of battery and, thus, the expected sum rate tends to zero.

Figure 4.5 shows the cumulative distribution function (CDF) of the individual data rates of the users in the system. The CDF of the no SWIPT case has a particular shape due to the fact that many users obtain zero data rate as they run out of battery. In this figure we clearly see the benefits of the proposed user selection schemes compared to the other approaches as low data rate percentiles as well as high data rate percentiles are much better for the proposed strategies.

Figure 4.6 depicts the evolution of the battery levels of all users in the system. We can observe that for the round robin scheme, users reach their maximum battery capacity. This is because the data rates achieved are low and, thus, users use little energy for decoding. Then, at the top-right figure we have the case where random scheduling is considered. In this case we see how the battery evolution of all users evolve randomly as in each frame new users are scheduled in a random fashion. Due to the battery overflows that some users experience and the randomness in the selection, this approach, as happens with the round robin scheme, are not very efficient in terms of aggregate throughput (as we saw before). The last two figures depict the battery evolution of the two proposed schemes. We observe that in both cases the battery levels of the users are substantially lower than the ones observed in previous schemes. This reduction in battery levels is related with the large throughput achieved by the users.

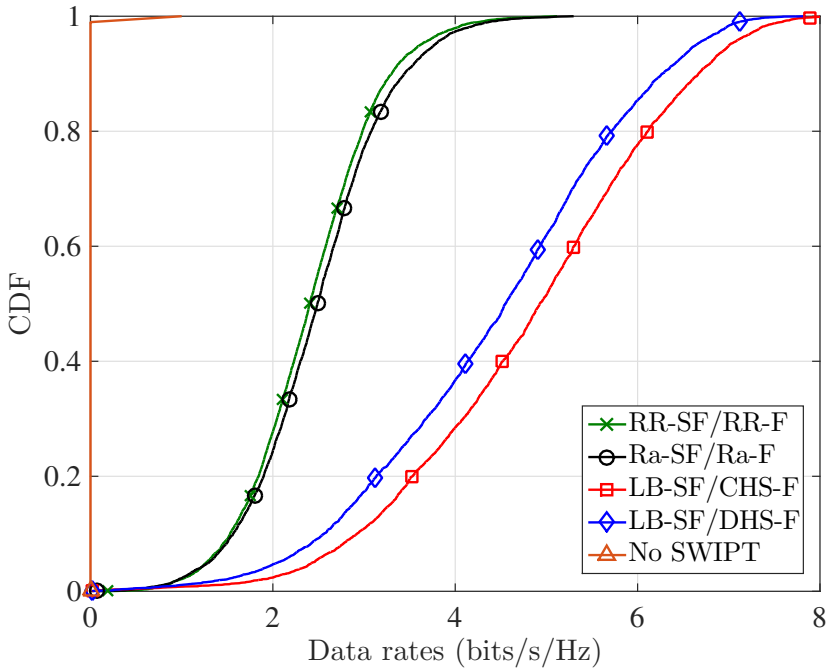


Figure 4.5: CDF of the individual data rates of all the users in the system for the different approaches.

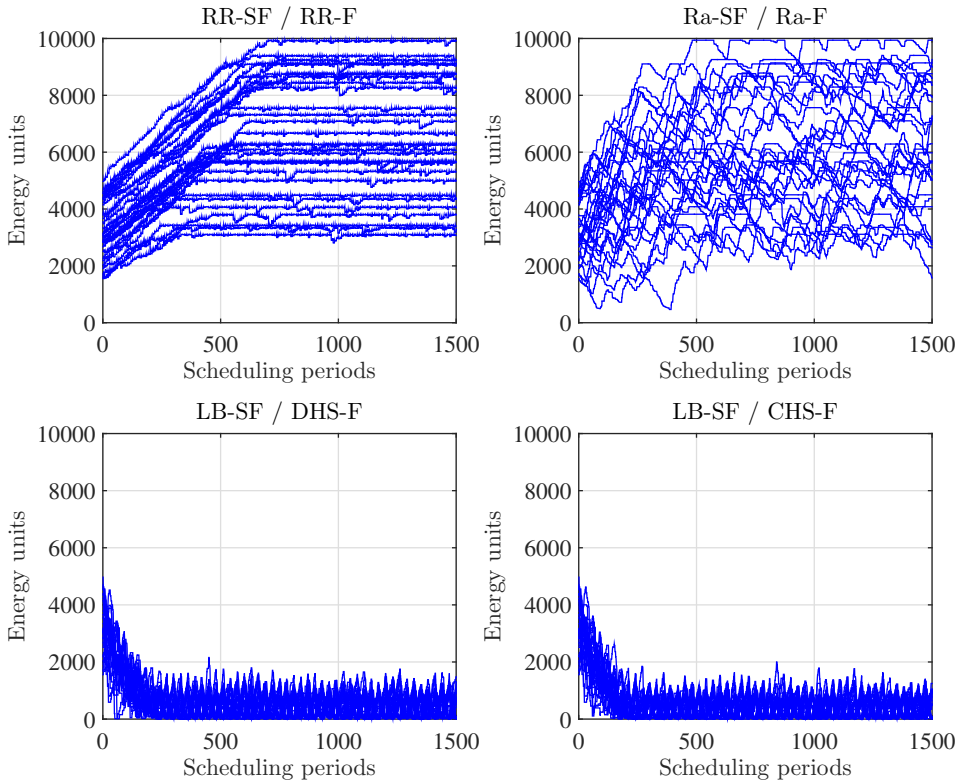


Figure 4.6: Time evolution of the battery levels of all users in the system for the different approaches.

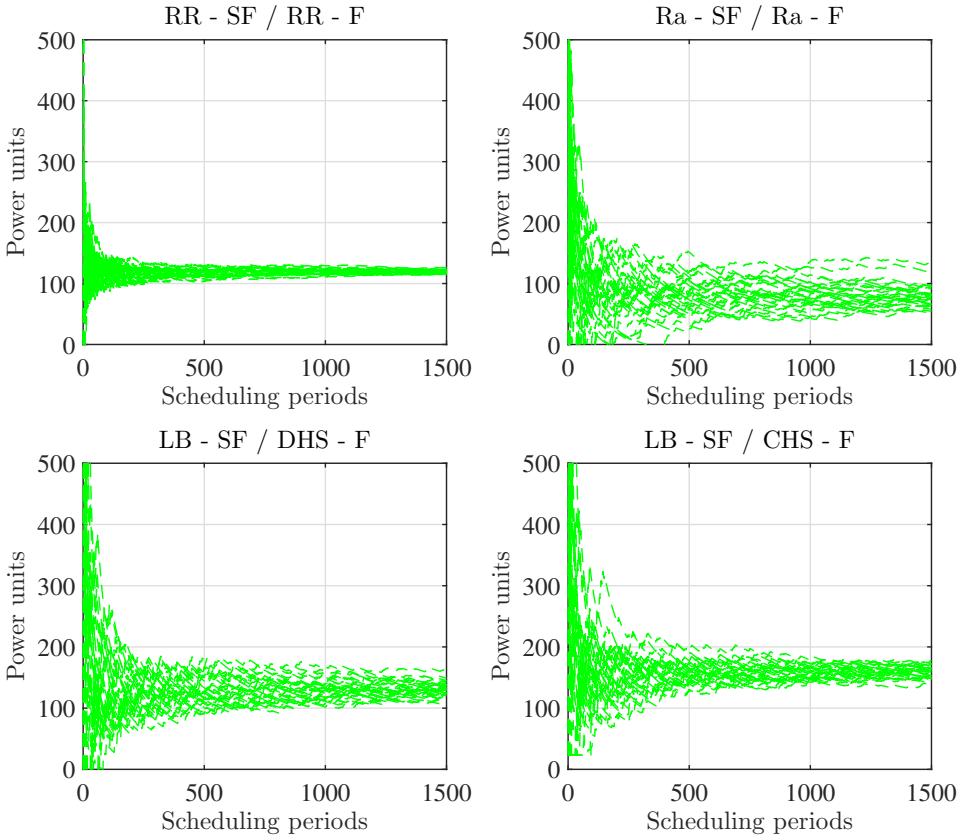


Figure 4.7: Average harvested power of all the users in the system for the different approaches (in power units).

Finally, in Figure 4.7 we depict the average evolution of the harvested power. It is interesting to note how all users tend to converge to a certain point (or the vicinity of a point). This is due to the fact that if a user is receiving much power, then its battery will increase and that will make the user more eligible to receive data making the harvesting decrease; whereas if a user has low energy in its battery, then it is directly selected to be included in set $\mathcal{U}_E^S$. We observe that the more complex approach CHS is able to provide the users with larger harvested power compared to the less complex approach, DHS.

4.3.5.2 System Performance Simulations

In the next figures, we will show the performance of the system obtained once the algorithms have converged (i.e., after 1500 frames). The first two figures, Figures 4.8 and 4.9, show the system performance considering that half of the users are at a relative distance to the BS greater than the other half of the users. In particular, Figure 4.8 presents the sum of the expected sum rate for the four schemes for four different relative distances. As expected, the sum rate decreases as the distance to the BS increases. On the other hand, Figure 4.9 shows the sum of the expected harvested power as a function of the relative distance. We see that if half of the users are four times farther away from the BS, the loss in harvested power is from 25% to 50%, and that the relative loss is lower for the proposed schemes.

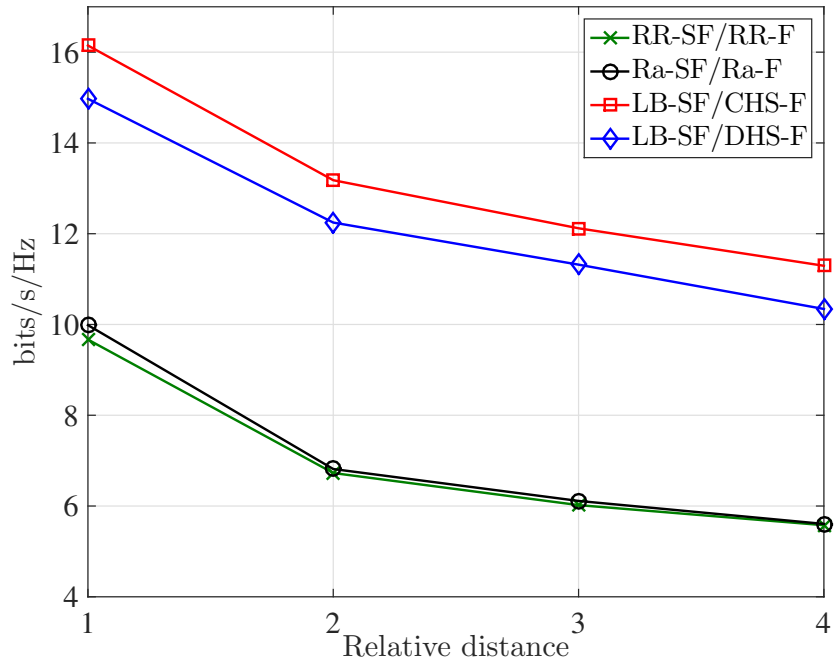


Figure 4.8: Expected system sum rate as a function of the distance to the BS.

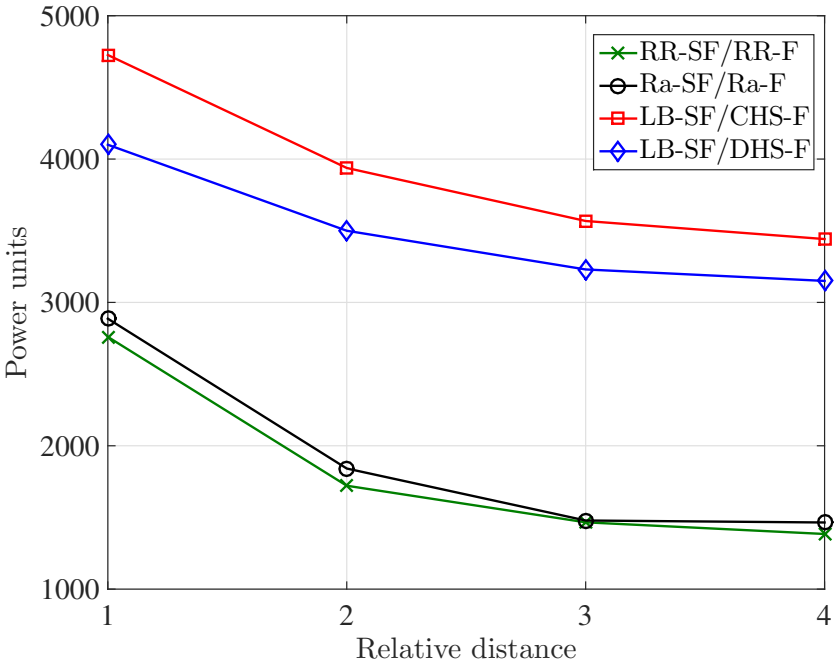


Figure 4.9: Sum of the expected harvested powers by all users (in power units) as a function of the distance to the BS.

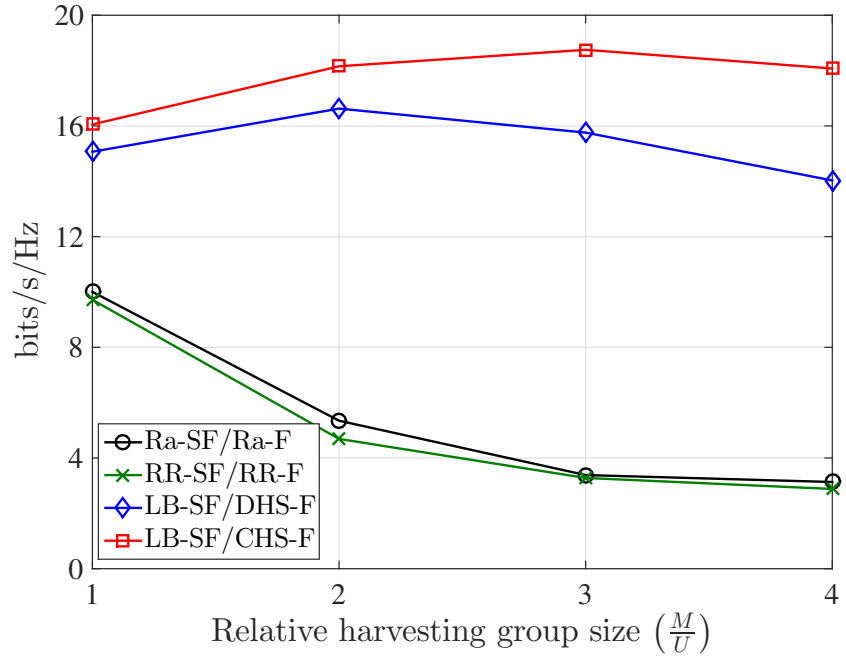


Figure 4.10: Expected system sum rate as a function of the relative size of the harvesting user group.

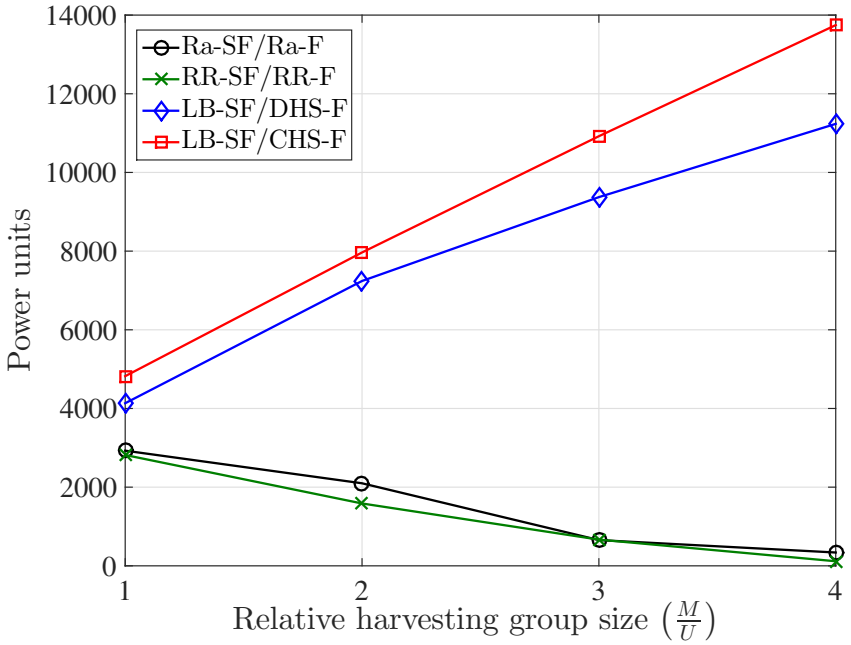


Figure 4.11: Sum of the expected harvested powers by all users (in power units) as a function of the relative size of the harvesting user group.

The last two figures, Figures 4.10 and 4.11, show the performance of the system when the size of the harvesting group increases in relative terms when compared to the size of the information group, i.e., when $\frac{M}{U}$ increases. This phenomenon is interesting to evaluate since the harvesting users appear in the constraints and they affect negatively the aggregated sum rate (see trade-off in Section 4.3.2.4). However, if many users are introduced in the harvesting set, then their batteries will recharge faster and they will be able to receive higher data rates. This is the compromise that is analyzed in the figures. First, in Figure 4.10 we see the expected aggregated sum rate. As we see, for the two benchmarking approaches, Ra and RR, the sum rate decreases as $\frac{M}{U}$ increases. This is because the harvesting users are selected without considering the impact that they have on the objective function and, therefore, if more harvesting users are considered in the optimization problem, lower sum rate will be achieved. In those cases, the optimization problem turns out to be infeasible many times and, therefore, the energy collected by all users also decreases, see Figure 4.11. On the other hand, the aggregated sum rate increases a bit for $\frac{M}{U} = 2$ for the proposed strategies. This is thanks to the fact that harvesting users are selected very efficiently and, thus, the constraints associated to them are not active, i.e., they do not affect the optimum value of the objective function. Besides as more users are able to recharge their batteries (see Figure 4.11), they can decode higher rates in future frames. Nonetheless, from a given size $\frac{M}{U}$ on, the system sum rate starts to decrease as the harvesting constraints become active although the problem is always feasible and users recharge their batteries as it is indirectly depicted in Figure 4.11.

4.3.6 Harvesting Management Strategies Based on Convex Sensitivity Theory

In the previous sections, we have considered that the minimum powers to be harvested, i.e., $\{Q_j\}$, $\forall j \in \mathcal{U}_E$ were known and fixed. However, the particular values of such constants affect considerably the system performance, i.e., the weighted sum rate. In Section 4.3.2.4, we presented the multidimensional trade-off that there exists between the sum rate and the individual harvesting constraints. In this section we will develop an approach to configure (i.e., recalculate) such harvesting constants $\{Q_j\}$, $\forall j \in \mathcal{U}_E$ under a pre-established target weighted sum rate.

In situations where the original problem is feasible but the sum rate obtained is not enough, the system may be forced to relax (decrease) the energy harvesting constraints so that the overall sum rate is enhanced. The idea is to identify which are the harvesting constraints that produce the largest enhancement of sum rate when they are reduced and to apply a reduction on them. On the other hand, if the target sum rate is below the one achieved, we could spend more resources on recharging the batteries of the harvesting users. In this case, a strategy for increasing the harvesting constants $\{Q_j\}$ is also needed. Ideally, we would like to modify the harvesting constants that accept a larger positive change and yield a small sum rate loss.

In order to identify the constraints to be changed, we use the theory of perturbation analysis from convex optimization theory [Ber99]. It is well-known that the Lagrange multipliers (dual variables) provide information about the sensitivity of the objective function w.r.t. the perturbations in the constraints. Let $\mathbf{q}^0$ be the vector of initial power harvesting constraints, i.e., $\mathbf{q}^0 = [Q_1^0, Q_2^0, \ldots, Q_M^0]^T$. Let $p^\star(\mathbf{q}^0)$ be the optimal value of problem (4.14), that is, $f_0(\mathbf{S}^\star) = p^\star(\mathbf{q}^0)$, where $f_0(\cdot)$ denotes the objective function in problem (4.14) and $\mathbf{S}^\star = (\mathbf{S}_i^\star)_{\forall i \in \mathcal{U}_I}$. From [Ber99] we know that the function $p^\star(\mathbf{q})$ is concave w.r.t. $\mathbf{q}$ where $\mathbf{q} = [Q_1, \ldots, Q_M]^T$ is the power harvesting perturbed vector defined as $\mathbf{q} = \mathbf{q}^0 + \Delta\mathbf{q}$, where $\Delta\mathbf{q} = [\Delta Q_1, \ldots, \Delta Q_M]^T$ being ΔQ_j a small change in the initial Q_j^0. Given this, we have that the optimal objective value of the relaxed problem can be upper bounded as

$$p^\star(\mathbf{q}) \leq p^\star(\mathbf{q}^0) + \nabla_{\mathbf{q}}\, p^\star(\mathbf{q}^0)^T(\mathbf{q} - \mathbf{q}^0). \tag{4.28}$$

Then, applying the following result from local sensitivity [Ber99],

$$\frac{\partial p^\star(\mathbf{q}^0)}{\partial Q_i} = -\lambda_i^\star(\mathbf{q}^0), \quad \text{with } \lambda_i^\star(\mathbf{q}^0) \geq 0,\ \forall i, \tag{4.29}$$

it follows that

$$\nabla_{\mathbf{q}}\, p^\star(\mathbf{q}^0) = \left[\frac{\partial p^\star(\mathbf{q}^0)}{\partial Q_1} \;\; \frac{\partial p^\star(\mathbf{q}^0)}{\partial Q_2} \;\; \ldots \;\; \frac{\partial p^\star(\mathbf{q}^0)}{\partial Q_M}\right]^T \tag{4.30}$$

$$= -\left[\lambda_1^\star(\mathbf{q}^0) \;\; \lambda_2^\star(\mathbf{q}^0) \;\; \ldots \;\; \lambda_M^\star(\mathbf{q}^0)\right]^T \tag{4.31}$$

$$= -\boldsymbol{\lambda}^\star(\mathbf{q}^0), \tag{4.32}$$

and the expression for the relaxed problem fulfills the following inequality defined by a hyperplane:

$$p^\star(\mathbf{q}) \leq p^\star(\mathbf{q}^0) - \boldsymbol{\lambda}^\star(\mathbf{q}^0)^T(\mathbf{q} - \mathbf{q}^0). \tag{4.33}$$

Now let us define the target sum rate as r_t and let us assume throughout the section that[12] $r_t > p^\star(\mathbf{q}^0)$. We would like to find a vector $\mathbf{q}$ such that $r_t = p^\star(\mathbf{q})$, but since $p^\star(\mathbf{q})$ is not known, we force r_t to be equal to the upper bound in (4.33):

$$r_t = p^\star(\mathbf{q}^0) - \boldsymbol{\lambda}^\star(\mathbf{q}^0)^T(\mathbf{q} - \mathbf{q}^0). \tag{4.34}$$

However, since $p(\cdot)^\star$ is a concave function, the solution obtained $p^\star(\mathbf{q})$ will be indeed below the desired sum rate, i.e., $p^\star(\mathbf{q}) \leq r_t$. In order to get a very close solution, that is $p^\star(\mathbf{q}) \approx r_t$, we must proceed iteratively by applying successive perturbations on vector $\mathbf{q}$ in a way similar to the well-known Newton's method [Ber99]. Before presenting the iterative algorithm, let us present different approaches (modeled as convex optimization problems) of how we can compute the new

[12] In case we had $r_t < p^\star(\mathbf{q}^0)$, then we should modify slightly the optimization problems that will be presented later in this section in order to increase the initial harvesting constraints until $r_t = p^\star(\mathbf{q})$.

relaxed power harvesting parameters $\{Q_j\}$ since, as we are referring to a vector of variables, there exist different ways to update the vector $\mathbf{q}$ that yield the same sum rate solution.

The first approach we propose is the simplest one. In this case, we fix the perturbed vector $\mathbf{q}$ to be a scaled version of the original vector, that is, $\mathbf{q} = \alpha\mathbf{q}^0$. In such a case, all the power harvesting constraints are reduced proportionally by the same amount. We seek to find the maximum value of α that produces the perturbed vector to yield the desired sum rate. Let us define $\tilde{r}_t = p^\star(\mathbf{q}^0) + \boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^0 - r_t$ and assume that $\tilde{r}_t \geq 0$, otherwise we cannot find any feasible vector $\mathbf{q}$, i.e., any $\mathbf{q} \succeq 0$. The problem is formulated as follows:

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \alpha \\ \text{subject to} \quad & C1 : \alpha\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^0 \leq \tilde{r}_t \\ & C2 : \alpha \geq 0. \end{aligned} \tag{4.35}$$

Lemma 4.1. *The optimal solution of problem* (4.35) *and the optimal perturbed vector are given by*

$$\alpha^\star = \frac{\tilde{r}_t}{\boldsymbol{\lambda}^\star(\boldsymbol{q}^0)^T\boldsymbol{q}^0}, \quad \boldsymbol{q}^\star = \frac{\tilde{r}_t}{\boldsymbol{\lambda}^\star(\boldsymbol{q}^0)^T\boldsymbol{q}^0}\boldsymbol{q}^0. \tag{4.36}$$

Proof. See Appendix 4.C. ■

Now, we propose a different approach to compute the perturbed vector $\mathbf{q}$. In this approach, we let the harvesting constraints have different relaxations and the objective is to minimize the sum of the harvesting reduction, i.e., $\|\Delta\mathbf{q}\|_1 = \|\mathbf{q} - \mathbf{q}^0\|_1$. The problem is formulated as follows:

$$\begin{aligned} \underset{\mathbf{q}}{\text{minimize}} \quad & \|\mathbf{q} - \mathbf{q}^0\|_1 \\ \text{subject to} \quad & C1 : \boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q} \leq \tilde{r}_t \\ & C2 : \mathbf{q} \succeq \mathbf{0}. \end{aligned} \tag{4.37}$$

The optimal solution of previous problem is given in the following result.

Lemma 4.2. *Let n be the index corresponding to the maximum Lagrange multiplier*[13]*, i.e., $\lambda_n^\star > \lambda_m^\star$, $\forall m \neq n$. The optimal solution of problem* (4.37) *is given by*

$$q_n^\star = \frac{1}{\lambda_n^\star}\left(\tilde{r}_t - \sum_{i\neq n}\lambda_i^\star Q_i^0\right), \qquad q_m^\star = Q_m^0 \quad \forall m \neq n. \tag{4.38}$$

Proof. See Appendix 4.D. ■

[13] We have assumed that there is just one maximum Lagrange multiplier. In case there were more than just one, we may choose one randomly.

As it can be seen, the optimal solution applies the harvesting power reduction to the user having the largest Lagrange multiplier associated with its harvesting constraint whereas the rest of the users remain with the same harvested power.

The final proposed approach tries to be fair in terms of harvested reduction. The fairness in achieved by considering the objective function to be the maximization of the minimum q_i. The reformulated (differentiable) problem is

$$\begin{aligned} \underset{\mathbf{q},\, t}{\text{maximize}} \quad & t \\ \text{subject to} \quad & C1 : t\mathbf{1} \preceq \mathbf{q} \\ & C2 : \boldsymbol{\lambda}^{\star}(\mathbf{q}^0)^T\mathbf{q} \leq \tilde{r}_t \\ & C3 : \mathbf{q} \succeq \mathbf{0}. \end{aligned} \tag{4.39}$$

Lemma 4.3. *The optimal solution of problem* (4.39) *is given by*

$$\boldsymbol{q}^{\star} = \frac{\tilde{r}_t}{\boldsymbol{\lambda}^{\star}(\boldsymbol{q}^0)^T\boldsymbol{1}}\boldsymbol{1}. \tag{4.40}$$

Proof. See Appendix 4.E. ■

As it can be seen, due to the maximin approach in problem (4.39), all users end up with the same perturbed power constraint. As a consequence, some users could end up with more harvested energy than the initial one (i.e., $q_j^{\star} > Q_j^0$ for some j).

As it was commented before, the three previous approaches only yield a solution such that the actual rate $r_t \geq p^{\star}(\mathbf{q})$ due to the concavity of function $p^{\star}(\cdot)$. For this reason, it is not enough with just one iteration and we have to apply the previous algorithm iteratively to get a closer solution to the target sum rate. Let us denote the obtained perturbed vector and the sum rate at iteration k by $\mathbf{q}^{(k)}$ and $r^{(k)} = p^{\star}(\mathbf{q}^{(k)})$, respectively. Let us introduce the parameter ϵ that trades-off the speed of convergence and the solution accuracy. The idea behind the iterative algorithm, described in Algorithm 4.7, is to use the previous procedures ((4.36), (4.38), and (4.40)) but with different iterations over Q_j, starting with Q_j^0.

4.3.6.1 Numerical Simulations of the Harvesting Management Strategies

In this section we present illustrative examples of the behavior of the different strategies developed in the previous section for a given particular frame. Let us consider that $C_i(t)\ \forall i \in \mathcal{U}_I$ is high enough so that it can be assumed that $R_i^{\star} < R_{\max,i}$. The setup is a BS with four transmit antennas, and two information users and two harvesting users with two antennas each. The maximum transmission power at the BS is $P_T = 50$ W. The entries of the matrix channels are generated from a complex Gaussian distribution with zero mean and unit variance. The values

Algorithm 4.7 Algorithm for adjusting the harvesting constraints

1: $k = 0$, $\mathbf{q}^{(k)} = (Q_1^{(k)}, \ldots, Q_M^{(k)})^T$
2: solve problem (4.14) $\longrightarrow r^{(k)} = p^\star(\mathbf{q}^{(k)})$, $\boldsymbol{\lambda}^\star(\mathbf{q}^{(k)})$
3: **while** $(r^{(k)} < r_t - \epsilon)$
4: obtain $\mathbf{q}^{(k+1)}$ following any strategy from (4.36), (4.38), or (4.40)
5: solve problem (4.14) $\longrightarrow r^{(k+1)} = p^\star(\mathbf{q}^{(k+1)})$, $\boldsymbol{\lambda}^\star(\mathbf{q}^{(k+1)})$
6: update $k \longleftarrow k + 1$
7: **end while**
8: **end algorithm**

of the initial minimum power to be harvested are $Q_1^0 = 33$ power units and $Q_2^0 = 19$ power units. The target weighted sum rate is $r_t = 8.5$ bits/s/Hz where the weights are $\omega_j = 1$, $\forall j$.

Figure 4.12 depicts the 3D rate-power surface of the achieved sum rate as a function of different values of $\{Q_j\}$. In the figure, we have considered the solution based on (4.39). The blue dot represents the achieved sum rate for the particular values of Q_1 and Q_2 assigned. In the figure, the black line represents the target sum rate, which in this particular case is greater than the initial case. In the plot, we show the different iterations that the Newton-like proposed algorithm presented in Algorithm 4.7 produces. As it can be seen, in just 3 iterations we obtain a solution close to the target sum rate with an error lower than 10^{-6} bits/s/Hz. Note that, when the algorithm converges, both users end up with the same amount of power to be harvested (approximately 20 power units each one).

The performance of the solution based on (4.35) is presented in Figure 4.13. In this case, the figure shows the contour lines of the 3D rate-power surface in order to better visualize the behavior of the algorithm. Also in this case, just 3 iterations are enough to yield a solution in the neighborhood of the target rate. Now, both harvesting users decrease their harvesting requirements by the same amount.

Finally, the behavior of the algorithm based on (4.37) is shown in Figure 4.14. As expected, just the user who has the largest harvesting requirement is modified while the other one is left with its initial value.

4.4 Part II: MM-Based Transmit Covariance Optimization Techniques

In this section, we are going to develop optimization strategies that provide local solutions to the nonconvex transmit covariance optimization problem that arises in multiuser MIMO SWIPT

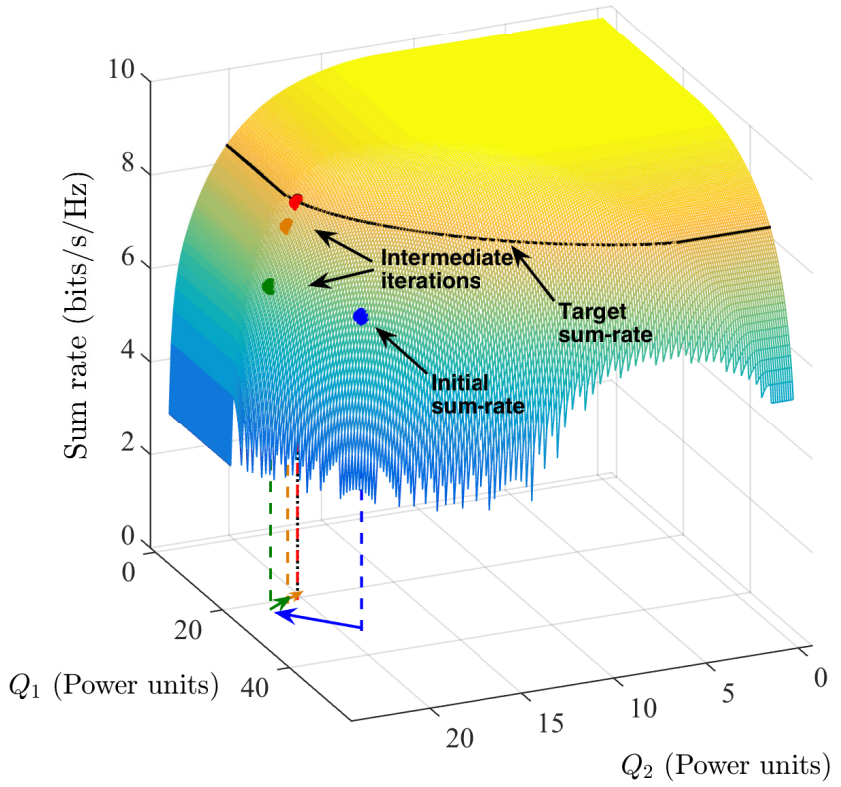


Figure 4.12: Performance of the proposed algorithm with minimum energy management based on (4.39).

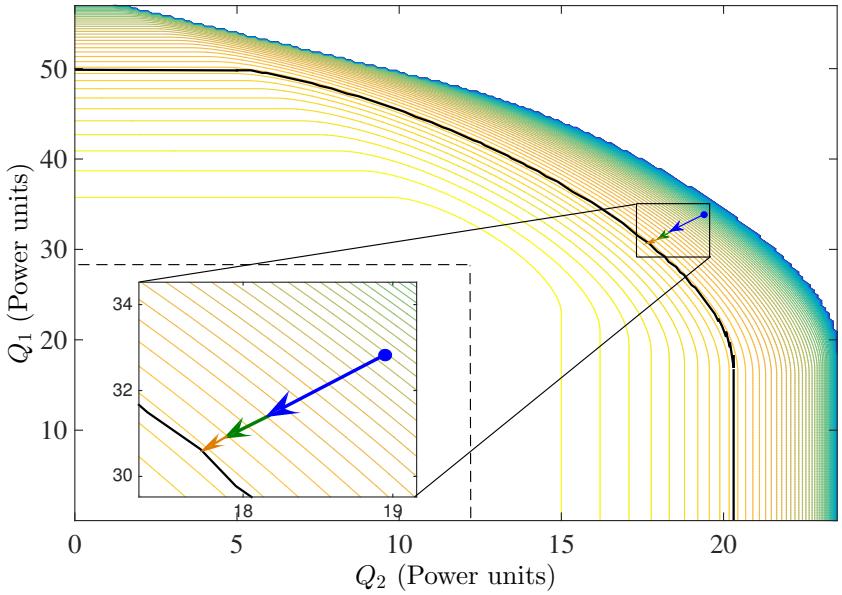


Figure 4.13: Performance of the proposed algorithm based on (4.35).

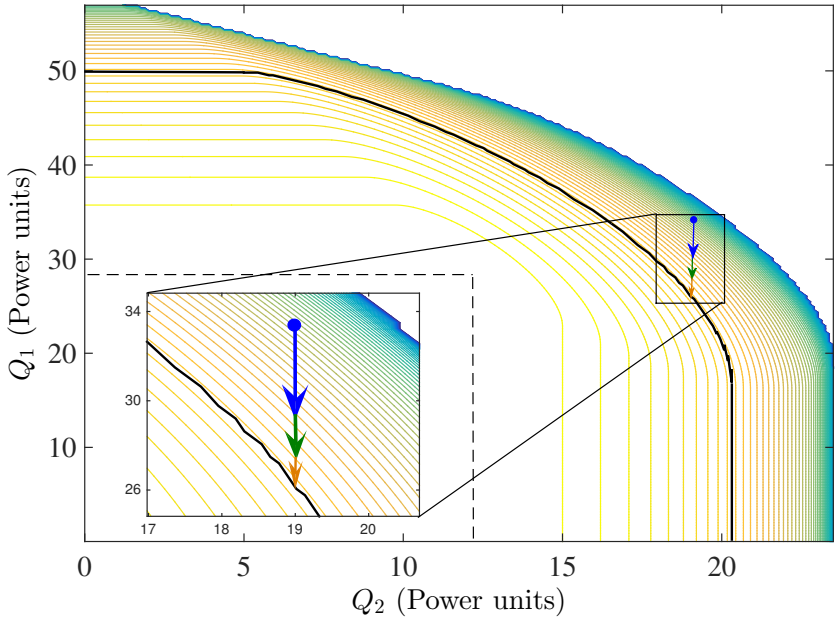


Figure 4.14: Performance of the proposed algorithm based on (4.37).

networks. In this setup, now, we do not apply the BD strategy and thus interference among users may be present, making the whole problem challenging and difficult to solve. We will consider that the user groups, $\mathcal{U}_I$ and $\mathcal{U}_E$, are fixed and known, i.e., the analysis will be focused on just one particular frame. In this section, we are going to generalize the formulation presented in the first part of this chapter and consider that both rates and harvested powers are optimized simultaneously. As we will see, the formulation considered in the first part of this chapter is a particular solution of the general formulation presented in this section.

4.4.1 Problem Formulation

The transmitter design that we propose in this section is modeled as a nonconvex multi-objective optimization problem. The goal is to maximize simultaneously the individual data rates and the harvested powers of the information and harvesting users, respectively. Given this and the previous system model, the optimization problem is written as

$$\begin{aligned} \underset{\{\mathbf{S}_i\}}{\text{maximize}} \quad & \Big((R_n(\mathbf{S}))_{n\in\mathcal{U}_I}, (E_m(\mathbf{S}))_{m\in\mathcal{U}_E} \Big) \\ \text{subject to} \quad & C1: \sum_{i\in\mathcal{U}_I} \text{Tr}(\mathbf{S}_i) \leq P_T \\ & C2: \mathbf{S}_i \succeq 0, \qquad\qquad \forall i \in \mathcal{U}_I, \end{aligned} \tag{4.41}$$

where we define the $|\mathcal{U}_I|-$tuple $\mathbf{S} \triangleq (\mathbf{S}_i)_{\forall i \in \mathcal{U}_I} = (\mathbf{S}_1, \ldots, \mathbf{S}_{|\mathcal{U}_I|})$, the data rate expression, $R_n(\mathbf{S})$, is given by

$$R_n(\mathbf{S}) = \log\det\left(\mathbf{I} + \mathbf{H}_n\mathbf{S}_n\mathbf{H}_n^H\mathbf{\Omega}_n^{-1}(\mathbf{S}_{-n})\right) \tag{4.42}$$

$$= \log\det\left(\mathbf{\Omega}_n(\mathbf{S}_{-n}) + \mathbf{H}_n\mathbf{S}_n\mathbf{H}_n^H\right) - \log\det\left(\mathbf{\Omega}_n(\mathbf{S}_{-n})\right) \tag{4.43}$$

$$= \underbrace{\log\det\left(\mathbf{I} + \mathbf{H}_n\bar{\mathbf{S}}\mathbf{H}_n^H\right)}_{\triangleq\, s_n(\mathbf{S})} - \underbrace{\log\det\left(\mathbf{\Omega}_n(\mathbf{S}_{-n})\right)}_{\triangleq\, g_n(\mathbf{\Omega}_n(\mathbf{S}_{-n}))}, \tag{4.44}$$

being $\bar{\mathbf{S}} = \sum_{k\in\mathcal{U}_I}\mathbf{S}_k$, where $\Omega_n(\mathbf{S}_{-n})$ was defined in (4.2), and the harvested power, $E_m(\mathbf{S})$, is given by

$$E_m(\mathbf{S}) = \sum_{i\in\mathcal{U}_I}\mathrm{Tr}(\mathbf{H}_m\mathbf{S}_i\mathbf{H}_m^H). \tag{4.45}$$

The previous multi-objective optimization problem in (4.41) is not convex due the objective functions (in particular, due to $\mathbf{\Omega}_i(\mathbf{S}_{-i})$) and is difficult to solve. In order to find Pareto optimal points, we can reformulate it by using any of the different techniques that we presented in Section 2.3. In the following, we propose two approaches based on the weighted sum method and on the hybrid method. For convenience, we start with the hybrid method as it is the one that has received the most attention in the literature [Zha13], [Liu13a] and it is the one considered in the first part of this chapter. Note, however, in that literature, the interference term in (4.42), $\mathbf{\Omega}_i(\mathbf{S}_{-i})$, is assumed to be removed by the transmission strategy through the application of BD. This assumption makes the problem convex and hence easier to solve.

4.4.1.1 Hybrid-Based Formulation to Solve (4.41)

In the hybrid approach some of the objective functions are collapsed into a single objective by means of scalarization and some of the objective functions are added as constraints (see Section 2.3.3.2). In particular, the data rates are left in the objective whereas the harvesting constraints are included as individual harvesting constraints. With this particular formulation, we are able to guarantee a minimum value for the power to be harvested by the harvesting users. Given that, problem (4.41) is formulated as

$$\begin{aligned}
\underset{\{\mathbf{S}_i\}}{\text{maximize}} \quad & \sum_{i\in\mathcal{U}_I}\omega_i\log\det\left(\mathbf{I} + \mathbf{H}_i\bar{\mathbf{S}}\mathbf{H}_i^H\right) - \omega_i\log\det\left(\mathbf{\Omega}_i(\mathbf{S}_{-i})\right) && (4.46)\\
\text{subject to} \quad & C1: \sum_{i\in\mathcal{U}_I}\mathrm{Tr}(\mathbf{H}_j\mathbf{S}_i\mathbf{H}_j^H) \geq Q_j, && \forall j \in \mathcal{U}_E\\
& C2: \sum_{i\in\mathcal{U}_I}\mathrm{Tr}(\mathbf{S}_i) \leq P_T\\
& C3: \mathbf{S}_i \succeq 0, && \forall i \in \mathcal{U}_I.
\end{aligned}$$

For simplicity in the notation, let us define the feasible set $\mathcal{S}_1$ as

$$\mathcal{S}_1 \triangleq \left\{ \mathbf{S} : \sum_{i \in \mathcal{U}_I} \text{Tr}(\mathbf{H}_j \mathbf{S}_i \mathbf{H}_j^H) \geq Q_j, \, \forall j \in \mathcal{U}_E, \, \sum_{i \in \mathcal{U}_I} \text{Tr}(\mathbf{S}_i) \leq P_T, \, \mathbf{S}_i \succeq 0, \forall i \in \mathcal{U}_I \right\}. \tag{4.47}$$

For a set of fixed harvesting constraints, the convex hull of the rate region can be obtained by varying the values of ω_i [Gol03]. In addition, we can use the values of the weights to assign priorities to some users if user scheduling is to be implemented, following, for example, the PF criterion [Liu10], [And01]. Note the similarities of problem (4.46) with the single user case presented in [Zha13] and its extension to the multiuser case presented in the first part of this chapter. As commented before, the novelty is that we do not force the transmitter to cancel the interference generated among the information users (using, for example, BD constraints [Spe04]) and, thus, we allow the system to have more degrees of freedom to improve the system throughput and the harvested power simultaneously. Later in Section 4.4.2.1, we will present a method based on MM to solve the previous nonconvex problem (see Section 2.4 for a review on the MM method), although only a local suboptimum will be achieved.

4.4.1.2 Weighted Sum-Based Formulation to Solve (4.41)

In situations where the exact amount of power to be harvested by harvesting users is not needed, we can also obtain Pareto optimal points by means of the simpler weighted-sum method. In this case, we can assign priorities so that some users tend to harvest more power than others, although the exact amounts cannot be controlled. As we will see later in this chapter, the positive aspect is that, with this new formulation, the overall problem is much easier to solve. The transmitter design is obtained through the following nonconvex optimization problem:

$$\begin{aligned} \underset{\{\mathbf{S}_i\}}{\text{maximize}} \quad & \sum_{i \in \mathcal{U}_I} \omega_i \log\det\left(\mathbf{I} + \mathbf{H}_i \bar{\mathbf{S}} \mathbf{H}_i^H\right) - \omega_i \log\det\left(\mathbf{\Omega}_i(\mathbf{S}_{-i})\right) + \sum_{j \in \mathcal{U}_E} \sum_{i \in \mathcal{U}_I} \alpha_j \, \text{Tr}(\mathbf{H}_j \mathbf{S}_i \mathbf{H}_j^H) \\ \text{subject to} \quad & C1 : \sum_{i \in \mathcal{U}_I} \text{Tr}(\mathbf{S}_i) \leq P_T \\ & C2 : \mathbf{S}_i \succeq 0, \quad \forall i \in \mathcal{U}_I, \end{aligned} \tag{4.48}$$

where α_j are some real non-negative weights. For simplicity in the notation, let us define the feasible set $\mathcal{S}_2$ as

$$\mathcal{S}_2 \triangleq \left\{ \mathbf{S} : \sum_{i \in \mathcal{U}_I} \text{Tr}(\mathbf{S}_i) \leq P_T, \, \mathbf{S}_i \succeq 0, \forall i \in \mathcal{U}_I \right\}. \tag{4.49}$$

As we will show later in Section 4.4.2.2, the resolution of (4.48) is easier than the resolution of (4.46). Hence, there is a trade-off in terms of speed of convergence of the algorithms and in terms of the harvested power control since, as we introduced before, in (4.46) the transmitter can fully control the amount of power to be harvested by the users whereas in (4.48) the transmitter can

only assign some priorities of the power to be harvested among the users through the weights α_j.

4.4.2 MM-based Techniques to Solve Problem (4.41)

In this section, we present a method based on the MM philosophy to solve problems (4.46) and (4.48). Since the original problems (4.46) and (4.48) are nonconvex, we apply the MM method and reformulate each surrogate (inner) problem and make them convex. This reformulation will follow two steps. In the first step, problems (4.46) and (4.48) will be convexified by using a linear approximation of the nonconcave terms. This is the approach taken in papers such as [Hon16], [Scu14], and [You14]. Instead of solving the reformulated (convex) problem, in the second step, we design a quadratic approximation of the remaining concave terms in order to find a surrogate problem easier to solve. Finally, we apply the MM method to the quadratic reformulation.

As benchmarks for comparison, we will consider the case of just convexifying linearly the nonconcave terms, approach taken in the literature, and also consider just a gradient method applied directly to the nonconvex problems (4.46) and (4.48).

Although the mathematical developments of the proposed MM approaches are more tedious than the approaches usually taken in the literature, the resulting algorithms are faster.

4.4.2.1 Approach to Solve the Hybrid Formulation in (4.46)

As we introduced before, we need to reformulate the original nonconvex problem (4.46) and make it convex. This will be done in two steps. Motivated by the work in [Scu14], in this first step, we derive a linear approximation for the nonconcave term (the rightmost one) of the objective function of (4.46), i.e., $f_0(\mathbf{S}) = \sum_{i\in\mathcal{U}_I} \omega_i s_i(\mathbf{S}) - \omega_i g_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}))$, in such a way that the modified problem is convex[14]. In order to find a concave lower bound of $f_0(\mathbf{S})$, $g_i(\cdot)$ can be upper bounded linearly at point $\mathbf{\Omega}_i^{(0)} = \sum_{\substack{k\in\mathcal{U}_I \\ k\neq i}} \mathbf{H}_i\mathbf{S}_k^{(0)}\mathbf{H}_i^H + \mathbf{I}$ as

$$g_i(\mathbf{\Omega}_i(\mathbf{S}_{-i})) \leq g_i\left(\mathbf{\Omega}_i^{(0)}\right) + \mathrm{Tr}\left(\left(\mathbf{\Omega}_i^{(0)}\right)^{-1}\left(\mathbf{\Omega}_i(\mathbf{S}_{-i}) - \mathbf{\Omega}_i^{(0)}\right)\right) \tag{4.50}$$

$$= \text{constant} + \mathrm{Tr}\left(\left(\mathbf{\Omega}_i^{(0)}\right)^{-1}\mathbf{\Omega}_i(\mathbf{S}_{-i})\right) \tag{4.51}$$

$$\triangleq \hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(0)}). \tag{4.52}$$

Even though problem (4.46) reformulated with the previous upper bound $\hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(0)})$ is convex, we want to go one step further and apply a quadratic lower bound for the leftmost term of $f_0(\mathbf{S})$, i.e., $s_i(\mathbf{S})$ in a way that the overall lower bound fulfills conditions (A1) – (A4)

[14] In fact, by applying the linear approximation, the overall objective function becomes concave.

presented before in Section 2.4 and the MM method can be invoked. Note that the upper bound $\hat{g}_i(\boldsymbol{\Omega}_i(\mathbf{S}_{-i}), \boldsymbol{\Omega}_i^{(0)})$ already fulfills the four conditions, (A1) – (A4). The idea of implementing this quadratic bound is to find a surrogate problem that is much simpler and easier to solve than the one obtained by just considering the linear bound[15] $\hat{g}_i(\boldsymbol{\Omega}_i(\mathbf{S}_{-i}), \boldsymbol{\Omega}_i^{(0)})$.

We now focus the attention on deriving the surrogate function for the leftmost term of $f_0(\mathbf{S})$, i.e., $s_i(\mathbf{S})$. In order for the surrogate problem to be easily solved, we force the surrogate function of $s_i(\mathbf{S})$ around $\bar{\mathbf{S}}^{(0)}$ to be quadratic, where $\bar{\mathbf{S}}^{(0)} = \sum_{k\in\mathcal{U}_I} \mathbf{S}_k^{(0)}$ and $\mathbf{S}_k^{(0)}$ represents the solution of the MM algorithm at the previous iteration. By doing this, as will be apparent later, the overall surrogate problem can be formulated as an SDP optimization problem, which can be easily solved.

Proposition 4.3. *A valid surrogate function for the function* $s_i(\bar{\boldsymbol{S}}) = \log\det\left(\boldsymbol{I} + \boldsymbol{H}_n\bar{\boldsymbol{S}}\boldsymbol{H}_n^H\right)$ *that satisfies conditions* $(A1) - (A4)$ *is*

$$\hat{s}_i(\bar{\boldsymbol{S}}, \bar{\boldsymbol{S}}^{(0)}) \triangleq \mathrm{Tr}\left(\boldsymbol{J}_i\bar{\boldsymbol{S}}\right) + \mathrm{Tr}\left(\bar{\boldsymbol{S}}^H\boldsymbol{M}_i\bar{\boldsymbol{S}}\right) + \kappa_1, \quad \forall\bar{\boldsymbol{S}},\ \bar{\boldsymbol{S}}^{(0)} \in \mathcal{S}_+^{n_T}, \tag{4.53}$$

with matrices $\boldsymbol{J}_i = \boldsymbol{G}_i - \bar{\boldsymbol{S}}^{(0),H}\boldsymbol{M}_i - \boldsymbol{M}_i\bar{\boldsymbol{S}}^{(0)}$, $\boldsymbol{G}_i = \boldsymbol{H}_i^H\left(\boldsymbol{I} + \boldsymbol{H}_i\bar{\boldsymbol{S}}^{(0)}\boldsymbol{H}_i^H\right)^{-1}\boldsymbol{H}_i$ *and* $\boldsymbol{M}_i = -\gamma_i\boldsymbol{I}$, *being* $\gamma_i \geq \frac{1}{2}\lambda_{\max}^2(\boldsymbol{H}_i^H\boldsymbol{H}_i)$, κ_1 *contains some terms that do not depend on* $\boldsymbol{S}$, *and* $\mathcal{S}_+^{n_T}$ *denotes the set of positive semidefinite matrices.*

Proof. See Appendix 4.F. ■

Let us now reformulate the optimization problem in (4.46) with the surrogate function $\hat{s}_i(\bar{\mathbf{S}}, \bar{\mathbf{S}}^{(0)}) - \hat{g}_i(\boldsymbol{\Omega}_i(\mathbf{S}_{-i}), \boldsymbol{\Omega}_i^{(0)})$:

$$\mathrm{Tr}\left(\mathbf{E}_i\bar{\mathbf{S}}\right) + \mathrm{Tr}\left(\bar{\mathbf{S}}^H\mathbf{M}_i\bar{\mathbf{S}}\right) + \mathrm{Tr}\left(\mathbf{R}_i\mathbf{S}_i\right) + \kappa_2, \tag{4.54}$$

where $\mathbf{R}_i = \mathbf{H}_i^H\left(\boldsymbol{\Omega}_i^{(0)}\right)^{-1}\mathbf{H}_i \in \mathbb{C}^{n_T\times n_T}$, $\mathbf{E}_i = \mathbf{J}_i - \mathbf{R}_i$, and κ_2 contains some terms that do not depend on $\mathbf{S}$. Thus, problem (4.46) can be reformulated as

$$\begin{aligned} \underset{\{\mathbf{S}_i\}}{\text{maximize}} \quad & \sum_{i\in\mathcal{U}_I}\omega_i\left(\mathrm{Tr}\left(\mathbf{E}_i\bar{\mathbf{S}}\right) + \mathrm{Tr}\left(\bar{\mathbf{S}}^H\mathbf{M}_i\bar{\mathbf{S}}\right) + \mathrm{Tr}\left(\mathbf{R}_i\mathbf{S}_i\right)\right) - \rho\left\|\mathbf{S}_i - \mathbf{S}_i^{(0)}\right\|_F^2 \\ \text{subject to} \quad & \mathbf{S} \in \mathcal{S}_1, \end{aligned} \tag{4.55}$$

where we have added a proximal quadratic term to the surrogate function in which ρ is any non-negative constant that can be tuned by the algorithm. This term that can be adjusted through numerical simulation, provides more flexibility in the algorithm design stage and may

[15]The surrogate problem obtained by just applying the bound $\hat{g}_i(\boldsymbol{\Omega}_i(\mathbf{S}_{-i}), \boldsymbol{\Omega}_i^{(0)})$ will be used as benchmark. The specific mathematical details of the optimization problem and the algorithm will be described in Appendix 4.I.

help to speed up the convergence. By performing some mathematical manipulations, we are able to obtain the following result:

Proposition 4.4. *The optimization problem presented in* (4.46) *can be solved based on the MM method by solving recursively the following SDP problem:*

$$\begin{aligned} \underset{\{\boldsymbol{S}_i\},\, \boldsymbol{s}, t}{\textit{minimize}} \quad & t & (4.56)\\ \textit{subject to} \quad & C1: \begin{bmatrix} t\boldsymbol{I} & \tilde{\boldsymbol{C}}^{\frac{1}{2}}\boldsymbol{s}-\boldsymbol{c} \\ \left(\tilde{\boldsymbol{C}}^{\frac{1}{2}}\boldsymbol{s}-\boldsymbol{c}\right)^H & 1 \end{bmatrix} \succeq 0 \\ & C2: \boldsymbol{T}_i\boldsymbol{s} = \operatorname{vec}(\boldsymbol{S}_i), \qquad \forall i \in \mathcal{U}_I \\ & C3: \boldsymbol{S} \in \mathcal{S}_1, \end{aligned}$$

where $\boldsymbol{s} = \left[\operatorname{vec}(\boldsymbol{S}_1)^T \operatorname{vec}(\boldsymbol{S}_2)^T \dots \operatorname{vec}(\boldsymbol{S}_N)^T\right]^T \in \mathbb{C}^{n_T n_T |\mathcal{U}_I| \times 1}$*, t is a dummy variable, and* $\tilde{\boldsymbol{C}}^{\frac{1}{2}}$*,* $\boldsymbol{T}_i$*, and* $\boldsymbol{c}$ *are some constant matrices and vectors computed as shown in Appendix 4.G. Vector* $\boldsymbol{c}$ *depends on matrix* $\bar{\boldsymbol{S}}^{(0)}$*.*

Proof. See Appendix 4.G. ■

The final algorithm is presented in Algorithm 4.8.

Algorithm 4.8 Algorithm for solving problem (4.46)

1: initialize $\mathbf{S}^{(0)} \in \mathcal{S}_1$. Set $k = 0$
2: **repeat**
3: compute $\mathbf{c}$ with $\mathbf{S}^{(k)}$, given in (4.101)
4: generate the $(k+1)$-tuple $(\mathbf{S}_i^\star)_{\forall i \in \mathcal{U}_I}$ by solving the SDP in (4.56)
5: set $\mathbf{S}_i^{(k+1)} = \mathbf{S}_i^\star$, $\forall i \in \mathcal{U}_I$, and set $k = k+1$
6: **until** convergence is reached
7: **end algorithm**

4.4.2.2 Approach to Solve the Sum Method Formulation in (4.48)

Let us start the development by reformulating problem (4.48) as

$$\begin{aligned} \underset{\{\mathbf{S}_i\}}{\text{maximize}} \quad & \sum_{i \in \mathcal{U}_I} \omega_i \left(s_i(\mathbf{S}) - g_i(\boldsymbol{\Omega}_i(\mathbf{S}_{-i}))\right) + \sum_{i \in \mathcal{U}_I} \operatorname{Tr}(\mathbf{R}_H \mathbf{S}_i) & (4.57)\\ \text{subject to} \quad & \mathbf{S} \in \mathcal{S}_2, \end{aligned}$$

where $\mathbf{R}_H = \sum_{j\in\mathcal{U}_E} \alpha_j \mathbf{H}_j^H \mathbf{H}_j$.The middle and the rightmost terms of the objective function of (4.57) are convex (in fact they are linear) whereas the leftmost term is not convex. Let us apply the same procedure that we applied before but with a slight modification. Previously in (4.52), we found that $g_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}))$ could be approximated by $\hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(0)}) = \mathrm{Tr}\left(\left(\mathbf{\Omega}_i^{(0)}\right)^{-1}\mathbf{\Omega}_i(\mathbf{S}_{-i})\right)$ (omitting the constant term). Now, as the objective function is different from the one in problem (4.46), the goal is to find a surrogate function (i.e., lower bound) for the function $s_i(\mathbf{S})$ that allows us to find efficiently a solution for the surrogate problem (i.e., the inner iteration defined by the MM algorithm).

Proposition 4.5. *A valid surrogate function for the function* $s_i(\boldsymbol{S}) = \log\det\left(\boldsymbol{I} + \boldsymbol{H}_i \sum_{k\in\mathcal{U}_I} \boldsymbol{S}_k \boldsymbol{H}_i^H\right)$ *that satisfies conditions* $(A1)-(A4)$ *is*

$$\hat{s}_i(\boldsymbol{S}, \boldsymbol{S}^{(0)}) \triangleq \sum_{\ell\in\mathcal{U}_I} \mathrm{Tr}\left(\boldsymbol{J}_i \boldsymbol{S}_\ell\right) + \sum_{\ell\in\mathcal{U}_I} \mathrm{Tr}\left(\boldsymbol{S}_\ell^H \boldsymbol{M}_i \boldsymbol{S}_\ell\right) + \kappa_3, \quad \forall \boldsymbol{S}_\ell,\, \boldsymbol{S}_\ell^{(0)} \in \mathcal{S}_+^{n_T}, \tag{4.58}$$

with matrices $\boldsymbol{J}_i = \boldsymbol{G}_i - \boldsymbol{S}_\ell^{(0),H}\boldsymbol{M}_i - \boldsymbol{M}_i\boldsymbol{S}_\ell^{(0)}$, $\boldsymbol{G}_i = \boldsymbol{H}_i^H\left(\boldsymbol{I} + \boldsymbol{H}_i\sum_{k\in\mathcal{U}_I}\boldsymbol{S}_k^{(0)}\boldsymbol{H}_i^H\right)^{-1}\boldsymbol{H}_i$, *and* $\boldsymbol{M}_i = -\xi_i\boldsymbol{I}$, *being* $\xi_i \geq \frac{1}{2}|\mathcal{U}_I|^2\lambda_{\max}^2(\boldsymbol{H}_i^H\boldsymbol{H}_i)$, *and* κ_3 *contains the constant terms that do not depend on* $\boldsymbol{S}$.

Proof. See Appendix 4.H. ■

Remark 4.1. *Note that the two surrogate functions* (4.53) *and* (4.58) *have the same form but with a difference in the quadratic term. Notice that surrogate function* (4.58) *is tighter than* (4.53) *and with cross-products. As will be shown later, this will allow us to decouple the optimization problem for each information user* i *and, thus, solve all problems in parallel. On the other hand, thanks to the fact that surrogate function* (4.53) *is looser than* (4.58)*, a faster convergence can be obtained than if surrogate function* (4.58) *were to be applied in problem* (4.46).

Let us now reformulate problem (4.57) with the lower bound that we just found (omitting the constant terms):

$$\begin{aligned}\underset{\{\mathbf{S}_i\}}{\text{maximize}} \quad & \sum_{i\in\mathcal{U}_I}\mathrm{Tr}\left(\check{\mathbf{J}}_i\mathbf{S}_i\right) + \sum_{i\in\mathcal{U}_I}\mathrm{Tr}\left(\mathbf{S}_i^H\check{\mathbf{M}}\mathbf{S}_i\right) - \sum_{i\in\mathcal{U}_I}\mathrm{Tr}\left(\mathbf{R}_i\sum_{\substack{k\in\mathcal{U}_I\\k\neq i}}\mathbf{S}_k\right) + \sum_{i\in\mathcal{U}_I}\mathrm{Tr}(\mathbf{R}_H\mathbf{S}_i) \\ \text{subject to} \quad & \mathbf{S}\in\mathcal{S}_2,\end{aligned} \tag{4.59}$$

where $\check{\mathbf{J}}_i = \check{\mathbf{G}} - \mathbf{S}_i^{(0),H}\check{\mathbf{M}} - \check{\mathbf{M}}\mathbf{S}_i^{(0)}$, with $\check{\mathbf{M}} = \sum_{k\in\mathcal{U}_I}\omega_k\mathbf{M}_k$ and $\check{\mathbf{G}} = \sum_{k\in\mathcal{U}_I}\omega_k\mathbf{G}_k$. Note that we have arranged the indices to make the notation easier to follow and consistent with the original notation. We can further simplify the objective function by grouping terms considering that matrix $\check{\mathbf{M}}$ is proportional to the identity matrix, i.e., $\check{\mathbf{M}} = -\beta\mathbf{I}$, being $\beta \geq$

$\frac{1}{2}|\mathcal{U}_I|^2 \sum_{k\in\mathcal{U}_I} \omega_k \lambda_{\max}^2(\mathbf{H}_k^H \mathbf{H}_k)$:

$$\begin{aligned} \underset{\{\mathbf{S}_i\}}{\text{minimize}} \quad & \beta \sum_{i\in\mathcal{U}_I} \mathrm{Tr}\left(\mathbf{S}_i^H \mathbf{S}_i\right) - \sum_{i\in\mathcal{U}_I} \mathrm{Tr}\left(\mathbf{F}_i \mathbf{S}_i\right) \\ \text{subject to} \quad & \mathbf{S} \in \mathcal{S}_2, \end{aligned} \tag{4.60}$$

where

$$\mathbf{F}_i = \check{\mathbf{J}}_i - \sum_{\substack{k\in\mathcal{U}_I \\ k\neq i}} \mathbf{R}_k + \mathbf{R}_H. \tag{4.61}$$

Note that we have changed the sign of the objective function and reformulated the problem as a minimization one. The idea is to find a closed-form expression for the optimum covariance matrices $\{\mathbf{S}_i\}$. If we dualize constraint $C1$ and form a partial Lagrangian, we obtain the following optimization problem:

$$\begin{aligned} \underset{\{\mathbf{S}_i\}}{\text{minimize}} \quad & \beta \sum_{i\in\mathcal{U}_I} \mathrm{Tr}\left(\mathbf{S}_i^H \mathbf{S}_i\right) - \sum_{i\in\mathcal{U}_I} \mathrm{Tr}\left(\mathbf{W}_i(\mu) \mathbf{S}_i\right) \\ \text{subject to} \quad & \mathbf{S}_i \succeq 0, \quad \forall i \in \mathcal{U}_I, \end{aligned} \tag{4.62}$$

where $\mathbf{W}_i(\mu) = \mathbf{F}_i - \mu\mathbf{I}$, being $\mu \geq 0$ the Lagrange multiplier associated with constraint $C1$ of problem (4.57). The previous problem is clearly separable for each user i. Thus, for each information user, problem (4.62) is equivalent to solving the following projection problem:

$$\begin{aligned} \underset{\mathbf{S}_i}{\text{minimize}} \quad & \left\| \sqrt{\beta}\mathbf{S}_i - \check{\mathbf{W}}_i(\mu) \right\|_F \\ \text{subject to} \quad & \mathbf{S}_i \succeq 0, \end{aligned} \tag{4.63}$$

where $\check{\mathbf{W}}_i(\mu) = \frac{1}{2\sqrt{\beta}} \mathbf{W}_i(\mu) = \frac{1}{2\sqrt{\beta}} (\mathbf{F}_i - \mu\mathbf{I})$. The previous result is very nice as the solution of (4.63) is simple and elegant, thanks to the fact that problem (4.63) is a projection onto the semidefinite cone and has closed-form solution [Hen11]. Let the EVD of matrix $\mathbf{F}_i$ be $\mathbf{F}_i = \mathbf{U}_{F_i} \mathbf{\Lambda}_{F_i} \mathbf{U}_{F_i}^H$. The explicit expression of $\mathbf{S}_i^\star(\mu)$ is, thus, given by

$$\mathbf{S}_i^\star(\mu) = \frac{1}{\sqrt{\beta}} [\check{\mathbf{W}}_i(\mu)]^+ = \frac{1}{2\beta} [\mathbf{F}_i - \mu\mathbf{I}]^+ = \frac{1}{2\beta} \mathbf{U}_{F_i}^H [\mathbf{\Lambda}_{F_i} - \mu\mathbf{I}]^+ \mathbf{U}_{F_i}, \quad \forall i \in \mathcal{U}_I, \tag{4.64}$$

where $\lambda_k([\mathbf{X}]^+) = \min(0, \lambda_k(\mathbf{X}))$, being $\lambda_k(\mathbf{X})$ the k-th eigenvalue of matrix $\mathbf{X}$. Now it remains to compute the optimal Lagrange multiplier μ. This can be found by means of the simple bisection method fulfilling $\sum_{i\in\mathcal{U}_I} \mathrm{Tr}\left([\mathbf{\Lambda}_{F_i} - \mu\mathbf{I}]^+\right) = 2\beta P_T$. It turns out that, at each inner iteration, we need to compute a single EVD per information user, that is, the EVD of $\mathbf{F}_i$, and a few iterations to find the optimal multiplier μ. Note that the surrogate problem can be solved straightforwardly with the previous steps. The final algorithm looks as follows:

Algorithm 4.9 Algorithm for solving problem (4.48)

1: initialize $\mathbf{S}^{(0)} \in \mathcal{S}_2$. Set $k = 0$
2: **repeat**
3: compute $\mathbf{F}_i$ with matrix $\mathbf{S}_i^{(k)}$, $\forall i \in \mathcal{U}_I$, given in (4.61)
4: compute EVD of $\mathbf{F}_i = \mathbf{U}_{F_i}\mathbf{\Lambda}_{F_i}\mathbf{U}_{F_i}^H, \quad \forall i \in \mathcal{U}_I$
5: compute $\mu^\star$ such that $\sum_{i\in\mathcal{U}_I} \mathrm{Tr}\left([\mathbf{\Lambda}_{F_i} - \mu^\star\mathbf{I}]^+\right) = 2\beta P_T$
6: compute $\mathbf{S}_i^\star(\mu^\star) = \frac{1}{2\beta}[\mathbf{F}_i - \mu^\star\mathbf{I}]^+, \quad \forall i \in \mathcal{U}_I$
7: set $\mathbf{S}_i^{(k+1)} = \mathbf{S}_i^\star(\mu^\star), \forall i \in \mathcal{U}_I$, and set $k = k + 1$
8: **until** convergence is reached
9: **end algorithm**

4.4.2.3 Approaches Used as Benchmarks for Performance Comparison

As the problem introduced in (4.41) has not been addressed before in the literature, there are not specific benchmarks to compare our approaches with. For this reason, in this section, we propose some benchmark algorithms that will be used in the simulation section to compare the performance of the proposed MM approaches. These benchmarks are:

- Gradient-based algorithms based on [Boy08, Sec. 7] applied directly to the nonconvex problems (4.46) and (4.48). As an example, the gradients for the problem (4.46) are presented in Appendix 4.J.
- MM approaches considering only the linear approximation (4.52), i.e., $\hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(0)})$, but it was applied to problems (4.46) and (4.48) to find an overall quadratic lower bound. The specific optimization problems and algorithms can be found in Appendix 4.I.

4.4.3 Numerical Simulations of the MM Strategies and Comparison with the BD-Based Techniques

In this section, we evaluate the performance of the previous algorithms. In the first part of this section, we present some convergence and computational time results. For the simulations, we consider a system composed of 1 transmitter with 6 antennas, and 3 information users and 3 harvesting users with 2 antennas each. In the second part of the section, we show the performance of the proposed methods compared to the classical BD-based approach in terms of rates and harvested powers. In this case, for ease of presenting the information, we assume a system composed of 1 transmitter with 4 antennas, and 2 information users and 2 harvesting users with 2 antennas each. The simulation parameters common to both scenarios are the following: the

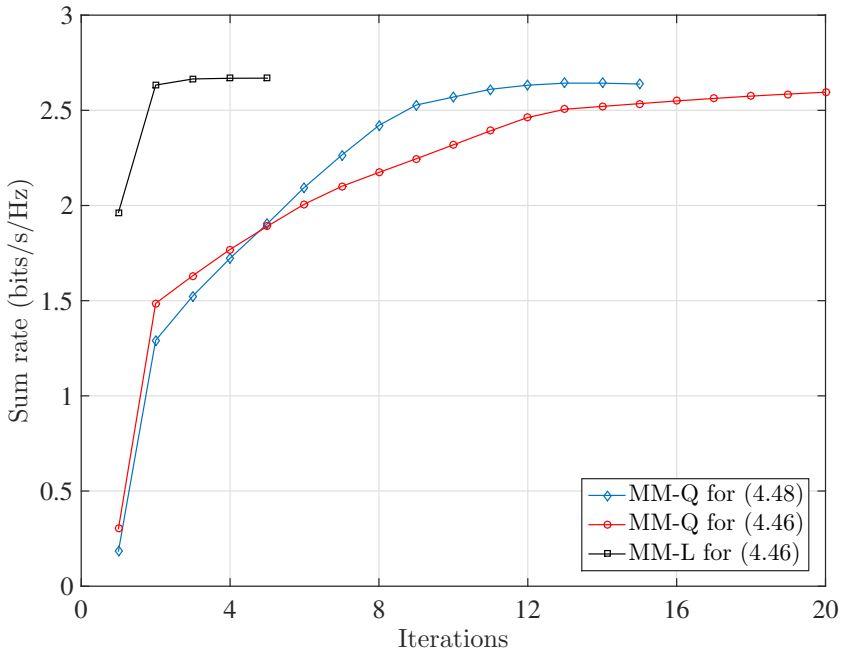


Figure 4.15: Convergence of the system sum rate vs number of iterations for three different approaches.

maximum radiated power is $P_T = 1$ W; the channel matrices are generated randomly with i.i.d. entries distributed according to $\mathcal{CN}(0, 1)$; and the weights ω_i are set to 1.

4.4.3.1 Convergence Evaluation

In this subsection, we evaluate the convergence behavior and the computational time of the methods presented in Sections 4.4.2.1 and 4.4.2.2 and the benchmark approaches presented in Section 4.4.2.3. The benchmark method for problem (4.48) presented in Appendix 4.I will not be evaluated as it is clearly worse[16] than the one presented in Section 4.4.2.2. In the figures, the legend is interpreted as follows: 'MM-L for (4.46)' refers to the method developed in Appendix 4.I for problem (4.46), 'MM-Q for (4.46)' refers to the method in Section 4.4.2.1, and 'MM-Q for (4.48)' refers to the method in Section 4.4.2.2. In order to compare all methods, we set the values of α_j and the values of Q_j so that the same system sum rate is achieved. These values are: $\boldsymbol{\alpha} = [1, 5, 10]$, and $\mathbf{Q} = [3.8, 7.2, 6.4]$ power units. Software package CVX is used to solve problem (4.129) [Gra13], and SeDuMi solver is used to solve problem (4.56) [Stu].

Figure 4.15 presents the sum rate convergence as a function of iterations. The three approaches converge to the same sum rate value but requiring a different number of iterations. In fact, the required number of iterations depends on how well the surrogate function approximates the original function. Note that, the surrogate function used in the 'MM-L for (4.46)' approach is the one that best approximates the objective function and, thus, fewer iterations are needed.

Figure 4.16 shows the computational time required by the three previous methods. We

[16]However, it was included in the paper for the sake of completeness

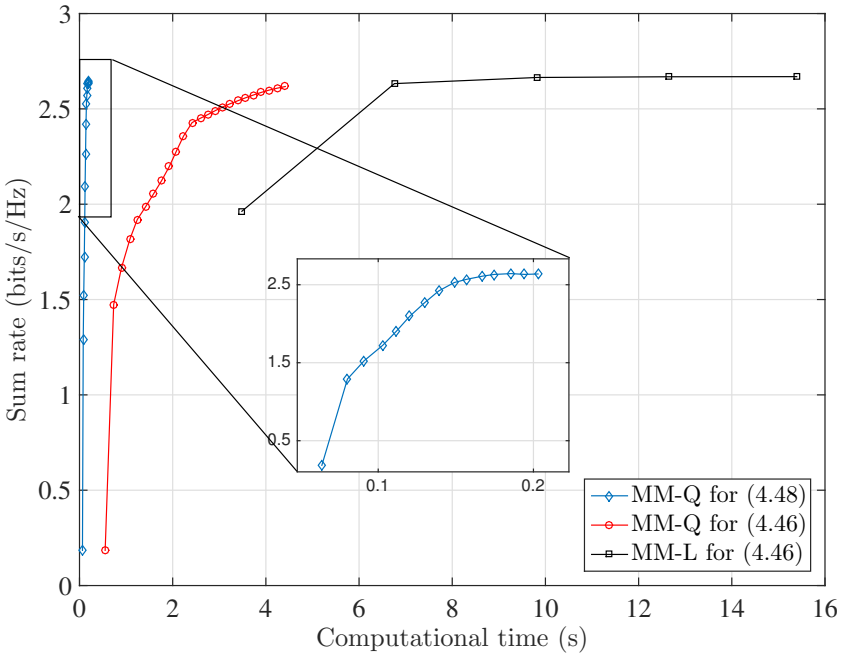


Figure 4.16: Convergence of the system sum rate vs computational time for three different approaches.

see that the 'MM-Q for (4.48)' method converges much faster than the other two approaches, as expected. The 'MM-Q for (4.46)' approach requires more iterations than the 'MM-L for (4.46)' approach but each iteration is solved faster since a specific algorithm can be employed to solve the convex optimization problem. Hence, if harvesting constraints are required to be fully controlled, i.e., the hybrid approach is required (Section 4.4.1.1), then the quadratic surrogate function, i.e., 'MM-Q for (4.46)' algorithm, is the best option.

For the sake of comparison and completeness, we also show in Figures 4.17 and 4.18 the convergence and the computational time for the gradient-like benchmark approaches. The plot legend reads as follows: 'GRAD for (4.46)' and 'GRAD for (4.48)' refers to a gradient approach applied to problems (4.46) and (4.48), respectively. 'all ones' and 'identity' mean that covariance matrices are initialized using an all ones matrix and the identity matrix, respectively. Results show that the proposed MM approaches are one to two orders of magnitude faster than the gradient-based methods.

4.4.3.2 Performance Evaluation

In this section, we evaluate the performance of the MM approach as compared to classical BD strategy considered in the literature. In order to show how harvesting users at different distances affect the performance, we have generated channel matrices with different norms. We would like to emphasize that, as the noise and channels are normalized, we will refer to the powers harvested by the receivers in terms of power units instead of Watts.

Figures 4.19 and 4.20 show the rate-power surface, that is, the multidimensional trade-off

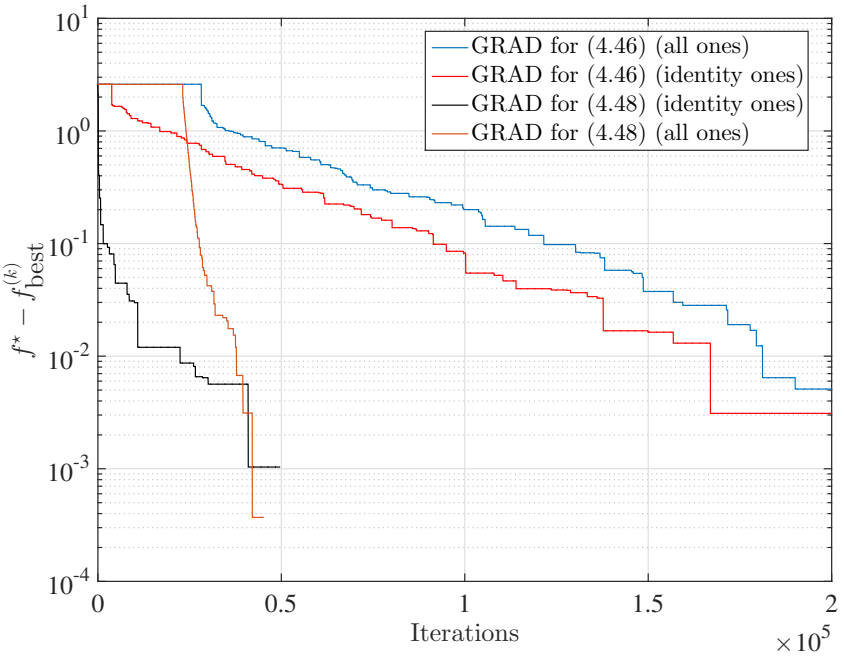


Figure 4.17: Convergence of the system sum rate vs iterations for a gradient approach for constrained optimization.

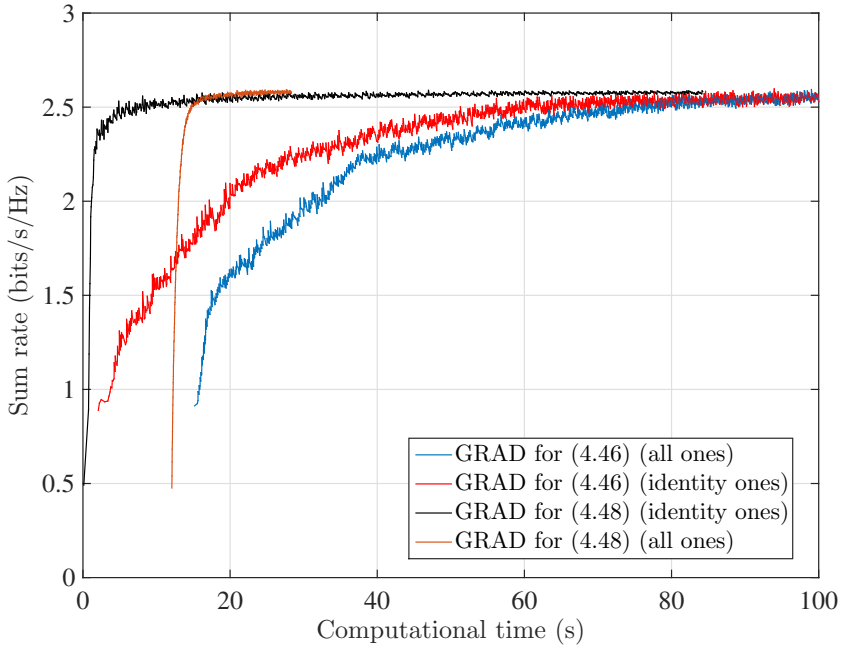


Figure 4.18: Convergence of the system sum rate vs computational time for a gradient approach for constrained optimization.

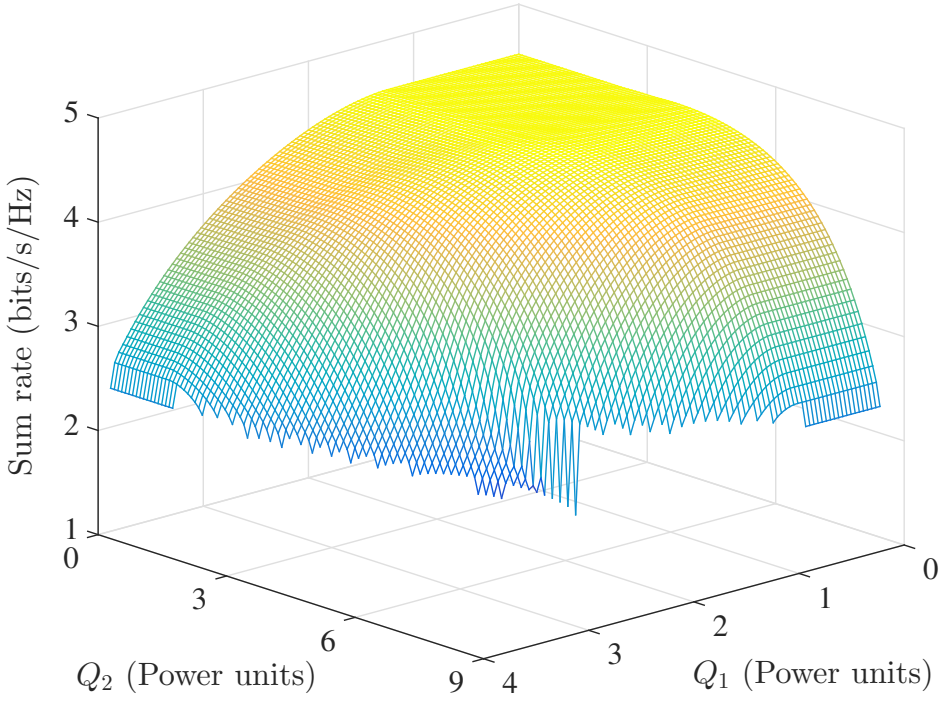


Figure 4.19: Rate-power surface for the MM method.

between the system sum rate and the powers to be collected by harvesting users (see Section 4.3.2.4). As we see, the MM approach outperforms the BD strategy in both terms, sum rate and harvested powers. The maximum system sum rate obtained with the MM approach when Q_1 and Q_2 are set to 0 is 4.5 bit/s/Hz, wheres the sum rate obtained with the BD approach is 2.75 bit/s/Hz. The rate-power surfaces are generated by varying the values of $\{Q_j\}$ in problem (4.46) or, equivalently, by varying the values of $\{\alpha_j\}$ in problem (4.48). A way to reduce the computational complexity associated to the generation of the rate-power surface is to use as initialization point the solution that was obtained for the previous values of $\{Q_j\}$ or $\{\alpha_j\}$ to generate the new value of the curve [Boy10]. Note, however, that the whole rate-power surface need not be generated for each transmission as it is just the representation of the existing rate-power tradeoff.

In order to clearly see the benefits in terms of collected power, Figures 4.21 and 4.22 show the contour plots of the previous 3D plots. We observe that, through the application of the MM-based techniques, users collect roughly 50% more power than the power collected by users when applying the BD strategy.

Finally, Figure 4.23 presents the rate-region of the MM approach for different values of $\{Q_j\}$. The same value of Q_j is set to the two harvesting users. In this case, we vary the values of ω_i to achieve the whole contour of the rate regions. We observe that, the larger the harvesting constraints, the smaller the rate-region, as expected. However, the relation between the harvesting constraints and the rate-region is not linear. As the harvesting constraints increase, a small change in the $\{Q_j\}$ produces a large reduction of the rate-region. This is because the

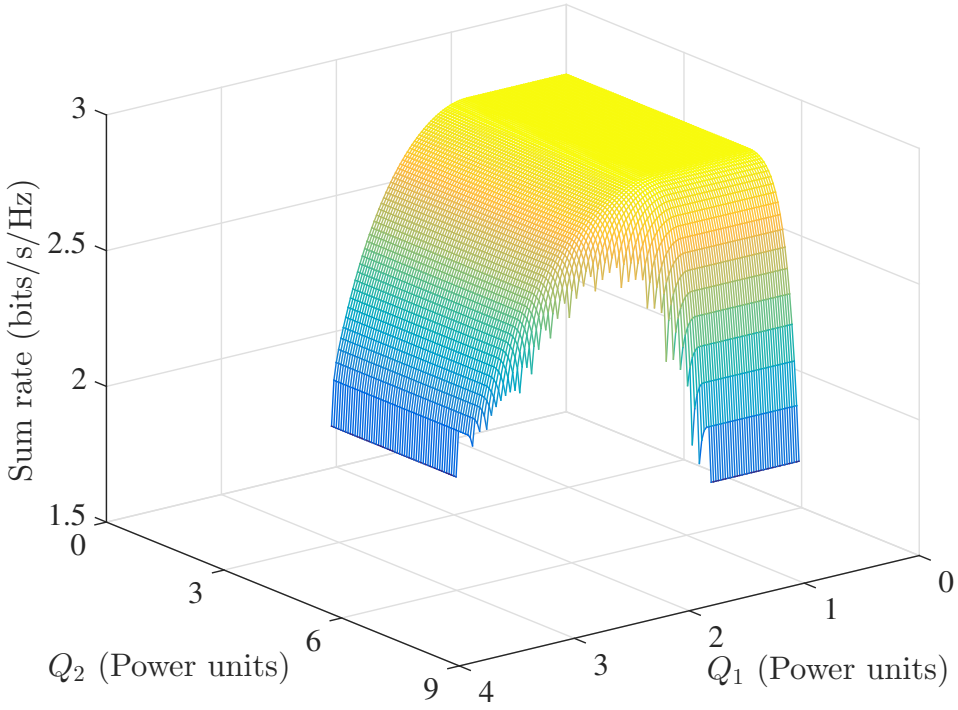


Figure 4.20: Rate-power surface for the BD method.

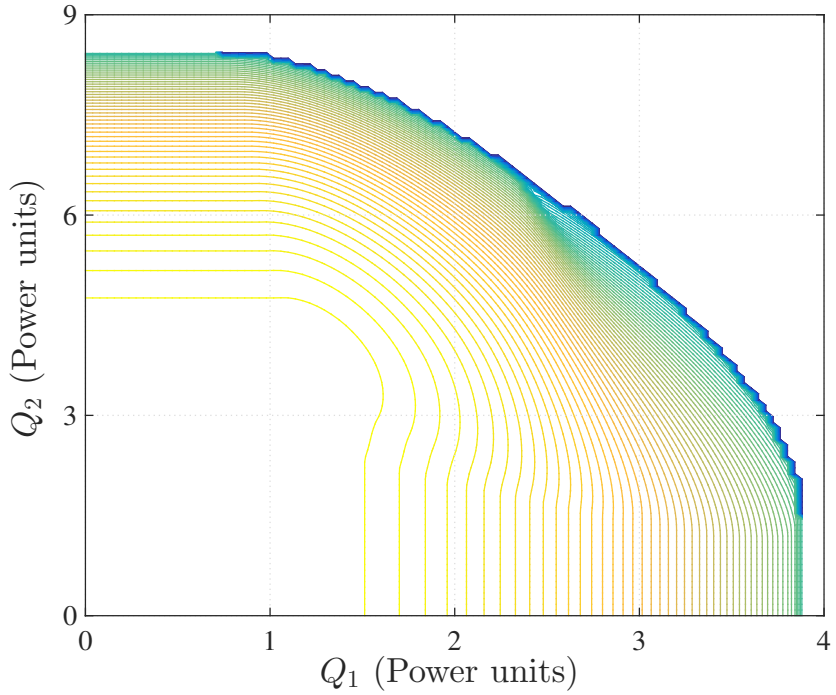


Figure 4.21: Contour of rate-power surface for the MM method.

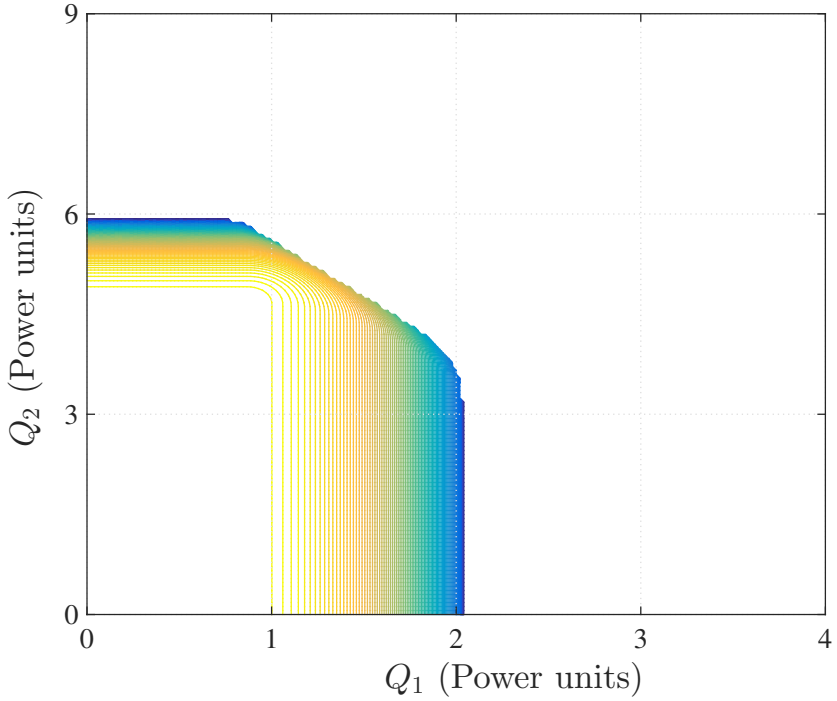


Figure 4.22: Contour of rate-power surface for the BD method.

3D rate-power surfaces presented before are not planes.

4.5 Chapter Summary and Conclusions

This chapter has studied the transmit covariance design that arises in multiuser multi-stream broadcast MIMO SWIPT networks.

In the first part of the chapter, we used the BD strategy in which interference among users is pre-canceled at the transmitter so that the resulting transmit covariance problem is convex. In this scenario, we derived the particular structure of the optimal transmit covariance matrices and particularized the scenario where only information or harvesting users were present in the system and where both types of users coexist in the system. When only harvesting users were considered, the problem was reformulated as a feasibility problem and we proposed some strategies for the cases where the original problem turned out to be infeasible. Then, we addressed the multidimensional trade-off between the sum-rate and the harvesting constraints in the general case. We showed that energy beamforming was optimal in case that the power harvested by one particular user was to be maximized. Later, we presented some user grouping techniques that allow the BS to select which users are better suited for information and which ones for battery replenishment in each particular frame for the case where both types of users are present in the system. We proposed two different scheduling techniques based on a different level of computational complexity. In the first approach, we selected the information and the harvesting users separately. In the second approach, the selection of both types of users was performed

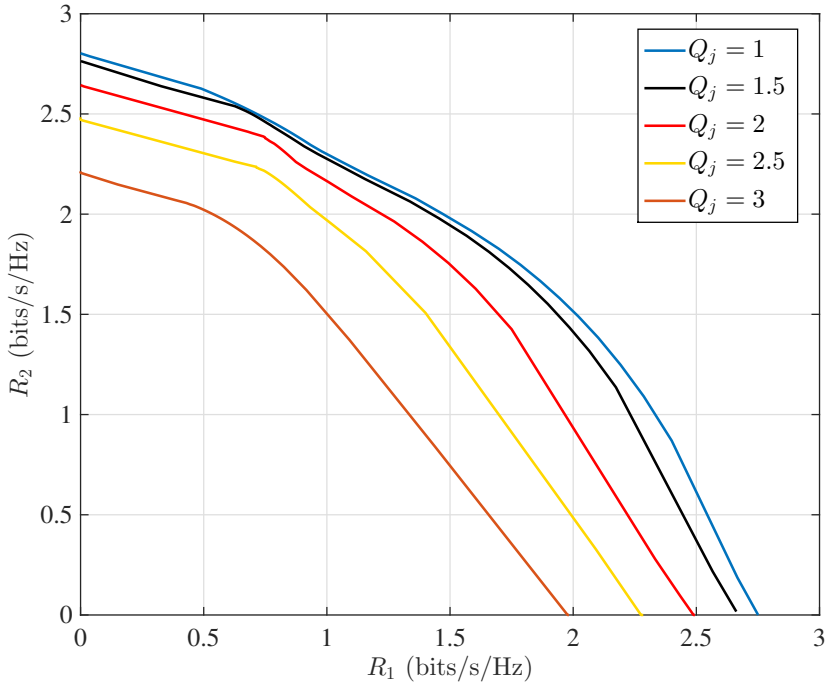


Figure 4.23: Rate region for different values of Q_j (in power units).

jointly. Simulation results showed that the aggregated throughput can be considerably improved if the proposed grouping strategy is implemented when the results are compared with traditional scheduling approaches. Finally, we have proposed different strategies for managing the minimum energy to be harvested. The procedures were derived from the sensitivity analysis of duality theory where we considered the effect on the system performance increase or decrease when adjusting the harvesting constraints.

In the second part of the chapter, we have presented a method to solve the difficult nonconvex problem that arises in multiuser multi-stream broadcast MIMO SWIPT networks when BD is not forced, i.e., interference among users is not cancelled. We formulated the general SWIPT problem as a multi-objective optimization problem, in which rates and harvested powers were to be optimized simultaneously. Then, we proposed two different formulations to obtain solutions of the general multi-objective optimization problem depending on the desired level of control of the power to be harvested. In the first approach, the transmitter was able to control the specific amount of power to be harvested by each user whereas in the second approach only the priorities to harvest power among the different users could be controlled. Both (nonconvex) formulations were solved based on the MM approach. We derived a convex approximation for two nonconvex objectives and developed two different algorithms. Simulation results showed that the proposed methods outperform the classical BD in terms of both system sum rate and power collected by users by a factor of approximately 50%. Moreover, the computational time needed to achieve convergence was shown to be really low for the approach in which the transmitter could only control the priorities of powers to be harvested (around two orders of magnitude lower than a gradient-like approach).

4.A Proof of Theorem 4.1

The Lagrangian of problem (4.23) is

$$\mathcal{L}(\tilde{\bar{\mathbf{S}}};\boldsymbol{\lambda},\mu) = -\sum_{i\in\mathcal{U}_I}\omega_i \log\det\left(\mathbf{I}+\hat{\mathbf{H}}_i\tilde{\mathbf{S}}_i\hat{\mathbf{H}}_i^H\right)+\sum_{j\in\mathcal{U}_E}\lambda_j\left(Q_j-\sum_{i\in\mathcal{U}_I}\mathrm{Tr}(\hat{\mathbf{H}}_{ji}\tilde{\mathbf{S}}_i\hat{\mathbf{H}}_{ji}^H)\right) + \mu\left(\sum_{i\in\mathcal{U}_I}\mathrm{Tr}(\tilde{\mathbf{S}}_i)-P_T\right) \tag{4.65}$$

where $\tilde{\bar{\mathbf{S}}} = (\tilde{\mathbf{S}}_i)_{\forall i\in\mathcal{U}_I}$ and we have omitted constraint $C3$. The previous Lagrangian can be manipulated and transformed into

$$\mathcal{L}(\tilde{\bar{\mathbf{S}}};\boldsymbol{\lambda},\mu) = -\sum_{i\in\mathcal{U}_I}\omega_i \log\det\left(\mathbf{I}+\hat{\mathbf{H}}_i\tilde{\mathbf{S}}_i\hat{\mathbf{H}}_i^H\right)+\sum_{i\in\mathcal{U}_I}\mathrm{Tr}\left(\mathbf{A}_i\tilde{\mathbf{S}}_i\right)+G, \tag{4.66}$$

where $G=\sum_{j\in\mathcal{U}_E}\lambda_j Q_j-\mu P_T$ and $\mathbf{A}_i=\mu\mathbf{I}-\sum_{j\in\mathcal{U}_E}\lambda_j\hat{\mathbf{H}}_{ji}^H\hat{\mathbf{H}}_{ji}$. The dual function of problem (4.23) is defined as $g(\boldsymbol{\lambda},\mu)=\min_{\tilde{\mathbf{S}}_i\succeq 0}\mathcal{L}(\tilde{\bar{\mathbf{S}}};\boldsymbol{\lambda},\mu)$.

Proposition 4.6. *In order to have a bounded solution of the dual function $g(\boldsymbol{\lambda},\mu)$, matrix $\boldsymbol{A}_i$ must be $\boldsymbol{A}_i\succ 0\ \forall i$, otherwise $g(\boldsymbol{\lambda},\mu)$ is unbounded below, i.e., $g(\boldsymbol{\lambda},\mu)=-\infty$.*

Proof. See Appendix 4.B. ■

Thanks to the fact that matrices $\{\mathbf{A}_i\}$ are positive definite, we can assure that they can be decomposed as $\mathbf{A}_i=\mathbf{A}_i^{1/2}\mathbf{A}_i^{1/2}$ and they always have inverse. Thus, by calling $\hat{\mathbf{S}}_i=\mathbf{A}_i^{1/2}\tilde{\mathbf{S}}_i\mathbf{A}_i^{1/2}$, the dual function can be expressed as

$$g(\boldsymbol{\lambda},\mu)=\min_{\hat{\mathbf{S}}_i\succeq 0}\left\{-\sum_{i\in\mathcal{U}_I}\omega_i\log\det\left(\mathbf{I}+\hat{\mathbf{H}}_i\mathbf{A}_i^{-1/2}\hat{\mathbf{S}}_i\mathbf{A}_i^{-1/2}\hat{\mathbf{H}}_i^H\right)+\sum_{i\in\mathcal{U}_I}\mathrm{Tr}\left(\hat{\mathbf{S}}_i\right)+G\right\}. \tag{4.67}$$

The dual function in (4.67) can be recognized to be equivalent to the dual function of the classical maximization of the sum rate with a power constraint, where the optimum covariance matrix $\hat{\mathbf{S}}_i$ diagonalizes the equivalent channel $\hat{\mathbf{H}}_i\mathbf{A}_i^{-1/2}$ [Zha10], i.e., $\hat{\mathbf{S}}_i=\hat{\mathbf{V}}_i\hat{\mathbf{D}}_i\hat{\mathbf{V}}_i^H$, where $\hat{\mathbf{D}}_i$ is the power allocation matrix and its components are computed following the water-filling policy [Cov06]. Finally, it is straightforward to show that the precoder $\mathbf{B}_i$ matrix with dimensions $n_T\times n_{S_i}$ corresponding to such covariance matrix is

$$\mathbf{B}_i^{\star}=\tilde{\mathbf{V}}_i^{(0)}\mathbf{A}_i^{-1/2}\hat{\mathbf{V}}_i\hat{\mathbf{D}}_i^{1/2}. \tag{4.68}$$

4.B Proof of Proposition 4.6

Let the EVD of $\mathbf{A}_i$ be $\bar{\mathbf{U}}_i\bar{\mathbf{\Gamma}}_i\bar{\mathbf{U}}_i^H$, where $\bar{\mathbf{\Gamma}}_i$ contains the eigenvalues in decreasing order w.l.o.g. Then, the second term of the Lagrangian in (4.66) is $\sum_{i\in\mathcal{U}_I}\text{Tr}\left(\bar{\mathbf{\Gamma}}_i\bar{\mathbf{U}}_i^H\tilde{\mathbf{S}}_i\bar{\mathbf{U}}_i\right)$. Now, calling $\bar{\mathbf{S}}_i = \bar{\mathbf{U}}_i^H\tilde{\mathbf{S}}_i\bar{\mathbf{U}}_i$ (note that $\bar{\mathbf{S}}_i \succeq 0 \iff \tilde{\mathbf{S}}_i \succeq 0$) and $\hat{\bar{\mathbf{H}}}_i = \hat{\mathbf{H}}_i\bar{\mathbf{U}}_i$ we have that $g(\boldsymbol{\lambda},\mu) = \min_{\bar{\mathbf{S}}_i\succeq 0} -\sum_{i\in\mathcal{U}_I}\omega_i\log\det\left(\mathbf{I}+\hat{\bar{\mathbf{H}}}_i\bar{\mathbf{S}}_i\hat{\bar{\mathbf{H}}}_i^H\right) + \sum_{i\in\mathcal{U}_I}\text{Tr}\left(\bar{\mathbf{\Gamma}}_i\bar{\mathbf{S}}_i\right) + G$. Let us take the particular structure for the covariance matrix $\bar{\mathbf{S}}_i$ being diagonal with all the elements equal 0 except the last one which is equal P, i.e., $\bar{\mathbf{S}}_i = \text{Diag}(0,\ldots,P)$. Then, denoting $L_i = n_T - n_R + n_{R_i}$, the first term of the dual function becomes $-\sum_{i\in\mathcal{U}_I}\omega_i\log\left(1+P\|[\hat{\bar{\mathbf{H}}}_i]_{:,L_i}\|^2\right)$, where $[\hat{\bar{\mathbf{H}}}_i]_{:,L_i}$ denotes the L_i-th column of $\hat{\bar{\mathbf{H}}}_i$. Since matrix $\hat{\bar{\mathbf{H}}}_i$ is formed by unitary rotations of a random matrix with i.i.d. entries, we can assure with probability equal to 1 that $\|[\hat{\bar{\mathbf{H}}}_i]_{:,L_i}\|^2 \neq 0$. As a conclusion, the first term of the Lagrangian is negative and decreases without bound as P increases. Let us have a look at the second term. If matrix $\mathbf{A}_i$ is not positive definite, i.e., if the lowest element (and, thus, the last component) in the diagonal of $\bar{\mathbf{\Gamma}}_i$ is not positive, then the second term of the Lagrangian either is negative and decreases without bound as $P\to\infty$ or is zero. In both cases, and taking into account the behavior of the first term of the Lagrangian as P tends to infinity, it is concluded that the dual function is equal to $-\infty$. Thus, the only possible solution so that $g(\boldsymbol{\lambda},\mu)\neq-\infty$ is that $\bar{\mathbf{\Gamma}}_i$ has all the diagonal elements strictly positive and, thus, $\mathbf{A}_i \succ 0$.

4.C Proof of Lemma 4.1

The Lagrangian of the problem is $\mathcal{L}(\alpha;\mu,\sigma) = -\alpha + \mu\left(\alpha\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^0 - \tilde{r}_t\right) - \sigma\alpha$. From the KKT conditions, if we take the partial derivative w.r.t. α we get $\frac{\partial\mathcal{L}(\alpha;\mu,\sigma)}{\partial\alpha} = -1+\mu\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^0-\sigma = 0$. By inspection we have that if $\alpha^\star = 0$, then $\alpha^\star\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^0 = 0 < \tilde{r}_t$ and, thus, $\mu^\star = 0$. But if $\mu^\star = 0$, then $\sigma^\star = -1$, which is not possible. Thus, $\alpha^\star > 0$ which implies that $\sigma^\star = 0$ (from the complementary slackness $\alpha^\star\sigma^\star = 0$). As a consequence, $\mu^\star = \frac{1}{\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^0} > 0$. Finally, substituting this value into the complementary slackness $\mu^\star\left(\alpha^\star\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^0 - \tilde{r}_t\right) = 0$, yields $\alpha^\star = \frac{\tilde{r}_t}{\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^0}$.

4.D Proof of Lemma 4.2

Before attempting to obtain the optimum solution, let us characterize the sign of the perturbation $\mathbf{q}-\mathbf{q}^0$. We claim that, at the optimum, $\mathbf{q}^\star \preceq \mathbf{q}^0$, that is, $q_i^\star \leq Q_i^0$, $\forall i$. The proof is straightforward. Suppose $q_i > Q_i^0$. Then, constraint $C1$ can be rewritten as $\sum_{j\neq i}\lambda_j^\star q_j \leq \tilde{r}_t - \lambda_i^\star q_i$. Since $\lambda_i^\star \geq 0$, then we could reduce q_i (for example assign $q_i = Q_i^0$), the objective function would decrease its value and constraint $C1$ would become looser. We can proceed similarly by induction with the rest of variable to complete the proof. Thanks to this claim, we know the

sign of the derivative of the term $\|\mathbf{q}-\mathbf{q}^0\|_1$ but we need to add explicitly $\mathbf{q}^\star \preceq \mathbf{q}^0$ to the original optimization problem. Thus, problem (4.37) is modified as

$$\begin{aligned} \underset{\mathbf{q}}{\text{minimize}} \quad & \|\mathbf{q}-\mathbf{q}^0\|_1 \\ \text{subject to} \quad & C1: \boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q} \leq \tilde{r}_t \\ & C2: \mathbf{q} \preceq \mathbf{q}^0 \\ & C3: \mathbf{q} \succeq \mathbf{0}. \end{aligned} \tag{4.69}$$

The Lagrangian of problem (4.69) is $\mathcal{L}(\mathbf{q};\boldsymbol{\gamma},\mu) = \|\mathbf{q}-\mathbf{q}^0\|_1 + \mu\left(\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}-\tilde{r}_t\right)+\boldsymbol{\gamma}^T(\mathbf{q}-\mathbf{q}^0)$. Let $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_M)$ (where we have omitted constraint $C3$ as the solution will turn out to satisfy that constraint). Then, taking the derivative w.r.t. the primal variables yields $\frac{\partial\mathcal{L}(\mathbf{q};\boldsymbol{\gamma},\mu)}{\partial q_j} = -1+\mu^\star\lambda_j^\star+\gamma_j^\star = 0 \Longrightarrow 1 = \mu^\star\lambda_j^\star+\gamma_j^\star$. If $q_i < Q_i^0$, $\forall i \Longrightarrow \gamma_i^\star = 0\ \forall i$. Then, it implies that $1=\mu^\star\lambda_i^\star\ \forall i$ which is not possible since in general $\lambda_i^\star \neq \lambda_j^\star$. By inspection, we can claim that there is only one possible $\gamma_i^\star$ that could be equal to 0. Let $q_i < Q_i^0$ and $q_j = Q_j^0\ \forall j \neq i$. Then $\gamma_i^\star = 0$ and $\gamma_j^\star > 0$. We have that $\mu^\star = \frac{1}{\lambda_i^\star}$ and substituting back, $1 = \frac{\lambda_j^\star}{\lambda_i^\star}+\gamma_j^\star$. Since $\gamma_j^\star > 0 \Longrightarrow \lambda_i^\star > \lambda_j^\star$ for the previous equation to be true. Finally, we have that $q_j^\star = Q_j^0$ and from constraint $C1$, $q_i^\star = \frac{1}{\lambda_i^\star}\left(\tilde{r}_t - \sum_{j\neq i}\lambda_j^\star Q_j^0\right)$.

4.E Proof of Lemma 4.3

The Lagrangian of the problem is $\mathcal{L}(t,\mathbf{q};\boldsymbol{\nu},\boldsymbol{\gamma},\mu) = -t+\boldsymbol{\nu}^T(t\mathbf{1}-\mathbf{q})+\mu\left(\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}-\tilde{r}_t\right)-\boldsymbol{\gamma}^T\mathbf{q}$. Let us for the moment omit the positivity constraints ($\boldsymbol{\gamma}^T\mathbf{q}$) since (as it will be shown later) the solution will automatically satisfy them. Taking the derivatives w.r.t. the primal variables yields, $\frac{\partial\mathcal{L}(t,\mathbf{q};\boldsymbol{\nu},\mu)}{\partial t} = -1+\boldsymbol{\nu}^{\star T}\mathbf{1} \Longrightarrow \boldsymbol{\nu}^{\star T}\mathbf{1}=1$ and $\frac{\partial\mathcal{L}(t,\mathbf{q};\boldsymbol{\nu},\mu)}{\partial q_j} = -\nu_j^\star+\mu^\star\lambda_j^\star(\mathbf{q}^0)=0 \Longrightarrow \nu_j^\star = \mu^\star\lambda_j^\star(\mathbf{q}^0)$. Summing at both sides, $\boldsymbol{\nu}^{\star T}\mathbf{1} = \mu^\star\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{1}$ and so $\mu^\star = \frac{1}{\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{1}}$. In this case, $\nu_j^\star = \frac{\lambda_j^\star(\mathbf{q}^0)}{\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{1}} > 0$ and from the complementary slackness it is implied that $t^\star\mathbf{1}=\mathbf{q}^\star$. Multiplying both sides by $\boldsymbol{\lambda}^\star(\mathbf{q}^0)$ yields $t^\star\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{1} = \boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{q}^\star \leq \tilde{r}_t$, and hence $t^\star\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{1} \leq \tilde{r}_t$. Given that, the maximum value of t is attained at $t^\star = \frac{\tilde{r}_t}{\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{1}}$. Finally, we have $\mathbf{q}^\star = \frac{\tilde{r}_t}{\boldsymbol{\lambda}^\star(\mathbf{q}^0)^T\mathbf{1}}\mathbf{1}$ which fulfills the positivity constraint.

4.F Proof of Proposition 4.3

The proposed quadratic surrogate function of $s_i(\bar{\mathbf{S}})$ has the following form:

$$\begin{aligned} \hat{s}_i(\bar{\mathbf{S}},\bar{\mathbf{S}}^{(0)}) \triangleq \log\det\left(\mathbf{I}+\mathbf{H}_i\bar{\mathbf{S}}^{(0)}\mathbf{H}_i^H\right) + \mathrm{Re}\Big\{\mathrm{Tr}\left(\mathbf{G}_i\left(\bar{\mathbf{S}}-\bar{\mathbf{S}}^{(0)}\right)\right)\Big\} \\ + \mathrm{Tr}\left(\left(\bar{\mathbf{S}}-\bar{\mathbf{S}}^{(0)}\right)^H\mathbf{M}_i\left(\bar{\mathbf{S}}-\bar{\mathbf{S}}^{(0)}\right)\right) \leq \log\det\left(\mathbf{I}+\mathbf{H}_i\bar{\mathbf{S}}\mathbf{H}_i^H\right), \quad \forall\bar{\mathbf{S}},\,\bar{\mathbf{S}}^{(0)} \in \mathcal{S}_+^{n_T} \end{aligned} \tag{4.70}$$

where matrices $\mathbf{G}_i \in \mathbb{C}^{n_T \times n_T}$ and $\mathbf{M}_i \in \mathbb{C}^{n_T \times n_T}$ need to be found such that conditions (A1) through (A4) are satisfied (see Section 2.4), and $\mathrm{Re}\{x\}$ denotes the real part of x. Note that (A1) and (A4) are already satisfied. Only (A2) and (A3) must be ensured.

Let us start by proving condition (A3). Let $\bar{\mathbf{S}}^{(0)}$ and $\bar{\mathbf{S}}^{(1)}$ be two positive semidefinite matrices, i.e, $\bar{\mathbf{S}}^{(0)}, \bar{\mathbf{S}}^{(1)} \in \mathcal{S}_+^{n_T}$. Then, the directional derivative of the surrogate function $\hat{s}_i(\bar{\mathbf{S}}, \bar{\mathbf{S}}^{(0)})$ in (4.70) at $\bar{\mathbf{S}}^{(0)}$ with direction $\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}$ is given by:

$$\lim_{\lambda \to 0} \frac{\hat{s}_i\left(\bar{\mathbf{S}}^{(0)} + \lambda\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right), \bar{\mathbf{S}}^{(0)}\right) - \hat{s}_i\left(\bar{\mathbf{S}}^{(0)}, \bar{\mathbf{S}}^{(0)}\right)}{\lambda} = \mathrm{Re}\Big\{ \mathrm{Tr}\left(\mathbf{G}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\Big\}. \quad (4.71)$$

Now, let us compute the directional derivative of the term $\log\det\left(\mathbf{I} + \mathbf{H}_i \bar{\mathbf{S}} \mathbf{H}_i^H\right)$,

$$\lim_{\lambda \to 0} \frac{\log\det\left(\mathbf{I} + \mathbf{H}_i\left(\bar{\mathbf{S}}^{(0)} + \lambda\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\mathbf{H}_i^H\right) - \log\det\left(\mathbf{I} + \mathbf{H}_i \bar{\mathbf{S}}^{(0)} \mathbf{H}_i^H\right)}{\lambda} \quad (4.72)$$

$$= \mathrm{Tr}\left(\mathbf{H}_i^H\left(\mathbf{I} + \mathbf{H}_i \bar{\mathbf{S}}^{(0)} \mathbf{H}_i^H\right)^{-1} \mathbf{H}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right), \quad (4.73)$$

where we have used $\mathrm{d}\log\det(\mathbf{X}) = \mathrm{Tr}(\mathbf{X}^{-1}\mathrm{d}\mathbf{X})$ [Mag88]. Hence, by applying condition (A3), the two directional derivatives must be equal,

$$\mathrm{Re}\Big\{ \mathrm{Tr}\left(\mathbf{G}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\Big\} = \mathrm{Tr}\left(\mathbf{H}_i^H\left(\mathbf{I} + \mathbf{H}_i \bar{\mathbf{S}}^{(0)} \mathbf{H}_i^H\right)^{-1} \mathbf{H}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right), \quad (4.74)$$

from which we are able to identify matrix $\mathbf{G}_i$ as

$$\mathbf{G}_i = \mathbf{H}_i^H\left(\mathbf{I} + \mathbf{H}_i \bar{\mathbf{S}}^{(0)} \mathbf{H}_i^H\right)^{-1} \mathbf{H}_i, \qquad \mathbf{G}_i = \mathbf{G}_i^H. \quad (4.75)$$

Note that as matrix $\mathbf{G}_i$ is hermitian, the real operator is no longer needed since the trace of the product of two hermitian matrices is real. In order to prove condition (A2), it suffices to show that for each linear cut in any direction, the surrogate function is a lower bound. Let $\bar{\mathbf{S}} = \bar{\mathbf{S}}^{(0)} + \mu\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)$, $\forall \mu \in [0,1]$. Then, it suffices to show

$$\log\det\left(\mathbf{I} + \mathbf{H}_i \bar{\mathbf{S}}^{(0)} \mathbf{H}_i^H\right) + \mu\,\mathrm{Tr}\left(\mathbf{G}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right) + \mu^2\,\mathrm{Tr}\left(\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)^H \mathbf{M}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)$$
$$\leq \log\det\left(\mathbf{I} + \mathbf{H}_i\left(\bar{\mathbf{S}}^{(0)} + \mu\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\mathbf{H}_i^H\right), \quad \forall \bar{\mathbf{S}}^{(1)}, \bar{\mathbf{S}}^{(0)} \in \mathcal{S}_+^{n_T}, \forall \mu \in [0,1]. \quad (4.76)$$

Since the left hand side of (4.76) is concave w.r.t. μ and (A1) holds true, a sufficient condition is that the second derivative of the left hand side of (4.76) must be lower than or equal to the

second derivative of the right hand side of (4.76) for any $\mu \in [0,1]$ and any $\bar{\mathbf{S}}^{(1)}$, $\bar{\mathbf{S}}^{(0)} \in \mathcal{S}_+^{n_T}$:

$$2\,\mathrm{Tr}\left(\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)^H \mathbf{M}_i \left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right) \leq \frac{\partial^2}{\partial \mu^2} \log\det\left(\mathbf{I} + \mathbf{H}_i\left(\bar{\mathbf{S}}^{(0)} + \mu\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\mathbf{H}_i^H\right)\Bigg|_{\forall \bar{\mathbf{S}}^{(1)}, \bar{\mathbf{S}}^{(0)} \in \mathcal{S}_+^{n_T}, \forall \mu \in [0,1]} \tag{4.77}$$

Let us compute the second derivative of the right hand side of (4.77). The first derivative is given by

$$\frac{\partial}{\partial \mu} \log\det\left(\mathbf{I} + \mathbf{H}_i\left(\bar{\mathbf{S}}^{(0)} + \mu\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\mathbf{H}_i^H\right) = \mathrm{Tr}\left(\left(\mathbf{I} + \mathbf{H}_i\left(\bar{\mathbf{S}}^{(0)} + \mu\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\mathbf{H}_i^H\right)^{-1} \mathbf{H}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\mathbf{H}_i^H\right), \tag{4.78}$$

and the second derivative is given by

$$\frac{\partial^2}{\partial \mu^2} \log\det\left(\mathbf{I} + \mathbf{H}_i\left(\bar{\mathbf{S}}^{(0)} + \mu\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\mathbf{H}_i^H\right) = -\,\mathrm{Tr}\left(\mathbf{A}_i^{-1}\mathbf{H}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\mathbf{H}_i^H\mathbf{A}_i^{-1}\mathbf{H}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\mathbf{H}_i^H\right), \tag{4.79}$$

where we have used the identity $\mathrm{d}\mathbf{X}^{-1} = -\mathbf{X}^{-1}\mathrm{d}\mathbf{X}\mathbf{X}^{-1}$ [Mag88] and matrix $\mathbf{A}_i \in \mathbb{C}^{n_{R_i} \times n_{R_i}}$ is defined as $\mathbf{A}_i = \mathbf{I} + \mathbf{H}_i\left(\bar{\mathbf{S}}^{(0)} + \mu\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)\mathbf{H}_i^H$.

We need to manipulate the previous expressions. To this end, let us define matrix $\mathbf{P}_i = \mathbf{H}_i^H\mathbf{A}_i^{-1}\mathbf{H}_i \in \mathbb{C}^{n_T \times n_T}$ and let us vectorize the result found in (4.79):

$$\begin{aligned}
&\mathrm{Tr}\left(\mathbf{P}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\mathbf{P}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right) \\
&\quad= \mathrm{vec}\left(\left(\mathbf{P}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right)^T\right)^T \mathrm{vec}\left(\mathbf{P}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right) && (4.80)\\
&\quad= \mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)^T\mathbf{P}_i^T\right)^T \mathrm{vec}\left(\mathbf{P}_i\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)\right) && (4.81)\\
&\quad= \mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)^T\right)^T \left(\mathbf{I} \otimes \mathbf{P}_i^T\right)\left(\mathbf{I} \otimes \mathbf{P}_i\right)\mathrm{vec}\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right) && (4.82)\\
&\quad= \mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right)^T\right)^T \left(\mathbf{I} \otimes \mathbf{P}_i^T\mathbf{P}_i\right)\mathrm{vec}\left(\bar{\mathbf{S}}^{(1)} - \bar{\mathbf{S}}^{(0)}\right), && (4.83)
\end{aligned}$$

where in (4.80) we have used $\mathrm{Tr}(\mathbf{AB}) = \mathrm{vec}(\mathbf{A}^T)^T\mathrm{vec}(\mathbf{B})$, in (4.82) we have used $\mathrm{vec}(\mathbf{AB})^T = \mathrm{vec}(\mathbf{A})^T(\mathbf{I} \otimes \mathbf{B})$ and $\mathrm{vec}(\mathbf{AB}) = (\mathbf{I} \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$, and in (4.83) we have used the identity $(\mathbf{A} \otimes$

$\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=(\mathbf{AC})\otimes(\mathbf{BD})$. Let us now vectorize the left hand side of (4.77):

$$2\,\mathrm{Tr}\left(\left(\bar{\mathbf{S}}-\bar{\mathbf{S}}^{(0)}\right)^{H}\mathbf{M}_i\left(\bar{\mathbf{S}}-\bar{\mathbf{S}}^{(0)}\right)\right)$$

$$=2\mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right)^{*}\right)^{T}(\mathbf{I}\otimes\mathbf{M}_i)\mathrm{vec}\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right) \tag{4.84}$$

$$=2\mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right)^{T}\right)^{T}(\mathbf{I}\otimes\mathbf{M}_i)\mathrm{vec}\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right), \tag{4.85}$$

where in (4.85) we have used the fact that $\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}$ is hermitian and $\mathrm{Tr}(\mathbf{ABC})=\mathrm{vec}(\mathbf{A}^T)^T(\mathbf{I}\otimes\mathbf{B})\mathrm{vec}(\mathbf{C})$. Finally, condition (4.77) is equivalent to

$$2\mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right)^{T}\right)^{T}(\mathbf{I}\otimes\mathbf{M}_i)\mathrm{vec}\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right) \tag{4.86}$$

$$\leq -\mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right)^{T}\right)^{T}\left(\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\right)\mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right)\right),$$

which can be expressed as

$$2\mathrm{vec}\left(\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right)^{T}\right)^{T}\left[(\mathbf{I}\otimes\mathbf{M}_i)+\frac{1}{2}\left(\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\right)\right]\mathrm{vec}\left(\bar{\mathbf{S}}^{(1)}-\bar{\mathbf{S}}^{(0)}\right)\leq 0. \tag{4.87}$$

A sufficient condition for the previous expression is:

$$(\mathbf{I}\otimes\mathbf{M}_i)+\frac{1}{2}\left(\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\right)=\mathbf{I}\otimes\left(\mathbf{M}_i+\frac{1}{2}\mathbf{P}_i^T\mathbf{P}_i\right)\preceq 0, \tag{4.88}$$

which means that

$$\mathbf{M}_i+\frac{1}{2}\mathbf{P}_i^T\mathbf{P}_i\preceq 0. \tag{4.89}$$

Now, if we set $\mathbf{M}_i=\alpha\mathbf{I}$ (note that this is a particular simple solution), (4.89) is satisfied if

$$\alpha\leq-\frac{1}{2}\lambda_{\max}\left(\mathbf{P}_i^T\mathbf{P}_i\right), \tag{4.90}$$

where $\lambda_{\max}(\mathbf{X})$ is the maximum eigenvalue of matrix $\mathbf{X}$. Now, let us introduce the following result:

Theorem 4.2 ([Wan92])**.** *Let $\boldsymbol{A}$, $\boldsymbol{B}\in\mathbb{C}^{n\times n}$, assume that $\boldsymbol{A}$ is positive definite, and assume that $\boldsymbol{B}$ is positive definite. Let $\lambda_i(\boldsymbol{A})$ be the i-th eigenvalue of matrix $\boldsymbol{A}$ such that $\lambda_1(\boldsymbol{A})\geq\lambda_2(\boldsymbol{A})\geq\cdots\geq\lambda_n(\boldsymbol{A})$. Then, for all $i,j,k\in\{1,\ldots,n\}$ such that $j+k\leq i+1$,*

$$\lambda_i(\boldsymbol{AB})\leq\lambda_j(\boldsymbol{A})\lambda_k(\boldsymbol{B}). \tag{4.91}$$

In particular, for all $i=1,\ldots,n$,

$$\lambda_i(\boldsymbol{A})\lambda_n(\boldsymbol{B})\leq\lambda_i(\boldsymbol{AB})\leq\lambda_i(\boldsymbol{A})\lambda_1(\boldsymbol{B}). \tag{4.92}$$

Thanks to the previous result, it is sufficient to take $\alpha \leq -\frac{1}{2}\lambda_{\max}^2(\mathbf{P}_i)$. Now, let the SVD of $\mathbf{H}_i$ be $\mathbf{H}_i = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{V}_i^H$. From this, we can upper bound $\lambda_{\max}(\mathbf{P}_i) = \lambda_{\max}\left(\mathbf{H}_i^H\mathbf{A}_i^{-1}\mathbf{H}_i\right) = \lambda_{\max}\left(\mathbf{\Sigma}_i\mathbf{V}_i^H\mathbf{A}_i^{-1}\mathbf{V}_i\mathbf{\Sigma}_i\right) \leq \sigma_{\max}^2(\mathbf{H}_i)\lambda_{\min}^{-1}(\mathbf{A}_i)$, where $\sigma_{\max}(\mathbf{X})$ is the maximum singular value of matrix $\mathbf{X}$. Because matrix $\mathbf{A}$ is positive definite with $\lambda_{\min}(\mathbf{A}_i) \geq 1$, a stricter condition is

$$\alpha \leq -\frac{1}{2}\sigma_{\max}^4(\mathbf{H}_i), \tag{4.93}$$

and thus, a possible matrix $\mathbf{M}_i$ satisfying conditions (A1) – (A4) is finally

$$\mathbf{M}_i = -\frac{1}{2}\sigma_{\max}^4(\mathbf{H}_i)\mathbf{I} = -\frac{1}{2}\lambda_{\max}^2(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{I}. \tag{4.94}$$

4.G Proof of Proposition 4.4

Let us start by vectorizing the surrogate function in (4.54):

$$\begin{aligned}\hat{R}_i(\mathbf{S},\mathbf{S}^{(0)}) &= \hat{s}_i(\bar{\mathbf{S}},\bar{\mathbf{S}}^{(0)}) - \hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}),\mathbf{\Omega}_i^{(0)}) \\ &= \mathrm{vec}\left(\bar{\mathbf{S}}^T\right)^T(\mathbf{I}\otimes\mathbf{M}_i)\,\mathrm{vec}\left(\bar{\mathbf{S}}\right) + \mathbf{e}_i^T\mathrm{vec}\left(\bar{\mathbf{S}}\right) + \mathbf{r}_i^T\mathrm{vec}\left(\mathbf{S}_i\right) + \kappa_2, \end{aligned}\tag{4.95}$$

where $\mathbf{e}_i = \mathrm{vec}\left(\mathbf{E}_i^T\right) \in \mathbb{C}^{n_Tn_T\times 1}$, $\mathbf{r}_i = \mathrm{vec}\left(\mathbf{R}_i^T\right) \in \mathbb{C}^{n_Tn_T\times 1}$, and κ_2 contains some constant terms that do not depend on $\{\mathbf{S}_i\}$. Let $\mathbf{s} = \left[\mathrm{vec}(\mathbf{S}_1)^T\mathrm{vec}(\mathbf{S}_2)^T\ldots\mathrm{vec}(\mathbf{S}_{|\mathcal{U}_I|})^T\right]^T \in \mathbb{C}^{n_Tn_T|\mathcal{U}_I|\times 1}$. Note that $\mathrm{vec}\left(\bar{\mathbf{S}}\right) = \mathbf{T}\mathbf{s}$, where $\mathbf{T} \in \mathbb{C}^{n_Tn_T\times n_Tn_T|\mathcal{U}_I|}$ is composed of $|\mathcal{U}_I|$ identity matrices of size $n_Tn_T \times n_Tn_T$, i.e., $\mathbf{T} = [\mathbf{I}\ \mathbf{I}\ \ldots\ \mathbf{I}]$. Now, we can rewrite (4.95) as

$$\hat{R}_i(\mathbf{S},\mathbf{S}^{(0)}) = \mathbf{s}^H\mathbf{T}^H(\mathbf{I}\otimes\mathbf{M}_i)\mathbf{T}\mathbf{s} + \mathbf{e}_i^T\mathbf{T}\mathbf{s} + \mathbf{r}_i^T\mathrm{vec}\left(\mathbf{S}_i\right) + \kappa_2. \tag{4.96}$$

We know proceed to formulate the objective function (denoted by $\bar{f}_0(\mathbf{S},\mathbf{S}^{(0)})$) of problem (4.46) but substituting the bound that we just computed and considering the proximal term. If we incorporate all the terms (but omitting the constant ones) we have

$$\bar{f}_0(\mathbf{S},\mathbf{S}^{(0)}) = \sum_{i\in\mathcal{U}_I}\omega_i\left(\mathbf{s}^H\mathbf{T}^H(\mathbf{I}\otimes\mathbf{M}_i)\mathbf{T}\mathbf{s} + \mathbf{e}_i^T\mathbf{T}\mathbf{s} + \mathbf{r}_i^T\mathrm{vec}\left(\mathbf{S}_i\right)\right) - \rho\left\|\mathbf{S}_i - \mathbf{S}_i^{(0)}\right\|_F^2 \tag{4.97}$$

$$= \mathbf{s}^H\mathbf{T}^H\tilde{\mathbf{M}}\mathbf{T}\mathbf{s} + \tilde{\mathbf{e}}^T\mathbf{T}\mathbf{s} + \hat{\mathbf{r}}^T\mathbf{s} - \rho\mathbf{s}^H\mathbf{s} + \rho\mathbf{s}^{(0),H}\mathbf{s} + \rho\mathbf{s}^H\mathbf{s}^{(0)} - \rho\mathbf{s}^{(0),H}\mathbf{s}^{(0)}, \tag{4.98}$$

where $\tilde{\mathbf{M}} = \sum_{i\in\mathcal{U}_I}\omega_i(\mathbf{I}\otimes\mathbf{M}_i) \in \mathbb{C}^{n_Tn_T\times n_Tn_T}$, $\tilde{\mathbf{e}} = \sum_{i\in\mathcal{U}_I}\omega_i\mathbf{e}_i$, $\hat{\mathbf{r}} = \left[\mathbf{r}_1^T\,\mathbf{r}_2^T\,\ldots\,\mathbf{r}_{|\mathcal{U}_I|}^T\right]^T \in \mathbb{C}^{n_Tn_T|\mathcal{U}_I|\times 1}$, and $\mathbf{s}^{(0)} = \left[\mathrm{vec}(\mathbf{S}_1^{(0)})^T\mathrm{vec}(\mathbf{S}_2^{(0)})^T\ldots\mathrm{vec}(\mathbf{S}_{|\mathcal{U}_I|}^{(0)})^T\right]^T \in \mathbb{C}^{n_Tn_T|\mathcal{U}_I|\times 1}$. Now taking into account that the objective function $\bar{f}_0(\mathbf{S},\mathbf{S}^{(0)})$ must be real and combining terms (omitting terms that do not depend on $\mathbf{s}$) we obtain

$$\bar{f}_0(\mathbf{S},\mathbf{S}^{(0)}) = \mathbf{s}^H\mathbf{C}\mathbf{s} + \mathbf{b}^T\mathbf{s} + \mathbf{s}^H\mathbf{b}^*, \tag{4.99}$$

where $\mathbf{b}^T = \frac{1}{2}\tilde{\mathbf{e}}^T\mathbf{T} + \frac{1}{2}\hat{\mathbf{r}}^T + \rho\mathbf{s}^{(0),H} \in \mathbb{C}^{1\times n_T n_T|\mathcal{U}_I|}$ and matrix $\mathbf{C}$ is $\mathbf{C} = \mathbf{T}^H\tilde{\mathbf{M}}\mathbf{T} - \rho\mathbf{I} \in \mathbb{C}^{n_T n_T|\mathcal{U}_I|\times n_T n_T|\mathcal{U}_I|}$. For convenient purposes, let us change the sign of $\bar{f}_0(\mathbf{S},\mathbf{S}^{(0)})$ such that $\bar{\bar{f}}_0(\mathbf{S},\mathbf{S}^{(0)}) = -\bar{f}_0(\mathbf{S},\mathbf{S}^{(0)}) = \mathbf{s}^H\tilde{\mathbf{C}}\mathbf{s} - \mathbf{b}^T\mathbf{s} - \mathbf{s}^H\mathbf{b}^*$, where $\tilde{\mathbf{C}} = -\mathbf{C} \succeq 0$. Finally, we can equivalently rewrite the objective function as the following expression (with this new reformulation, the objective is to minimize $\bar{\bar{f}}_0(\mathbf{S},\mathbf{S}^{(0)})$ instead of maximizing it):

$$\bar{\bar{f}}_0(\mathbf{S},\mathbf{S}^{(0)}) = \|\tilde{\mathbf{C}}^{\frac{1}{2}}\mathbf{s} - \mathbf{c}\|_2^2, \tag{4.100}$$

where

$$\mathbf{c} = \tilde{\mathbf{C}}^{-\frac{1}{2}}\mathbf{b}^* \in \mathbb{C}^{n_T n_T|\mathcal{U}_I|\times 1}. \tag{4.101}$$

Note that the term $\mathbf{c}^H\mathbf{c}$ does not affect the optimum value of the optimization variables as this term does not depend on $\mathbf{s}$. Now, we can reformulate the optimization problem presented in (4.46) as

$$\begin{aligned} \underset{\{\mathbf{S}_i\},\mathbf{s}}{\text{minimize}} \quad & \|\tilde{\mathbf{C}}^{\frac{1}{2}}\mathbf{s} - \mathbf{c}\|_2^2 \\ \text{subject to} \quad & C1: \mathbf{T}_i\mathbf{s} = \text{vec}\left(\mathbf{S}_i\right), \quad \forall i \in \mathcal{U}_I \\ & C2: \mathbf{S} \in \mathcal{S}_1, \end{aligned} \tag{4.102}$$

where $\mathbf{T}_i = [\underbrace{\mathbf{0},\mathbf{0},\ldots,\mathbf{0}}_{i-1},\mathbf{I},\mathbf{0},\ldots,\mathbf{0}] \in \mathbb{R}^{n_T n_T\times n_T n_T|\mathcal{U}_I|}$ is composed of zero matrices of dimension $n_T n_T \times n_T n_T$ with an identity matrix at the i-th position. Problem (4.102) can further be reformulated as

$$\begin{aligned} \underset{\{\mathbf{S}_i\},\mathbf{s},t}{\text{minimize}} \quad & t \\ \text{subject to} \quad & C1: \|\tilde{\mathbf{C}}^{\frac{1}{2}}\mathbf{s} - \mathbf{c}\|_2 \le t \\ & C2: \mathbf{T}_i\mathbf{s} = \text{vec}\left(\mathbf{S}_i\right), \quad \forall i \in \mathcal{U}_I \\ & C3: \mathbf{S} \in \mathcal{S}_1, \end{aligned} \tag{4.103}$$

and, finally, as the following SDP optimization problem

$$\begin{aligned} \underset{\{\mathbf{S}_i\},\mathbf{s},t}{\text{minimize}} \quad & t \\ \text{subject to} \quad & C1: \begin{bmatrix} t\mathbf{I} & \tilde{\mathbf{C}}^{\frac{1}{2}}\mathbf{s} - \mathbf{c} \\ \left(\tilde{\mathbf{C}}^{\frac{1}{2}}\mathbf{s} - \mathbf{c}\right)^H & 1 \end{bmatrix} \succeq 0 \\ & C2: \mathbf{T}_i\mathbf{s} = \text{vec}\left(\mathbf{S}_i\right), \qquad \forall i \in \mathcal{U}_I \\ & C3: \mathbf{S} \in \mathcal{S}_1. \end{aligned} \tag{4.104}$$

4.H Proof of Proposition 4.5

The proposed quadratic surrogate function of $s_i(\mathbf{S})$ has the following form:

$$\begin{aligned}\hat{s}_i(\mathbf{S},\mathbf{S}^{(0)}) \triangleq \log\det\left(\mathbf{I}+\mathbf{H}_i\sum_{k\in\mathcal{U}_I}\mathbf{S}_k^{(0)}\mathbf{H}_i^H\right)+\sum_{\ell\in\mathcal{U}_I}\mathrm{Re}\Big\{\mathrm{Tr}\left(\mathbf{G}_{\ell i}\left(\mathbf{S}_\ell-\mathbf{S}_\ell^{(0)}\right)\right)\Big\}\\ +\sum_{\ell\in\mathcal{U}_I}\mathrm{Tr}\left(\left(\mathbf{S}_\ell-\mathbf{S}_\ell^{(0)}\right)^H\mathbf{M}_{\ell i}\left(\mathbf{S}_\ell-\mathbf{S}_\ell^{(0)}\right)\right)\\ \leq \log\det\left(\mathbf{I}+\mathbf{H}_i\sum_{k\in\mathcal{U}_I}\mathbf{S}_k\mathbf{H}_i^H\right),\quad \forall\mathbf{S}_\ell,\,\mathbf{S}_\ell^{(0)}\in\mathcal{S}_+^{n_T},\end{aligned}\tag{4.105}$$

where matrices $\mathbf{G}_i\in\mathbb{C}^{n_T\times n_T}$ and $\mathbf{M}_i\in\mathbb{C}^{n_T\times n_T}$ need to be found such that conditions (A1) through (A4) are satisfied. Note that (A1) and (A4) are already satisfied. Only (A2) and (A3) must be ensured.

Let us start with condition (A3). Let $\mathbf{S}_\ell^{(0)}$, $\mathbf{S}_\ell^{(1)}\in\mathcal{S}_+^{n_T}$, $\forall\ell$. Then, the directional derivative of the surrogate function $\hat{s}_i(\mathbf{S},\mathbf{S}^{(0)})$ in (4.105) at $\mathbf{S}_\ell^{(0)}$ with direction $\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}$ is given by

$$\lim_{\lambda\to0}\frac{\hat{s}_i\left(\mathbf{S}_\ell^{(0)}+\lambda\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right),\mathbf{S}^{(0)}\right)-\hat{s}_i\left(\mathbf{S}_\ell^{(0)},\mathbf{S}^{(0)}\right)}{\lambda}=\sum_{\ell\in\mathcal{U}_I}\mathrm{Re}\Big\{\mathrm{Tr}\left(\mathbf{G}_{\ell i}\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)\right)\Big\},\tag{4.106}$$

and the directional derivative of the right hand side of (4.105) at $\mathbf{S}_\ell^{(0)}$ with direction $\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}$ is given by

$$\mathrm{Tr}\left(\mathbf{H}_i^H\left(\mathbf{I}+\mathbf{H}_i\sum_{k\in\mathcal{U}_I}\mathbf{S}_k^{(0)}\mathbf{H}_i^H\right)^{-1}\mathbf{H}_i\left(\sum_{\ell\in\mathcal{U}_I}\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)\right)\right)\tag{4.107}$$

$$=\sum_{\ell\in\mathcal{U}_I}\mathrm{Tr}\left(\mathbf{H}_i^H\left(\mathbf{I}+\mathbf{H}_i\sum_{k\in\mathcal{U}_I}\mathbf{S}_k^{(0)}\mathbf{H}_i^H\right)^{-1}\mathbf{H}_i\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)\right).\tag{4.108}$$

From (4.106) and (4.108), we identify the matrices $\mathbf{G}_{\ell i}$ as

$$\mathbf{G}_{\ell i}=\mathbf{H}_i^H\left(\mathbf{I}+\mathbf{H}_i\sum_{k\in\mathcal{U}_I}\mathbf{S}_k^{(0)}\mathbf{H}_i^H\right)^{-1}\mathbf{H}_i,\quad \mathbf{G}_{\ell i}=\mathbf{G}_{\ell i}^H,\tag{4.109}$$

where we find that all matrices $\mathbf{G}_{\ell i}$ for a given user i can be the same, $\mathbf{G}_i=\mathbf{G}_{\ell i}$ (i.e., they do not depend on ℓ).

Now, we seek to find matrices $\{\mathbf{M}_{\ell i}\}$ based on condition (A2). To this end, we follow the same procedure presented before. We make linear cuts in each possible direction and apply the condition over the second derivative (see (4.77)). The second derivative of the left hand side of

(4.105) is given by

$$2\sum_{\ell\in\mathcal{U}_I}\mathrm{Tr}\left(\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)^H\mathbf{M}_{\ell i}\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)\right) \tag{4.110}$$

$$=2\sum_{\ell\in\mathcal{U}_I}\mathrm{vec}\left(\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)^T\right)^T(\mathbf{I}\otimes\mathbf{M}_{\ell i})\,\mathrm{vec}\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right), \tag{4.111}$$

and the second derivative of the right hand side is given by

$$-\mathrm{vec}\left(\left(\sum_{\ell\in\mathcal{U}_I}\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)\right)^T\right)^T(\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i)\,\mathrm{vec}\left(\sum_{\ell\in\mathcal{U}_I}\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)\right), \tag{4.112}$$

where $\mathbf{P}_i=\mathbf{H}_i^H\left(\mathbf{I}+\mathbf{H}_i\left(\sum_{\ell\in\mathcal{U}_I}\left(\mathbf{S}_\ell^{(0)}+\mu\left(\mathbf{S}_\ell^{(1)}-\mathbf{S}_\ell^{(0)}\right)\right)\right)\mathbf{H}_i^H\right)^{-1}\mathbf{H}_i$, being $\mu\in[0,1]$. Let $\mathbf{s}=\left[\mathrm{vec}\left(\mathbf{S}_1^{(1)}-\mathbf{S}_1^{(0)}\right)^T\mathrm{vec}\left(\mathbf{S}_2^{(1)}-\mathbf{S}_2^{(0)}\right)^T\cdots\mathrm{vec}\left(\mathbf{S}_{|\mathcal{U}_I|}^{(1)}-\mathbf{S}_{|\mathcal{U}_I|}^{(0)}\right)^T\right]^T$. Let us introduce the following block diagonal matrix

$$\tilde{\mathbf{M}}_i=\begin{bmatrix}\mathbf{I}\otimes\mathbf{M}_{1i} & \mathbf{0} & \dots & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\otimes\mathbf{M}_{2i} & & \vdots\\ \vdots & & \ddots & \mathbf{0}\\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}\otimes\mathbf{M}_{|\mathcal{U}_I|i}\end{bmatrix}. \tag{4.113}$$

Then we have that the following condition should be fulfilled:

$$2\mathbf{s}^H\tilde{\mathbf{M}}_i\mathbf{s}+\mathbf{s}^H\mathbf{T}^H\left(\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\right)\mathbf{T}\mathbf{s}\leq 0, \tag{4.114}$$

where $\mathbf{T}\in\mathbb{C}^{n_Tn_T\times n_Tn_T|\mathcal{U}_I|}$ is composed of $|\mathcal{U}_I|$ identity matrices of size $n_Tn_T\times n_Tn_T$, i.e., $\mathbf{T}=[\mathbf{I}\;\mathbf{I}\;\dots\;\mathbf{I}]$, which means that

$$\tilde{\mathbf{M}}_i+\frac{1}{2}\mathbf{T}^H\left(\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\right)\mathbf{T}\preceq 0. \tag{4.115}$$

Note that matrix $\mathbf{T}^H\left(\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\right)\mathbf{T}$ is

$$\mathbf{T}^H\left(\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\right)\mathbf{T}=\begin{bmatrix}\mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i & \mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i & \dots & \mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\\ \mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i & \mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i & & \vdots\\ \vdots & & \ddots & \\ \mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i & \dots & & \mathbf{I}\otimes\mathbf{P}_i^T\mathbf{P}_i\end{bmatrix}, \tag{4.116}$$

From the previous conditions we can see that all matrices $\mathbf{M}_{\ell i}$ will be the same for user i, i.e., $\mathbf{M}_{\ell i}=\mathbf{M}_i$, $\forall\ell$. Now if we choose the particular structure $\mathbf{M}_i=\alpha_i\mathbf{I}$, then condition (4.115)

is equivalent to

$$\alpha_i \mathbf{I} + \frac{1}{2}\mathbf{T}^H \left(\mathbf{I} \otimes \mathbf{P}_i^T \mathbf{P}_i\right) \mathbf{T} \preceq 0. \tag{4.117}$$

Now, condition (4.117) is equivalent to

$$\alpha_i \mathbf{g}^H \mathbf{g} \leq -\frac{1}{2}\mathbf{g}^H \mathbf{T}^H \left(\mathbf{I} \otimes \mathbf{P}_i^T \mathbf{P}_i\right) \mathbf{T}\mathbf{g}, \quad \forall \mathbf{g}. \tag{4.118}$$

If we propose a value of α such that

$$\alpha_i \mathbf{g}^H \mathbf{g} \leq -\frac{1}{2}\|\mathbf{T}\mathbf{g}\|_2^2 \lambda_{\max}\left(\mathbf{I} \otimes \mathbf{P}_i^T \mathbf{P}_i\right), \quad \forall \mathbf{g}, \tag{4.119}$$

$$\alpha_i \mathbf{g}^H \mathbf{g} \leq -\frac{1}{2}\|\mathbf{T}\mathbf{g}\|_2^2 \lambda_{\max}\left(\mathbf{P}_i^T \mathbf{P}_i\right), \quad \forall \mathbf{g}. \tag{4.120}$$

are fulfilled, this ensures that (4.118) is fulfilled. Therefore, the conditions over α shown in (4.119) and (4.120) are sufficient conditions to fulfill (4.117). Now, the term $\|\mathbf{T}\mathbf{g}\|_2^2$ can be further simplified. Based on the structure of matrix $\mathbf{T}$, we have that

$$\begin{aligned}
\|\mathbf{T}\mathbf{g}\|_2^2 &= \sum_{i=1}^{n_T n_T} |\mathbf{g}_i + \mathbf{g}_{i+n_T n_T+1} + \ldots + \mathbf{g}_{i+n_T n_T(|\mathcal{U}_I|-1)+1}|^2, && (4.121)\\
&\leq \sum_{i=1}^{n_T n_T} ||\mathcal{U}_I| \max\{\mathbf{g}_i, \mathbf{g}_{i+n_T n_T+1}, \ldots, \mathbf{g}_{i+n_T n_T(|\mathcal{U}_I|-1)+1}\}|^2, && (4.122)\\
&\leq \sum_{i=1}^{n_T n_T} |\mathcal{U}_I|^2 \left(|\mathbf{g}_i|^2 + |\mathbf{g}_{i+n_T n_T+1}|^2 + \ldots + |\mathbf{g}_{i+n_T n_T(|\mathcal{U}_I|-1)+1}|^2\right), && (4.123)\\
&= |\mathcal{U}_I|^2 \sum_{i=1}^{n_T n_T |\mathcal{U}_I|} |\mathbf{g}_i|^2, && (4.124)\\
&= |\mathcal{U}_I|^2 \|\mathbf{g}\|_2^2. && (4.125)
\end{aligned}$$

Thus, a sufficient condition to fulfill (4.120) is

$$\alpha_i \|\mathbf{g}\|_2^2 \leq -\frac{1}{2}|\mathcal{U}_I|^2 \|\mathbf{g}\|_2^2 \lambda_{\max}\left(\mathbf{P}_i^T \mathbf{P}_i\right), \quad \forall \mathbf{g}, \tag{4.126}$$

and, finally,

$$\alpha_i \leq -\frac{1}{2}|\mathcal{U}_I|^2 \lambda_{\max}\left(\mathbf{P}_i^T \mathbf{P}_i\right) \leq -\frac{1}{2}|\mathcal{U}_I|^2 \lambda_{\max}^2(\mathbf{H}_i^H \mathbf{H}_i). \tag{4.127}$$

Hence, a possible matrix $\mathbf{M}_i$ satisfying assumptions (A1) – (A4) is, finally,

$$\mathbf{M}_i = -\frac{1}{2}|\mathcal{U}_I|^2 \lambda_{\max}^2(\mathbf{H}_i^H \mathbf{H}_i)\mathbf{I}. \tag{4.128}$$

Algorithm 4.10 Algorithm for solving problem (4.46)

1: initialize $\mathbf{S}^{(0)} \in \mathcal{S}_1$. Set $k = 0$
2: **repeat**
3: generate the $(k+1)$-th tuple $(\mathbf{S}_i^\star)_{\forall i \in \mathcal{U}_I}$ by solving (4.129)
4: set $\mathbf{S}_i^{(k+1)} = \mathbf{S}_i^\star$, $\forall i \in \mathcal{U}_I$, and set $k = k+1$
5: **until** convergence is reached
6: **end algorithm**

4.I Benchmark Formulations and Algorithms Based on the MM Method

In this appendix, we are going to describe the benchmarks based on the works in [Hon16], [Scu14], and [You14]. We start with the benchmark for problem (4.46).

Note that the upper bound $\hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(0)})$ can be used to build a lower bound of $f_0(\bar{\mathbf{S}})$ that fulfills the four conditions (A1) − (A4) presented before in Section 2.4.

By applying a successive approximation of $f_0(\cdot)$ through the application of the previous surrogate function, i.e., $\hat{f}_0(\mathbf{S}, \mathbf{S}^{(k)}) = \sum_{i \in \mathcal{U}_I} \omega_i s_i(\mathbf{S}) - \omega_i \hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(k)}) - \rho \left\| \mathbf{S}_i - \mathbf{S}_i^{(k)} \right\|_F^2$, where $\mathbf{S}^{(k)} \triangleq (\mathbf{S}_i^{(k)})_{\forall i \in \mathcal{U}_I}$, for different evaluation points, we obtain an iterative algorithm based on the MM approach that converges to a stationary point (or local optimum) of the original problem (4.46). Note that we have considered a proximal-like term. Given this, the convex optimization problem to solve is

$$\begin{aligned} \underset{\{\mathbf{S}_i\}}{\text{maximize}} \quad & \sum_{i \in \mathcal{U}_I} \omega_i s_i(\mathbf{S}) - \omega_i \hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(k)}) - \rho \left\| \mathbf{S}_i - \mathbf{S}_i^{(k)} \right\|_F^2 \\ \text{subject to} \quad & \mathbf{S} \in \mathcal{S}_1. \end{aligned} \tag{4.129}$$

We must proceed iteratively until convergence is reached. The procedure is presented in Algorithm 4.10.

Let us now continue with the benchmark for problem (4.48). If we apply the bound from (4.52), i.e., $\hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(0)})$, problem (4.48) can be solved by solving consecutively the following problem:

$$\begin{aligned} \underset{\{\mathbf{S}_i\}}{\text{maximize}} \quad & \sum_{i \in \mathcal{U}_I} \omega_i s_i(\mathbf{S}) - \omega_i \hat{g}_i(\mathbf{\Omega}_i(\mathbf{S}_{-i}), \mathbf{\Omega}_i^{(k)}) + \text{Tr}(\mathbf{R}_H \mathbf{S}_i) - \rho \left\| \mathbf{S}_i - \mathbf{S}_i^{(k)} \right\|_F^2 \\ \text{subject to} \quad & \mathbf{S} \in \mathcal{S}_2. \end{aligned} \tag{4.130}$$

As problem (4.130) is convex, the MM method can be invoked to obtain a local optimum of problem (4.48), following the same procedure as we did before for problem (4.129).

4.J Gradients of Problem (4.46)

Let us start with the gradient of the objective function. As the covariance matrices $\mathbf{S}_i$ have a particular structure (they are positive semidefinite matrices, i.e., $\mathbf{S}_i \succeq 0$, we have to follow the steps presented in [Mag88] to obtain the desired gradient. We will first consider that matrices $\mathbf{S}_i$ are unpatterned matrices denoted by $\tilde{\mathbf{S}}_i$. Then, we will particularize the results for the specific pattern they have. Given this, differential of $f_0(\tilde{\mathbf{S}})$ w.r.t. $\tilde{\mathbf{S}}_\ell$ is

$$\begin{aligned} \mathrm{d}f_0(\tilde{\mathbf{S}}) = & \sum_{i\in\mathcal{U}_I} \omega_i \operatorname{Tr}\left(\left(\mathbf{H}_i \sum_{k\in\mathcal{U}_I} \tilde{\mathbf{S}}_k \mathbf{H}_i^H + \mathbf{I}\right)^{-1} \mathbf{H}_i \mathrm{d}\tilde{\mathbf{S}}_\ell \mathbf{H}_i^H\right) \\ & - \sum_{\substack{i\in\mathcal{U}_I \\ i\neq\ell}} \omega_i \operatorname{Tr}\left(\left(\mathbf{\Omega}_i\right)^{-1} \mathbf{H}_i \mathrm{d}\tilde{\mathbf{S}}_\ell \mathbf{H}_i^H\right), \end{aligned} \tag{4.131}$$

where we have used $\mathrm{d}\log\det(\mathbf{X}) = \operatorname{Tr}(\mathbf{X}^{-1}\mathrm{d}\mathbf{X})$ [Mag88], and the gradient w.r.t. $\tilde{\mathbf{S}}_\ell$ and $\tilde{\mathbf{S}}_\ell^*$ are given, thus, by

$$\nabla_{\tilde{\mathbf{S}}_\ell} f_0(\tilde{\mathbf{S}}) = \sum_{i\in\mathcal{U}_I} \omega_i \mathbf{H}_i^T \left(\mathbf{H}_i \sum_{k\in\mathcal{U}_I} \tilde{\mathbf{S}}_k \mathbf{H}_i^H + \mathbf{I}\right)^{-T} \mathbf{H}_i^* - \sum_{\substack{i\in\mathcal{U}_I \\ i\neq\ell}} \omega_i \mathbf{H}_i^T \left(\mathbf{\Omega}_i\right)^{-T} \mathbf{H}_i^*. \tag{4.132}$$

$$\nabla_{\tilde{\mathbf{S}}_\ell^*} f_0(\tilde{\mathbf{S}}) = \mathbf{0}. \tag{4.133}$$

Now, for the particular case of having Hermitian matrices, the following relation holds

$$\nabla_{\mathbf{S}_\ell} f_0\left(\mathbf{S}\right) = \left[\nabla_{\tilde{\mathbf{S}}_\ell} f_0\left(\tilde{\mathbf{S}}\right) + \left(\nabla_{\tilde{\mathbf{S}}_\ell^*} f_0\left(\tilde{\mathbf{S}}\right)\right)^T\right]_{\tilde{\mathbf{S}}_\ell = \mathbf{S}_\ell}, \tag{4.134}$$

and, since $\mathbf{S}_\ell^* = \mathbf{S}_\ell^T$, it follows that

$$\nabla_{\mathbf{S}_\ell^*} f_0\left(\mathbf{S}\right) = \nabla_{\mathbf{S}_\ell^T} f_0\left(\mathbf{S}\right) = \left(\nabla_{\mathbf{S}_\ell} f_0\left(\mathbf{S}\right)\right)^T \tag{4.135}$$

$$= \left[\nabla_{\tilde{\mathbf{S}}_\ell^*} f_0\left(\tilde{\mathbf{S}}_i\right) + \left(\nabla_{\tilde{\mathbf{S}}_\ell} f_0\left(\tilde{\mathbf{S}}_i\right)\right)^T\right]_{\tilde{\mathbf{S}}_\ell = \mathbf{S}_\ell}. \tag{4.136}$$

Finally, from (4.132), (4.133), and (4.136), it follows that the gradient of $f_0\left(\mathbf{S}\right)$ w.r.t. $\mathbf{S}_\ell^*$ is given by

$$\nabla_{\mathbf{S}_\ell^*} f_0\left(\mathbf{S}\right) = \sum_{i\in\mathcal{U}_I} \omega_i \mathbf{H}_i^H \left(\mathbf{H}_i \sum_{k\in\mathcal{U}_I} \mathbf{S}_k \mathbf{H}_i^H + \mathbf{I}\right)^{-1} \mathbf{H}_i - \sum_{\substack{i\in\mathcal{U}_I \\ i\neq\ell}} \omega_i \mathbf{H}_i^H \left(\mathbf{\Omega}_i\right)^{-1} \mathbf{H}_i. \tag{4.137}$$

Now we follow the same procedure for the constraints. Note that imposing that matrix $\mathbf{S}_i$ is positive semidefinite is the same as imposing that the eigenvalues of $\mathbf{S}_i$ are all non-negative. Given this, the differentials of constraints $C1, C2$, and $C3$ w.r.t. the unpatterned matrix $\tilde{\mathbf{S}}_\ell$ are given by

$$\mathrm{d}C1 = -\operatorname{Tr}\left(\mathbf{H}_j \mathrm{d}\tilde{\mathbf{S}}_\ell \mathbf{H}_j^H\right), \quad \forall j \in \mathcal{U}_E \tag{4.138}$$

$$\mathrm{d}C2 = \operatorname{Tr}\left(\mathrm{d}\tilde{\mathbf{S}}_\ell\right) \tag{4.139}$$

$$\mathrm{d}C3 = \mathrm{d}\lambda_\ell^{(k)}\left(\tilde{\mathbf{S}}_\ell\right) = -\mathbf{v}_\ell^{H(k)} \mathrm{d}\tilde{\mathbf{S}}_\ell \mathbf{v}_\ell^{(k)}, \quad \forall \ell \in \mathcal{U}_I, \forall k, \tag{4.140}$$

where $\lambda_\ell^{(k)}(\cdot)$ is the k-th eigenvalue of the ℓ-th covariance matrix and $\mathbf{v}_\ell^{(k)}$ is the eigenvector associated with the k-th eigenvalue. Note that we have used the identity $\mathrm{d}\lambda_i(\mathbf{X}) = \mathbf{v}_i^H \mathrm{d}\mathbf{X} \mathbf{v}_i$ [Mag88]. The gradients w.r.t. $\tilde{\mathbf{S}}_\ell$ and $\tilde{\mathbf{S}}_\ell^*$ are:

$$\nabla_{\tilde{\mathbf{S}}_\ell} C1 = -\mathbf{H}_j^T \mathbf{H}_j^*, \quad \forall j \in \mathcal{U}_E \tag{4.141}$$

$$\nabla_{\tilde{\mathbf{S}}_\ell} C2 = \mathbf{I} \tag{4.142}$$

$$\nabla_{\tilde{\mathbf{S}}_\ell} \lambda_\ell^{(k)}(\mathbf{S}_\ell) = -\mathbf{v}_\ell^{*(k)} \mathbf{v}_\ell^{T(k)}, \quad \forall \ell \in \mathcal{U}_I, \quad \forall k \tag{4.143}$$

$$\nabla_{\tilde{\mathbf{S}}_\ell^*} C1 = \mathbf{0}, \quad \forall j \in \mathcal{U}_E \tag{4.144}$$

$$\nabla_{\tilde{\mathbf{S}}_\ell^*} C2 = \mathbf{0} \tag{4.145}$$

$$\nabla_{\tilde{\mathbf{S}}_\ell^*} \lambda_\ell^{(k)}(\mathbf{S}_\ell) = \mathbf{0}, \quad \forall \ell \in \mathcal{U}_I, \quad \forall k. \tag{4.146}$$

Finally, from (4.142)-(4.146), and (4.136) it follows that the gradients w.r.t. $\mathbf{S}_\ell^*$ are given by

$$\nabla_{\mathbf{S}_\ell^*} C1 = -\mathbf{H}_j^H \mathbf{H}_j, \quad \forall j \in \mathcal{U}_E \tag{4.147}$$

$$\nabla_{\mathbf{S}_\ell^*} C2 = \mathbf{I} \tag{4.148}$$

$$\nabla_{\mathbf{S}_\ell^*} \lambda_\ell^{(k)}(\mathbf{S}_\ell) = -\mathbf{v}_\ell^{(k)} \mathbf{v}_\ell^{H(k)}, \quad \forall \ell \in \mathcal{U}_I, \quad \forall k. \tag{4.149}$$

Part III

Energy Harvesting Techniques at the Transmitter Side

Chapter 5

Energy Dimensioning Methodology and Dynamic Base Station On-Off Mechanisms for Sustainable Wireless Networks

5.1 Introduction

Remote rural areas, characterized by low population densities, have generally been disregarded by cellular operators because classical technologies do not assure the return of investment. The difficult access to these areas, the high cost derived from the required network equipment and the energy consumption, and the low density of customers are the main factors resulting in low profit business. On the other hand, cellular communications have become essential for human development as they provide a way to have internet access, a key factor to favor the improvement of the economy in addition to being the basis for key social services, such as remote-health care programs. Therefore, finding innovative solutions to connect these areas is an utmost necessity and is becoming one of the strategy objectives of many Governments.

Frequently, access to the fixed electric power grid in such isolated areas is difficult (usually due to geographical or economic reasons). For this reason, an appealing solution would be to use BSs that are entirely powered by energy harvesting sources and batteries to make the whole access network energy sustainable and low-cost. There are different energy harvesting technologies available (see [Par05]), but the one considered in this work is based on solar energy. In this framework, any deployment strategy that reduces the energy consumption becomes also a fundamental element to reduce costs. In this sense, switching off BSs whenever possible contributes to reduce the energy consumption. These energy savings are directly translated into a more efficient energetic dimensioning (i.e., reduction of the capacities of the batteries and the sizes of solar cells) and, thus, a reduction of the capital expenditures (CAPEX). Usually, the

energy dimensioning is devised assuming that the BSs are always active, but this is not always necessary if a criterion for on/off switching is designed properly while guaranteeing a minimum pre-defined QoS in terms, for example, a maximum blocking probability of incoming calls. Note that switching off a particular BS affects the network topology in the sense that the traffic that was originally served by the BS to be switched off has to be either transferred to nearby BSs or dropped. One important characteristic of the traffic profile, both for voice and data services, is that the rate of traffic generation is not constant throughout the day. This characteristic could be exploited to define a strategy based on a dynamic selection of the BSs to be switched off when the traffic load is low.

5.1.1 Related Work

Based on the previous motivations, we propose a procedure for the dimensioning of the energy units (solar panels and battery sizes) and for switching off BSs whenever possible (without affecting the QoS of the users in the network) in order to be able to deploy smaller solar panels and batteries and, thus, to reduce the cost of the equipment. In the literature, there are some works dealing with the problem of switching on/off BSs. For example, in [Soh13] a strategy is developed to decrease the energy consumption by switching off BSs when the activity is low under the constraint of keeping the coverage unaltered. The strategies are developed within the framework of stochastic geometry and, therefore, are well suited for the case of having many BSs at random positions, which does not fit the rural scenarios considered in this work where the number of deployed BSs is rather low. In [Oh10], a strategy is presented taking into account that the traffic profile is time varying and under the objective of minimizing the energy consumption of the network assuming that there are many BSs uniformly distributed within the area of interest. [Guo12] defines different configuration for the BSs (active mode, sleep mode, expanded coverage, etc.) in order to develop a switching strategy between the states depending on the instantaneous traffic load. This is achieved by expanding the coverage areas of the BSs that remain active. In [Bou12a], [Bou12b], the authors propose a sleeping algorithm for the BSs assuming that the distances between the mobile terminals and their associated BSs are known. A more complex problem is analyzed in [Bou13], where a scenario with several BSs from different operators are considered. That paper introduces the cost that has to be paid by an operator when its subscribers have to be served by another operator due to the fact that some BSs have been switched off. It is important to remark that in the previous works, the decision to switch off BSs was based only on the traffic demand. As opposed to that, since in the work presented here the BSs are powered with finite batteries, the decision to switch off BSs must be carried out under the criterion of minimizing the overall energy consumption.

5.1.2 Main Contribution

The access network configuration considered in this chapter is based on two BSs placed at the same cell-site, i.e., at the same tower, with coverage areas fully overlapped, and operating at different frequencies so that they do not interfere with each other[1]. The main contributions of this chapter w.r.t. the literature is:

- Proposal of a methodology for dimensioning the energy units, i.e., solar panels and batteries. The output of the dimensioning strategy is the required number of units along with their sizes.

- Development of a dynamic on/off switching strategy that provides a size reduction of the energy units. The decision to switch off one of the BSs is based on the required power and not just on the traffic demand.

- Two different approaches to determine the on/off switching threshold are derived: a deterministic approach, where we assume that full knowledge of the traffic profile is available, and a more realistic statistical robust approach, which accounts for possible errors in the traffic estimation and modeling. Both approaches are derived for the case of having a single type of traffic (voice) and mixed traffic (voice and data).

- Algorithms are validated using extensive network simulations, considering a real scenario with the corresponding channel attenuations given by the specific location, and with a real network model and equipment parameters.

5.1.3 Organization of the Chapter

The remainder of the chapter is organized as follows. In Section 5.2, the system model is described. In Section 5.3, we derive the procedure for the dimensioning of the energy units. Section 5.4 presents the methodology for switching on/off BSs considering two different approaches: a deterministic approach assuming perfect knowledge of the traffic profile and a robust approach for situations where the traffic profile is partially known. Section 5.5 presents numerical results and, finally, conclusions are drawn in Section 5.6.

Figure 5.1: Scenario with two BSs placed at the same communications tower with fully overlapped coverage areas.

5.2 System Model

As commented in the introduction, we consider a system with two BSs placed at the same location, i.e., sharing the same cell-site[2]. Let us also consider that their coverage areas are fully overlapped and a different frequency is configured in each BS so that they do not interfere with each other (see Figure 5.1).

This setup is usually required in scenarios where there is a peak traffic of demand in a specific location and there are not enough radio resources available to be allocated in a single BS to cope with such users' demands. As commented in the introduction, we assume that the BSs are provided with a finite battery and with solar panels that allow them to recharge their batteries.

As it is widely known, the evolution of the traffic throughout the day is not stationary, i.e., the traffic is high at particular peak hours and considerably low at nights. As a consequence, it may be reasonable to switch off one of the BSs if the required power for having the two BSs on is higher, and only one BS is enough to provide the demanded QoS.

Later in this chapter, we will present a switching off procedure that provides energy savings (and CAPEX reduction). For the moment, let us mention that in order to determine whether a BS should be switched off or not, we need to measure and compare the required power that is needed for both configurations, i.e., a single BS or two BSs, to serve the users with a specific

[1]Note that the methodology that we present in this chapter can also be applied to the case in which a single BS with two different carriers having coverage areas fully overlapped is deployed. As we will explain later, the work developed in this chapter is part of a project in which a real network was deployed and, due to several reasons, the solution of having two different BSs at the same tower was chosen in this case. For this reason, we will consider that network configuration throughout the chapter.

[2]The procedure described in this chapter is valid not only for two BSs at the same location, but for any number of BSs placed at the same site. We consider only two BSs for the sake of clarity in the notation.

(a) Aerial view of a target location. (b) Aerial view of a target location.

Figure 5.2: Real landscape of the scenarios under consideration in this work. These locations correspond to rural areas in the Amazon forest of Perú.

traffic demand since the configuration of a single BS is only admissible if one BS can provide the service required by the users with a blocking probability lower than a pre-established value related with the QoS. The estimation of the required power can be obtained through measurement campaigns or with Monte Carlo simulations if the specifications of the deployment are known (e.g., type of BSs, average number of active users, density of users per unit area, etc.). We will provide some steps towards the computation of such powers as they are needed for the dimensioning of the energy units [Ren11].

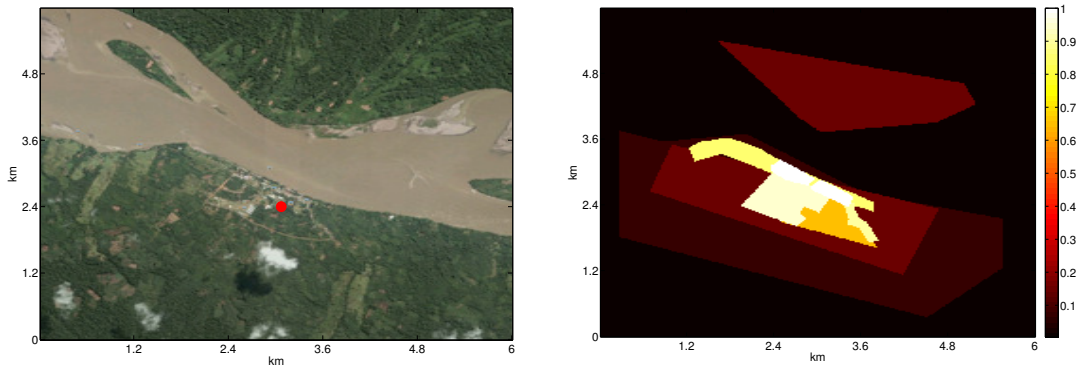
The work performed in this chapter has been developed within the framework of the European project TUCAN3G (http://www.ict-tucan3g.eu). The goal of this project is to deploy a network of BSs to provide 3G coverage in some rural areas in the Amazon forest in Perú. In this chapter, we will assume that the access network dimensioning has already been done (see [tuc13] for the methodology and results) and that a specific setup, (e.g., BS type, number of available circuits, number of codes, coverage area, user distribution, QoS, etc.) is already pre-established. In the following, we will show a few representative figures regarding the scenario under consideration in which the energy dimensioning methodology and the switching strategy have been implemented.

Figure 5.2 depicts the view of the landscape of the real locations considered in this work. This image was taken from the telecommunications towers where the BSs are placed.

The real BSs being deployed as well as the batteries and the solar panels are presented in Figure 5.3. In the picture, the white box represents the battery that is used to power the BSs and the rectangular white instrument on top of the picture corresponds to the solar panel. In the picture, we can also see a third antenna that is basically acting as a wireless backhaul link that connects several towns of the area with the cellular core network. For the specifics of the network deployment, see http://www.ict-tucan3g.eu.

Figures 5.4 and 5.5 present the satellite view corresponding to two specific towns (Santa

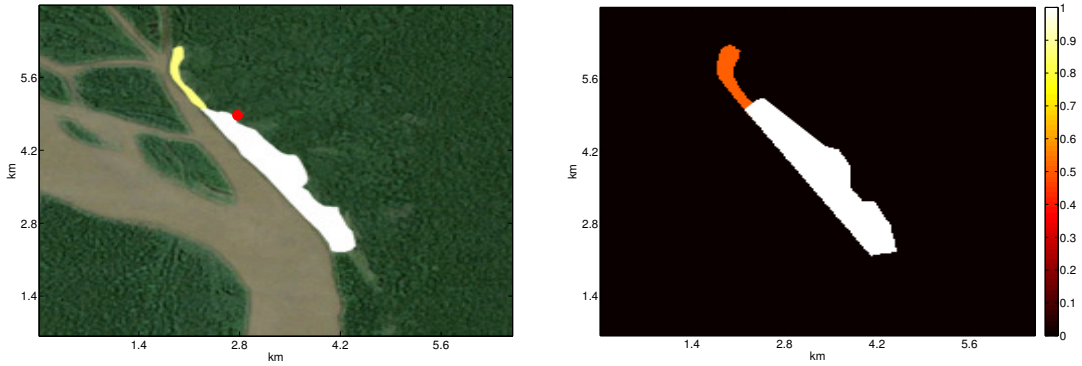
Figure 5.3: Picture showing the real BS, battery, and solar panel employed in the deployment.



(a) Satellite view of Santa Clotilde.

(b) Expected traffic distribution in Santa Clotilde.

Figure 5.4: Satellite view and traffic distribution of one of the target scenarios, Santa Clotilde, in the forest in Perú.

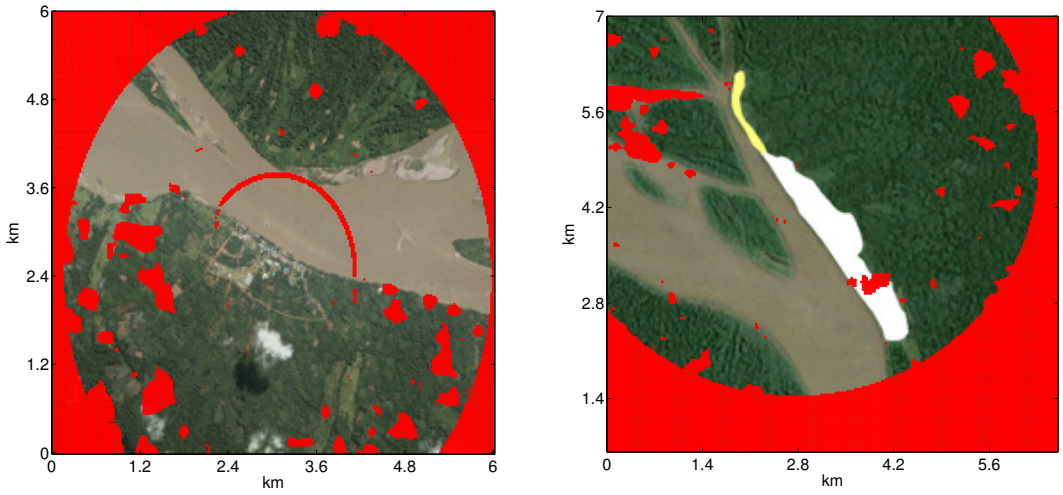


(a) Satellite view of Tuta Pisco.

(b) Expected traffic distribution in Tuta Pisco.

Figure 5.5: Satellite view and traffic distribution of one of the target scenarios, Tuta Pisco, in the forest in Perú.

Clotilde and Tuta Pisco) with the exact position of the BS and the expected traffic distribution over the area.



(a) Santa Clotilde.

(b) Tuta Pisco.

Figure 5.6: Provided coverage with two BSs deployed in two different target locations.

Finally, Figure 5.6 presents the coverage provided in both locations, Santa Clotilde and Tuta Pisco, based on the network access dimensioning that guarantees the minimum QoS in the 95% of the region when 2 BSs are deployed. Red spots in the figures represent zones where QoS is not provided. For a more detailed analysis, see [tuc13].

5.3 Energy Provision and Energy Systems Dimensioning

The goal of this section is to determine the number of required solar panels and batteries denoted as n_{SP} and n_B, respectively, to provide energy to the BSs. In order to carry out this, we need to compute the power required by both setups, i.e., considering that just one BS is active and that two BSs are active. For the computation of this power it is required to know how the access network has been modeled. Without going into much details (the complete description can be found in [tuc13]), let us say that the access network dimensioning has been carried out using a Markov chain model. The objective of the access network dimensioning is to configure different access network parameters (number of wideband code division multiple access (WCDMA) codes and circuits) in order to minimize the required backhaul bandwidth, while satisfying the access network performance constraints in terms of a maximum blocking probability. Regarding the Markov model, each state of the model represents the simultaneous voice and data connections given by n and m, respectively, at each BS. The links between the states depend on the average aggregated rates of incoming calls and packets and the coverage probability. In this sense, the coverage probability is defined as the probability of being able to have enough power to accommodate a new incoming voice call or data packet. In the following, let us assume that the possible states of the Markov chain are fixed and known, and that they compound the region $\mathcal{S}$. For specific details of the Markov modeling, see [tuc13]. Let n and m denote number of voice users and data users served by one of the two BS, respectively, and let k and l denote number

of voice users and data users served by the other BS. In this sense, (n, m, k, l) defines a state in the Markov chain.

5.3.1 Energy Computation

For the moment, let us consider that there is only one BS active and that there is a maximum number of voice and data users allowed in the network given by $(C_{\mathrm{v}}, C_{\mathrm{d}})$. Let $\lambda_{\mathrm{v}}(h)$ and $\lambda_{\mathrm{d}}(h)$ be the functions representing the average aggregated arrival call rate for voice users and data users at hour h, respectively. As the call rates vary through the day, the state probabilities will also be time varying functions, denoted as $P_{n,m}(h)$. Now, let $f\left(P_R|n,m\right)$ denote the PDF of the power radiated (P_R) by the BS, conditioned to having n voice and m data users simultaneously connected in the system (note that, the average power transmitted by the BS, $\bar{P}_i^{\mathrm{BS}}$, will depend on the users' positions within the coverage area). Notice that $f\left(P_R|n,m\right), \forall (n,m) \in \mathcal{S}$ should be obtained through Monte Carlo simulations since an analytical expression does not exist. Therefore, the PDF of the total radiated power conditioned to a particular hour of a day can be expressed as

$$f\left(P_R|h\right) = \sum_{(n,m)\in\mathcal{S}} f\left(P_R|n,m\right) P_{n,m}(h), \tag{5.1}$$

and the average power radiated by the BS in one particular hour is given by

$$P_{\mathrm{rad}}(h) = \int P_R \, f\left(P_R|h\right) \mathrm{d}P_R. \tag{5.2}$$

In order to dimension the energy units, we need to model the total BS power consumption. The model considered in this chapter is [Aue11]:

$$P_c = \begin{cases} N_{TRX}\left(P_0 + \Delta_p P_{\mathrm{rad}}\right), & 0 < P_{\mathrm{rad}} \leq P_{\max}, \\ N_{TRX} P_{\mathrm{sleep}}, & P_{\mathrm{rad}} = 0, \end{cases} \tag{5.3}$$

where $P_{\max}$ is the maximum radiated power, P_0 is the fixed power consumed by the components of the RF chain, the cooling system, and the baseband processing, Δ_p is a constant, and N_{TRX} is the number of RF chains. P_{sleep} is the power consumed by the BS when it is in off mode. Given that, the total average energy consumption (in W·h) is (assuming that the BS is on for the whole day):

$$L = \sum_{h=0}^{23} P_c(h) = 24 \cdot N_{TRX} P_0 + N_{TRX} \Delta_p \sum_{h=0}^{23} \int P_R \, f\left(P_R|h\right) \, \mathrm{d}P_R. \tag{5.4}$$

The previous development was done assuming that only one BS was active. However, the two BSs will be needed active at some hours during the day and, thus, we need to consider this

for the dimensioning of the energy units. If the two BSs are switched on, the state probabilities are denoted as $P_{n,m,k,l}(h)$. The PDF of the total radiated power conditioned to a particular hour of a day can be expressed as

$$f\left(P_R|h\right) = \sum_{(n,m,k,l)\in\mathcal{S}} f\left(P_R|n,m,k,l\right) \cdot P_{n,m,k,l}(h), \tag{5.5}$$

where $f\left(P_R|n,m,k,l\right)$ denotes the PDF of the power radiated by the BS conditioned to having n and k simultaneous voice users, and m and l simultaneous data users connected to both BSs. Then, the average energy consumption can be computed also using (5.4).

5.3.2 Energy Dimensioning

The energy dimensioning procedure followed in this chapter is based on [Ren11]. Considering the average energy consumption obtained in (5.4), we have that the total energy that the solar panels have to generate within one day is:

$$E_{SP} = L(1+\eta_G)f_c, \tag{5.6}$$

where η_G denotes the losses due to the inefficiency of the solar panels (around 10%) and f_c is a correction factor ($f_c = 1.3$) introduced as the solar panels have to generate energy for the system and for the batteries themselves that are accumulating the energy. The number of solar panels depends on the solar radiation in the place of interest and on the nominal power of the solar panel, P_{nom}, which is obtained assuming that the solar radiation is $1,000$ W/m^2. Given that, the number of necessary solar panels is:

$$n_{SP} = \frac{L(1+\eta_G)f_c}{P_{\text{nom}}\frac{G_{\text{dm}}}{1{,}000}}, \tag{5.7}$$

where G_{dm} is taken as the average daily solar radiation during the worst month of the year. The capacity of the batteries (in W·h) should be enough to provide electrical energy to the system to be up to N_{da} days without any charging procedure. Therefore the capacity of the batteries must satisfy:

$$C_B = L(1+\eta_G)\frac{N_{\text{da}}}{P^d_{\text{max}}}, \tag{5.8}$$

where P^d_{max} is a parameter to impose that a battery should have at least 20% of its maximum capacity ($P^d_{\text{max}} = 0.8$). Finally, if C_1 is the capacity of a single battery, then, the number of required batteries becomes

$$n_B = \left\lceil \frac{C_B}{C_1} \right\rceil, \tag{5.9}$$

where $\lceil\cdot\rceil$ is the ceiling operator.

5.4 BS On/Off Switching Strategies

In this section, we introduce a technique to dynamically switch on and off one of the two BSs. As it was commented in the introduction, the BSs are assumed to be equipped solely with solar panels. In order to have a network as sustainable as possible, the decision to switch off one of the BSs must be based on the required power and not just on the traffic demand. The reason is that, for example, a situation could be possible where all the traffic could be served with just one BS (if there are enough resources to be allocated among the users), but it could happen that in order to compensate the intra-cell interference (we are assuming a WCDMA system), two BSs operating at two different frequencies would require less power to serve such traffic (we also have to take into account that the BS has a fixed power consumption due to the electronic equipment, see (5.3)). Thus, if the power increase to overcome interference is higher than the fixed power consumption of the BS, then, the configuration with two BSs would be optimal in terms of energy reduction. This is just an example, but other situations may be possible.

In order to determine whether a BS should be switched off or not, we need to measure and compare the required power that is needed for both configurations, i.e., a single BS or two BSs, to serve the users with a specific traffic demand. Note that the configuration of single BS is only admissible if with one BS we can provide the service required by the users with a given QoS.

We will present two different approaches: a deterministic approach, where we consider that full knowledge of the traffic profile is available, and a statistical robust strategy that accounts for possible error traffic modeling.

In the following, we present the deterministic procedure developed in this chapter to obtain the switching threshold when we have a single type of traffic (voice) and mixed traffic (voice and data).

5.4.1 Deterministic Switching Strategies

5.4.1.1 Single Type of Traffic

In this first approach we assume that the expected traffic demands are completely known. Let $\lambda_{0v}[m]$ represent the traffic load (in calls/second) corresponding to the m-th time instant within the day. This corresponds to a discretized version of the aggregated arrival call rate $\lambda_{\mathrm{v}}(h)$ since time instant m could be in general different than h. In Algorithm 5.1 we present the procedure to calculate the threshold λ_{TH}^{v} to be applied to $\lambda_{0v}[m]$ for switching on/off a BS when the traffic profile is assumed to be known. Note that, when there is only a single type of traffic, the threshold λ_{TH} is just a scalar value that needs to be compared with the single traffic demand. Figure 5.7 presents the threshold computation with a graphic example. In the left figure, we

Algorithm 5.1 Threshold computation for switching on/off BSs with single traffic

1: Compute the mean power required by the two configurations (one and two BSs) for all possible traffic rates (λ_0), $P_{1\text{BS}}(\lambda_0)$ for 1 BS and $P_{2\text{BS}}(\lambda_0)$ for 2 BSs

2: Let λ be the maximum traffic that can be supported with one BS fulfilling the maximum blocking probability constraint. If $P_{1\text{BS}}(\lambda) < P_{2\text{BS}}(\lambda)$, then $\lambda_{TH} = \lambda$. Otherwise, λ_{TH} is the value of λ for which $P_{1\text{BS}}(\lambda) = P_{2\text{BS}}(\lambda)$.

3: Switch off one of the two BSs in all time instants m where $\lambda_{TH} \geq \lambda_0$.

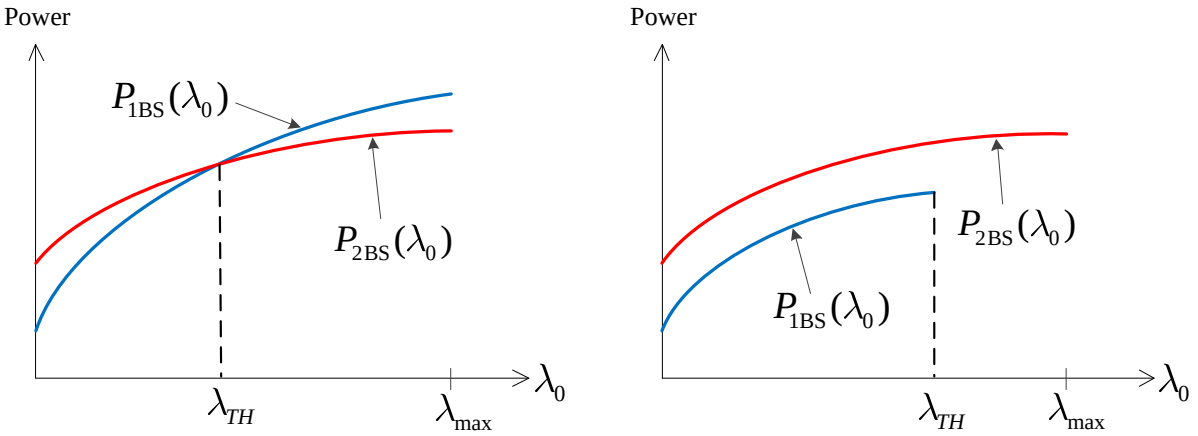


Figure 5.7: Determination of the switching threshold for a single type of traffic.

have the case where with a single BS we can serve the maximum traffic experienced at the peak hour, denoted by $\lambda_{\max}$. In the right figure, we have the scenario where with a single BS we cannot serve the whole traffic required by the users at all time instants. In this case, both curves of power consumption may not cross each other as the figure presents. If that happens, then the switching threshold is determined by the maximum traffic that can be served with a single BS.

5.4.1.2 Two Types of Traffic

In this case we consider that we have two different types of traffic, voice and data, denoted as $\lambda_{0T}[m] = (\lambda_{0d}[m], \lambda_{0v}[m])$. Now $\lambda_{0d}[m]$ represents the discretized version of the aggregated arrival packet rate $\lambda_{\text{D}}(h)$. Note that $\lambda_{0d}[m]$ is different for the UL and the DL. Therefore, we will only consider the DL $\lambda_{0d}[m]$ to design the switching threshold. In Algorithm 5.2 we present the procedure to calculate the threshold frontier $\boldsymbol{\lambda}_{TH}$ to decide whether for a given traffic $\lambda_{0T}[m]$ we have to switch on or off a BS. Now, the threshold $\boldsymbol{\lambda}_{TH}$ is a frontier in a two dimensional plane composed of tuples of voice and data traffic demands, (λ_{0d}, λ_{0v}). The set of points belonging to the threshold frontier are calculated as described in Algorithm 5.2. Figure 5.8 depicts an example of a graphic representation of a threshold frontier.

Algorithm 5.2 Threshold computation for switching on/off BSs with two traffics

1: Define $\boldsymbol{\lambda}_{TH} = \emptyset$

2: Compute the mean power required by the two configurations (one and two BSs) for all possible traffic rates $(\lambda_{0v}, \lambda_{0d})$, $P_{1\text{BS}}(\lambda_{0v}, \lambda_{0d})$ and $P_{2\text{BS}}(\lambda_{0v}, \lambda_{0d})$.

3: For a given λ_{0v} let λ_{0d}^{m} be the maximum data traffic rate that can be supported with one BS fulfilling the maximum blocking probability constraint.
Let $\lambda = (\lambda_{0v}, \lambda_{0d}^{m})$. If $P_{1\text{BS}}(\lambda) < P_{2\text{BS}}(\lambda)$ then $\boldsymbol{\lambda}_{TH} \longleftarrow \boldsymbol{\lambda}_{TH} \cup \lambda$.
Otherwise, λ_{0d}, $(\lambda = (\lambda_{0v}, \lambda_{0d}))$, is the data traffic value for which $P_{1\text{BS}}(\lambda) = P_{2\text{BS}}(\lambda)$.
Then, $\boldsymbol{\lambda}_{TH} \longleftarrow \boldsymbol{\lambda}_{TH} \cup \lambda$. Repeat for all possible λ_{0v}.

4: At a given time instant where the traffic loads are $(\lambda_{0v}, \lambda_{0d})$, one of the BSs will be switched off if there exists a point in the threshold frontier, $(\lambda_{0v,T}, \lambda_{0d,T}) \in \boldsymbol{\lambda}_{TH}$, such that $\lambda_{0v,T} \geq \lambda_{0v}$ and $\lambda_{0d,T} \geq \lambda_{0d}$.

5.4.2 Robust Switching Strategy

In the previous subsection, we considered that the value of the rate of the traffic generation was perfectly known, i.e., $\lambda_{0v}[m]$ and $\lambda_{0d}[m]$. However, in some situations, it is very difficult to obtain such information with a high precision. For already deployed networks, measurements campaigns can be performed to obtain the current traffic profile throughout the day. Nevertheless, the expected value of traffic is something that is not constant over time and usually depends on whether there is an special event in that location, or even on the particular day of the week. As a consequence, slight variations of the nominal traffic profile will be experienced. On the other hand, in locations where the network is not already deployed, the uncertainty is even higher. Network planning engineers usually take a nominal traffic in a similar already deployed network. This implies the risk that such traffic information is not an accurate approximation of the actual traffic profile of the new location. For these reasons, in this section, we propose a new model of the traffic profile to cope with such uncertainties and make the final switching off procedure robust against all possible uncertainties in the traffic model. For simplicity in the notation and the mathematical developments, throughout this section, we will consider that there is only one type of traffic and some ideas on how to extend the methodology to mixed traffics will be presented later.

Let us consider that the traffic profile is a stochastic discrete process modeled as a deterministic mean component plus a random component as follows:

$$\lambda[m] = \lambda_0[m] + p[m], \tag{5.10}$$

where $\lambda_0[m]$ corresponds to the available mean (nominal) profile and $p[m]$ is a zero mean white

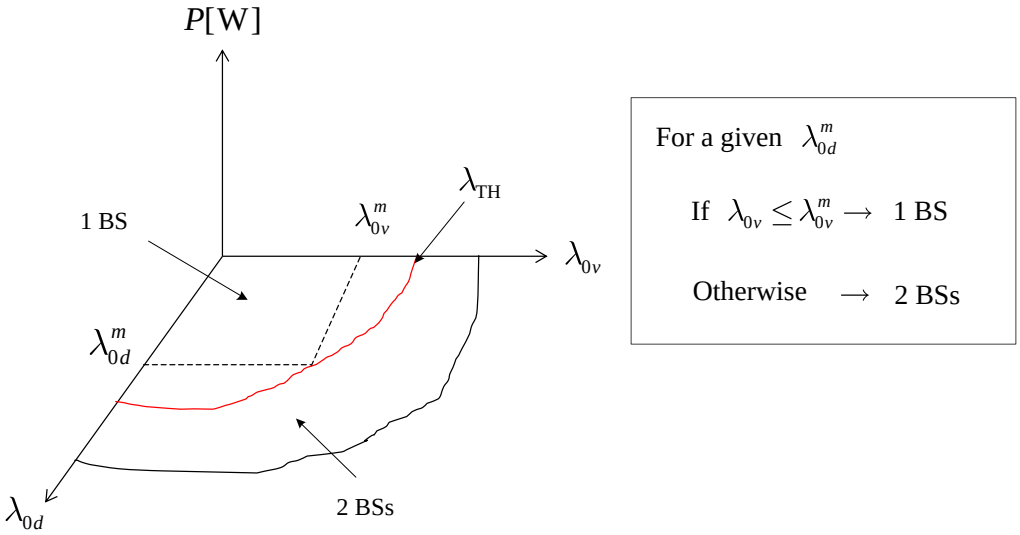


Figure 5.8: Graphic representation of the threshold computation for the deterministic case with two types of traffic.

Gaussian process with variance σ_p^2 (in reality, σ_p^2 may also vary over the day, i.e, $\sigma_p^2[m]$, but for the ease of notation, we will consider it constant). The stochastic process $p[m]$ incorporates all possible uncertainties presented before.

The goal of this section is to obtain a new threshold for switching on/off the BSs fulfilling a given predetermined target outage probability. The decision of shutting down the BS will be based on the comparison of an estimate of $\lambda[m]$ with the new threshold. In order to estimate the traffic being generated, we have to define a time window where we will measure such demanded traffic. Let M be the number of time instants that are considered for the measurement of the traffic and T be the separation time between consecutive time instants.

Let us consider the following definition:

$$\bar{\lambda}_t \triangleq T \sum_{m=t}^{t+M-1} \lambda[m] = T \left(\sum_{m=t}^{t+M-1} \lambda_0[m] + \sum_{m=t}^{t+M-1} p[m] \right) \bar{\lambda}_{0t} + \bar{p}_t, \tag{5.11}$$

where $\bar{p}_t \sim \mathcal{N}(0, \bar{\sigma}_p^2)$, $\bar{\sigma}_p^2 = MT^2\sigma_p^2$. In order to avoid $\bar{\lambda}_t$ being negative, we truncate the Gaussian distribution of $\bar{p}_t$ in a way that $-\bar{\lambda}_{0t} < \bar{p}_t < \infty$. Then, $\bar{p}_t$ has a truncated normal distribution with PDF given by

$$f\left(\bar{p}_t, \bar{\sigma}_p^2, -\bar{\lambda}_{0t}\right) = \frac{1}{\Phi_t \sqrt{2\pi}\bar{\sigma}_p} \mathrm{e}^{-\frac{\bar{p}_t^2}{2\bar{\sigma}_p^2}}, \quad -\bar{\lambda}_{0t} < \bar{p}_t < \infty, \tag{5.12}$$

where $\Phi_t \triangleq \frac{1}{\sqrt{2\pi}\bar{\sigma}_p} \int_{-\bar{\lambda}_{0t}}^{\infty} \mathrm{e}^{-\frac{\bar{p}_t^2}{2\bar{\sigma}_p^2}} \, \mathrm{d}\bar{p}_t$.

5.4.2.1 Traffic Profile Estimation

Let us consider that we have N measurements (for example, from N previous days) of the number of generated calls, denoted as k_{tn}, $n = 1, \ldots, N$, during the same time period to estimate $\bar{\lambda}_t$ (k_{tn} corresponds to the number of calls measured within the window time $[t, t+M-1]$ of the day n). We assume that such variable is Poisson distributed where the probability of a given value of k_{tn} is given by $\Pr(k_{tn}) = \prod_{n=1}^{N} \mathrm{e}^{-\bar{\lambda}_t} \frac{(\bar{\lambda}_t)^{k_{tn}}}{k_{tn}!}$. We formulate the maximum likelihood (ML) estimator of $\bar{\lambda}_t$ given the previous observations as [Kay93]

$$\hat{\lambda}_{ML}(t) = \arg\max_{\bar{\lambda}_t} \left(\prod_{n=1}^{N} \mathrm{e}^{-\bar{\lambda}_t} \frac{(\bar{\lambda}_t)^{k_{tn}}}{k_{tn}!} \right). \tag{5.13}$$

Given that, the ML estimator $\hat{\lambda}_{ML}(t)$ can be easily obtained as

$$\hat{\lambda}_{ML}(t) = \frac{1}{N} \sum_{n=1}^{N} k_{tn}. \tag{5.14}$$

Notice that, in the previous expression, $\sum_{n=1}^{N} k_{tn}$ is Poisson distributed[3] with expected value $(\bar{\lambda}_{0t} + \bar{p}_t)N$.

5.4.2.2 Threshold Computation Based on Outage Probability

Given the previous traffic estimate, the Bayesian robust outage condition $\bar{P}_{\text{out}}$ is defined as

$$\bar{P}_{\text{out}} = \mathbb{E}_{\bar{p}_{t_w}} \left[\mathbb{P} \left(\hat{\lambda}_{ML}(t_w) \leq \bar{\lambda}'_{TH} \,\middle|\, \bar{\lambda}_{t_w} \right) \right] \leq \theta, \tag{5.15}$$

where $\bar{\lambda}_{t_w} = \bar{\lambda}_{0t_w} + \bar{p}_{t_w}$, θ is the outage probability with which we tolerate that 1 BS goes off when, in reality, the two BSs should be on, $\bar{\lambda}'_{TH}$ is the new threshold to be computed, and t_w is the worst case time instant that defines the time window as

$$t_w = \arg\min_{t} \left\{ \sum_{m=t}^{t+M-1} \lambda[m] \,\middle|\, \exists\, t' \in [t+M-1,\, t+M), \text{ with } \lambda(t') > \lambda_{TH} \right\}, \tag{5.16}$$

where λ_{TH} is the deterministic threshold found in the previous section. As the outage condition should be fulfilled for the whole traffic profile over the day, we have to find the time window where the estimate $\hat{\lambda}_{ML}(t)$ will be most susceptible to originate situations in which the BS should be on but the estimate decides to turn it off. Notice that a trivial solution of the previous outage probability is to set $\bar{\lambda}'_{TH} = 0$. However, in this situation both BSs will be always on and no energy saving is possible. For that reason, the threshold $\bar{\lambda}'_{TH}$ should be the maximum value

[3] Recall that if $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\nu)$, then $Z = X + Y \sim \text{Poisson}(\lambda + \nu)$.

that fulfills the following condition:

$$\mathbb{E}_{\bar{p}_{t_w}}\left[\mathbb{P}\left(\sum_{n=1}^{N} k_{t_w n} \leq \bar{\lambda}_{TH}^{'} N \,\middle|\, \bar{\lambda}_{t_w}\right)\right] \leq \theta. \tag{5.17}$$

Note that (5.17) has a unique solution because it is an increasing monotonic function in $\bar{\lambda}_{TH}^{'}$ since we are averaging a CDF (the sum of monotone functions yields a monotone function). The last step requires the normalization: $\lambda_{TH}^{'} = \frac{\bar{\lambda}_{TH}^{'}}{MT}$ to get units of number of calls/s. Let us present the following result:

Proposition 5.1. *The Bayesian outage probability, $\bar{P}_{\text{out}}$, can be approximated in closed-form solution by the following expression:*

$$\begin{aligned}\mathbb{E}_{\bar{p}_{t_w}}\left[\mathbb{P}\left(\sum_{n=1}^{N} k_{t_w n} \leq \bar{\lambda}_{TH}^{'} N \,\middle|\, \bar{\lambda}_{0t_w} + \bar{p}_{t_w}\right)\right] \approx \\ K\mathcal{Q}\left(\frac{-\bar{\lambda}_{0t_w}}{\bar{\sigma}_p}\right)\Gamma(x) - K\,\gamma(x, N\bar{\lambda}_{0t_w}) - K_1\Delta_1\,\Psi\left(\bar{p}_{t_w}, \mu_1, \frac{1}{\sqrt{2}}, \bar{\lambda}_0, \infty, x-1\right) \\ - K_2\Delta_2\,\Psi\left(\bar{p}_{t_w}, \mu_2, \frac{1}{\sqrt{2}}, \bar{\lambda}_{0t_w}, \infty, x-1\right) + K_1\Delta_3\,\Psi\left(\bar{p}_{t_w}, \mu_1, \frac{1}{\sqrt{2}}, 0, \bar{\lambda}_{0t_w}, x-1\right) \\ + K_2\Delta_4\,\Psi\left(\bar{p}_{t_w}, \mu_2, \frac{1}{\sqrt{2}}, 0, \bar{\lambda}_{0t_w}, x-1\right),\end{aligned} \tag{5.18}$$

where $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \mathrm{e}^{\frac{u^2}{2}}\,\mathrm{d}u$ is the Gaussian $\mathcal{Q}$-function [Chi03], $x = \lceil \bar{\lambda}_{TH}^{'} N \rceil + 1$, $\Gamma(x) = \int_0^\infty s^{x-1}\mathrm{e}^{-s}\,\mathrm{d}s$ is the Gamma function [Abr72], $\gamma(n,x) = \int_0^x s^{n-1}\mathrm{e}^{-s}\,\mathrm{d}s$ is the lower incomplete Gamma function [Abr72], $\Delta_1 = \frac{1}{\sqrt{\pi}}\int_{\bar{\lambda}_0}^\infty \mathrm{e}^{-(m-\mu_1)^2}\,\mathrm{d}m$, $\Delta_2 = \frac{1}{\sqrt{\pi}}\int_{\bar{\lambda}_0}^\infty \mathrm{e}^{-(m-\mu_2)^2}\,\mathrm{d}\bar{m}$, $\Delta_3 = \frac{1}{\sqrt{\pi}}\int_0^{\bar{\lambda}_0} \mathrm{e}^{-(m-\mu_1)^2}\,\mathrm{d}\bar{m}$, $\Delta_4 = \frac{1}{\sqrt{\pi}}\int_0^{\bar{\lambda}_0} \mathrm{e}^{-(m-\mu_2)^2}\,\mathrm{d}\bar{m}$, $K = \frac{1}{(x-1)!\Phi_{t_w}}$, $K_1 = \frac{N^x \mathrm{e}^{-N\bar{\lambda}_{0t_w}+N^2\bar{\sigma}_p^4}\sqrt{\pi}}{12(x-1)!\Phi_{t_w}}$, $K_2 = \frac{N^x \mathrm{e}^{-N\bar{\lambda}_{0t_w}+\frac{9N^2\bar{\sigma}_p^4}{16}}\sqrt{\pi}}{4(x-1)!\Phi_{t_w}}$, $\mu_1 = -(\bar{\sigma}_p^2 N - \bar{\lambda}_{0t_w})$, $\mu_2 = -\left(\frac{3\bar{\sigma}_p^2 N}{4} - \bar{\lambda}_{0t_w}\right)$, and $\Psi(\cdot)$ is the moment of a truncated Gaussian variable as defined in Appendix 5.B.

Proof. See Appendix 5.A. ■

5.4.2.3 Extension to Two Types of Traffic

Previously, we assumed in the derivation of the Bayesian robust threshold that there was only one type of traffic or, what it is equivalent, that there were two types of traffic and that the load corresponding to one of them was known perfectly without uncertainty.

In case that the traffic loads for both kinds of traffic are known imperfectly, we should define a robust Bayesian threshold frontier by means of extending the previous integrals to include simultaneously the uncertainty model for both traffics. Such integrals are, however, extremely complicated and should be solved resorting to complex numerical methods.

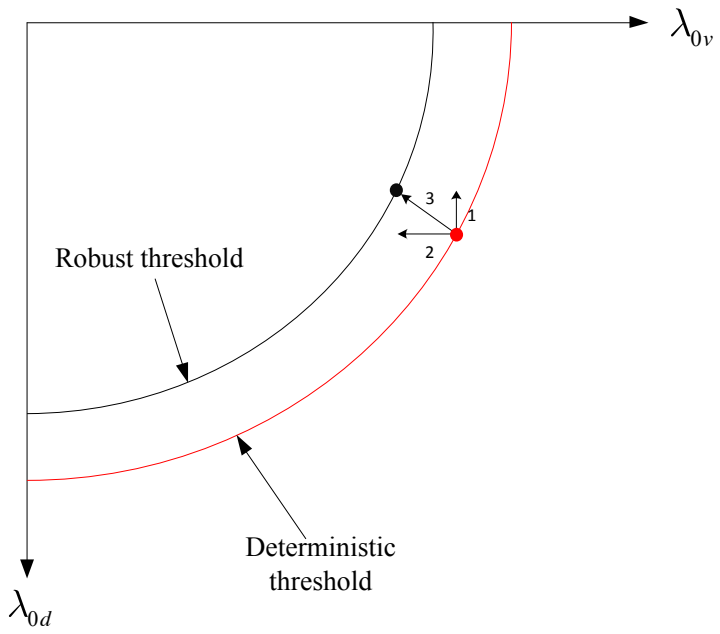


Figure 5.9: Graphic representation of the threshold computation for the robust case with two types of traffic.

A simpler methodology to derive a suboptimum robust threshold frontier could be based on the following brief idea:

1. Assuming that the data traffic load is known perfectly, calculate the robust threshold for the voice traffic for each possible value of data traffic load using the methodology presented previously for one kind of traffic.

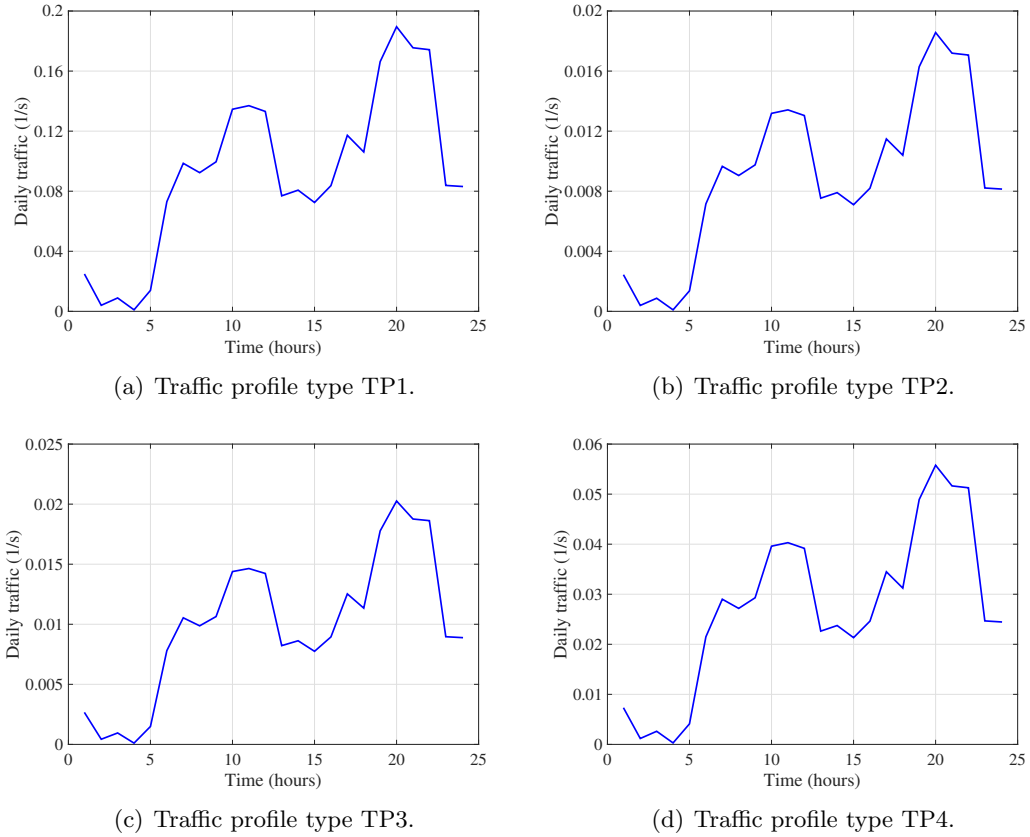
2. Assuming that the voice traffic load is known perfectly, calculate the robust threshold for the data traffic for each possible value of voice traffic load using the methodology presented previously for one kind of traffic.

3. Combine both thresholds shifts as a resultant vector movement in a two dimensional plane.

The previous methodology is described for clarity in Figure 5.9. In the figure, we plot the deterministic threshold for two types of traffic and the new robust threshold. Note that, in this example, the robust threshold is more conservative than the deterministic threshold but, in fact, this will depend on the value of outage probability that we set.

5.5 Numerical Simulations

In this section, we provide simulation results for the dimensioning of the energy units and the results of the on/off switching strategy. The considered batteries have the following specifications: 12 V, 100 Ah, and a capacity $C_1 = 1,200$ Ah $\times$ V. The solar panels have a nominal power



(a) Traffic profile type TP1.

(b) Traffic profile type TP2.

(c) Traffic profile type TP3.

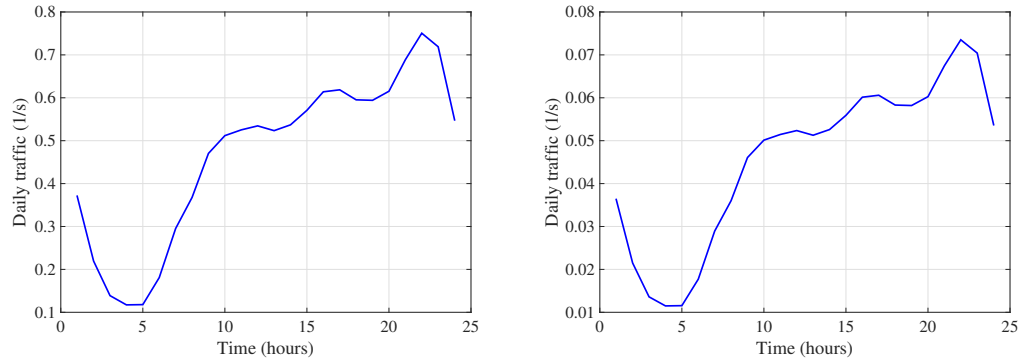
(d) Traffic profile type TP4.

Figure 5.10: Daily traffic profile of the 4 different towns considered for voice services.

$P_{\text{nom}} = 85$ W. We consider that $N_{\text{da}} = 3$. In this section, we have three different types of BSs with different maximum radiated powers $P_{\text{max}}^{\text{BS}}$ given by $\text{BS}_a = 20$ dBm, $\text{BS}_b = 13$ dBm, and $\text{BS}_c = 24$ dBm. We assume that we have 4 different traffic profiles denoted as TP1, TP2, TP3, and TP4 which correspond to 4 different predicted traffic demands of real towns. The plots of the traffic evolution of the four different profiles are shown in Figure 5.10 and Figure 5.11 for the voice and data traffic daily evolution, respectively. As can be seen the shape of the traffics among the towns are the same but the magnitudes are different as they depend on the population. We also assume that the dimensioning of the energy units must be performed considering the traffic estimation forecast for the incoming 5 years (voice and data traffic intensities increase by 180% in the second year, by 4% in the third year, and by 2% in the fourth and fifth year after the deployment). For particular details, see [tuc13]. The parameters of the power consumption model [Aue11] are set to $N_{TRX} = 1$, $P_0 = 4.8$ W, $\Delta_p = 8$, and $P_{\text{sleep}} = 2.9$ W.

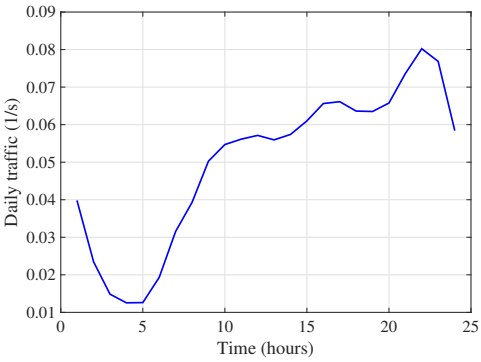
Figure 5.12 depicts the traffic profile considering only voice traffic for a specific location and the corresponding threshold for switching on/off one of the two available BSs. The type of BS considered is BS_a. As we can see, the two BSs are needed only a few hours during the day (30% of the total time). For the rest of the hours, only one BS is enough. Thus, potential energy savings can be obtained.

Figure 5.13 depicts the reduction of the solar panel size in percentage terms compared to

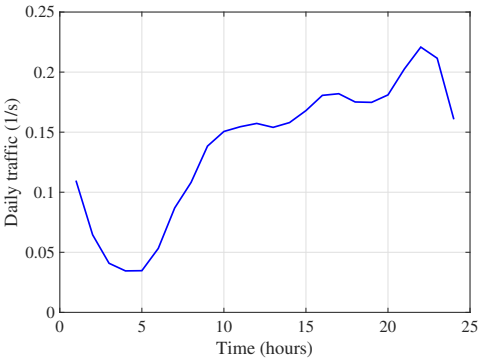


(a) Traffic profile type TP1.

(b) Traffic profile type TP2.



(c) Traffic profile type TP3.



(d) Traffic profile type TP4.

Figure 5.11: Daily traffic profile of the 4 different towns considered for data services.

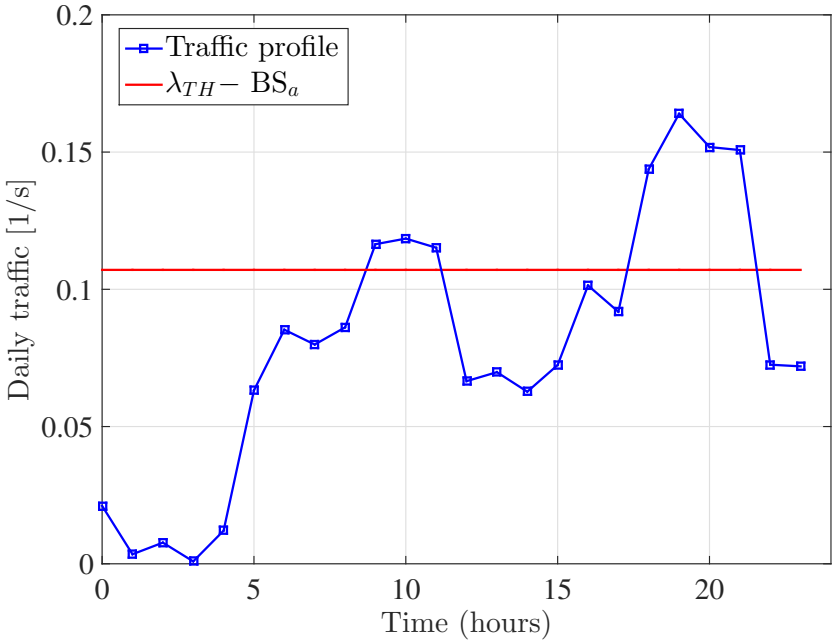


Figure 5.12: On/off switching threshold example for a single traffic profile during the first year with traffic type TP1.

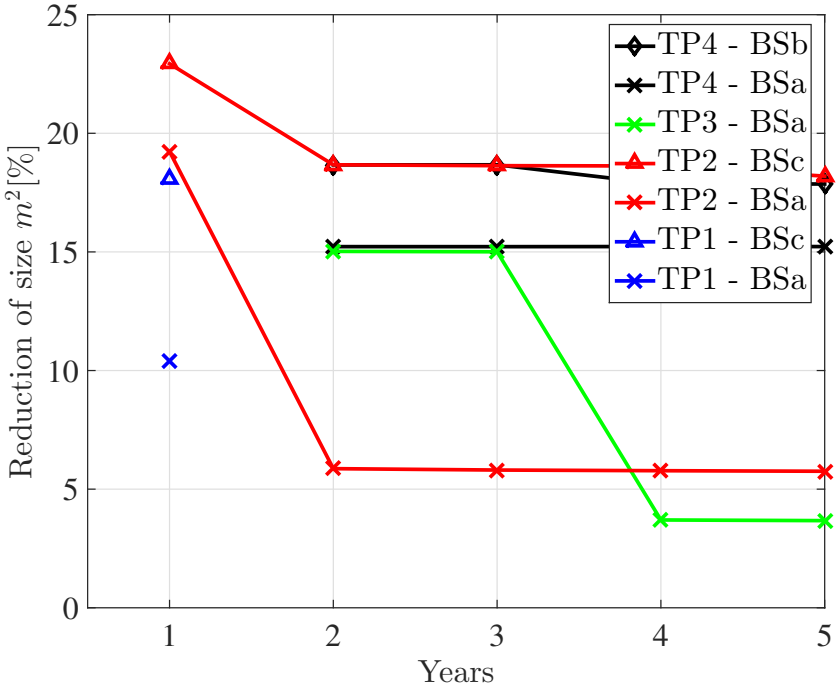


Figure 5.13: Solar panel size reduction for different traffic profiles and different types of BS within the first 5 years.

the case where two BSs are considered to be always active if only voice traffic is considered. The results are represented as a function of the estimated traffic evolution in 5 years for different traffic profiles and different types of BS. Battery size is linearly proportional to the solar panel size and, thus, the experienced reduction is the same in both cases. As it can be observed, for some traffic profiles the amount of reduction is around 15-20% for the first years which directly translates into a CAPEX reduction.

Figure 5.14 presents the reduction of the consumed power in percentage terms compared to the case where two BSs are considered to be always active if only voice traffic is considered. The results are represented as a function of the estimated traffic evolution in 5 years for different traffic profiles and different types of BS. As it can be observed, for some traffic profiles the amount of reduction is around 12-16%. This directly impacts the reduction of the solar panel sizes and the battery sizes as we previously showed in Figure 5.13.

Figure 5.15 depicts the threshold region that determines the tuples of voice and data traffic rates that are required to switch off one of the BS. The corresponding hours of the day in which a specific BS should be switched on or off is shown in Figure 5.16. For the specific traffic profile TP1 and using BS type BSc, the two BSs are needed only 10 hours per day during the first year.

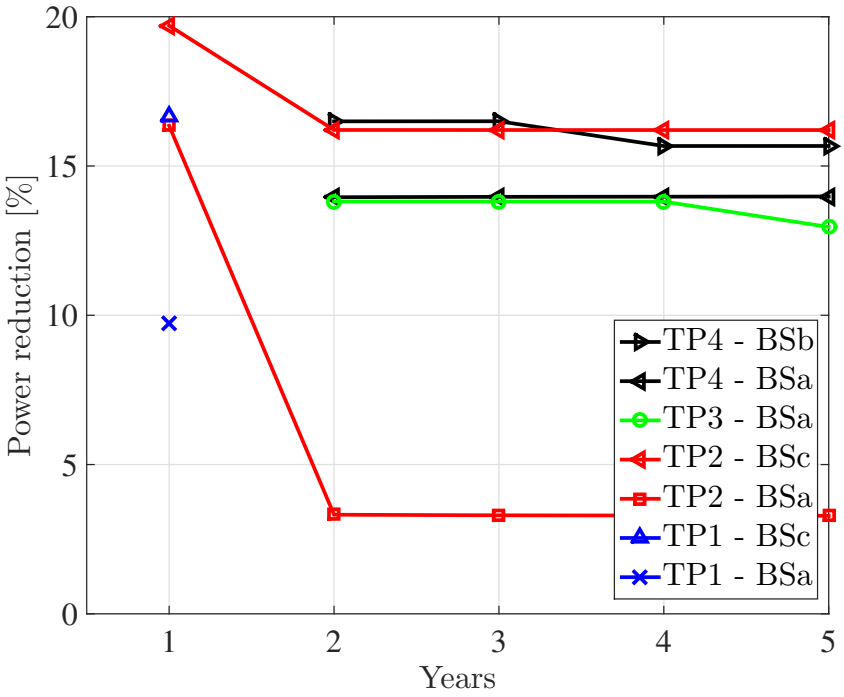


Figure 5.14: Power reduction for different traffic profiles and different types of BS within the first 5 years.

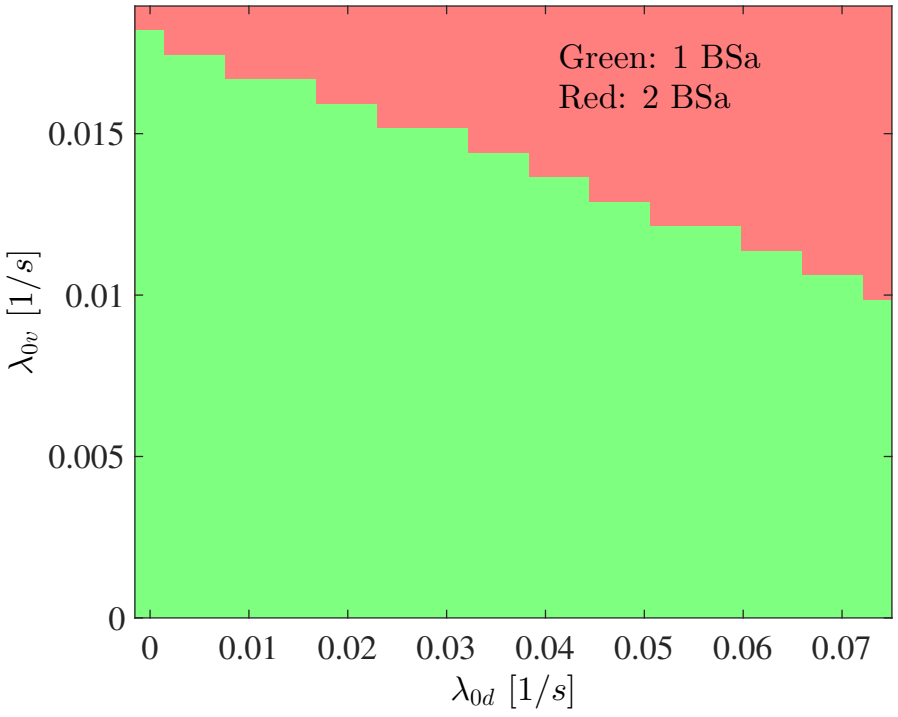


Figure 5.15: Region of (λ-threshold) for mixed traffic for a specifc traffic profile for the first year.

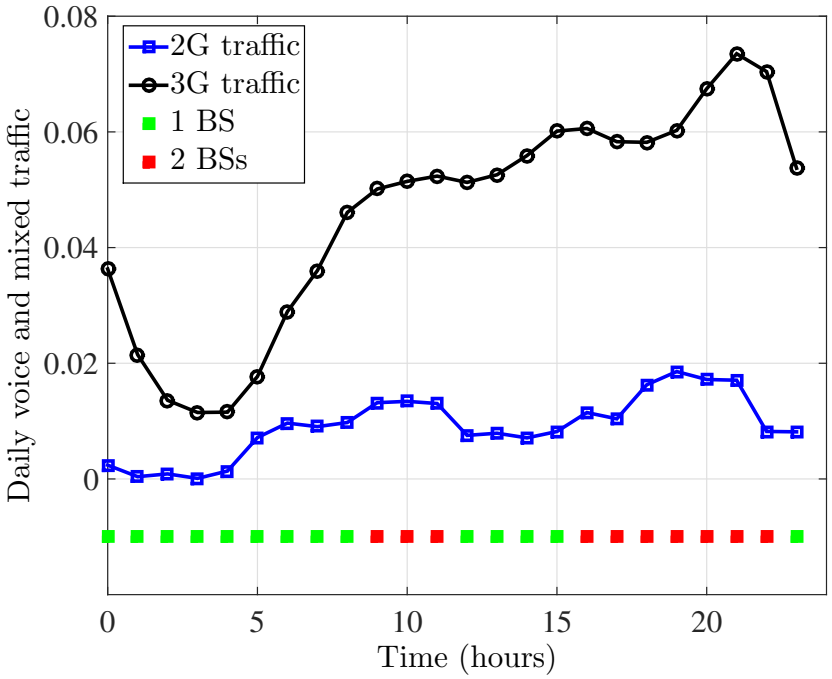


Figure 5.16: Hours of the day needing 1 or 2 active BSs for a specific traffic profile for the first year.

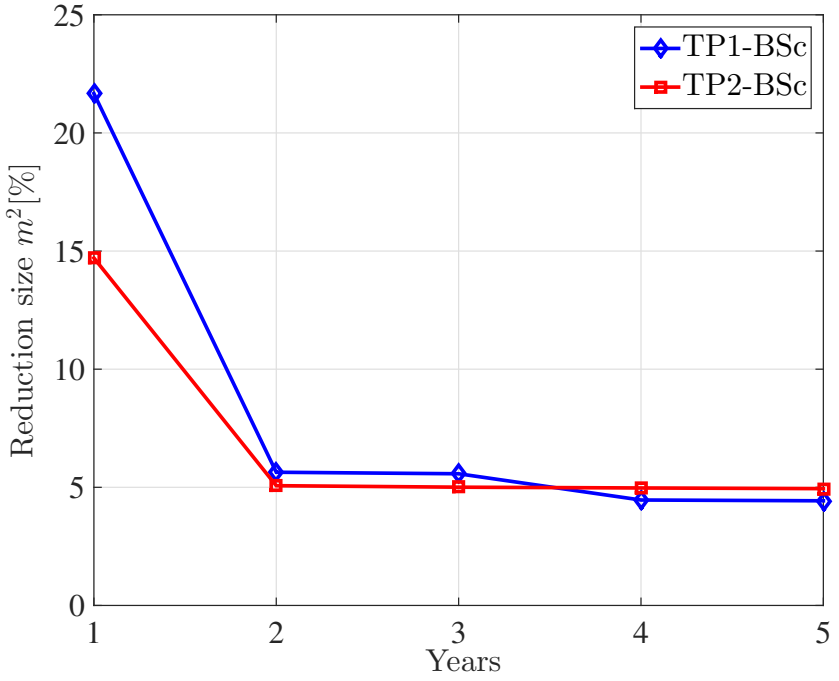


Figure 5.17: Solar panel size reduction for mixed traffic and two different traffic profiles.

Figure 5.17 depicts the reduction of the solar panel size in percentage if on/off strategy is implemented compared to the case where the two BSs are considered to be always active for mixed traffic demands. As it can be seen, as the traffic increases substantially during the second year, the configuration with two BSs on is required more hours during the day and, thus, only 5% of savings is possible. Of course, the savings are influenced by the forecast of the traffic for

the next years. In areas where the traffic is considered to be more or less static, then, we would see larger savings for all the years.

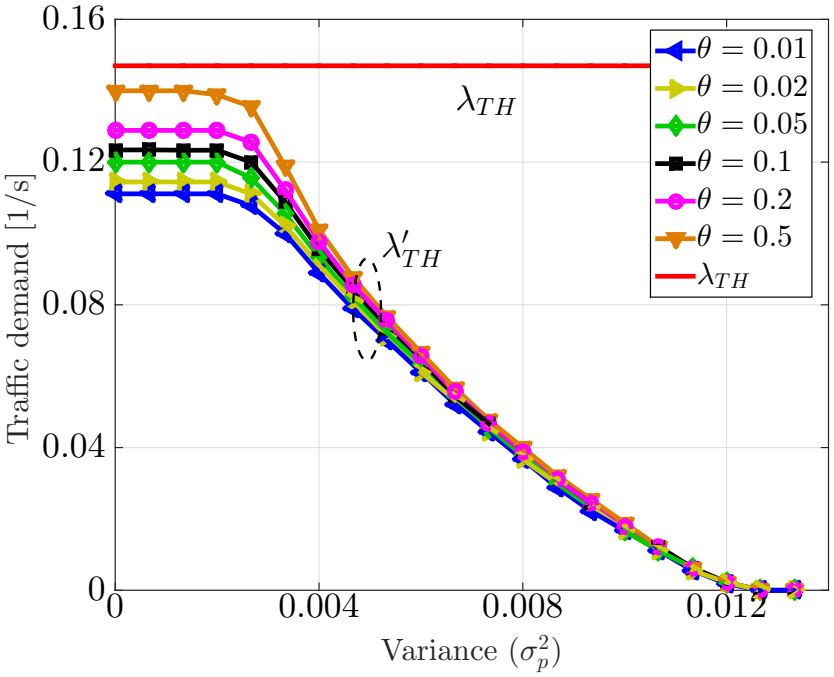


Figure 5.18: Bayesian threshold as a function of the variance σ_p^2 for different outages.

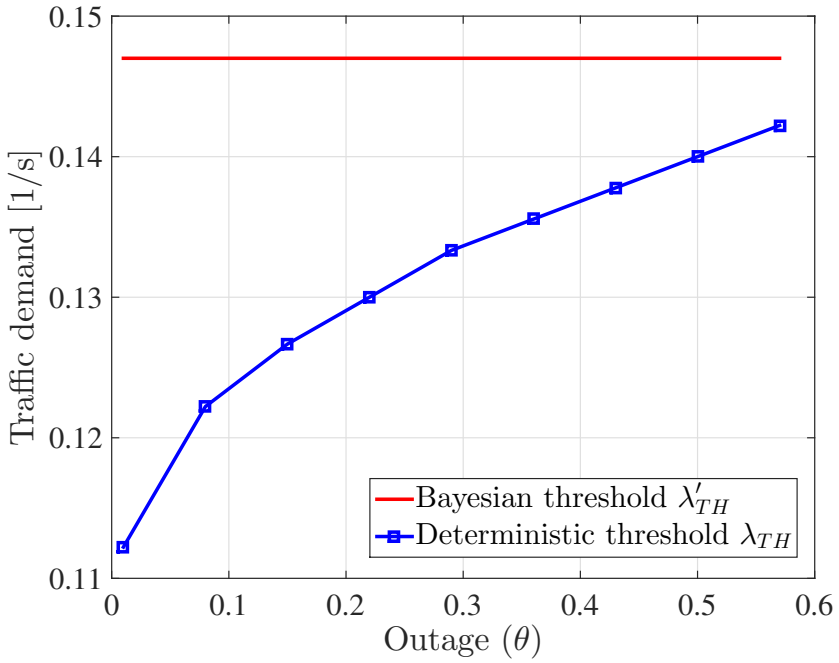


Figure 5.19: Bayesian threshold as a function of the outage probability for $\sigma_p^2 = 0$.

Figure 5.18 shows the threshold computed using the Bayesian methodology for a given variance of $p[n]$ and for different outage probabilities. We also show the deterministic threshold (when the traffic profile is perfectly known). We can see that, if we allow a higher outage

probability, then the threshold increases. For the particular case of $\theta = 0$, the threshold should be set to zero, i.e., $\lambda'_{TH} = 0$. Another effect that we can see and also expect is that for larger variances, we cannot trust the estimation and, thus, the threshold must be reduced to guarantee the same outage. We also see that for large variances, the outage probability does not have an important impact on the threshold calculation.

Finally, Figure 5.19 depicts the threshold computation for different values of outage and for variance $\sigma_p^2 = 0$ (deterministic threshold). This would correspond to the ideal case where we know perfectly the model. One thing that we could see is that, for $\theta = 0.5$ the Bayesian threshold does not yield the deterministic threshold as expected. This is because, in the Bayesian approach, the traffic profile is estimated with a finite window time. This estimation incurs errors that makes the computation of the Bayesian threshold be more conservative than the deterministic one.

5.6 Chapter Summary and Conclusions

In this chapter we have presented a methodology for dimensioning the energy units, e.g., solar panels and batteries for powering BSs, with special impact in rural scenarios where the access to the electric grid is impossible or very expensive. The scenario under consideration for these rural areas was based on two BSs placed at the same site, with fully overlapped coverage areas and using two different frequencies. Because the daily traffic profile is not constant, we have also provided a methodology for switching on and off one of the BSs in order to reduce the energy consumption and, thus, deploy smaller solar panels and fewer number of batteries. We proposed a decision strategy where we had perfect knowledge of the traffic profile and a robust Bayesian strategy in order to account for possible error modeling in the traffic profile information.

Simulations were performed with real data for a real network deployment that is being designed for isolated rural areas in the Amazon forest of Perú. Results showed that proposed solution can be a sustainable and economical solution to provide cellular services in outdoor isolated rural scenarios.

5.A Proof of Proposition 5.1

Let us start with the computation of the probability, $\mathbb{P}\left(\sum_{n=1}^{N} k_{t_w n} \leq \bar{\lambda}'_{TH} N \,\middle|\, \bar{\lambda}_{t_w} = \bar{\lambda}_{0t_w} + \bar{p}_{t_w}\right)$. Recall that $\sum_{n=1}^{N} k_{t_w n}$ is Poisson distributed with expected value $(\bar{\lambda}_{0t_w} + \bar{p}_{t_w})N$. Then, the CDF of such Poisson random variable can be defined as [Hai67]

$$\mathbb{P}\left(\sum_{n=1}^{N} k_{t_w n} \leq \bar{\lambda}'_{TH} N \,\middle|\, \bar{\lambda}_{t_w}\right) = \frac{\Gamma\left(\lceil \bar{\lambda}'_{TH} N \rceil + 1, (\bar{\lambda}_{0t_w} + \bar{p}_{t_w})N\right)}{\lceil \bar{\lambda}'_{TH} N \rceil !}, \tag{5.19}$$

where $\lceil \cdot \rceil$ is the ceiling operator and $\Gamma(n, x) = \int_x^\infty s^{n-1} \mathrm{e}^{-s}\, \mathrm{d}s$ is the upper incomplete Gamma function [Abr72]. Let us drop the time index throughout the development for ease of notation. Let us define $x \triangleq \lceil \bar{\lambda}'_{TH} N \rceil + 1$. Thus, the previous expression can be expressed as

$$\mathbb{P}\left(\sum_{n=1}^{N} k_{tn} \leq \bar{\lambda}'_{TH} N \,|\, \bar{\lambda}_0 + \bar{p}\right) = \frac{1}{(x-1)!}\Gamma\left(x, (\bar{\lambda}_0 + \bar{p})N\right) \tag{5.20}$$

If we take the expectation of the right hand side of (5.20) w.r.t. the unknown parameter $\bar{p}$, we obtain:

$$\mathbb{E}_{\bar{p}}[\,\dots\,] = \frac{1}{(x-1)!}\int_{-\infty}^{\infty} \Gamma\left(x, (\bar{\lambda}_0 + \bar{p})N\right) f\left(\bar{p}, 0, \bar{\sigma}_p^2, -\bar{\lambda}_0, \infty\right)\, \mathrm{d}\bar{p} \tag{5.21}$$

$$= \frac{1}{(x-1)!\Phi}\int_{-\bar{\lambda}_0}^{\infty} \Gamma\left(x, (\bar{\lambda}_0 + \bar{p})N\right) \frac{1}{\sqrt{2\pi}\bar{\sigma}_p} \mathrm{e}^{-\frac{\bar{p}^2}{2\bar{\sigma}_p^2}}\, \mathrm{d}\bar{p} \tag{5.22}$$

$$= \frac{1}{(x-1)!\Phi\sqrt{2\pi}\bar{\sigma}_p}\int_{-\bar{\lambda}_0}^{\infty}\int_{(\bar{\lambda}_0+\bar{p})N}^{\infty} s^{x-1}\mathrm{e}^{-s}\mathrm{e}^{-\frac{\bar{p}^2}{2\bar{\sigma}_p^2}}\, \mathrm{d}s\, \mathrm{d}\bar{p}. \tag{5.23}$$

Now, if we integrate first w.r.t. $\bar{p}$:

$$\mathbb{E}_{\bar{p}}[\,\dots\,] = \frac{1}{(x-1)!\Phi\sqrt{2\pi}\bar{\sigma}_p}\int_0^{\infty}\int_{-\bar{\lambda}_0}^{\frac{s}{N}-\bar{\lambda}_0} s^{x-1}\mathrm{e}^{-s}\mathrm{e}^{-\frac{\bar{p}^2}{2\bar{\sigma}_p^2}}\, \mathrm{d}\bar{p}\, \mathrm{d}s \tag{5.24}$$

$$= \frac{1}{(x-1)!\Phi}\int_0^{\infty} s^{x-1}\mathrm{e}^{-s}\left(\int_{-\bar{\lambda}_0}^{\frac{s}{N}-\bar{\lambda}_0} \frac{1}{\sqrt{2\pi}\bar{\sigma}_p}\mathrm{e}^{-\frac{\bar{p}^2}{2\sigma_p^2 T^2}}\, \mathrm{d}\bar{p}\right) \mathrm{d}s \tag{5.25}$$

$$= \frac{1}{(x-1)!\Phi}\int_0^{\infty} s^{x-1}\mathrm{e}^{-s}\left[\mathcal{Q}\left(\frac{-\bar{\lambda}_0}{\bar{\sigma}_p}\right) - \mathcal{Q}\left(\frac{\frac{s}{N}-\bar{\lambda}_0}{\bar{\sigma}_p}\right)\right] \mathrm{d}s \tag{5.26}$$

$$= \frac{1}{(x-1)!\Phi}\mathcal{Q}\left(\frac{-\bar{\lambda}_0}{\bar{\sigma}_p}\right)\Gamma(x) - \underbrace{\frac{1}{(x-1)!\Phi}\int_0^{\infty} s^{x-1}\mathrm{e}^{-s}\mathcal{Q}\left(\frac{\frac{s}{N}-\bar{\lambda}_0}{\bar{\sigma}_p}\right) \mathrm{d}s}_{\mathcal{I}}, \tag{5.27}$$

where $\mathcal{Q}(\cdot)$ is the Gaussian $\mathcal{Q}$-function defined as $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \mathrm{e}^{-\frac{u^2}{2}}\, \mathrm{d}u$ [Chi03] and $\Gamma(x) = \int_0^\infty s^{x-1}\mathrm{e}^{-s}\, \mathrm{d}s$ is the Gamma function [Abr72].

In order to be able to compute the right hand side of (5.24), we will apply the following tight bound on the function $\mathcal{Q}(x)$ [Chi03]:

$$\mathcal{Q}(x) \approx \frac{1}{12}\mathrm{e}^{-\frac{x^2}{2}} + \frac{1}{4}\mathrm{e}^{-\frac{2x^2}{3}}, \quad x > 0. \tag{5.28}$$

Before integrating $\mathcal{I}$, we perform the following change of variables: $t = \frac{\frac{S}{N}-\bar{\lambda}_0}{\bar{\sigma}_p} \longrightarrow s = \bar{\sigma}_p N t + N\bar{\lambda}_0$. Thus, $\mathrm{d}s = \bar{\sigma}_p N\,\mathrm{d}t$, and the new limits of integration are $\left\{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}, \infty\right\}$. Hence,

$$\mathcal{I} = \frac{1}{(x-1)!\Phi}\int_0^\infty s^{x-1}\mathrm{e}^{-s}\mathcal{Q}\left(\frac{\frac{s}{N}-\bar{\lambda}_0}{\bar{\sigma}_p}\right)\mathrm{d}s \tag{5.29}$$

$$= \frac{\bar{\sigma}_p N}{(x-1)!\Phi}\int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^\infty (N(\bar{\sigma}_p t+\bar{\lambda}_0)^{x-1}\mathrm{e}^{-\bar{\sigma}_p N t}\mathrm{e}^{-N\bar{\lambda}_0}\mathcal{Q}\left(t\right)\,\mathrm{d}t \tag{5.30}$$

$$= \frac{\bar{\sigma}_p N^x \mathrm{e}^{-N\bar{\lambda}_0}}{(x-1)!\Phi}\int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^\infty (\bar{\sigma}_p t+\bar{\lambda}_0)^{x-1}\mathrm{e}^{-\bar{\sigma}_p N t}\mathcal{Q}\left(t\right)\,\mathrm{d}t. \tag{5.31}$$

Let us define $\bar{K} \triangleq \frac{\bar{\sigma}_p N^x \mathrm{e}^{-N\bar{\lambda}_0}}{(x-1)!\Phi}$. Note that the limits of integration takes negatives values. However, the bound of the function $\mathcal{Q}(x)$ only accepts positive arguments (see (5.28)). But, as $\mathcal{Q}(-x) = 1 - \mathcal{Q}(|x|)$, then,

$$\mathcal{I} = \bar{K}\int_0^\infty (\bar{\sigma}_p t+\bar{\lambda}_0)^{x-1}\mathrm{e}^{-\bar{\sigma}_p N t}\mathcal{Q}(t)\,\mathrm{d}t + \bar{K}\int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^0 (\bar{\sigma}_p t+\bar{\lambda}_0)^{x-1}\mathrm{e}^{-\bar{\sigma}_p N t}\mathcal{Q}\left(t\right)\,\mathrm{d}t \tag{5.32}$$

$$= \underbrace{\bar{K}\int_0^\infty (\bar{\sigma}_p t+\bar{\lambda}_0)^{x-1}\mathrm{e}^{-\bar{\sigma}_p N t}\mathcal{Q}(t)\,\mathrm{d}t}_{\mathcal{L}}$$

$$+ \underbrace{\bar{K}\int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^0 (\bar{\sigma}_p t+\bar{\lambda}_0)^{x-1}\mathrm{e}^{-\bar{\sigma}_p N t}\left(1-\mathcal{Q}\left(|t|\right)\right)\,\mathrm{d}t}_{\mathcal{R}}, \tag{5.33}$$

and now we can apply the bound on both terms of the previous expression. The left hand side, $\mathcal{L}$, is approximated by

$$\mathcal{L} \approx \frac{\bar{K}}{12}\int_0^\infty (\bar{\sigma}_p t+\bar{\lambda}_0)^{x-1}\mathrm{e}^{-\bar{\sigma}_p N t}\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t + \frac{\bar{K}}{4}\int_0^\infty (\bar{\sigma}_p t+\bar{\lambda}_0)^{x-1}\mathrm{e}^{-\bar{\sigma}_p N t}\mathrm{e}^{-\frac{2t^2}{3}}\,\mathrm{d}t. \tag{5.34}$$

Now, performing the change of variable $m = \bar{\sigma}_p t+\bar{\lambda}_0, \quad t = \frac{m-\bar{\lambda}_0}{\bar{\sigma}_p}, \quad \mathrm{d}t = \frac{1}{\bar{\sigma}_p}\,\mathrm{d}m$, and completing the squares, we end up with

$$\mathcal{L} \approx \frac{\bar{K}\mathrm{e}^{\bar{\sigma}_p^4 N^2}\sqrt{\pi}}{12\bar{\sigma}_p}\int_{\bar{\lambda}_0}^\infty \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{2}}} m^{x-1}\mathrm{e}^{-(m-\mu_1)^2}\,\mathrm{d}m$$

$$+ \frac{\bar{K}\mathrm{e}^{\frac{9\bar{\sigma}_p^4 N^2}{16}}\sqrt{\pi}}{4\bar{\sigma}_p}\int_{\bar{\lambda}_0}^\infty \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{2}}} m^{x-1}\mathrm{e}^{-(m-\mu_2)^2}\,\mathrm{d}m, \tag{5.35}$$

where $\mu_1 = -(\bar{\sigma}_p^2 N - \bar{\lambda}_0)$ and $\mu_2 = -\left(\frac{3\bar{\sigma}_p^2 N}{4} - \bar{\lambda}_0\right)$. Note that both integrals are the $(x-1)$-

th moment of a truncated Gaussian random variable with mean μ_1 and μ_2, respectively, and variance $\frac{1}{\sqrt{2}}$.

Now, let us continue with the right hand side of $\mathcal{I}$, i.e, $\mathcal{R}$:

$$\mathcal{R} \approx \bar{K} \int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^{0} (\bar{\sigma}_p t + \bar{\lambda}_0)^{x-1} \mathrm{e}^{-\bar{\sigma}_p N t} \,\mathrm{d}t - \bar{K} \int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^{0} (\bar{\sigma}_p t + \bar{\lambda}_0)^{x-1} \mathrm{e}^{-\bar{\sigma}_p N t} \mathcal{Q}(|t|) \,\mathrm{d}t \tag{5.36}$$

$$= \bar{K} \int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^{0} (\bar{\sigma}_p t + \bar{\lambda}_0)^{x-1} \mathrm{e}^{-\bar{\sigma}_p N t} \,\mathrm{d}t - \frac{\bar{K}}{12} \int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^{0} (\bar{\sigma}_p t + \bar{\lambda}_0)^{x-1} \mathrm{e}^{-\bar{\sigma}_p N t} \mathrm{e}^{-\frac{t^2}{2}} \,\mathrm{d}t$$

$$- \frac{\bar{K}}{4} \int_{-\frac{\bar{\lambda}_0}{\bar{\sigma}_p}}^{0} (\bar{\sigma}_p t + \bar{\lambda}_0)^{x-1} \mathrm{e}^{-\bar{\sigma}_p N t} \mathrm{e}^{-\frac{2t^2}{3}} \,\mathrm{d}t. \tag{5.37}$$

Performing the change of variable on all integrals $m = \bar{\sigma}_p t + \bar{\lambda}_0$, $\mathrm{d}t = \frac{1}{\bar{\sigma}_p} \mathrm{d}m$, and completing the squares, and a second change of variable on the first integral, $mN = \tilde{m}$, $\mathrm{d}m = \frac{1}{N} \mathrm{d}\tilde{m}$, it yields

$$\mathcal{R} \approx \frac{\bar{K} \mathrm{e}^{\bar{\lambda}_0 N}}{\bar{\sigma}_p N^x} \gamma(x, N\bar{\lambda}_0) - \frac{\bar{K} \mathrm{e}^{N^2 \bar{\sigma}_p^4} \sqrt{\pi}}{12 \bar{\sigma}_p} \int_0^{\bar{\lambda}_0} \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{2}}} m^{x-1} \mathrm{e}^{-(m-\mu_1)^2} \,\mathrm{d}m$$

$$- \frac{\bar{K} \mathrm{e}^{\frac{9N^2 \bar{\sigma}_p^4}{16}} \sqrt{\pi}}{4 \bar{\sigma}_p} \int_0^{\bar{\lambda}_0} \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{2}}} m^{x-1} \mathrm{e}^{-(m-\mu_2)^2} \,\mathrm{d}m, \tag{5.38}$$

where μ_1 and μ_2 where previously defined and $\gamma(s, x) = \int_0^x t^{s-1} \mathrm{e}^{-t} \,\mathrm{d}t$ is the lower incomplete Gamma function [Abr72]. Let us introduce the following definitions:

$$K \triangleq \frac{1}{(x-1)! \Phi} \tag{5.39}$$

$$K_1 \triangleq \frac{N^x \mathrm{e}^{-N\bar{\lambda}_0 + N^2 \bar{\sigma}_p^4} \sqrt{\pi}}{12 (x-1)! \Phi} \tag{5.40}$$

$$K_2 \triangleq \frac{N^x \mathrm{e}^{-N\bar{\lambda}_0 + \frac{9N^2 \bar{\sigma}_p^4}{16}} \sqrt{\pi}}{4 (x-1)! \Phi} \tag{5.41}$$

$$\Delta_1 \triangleq \frac{1}{\sqrt{\pi}} \int_{\bar{\lambda}_0}^{\infty} \mathrm{e}^{-(m-\mu_1)^2} \,\mathrm{d}m \tag{5.42}$$

$$\Delta_2 \triangleq \frac{1}{\sqrt{\pi}} \int_{\bar{\lambda}_0}^{\infty} \mathrm{e}^{-(m-\mu_2)^2} \,\mathrm{d}\bar{m} \tag{5.43}$$

$$\Delta_3 \triangleq \frac{1}{\sqrt{\pi}} \int_{0}^{\bar{\lambda}_0} \mathrm{e}^{-(m-\mu_1)^2} \,\mathrm{d}\bar{m} \tag{5.44}$$

$$\Delta_4 \triangleq \frac{1}{\sqrt{\pi}} \int_{0}^{\bar{\lambda}_0} \mathrm{e}^{-(m-\mu_2)^2} \,\mathrm{d}\bar{m}. \tag{5.45}$$

By defining the m-th moment of a truncated Gaussian distribution as (see Appendix 5.B)

$$\Psi(p, \mu, \sigma, x, y, m) = \frac{1}{\frac{1}{\sqrt{2\pi}\sigma} \int_x^y \mathrm{e}^{-\frac{(p-\mu)^2}{2\sigma^2}} \,\mathrm{d}p} \frac{1}{\sqrt{2\pi}\sigma} \int_x^y p^m \mathrm{e}^{-\frac{(p-\mu)^2}{2\sigma^2}} \,\mathrm{d}p, \tag{5.46}$$

we, finally, obtain

$$\begin{aligned}\mathbb{E}_{\bar{p}}\left[\mathbb{P}\left(\sum_{n=1}^{N} k_{t_w n} \leq \bar{\lambda}'_{TH} N \,\middle|\, \bar{\lambda}_{0t} + \bar{p}_t\right)\right] \approx \\ + K\mathcal{Q}\left(\frac{-\bar{\lambda}_0}{\bar{\sigma}_p}\right)\Gamma(x) - K_1\Delta_1\,\Psi\left(\bar{p}, \mu_1, \frac{1}{\sqrt{2}}, \bar{\lambda}_0, \infty, x-1\right) \\ - K_2\Delta_2\,\Psi\left(\bar{p}, \mu_2, \frac{1}{\sqrt{2}}, \bar{\lambda}_0, \infty, x-1\right) - K\,\gamma(x, N\bar{\lambda}_0) \\ + K_1\Delta_3\,\Psi\left(\bar{p}, \mu_1, \frac{1}{\sqrt{2}}, 0, \bar{\lambda}_0, x-1\right) + K_2\Delta_4\,\Psi\left(\bar{p}, \mu_2, \frac{1}{\sqrt{2}}, 0, \bar{\lambda}_0, x-1\right).\end{aligned} \tag{5.47}$$

5.B Computation of Moments of Truncated Gaussian Random Variables

In this appendix, we provide the formal definition of the moment of a truncated Gaussian random variable, also known as the partial moment, and a procedure to easily compute such partial moments. For more details, the reader is referred to [Dhr05] for the particular case of Gaussian random variables, and [Bro13] for the general case of any distribution.

Let $F(u)$ and $f(u)$ denote the CDF and PDF of u. Let $u \sim \mathcal{N}(\mu, \sigma)$. The m-th partial moment of the random variable u is given by

$$\mathbb{E}\left[u^m \,|\, u \leq k\right] = \frac{1}{F(h)} \int_{-\infty}^{k} u^m f(u)\,\mathrm{d}u, \tag{5.48}$$

where $h = \frac{k-\mu}{\sigma}$. Let $\xi = \frac{u-\mu}{\sigma}$. Based on the previous definition, we are able to obtain a recursion for the calculation of the m-th partial moment as follows:

$$\mathbb{E}\left[u^m \,|\, u \leq k\right] = \sum_{r=0}^{m} \binom{m}{r} \mu^{m-r} \sigma^r I_r, \tag{5.49}$$

where

$$I_r = \frac{1}{F(h)} \int_{-\infty}^{h} \xi^r f(\xi)\,\mathrm{d}\xi = -h^{r-1}\frac{f(h)}{F(h)} + (r-1)I_{r-2}, \tag{5.50}$$

and the initial conditions are $I_0 = 1,\ I_1 = -\frac{f(h)}{F(h)}$.

Chapter 6

Stochastic Resource Allocation with Backhaul and Energy Constraints

6.1 Introduction

DL and UL radio resource allocation strategies for a system with limited backhaul capacity are considered. In the proposed system, the BS is equipped with a finite battery recharged by an energy harvester. Although backhaul availability has been taken for granted in conventional systems, backhaul is, in general, a limited resource. This is the case of the deployment planned in the European project TUCAN3G (already presented in the previous chapter, http://www.ict-tucan3g.eu). This project studies, from both the technological and socio-economical perspectives, the progressive introduction of mobile telephony and data services in isolated rural areas in developing countries. In particular, remote locations in Perú are considered. In such locations, three main challenges arise: backhaul capacity, cost of BSs, and business models adapted to people with low incomes. The solution adopted in TUCAN3G consists of an access network based on 3G femtocells (and its evolution to 4G) empowered by solar panels of limited size in outdoor scenarios, as well as WiFi-LD (WiFi for Long Distances) - WiMAX - VSAT heterogeneous backhauling.[1]

The resource allocation strategy considered in this chapter is developed with this scenario in mind, for which the limited capacity of the backhaul may have a huge impact on performance. The proposed strategy is described for 3G femtocells (based on WCDMA) as it is the solution initially considered in TUCAN3G (since it allows the use of cheap BSs and terminals). Nevertheless, the concept and methodology proposed in the chapter can be extended to 4G femtocells (based on LTE), as will be described below in the problem formulation.

In addition to the backhaul limitation, the energy available at the BS may be a very limited

[1]The WiFi-LD network is already deployed and is currently in use to provide connectivity to health centers in remote areas of Perú. It will also be used as a backhaul to provide 3G connectivity (voice and data) to the general population in the area, once the access network is deployed. Such a solution meets the low energy consumption and low maintenance/installation cost constraints required in the project while allowing for an easy and progressive network upgrade as traffic demand increases.

resource as well. If the BS is only powered with batteries (as may happen in rural environments, for instance in the TUCAN3G deployments described above), then the battery status as well as the harvesting capabilities (if any) should also be explicitly considered in the scheduling strategy if we want to optimize the performance subject to the energy limitations. In the scenario that we will consider in this chapter, the BS will be powered only by a limited battery and an energy harvesting device, e.g., solar panels that will recharge the batteries [Par05].

6.1.1 Related Work

With the advent of heterogeneous networks consisting of large and small cells, backhaul capacity limitations have been considered in the recent literature. For example, in [Cho11], the authors developed a strategy to design the precoder and the power allocation in a DL scenario considering limited backhaul capacity. In [Zho13a], the authors proposed a simple scheme that performs Wyner-Ziv compress-and-forward relaying on a per-BS basis in an UL multi-cell scenario where the BSs are connected to a centralized processor via rate-limited backhaul links. In [Mar11a], a strategy is developed to efficiently manage the backhaul capacity among a group of picocells. Specifically, a backhaul scheduling approach is proposed based on traffic demands along with an underlying optimum PHY layer transmission scheme that maximizes the picocell utility. Sum-rate optimization with limited backhaul capacity in a network-MIMO setup and in a coordinated multipoint (CoMP) setup was considered in [Sol11] and [Yu14]. A joint beamforming and clustering strategy was presented in [Dai14]. The scenario of that work is a DL network-MMO scenario, where BSs are connected to a central processor with rate-limited backhaul links. A heuristic scheme that jointly optimizes user scheduling and power control was proposed in [Xu13] where cooperation among BSs via capacity-limited backhaul links was considered. Finally, a joint user association and resource allocation strategy in a multi-cell heterogeneous network was presented in [Cui14], where each BS was provided with a limited backhaul capacity link.

In the previous works, the backhaul capacity limitation is introduced by imposing a maximum instantaneous aggregate traffic constraint. However, limiting the sum-rate instantaneously at each specific frame to match the instantaneous backhaul rate may hamper the performance of the system in terms of the achievable long-term rates. In these circumstances, it seems less limiting to use high data rates in the access network whenever the channel conditions allow (possibly using greater instantaneous values than the average constraint imposed by the backhaul) provided that the average backhaul rate constraint is met when averaging the traffic served. That is the strategy that we follow in this dissertation and it is the main difference w.r.t. the works presented before. Note that the backhaul constraint in terms of average traffic is suitable if we assume that queues are implemented at the entrance of the access network. Note also that this relaxed constraint will increase latency on the backhaul, which we do not consider in our analysis.

6.1.2 Main Contribution

In this chapter, we propose some resource allocation strategies for the single-cell scenario, in which there exists a backhaul connection that constrains the data rate at the access network. The backhaul constraint is assumed to be long-term and, thus, ergodic optimization techniques are employed to tackle the problem. Online resource allocation strategies for the DL as well as for the UL have been developed. The main contributions are:

- Proposal of a fair online scheduling algorithm considering a long-term backhaul constraint, the battery status of the BS, and the energy that is being harvested.
- An online strategy based on ergodic optimization (also known as stochastic approximation) is developed.
- Two different types of users, voice users and data users, are assumed to coexist in the network. Each type of users demands a different QoS.
- Resource allocation strategies are proposed for both, the DL and the UL scenarios.

6.1.3 Organization of the Chapter

The remainder of the chapter is organized as follows. In Section 6.2, we describe the resource allocation strategy for the DL scenario and present some numerical results. In Section 6.3, we develop the resource allocation procedure for the UL connections and show a numerical evaluation of the proposed strategy. Finally, Section 6.4 presents some final thoughts and concluding remarks.

6.2 Stochastic Resource Allocation for the Downlink Scenario

In this section, we develop a resource allocation strategy for the DL setup where we consider a backhaul capacity constraint and the BS to be powered with a finite battery and an energy harvesting source that recharges it. The resource allocation will be modeled as an optimization problem and will be based on assigning powers and codes to voice users and data users.

6.2.1 System Model

6.2.1.1 System Description

Let us consider a DL scenario composed of a single BS and several users. Because we focus on providing 3G connectivity, the system is based on WCDMA technology and that two different

types of users coexist: voice users and data users. The set of voice and data users are defined by $\mathcal{U}_V$ and $\mathcal{U}_D$, respectively, and it is assumed that voice users request a fixed service rate whereas data users request a flexible service rate.

Users in WCDMA are multiplexed using codes [Gol05]. We assume that the network operator has already reserved a set of codes for the voice users and that the remaining codes are to be allocated among the data users. Thus, the amount of available codes in each set is known and fixed at the BS.

The BS is powered only with a battery and an energy harvester. The energy harvester allows the BS to collect energy from the environment and recharge the battery (for example, solar panels). This is especially important in rural areas, where the access to the power grid may be impossible or too expensive. We consider that only causal information is available for the resource allocation strategy, i.e., only information of the past and current harvesting collections and battery dynamics will be available to execute the scheduling strategy at each particular frame, yielding to an *online approach*.

One of the novelties of this work is that we account for a maximum backhaul rate constraint. However, instead of limiting the instantaneous access network data rates as the maximum flow allowed by the backhaul, as in [Cho11], [Zho13a], and [Cui14], we limit the average throughput served by the access network. That means that we allow the instantaneous rate in the access wireless links to surpass the backhaul limitation at certain time instants. This can be done whenever we have queues at the entrance of the access network and such queues are stable (which, in fact, is guaranteed by imposing that the average aggregated rate is not higher than the backhaul capacity). As we want to incorporate fairness, in the resource allocation problem the long-term backhaul capacity is equally divided among the users with the same type of service. Accordingly, we consider in the access network resource allocation problem that the backhaul capacity is equally divided among the users with the same type of service. Figure 6.1 presents the system architecture of the target rural scenario.

6.2.1.2 Power Consumption Model and Battery Dynamics

In this subsection, we introduce the power consumption and battery models considered in this chapter. The overall power consumption at the BS is modeled as the addition of the radiated power, which is divided into the power devoted to pilot channels (P_{CPICH}, assumed to be fixed) and the power consumed by the traffic channels ($P_{\text{BS}}(t)$), and a fixed power consumed by the electronics of the BS (P_{c}), where t denotes the frame. The model considered for the last term is based on [Aue11] and includes the power consumption of the RF chains, the baseband power consumption, and the consumption of the cooling systems. The maximum traffic power will depend on the current battery level of the BS, as will be described in more detail later.

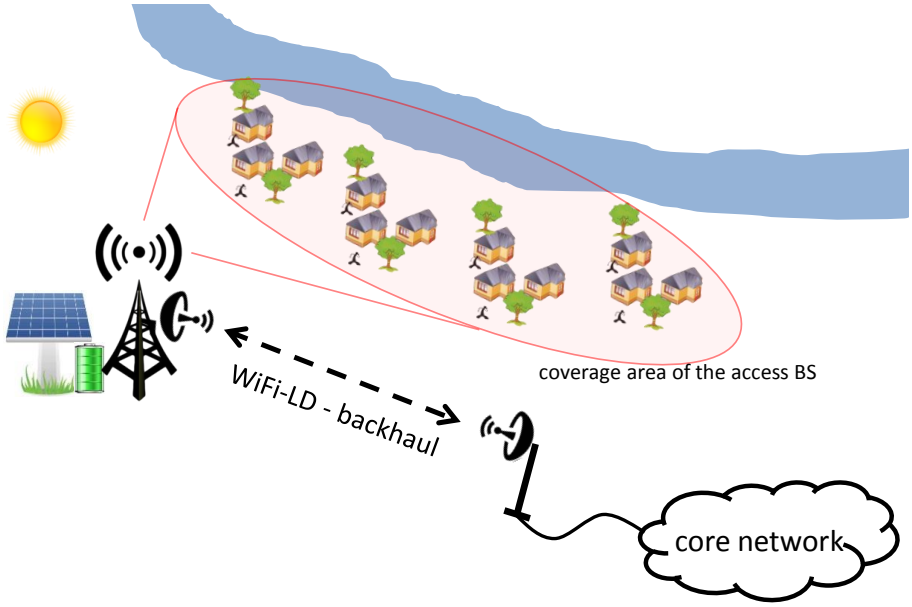


Figure 6.1: Architecture of the target rural scenario under consideration in the chapter. The BS is powered with a solar panel and a battery and the backhaul considered is based on WiFi-LD. The specific details of the real deployment as well as the location will be explained in the simulation section.

The overall energy consumption by the BS during the t-th frame is

$$E(t) \triangleq T_f \cdot \left(P_{\text{CPICH}} + P_{\text{BS}}(t) + P_{\text{c}}\right), \quad \forall t, \tag{6.1}$$

where T_f is the duration of the frame. Due to physical constraints of the amplifiers of the BS, the amount of power available for traffic services is limited, and it is denoted as $P_{\text{BS}}^{\max}$, so $P_{\text{BS}}(t) \leq P_{\text{BS}}^{\max}$.

Let $C(t)$ be the energy stored at the battery of the BS at the beginning of the frame[2] t. Then at period $t+1$, the battery level is updated in general as

$$C(t+1) = f(C(t), E(t), H(t)), \quad \forall t, \tag{6.2}$$

where $H(t)$ is the energy harvested in Joules during the frame t and the function $f(\cdot) : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ depends upon the battery dynamics, such as storage efficiency and memory effects. A common practice is to consider the following battery update:

$$C(t+1) = \left(C(t) - E(t) + H(t)\right)_0^{C_{\max}}, \quad \forall t, \tag{6.3}$$

where the projection of x onto the interval $[a, b]$ accounts for possible battery overflows and assures that the battery levels are non-negative, and $C_{\max}$ is the battery capacity. Notice that the whole harvesting collected during period t is assumed to be available in the battery at the end of the period for simplicity. In general, the total energy consumed by the BS during one

[2]Throughout the chapter, we will refer to the time instants as frames. Note that a particular frame corresponds to a scheduling period in which the resources are allocated.

period will be limited by a function of the current battery level as

$$T_f \cdot (P_{\text{CPICH}} + P_{\text{BS}}(t) + P_{\text{c}}) \leq g(C(t)), \quad \forall t, \tag{6.4}$$

where the function $g(\cdot)$ is defined as $g(C(t)) \triangleq \min\{T_f \left(P_{\text{CPICH}} + P_{\text{BS}}^{\max} + P_{\text{c}}\right), w(C(t))\}$, and $w(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ a generic continuous increasing function that satisfies $w(C(t)) \leq C(t)$, $\forall t$. For example, if all the battery is allowed to be spent during one particular frame, then $w(C(t)) = C(t)$. Nevertheless, the approach followed in this chapter is the same as the one applied in Chapter 3 of this dissertation in which we limit the battery amount that can be used in a particular frame in order to keep more energy in the battery over long periods of time. Thus, only a given fraction of the battery is allowed to be used during a particular frame, i.e.,

$$w(C(t)) = \alpha \cdot C(t), \quad 0 \leq \alpha \leq 1. \tag{6.5}$$

As we saw in Chapter 3, that value of α provides a trade-off between achieving larger sum rate at the beginning of the transmission and achieving larger aggregate sum rate throughout longer transmissions. In fact, there is an optimum value of α that can be obtained through numerical simulations that maximizes the aggregate sum rate.

6.2.1.3 Energy Harvesting Model

We assume a discretized model for the energy arrivals [Yan12a] where $H(t)$ is modeled as an ergodic Bernoulli process (which is a particular case of a Markov chain). As a result, only two values of harvested energy are possible, i.e., $H(t) \in \{0, e\}$, where e is the amount of Joules contained in an energy packet. The probability of receiving an energy harvesting packet during one frame depends on the actual harvesting intensity (in the case of solar energy, it depends on the particular hour of the day) and is denoted by $p(t)$. Note that a higher value of $p(t)$ will be obtained in frames where the harvesting intensity is higher, e.g., during solar presence such as during the day, and a lower value of $p(t)$ will be obtained during periods of solar absence, such as during the night.

6.2.1.4 System Assumptions

Let us collect all the channel gains, h_k, that includes the antenna gains, the path loss, and the fading, in $\mathbf{h} = \{h_k, \forall k \in \mathcal{U}_V \cup \mathcal{U}_D\}$. Generally, the wireless channels depend on the specific frame, $\mathbf{h}(t)$, as they vary over time but, for simplicity in the notation, we will just refer to them as $\mathbf{h}$ throughout the chapter. The traffic power, $P_{\text{BS}}(t)$ in (6.1), can be split into power for voice and data connections as $P_{\text{BS}}(t) = \sum_{k\in\mathcal{U}_V} \check{p}_k(\mathbf{h}) + \sum_{k\in\mathcal{U}_D} p_k(\mathbf{h})$, where $\check{p}_j(\mathbf{h})$ and $p_k(\mathbf{h})$ are the instantaneous powers corresponding to the transmission toward the j-th and k-th voice and data user, respectively. Let $P_{\text{RAD}}(t) = P_{\text{BS}}(t) + P_{\text{CPICH}}$ be the overall radiated power by the BS.

The voice users request a fixed data rate and we assume that just one WCDMA code is assigned to them. This is translated into a minimum SINR requirement as follows:

$$\frac{M_V \check{p}_k(\mathbf{h}) h_k}{\theta(P_{\mathrm{RAD}}(t) - \check{p}_k(\mathbf{h}))h_k + \sigma^2} \geq \Gamma, \quad \forall k \in \mathcal{U}_V, \tag{6.6}$$

where M_V is the spreading factor for voice codes, θ is the orthogonality factor among DL codes [Gol05], and σ^2 is the noise power. For simplicity in the notation and tractability, we will consider the following approximation[3]:

$$\theta(P_{\mathrm{RAD}}(t) - \check{p}_k(\mathbf{h}))h_k + \sigma^2 \approx \theta P_{\mathrm{RAD}}(t) h_k + \sigma^2. \tag{6.7}$$

On the other hand, the data users request a flexible service rate. The instantaneous throughput in the wireless access channel during one particular frame achieved by the k-th user, $r_k(\mathbf{h})$, is upper bounded by the maximum achievable rate that the access network is able to provide, which is formulated as

$$r_k(\mathbf{h}) \leq n_k(\mathbf{h}) \frac{W}{M_D} \log_2 \left(1 + \frac{M_D p_k(\mathbf{h}) h_k}{n_k(\mathbf{h})(\theta P_{\mathrm{RAD}}(t) h_k + \sigma^2)}\right), \tag{6.8}$$

where M_D is the spreading factor for data codes, W is the chip rate, and $n_k(\mathbf{h})$ is the number of codes assigned to user k. Notice that we have also approximated the denominator within the logarithm as in (6.7).

6.2.2 Problem Formulation and Resolution

Let us introduce the following set of definitions: $\mathbf{r} \triangleq \{r_k(\mathbf{h}),\ \forall k \in \mathcal{U}_D\}$, $\check{\mathbf{p}} \triangleq \{\check{p}_k(\mathbf{h}),\ \forall k \in \mathcal{U}_V\}$, $\mathbf{p} \triangleq \{p_k(\mathbf{h}),\ \forall k \in \mathcal{U}_D\}$, $\mathbf{n} \triangleq \{n_k, (\mathbf{h})\ \forall k \in \mathcal{U}_D\}$. We formulate an optimization problem for the resource allocation strategy with backhaul and energy constraints to be executed at the beginning of each particular frame, which involves finding the optimum resource allocation variables, $\mathbf{r}$, $\check{\mathbf{p}}$, $\mathbf{p}$, and $\mathbf{n}$ that maximize the minimum of the expected throughputs (note that if a scheduling criterion different from the maximin approach is to be taken, problem (6.9) could be extended

[3]If the number of users is relatively high, then $P_{\mathrm{RAD}}(t) \gg \check{p}_k$, and the approximation is fair. In any case, the approximation provides a lower bound of the actual SINR value.

by just reformulating the objective function accordingly):

$$
\begin{aligned}
\underset{\mathbf{r},\check{\mathbf{p}},\mathbf{p},\mathbf{n},P_{\mathrm{RAD}}(t)}{\text{maximize}} \quad & \min_{k\in\mathcal{U}_D} \ \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})] \qquad (6.9)\\
\text{subject to} \quad & C1: \frac{M_V \check{p}_k(\mathbf{h}) h_k}{\theta P_{\mathrm{RAD}}(t) h_k + \sigma^2} \geq \Gamma, \quad \forall k \in \mathcal{U}_V \\
& C2: \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})] \leq \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}, \quad \forall k \in \mathcal{U}_D \\
& C3: r_k(\mathbf{h}) \leq n_k(\mathbf{h}) \frac{W}{M_D} \log_2 \left(1 + \frac{M_D p_k(\mathbf{h}) h_k}{n_k(\mathbf{h})(\theta P_{\mathrm{RAD}}(t) h_k + \sigma^2)} \right), \quad \forall k \in \mathcal{U}_D \\
& C4: T_f \left(\sum_{k\in\mathcal{U}_V} \check{p}_k(\mathbf{h}) + \sum_{k\in\mathcal{U}_D} p_k(\mathbf{h}) \right) \leq \phi\left(C(t)\right) \\
& C5: \sum_{k\in\mathcal{U}_D} n_k(\mathbf{h}) \leq N_{\max} \\
& C6: r_k(\mathbf{h}) \geq 0,\ p_k(\mathbf{h}) \geq 0,\ n_k(\mathbf{h}) \geq 0, \quad \forall k \in \mathcal{U}_D \\
& C7: P_{\mathrm{RAD}}(t) = \sum_{k\in\mathcal{U}_V} \check{p}_k(\mathbf{h}) + \sum_{k\in\mathcal{U}_D} p_k(\mathbf{h}) + P_{\mathrm{CPICH}},
\end{aligned}
$$

where ξ, ($\xi > 1$), is an overhead considered for the data transmissions to be sent through the backhaul, $\check{R}_{BH}(|\mathcal{U}_V|)$ is the backhaul capacity used by the voice users[4], $|\mathcal{U}_V|$ being the number of voice users, R_{BH} is the overall backhaul capacity, Γ is the target SINR for the voice users, the function $\phi(\cdot)$ is related to $g(\cdot)$ in (6.4) as $\phi\left(C(t)\right) = g(C(t)) - T_f \cdot (P_{\mathrm{CPICH}} + P_c)$, and $N_{\max}$ is the number of available codes for the data users. Although all the variables in the optimization problem (6.9) depend on the frame t, we only keep such explicit dependence w.r.t. time in the variable $P_{\mathrm{RAD}}(t)$ to make explicit that the temporal evolution of the battery levels has a direct impact on the maximum power to be spent for the voice and data traffic, which is not constant over time.

It is important to realize that problem (6.9) may not be feasible due to constraint $C1$ as it may happen that there could not be enough power to satisfy all the target SINRs simultaneously. However, let us consider initially through the development that the problem is feasible (the feasibility condition will be developed later on). Notice that, at the optimum, $C4$ is attained with equality; otherwise, we could re-scale all the power variables with a common positive factor higher than 1 until $C4$ is fulfilled with equality. This would increase the objective function and all the other constraints would still be fulfilled. Because of this, we can assume that the optimum value of $P_{\mathrm{RAD}}(t)$ is $P^{\star}_{\mathrm{RAD}}(t) = \frac{\phi(C(t))}{T_f} + P_{\mathrm{CPICH}}$ and we can eliminate constraint $C7$ from problem (6.9). Constraint $C2$ states that the average throughput that a user is experiencing in the access network should not exceed the maximum backhaul rate assigned to this user (every user has been already assigned a portion of the backhaul, as commented above). If this is not the case,

[4]The overall backhaul capacity required to provide voice service generally depends on the current number of voice users being served. In some cases, voice users can be jointly encoded and, thus, the overall overhead for voice users may be reduced as the number of voice users increases. Anyway, in the problem formulation and the for the sake of generality, we just use the notation $\check{R}_{BH}(|\mathcal{U}_V|)$.

then $C2$ could be rewritten as $\sum_{k\in\mathcal{U}_D}\mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})] \leq \frac{R_{BH}-\check{R}_{BH}(|\mathcal{U}_V|)}{\xi}$. In any case, notice that the instantaneous rates allocated to one user in the access network can be higher in some frames than the maximum backhaul per-user rate $\left(\frac{R_{BH}-\check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}\right)$ thanks to the fact that queues are considered at the entrance of the access network. The average rate constraint $C2$ assures that the queues will be stable [Mar12].

Notice that the problem is separable into voice and data users without loss of optimality. This is because the voice users do not affect explicitly the optimal value of the objective function. Thus, we can obtain the optimum power variables for the voice users (i.e., the minimum power required to satisfy constraints $C1$) and, then, assign the rest of the resources to the data users. Hence, we will start by analyzing the resource allocation for the voice users in Section 6.2.2.1.

Although the previous problem and the resolution that will be presented in the following sections assume that the access technology is WCDMA, let us now provide the guidelines to reformulate the problem presented in (6.9) for an LTE system, which uses OFDMA as the underlying physical multiple access technology. The objective function could be expressed as it is. Constraints $C2$, $C4$, and $C6$ would remain equal (in fact, in $C6$ $n_k(\mathbf{h})$ would have to be changed by the variable representing the number of carriers). Constraint $C1$ needs to be modified. Neither the code gain (M_V) nor the intra-cell interference (as the access is now orthogonal) should be considered. We need to reformulate $C3$ completely. A possible candidate for the power-rate expression would be $\sum_{j=1}^{N}\beta_k^j W\log\left(1+\frac{p_k^j(\mathbf{h})h_k^j}{\sigma^2}\right)$, where β_k^j is a new binary optimization variable that takes a value equal to 1 if carrier j is assigned to user k and is 0 otherwise, $p_k^j(\mathbf{h})$ represents the power per carrier, and W is the carrier bandwidth. Finally, $C5$ and $C7$ would not be present. New constraints would have to be added: $\sum_{\forall j,\forall k}\beta_k^j \leq N$, where N is number of available carriers, $\sum_{\forall k}\beta_k^j = 1$, $\forall j$, and $\beta_k^j = \{0,1\}$.

6.2.2.1 Resource Allocation for Voice Users

Voice users must satisfy a minimum SINR constraint that is related to the target data rate service:

$$\frac{M_V\check{p}_k(\mathbf{h})h_k}{\theta P^{\star}_{\mathrm{RAD}}(t)h_k+\sigma^2} \geq \Gamma, \quad \forall k\in\mathcal{U}_V. \tag{6.10}$$

It is straightforward to obtain the optimum power allocation for each voice user as follows (realizing that at the optimum, constraints $C1$ are fulfilled with equality):

$$\check{p}_k^{\star}(\mathbf{h}) = \frac{\Gamma(\theta P^{\star}_{\mathrm{RAD}}(t)h_k+\sigma^2)}{M_V h_k}, \quad \forall k\in\mathcal{U}_V. \tag{6.11}$$

At this point, we could check the feasibility of (6.9). The problem is feasible if

$$T_f\sum_{k\in\mathcal{U}_V}\check{p}_k^{\star}(\mathbf{h}) \leq \phi(C(t)), \tag{6.12}$$

which could also be written only in terms of the channels of the voice users, the current battery level, and some constants as follows:

$$\sum_{k\in\mathcal{U}_V}\frac{1}{h_k}\leq\kappa_1\phi(C(t))-\kappa_2, \tag{6.13}$$

where $\kappa_1=\frac{M_V-|\mathcal{U}_V|\theta\Gamma}{\sigma^2T_f\Gamma}$ and $\kappa_2=\frac{|\mathcal{U}_V|\theta P_{\text{CPICH}}}{\sigma^2}$. If the problem is not feasible, then we should consider either reducing the minimum SINR requirements (which would increase the constant term κ_1), dropping out some voice users in the frame, or increasing $\phi(C(t))$ by taking a higher value for α, but always guaranteeing that the maximum radiated constraint $P_{\text{BS}}^{\max}$ is not exceeded.

6.2.2.2 Resource Allocation for Data Users

Now that we have considered the voice users, we can tackle the resource allocation problem for the data users by solving problem (6.9). Note that problem (6.9) is convex once we know $P_{\text{BS}}^{\star}(t)$. To solve problem (6.9), we will reformulate it by introducing the slack variable s, which preserves convexity [Boy04], as

$$\begin{aligned}\underset{s,\,\mathbf{r},\,\mathbf{p},\,\mathbf{n}}{\text{maximize}}\quad & s \\ \text{subject to}\quad & C2,\ldots,C6 \text{ of problem (6.9)} \\ & C8: s\leq\mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})],\quad \forall k\in\mathcal{U}_D \\ & C9: 0\leq s\leq\frac{R_{BH}-\check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}.\end{aligned} \tag{6.14}$$

Notice that we have introduced an additional constraint, $C9$. As it is clear from the formulation, this constraint does not affect the optimum solution, but it will help in the numerical search of the optimum value. Notice also that the previous optimization problem is time-coupled (we require the future channel realizations due to the expectation operator appearing in $C8$). In order to deal with such a difficult problem involving expectations, we propose to use a *stochastic approximation* [Rib10b]. In this approach, the constraints involving expectations are dualized, and their Lagrange multipliers are estimated stochastically at each period.

Let us start by dualizing constraint $C8$. Let $\boldsymbol{\lambda}\triangleq\{\lambda_k,\forall k\in\mathcal{U}_D\}$ be the vector of Lagrange multipliers associated with $C8$. The partial Lagrangian is given by $\mathcal{L}_{C8}(s;\boldsymbol{\lambda})=-s+\sum_{k\in\mathcal{U}_D}\lambda_k\left(s-\mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})]\right)$. In order to find the optimum s we have to perform the following minimization:

$$\underset{0\leq s\leq\frac{R_{BH}-\check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}}{\text{minimize}}\quad \mathcal{L}_{C8}(s;\boldsymbol{\lambda}). \tag{6.15}$$

According to [Rib10b], when the objective function is linear in the optimization variable, the stochastic primal-dual algorithms present some numerical problems. This can be avoided by transforming the objective function through the introduction of a general differentiable monotonically increasing cost function $U(\cdot)$ (e.g., the logarithm). Note that the use of this function does not modify the optimal value of the optimization variables (i.e., the solution is the same). Given that, setting the gradient to zero, $\nabla_s \mathcal{L}_{C8}(s;\boldsymbol{\lambda}) = 0$ and solving yields:

$$s^\star(\boldsymbol{\lambda}) = \left((\dot{U})^{-1} \left(\sum_{k \in \mathcal{U}_D} \lambda_k \right) \right)_0^{\frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}}, \tag{6.16}$$

where $\dot{U}(\cdot)$ is the derivative of $U(\cdot)$ and $(\dot{U})^{-1}(\cdot)$ is the inverse function of $\dot{U}(\cdot)$. Once we know $s^\star$, problem (6.14) is updated as follows (where we have skipped in the objective function the term that does not depend on the optimization variables remaining in the optimization problem):

$$\begin{aligned} \underset{\mathbf{r},\,\mathbf{p},\,\mathbf{n}}{\text{maximize}} \quad & \sum_{k \in \mathcal{U}_D} \lambda_k \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})] \\ \text{subject to} \quad & C2, \ldots, C6 \text{ of problem (6.9)}. \end{aligned} \tag{6.17}$$

Now, we proceed to dualize constraint $C2$. Let $\boldsymbol{\mu} \triangleq \{\mu_k, \forall k \in \mathcal{U}_D\}$ be the vector of Lagrange multipliers associated with $C2$. The partial Lagrangian is

$$\mathcal{L}_{C2}(r_k(\mathbf{h});\boldsymbol{\lambda},\boldsymbol{\mu}) = - \sum_{k \in \mathcal{U}_D} \lambda_k \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})] + \sum_{k \in \mathcal{U}_D} \mu_k \left(\mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})] - \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|} \right) \tag{6.18}$$

$$= - \mathbb{E}_{\mathbf{h}} \left[\sum_{k \in \mathcal{U}_D} (\lambda_k - \mu_k) r_k(\mathbf{h}) \right] - \sum_{k \in \mathcal{U}_D} \mu_k \left(\frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|} \right). \tag{6.19}$$

For given Lagrange multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, the optimization problem (6.14) is reformulated equivalently as (where we have skipped again in the objective function the term that does not depend on the optimization variables remaining in the optimization problem):

$$\begin{aligned} \underset{\mathbf{r},\,\mathbf{p},\,\mathbf{n}}{\text{maximize}} \quad & \sum_{k \in \mathcal{U}_D} (\lambda_k - \mu_k) r_k(\mathbf{h}) \\ \text{subject to} \quad & C3, \ldots, C6 \text{ of problem (6.9)}. \end{aligned} \tag{6.20}$$

Notice that the expectations are no longer present in the formulation because the remaining constraints $C3 - C6$ are applied to instantaneous resource allocation variables (without expectations) and, therefore, the maximization of the expected value of the objective function w.r.t. $\mathbf{r}$, $\mathbf{p}$, and $\mathbf{n}$ in the current frame is the same as the maximization of the term within the expectation. The problem now resides in the computation of the optimum Lagrange multipliers which requires knowing the statistics of $r_k(\mathbf{h})$. If we solve the dual problem of (6.20), i.e., $\sup_{\boldsymbol{\lambda} \succeq 0,\, \boldsymbol{\mu} \succeq 0} \inf_{\mathbf{r},\,\mathbf{p},\,\mathbf{n}} \mathcal{L}(\mathbf{r},\,\mathbf{p},\,\mathbf{n};\boldsymbol{\lambda},\boldsymbol{\mu})$, $\mathcal{L}(\mathbf{r},\,\mathbf{p},\,\mathbf{n};\boldsymbol{\lambda},\boldsymbol{\mu})$ is the Lagrangian of problem (6.20), using

a gradient approach, the optimum multipliers could be calculated recursively as [Boy04]

$$\lambda_k^{(q+1)} = \left(\lambda_k^{(q)} + \epsilon\left(s^\star(\boldsymbol{\lambda}^{(q)}) - \mathbb{E}_{\mathbf{h}}\left[r_k^\star\left(\mathbf{h};\boldsymbol{\lambda}^{(q)},\boldsymbol{\mu}^{(q)}\right)\right]\right)\right)_0^\infty, \quad \forall k, \tag{6.21}$$

$$\mu_k^{(q+1)} = \left(\mu_k^{(q)} + \epsilon\left(\mathbb{E}_{\mathbf{h}}\left[r_k^\star\left(\mathbf{h};\boldsymbol{\lambda}^{(q)},\boldsymbol{\mu}^{(q)}\right)\right] - \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}\right)\right)_0^\infty, \ \forall k, \tag{6.22}$$

where ϵ is the step size. Note that it is not possible to compute the value of the Lagrange multipliers in real time and then solve (6.20), as they depend on the statistics of $r_k(\mathbf{h})$ that is a function not known a priori (it is the solution of the optimization problem itself). In this situation, we propose to follow a stochastic approximation [Rib10b] and eliminate this uncertainty constraint by estimating the multipliers stochastically at each frame (with a noisy instantaneous unbiased estimate of the gradient) as follows (note that this philosophy is similar to the instantaneous estimation of the gradient in the classical LMS algorithm [Hay02]):

$$\lambda_k(t+1) = \left(\lambda_k(t) + \epsilon\left(s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h};\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))\right)\right)_0^\infty, \quad \forall k, \tag{6.23}$$

$$\mu_k(t+1) = \left(\mu_k(t) + \epsilon\left(r_k^\star(\mathbf{h};\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t)) - \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}\right)\right)_0^\infty, \quad \forall k. \tag{6.24}$$

Note that we have assumed that $s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h};\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))$ and $r_k^\star(\mathbf{h};\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t)) - \frac{R_{BH}-\check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}$ are stochastic subgradients of the dual function. In the following, we present a result that shows that $s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h};\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t))$ and $r_k^\star(\mathbf{h};\boldsymbol{\lambda}(t),\boldsymbol{\mu}(t)) - \frac{R_{BH}-\check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}$ are indeed stochastic subgradients of the dual function. We first introduce some notation. Let us collect the Lagrange multipliers in $\boldsymbol{\Psi}_k(t) = [\lambda_k(t),\, \mu_k(t)]^T$ and denote the stochastic subgradient by $\mathbf{t}_k(t) = [t_{k1}(t),\, t_{k2}(t)]^T$ with components

$$t_{k1}(t) = s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t)), \tag{6.25}$$

$$t_{k2}(t) = r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t))) - \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}. \tag{6.26}$$

Let the dual function of problem (6.14), after dualizing constraints $C2$ and $C8$, be defined as

$$\begin{aligned} g_k(\boldsymbol{\Psi}_k) \triangleq \underset{s,\mathbf{r},\mathbf{p},\mathbf{n}}{\text{minimize}} \quad & \mathcal{L}_k(s, r_k(\mathbf{h}); \boldsymbol{\Psi}_k) \\ \text{subject to} \quad & C3, \ldots, C6 \text{ of problem (6.9)} \\ & C9: 0 \le s \le \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}, \end{aligned} \tag{6.27}$$

where the Lagrangian is defined as

$$\begin{aligned} \mathcal{L}_k(s, r_k(\mathbf{h}); \boldsymbol{\Psi}_k) \triangleq & -s + \sum_{k\in\mathcal{U}_D} \lambda_k \left(s - \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})]\right) \\ & + \sum_{k\in\mathcal{U}_D} \mu_k \left(\mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})] - \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}\right). \end{aligned} \tag{6.28}$$

Proposition 6.1. *Given $\boldsymbol{\Psi}_k(t)$, the expected value of the stochastic subgradient $\boldsymbol{t}_k(t)$ is a subgradient of the dual function $g_k(\boldsymbol{\Psi}_k)$, i.e., $\forall \boldsymbol{\Psi}_k \geq \boldsymbol{0}$,*

$$\mathbb{E}_{\boldsymbol{h}}\left[\boldsymbol{t}_k^T(t) \,|\, \boldsymbol{\Psi}_k(t)\right](\boldsymbol{\Psi}_k - \boldsymbol{\Psi}_k(t)) \geq g_k(\boldsymbol{\Psi}_k) - g_k(\boldsymbol{\Psi}_k(t)). \tag{6.29}$$

In particular,

$$\mathbb{E}_{\boldsymbol{h}}\left[\boldsymbol{t}_k^T(t) \,|\, \boldsymbol{\Psi}_k(t)\right](\boldsymbol{\Psi}_k^\star - \boldsymbol{\Psi}_k(t)) \geq D_k - g_k(\boldsymbol{\Psi}_k(t)) \geq 0, \tag{6.30}$$

where $D_k \triangleq \max_{\boldsymbol{\Psi}_k \geq \boldsymbol{0}} g_k(\boldsymbol{\Psi}_k)$ and $\boldsymbol{\Psi}_k^\star \triangleq \arg\max_{\boldsymbol{\Psi}_k \geq \boldsymbol{0}} g_k(\boldsymbol{\Psi}_k)$.

Proof. See Appendix 6.A. ■

The previous proposition states that the average of the stochastic subgradient $\mathbf{t}_k(t)$ is a subgradient of the dual function. Since $\mathbb{E}_{\mathbf{h}}[\mathbf{t}_k(t) \,|\, \boldsymbol{\Psi}_k(t)]$ points towards $\boldsymbol{\Psi}_k^\star$ (the angle between $\mathbb{E}_{\mathbf{h}}\left[\mathbf{t}_k^T(t) \,|\, \boldsymbol{\Psi}_k(t)\right]$ and $\boldsymbol{\Psi}_k^\star - \boldsymbol{\Psi}_k(t)$ is lower than 90^{o}), it is not difficult to prove that $\boldsymbol{\Psi}_k(t)$ eventually approaches $\boldsymbol{\Psi}_k^\star$ [Sho85]. Thus, as $\mathbf{t}_k(t)$ points towards the set of optimal dual variables $\boldsymbol{\Psi}_k^\star$ in average, it is reasonable to expect that stochastic subgradient ascent iterations in (6.23) and (6.24) will also approach $\boldsymbol{\Psi}_k^\star$ in some sense. For more details, please see Section 2.2.

The advantages of the stochastic techniques are threefold: *i*) the computational complexity of the stochastic technique is significantly lower than that of their off-line counterparts; *ii*) stochastic approaches can deal with non-stationarity environments; *iii*) the distribution of the involved random variables $\mathbf{h}$ is not required.

Once we update the values of the Lagrange multipliers, problem (6.20) can be solved using, for example, a primal-dual approach. Notice that constraint $C3$ can be put directly in the objective function as, at the optimum, it is fulfilled with equality, i.e., $r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) = n_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) \frac{W}{M_D} \log_2\left(1 + \frac{M_D p_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) h_k}{n_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))(\theta P_{\text{RAD}}^\star(t) h_k + \sigma^2)}\right)$. Thus, the resource allocation problem to be solved at the beginning of the frame t is given by

$$\begin{aligned} \underset{\mathbf{p},\,\mathbf{n}}{\text{maximize}} \quad & \sum_{k \in \mathcal{U}_D} (\lambda_k(t) - \mu_k(t)) n_k \frac{W}{M_D} \log_2\left(1 + \frac{M_D p_k h_k}{n_k(\theta P_{\text{RAD}}^\star(t) h_k + \sigma^2)}\right) \\ \text{subject to} \quad & C4, \ldots, C6 \text{ of problem (6.9)}. \end{aligned} \tag{6.31}$$

Problem (6.31) can be solved as described in Appendix 6.B.

It can be shown that the sample average of the stochastic rates, $r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))$, satisfies all the constraints in (6.9) and incurs minimal performance loss relative to the optimal (off-line) solution of (6.9). This can be stated rigorously as follows: let us define $F(t) \triangleq \min_{k \in \mathcal{U}_D} \frac{1}{t} \sum_{\tau=1}^{t} r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(\tau), \boldsymbol{\mu}(\tau))$ and $f^\star$ as the minimum value of the objective function in (6.9). Then, it holds with probability one that as $t \to \infty$: *i*) the solution is feasible; and *ii*) $F(t) \leq f^\star + \delta(\epsilon)$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. A proof of this result is not presented here due

to space limitations but it can be derived following [Rib10b]. Let us introduce some important remarks here regarding problem (6.31):

Remark 1: the values of $\{\lambda_k\}$ measure how far the average rate $s^\star(\boldsymbol{\lambda}(t))$ is from the instantaneous rates served to the users in the access network. If the quality of the channels or the available powers are such that the instantaneous rates served in the access network are far from the target average rate $s^\star(\boldsymbol{\lambda}(t))$ for all the users, the sum of $\{\lambda_k\}$ will increase (see (6.23)) and the system will reduce the target average rate $s^\star(\boldsymbol{\lambda}(t))$ (see (6.16)).

Remark 2: in the access network, the values of $\{\lambda_k\}$ are in charge of ensuring that all average rates tend to grow simultaneously (maximin objective function). Note that, if at any temporal period $s^\star(\boldsymbol{\lambda}(t)) > r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))$, then $\lambda_k(t+1)$ grows (see (6.23)) and the priority of the k-th user to be served increases.

Remark 3: if at any temporal period $r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) > \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}$, then $\mu_k(t+1) > \mu_k(t)$ (see (6.24)). For a fixed set of $\{\lambda_k\}$, if $\lambda_k(t+1) - \mu_k(t+1)$ decreases, the user will have a lower priority to be served in the next period. The same reasoning could be applied if $r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) < \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}$ to deal with the inverse situation.

6.2.2.3 Overall Resource Allocation Algorithm for the Downlink

In this subsection, we present the overall algorithm that solves the resource allocation for the voice and data users based on the approaches presented in previous sections. This algorithm should be solved by the BS at the beginning of each frame (whose duration is chosen usually according to the channel dynamics). The algorithm is summarized in Table 6.1. Notice that in the algorithm, steps 7 to 15 correspond to the steps presented in Appendix 6.B to solve the convex optimization problem (6.31). Note also that the computational burden of the proposed scheme is similar to the one of the conventional PF approach [Wan07]. In steps 10 and 11, we need to solve two water-filling-like expressions, and the rest of the steps are just simple updates.

6.2.3 Numerical Simulations

In this section we evaluate the performance of the proposed strategy. The scenario under consideration is composed of 1 BS, 3 voice users, and 6 data users. The maximum radiated power is $P_{\text{BS}}^{\max} = 9$ dBm, the pilot power is $P_{\text{CPICH}} = 4$ dBm (which represents the 13% of the maximum radiated power, as we considered in [tuc13]), and the fixed power is $P_c = 3$ dBm (considering the model in [Aue11], which was applied in [tuc13]). The number of available codes for data transmission services is $N_{\max} = 15$. All the users are mobile with a speed of 3 m/s. The instantaneous channel gain, h_k, incorporates antenna gains, Rayleigh fading with unitary gain, and the path loss. The path losses correspond to a town in Perú known as Tuta Pisco (see

Algorithm 6.1 Algorithm for solving resource allocation problem (6.9)

initialize $\lambda_k(t) \geq 0$, $\mu_k(t) \geq 0$, $\forall k \in \mathcal{U}_D$

compute $P^{\star}_{\mathrm{RAD}}(t) = \frac{\phi(C(t))}{T_f} + P_{\mathrm{CPICH}}$

Voice users

  compute $\check{p}^{\star}_k(\mathbf{h}) = \frac{\Gamma(\theta P^{\star}_{\mathrm{RAD}}(t) h_k + \sigma^2)}{M_V h_k}, \quad \forall k \in \mathcal{U}_V$

  if $T_f \sum_{k \in \mathcal{U}_V} \check{p}^{\star}_k(\mathbf{h}) > \phi(C(t)) \quad \longrightarrow$ drop some voice users or reduce Γ, go to 3

Data users

  repeat

    initialize $\mathbf{n} \succeq 0$

    repeat

      $p_k^{(q,k+1)} = p^{\star}_k\left(\mathbf{n}^{(q,k)}, \beta^{(q)}, \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)\right)$ using (6.63), $\forall k \in \mathcal{U}_D$

      $n_k^{(q,k+1)} = n^{\star}_k\left(\mathbf{n}^{(q,k)}, \mathbf{p}^{(q,k+1)}, \varphi^{(q)}, \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)\right)$ using (6.64), $\forall k \in \mathcal{U}_D$

    until $p_k^{(q,k+1)}$ and $n_k^{(q,k+1)}$ converge

    update variables, $\beta^{(q+1)}$ and $\varphi^{(q+1)}$, using $p_k^{(q)}$ and $n_k^{(q)}$ with (6.66) and (6.67)

  until $\beta^{(q+1)}$ and $\varphi^{(q+1)}$ converge

  compute $r^{\star}_k(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))$ with $p^{\star}_k(\mathbf{n}, \beta)$ and $n^{\star}_k(\mathbf{p}, \mathbf{n}, \varphi)$

  update (dualized) primal variable:

  $s^{\star}(\boldsymbol{\lambda}(t)) = \left((\dot{U})^{-1}\left(\sum_{k \in \mathcal{U}_D} \lambda_k(t)\right)\right)_0^{\frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}}$

  update stochastic dual variables:

  $\lambda_k(t+1) = \left(\lambda(t) + \epsilon\left(s^{\star}(\boldsymbol{\lambda}(t)) - r^{\star}_k(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))\right)\right)_0^{\infty}$

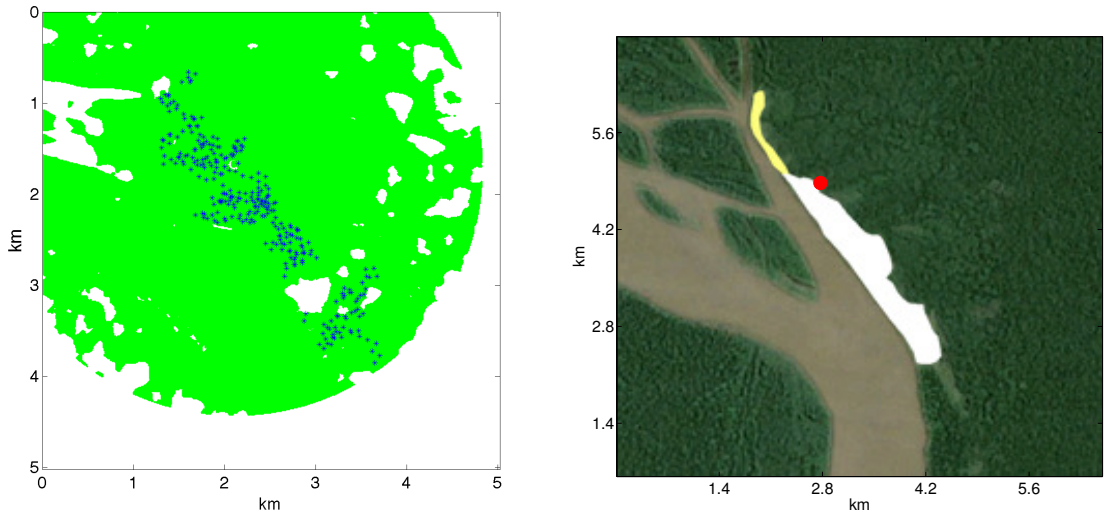
  $\mu_k(t+1) = \left(\mu(t) + \epsilon\left(r^{\star}_k(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) - \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}\right)\right)_0^{\infty}$

Battery update

  $C(t+1) = \left(C(t) - E(t) + H(t)\right)_0^{C_{\max}}$

$t \longleftarrow t+1$ and go to 2

end algorithm



(a) Snapshot of the user distribution in town Tuta Pisco.

(b) Aerial view of Tuta Pisco.

Figure 6.2: Snapshot of the user distribution and the aerial view of the town Tuta Pisco located in the forest in Perú.

details in [tuc13]). Figure 6.2 shows a snapshot of the user distribution and the aerial view of the town Tuta Pisco. The orthogonality factor is $\theta = 0.35$. The code gain of data codes $M_D = 16$ and the minimum SINR normalized by the code gain for voice users is $\frac{\Gamma}{M_V} = -13.7$ dB, which corresponds to a rate of 12.2 Kbps. The noise power is $\sigma^2 = -102$ dBm. The battery capacity is $C_{\max} = 410$ μJ, the energy packet size is $e = 30$ μJ, and $\alpha = 0.3$, unless otherwise stated. The frame duration for the data users and voice users are 2ms and 20ms, respectively, thus, $T_f = 2$ ms. The utility function is $U(\cdot) = \log(\cdot)$. Two backhaul capacities have been considered in the simulations: $R_{BH} = 2$ Mbps and $R_{BH} = 500$ Kbps. The amount of backhaul capacity required by the 3 voice users considered in this deployment is $\check{R}_{BH}(|\mathcal{U}_V|) = 173$ Kbps. The overhead for the data transmissions is $\xi = 1.2$. The step size for the update of the stochastic multipliers is $\epsilon = 10^{-3}$. For a more detailed description of the simulation parameters see [tuc13].

In the simulations, we consider as a benchmark the case where the BS is connected to the electric grid (which means equivalently that the battery remains full of energy for the whole simulation). For comparison purposes, we also show the resource allocation of the proposed strategy and the PF strategy both with an instantaneous per-user backhaul constraint $r_k(\mathbf{h}) \leq \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}$. Thanks to this, we can compare the performance of the stochastic maximin strategy with well-known scheduling strategies and we can directly measure the improvement of the proposed stochastic scheme when we allow instantaneous access data rates to surpass the backhaul capacity. The effective length of the exponential window in the PF scheme has been set to $T_c = 500$ [Wan07][5]. In the figures, PE refers to the solution of Algorithm 6.1, PI refers to the strategy from (6.9) but replacing constraint $C2$ by an instantaneous backhaul constraint, and PF refers to the PF scheme.

[5]The weights of the PF scheduler are calculated as $\omega_k(t) = \frac{1}{T_k(t)}$, where $T_k(t)$ is the average throughput of user k computed as $T_k(t) = (1 - \frac{1}{T_c})T_k(t-1) + \frac{1}{T_c} r_k(t)$.

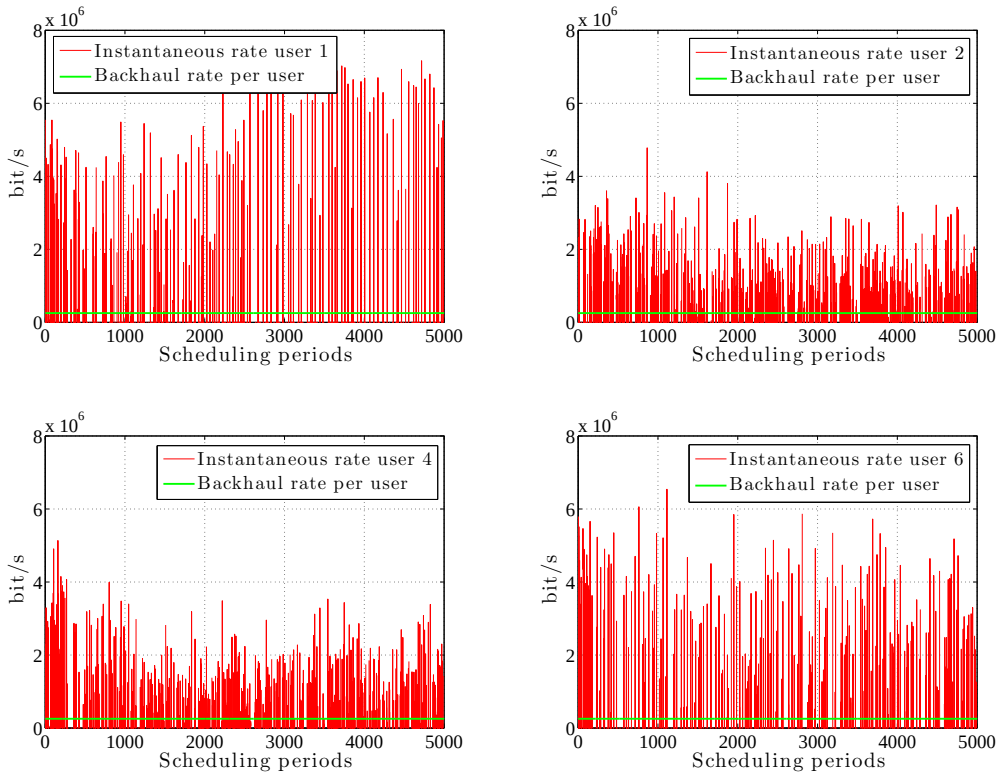


Figure 6.3: Time evolution of the instantaneous data rates served at the access network and the backhaul capacity limitation per user with a backhaul capacity of 2 Mbps.

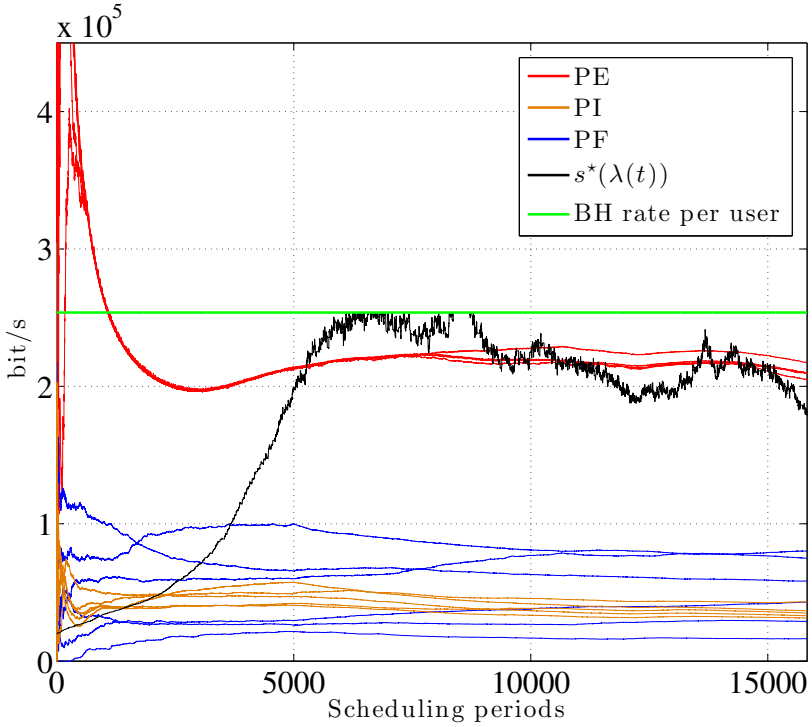


Figure 6.4: Time evolution of the data rates for the different approaches and the backhaul capacity per user when the BS is connected to the electric grid with a backhaul capacity of 2 Mbps.

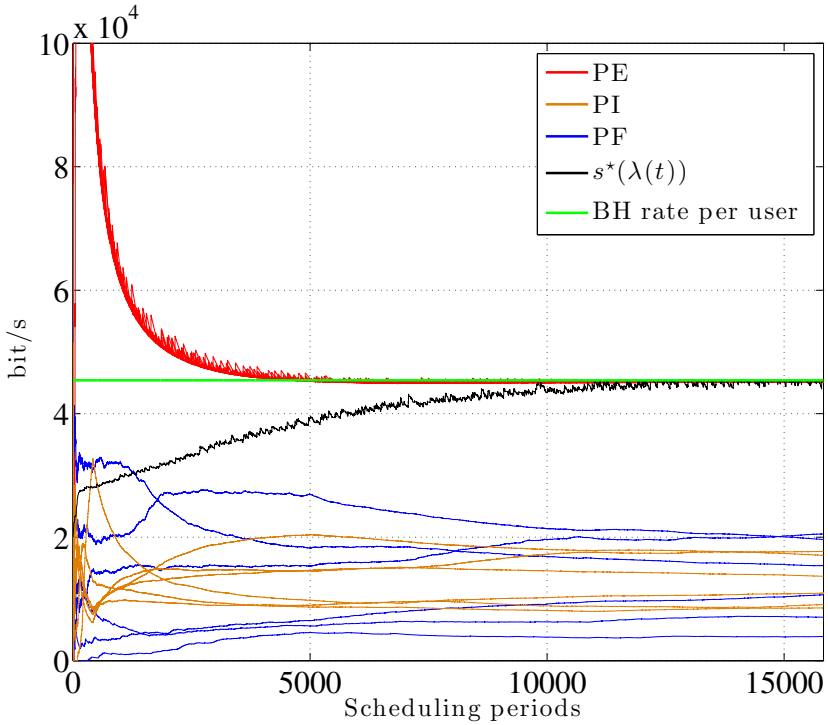


Figure 6.5: Time evolution of the data rates for the different approaches and the backhaul capacity per user when the BS is connected to the electric grid with a backhaul capacity of 500 Kbps.

Figure 6.3 presents the instantaneous data rates served at the access network of four data users out of the six. In this case, the BS is connected to the electric power grid. As we can see, the instantaneous rates are able to exceed the backhaul capacity in particular frames whereas, at the same time, the average rates fulfills the maximum backhaul capacity as it is shown in Figure 6.4.

Figure 6.4 shows the time evolution of the expected data rates of the three approaches. At any time instant the expected rates have been estimated using $r_k(t) = \frac{1}{t}\sum_{\tau=1}^{t} r_k(\tau)$. We also plot the time evolution of $s^\star(\boldsymbol{\lambda}(t))$ and the per-data user backhaul rate. In this case, the BS is connected to the electric grid. The backhaul capacity is $R_{BH} = 2$ Mbps. Initially, we assume that the queues at the access network are sufficiently full so that all the bits demanded by the users are served. This allows the initial average rates to violate the backhaul capacity constraint for a short period of time (see the initial transient in the figure). This is due to the stochastic approximation of the multipliers but, in any case, when the average rates converge, they fulfill all the constraints of the original problem. As we can also see in the figure, the limitation of the rates comes from the limited resources available at the access network, i.e., the power and the codes, as the backhaul capacity is not reached. It should be also emphasized that the proposed stochastic approach provides a solution that introduces more fairness when compared with the PF approach as the average rates for the different users are quite similar. Figure 6.5 depicts the same curves but now considering a backhaul capacity of $R_{BH} = 500$ Kbps. As we can see, now the system is limited by the backhaul and not by the limited resources of the access network. It

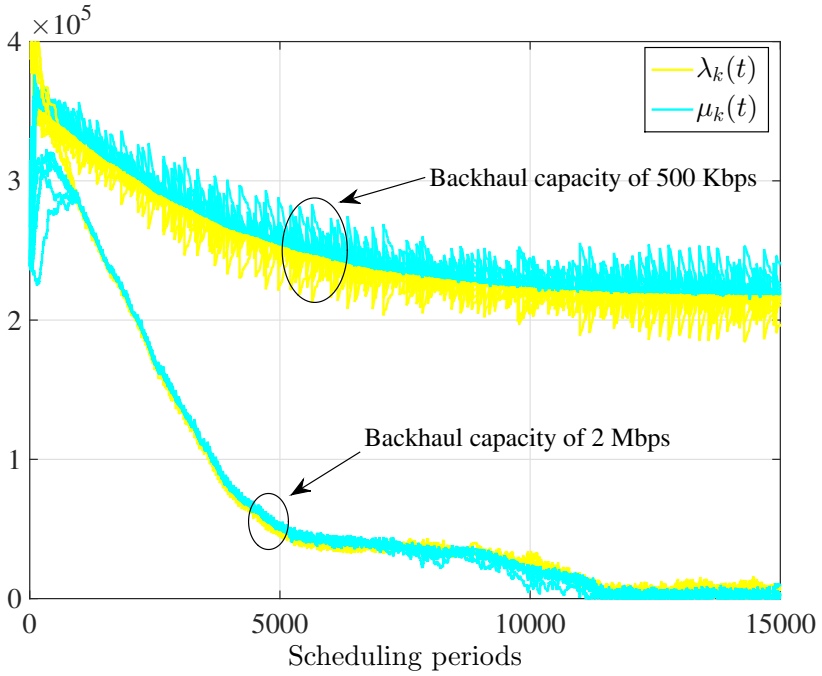


Figure 6.6: Time evolution of the stochastic Lagrange multipliers for different backhaul capacities.

can be appreciated in both figures the improvement of the data rates when considering average backhaul constraints instead of instantaneous backhaul constraints.

Figure 6.6 shows the evolution of the stochastic estimates of the Lagrange multipliers, i.e., $\lambda_k(t)$ and $\mu_k(t)$, for the cases where we have a backhaul capacity of $R_{BH} = 2$ Mbps and the case of having a backhaul capacity of $R_{BH} = 500$ Kbps. From duality theory, we know that if the backhaul constraint is not active, i.e., if the expected rates are below the backhaul capacity, then the optimum value of the multipliers is zero. This is what we see in the figure for the case of having a backhaul capacity of 2 Mbps. On the other hand, if the system is limited by the backhaul, then the optimum Lagrange multipliers are generally not zero as the corresponding constraints become active. In the figure we see that the multipliers converge to a non-zero value. The convergence of the multipliers states that the overall stochastic approach is working properly.

Figure 6.7 depicts the expected rate of the worst user (shown in the left figure) and the sum of the average data rates $\left(\frac{1}{t}\sum_{\tau=1}^{t}\sum_{k\in\mathcal{U}_D} r_k(\tau)\right)$ (right figure) as a function of the overall backhaul capacity (R_{BH}) for the different approaches when the BS is connected to the grid and when the BS has a finite battery with different harvesting intensities p. Concerning the comparison between our strategy and the PF approach, we can see that the proposed scheme provides a greater data rate to the worst-user rate in all configurations and for all backhaul bandwidths. Here we emphasize again the great improvement in terms of worst-rate and sum-rate that the stochastic with average backhaul constraints provides. In other words, our proposed solution achieves much more fairness as shown in Figure 6.4 and Figure 6.5. This is expected since we are considering explicitly the maximin criterion in the proposed resource allocation strategy.

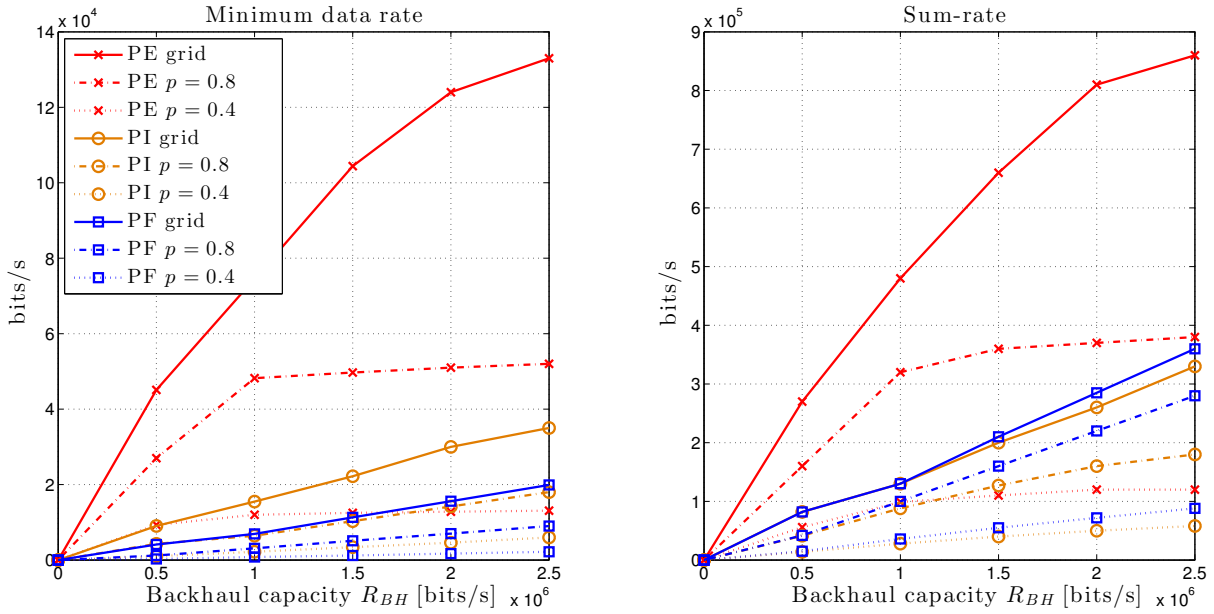


Figure 6.7: Sum rates as a function of backhaul capacity for different approaches and different probabilities of energy packet p.

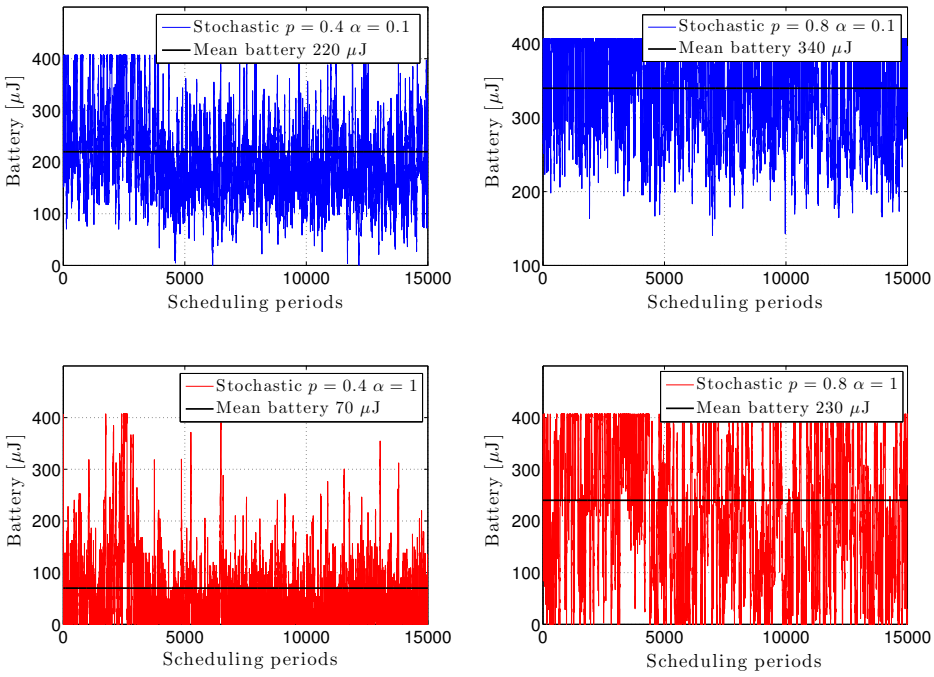


Figure 6.8: Battery evolution of the proposed stochastic approach and the PF with sum constraint with a probability of energy packet $p = 0.4$ and $p = 0.8$ and for $\alpha = 0.1$ and $\alpha = 1$.

Finally, Figure 6.8 depicts the time evolution of the instantaneous battery level of the BS when the electric grid is not available for the stochastic approach. We assume that the probability of receiving an energy packet during one frame is $p = 0.4$ and $p = 0.8$ and two values of α have been considered: $\alpha = 0.1$ and $\alpha = 1$. Recall that $\alpha = 1$ means that all the battery could be used during one particular frame (if the maximum power than can be radiated by the BS allows it). In can be proved that if no radiated power limitation exists at the BS and the battery never reaches its maximum value, then theoretically the expected value of the battery is given by $\hat{b} = \frac{\mathbb{E}[H(t)]}{\alpha} = \frac{p \cdot e}{\alpha}$ (see Section 3.4.5). However, due to the maximum power radiation at the BS and the battery overflows, the previous expression yields a lower bound of the true expected battery level, i.e., $\hat{b} \leq \lim_{t\to\infty} \mathbb{E}[C(t)]$. For example, if $p = 0.8$, $\alpha = 0.1$, then $\hat{b} = 240$ μJ, but the figures shows $\lim_{t\to\infty} \mathbb{E}[C(t)] = 340$ μJ.

6.3 Stochastic Resource Allocation for the Uplink Scenario

In the previous section, we developed resource allocation strategies for the DL setup where we considered a backhaul capacity constraint and the BS was powered with a finite battery and an energy harvesting source that recharged it. Now, in this section, based on the same setup, we develop a resource allocation strategy for the UL scenario. The resource allocation will also be based on assigning powers and codes to voice users and data users.

6.3.1 System Model and Assumptions

Let us consider an UL system composed of a single BS and several users. We also consider that there is a maximum backhaul rate constraint for the UL connections in terms of the average throughput.

Following the notation that we employed for the DL, let us collect all the channel gains, h_k that includes the antenna gains, the path loss, and the fading, in $\mathbf{h} = \{h_k, \forall k \in \mathcal{U}_V \cup \mathcal{U}_D\}$. Let $\check{p}_j(\mathbf{h})$ and $p_k(\mathbf{h})$ be the instantaneous powers corresponding to the transmission from the j-th and k-th voice and data user, respectively, and $n_k(\mathbf{h})$ be the number of codes assigned to the k-th data user.

The set of voice users request a fixed data rate and we assume that just one WCDMA code is assigned to them. This is translated into a minimum SINR requirement at the BS as follows:

$$\frac{M_V \check{p}_k(\mathbf{h}))h_k}{\sum_{\ell\in\mathcal{U}_D} p_\ell(\mathbf{h})h_\ell + \sum_{m\in\mathcal{U}_V, m\neq k} \check{p}_m(\mathbf{h}) + \sigma^2} \geq \Gamma, \quad \forall k \in \mathcal{U}_V, \tag{6.32}$$

where M_V is the spreading factor for voice codes, σ^2 is the noise power of the BS, and Γ is the target SINR. It is important to emphasize that the target SINR has to be achieved with

equality (any other solution that fulfills the SINR constraint with strict inequality implies a power spending higher than necessary and, consequently, higher levels of interference). Note also that the orthogonality factor does not appear in the previous expression since in UL we assume that users are not synchronized in time.

On the other hand, the set of data users request a flexible service rate. The instantaneous throughput in the wireless access channel achieved during one particular frame by the k-th user, $r_k(\mathbf{h})$, is upper bounded by the maximum achievable rate that the access network is able to provide, which is formulated as

$$r_k(\mathbf{h}) \leq n_k(\mathbf{h}) \frac{W}{M_D} \log_2 \left(1 + \frac{M_D p_k(\mathbf{h}) h_k}{n_k \left(\sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell - \frac{p_k(\mathbf{h}) h_k}{n_k(\mathbf{h})} + \sum_{m \in \mathcal{U}_V} \check{p}_m(\mathbf{h}) + \sigma^2 \right)} \right), \quad (6.33)$$

where M_D is the spreading factor for data codes, W is the chip rate. Following the same approach as we did for the DL, we will approximate the denominator of (6.33) by

$$\sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell - \frac{p_k(\mathbf{h}) h_k}{n_k(\mathbf{h})} + \sum_{m \in \mathcal{U}_V} \check{p}_m(\mathbf{h}) \sigma^2 \approx \sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell + \sum_{m \in \mathcal{U}_V} \check{p}_m(\mathbf{h}) + \sigma^2, \quad (6.34)$$

which is a reasonable assumption if the number of users is relatively high. In any case, when using the approximation, the previous obtained rate corresponds to a lower bound of the achievable rate.

The number of data codes assigned to data users has to fulfill the following condition:

$$n_k(\mathbf{h}) \leq N_{\max}^{(k)}, \quad \forall k \in \mathcal{U}_D. \quad (6.35)$$

Notice that now each user has an independent constraint in terms of maximum number of data codes to be used. This is so, as each user is allocated a different scrambling code. By following the approximation in (6.34), the right-hand side in (6.33) is an increasing function of $n_k(\mathbf{h})$. This implies that it is optimum to use all data codes available for all data users (note that this is one of the main differences w.r.t. the DL case):

$$n_k^\star(\mathbf{h}) = N_{\max}^{(k)}, \quad \forall k \in \mathcal{U}_D. \quad (6.36)$$

Given the previous result, only the powers for the voice and data users must be allocated. As happened for the DL case, the structure of the optimization problem allows for the separation of the voice users and the data users without loosing optimality. For that reason, we will first develop the resource allocation strategy for the voice users and then for the data users. The overall problem formulation is not presented for brevity, but it will be apparent when presenting the formulation of the resource allocation strategy for the data users.

6.3.2 Problem Formulation and Resolution

6.3.2.1 Resource Allocation for the Voice Users

Let us define the following variable that takes into account the noise in addition to the received power corresponding to the data connections:

$$\sigma_{int}^2 = \sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell + \sigma^2. \tag{6.37}$$

According to this, the set of equations presented in (6.32) can be written in matrix form as follows (each row corresponds to each voice user, where we have taken the following order in the numbering: $k = 1, 2, ..., |\mathcal{U}_V|$, being $|\mathcal{U}_V|$ the total number of active voice users):

$$\begin{bmatrix} M_V & -\Gamma & -\Gamma & \dots & -\Gamma \\ -\Gamma & M_V & -\Gamma & \dots & -\Gamma \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\Gamma & -\Gamma & -\Gamma & \dots & M_V \end{bmatrix} \times \begin{bmatrix} \check{p}_1(\mathbf{h}) h_1 \\ \check{p}_2(\mathbf{h}) h_2 \\ \vdots \\ \check{p}_{|\mathcal{U}_V|}(\mathbf{h}) h_{|\mathcal{U}_V|} \end{bmatrix} = \sigma_{int}^2 \Gamma \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Notice that all the previous equations are completely symmetric w.r.t. users. This implies that the power allocated to voice users are inversely proportional to the channels, i.e., the powers received at the BS from all voice users must be equal:

$$\check{p}_k(\mathbf{h}) = \frac{\vartheta}{h_k}, \quad \forall k \in \mathcal{U}_V, \tag{6.38}$$

where

$$\vartheta = \frac{\sigma_{int}^2 \Gamma}{M_V - \Gamma(|\mathcal{U}_V| - 1)}. \tag{6.39}$$

As opposed to the DL case, in the UL scenario, the transmit power constraints are individual, i.e., on a per-user basis. That means that if $\check{p}_k(\mathbf{h}) = \frac{\vartheta}{h_k} > P_T^{(k)}$ for some voice user (where $P_T^{(k)}$ represents the maximum transmission power for the k-th voice user), then the SINR constraints cannot be fulfilled and some users should be dropped off from the system. Notice that if we want to assure that all voice users achieve its minimum SINR, we have to impose a constraint on the interference that the data users generate to the voice users as follows:

$$\vartheta = \frac{\sigma_{int}^2 \Gamma}{M_V - \Gamma(|\mathcal{U}_V| - 1)} = \left(\sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell + \sigma^2 \right) \frac{\Gamma}{M_V - \Gamma(|\mathcal{U}_V| - 1)} \leq \min_{k \in \mathcal{U}_V} P_T^{(k)} h_k, \tag{6.40}$$

or, equivalently,

$$\sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell \leq \left(\min_{k \in \mathcal{U}_V} P_T^{(k)} h_k \right) \frac{M_V - \Gamma(|\mathcal{U}_V| - 1)}{\Gamma} - \sigma^2. \tag{6.41}$$

Note that (6.41) is a constraint on the maximum power radiated by all data users simultaneously, i.e., sum-power constraint. As a result, (6.41) must be added in the resource allocation for data users if we want to assure that voice users receive the service they demand.

6.3.2.2 Resource Allocation for the Data Users

Now, we can proceed to obtain the optimum power allocation for the data users. Before presenting the optimization problem, let us formulate the achievable rate in terms of the previous results:

$$\begin{aligned} r_k(\mathbf{h}) &\leq \frac{N_{\max}^{(k)} W}{M_D} \log_2 \left(1 + \frac{M_D p_k(\mathbf{h}) h_k}{N_{\max}^{(k)} \left(\sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell - \frac{p_k(\mathbf{h}) h_k}{n_k(\mathbf{h})} + \frac{\sigma_{int}^2 \Gamma |\mathcal{U}_V|}{M_V - \Gamma(|\mathcal{U}_V| - 1)} + \sigma^2 \right)} \right) \\ &= \frac{N_{\max}^{(k)} W}{M_D} \log_2 \left(1 + \frac{M_D p_k(\mathbf{h}) h_k}{N_{\max}^{(k)} c_V \left(\sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell + \sigma^2 \right)} \right), \end{aligned} \tag{6.42}$$

where c_V is defined as

$$c_V = 1 + \frac{\Gamma |\mathcal{U}_V|}{M_V - \Gamma(|\mathcal{U}_V| - 1)}. \tag{6.43}$$

Given the previous definitions and taking the same objective function as the one considered in the DL formulation (see Section 6.2.2), in which we optimize the expected data rate of the user that achieves the worst rate, we can formulate the resource allocation strategy for the data users for the UL connections as

$$\begin{aligned} \underset{\mathbf{r}, \mathbf{p}, s}{\text{maximize}} \quad & U(s) \\ \text{subject to} \quad & C1: s \leq \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})], \quad \forall k \in \mathcal{U}_D \\ & C2: \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h})] \leq \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}, \quad \forall k \in \mathcal{U}_D \\ & C3: r_k(\mathbf{h}) \leq N_{\max}^{(k)} \frac{W}{M_D} \log_2 \left(1 + \frac{M_D p_k(\mathbf{h}) h_k}{N_{\max}^{(k)} c_V \left(\sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell + \sigma^2 \right)} \right), \quad \forall k \in \mathcal{U}_D \\ & C4: \sum_{\ell \in \mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell \leq \left(\min_{k \in \mathcal{U}_V} P_T^{(k)} h_k \right) \frac{|\mathcal{U}_V|}{c_V - 1} - \sigma^2. \\ & C5: p_k \leq P_T^{(k)}, \quad \forall k \in \mathcal{U}_D. \end{aligned} \tag{6.44}$$

Now, we proceed similar as we did for the DL case. We dualize constraints $C1$ and $C2$ of

problem (6.44) and obtain the following equivalent problem:

$$\begin{aligned}
\underset{\mathbf{p}}{\text{maximize}} \quad & \sum_{k\in\mathcal{U}_D} (\lambda_k - \mu_k) N_{\max}^{(k)} \frac{W}{M_D} \log_2\left(1 + \frac{M_D p_k(\mathbf{h}) h_k}{N_{\max}^{(k)} c_V \left(\sum_{\ell\in\mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell + \sigma^2\right)}\right) \\
\text{subject to} \quad & C4: \sum_{\ell\in\mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell \leq \left(\min_{k\in\mathcal{U}_V} P_T^{(k)} h_k\right) \frac{|\mathcal{U}_V|}{c_V - 1} - \sigma^2 \\
& C5: p_k \leq P_T^{(k)}, \quad \forall k \in \mathcal{U}_D,
\end{aligned} \tag{6.45}$$

where λ_k and μ_k are the dual variables associated to constraints $C1$ and $C2$. Also in this case we follow the approach based on stochastic optimization theory in order to obtain estimates of λ_k and μ_k, denoted as $\lambda_k(t)$ and $\mu_k(t)$, as follows:

$$\lambda_k(t+1) = \left(\lambda_k(t) + \epsilon\left(s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))\right)\right)_0^\infty, \quad \forall k, \tag{6.46}$$

$$\mu_k(t+1) = \left(\mu_k(t) + \epsilon\left(r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) - \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}\right)\right)_0^\infty, \quad \forall k, \tag{6.47}$$

where

$$r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) = \frac{N_{\max}^{(k)} W}{M_D} \log_2\left(1 + \frac{M_D p_k^\star(\mathbf{h}) h_k}{N_{\max}^{(k)} c_V \left(\sum_{\ell\in\mathcal{U}_D} p_\ell^\star(\mathbf{h}) h_\ell + \sigma^2\right)}\right), \quad \forall k \in \mathcal{U}_D, \tag{6.48}$$

being $p_k^\star(\mathbf{h})$ the optimum powers obtained from the optimization problem (6.45). In Proposition 6.1 in the previous section, it was proved that the proposed stochastic subgradients, that is, $s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))$ and $r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) - \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}$, were valid subgradients of the dual function. It is important to note that problem (6.45) is not convex due to the interference term in the rate expression. Given that, let us present the methodology employed to find the optimum power allocation. First, we introduce introduce the following constraint that is to be added without loss of optimality:

$$\sum_{\ell\in\mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell \leq \sum_{\ell\in\mathcal{U}_D} P_T^{(\ell)} h_\ell. \tag{6.49}$$

With this, we guarantee that either constraint $C4$ of problem (6.45) or (6.49) will be active at the optimum. Secondly, we introduce the following slack variable q that is defined as

$$q = \sum_{\ell\in\mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell. \tag{6.50}$$

Having introduced the previous constraint and variable, we may write problem (6.45) as follows:

$$\begin{aligned}
\underset{\mathbf{p},\,q}{\text{maximize}} \quad & \sum_{k\in\mathcal{U}_D} (\lambda_k(t)-\mu_k(t)) N_{\max}^{(k)} \frac{W}{M_D} \log_2\left(1+\frac{M_D p_k(\mathbf{h}) h_k}{N_{\max}^{(k)} c_V\,(q+\sigma^2)}\right) \qquad (6.51)\\
\text{subject to} \quad & C4: q \le \left(\min_{k\in\mathcal{U}_V} P_T^{(k)} h_k\right) \frac{|\mathcal{U}_V|}{c_V-1} - \sigma^2\\
& C5: p_k \le P_T^{(k)}, \quad \forall k \in \mathcal{U}_D,\\
& C6: q \le \sum_{\ell\in\mathcal{U}_D} P_T^{(\ell)} h_\ell, \quad \forall k \in \mathcal{U}_D,\\
& C7: q = \sum_{\ell\in\mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell.
\end{aligned}$$

The previous optimization problem (6.51) is concave w.r.t. the set of powers $\{p_k(\mathbf{h})\}$ for a fixed value of q. This means that we can always find efficiently the optimum value of $\{p_k(\mathbf{h})\}$ for a certain value of q (see, for example, the primal-dual method explained in Appendix 6.B that was used to solve the DL problem). Unfortunately, problem (6.51) is not jointly concave in $\{p_k(\mathbf{h})\}$ and q, so there is not an efficient method to obtain both $\{p_k^\star(\mathbf{h})\}$ and $q^\star$ simultaneously. For this reason, we propose a suboptimum approach to solve problem (6.51) where we perform the optimization in two stages: we first fix q and obtain $\{p_k^\star(\mathbf{h};q)\}$, then we change the value of q within its range and solve the problem again. The final solution is the one that provides the largest sum-rate for all the values of q configured. The intuition behind the algorithm is simple. It is based on a exhaustive search approach for the value of q where the range of q has been quantized into small steps so as to provide an algorithm with finite iterations. So, if we want N iterations and the range of q is $q \in [0, Q_{\max}]$, then the step size is $\Delta = Q_{\max}/N$. The smaller the value of Δ, the better the precision. If fact, if $N \longrightarrow \infty \Longrightarrow \Delta \longrightarrow 0$, then the proposed algorithm provides the optimum solution of problem (6.51). The algorithm is presented Algorithm 6.2.

6.3.2.3 Overall Resource Allocation Algorithm for the Uplink

In this subsection, we just present the overall algorithm to solve the resource allocation for the voice and data users based on the approaches presented in previous sections. The stochastic updates are also presented. The algorithm is summarized in Algorithm 6.3.

6.3.3 Numerical Simulations

In this section we evaluate the performance of the proposed strategy. The scenario is composed of 1 BS, and 3 voice users and 6 data users. The number of available codes is $N_{\max} = 13$. All the users are mobile with a speed of 3 m/s. The instantaneous channel gain, h_k, incorporates antenna gains, Rayleigh fading with unitary gain, and a real path loss of a town in Perú known

Algorithm 6.2 Algorithm for solving resource allocation problem (6.51)

1: set number of iterations N

2: define vector $\check{\mathbf{q}} = [\check{q}_1, \ldots, \check{q}_N]$ with
$0 \leq \check{q}_i \leq \min\left\{\left(\min_{k\in\mathcal{U}_V} P_T^{(k)} h_k\right) \frac{|\mathcal{U}_V|}{c_V - 1} - \sigma^2, \sum_{\ell\in\mathcal{U}_D} P_T^{(\ell)} h_\ell\right\}$

3: compute vector $\mathbf{u} = [u^\star(\check{q}_1), \ldots, u^\star(\check{q}_N)]$ where each component is the solution of the following convex problem:

4: $u^\star(\check{q}_i) = \text{maximize}_{\mathbf{p}} \quad \sum_{k\in\mathcal{U}_D} (\lambda_k(t) - \mu_k(t)) N_{\max}^{(k)} \frac{W}{M_D} \log_2\left(1 + \frac{M_D p_k(\mathbf{h}) h_k}{N_{\max}^{(k)} c_V (\check{q}_i + \sigma^2)}\right)$
$\text{subject to} \quad p_k \leq P_T^{(k)}, \quad \forall k \in \mathcal{U}_D \text{ and } \sum_{\ell\in\mathcal{U}_D} p_\ell(\mathbf{h}) h_\ell = \check{q}_i$

5: select $\check{q}^\star = \arg\max_{\check{q}_i} \; u^\star(\check{q}_i)$

6: select the powers as the value of $\left\{0 \leq p_k(\mathbf{h}) \leq P_T^{(k)}\right\}, \forall k \in \mathcal{U}_D$ that achieve $u^\star(\check{q}^\star)$

7: **end algorithm**

as Tuta Pisco (see details in [tuc13]). The code gain of data codes $M_D = 16$ and the minimum SINR normalized by the code gain for voice users is, $\frac{\Gamma}{M_V} = -13.7$ dB which corresponds to a rate of 12.2 Kbps. The noise power is $\sigma^2 = -102$ dBm. The frame for the data users and voice users are 2 ms and 20 ms, respectively. The utility function is $U(\cdot) = \log(\cdot)$. The amount of backhaul capacity required by the 3 voice users is $\check{R}_{BH}(|\mathcal{U}_V|) = 90$ Kbps. The overhead for the data transmissions is $\xi = 1.2$. The step size for the update of the stochastic multipliers is $\epsilon = 10^{-3}$. For a more detailed description of the simulation parameters see [tuc13].

For comparison purposes, we also show the resource allocation of the PF strategy [Wan07] with an instantaneous per-user backhaul constraint, $r_k(\mathbf{h}) \leq \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}$, and an instantaneous sum constraint, $\sum_{k\in\mathcal{U}_D} r_k(\mathbf{h}) \leq \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi}$. Note that for the DL case, we considered different comparative approaches. Instead of the instantaneous sum constraint, we assumed a PF with an expected per-user constraint. The effective length of the exponential window in the PF has been set to $T_c = 500$ [Wan07].

Figure 6.9 and Figure 6.10 show the time evolution of the expected rates of the proposed stochastic scheduler and the PF scheduler. We also plot the time evolution of $s^\star(\boldsymbol{\lambda}(t))$ and the per-data user backhaul rate. The total backhaul capacity considered is 6 Mbps in Figure 6.9 and is 2 Mbps in Figure 6.10. Initially, we assume that the queues at the access network are sufficiently full so that all the bits demanded by the users are served. This makes the initial average rates violate the backhaul capacity constraint for a short period of time (see the initial transient in the figure). This is due to the stochastic approximation of the multipliers but, in any case, when the average rates converge, they fulfill all the constraints of the original problem.

As we can also see in Figure 6.9, the limitation of the rates comes from the limited resources available at the access network, i.e., the power and the codes, as the backhaul capacity is not

Algorithm 6.3 Algorithm for solving the resource allocation strategy for the UL connections

1: initialize $\lambda_k(t) \geq 0$, $\mu_k(t) \geq 0$, $\forall k \in \mathcal{U}_D$

2: set $n_k^\star(\mathbf{h}) = N_{\max}^{(k)}, \quad \forall k \in \mathcal{U}_D$

3: **Data users**

4: compute $r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))$ with $p_k^\star(\mathbf{h})$ and $n_k^\star(\mathbf{h})$ from (6.51) using Algorithm 6.2

5: update (dualized) primal variable:

6: $s^\star(\boldsymbol{\lambda}(t)) = \left((\dot{U})^{-1}\left(\sum_{k\in\mathcal{U}_D} \lambda_k(t)\right)\right)_0^{\frac{R_{BH}-\check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}}$

7: update stochastic dual variables:

8: $\lambda_k(t+1) = (\lambda(t) + \epsilon\,(s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))))_0^\infty$

9: $\mu_k(t+1) = \left(\mu(t) + \epsilon\left(r_k^\star(\mathbf{h}; \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) - \frac{R_{BH}-\check{R}_{BH}(|\mathcal{U}_V|)}{\xi|\mathcal{U}_D|}\right)\right)_0^\infty$

10: **Voice users**

11: compute $\sigma_{int}^2 = \sum_{\ell\in\mathcal{U}_D} p_\ell^\star(\mathbf{h}) h_\ell + \sigma^2$

12: compute $\check{p}_k^\star(\mathbf{h}) = \frac{\sigma_{int}^2 \Gamma}{h_k(M_V - \Gamma(|\mathcal{U}_V|-1))}, \quad \forall k \in \mathcal{U}_V$

13: $t \longleftarrow t+1$ and go to 4

14: **end algorithm**

reached. It should be also emphasized that the proposed stochastic approach provides a solution that introduces more fairness when compared with the PF approach as the average rates for the different users are quite similar. Considering now Figure 6.10, the limitations comes from backhaul as the expected rates converge to the maximum per-user backhaul capacity.

Figure 6.11 shows the total rate served to data users as a function of the backhaul capacity. Note that the stochastic approach performs slightly worse than the two PF schedulers when the system is limited by the access network and not by the backhaul network but works better than the PF with individual constraints for some backhaul capacities. However, the stochastic scheduler offers a greater fairness, so the rate for the worst case user is better for the stochastic scheduler than for the other approaches, as it is shown in Figure 6.12.

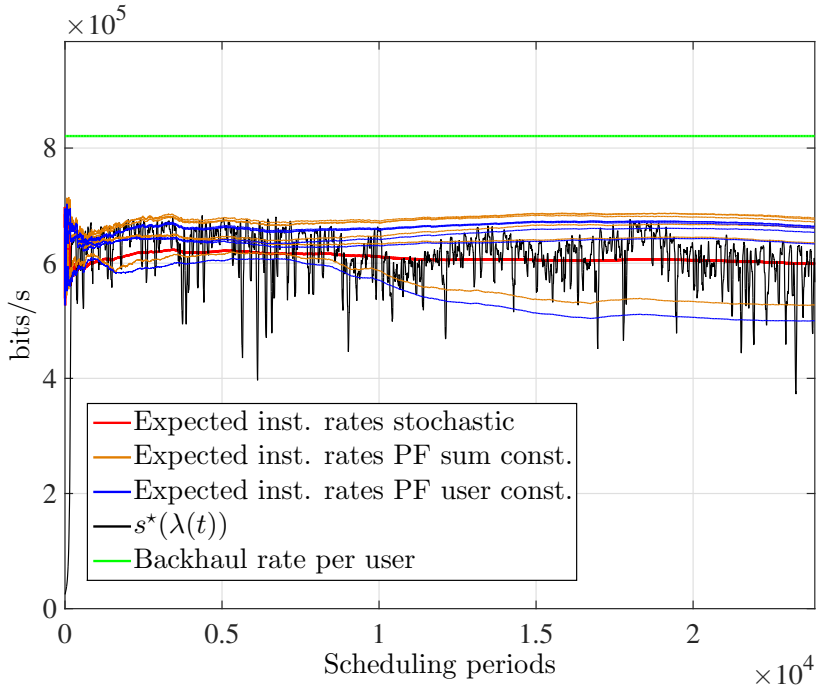


Figure 6.9: Average bit rates per data user served in the air interface by different schedulers for a total backhaul capacity of 6 Mbps.

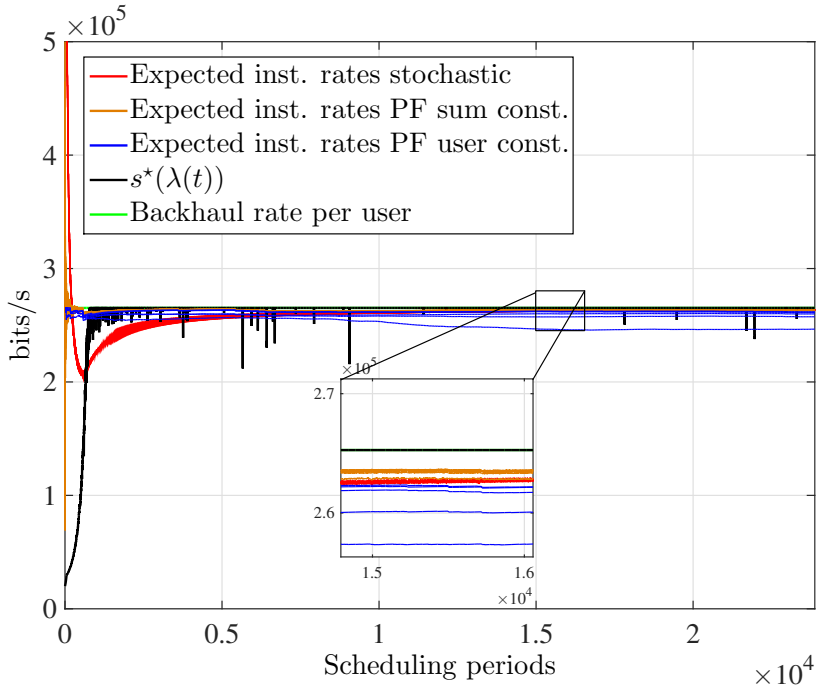


Figure 6.10: Average bit rates per data user served in the air interface by different schedulers for a total backhaul capacity of 2 Mbps.

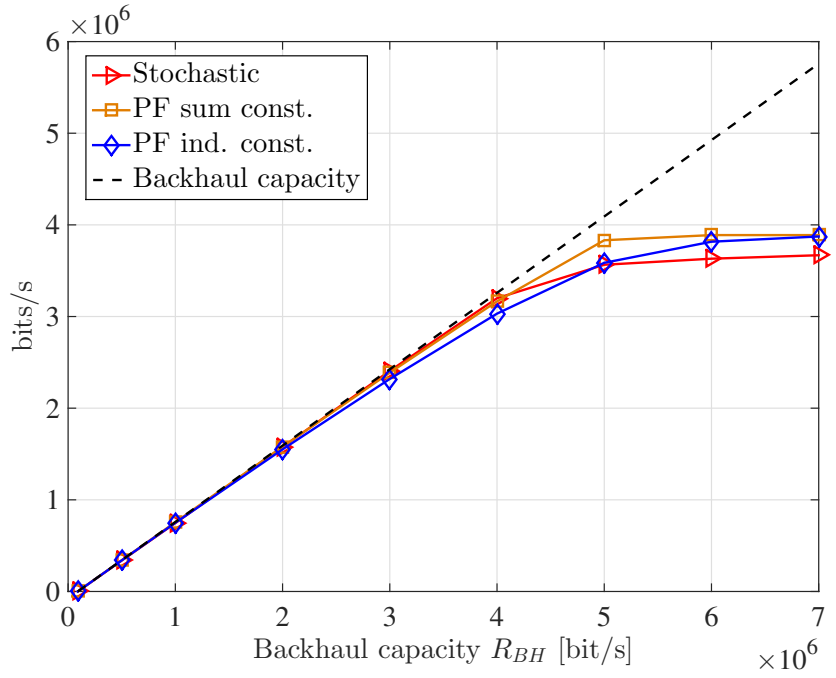


Figure 6.11: Sum-rate served in the air interface for data users versus the total backhaul capacity.

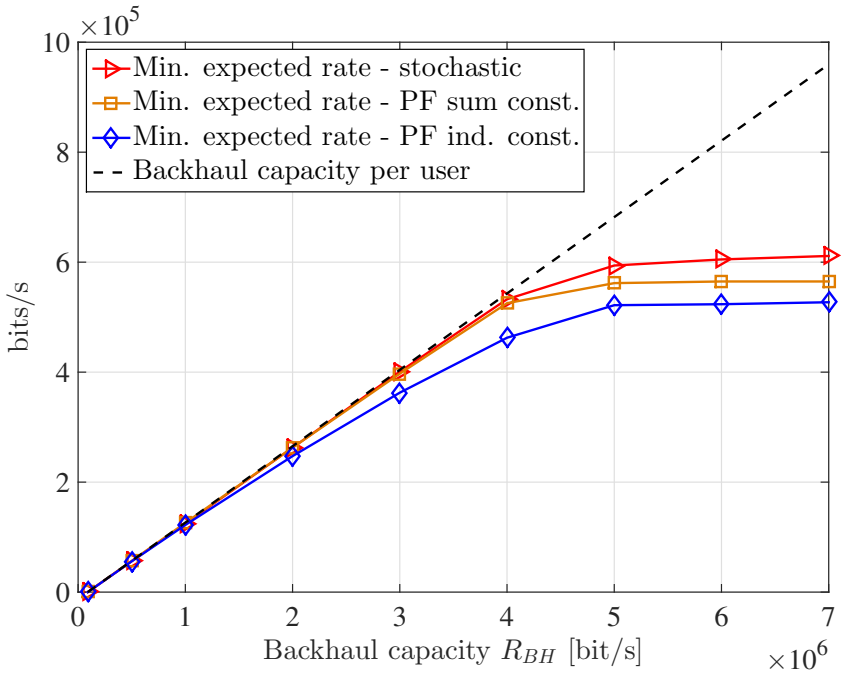


Figure 6.12: Rate served in the air interface for the worst case data user versus the total backhaul capacity.

6.4 Chapter Summary and Conclusions

In this chapter, we have proposed a resource allocation strategy for the DL and the UL setup based on the maximization of the minimum average data rate in a WCDMA system. By the use of stochastic optimization tools, we are able to deal with expected rates both in the objective

function and in the backhaul capacity constraint, allowing the access network to offer higher rates by taking advantage of good instantaneous wireless channel conditions. We assumed that there were buffers that allow to have larger instantaneous rates at the access network than the capacity rate of the backhaul, even though, in average terms, the rates achieved at the access network do not exceed that of the backhaul. This effect implies some traffic delays that could be evaluated in future works. We have also assumed that the BS is powered with a finite battery that is able to be recharged by means of an energy harvesting source. The dynamics of the energy harvesting, the energy spending, and the battery have also been taken into account explicitly in the proposed resource allocation problem. Simulations results showed that the proposed approach achieves more fairness among the users when compared to the traditional PF strategy, and provides greater worst-user rate and sum-rate if an average backhaul constraint is considered instead of an instantaneous constraint.

6.A Proof of Proposition 6.1

Let us start the proof by noticing that, in the Lagrangian $\mathcal{L}_k(s, r_k(\mathbf{h}); \mathbf{\Psi}_k(t))$, the terms involving s and $r_k(\mathbf{h})$ are decoupled. Consequently, the maximization of $\mathcal{L}_k(s, r_k(\mathbf{h}); \mathbf{\Psi}_k(t))$ in (6.27) required to obtain the value of the dual function $g_k(\mathbf{\Psi}_k(t))$ can be performed as the maximization of the separate terms w.r.t. s and $r_k(\mathbf{h})$. For simplicity, let $\tilde{R}_{BH} \triangleq \frac{R_{BH} - \check{R}_{BH}(|\mathcal{U}_V|)}{\xi |\mathcal{U}_D|}$. Therefore, $g_k(\mathbf{\Psi}_k(t))$ can be expressed as

$$
\begin{aligned}
g_k(\mathbf{\Psi}_k(t)) = & \min_{0 \le s \le \tilde{R}_{BH}} \left\{ -s + \sum_{k \in \mathcal{U}_D} \lambda_k(t) s \right\} \\
& + \min_{r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) \ge 0} \left\{ - \sum_{k \in \mathcal{U}_D} \lambda_k(t) \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) \,|\, \mathbf{\Psi}_k(t)] \right. \\
& \qquad \left. + \sum_{k \in \mathcal{U}_D} \mu_k(t) \mathbb{E}_{\mathbf{h}}[r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) \,|\, \mathbf{\Psi}_k(t)] + \sum_{k \in \mathcal{U}_D} \mu_k(t) \tilde{R}_{BH} \right\} \\
& \text{subject to} \quad C3, \ldots, C6 \text{ of problem (6.9)}
\end{aligned}
\tag{6.52}
$$

or equivalently as

$$
\begin{aligned}
g_k(\mathbf{\Psi}_k(t)) = & \min_{0 \le s \le \tilde{R}_{BH}} \left\{ -s + \sum_{k \in \mathcal{U}_D} \lambda_k(t) s \right\} + \min_{r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) \ge 0} \left\{ \mathbb{E}_{\mathbf{h}} \left[- \sum_{k \in \mathcal{U}_D} \lambda_k(t) r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) \right. \right. \\
& \left. \left. + \sum_{k \in \mathcal{U}_D} \mu_k(t) r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) + \sum_{k \in \mathcal{U}_D} \mu_k(t) \tilde{R}_{BH} \,\middle|\, \mathbf{\Psi}_k(t) \right] \right\} \\
& \text{subject to} \quad C3, \ldots, C6 \text{ of problem (6.9)}
\end{aligned}
\tag{6.53}
$$

The expected value is conditioned w.r.t. $\mathbf{\Psi}_k(t)$ because $\mathbf{\Psi}_k$ is deterministic in (6.27) but random in (6.53). Due to the linearity of the expectation operator $\mathbb{E}_{\mathbf{h}}[\cdot]$, the maximum over functions is equal to the expected value of the maximum w.r.t. the individual function values. Thanks to this, we can rewrite (6.53) as

$$
\begin{aligned}
g_k(\mathbf{\Psi}_k(t)) = & \min_{0 \le s \le \tilde{R}_{BH}} \left\{ -s + \sum_{k \in \mathcal{U}_D} \lambda_k(t) s \right\} + \mathbb{E}_{\mathbf{h}} \left[\min_{r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) \ge 0} \left\{ - \sum_{k \in \mathcal{U}_D} \lambda_k(t) r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) \right. \right. \\
& \left. \left. + \sum_{k \in \mathcal{U}_D} \mu_k(t) r_k(\mathbf{h}; \mathbf{\Psi}_k(t)) + \sum_{k \in \mathcal{U}_D} \mu_k(t) \tilde{R}_{BH} \right\} \,\middle|\, \mathbf{\Psi}_k(t) \right] \\
& \text{subject to} \quad C3, \ldots, C6 \text{ of problem (6.9)}
\end{aligned}
\tag{6.54}
$$

Note that the minimization over s and $r_k(\mathbf{h}; \mathbf{\Psi}_k(t))$ in (6.54) coincide with the primal iteration minimizations in (6.15) and (6.20). Therefore, $s^\star(\boldsymbol{\lambda}(t))$ and $r_k^\star(\mathbf{h}; \mathbf{\Psi}_k(t))$ obtained from (6.15)

and (6.20) maximize the right hand side of (6.54) and, thus, (6.54) can be expressed as

$$g_k(\boldsymbol{\Psi}_k(t)) = -s^\star(\boldsymbol{\lambda}(t)) + \sum_{k\in\mathcal{U}_D} \lambda_k(t) s^\star(\boldsymbol{\lambda}(t)) + \mathbb{E}_{\mathbf{h}}\left[- \sum_{k\in\mathcal{U}_D} \lambda_k(t) r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t)) + \sum_{k\in\mathcal{U}_D} \mu_k(t) r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t)) + \sum_{k\in\mathcal{U}_D} \mu_k(t)\tilde{R}_{BH} \,\middle|\, \boldsymbol{\Psi}_k(t)\right] \tag{6.55}$$

Because $s^\star(\boldsymbol{\lambda}(t))$ is a deterministic function of $\boldsymbol{\Psi}_k(t)$, it follows that $s^\star(\boldsymbol{\lambda}(t)) = \mathbb{E}_{\mathbf{h}}[s^\star(\boldsymbol{\lambda}(t)) \,|\, \boldsymbol{\Psi}_k(t)]$. Using the previous fact and rearranging the terms in (6.55), we can obtain

$$g_k(\boldsymbol{\Psi}_k(t)) = -s^\star(\boldsymbol{\lambda}(t)) + \sum_{k\in\mathcal{U}_D} \lambda_k(t)\mathbb{E}_{\mathbf{h}}\left[s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t)) \,|\, \boldsymbol{\Psi}_k(t)\right] + \sum_{k\in\mathcal{U}_D} \mu_k(t)\mathbb{E}_{\mathbf{h}}\left[r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t)) - \tilde{R}_{BH} \,|\, \boldsymbol{\Psi}_k(t)\right] \tag{6.56}$$

Note that the definitions of stochastic subgradients in (6.25) and (6.26) coincide with the components inside the expectations in (6.56). It then follows

$$g_k(\boldsymbol{\Psi}_k(t)) = -s^\star(\boldsymbol{\lambda}(t)) + \mathbb{E}_{\mathbf{h}}\left[\mathbf{t}_k^T(t) \,|\, \boldsymbol{\Psi}_k(t)\right] \boldsymbol{\Psi}_k(t). \tag{6.57}$$

Consider now an arbitrary dual variable $\tilde{\boldsymbol{\Psi}}_k \geq \mathbf{0}$. By following the same steps in (6.52)-(6.54) but considering $\tilde{\boldsymbol{\Psi}}_k$, we can obtain the corresponding value of the dual function $g_k(\tilde{\boldsymbol{\Psi}}_k)$ given by the minimum of the Lagrangian $\mathcal{L}_k(s, r_k(\mathbf{h}); \tilde{\boldsymbol{\Psi}}_k)$ as

$$g_k(\tilde{\boldsymbol{\Psi}}_k) = \min_{0\leq s\leq \tilde{R}_{BH}} \left\{ -s + \sum_{k\in\mathcal{U}_D} \lambda_k s \right\} + \mathbb{E}_{\mathbf{h}}\left[\min_{r_k(\mathbf{h};\tilde{\boldsymbol{\Psi}}_k)\geq 0} \left\{ - \sum_{k\in\mathcal{U}_D} \lambda_k r_k(\mathbf{h};\tilde{\boldsymbol{\Psi}}_k) + \sum_{k\in\mathcal{U}_D} \mu_k r_k(\mathbf{h};\tilde{\boldsymbol{\Psi}}_k) + \sum_{k\in\mathcal{U}_D} \mu_k \tilde{R}_{BH} \right\} \,\middle|\, \boldsymbol{\Psi}_k(t)\right] \tag{6.58}$$

$$\text{subject to} \quad C3, \ldots, C6 \text{ of problem (6.9)}$$

where conditioning on $\boldsymbol{\Psi}_k(t)$ is irrelevant because all variables are independent of $\boldsymbol{\Psi}_k(t)$. Since the expression in (6.58) involves minimizations w.r.t. s and $r_k(\mathbf{h};\tilde{\boldsymbol{\Psi}}_k)$, an upper bound of $g_k(\tilde{\boldsymbol{\Psi}}_k)$ can be obtained by evaluating the minimands at $s^\star(\boldsymbol{\lambda}(t))$ and $r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t))$. Therefore,

$$g_k(\tilde{\boldsymbol{\Psi}}_k) \leq -s^\star(\boldsymbol{\lambda}(t)) + \sum_{k\in\mathcal{U}_D} \lambda_k s^\star(\boldsymbol{\lambda}(t)) + \mathbb{E}_{\mathbf{h}}\left[- \sum_{k\in\mathcal{U}_D} \lambda_k r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t)) + \sum_{k\in\mathcal{U}_D} \mu_k r_k^\star(\mathbf{h};\boldsymbol{\Psi}_k(t)) + \sum_{k\in\mathcal{U}_D} \mu_k \tilde{R}_{BH} \,\middle|\, \boldsymbol{\Psi}_k(t)\right]. \tag{6.59}$$

By reordering terms as we did in (6.56), we can rewrite the upper bound in (6.59) as

$$\begin{aligned} g_k(\tilde{\boldsymbol{\Psi}}_k) \leq \; & -s^\star(\boldsymbol{\lambda}(t)) + \sum_{k\in\mathcal{U}_D} \lambda_k \mathbb{E}_{\mathbf{h}} \left[s^\star(\boldsymbol{\lambda}(t)) - r_k^\star(\mathbf{h}; \boldsymbol{\Psi}_k(t)) \,|\, \boldsymbol{\Psi}_k(t) \right] \\ & + \sum_{k\in\mathcal{U}_D} \mu_k \mathbb{E}_{\mathbf{h}} \left[r_k^\star(\mathbf{h}; \boldsymbol{\Psi}_k(t)) - \tilde{R}_{BH} \,|\, \boldsymbol{\Psi}_k(t) \right]. \end{aligned} \tag{6.60}$$

Using the definitions of the stochastic subgradient as we followed going from (6.56) to (6.57), we end up with

$$g_k(\tilde{\boldsymbol{\Psi}}_k) \leq -s^\star(\boldsymbol{\lambda}(t)) + \mathbb{E}_{\mathbf{h}} \left[\mathbf{t}_k^T(t) \,|\, \boldsymbol{\Psi}_k(t) \right] \tilde{\boldsymbol{\Psi}}_k. \tag{6.61}$$

By subtracting (6.57) from (6.61) yields (6.29). Finally, (6.30) is a particular case of (6.29) with $\tilde{\boldsymbol{\Psi}}_k = \boldsymbol{\Psi}_k^\star$ and $g_k(\tilde{\boldsymbol{\Psi}}_k) = g_k(\boldsymbol{\Psi}_k^\star) = D_k$.

6.B Description of the Algorithm to Solve Problem (6.31)

In this appendix, we present the technique to solve (6.31) assuming that the Lagrange multipliers are known (therefore, we omit the explicit dependence of the optimization variables w.r.t. the stochastic Lagrange multipliers, $\boldsymbol{\lambda}(t)$, $\boldsymbol{\mu}(t)$, due to simplicity in the notation). Let β and φ be the Lagrange multipliers associated to constraints $C4$ and $C5$. There is no need to dualize constraint $C6$ because the solution will turn out to automatically satisfy it. The Lagrangian of problem (6.31) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{p}, \mathbf{n}; \beta, \varphi) = & \\ & - \sum_{k\in\mathcal{U}_D} (\lambda_k(t) - \mu_k(t)) n_k \frac{W}{M_D} \log_2 \left(1 + \frac{M_D p_k h_k}{n_k(\theta P_{\mathrm{RAD}}^\star(t) h_k + \sigma^2)} \right) \\ & + \beta \left(\sum_{k\in\mathcal{U}_D} p_k - \left(\frac{\phi(C(t))}{T_f} - \sum_{k\in\mathcal{U}_V} \check{p}_k \right) \right) + \varphi \left(\sum_{k\in\mathcal{U}_D} n_k - N_{\max} \right). \end{aligned} \tag{6.62}$$

For given Lagrange multipliers, β and φ, we need to minimize the Lagrangian w.r.t. the primal variables. As it will be shown next, the structure of $\mathcal{L}(\mathbf{p}, \mathbf{n}; \beta, \varphi)$ allows the minimization w.r.t. $\mathbf{p}$ and $\mathbf{n}$ to be found in closed-form. Because $\mathcal{L}(\mathbf{p}, \mathbf{n}; \beta, \varphi)$ is strictly convex and differentiable w.r.t. $\mathbf{p}$ and $\mathbf{n}$, minimization w.r.t. these variables requires to equating the corresponding partial derivatives of $\mathcal{L}(\mathbf{p}, \mathbf{n}; \beta, \varphi)$ to zero. Differentiating the Lagrangian w.r.t. the data powers, equating the derivative to zero and solving such expression for the data powers yields

$$p_k^\star(\mathbf{n}, \beta) = \left(\frac{(\lambda_k(t) - \mu_k(t)) n_k W}{\ln(2) \beta M_D} - \frac{n_k \left(\theta P_{\mathrm{RAD}}^\star(t) h_k + \sigma^2 \right)}{M_D h_k} \right)_0^\infty, \tag{6.63}$$

where the projection on the nonnegative orthant guarantees that constraint $C6$ is fulfilled. Proceeding similar with the optimum code allocation, we set the partial derivative of the Lagrangian

w.r.t. n_k, equate it to zero and solve such equation for the codes, yielding:

$$n_k^\star(\mathbf{p},\mathbf{n},\varphi) = \left(\frac{(\lambda_k(t)-\mu_k(t))WM_D p_k h_k \left((\theta P_{\mathrm{RAD}}^\star(t) h_k + \sigma^2)\ln(2)\right)^{-1}}{(\lambda_k(t)-\mu_k(t))W\log\left(1+\frac{M_D p_k h_k}{n_k(\theta P_{\mathrm{RAD}}^\star(t)h_k+\sigma^2)}\right) - M_D\varphi} - \frac{M_D p_k h_k}{\theta P_{\mathrm{RAD}}^\star(t)h_k+\sigma^2} \right)_0^\infty \tag{6.64}$$

where, also in this case, the projection on the nonnegative orthant guarantees that constraint $C6$ is fulfilled. Notice that a fixed-point iteration to compute the optimum code allocation, $n_k^\star(\varphi)$, can be used in this case.

Having obtained the optimum primal variables as a function of the Lagrange multipliers, we now seek to find the optimum Lagrange multipliers to obtain the global optimum primal variables. The approach we propose to find the optimum multipliers is based on the maximization of the dual function $D(\beta,\varphi)$ [Boy04], which is defined as the minimization of the Lagrangian w.r.t. the primal variables, i.e., $D(\beta,\varphi) \triangleq \inf_{\mathbf{p},\mathbf{n}} \mathcal{L}(\mathbf{p},\mathbf{n};\beta,\varphi) \equiv \mathcal{L}(\{p_k^\star(\mathbf{n};\beta)\},\{n_k^\star(\mathbf{p},\mathbf{n},\varphi)\},\beta,\varphi)$. Then, the the multipliers are obtained by solving the dual problem as

$$\begin{aligned} \underset{\beta,\varphi}{\text{maximize}} \quad & D(\beta,\varphi) \\ \text{subject to} \quad & \beta \geq 0, \quad \varphi \geq 0. \end{aligned} \tag{6.65}$$

Recall that the dual problem is always convex w.r.t. the dual variables and, thus, can be efficiently solved with a projected gradient method (if $D(\beta,\varphi)$ is differentiable) or a projected supergradient method if it is not differentiable [Ber99]. A valid supergradient for each particular dual variable is given by the constraint it is associated with [Ber99]. The update equations are given by

$$\beta^{(q+1)} = \left(\beta^{(q)} + \nu^{(q)} \left(\sum_{k\in\mathcal{U}_D} p_k^{(q)} - \left(\frac{\phi(C(t))}{T_f} - \sum_{k\in\mathcal{U}_V} \check{p}_k \right) \right) \right)_0^\infty, \tag{6.66}$$

$$\varphi^{(q+1)} = \left(\varphi^{(q)} + \nu^{(q)} \left(\sum_{k\in\mathcal{U}_D} n_k^{(q)} - N_{\max} \right) \right)_0^\infty, \tag{6.67}$$

where q indicates the iteration and the step size defined as $\nu^{(q)} = \frac{Q}{\sqrt{q}}\left(\|\nabla D\|_2\right)^{-1}$, being ∇D the overall supergradient of the dual function, is chosen such that the diminishing conditions are fulfilled, i.e., $\lim_{k\to\infty}\nu^{(q)} = 0$, $\sum_{k\in\mathcal{U}_D}\nu^{(q)} = \infty$ [Ber99]. Once we know the optimal dual variables, $\beta^\star$ and $\varphi^\star$, we can obtain the optimum power and code allocations, $p_k^\star(\beta^\star)$ and $n_k^\star(\varphi^\star)$. The proposed iterative algorithm is based on the primal-dual block coordinate descent method for the update of the primal variables p_k and n_k [Ber99].

Chapter 7

User Association for Load Balancing in Heterogeneous Networks Powered with Energy Harvesting Sources

7.1 Introduction

Multi-tier heterogeneous networks (HetNets) [Dam97] have emerged as a potential solution to increase the system capacity and coverage to match the surging traffic demands [Dam11], [Wu04]. In particular, there is a strong tendency to consider the deployment of small BSs such as, for example, picocells and femtocells, along with already deployed macrocells. Each of these different types of BSs differ substantially in terms of maximum transmit power, physical size, coverage areas, and cost, among others. HetNets enable a more flexible, easy to install, and economical deployment as new infrastructure can be deployed when necessary, thus reducing the deployment cost.

In this chapter, we are targeting scenarios where only energy harvesting devices, e.g. solar panels, will empower the BSs. An example of a potential scenario is a rural deployment, where access to the electric power grid is too expensive or impossible. In this framework, the European project TUCAN3G (http://www.ict-tucan3g.eu) aims at contributing to this task. This project studies, from both the technological and socio-economical perspectives, the progressive introduction of cellular services in remote rural areas. In particular, remote locations in Perú have been considered. In such locations, some main challenges arise: cost of BSs, cost of solar panels and batteries, cost of terminals, and business models adapted to people with low incomes. The solution adopted in TUCAN3G consists of an access network based on 3G BSs empowered by solar panels of limited size in outdoor scenarios.

Thanks to having small cells deployed within the coverage area of macrocells, efficient traffic balancing techniques can be implemented by dynamic association of some of the users to the small cells which will help to improve the overall network throughput. The well-known user

association considered in 3GPP standard is based on received maximum signal to interference and noise ratio (SINR). Nevertheless, such approach does not necessarily provide a good network load balancing measured in throughput as users tend to connect to the strongest BS (usually the macrocell) and, thus, drives it to a heavy loaded situation. As a consequence, a strategy that manages such user associations for load balancing is needed. More importantly, if the BSs are only powered with batteries, then the battery status as well as the harvesting capabilities should also be considered explicitly in the association strategy; otherwise the best BSs (those providing larger data rates) will run out of energy very quickly.

There are a few works in the literature dealing with the concept of user association for load balancing. However, most of them only consider macrocell networks (see for example [Son09], [San08], [Das97], [Yan04]) and only the most recent papers focus on HetNets scenarios. Concerning solutions for HetNets, authors in [Mad10] propose a solution for cell association managing the interference being generated among BSs in a LTE-like setup scenario. They propose simple heuristic techniques that show a significant improvement in the system performance. In [She13], authors propose a DL cell association based on dual coordinate descent method to achieve balancing and they adopt the maximization of the network utility following a proportional fairness objective. However, they assume that the power to be transmitted at the BS is fixed and cannot be optimized. User association for load balancing in HetNets is also addressed in [Ye13b]. In that paper, the authors propose a load measure based on long-term service data rate. Unlike the previous papers, there are a few works that present joint resource allocation and cell selection mechanisms. For example, in [Lee06] the joint resource allocation and BS assignment design in CDMA networks is considered. The authors develop a pricing-based distributed algorithm that considers congestion level of the BSs as well as the channel states of the mobile terminals. Authors in [Hon12] study the joint BS association and resource allocation in DL OFDMA networks assuming that both procedures are executed simultaneously at the same time. Nevertheless, we think the user association should be run at a different time scale than the scheduling as the mobility of the users is slower than the changes the wireless channels experience.

Regarding user association strategies with BSs powered by energy harvesting sources, only a few works are available in the literature. In [Dan14], authors propose a strategy that aims to maximize the number of accepted users and minimize the radio resource consumption where the available energy of BSs depends on the harvested energy in a certain period of time. A lexicographic minimization of on-grid energy consumption is proposed in [Dan15], which involves the optimization in both the space and time dimensions, due to the temporal and spatial dynamics of mobile traffic and green energy generation. Finally, the authors in [Zha15] propose a user association based on the concept of topology potential which takes the traffic load of users and the available renewable energy of BSs into consideration for energy-load tradeoffs.

In this chapter, we introduce several user association strategies that perform load balancing (based on user connections to BSs) among the different network tiers and that consider explicitly

the battery status of the BSs as well as the energy that is being harvested. We also consider that the coverage areas depend explicitly on the available energies at the batteries. In the first part of the chapter, we propose some approaches based on the greedy strategy followed in previous chapters of this dissertation in which the overall radiated power is limited by the current energy available at the battery. In the second part of this chapter, we propose a stochastic approach in which we estimate somehow the future impact of the current energy usage by using ergodic optimization techniques [Rib10b].

The organization of the rest of the chapter is as follows. Section 7.2 introduces the description of the system model and the assumptions considered. Then, in Section 7.3 we develop the user association strategy based on the greedy approach. We propose some low complexity user association solutions that are required in environments with high mobility and then we present a decentralized strategy to be executed among the users and the BSs with just local information. Later, in Section 7.4 we develop the user association strategy based on the ergodic optimization theory. Simulations results that evaluate the performance of the proposed approaches are presented in Section 7.5 and, finally, conclusions are drawn in Section 7.6.

7.2 System Model and Problem Formulation

7.2.1 System Description

Let us consider a DL cellular multi-tier SISO system composed of several BSs. Each of these BSs belongs to a particular BS class, each having different capabilities (transmission power, battery size, etc.), where each class is categorized as a tier. Each BS is indexed by i and the set of all BSs is denoted as $\mathcal{B}$. We consider that there are two types of users: users that demand a fixed service rate and users that request a flexible service rate. This could correspond to a cellular system comprised of voice users[1], that demand a fixed service rate, and data users, who are usually provided with a dynamic flexible rate depending on the system load and the channel conditions. Let us denote the set of voice users as $\mathcal{U}_V$ and the set of data users as $\mathcal{U}_D$. $\mathcal{U}_V^{(i)}$ and $\mathcal{U}_D^{(i)}$ represent the set of voice and data users that are associated to BS i, respectively. The set that contains all users in the system is denoted as $\mathcal{U}_T = \mathcal{U}_V \cup \mathcal{U}_D$ and the set of all users connected to BS i as $\mathcal{U}_T^{(i)} = \mathcal{U}_V^{(i)} \cup \mathcal{U}_D^{(i)}$.

We assume a WCDMA network [Gol05], [Lee06], which implies that users are multiplexed using codes. This assumption is required since the work developed in this chapter is part of the research carried out in the European project TUCAN3G (already presented in the previous chapters, http://www.ict-tucan3g.eu) in which a 3G network for rural areas must be sized and deployed. However, the methodology presented in this chapter could be easily extended to

[1]From now on, we will assume that the users that request a fixed service rate are voice users, but this set could indeed contain any type of service requiring fixed service rates.

consider 4G systems by just considering the specific rate-power function and multiplexing users with carriers instead of codes. In any case, we assume that the network operator has already reserved a set of codes for the voice users and the remaining codes are to be allocated among the data users. Thus, the amount of available codes in each set is known and fixed at each BS, a strategy followed commonly by network operators [Lai06].

Each BS is powered solely with its own harvesting source and battery. The energy harvesting sources allow the BSs to collect energy from the environment and recharge the batteries. We consider that, at a given moment of time, only information of the past and current harvesting and battery dynamics of the BSs will be available to perform the user association, yielding to an *online approach*.

Let us denote the time instants (epochs) between association periods by index $\tau \in \Upsilon$ and the duration of such epochs as T_{e}. Let us also denote the scheduling periods by $t \in \mathcal{T}$ with a duration of T_f. In general, the user association procedure should be executed in a time scale longer than the scheduling periods. This makes the association problem very challenging since the decision has to be taken considering, amongst others, the current CSI, but the channel may vary during the whole association period. As a consequence, the association decision should be implemented considering the expected throughput (over the channel realizations) within the duration of the association period.

We will assume that the allocation strategy is executed either by a central node of the network that is able to collect all the necessary information (channel information, state of the batteries, etc.) of all users and BSs or in a distributed fashion in which users and BSs individually execute a procedure with only local information, (i.e., BSs have information only about the channels from users that are within their coverage area and its own battery information). In the centralized architecture, after performing the association decision, this central node broadcasts the optimum decision to the BSs and each BS is responsible for informing the final set of users they are going to serve. Figure 7.1 represents all the interconnections and message passing information needed between the BSs and the coordinating entity. Some of the nomenclature presented in the figure will be apparent later as it appears in the notation. Nevertheless, the cell range expansion (CRE) and on/off procedures are neither presented nor evaluated in this chapter. The objective of this functionality would be to optimize the coverage area of all BSs, i.e., their pilot power, based on the current system load (denoted as $L_i(m)$) and the current battery levels of the BSs. The scheduling procedure will not be explicitly described in this chapter, but we will assume a particular scheduler that will be used to evaluate the performance of the system in the simulations section.

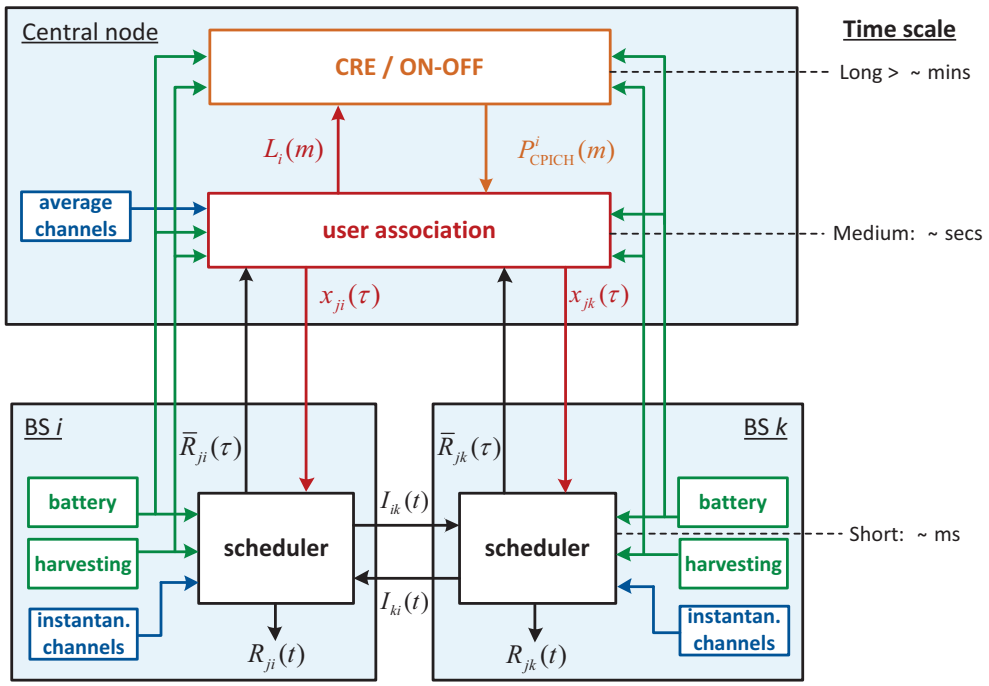


Figure 7.1: Centralized architecture consisting of resource allocation, user association, and CRE.

7.2.2 Power Consumption Model and Battery Dynamics

The power consumption at any BS is modeled as the addition of the radiated power, which is divided into the power devoted to the common DL channels, P^i_{CPICH}, the power consumed by the user-specific physical channels, $P^i_{\text{BS}}(t)$, and a fixed power consumed by the electronics and cooling, P^i_c. Let $P^i_{\text{RAD}}(t) = P^i_{\text{BS}}(t) + P^i_{\text{CPICH}}$ be the overall radiated power by the i-th BS.

We consider that the amount of power that can be used for traffic services is limited in each BS and is denoted by $P^{\max}_{\text{BS}_i}$, so $P^i_{\text{BS}}(t) \leq P^{\max}_{\text{BS}_i}\ \forall i$. Let $C_i(t)$ be the energy stored at the battery of the i-th BS at the beginning of scheduling period t and let $E_i(t)$ be the overall energy consumed during scheduling period t:

$$E_i(t) = T_f \cdot \left(P^i_{\text{CPICH}} + P^i_{\text{BS}}(t) + P^i_c\right). \tag{7.1}$$

Then at period $t+1$, the battery level is updated in general as [Ho12a]

$$C_i(t+1) = \left(C_i(t) - E_i(t) + H_i(t)\right)_0^{C^i_{\max}}, \quad t \in \mathcal{T}, \tag{7.2}$$

where $H_i(t)$ is the energy harvested in Joules during the whole scheduling period t, the projection of x onto the interval $[a, b]$ accounts for possible battery overflows and assures that the battery levels are non-negative, and $C^i_{\max}$ is the battery capacity. Notice that the whole harvesting collected during period t is assumed to be available in the battery at the end of the epoch for simplicity.

7.2.3 Energy Harvesting Model

We assume a discretized model for the energy arrivals [Yan12a], [Ho12a], [Gre13b] where $H_i(\tau)$ is modeled as an ergodic Bernoulli process. As a result, only two values of harvested energy are possible, i.e., $H_i(\tau) \in \{0, e_i\}$, where e_i is the amount of Joules contained in an energy packet. The probability of receiving an energy harvesting packet during one epoch depends on the actual harvesting intensity (in the case of solar energy, it depends on the particular hour of the day) and is denoted by $p_i(\tau)$. Note that a higher value of $p_i(\tau)$ will be obtained in epochs where the harvesting intensity is higher, e.g., during the day, and a lower value of $p_i(\tau)$ during the night or during cloudy days. Note that, in some cases, it is possible to predict or estimate partially the expected energy available to be harvested. This situation can be included in our formulation by adjusting our model parameters by knowing that $\mathbb{E}[H_i(\tau)] = p_i(\tau) \cdot e_i$. As a result, the majority of harvesting sources can be modeled with the previous stochastic model.

7.2.4 System Definitions

Let us define a set of indicator variables $x_{ji}(\tau) \in \{0, 1\}$ to denote whether a given user j is associated to a particular BS i as follows

$$x_{ji}(\tau) = \begin{cases} 1, & \text{if user } j \in \mathcal{U}_T^{(i)}, \\ 0, & \text{otherwise.} \end{cases} \tag{7.3}$$

Let $\mathbf{x}(\tau) \triangleq \{x_{ji}(\tau),\ j \in \mathcal{U}_T, i \in \mathcal{B}\}$ be the set containing the indicator variables. We consider that a user can only be connected to one BS at a given epoch. This is formulated through the following unique association constraint:

$$\sum_{i \in \mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j \in \mathcal{U}_T. \tag{7.4}$$

In order to be able to connect to a given BS, users need to receive its pilot signals above a minimum SINR threshold to estimate the channel. Thus, a particular user j will only be able to connect to the following set of BSs:

$$\mathcal{S}_j(t) = \left\{ i \in \mathcal{B} \;\middle|\; \frac{P^i_{\text{CPICH}}}{P^i_{\text{BS}}(t) + A_{ji}(t)} \geq \gamma_{\text{CPICH}} \right\} \subseteq \mathcal{B}, \tag{7.5}$$

where $A_{ji}(t) = \frac{\sigma_j^2 + I_{ji}(t)}{\tilde{h}_{ji}(t)}$, being σ_j^2 the noise power at j-th receiver, $I_{ji}(t)$ the inter-cell interference, i.e., $I_{ji}(t) = \sum_{m \in \mathcal{B}, m \neq i} \tilde{h}_{jm}(t) P^m_{\text{RAD}}(t)$, and $\tilde{h}_{ji}(t)$ the instantaneous channel between BS

i and user j. Given that, from (7.5) we obtain the following association constraints:

$$x_{ji}(\tau) = 0, \quad \forall i \notin \mathcal{S}_j(t). \tag{7.6}$$

The traffic power, $P^i_{\mathrm{BS}}(t)$ can be split into power for voice and data connections as $P^i_{\mathrm{BS}}(t) = \sum_{j\in\mathcal{U}_V^{(i)}} \tilde{\tilde{p}}_{ji}(t) + \sum_{k\in\mathcal{U}_D^{(i)}} \tilde{p}_{ki}(t)$, where $\tilde{\tilde{p}}_{ji}(t)$ and $\tilde{p}_{ki}(t)$ are the instantaneous powers corresponding to the transmission toward the j-th and k-th voice and data user connected to the i-th BS, respectively. The set of voice users request a fixed rate service and we assume that just one code is assigned to each of them. This is translated into a minimum SINR requirement (γ_j) as follows:

$$\sum_{i\in\mathcal{B}} \frac{x_{ji}(\tau)\tilde{\tilde{p}}_{ji}(t) M_V}{\theta\left(P^i_{\mathrm{RAD}}(t) - \tilde{\tilde{p}}_{ji}(t)\right) + A_{ji}(t)} \geq \gamma_j, \quad \forall j \in \mathcal{U}_V, \tag{7.7}$$

where M_V is the spreading factor for voice codes and θ is the orthogonality factor [Gol05]. The set of data users request a flexible rate service. As it was commented before, the user association procedure should be executed in a longer time scale compared to the coherence time of the channel. For this reason, the user association strategy should not take decisions based on the instantaneous channels, as they will vary over the entire epoch. As a consequence, we propose to take decisions based on the expected value of the throughputs (through the application of the rate function based on expected channels as will be shown next). Hence, the expected throughput achieved during one particular scheduling period by the i-th user connected to the j-th BS is

$$R_{ji}(t) = \mathbb{E}_{\tilde{\mathbf{h}}}\left[\mathbb{E}_{\tilde{\mathbf{p}},\tilde{\mathbf{n}}|\tilde{\mathbf{h}}}\left[\tilde{n}_{ji}(t)\log_2\left(1 + \frac{M_D\tilde{p}_{ji}(t)}{\tilde{n}_{ji}(t)(\theta\left(P^i_{\mathrm{RAD}}(t) - \tilde{p}_{ji}(t)\right) + A_{ji}(t))}\right)\right]\right], \tag{7.8}$$

where M_D is the spreading factor for data codes, $\tilde{\mathbf{p}} \triangleq \{\tilde{p}_{ji}\ j \in \mathcal{U}_D, i \in \mathcal{B}\}$ and $\tilde{\mathbf{n}} \triangleq \{\tilde{n}_{ji}\ j \in \mathcal{U}_D, i \in \mathcal{B}\}$ are the power and code allocation variables (random variables), and where we have assumed that the same power is allocated to all the codes of the same user. Note that the instantaneous allocated powers and codes depend upon the concrete scheduling policy and, generally, they do not only depend on the instantaneous channels, but also on other factors (e.g. queue states). That is why we model $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{n}}$ as random variables, and average w.r.t. them have to be applied in (7.8).

We use the power and code variables to measure the load that a given BS is experiencing so that a well load-balanced network can be obtained by a proper user association strategy.

7.2.5 Problem Formulation

In systems where variables with temporal evolution affect the long-term system performance, (such as, for example, the energy available in the battery that constrains the amount of power

that can be allocated among the users) it is important to take decisions considering not only the current impact on the overall system performance, but also taking into account the future impact. It is evident that if we spend much power in the current epoch and the conditions of the wireless channels are poor, we may be missing the opportunity to use better this energy in the near future when the channel conditions improve. In this sense, the user association procedure should be coupled in time, and associations that are being carried out at present time should consider the past knowledge as well as the future impact. For example, if a given BS is harvesting a lot of energy, this BS should accept more users as more energy will be available in the future. If, on the other hand, the energy is decaying fast, then an intelligent strategy would be to allocate such users to the neighboring BSs gracefully (an aggressive strategy would not be smart either since if there was a massive reallocation of users, the neighboring BSs would need a lot of energy to satisfy the users demands).

Another point to consider is the fact that the association strategy should run in real time. This implies that the association strategy must be implemented considering only current information of the channels, the batteries, and the harvesting being collected. Contrary, offline solutions that require future knowledge, i.e., non-causal knowledge of the channel, the batteries, the harvesting, etc., have been proposed extensively in situations where network nodes are provided with energy harvesting sources [Yan12a], [Oze11]. However, the assumption of knowing when the energy will be available in future instant times is usually hard to accept. Therefore, in this chapter we focus only on online approaches that are more suitable for a realistic implementation.

A formal way to model the dynamic behavior and consider only causal information consists in the use of DP techniques [Ber05]. DP has been shown to be a good mathematical tool to solve time-coupled problems when only past and current information is available, i.e., an online approach, and we desire to have a control over how the resources used in the current time instant can affect the future performance. Unfortunately, in such DP problems, the optimum value of the variables are functionals instead of scalars as in classical optimization. In addition, the computation of such functionals is usually extremely difficult and computationally challenging. In any case, let us present now the most general user association strategy and then propose some simplifications that will make the problem tractable and solvable with a reasonable computational complexity. The system performance under consideration is the sum-utility of the overall expected throughput of all data users in the system. For the sake of generality, we consider a general utility function denoted by $U_j(\cdot)$ for the j-th user. Let us introduce the following definitions: $\tilde{\mathbf{r}} = \{\tilde{R}_j,\ j \in \mathcal{U}_D\}$, $\tilde{\mathbf{p}}(t) = \{\tilde{p}_{ji}(t),\ j \in \mathcal{U}_D, i \in \mathcal{B}\}$, $\tilde{\tilde{\mathbf{p}}}(t) = \{\tilde{\tilde{p}}_{ji}(t),\ j \in \mathcal{U}_V, i \in \mathcal{B}\}$, $\mathbf{n}(t) = \{n_{ji}(t),\ j \in \mathcal{U}_D, i \in \mathcal{B}\}$, $\mathbf{x}(\tau) = \{x_{ji}(\tau),\ j \in \mathcal{U}_T, i \in \mathcal{B}\}$, and $\mathbf{p}_{\mathrm{RAD}}(t) \triangleq \{P^i_{\mathrm{RAD}}(t),\ i \in \mathcal{B}\}$. The association procedure is formulated through the following

optimization problem:

$$\underset{\substack{\tilde{\mathbf{r}},\tilde{\mathbf{p}}(t),\tilde{\tilde{\mathbf{p}}}(t),\\ \tilde{\mathbf{n}}(t),\mathbf{x}(\tau),\\ \mathbf{p}_{\mathrm{RAD}}(t)}}{\text{maximize}} \quad \sum_{j\in\mathcal{U}_D} U_j\left(\tilde{R}_j\right) \tag{7.9}$$

subject to

$$C1: \tilde{R}_j \leq \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{B}}\mathbb{E}_{\tilde{\mathbf{h}}}\left[\mathbb{E}_{\tilde{\mathbf{p}},\tilde{\mathbf{n}}|\tilde{\mathbf{h}}}\left[\tilde{n}_{ji}(t)\log_2\left(1+\frac{M_D\tilde{p}_{ji}(t)}{\tilde{n}_{ji}(t)(\theta\left(P^i_{\mathrm{RAD}}(t)-\tilde{p}_{ji}(t)\right)+A_{ji}(t))}\right)\right]\right], \forall j$$

$$C2: \sum_{i\in\mathcal{B}}\frac{M_V\tilde{\tilde{p}}_{ji}(t)}{\theta(P^i_{\mathrm{RAD}}(t)-\tilde{\tilde{p}}_{ji}(t))+A_{ji}(t)} \geq \gamma_j, \quad \forall j\in\mathcal{U}_V, \forall t\in\mathcal{T}$$

$$C3: \sum_{i\in\mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j\in\mathcal{U}_T, \forall\tau\in\Upsilon$$

$$C4: \sum_{j\in\mathcal{U}_D}\tilde{n}_{ji}(t) \leq n_D^{(i)}, \quad \forall i\in\mathcal{B}, \forall t\in\mathcal{T}$$

$$C5: T_f\left(\sum_{j\in\mathcal{U}_D}\tilde{p}_{ji}(t)+\sum_{j\in\mathcal{U}_V}\tilde{\tilde{p}}_{ji}(t)+P^i_{\mathrm{CPICH}}+P^i_c\right) \leq C_i(t), \quad \forall i\in\mathcal{B}, \forall t\in\mathcal{T}$$

$$C6: C_i(t+1) = (C_i(t)-E_i(t)+H_i(t))_0^{C^i_{\max}}, \quad \forall i\in\mathcal{B}, \forall t\in\mathcal{T}$$

$$C7: P^i_{\mathrm{RAD}}(t) = \sum_{j\in\mathcal{U}_D}\tilde{p}_{ji}(t)+\sum_{j\in\mathcal{U}_V}\tilde{\tilde{p}}_{ji}(t)+P^i_{\mathrm{CPICH}}, \quad \forall i\in\mathcal{B}, \forall t\in\mathcal{T}$$

$$C8: 0\leq\tilde{n}_{ji}(t)\leq x_{ji}(\tau)n_D^{(i)}, \quad \forall i\in\mathcal{B}, \forall j\in\mathcal{U}_D, \forall t\in\mathcal{T}$$

$$C9: 0\leq T_f\tilde{\tilde{p}}_{ji}(t)\leq x_{ji}(\tau)C_i(t), \quad \forall i\in\mathcal{B}, \forall j\in\mathcal{U}_V, \forall t\in\mathcal{T}$$

$$C10: x_{ji}(\tau)=0, \quad \forall i\notin\mathcal{S}_j(t), \forall j\in\mathcal{U}_T, \forall\tau\in\Upsilon$$

$$C11: x_{ji}(\tau)\in\{0,1\}, \quad \forall i\in\mathcal{B}, \forall j\in\mathcal{U}_T, \forall\tau\in\Upsilon$$

$$C12: \tilde{p}_{ji}(t)\geq 0, \quad \forall i\in\mathcal{B}, \forall j\in\mathcal{U}_D, \forall t\in\mathcal{T}.$$

where $|\mathcal{T}|$ is the cardinality of set $\mathcal{T}$ and $n_D^{(i)}$ is the number of available codes for data users at BS i. Notice that we have considered a general per-user utility function $U_j(\cdot)$ as long as it is an increasing, strictly concave, and continuously differentiable function [Wan07]. The problem presented in (7.9) is extremely difficult to solve and very challenging for the following reasons: *i*) the expected rate expression that appears in constraint $C1$ is very difficult to be handled; *ii*) it is time coupled through constraints $C1$ and $C6$; *iii*) the interference term in the rate expression makes the whole expression not concave; *iv*) the coverage areas depend on the radiated powers and, so, they should also be optimized jointly, and *v*) the association variables $x_{ji}(t)$ are integer variables, making the whole problem a combinatorial optimization problem. In this regard, all this complicating issues should be handled before presenting the association strategy. Since the user association algorithm must be executed regularly in real time, the solution of the previous problem based on DP will not be considered as the computational complexity is prohibitively high. Instead, in this chapter, we develop simpler and less complex association strategies that could be closer to a real implementation. In this regard, we are going to introduce two different

ways to simplify the previous problem. As we introduced before, one of the approaches will follow the greedy strategy presented in previous chapters of this dissertation. The basic idea is to decouple the problem in time and solve for each time instant independently. In the second approach, we will present a method based on ergodic optimization in which partial time-coupling is considered. We will introduce later how to deal with the other challenging issues exposed before.

7.3 Part I: Greedy-Based User Association Strategies

In this part of the chapter we propose some user association techniques based on the greedy approach followed in previous chapters of this dissertation. The simplifications of the challenging issues presented in Section 7.2.5 that we are considering in this section are the following:

- *i*) The expected rate expression given in (7.8) is very difficult to compute as it depends on the PDF of the powers, the codes, and the channels. Instead, an approximation of the average rate is considered in this chapter following the same criteria used in papers like [Ye13b], [Ye13a], [Wei13]. In this approximation, we consider average channels instead of instantaneous ones so that the channels now change from epoch to epoch, i.e., $h_{ji}(\tau) = \mathbb{E}_{\mathbf{h}}[\tilde{h}_{ji}(t)]$, which includes shadowing, path loss, and antenna gain. Now, all the variables change at the time scale given by τ and the expectations w.r.t. the powers and the codes are also removed since these new variables, $p_{ji}(\tau)$, $\check{p}_{ji}(\tau)$, and $n_{ji}(\tau)$, can be interpreted as average allocation variables, i.e., estimates of the average values of the instantaneous powers allocated by the scheduler during one epoch. By carrying out such approximation, the expected throughput that a data user can obtain during one particular epoch τ can be approximated by

$$\bar{R}_{ji}(\tau) = n_{ji}(\tau)\log_2\left(1 + \frac{M_D p_{ji}(\tau)}{n_{ji}(\tau)(\theta(P^i_{\mathrm{RAD}}(\tau) - p_{ji}(\tau)) + A_{ji}(\tau))}\right). \tag{7.10}$$

  Note that all other variables, such as $A_{ji}(\tau)$, and equations such as the battery update in (7.2) ($C_i(\tau)$) now depend on the average channels $h_{ji}(\tau)$. Additionally, in order to guarantee that constraints in (7.5) are still fulfilled (with a certain outage probability, as they cannot be fulfilled strictly for all possible channel realizations), we could increase the threshold γ_{CPICH} to make the constraints more restrictive.

- *ii*) The most important simplification is described next. We are transforming the DP problem (7.9) into a standard optimization problem. Regarding constraint $C1$, we will not consider the summation term over all time instants. Then, instead of having constraint $C6$ that couples the battery evolution and, hence, couples all allocation variables, our on-line user association strategy is executed at each epoch independently by allowing each BS to

spend just a given fraction of the energy available at the battery during that particular epoch. The amount of energy to be spent at each epoch can be further optimized offline (see Section 3.4.4). This can be seen as a greedy policy, where at each epoch we execute the association strategy with just past and current information of the battery levels, channel states, etc.

In general, the total energy consumed by a given BS during one epoch, i.e., $E_i(t)$, is limited by a function of the current battery level as

$$T_{\mathrm{e}} \cdot \left(P^i_{\mathrm{CPICH}} + P^i_{\mathrm{BS}}(\tau) + P^i_{\mathrm{c}}\right) \leq g_i(C_i(\tau)), \quad \forall i, \tau, \tag{7.11}$$

where the function $g_i(\cdot)$ is defined as $g_i(C_i(\tau)) \triangleq \min\{T_{\mathrm{e}} \cdot \left(P^i_{\mathrm{CPICH}} + P^{\max}_{\mathrm{BS}_i} + P^i_{\mathrm{c}}\right), w_i(C_i(\tau))\}$, being $w_i(\cdot)$ a generic monotonous increasing function that fulfills that $w_i(C_i(\tau)) \leq C_i(\tau)$, $\forall i, \tau$. For example, if all the energy in the battery is allowed to be spent during one particular epoch, then $w_i(C_i(\tau)) = C_i(\tau)$. Nevertheless, the approach that we follow is to limit the amount of energy that can be used in one particular epoch in order to spend it in a more conservative way. According to this, we consider that only a given fraction of the battery level is allowed to be spent during each epoch, i.e.,

$$w_i(C_i(\tau)) = \alpha_i \cdot C_i(\tau), \quad 0 \leq \alpha_i \leq 1, \quad \forall i \in \mathcal{B}. \tag{7.12}$$

- *iii*) For simplicity in the notation, in the rate expression in (7.10) we will approximate $\theta\left(P^i_{\mathrm{RAD}}(\tau) - p_{ji}(\tau)\right) + A_{ji}(\tau)$ by $\theta P^i_{\mathrm{RAD}}(\tau) + A_{ji}(\tau)$ throughout the chapter[2]. Notice that now $\bar{R}_{ji}(\tau)$ is jointly concave in $n_{ji}(\tau)$ and $p_{ji}(\tau)$ and $\bar{R}_{ji}(\tau) = 0$ if $n_{ji}(\tau) = 0$, for any $p_{ji}(\tau) \geq 0$.

- *iv*) Note that the power variables affect both the coverage (see (7.5)) and the data rates. However, it is extremely difficult to deal with both effects jointly. For this reason, we assume a worst-case approach when defining the coverage areas assuming that the interference is maximum[3]. The worst-case coverage sets are, therefore, defined as $\mathcal{S}_j(\tau) = \left\{ i \in \mathcal{B} \;\middle|\; \frac{P^i_{\mathrm{CPICH}}}{\bar{P}^i_{\mathrm{BS}}(C_i(\tau)) + \bar{A}_{ji}(\tau)} \geq \gamma_{\mathrm{CPICH}} \right\}$, where $\bar{A}_{ji}(\tau) = \frac{\sigma^2_j + \sum_{m \in \mathcal{B}, m \neq i} h_{jm}(\tau)(\bar{P}^m_{\mathrm{BS}}(B_m(\tau)) + P^m_{\mathrm{CPICH}})}{h_{ji}(\tau)}$. With this, the coverage areas only depend on the current battery and channel state.

- *v*) Due to the integrity of the variables $x_{ji}(\tau)$, it is a combinatorial problem, whose computational burden grows as $\mathcal{O}(|\mathcal{B}|^{|\mathcal{U}_T|})$. A typical approach to follow in this situation is to relax the association variables so that $x_{ji}(\tau) \in [0, 1]$, $\forall i, j$ [Ye13b]. Notice that such relaxation will result in another optimization problem with a different optimum objective value as the constraint set has been expanded. If the optimum solution of the relaxed problem is integer, this corresponds to the optimum solution of the original problem. Otherwise, we can, for example, round the relaxed association variables so that $x_{ji}(\tau) \in \{0, 1\}$.

[2] If the number of users is relatively high, then $P^i_{\mathrm{RAD}}(\tau) \gg p_{ji}(\tau)$, and the approximation is fair. In any case, the approximation provides a lower bound of the actual SINR value.

[3] It is a worst-case solution as the BSs may not necessarily be transmitting at full power all the time.

Now, we present the different user association techniques developed assuming the previous simplifications.

7.3.1 General Epoch by Epoch User Association Formulation

In this section, we develop the most general epoch-by-epoch user association strategy. Let us introduce the following definitions: $\mathbf{p}(\tau) \triangleq \{p_{ji}(\tau),\ j \in \mathcal{U}_D, i \in \mathcal{B}\}$, $\check{\mathbf{p}}(\tau) \triangleq \{\check{p}_{ji}(\tau),\ j \in \mathcal{U}_V, i \in \mathcal{B}\}$, and $\mathbf{n}(\tau) \triangleq \{n_{ji}(\tau),\ j \in \mathcal{U}_D, i \in \mathcal{B}\}$. Given that, we formulate the optimization problem for the association strategy to be solved at the beginning of each epoch, which involves finding the indicators $\mathbf{x}(\tau)$ corresponding to the association as well as the average resource allocation variables, $\mathbf{p}(\tau)$, $\check{\mathbf{p}}(\tau)$, and $\mathbf{n}(\tau)$, that maximizes the aggregate utility function as follows:

$$\underset{\substack{\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\\ \mathbf{x}(\tau),\mathbf{p}_{\mathrm{RAD}}(\tau)}}{\text{maximize}} \quad \sum_{j\in\mathcal{U}_D} U_j\left(\sum_{i\in\mathcal{B}} n_{ji}(\tau)\log_2\left(1+\frac{M_D p_{ji}(\tau)}{n_{ji}(\tau)(\theta P^i_{\mathrm{RAD}}(\tau)+A_{ji}(\tau))}\right)\right) \tag{7.13}$$

$$\begin{aligned}
\text{subject to} \quad & C1: \sum_{i\in\mathcal{B}} \frac{\check{p}_{ji}(\tau)M_V}{\theta P^i_{\mathrm{RAD}}(\tau)+A_{ji}(\tau)} \geq \gamma_j, \quad \forall j \in \mathcal{U}_V \\
& C2: \sum_{i\in\mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j \in \mathcal{U}_T \\
& C3: \sum_{j\in\mathcal{U}_D} n_{ji}(\tau) \leq n_D^{(i)}, \quad \forall i \in \mathcal{B} \\
& C4: \sum_{j\in\mathcal{U}_D} p_{ji}(\tau) + \sum_{j\in\mathcal{U}_V} \check{p}_{ji}(\tau) \leq \bar{P}^i_{\mathrm{BS}}(C_i(\tau)), \quad \forall i \in \mathcal{B} \\
& C5: P^i_{\mathrm{RAD}}(\tau) = \sum_{j\in\mathcal{U}_D} p_{ji}(\tau) + \sum_{j\in\mathcal{U}_V} \check{p}_{ji}(\tau) + P^i_{\mathrm{CPICH}}, \quad \forall i \in \mathcal{B} \\
& C6: 0 \leq n_{ji}(\tau) \leq x_{ji}(\tau) n_D^{(i)}, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D \\
& C7: 0 \leq \check{p}_{ji}(\tau) \leq x_{ji}(\tau)\bar{P}^i_{\mathrm{BS}}(C_i(\tau)), \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_V \\
& C8: x_{ji}(\tau) = 0, \quad \forall i \notin \mathcal{S}_j(\tau), \forall j \in \mathcal{U}_T \\
& C9: x_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_T \\
& C10: p_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D.
\end{aligned}$$

In the previous optimization problem, $\bar{P}^i_{\mathrm{BS}}(C_i(\tau)) \triangleq \left(g_i(C_i(\tau))/T_{\mathrm{e}} - \left(P^i_{\mathrm{CPICH}} + P^i_c\right)\right)_0^{\infty}$ is the maximum power that the BS can use for traffic and pilot channels taking into account that the overall radiated power is limited and also the current battery level. Note that, even though at a particular epoch just the current battery is present in the formulation, the past harvesting and battery spendings also affect the current performance of the epoch as they appear in $C_i(\tau)$ through (7.2) (even though in (7.2) we used the time dependence t). Note that constraint $C6$ assures that if $x^\star_{ji} = 0$, then $n^\star_{ji} = 0$ and the rate $R^\star_{ji} = 0$. In this case, the allocated power will also be zero, $p^\star_{ji} = 0$, as it does not improve the objective function but wastes power if $p^\star_{ji} > 0$. Note that, at the optimum, $C4$ is attained with equality. Otherwise, if $C4$ is not fulfilled with

equality, we could re-scale all the powers with a common positive factor higher than 1 until $C4$ is fulfilled with equality. This would increase the objective function and all the other constraints would still be fulfilled. Then $P^{i\star}_{\mathrm{RAD}}(\tau) = \bar{P}^{i}_{\mathrm{BS}}(C_i(\tau)) + P^{i}_{\mathrm{CPICH}}$ and we can eliminate constraint $C5$ in problem (7.13).

Note that problem (7.13) is a convex optimization problem that can be solved using standard numerical algorithms [Boy04]. However, less complex algorithms can be derived by making use of the analytical structure of the problem. According to this, we propose in the following a more efficient algorithm based on the dual problem and the subgradient method.

7.3.1.1 Primal-Dual Solution

In this subsection, we develop the association algorithm based on the dual problem of (7.13). The optimal solution will be presented first as a function of the Lagrange multipliers (or dual variables). A gradient-type scheme will later be developed for computing the optimum multipliers. Let $\boldsymbol{\nu} = \{\nu_j,\, j \in \mathcal{U}_V\}$, $\boldsymbol{\beta} = \{\beta_j,\, j \in \mathcal{U}_T\}$, $\boldsymbol{\mu} = \{\mu_i,\, i \in \mathcal{B}\}$, $\boldsymbol{\lambda} = \{\lambda_i,\, i \in \mathcal{B}\}$, $\boldsymbol{\pi} = \{\pi_{ji},\, j \in \mathcal{U}_D, i \in \mathcal{B}\}$, and $\boldsymbol{\xi} = \{\xi_{ji},\, j \in \mathcal{U}_V, i \in \mathcal{B}\}$ denote the vectors of dual variables associated to constraints $C1, \ldots, C4$, $C6$, and $C7$ in problem (7.13). We collect all the Lagrange multipliers in $\boldsymbol{\Psi} = \{\boldsymbol{\nu}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\xi}\}$. There is no need to dualize constraints $C8$, $C9$, and $C10$ because the solution will automatically satisfy them. Note that the Lagrange multipliers depend on the particular epoch and they should be time dependent. However, for simplicity in the notation, from now on we will omit the time dependence unless stated otherwise. The Lagrangian of the optimization problem (7.13) is

$$\begin{aligned}\mathcal{L}(\mathbf{p}, \check{\mathbf{p}}, \mathbf{n}, \mathbf{x}, \boldsymbol{\Psi}) = &- \sum_{j\in\mathcal{U}_D} U_j \left(\sum_{i\in\mathcal{B}} n_{ji} \log_2 \left(1 + \frac{M_D p_{ji}}{n_{ji}(\theta P^{i\star}_{\mathrm{RAD}} + A_{ji})} \right) \right) \\ &+ \sum_{j\in\mathcal{U}_V} \nu_j \left(\gamma_j - \sum_{i\in\mathcal{B}} \frac{\check{p}_{ji} M_V}{\theta P^{i\star}_{\mathrm{RAD}} + A_{ji}} \right) + \sum_{j\in\mathcal{U}_T} \beta_j \left(\sum_{i\in\mathcal{B}} x_{ji} - 1 \right) \\ &+ \sum_{i\in\mathcal{B}} \mu_i \left(\sum_{j\in\mathcal{U}_D} n_{ji} - n^{(i)}_D \right) + \sum_{i\in\mathcal{B}} \lambda_i \left(\sum_{j\in\mathcal{U}_D} p_{ji} + \sum_{j\in\mathcal{U}_V} \check{p}_{ji} - \bar{P}^{i}_{\mathrm{BS}}(B_i) \right) \\ &+ \sum_{j\in\mathcal{U}_D} \sum_{i\in\mathcal{B}} \xi_{ji} \left(\check{p}_{ji} - x_{ji} \bar{P}^{i}_{\mathrm{BS}}(B_i) \right) + \sum_{j\in\mathcal{U}_D} \sum_{i\in\mathcal{B}} \pi_{ji} \left(n_{ji} - x_{ji} n^{(i)}_D \right).\end{aligned} \tag{7.14}$$

The dual function is defined as the minimum of the Lagrangian w.r.t. the primal variables, i.e.,

$$D(\boldsymbol{\Psi}) = \inf_{\substack{\mathbf{p}\succeq\mathbf{0}, \check{\mathbf{p}}\succeq\mathbf{0}, \\ \mathbf{n}\succeq\mathbf{0}, \mathbf{x}\succeq\mathbf{0}}} \mathcal{L}(\mathbf{p}, \check{\mathbf{p}}, \mathbf{n}, \mathbf{x}, \boldsymbol{\Psi}). \tag{7.15}$$

The dual problem is the following optimization problem:

$$\underset{\substack{\boldsymbol{\nu}\succeq\mathbf{0},\boldsymbol{\mu}\succeq\mathbf{0},\boldsymbol{\lambda}\succeq\mathbf{0},\\ \boldsymbol{\pi}\succeq\mathbf{0},\boldsymbol{\xi}\succeq\mathbf{0}}}{\text{maximize}} \quad D(\boldsymbol{\Psi}). \tag{7.16}$$

Recall that the dual problem is always convex w.r.t. the dual variables and, thus, can be efficiently solved with a projected gradient method (if $D(\boldsymbol{\Psi})$ is differentiable) or a projected supergradient method if it is not differentiable [Ber99].

Strong duality holds in problem (7.13) and the solution obtained by solving the primal problem (7.13) and the dual problem (7.16) is the same [Ber99]. For a given set of Lagrange multipliers, $\{\boldsymbol{\nu},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\lambda},\boldsymbol{\pi},\boldsymbol{\xi}\}$, we need to minimize the Lagrangian w.r.t. the primal variables. As it will be shown next, the structure of $\mathcal{L}(\mathbf{p},\,\check{\mathbf{p}},\,\mathbf{n},\,\mathbf{x},\boldsymbol{\Psi})$ allows the minimization w.r.t. $\mathbf{p}$ and $\mathbf{n}$ to be found in closed-form. Because $\mathcal{L}(\mathbf{p},\,\check{\mathbf{p}},\,\mathbf{n},\,\mathbf{x},\boldsymbol{\Psi})$ is strictly convex and differentiable w.r.t. $\mathbf{p}$ and $\mathbf{n}$, minimization w.r.t. these variables requires to equating the corresponding partial derivatives of $\mathcal{L}(\mathbf{p},\,\check{\mathbf{p}},\,\mathbf{n},\,\mathbf{x},\boldsymbol{\Psi})$ to zero. Differentiating the Lagrangian w.r.t. the data powers, equating the derivatives to zero and solving such expression for the data powers yields

$$p^{\star}_{k\ell}(\mathbf{n},\mathbf{p},\boldsymbol{\Psi}) = \left(\frac{G_{k\ell}(\mathbf{n},\mathbf{p})n_{k\ell}}{\ln(2)\lambda_{\ell}} - \frac{(\theta P^{\ell\star}_{\mathrm{RAD}} + A_{k\ell})n_{k\ell}}{M_D}\right)^{\infty}_{0}, \tag{7.17}$$

where the projection guarantees that the allocated powers are nonnegative (constraint $C10$ in (7.13)) and $G_{k\ell}(\mathbf{n},\mathbf{p}) \triangleq \dot{U}_k\left(s_{k\ell} + n_{k\ell}\log_2\left(1+\frac{M_D p_{k\ell}}{n_{k\ell}(\theta P^{\ell\star}_{\mathrm{RAD}}+A_{k\ell})}\right)\right)$ where $\dot{U}_k(\cdot)$ is the derivative of function $U_k(\cdot)$ and $s_{k\ell} = \sum_{i\in\mathcal{B}/\{\ell\}} n_{ki}\log_2\left(1+\frac{M_D p_{ki}}{n_{ki}(\theta P^{i\star}_{\mathrm{RAD}}+A_{ki})}\right)$ is a constant term for all users connected to BS ℓ. Notice that the bisection method can be used to calculate numerically the optimum power allocation (7.17). Proceeding similarly in the calculation of the optimum code allocation, we calculate the partial derivatives of the Lagrangian w.r.t. n_{ji}, equate them to zero and solve such equations for the codes, obtaining:

$$n^{\star}_{k\ell}(\mathbf{n},\mathbf{p},\boldsymbol{\Psi}) = \begin{cases} \left(\dfrac{G_{k\ell}(\mathbf{n},\mathbf{p})M_D p_{k\ell}((\theta P^{\ell\star}_{\mathrm{RAD}}+A_{k\ell})\ln(2))^{-1}}{G_{k\ell}(\mathbf{n},\mathbf{p})\log_2\left(1+\frac{M_D p_{k\ell}}{n_{k\ell}(\theta P^{\ell\star}_{\mathrm{RAD}}+A_{k\ell})}\right)-\mu_{\ell}-\pi_{k\ell}} - \dfrac{M_D p_{k\ell}}{\theta P^{\ell\star}_{\mathrm{RAD}}+A_{k\ell}}\right)^{\infty}_{0} & \text{if } \ell\in\mathcal{S}_k, \\ 0 & \text{if } \ell\notin\mathcal{S}_k, \end{cases} \tag{7.18}$$

where the projection guarantees that the codes satisfy the nonnegative constraints $C6$ and the definition by parts of the function of the codes is due to constraint $C8$. Notice that, the bisection method can also be used in this case to compute the optimum code allocation, $n_{k\ell}$.

Unfortunately, the minimization of the Lagrangian w.r.t. the association variables x_{ji} and the voice powers $\check{p}_{ji}$ cannot be obtained by differentiating the Lagrangian as they appear through linear terms. In order to obtain the optimum power allocation for voice users and the association variables, we employ an iterative projected gradient approach to minimize the Lagrangian. The

gradient for the j-th voice user power connected to the i-th BS is given by

$$\nabla_{\check{p}_{ji}}\mathcal{L}(\mathbf{p},\,\check{\mathbf{p}},\,\mathbf{n},\,\mathbf{x},\boldsymbol{\Psi}) \triangleq s_{ji} = \frac{\nu_j M_D}{\theta P^{i\star}_{\mathrm{RAD}} + A_{ji}} + \lambda_i + \xi_{ji}, \quad \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}, \tag{7.19}$$

and the update equation of the projected gradient method is given by

$$\check{p}^{(k+1)}_{ji}(\tau) = \left(\check{p}^{(k)}_{ji}(\tau) - \delta^{(k)} s_{ji}(\tau)\right)_0^{\infty}, \quad \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}, \tag{7.20}$$

where k is the iteration index, the projection guarantees the nonnegativity constraints in (7.13), and $\delta^{(k)} = \frac{K}{\sqrt{k}\|\nabla\mathcal{L}\|_2}$ is the step size chosen such that the diminishing conditions are fulfilled, i.e., $\lim_{k\to\infty}\delta^{(k)} = 0,\ \sum_{k=1}^{\infty}\delta^{(k)} = \infty$, being $\nabla\mathcal{L}$ the gradient of the Lagrangian (w.r.t. all variables) [Ber99]. Note that we have introduced the time dependence in order to clearly show the different time scales. For the association variables, the gradient is given by

$$\nabla_{x_{ji}}\mathcal{L}(\mathbf{p},\,\check{\mathbf{p}},\,\mathbf{n},\,\mathbf{x},\boldsymbol{\Psi}) \triangleq t_{ji} = \beta_j - \xi_{ji}\bar{P}^i_{\mathrm{BS}}(B_i) - \pi_{ji} n_D^{(i)}, \quad \forall j \in \mathcal{U}_T, \forall i \in \mathcal{B}, \tag{7.21}$$

and the update equation of the projected method is

$$x^{(k+1)}_{ji}(\tau) = \left(x^{(k)}_{ji}(\tau) - \delta^{(k)} t_{ji}(\tau)\right)_0^{\infty}, \quad \forall j \in \mathcal{U}_T, \forall i \in \mathcal{B}. \tag{7.22}$$

At this point, we have the expressions of the optimum primal variables, either in closed-form in (7.17) and (7.18) or iteratively in (7.20) and (7.22), for given dual variables. If the dual variables were optimum, then the expressions for the primal variables would yield the optimum values. The optimum Lagrange multipliers can be obtained from the dual problem by maximizing the dual function in (7.16). As the dual function is concave and generally not differentiable, we can apply any subgradient-type algorithm to find the optimum solution. A valid supergradient for each particular dual variable is given by the constraint it is associated with [Ber99]. The update equations are given by (introducing the time dependence)

$$\nu_j^{(q+1)}(\tau) = \left(\nu_j^{(q)}(\tau) + \kappa^{(q)}\left(\gamma_j - \sum_{i\in\mathcal{B}}\frac{\check{p}_{ji}(\tau)M_V}{\theta P^{i\star}_{\mathrm{RAD}}(\tau) + A_{ji}(\tau)}\right)\right)_0^{\infty}, \forall j \in \mathcal{U}_V, \tag{7.23}$$

$$\beta_j^{(q+1)}(\tau) = \beta_j^{(q)}(\tau) + \kappa^{(q)}\left(\sum_{i\in\mathcal{B}} x_{ji}(\tau) - 1\right), \quad \forall j \in \mathcal{U}_T, \tag{7.24}$$

$$\mu_i^{(q+1)}(\tau) = \left(\mu_i^{(q)}(\tau) + \kappa^{(q)}\left(\sum_{j\in\mathcal{U}_D} n_{ji}(\tau) - n_D^{(i)}\right)\right)_0^{\infty}, \quad \forall i \in \mathcal{B}, \tag{7.25}$$

$$\lambda_i^{(q+1)}(\tau) = \left(\lambda_i^{(q)}(\tau) + \kappa^{(q)}\left(\sum_{j\in\mathcal{U}_D} p_{ji}(\tau) + \sum_{j\in\mathcal{U}_V}\check{p}_{ji}(\tau) - \bar{P}^i_{\mathrm{BS}}(C_i(\tau))\right)\right)_0^{\infty} \quad \forall i \in \mathcal{B}, \tag{7.26}$$

$$\xi_{ji}^{(q+1)}(\tau) = \left(\xi_{ji}^{(q)}(\tau) + \kappa^{(q)}\left(\check{p}_{ji}(\tau) - x_{ji}(\tau)\bar{P}^i_{\mathrm{BS}}(C_i(\tau))\right)\right)_0^{\infty}, \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}, \tag{7.27}$$

$$\pi_{ji}^{(q+1)}(\tau) = \left(\pi_{ji}^{(q)}(\tau) + \kappa^{(q)}\left(n_{ji}(\tau) - x_{ji}(\tau)n_D^{(i)}\right)\right)_0^{\infty}, \quad \forall j \in \mathcal{U}_D, \forall i \in \mathcal{B}, \tag{7.28}$$

where the projections guarantee the nonnegativity constraints of the dual variables in (7.16), q indicates the iteration index, and the step size is given by $\kappa^{(q)} = \frac{Q}{\sqrt{q}\|\nabla D\|_2}$, so that the diminishing conditions for the step size assure convergence (in the previous expression, ∇D denotes the supergradient of the dual function w.r.t. all variables). Once we know the optimal dual variables $\boldsymbol{\Psi}^\star$, we can obtain the optimum associations $\mathbf{x}^\star(\boldsymbol{\Psi}^\star)$. The last step requires the quantization or rounding of such variables since the solution of (7.13) will provide, in general, a value of $x_{ji}^\star \in [0,1]$, but, for the actual implementation of the user association, we require $x_{ji}^\star \in \{0,1\}$ with $\sum_{i\in\mathcal{B}} x_{ji}^\star = 1$, $\forall j$.

The proposed user association algorithm is based on the primal-dual block coordinate descent method for the update of the primal variables p_{ji} and n_{ji} (see Algorithm 7.1). The optimal association policy for any user j can be written in closed form using the indicator function as

$$x_{ji}^\star(\boldsymbol{\beta}^\star, \boldsymbol{\xi}^\star, \boldsymbol{\pi}^\star) = \mathbb{1}_{\{i=\arg\max_{i'}\{x_{ji'}(\boldsymbol{\beta}^\star,\boldsymbol{\xi}^\star,\boldsymbol{\pi}^\star)\}\}}. \tag{7.29}$$

If multiple maximums exist simultaneously in (7.29), we propose to select just one randomly.

7.3.2 Epoch by Epoch User Association: Low Complexity Solutions

In this section we develop some strategies with lower complexity than the solution presented in the previous section. The goal is to provide algorithms with a reduced computational burden that could be of interest in scenarios where there is high mobility and the association procedure has to be executed more frequently.

7.3.2.1 Association of Voice and Data Users Separately

The first approach to consider is to split the two sets of users, namely the sets of voice and data users, and solve the association procedure for each group separately, first the voice users and then the data users. This is a good approach if the users are more or less spread out throughout the network. The assumption would not be valid if most of the data users are concentrated around just one BS and the voice users are connected to that BS, even though they could be connected to the neighboring BSs due to their overlapping coverage areas. In any case, if the number of voice users is relatively low compared to the number of data users, such approximation is reasonably fair.

Voice Users Association

We first perform the association of voice users. As we saw earlier, each particular voice user has to fulfill a minimum quality constraint in terms of SINR to guarantee a fixed rate service:

$$\sum_{i\in\mathcal{B}} x_{ji}(\tau)\frac{M_V \check{p}_{ji}(\tau)}{\theta P_{\mathrm{RAD}}^{i\star}(\tau) + A_{ji}(\tau)} \geq \gamma_j, \quad \forall j \in \mathcal{U}_V. \tag{7.30}$$

Algorithm 7.1 Primal-dual general user association algorithm

input: $C_i(\tau)$, $\forall i \in \mathcal{B}$
repeat (index τ)
  compute $P^{i\star}_{\text{RAD}}(\tau) = \bar{P}^i_{\text{BS}}(C_i(\tau)) + P^i_{\text{CPICH}}$, $\forall i \in \mathcal{B}$
  calculate $\mathcal{S}_j(\tau)$, $\forall j \in \mathcal{U}_T$
  primal-dual algorithm:
    initialize $\boldsymbol{\nu}(\tau) \succeq 0, \boldsymbol{\beta}(\tau), \boldsymbol{\mu}(\tau) \succeq 0, \boldsymbol{\lambda}(\tau) \succeq 0, \boldsymbol{\pi}(\tau) \succeq 0, \boldsymbol{\xi}(\tau) \succeq 0$
    repeat (index q)
      initialize $\mathbf{n}(\tau) \succeq 0$
      repeat (index k)
        $p^{(q,k+1)}_{ji}(\tau) = p^{\star}_{ji}\left(\mathbf{n}^{(q,k)}, \mathbf{p}^{(q,k)}, \boldsymbol{\Psi}^{(q)}, \tau\right)$ using (7.17) $\quad \forall j, i$
        $n^{(q,k+1)}_{ji}(\tau) = n^{\star}_{ji}\left(\mathbf{n}^{(q,k)}, \mathbf{p}^{(q,k+1)}, \boldsymbol{\Psi}^{(q)}, \tau\right)$ using (7.18) $\quad \forall j, i$
      until $p^{(q,k+1)}_{ji}(\tau)$ and $n^{(q,k+1)}_{ji}(\tau)$ converge
      repeat (index k)
        $\check{p}^{(q,k+1)}_{ji}(\tau) = \left(\check{p}^{(q,k)}_{ji}(\tau) - \delta^{(k)} s^{(q)}_{ji}(\tau)\right)^{\infty}_{0} \quad \forall j, i$
      until $\check{p}^{(q,k+1)}_{ji}(\tau)$ converges
      repeat (index k)
        $x^{(q,k+1)}_{ji}(\tau) = \left(x^{(q,k)}_{ji}(\tau) - \delta^{(k)} t^{(q)}_{ji}(\tau)\right)^{\infty}_{0} \quad \forall j, i$
      until $x^{(q,k+1)}_{ji}(\tau)$ converges
      update the dual variables using $p^{(q)}_{ji}(\tau)$, $\check{p}^{(q)}_{ji}(\tau)$, $n^{(q)}_{ji}(\tau)$, and $x^{(q)}_{ji}(\tau)$
      with (7.23), (7.24), (7.25), (7.26), (7.27), and (7.28)
    until $\boldsymbol{\nu}^{(q+1)}(\tau), \boldsymbol{\beta}^{(q+1)}(\tau), \boldsymbol{\mu}^{(q+1)}(\tau), \boldsymbol{\lambda}^{(q+1)}(\tau), \boldsymbol{\pi}^{(q+1)}(\tau)$, and $\boldsymbol{\xi}^{(q+1)}(\tau)$ converge
  end of primal-dual algorithm
  quantization of $x_{ji}(\tau)$:
  $\forall j \in \mathcal{U}_T \longrightarrow i^{\star} = \arg\max_{i \in \mathcal{B}} x_{ji}(\tau)$
  $x^{\star}_{ji^{\star}}(\tau) = 1$, $x^{\star}_{ji}(\tau) = 0, \forall i \neq i^{\star}$
  update batteries:
  $C^i(\tau + 1) = (C_i(\tau) - E_i(\tau) + H_i(\tau))^{C^i_{\max}}_{0}$, $\forall i \in \mathcal{B}$
for all epochs $\tau \in \Upsilon$
end algorithm

In order not to waste power unnecessarily, the previous constraint should be tight at the optimum. Given that, in the following we present the procedure for the association of the voice users: for all users $j \in \mathcal{U}_V$, find the BS i' such that $\theta P^{i'\star}_{\text{RAD}}(\tau) + A_{ji'}(\tau)$ is minimum and $i' \in \mathcal{S}_j(\tau)$,

i.e., constraint $C8$ from problem (7.13) is not violated. According to this, the association of voice users can be written in closed form using the indicator function as

$$x^{\star}_{ji}(\tau) = \mathbb{1}_{\{i = \arg\min_{i'}\{\theta P^{i'\star}_{\mathrm{RAD}}(\tau) + A_{ji'}(\tau)\,|\,i' \in \mathcal{S}_j(\tau)\}\}}, \quad \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}. \tag{7.31}$$

The optimum power to be allocated to voice user j associated to BS i' is given by

$$\check{p}^{\star}_{ji'}(\tau) = \begin{cases} \gamma_j \frac{\theta P^{i'\star}_{\mathrm{RAD}}(\tau) + A_{ji'}(\tau)}{M_V}, & \text{if } x^{\star}_{ji}(\tau) = 1, \\ 0, & \text{if } x^{\star}_{ji}(\tau) = 0. \end{cases} \tag{7.32}$$

The association is feasible if $P^{(i)}_V(\tau) \triangleq \sum_{j \in \mathcal{U}_V} \check{p}^{\star}_{ji}(\tau) \leq \bar{P}^{i}_{\mathrm{BS}}(C_i(\tau))$, $\forall i \in \mathcal{B}$, otherwise there is not enough energy at the batteries to fulfill all the minimum SINR constraints. In such case, a few users should be dropped from the system or their target SINR, γ_j, should be reduced.

Data Users Association

Assuming that the voice user association is feasible, we can focus on the association of the data users. We model the association procedure through the following convex problem:

$$\begin{aligned} \underset{\mathbf{p}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau)}{\text{maximize}} \quad & \sum_{j \in \mathcal{U}_D} U_j \left(\sum_{i \in \mathcal{B}} n_{ji}(\tau) \log_2 \left(1 + \frac{M_D p_{ji}(\tau)}{n_{ji}(\tau)(\theta_j P^{i\star}_{\mathrm{RAD}}(\tau) + A_{ji}(\tau))} \right) \right) \\ \text{subject to} \quad & C1: \sum_{i \in \mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j \in \mathcal{U}_D \\ & C2: \sum_{j \in \mathcal{U}_D} n_{ji}(\tau) \leq n^{(i)}_D, \quad \forall i \in \mathcal{B} \\ & C3: \sum_{j \in \mathcal{U}_D} p_{ji}(\tau) \leq \bar{P}^{i}_{\mathrm{BS}}(C_i(\tau)) - P^{(i)}_V(\tau), \quad \forall i \in \mathcal{B} \\ & C4: 0 \leq n_{ji}(\tau) \leq x_{ji}(\tau) n^{(i)}_D, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D \\ & C5: x_{ji}(\tau) = 0, \quad \forall i \notin \mathcal{S}_j(\tau), \forall j \in \mathcal{U}_D \\ & C6: x_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D \\ & C7: p_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D. \end{aligned} \tag{7.33}$$

Notice the similarities of the previous optimization problem w.r.t. (7.13). We have eliminated the constraints involving the voice users and, now, the user set just includes the data users, $\mathcal{U}_D$, not the overall set of users, $\mathcal{U}_T$, as before. Notice that we have already relaxed the association variables making the problem convex. Strong duality also holds for this problem and, therefore, it can be solved using the primal-dual approach followed in the previous section. The details of the procedure employed to solved this problem will not be presented in the chapter. This association strategy will be referred as first-voice then-data (FVTD).

7.3.2.2 Fixed Resources per Data User

We can go one step further and reduce the complexity of the previous algorithm by considering that the number of codes or the power assigned to data users are fixed. In this case, only one variable (code or power) will be used as a load metric of the BS. There is a slight conceptual difference between these two approaches (fixed codes and fixed power) when compared to the previous approach or the general approach in (7.13). In the two previous cases, the association strategy did not force to assign resources to all users, i.e., some data users could be assigned to a particular BS but the average rate they are assigned could be zero. However, as in this case a given user is already assigned a portion of the total power or a certain number of codes, the associated rate will be greater than zero, i.e., the BS will spend resources as long as such user is associated to it.

Fixed Power Allocation per Data User

In this case, we consider that all users get a given portion of the power available at the BS, i.e., $p_i(\tau) \triangleq \frac{\bar{P}^i_{\text{BS}}(C_i(\tau)) - P^{(i)}_V(\tau)}{K_i}$, $\forall i \in \mathcal{B}$, where K_i is the fixed maximum number of users that the BS i wants to accept. In the most general case, this portion is different for each BS, where such portion together with the total available power constrains the maximum number of users that are allowed to be connected to each particular BS. The user association strategy is formulated through the following convex optimization problem:

$$
\begin{aligned}
\underset{\mathbf{n}(\tau),\mathbf{x}(\tau)}{\text{maximize}} \quad & \sum_{j\in\mathcal{U}_D} U_j\left(\sum_{i\in\mathcal{B}} n_{ji}(\tau)\log_2\left(1+\frac{M_D p_i(\tau)}{n_{ji}(\tau)(\theta P^{i\star}_{\text{RAD}}(\tau)+A_{ji}(\tau))}\right)\right) \qquad (7.34)\\
\text{subject to} \quad & C1: \sum_{i\in\mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j \in \mathcal{U}_D\\
& C2: \sum_{j\in\mathcal{U}_D} n_{ji}(\tau) \leq n^{(i)}_D, \quad \forall i \in \mathcal{B}\\
& C3: \sum_{j\in\mathcal{U}_D} x_{ji}(\tau) \leq \frac{\bar{P}^i_{\text{BS}}(C_i(\tau)) - P^{(i)}_V(\tau)}{p_i(\tau)}, \quad \forall i \in \mathcal{B}\\
& C4: 0 \leq n_{ji}(\tau) \leq x_{ji}(\tau) n^{(i)}_D, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D\\
& C5: x_{ji}(\tau) = 0, \quad \forall i \notin \mathcal{S}_j(\tau), \forall j \in \mathcal{U}_D\\
& C6: x_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D.
\end{aligned}
$$

We will not present the details of the solution of this problem for the sake of space as we have already presented the methodology for the previous two association problems. We will just evaluate the performance of the strategy in the simulations section.

Fixed Code Allocation per Data User

Now we consider that a given number of codes are assigned to each particular data user, i.e., $n_i \triangleq \frac{n^{(i)}_D}{K_i}$, $\forall i \in \mathcal{B}$, where K_i is again the maximum number of users that BS i wants to accept.

In this case, the power is the resource that is optimized. The association strategy is, thus, formulated as

$$\begin{aligned}
\underset{\mathbf{p}(\tau),\mathbf{x}(\tau)}{\text{maximize}} \quad & \sum_{j\in\mathcal{U}_D} U_j\left(\sum_{i\in\mathcal{B}} n_i \log_2\left(1+\frac{M_D p_{ji}(\tau)}{n_i(\theta P^{i\star}_{\text{RAD}}(\tau)+A_{ji}(\tau))}\right)\right) \qquad (7.35)\\
\text{subject to} \quad & C1: \sum_{i\in\mathcal{B}} x_{ji}(\tau)=1, \quad \forall j\in\mathcal{U}_D\\
& C2: \sum_{j\in\mathcal{U}_D} x_{ji}(\tau)\le \frac{n_D^{(i)}}{n_i}, \quad \forall i\in\mathcal{B}\\
& C3: \sum_{j\in\mathcal{U}_D} p_{ji}(\tau)\le \bar{P}^i_{\text{BS}}(C_i(\tau))-P_V^{(i)}(\tau), \quad \forall i\in\mathcal{B}\\
& C4: 0\le p_{ji}(\tau)\le x_{ji}(\tau)\left(\bar{P}^i_{\text{BS}}(C_i(\tau))-P_V^{(i)}(\tau)\right), \quad \forall i\in\mathcal{B},\forall j\in\mathcal{U}_D\\
& C5: x_{ji}(\tau)=0, \quad \forall i\notin\mathcal{S}_j(\tau),\forall j\in\mathcal{U}_D\\
& C6: x_{ji}(\tau)\ge 0, \quad \forall i\in\mathcal{B},\forall j\in\mathcal{U}_D.
\end{aligned}$$

Also in this case, we will not present the details of the solution of the problem as we have already presented the methodology for the previous two association problems. We will just evaluate the performance of the strategy in the numerical sections.

7.3.3 Distributed Algorithm

In this section, we extend the concept presented in previous sections where the association strategy was executed by a centralized entity that required the knowledge of all the parameters involved in the association decision, such as the channels of all users to all BSs, the current battery levels of all BSs, etc. Due to the complexity of the acquisition of such parameters and the complexity of the association algorithm itself, the centralized approach for solving such problems only allows a slow adaptation at relatively long timescales. Moreover, if the central entity is implemented in the core network, then physical links and an intensive exchange of messages among the central entity and the different tiers is required. Nevertheless, such connections are usually not feasible as in heterogeneous networks, BSs are deployed by operators and users (femtocells are deployed in home scenarios [Cha08]). In such situations, a low complexity distributed algorithm is desirable. In this section, we propose a distributed algorithm of the general problem presented in (7.13) via Lagrange dual decomposition [Pal06]. We can apply a dual decomposition to the original problem in (7.13) and solve it separately by users and BSs. This implies that only local information is required to solve the subproblems, i.e., users only need to know the available resources of the BSs they are able to connect to and the individual propagation channels with them.

Having introduced the motivation behind the distributed solution, we proceed now to explain

the steps of the procedure. As the objective function of the problem cannot be decomposed for each BS, the users will be the ones carrying out the computation of the association. The role of the BSs will be to provide certain *prices* to the users as a function of the association demands. Thus, the distributed solution will require a few iterations before the overall *agreement* is achieved [Pal06]. Because each user selects the preferred BS, the criterion to be employed for the voice users is the one explained in 7.3.2.1. As a consequence, the voice users and data users must select the BSs separately: first the voice users and then the data users.

Once the voice users have selected the BSs they prefer and the problem is feasible, we focus on the data users by applying a dual decomposition method to problem (7.33). The coupling constraints are $C2$ and $C3$ as we have just one constraint for all users not allowing to split the optimization problem for each user. This motivates us to dualize both constraints and decouple the problem as explained in the sequel. Let $\boldsymbol{\mu}(\tau) = \{\mu_i(\tau),\, i \in \mathcal{B}\}$ and $\boldsymbol{\lambda}(\tau) = \{\lambda_i(\tau),\, i \in \mathcal{B}\}$ be the Lagrange multipliers associated to constraints $C2$ and $C3$ and let $\mathbf{p}_j(\tau) = \{p_{ji}(\tau),\, i \in \mathcal{S}_j(\tau)\}$, $\mathbf{n}_j(\tau) = \{n_{ji}(\tau),\, i \in \mathcal{S}_j(\tau)\}$, and $\mathbf{x}_j(\tau) = \{x_{ji}(\tau),\, i \in \mathcal{S}_j(\tau)\}$. Then, the j-th data user must solve the following optimization problem (where, for simplicity in the notation and the resulting algorithm, we have eliminated the terms that do not depend on the optimization variables, i.e., constant terms, in the objective function as they do not affect the optimization variables):

$$
\begin{aligned}
\underset{\mathbf{p}_j(\tau),\, \mathbf{n}_j(\tau)\mathbf{x}_j(\tau)}{\text{maximize}} \quad & U_j\left(\sum_{i\in\mathcal{S}_j(\tau)} n_{ji}(\tau)\log_2\left(1+\frac{M_D p_{ji}(\tau)}{n_{ji}(\tau)(\theta P^{i\star}_{\text{RAD}}(\tau)+A_{ji}(\tau))}\right)\right) \\
& - \sum_{i\in\mathcal{S}_j(\tau)} \mu_i(\tau) n_{ji}(\tau) - \sum_{i\in\mathcal{S}_j(\tau)} \lambda_i(\tau) p_{ji}(\tau) \\
\text{subject to} \quad & C1: \sum_{i\in\mathcal{S}_j(\tau)} x_{ji}(\tau) = 1 \\
& C2: 0 \le n_{ji}(\tau) \le x_{ji}(\tau) n_D^{(i)}, \quad \forall i \in \mathcal{S}_j(\tau) \\
& C3: x_{ji}(\tau) \ge 0, \quad \forall i \in \mathcal{S}_j(\tau) \\
& C4: p_{ji}(\tau) \ge 0, \quad \forall i \in \mathcal{S}_j(\tau).
\end{aligned}
\tag{7.36}
$$

Notice that the knowledge of the Lagrange multipliers $\boldsymbol{\mu}(\tau)$ and $\boldsymbol{\lambda}(\tau)$ is required to solve the previous problem. Notice also that only local information is needed, which is represented by the set of available BSs for each particular user given by $\mathcal{S}_j(\tau)$ for the j-th user. The previous optimization problem is convex and can be solved using any standard procedure as the one explained in this document (primal-dual iteration) in subsection 7.3.1.1.

When the data users execute the previous association strategy, they convey the results of the optimum variables, ($\mathbf{p}_j^\star(\tau)$, $\mathbf{n}_j^\star(\tau)$, and $\mathbf{x}_j^\star(\tau)$), to the BSs. Then, the BSs are responsible for iterating to solve the dual problem to obtain asymptotically the optimum Lagrange multipliers

$\boldsymbol{\mu}(\tau)$ and $\boldsymbol{\lambda}(\tau)$ using

$$\mu_i^{(q+1)}(\tau) = \left(\mu_i^{(q)}(\tau) + \delta^{(q)} \left(\sum_{j \in \mathcal{U}_D^{(i)}} n_{ji}^{(q)}(\tau) - n_D^{(i)} \right) \right)_0^{\infty}, \quad \forall i \in \mathcal{B} \tag{7.37}$$

$$\lambda_i^{(q+1)}(\tau) = \left(\lambda_i^{(q)}(\tau) + \delta^{(q)} \left(\sum_{j \in \mathcal{U}_D^{(i)}} p_{ji}^{(q)}(\tau) - \bar{P}_{\mathrm{BS}}^{i}(C_i(\tau)) + P_V^{(i)}(\tau) \right) \right)_0^{\infty}, \quad \forall i \in \mathcal{B}. \tag{7.38}$$

The BSs broadcast the information of the updated values of the Lagrange multipliers and, then, users execute again the association procedure. As the dual problem is convex, any subgradient-type algorithm yields asymptotically the optimum solution. Because Slater's condition is satisfied, strong duality holds and the original problem (7.33) can be equivalently solved in a distributed fashion without coordination among the users or the BSs and the convergence is guaranteed.

The final distributed algorithm is shown in Algorithm 7.2. In the algorithm, $\delta^{(q)}$ is the step size. The multipliers $\boldsymbol{\mu}(\tau)$ and $\boldsymbol{\lambda}(\tau)$ work as messages between data users and BSs. They represent the *price* of the resources of the particular BS determined by the load situation. If, for example, the demand of code resources for a given BS i, $\sum_j n_{ji}(\tau)$, is larger than the maximum number that the BS can offer, $\sum_j n_{ji}(\tau) > n_D^{(i)}$ then, the price μ_i increases, which means that the users will have to pay more to utilize its code resources making the users consider other BS options with lower prices that would improve its objective function (and, thus, its throughput). The same reasoning can be applied to the power. If a given BS is underutilized, then its corresponding prices will be low making it appealing for the users achieving a well load-balanced network.

Algorithm 7.2 User association strategy based on the distributed algorithm

input: $C_i(\tau)$, $\forall i \in \mathcal{B}$

repeat (index τ)

compute $P^{i\star}_{\text{RAD}}(\tau) = \bar{P}^i_{\text{BS}}(C_i(\tau)) + P^i_{\text{CPICH}}$, $\forall i \in \mathcal{B}$

calculate $\mathcal{S}_j(\tau)$, $\forall j \in \mathcal{U}_T$

distributed algorithm:

voice users:

$x^\star_{ji}(\tau) = \mathbb{1}_{\{i = \arg\min_{i'}\{\theta P^{i'\star}_{\text{RAD}}(\tau) + A_{ji'}(\tau) \,|\, i' \in S_j(\tau)\}\}}, \quad \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}$

$\check{p}^\star_{ji}(\tau) = x^\star_{ji}(\tau) \frac{\gamma_j(\theta P^{i\star}_{\text{RAD}}(\tau) + A_{ji}(\tau))}{M_V}, \qquad \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}$

$P^{(i)}_V(\tau) \triangleq \sum_{j \in \mathcal{U}_V} \check{p}^\star_{ji}(\tau), \quad \forall i \in \mathcal{B}$

if $P^{(i)}_V(\tau) > \bar{P}^i_{\text{BS}}(C_i(\tau))$

find any j' such that $x^\star_{j'i}(\tau) = 1$

assign $x^\star_{j'i}(\tau) = 0$ and $x^\star_{j'k}(\tau) = \mathbb{1}_{\{k = \arg\min_{k'}\{\theta P^{k'\star}_{\text{RAD}}(\tau) + A_{j'k'}(\tau) \,|\, k' \in \mathcal{S}_{j'}(\tau) \setminus \{i\}\}\}}$

go to 8

end if

initialize $\boldsymbol{\mu}(\tau) \succeq 0, \boldsymbol{\lambda}(\tau) \succeq 0$

repeat (index q)

data users:

solve problem (7.36) and obtain $\mathbf{p}^{(q)}_j(\boldsymbol{\lambda}^{(q)}, \tau)$ and $\mathbf{n}^{(q)}_j(\boldsymbol{\mu}^{(q)}, \tau)$, $\forall j \in \mathcal{U}_D$

each user sends $p^{(q)}_{ji}(\boldsymbol{\lambda}^{(q)}, \tau)$ and $n^{(q)}_{ji}(\boldsymbol{\mu}^{(q)}, \tau)$ to the corresponding BSs

BSs:

update the dual variables using $\mathbf{p}^{(q)}_j(\tau)$ and $\mathbf{n}^{(q)}_j(\tau)$, and broadcast them:

$\mu^{(q+1)}_i(\tau) = \left(\mu^{(q)}_i(\tau) + \delta^{(q)} \left(\sum_{j \in \mathcal{U}^{(i)}_D} n^{(q)}_{ji}(\tau) - n^{(i)}_D\right)\right)^\infty_0, \quad \forall i \in \mathcal{B}$

$\lambda^{(q+1)}_i(\tau) = \left(\lambda^{(q)}_i(\tau) + \delta^{(q)} \left(\sum_{j \in \mathcal{U}^{(i)}_D} p^{(q)}_{ji}(\tau) - \bar{P}^i_{\text{BS}}(C_i(\tau)) + P^{(i)}_V(\tau)\right)\right)^\infty_0$

until $\boldsymbol{\mu}^{(q+1)}(\tau)$ and $\boldsymbol{\lambda}^{(q+1)}(\tau)$ converge

quantization of $x_{ji}(\tau)$:

$\forall j \in \mathcal{U}_D \longrightarrow i^\star = \arg\max_{i \in \mathcal{B}} x_{ji}(\tau)$

$x^\star_{ji^\star}(\tau) = 1$, $x^\star_{ji}(\tau) = 0, \forall i \neq i^\star$

update batteries:

$C_i(\tau + 1) = (C_i(\tau) - E_i(\tau) + H_i(\tau))^{C^i_{\max}}_0$, $\forall i \in \mathcal{B}$

for all epochs $\tau \in \Upsilon$

end algorithm

7.3.4 Asymptotic Analysis of the Battery Evolution

In this section, we present the asymptotic behavior of the battery evolution of the system. This analysis will help us understand how the system works when the number of epochs grows up to infinity. We will assume that function $g_i(\cdot)$ is linear (or affine) w.r.t. the battery, and that the BSs do not have physical limitations ($P_{\text{BS}_i}^{\max} = \infty$), i.e., $g_i(C_i(\tau)) = \alpha \cdot C_i(\tau)$, $\forall i$. This implies that the lower limit of the battery is never reached as long as $\mathbb{E}[H_i(\tau)] > 0$. Let us consider that the harvesting is stationary (otherwise no convergence is guaranteed unless the time window where the harvesting intensity does not change is sufficiently large). We will also assume for simplicity in the development that the battery never reaches its maximum limit, $C_{\max}^i$ and, thus, the battery dynamics equation that governs the battery evolution is given by

$$C_i(\tau+1) = C_i(\tau) - E_i(\tau) + H_i(\tau) = (1-\alpha)C_i(\tau) + H_i(\tau), \tag{7.39}$$

where in the right hand side of (7.39) we have assumed that $E_i(\tau) = \alpha_i \cdot C_i(\tau)$ (this is true because constraint $C4$ in (7.13) is fulfilled with equality). Recall that $H_i(\tau)$ is a Bernoulli stochastic process such that its components $\{H_i(\tau)\}_\tau$ are identically distributed (i.i.d.) random variables and, thus, $C_i(\tau)$ is also a stochastic process. If $0 < \alpha_i < 1$, then the recursive relation in (7.39) is stable since the input, $H_i(\tau)$, is bounded and the solution to (7.39) is given by

$$C_i(\tau+1) = \sum_{j=0}^{\infty} (1-\alpha_i)^j H_i(\tau - j), \tag{7.40}$$

where the previous sum converges in quadratic mean since $|1-\alpha_i| < 0$ implies that $\sum_{j=0}^{\infty} |(1-\alpha_i)^j| = \frac{1}{\alpha_i} < \infty$. Given this, the expected value of the battery in convergence is, thus, given by

$$\lim_{\tau\to\infty} \mathbb{E}[C_i(\tau)] = \frac{\mathbb{E}[H_i(\tau)]}{\alpha_i} = \frac{p_i \cdot e_i}{\alpha_i}. \tag{7.41}$$

This implies that the expected energy allowed to be extracted from the battery is limited by the energy collected through harvesting, $\mathbb{E}[H_i(\tau)]$, as $\mathbb{E}[g(C_i(\tau))] = \mathbb{E}[H_i(\tau)]$. As a result, the expected performance in terms of throughput does not depend on the value of α_i that is configured or the initial battery level[4]. The variance of the expected battery in convergence is

[4]This statement is true if the upper limit of the battery is never reached or $B_i^{\max} = \infty$. Notice that, if $\alpha \approx 0$, then overflows may occur, incurring a loss of energy and, thus, a loss in throughput performance. Therefore, in reality, there is an optimum value of $\alpha_i^\star > 0$ that achieves the maximum throughput performance. Unfortunately, an analytical expression of the optimum value of α has not been found and numerical simulations are needed (see Section 3.4.4).

expressed as

$$\text{var}(C_i(\tau)) = \text{var}\left(\sum_{j=0}^{\infty}(1-\alpha_i)^j H_i(\tau - j)\right) \tag{7.42}$$

$$= \sum_{j=0}^{\infty}(1-\alpha_i)^{2j}\text{var}(H_i(\tau - j)) \tag{7.43}$$

$$= \sum_{j=0}^{\infty}(1-\alpha_i)^{2j} p_i(1-p_i)e_i^2 \tag{7.44}$$

$$= \frac{p_i(1-p_i)e_i^2}{1-(1-\alpha_i)^2}. \tag{7.45}$$

7.4 Part II: Ergodic-Based User Association Strategies

In this part of the chapter we propose a different type of user association technique based on the stochastic optimization method [Rib10b]. Also in this section, we assume some simplifications to the challenging issues presented in Section 7.2.5 that will allow us to handle the problem. In particular, assumptions *i*, *iii*, *iv*, and *v* from Section 7.3 will be also considered in this section. The only thing that changes w.r.t. the greedy approach is how we handle now assumption *ii*. Now, instead of time-decoupling the constraints and solving the association problem at each epoch by allowing to spend a given fraction of the battery, we are going to introduce some learning mechanism that will allow to take decisions on how much energy should a given BS spend for transmission given the past and current harvested energy and also on the rate of change of such harvesting process, that is, on the current and future dynamics of the harvesting process. As now we allow some time coupling, we no longer require to eliminate the time coupling from constraint $C1$ of problem (7.9). Constraints $C5$ and $C6$ will be changed by a different set of constraints. $C1$ will be reformulated accordingly and $C7$ will no longer be needed.

7.4.1 Problem Formulation

In this section, we describe the ergodic user association procedure. As we assumed before, the system performance under consideration is the sum-utility of the mean throughput of all data users in the system. Let us introduce the following vector definitions: $\tilde{\mathbf{R}} = \{\tilde{R}_j,\ j \in \mathcal{U}_D\}$ and $\tilde{\mathbf{P}} \triangleq \{\tilde{P}_i,\ i \in \mathcal{B}\}$. The association procedure is formulated through the following optimization

problem:

$$\underset{\substack{\tilde{\mathbf{R}},\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\\ \mathbf{n}(\tau),\mathbf{x}(\tau),\tilde{\mathbf{P}}}}{\text{maximize}} \quad \sum_{j\in\mathcal{U}_D} U_j\left(\tilde{R}_j\right) \tag{7.46}$$

$$\begin{aligned}
\text{subject to} \quad & C1: \tilde{R}_j \leq \frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon}\sum_{i\in\mathcal{B}} n_{ji}(\tau)\log_2\left(1+\frac{M_D p_{ji}(\tau)}{n_{ji}(\tau)(\theta(\tilde{P}_i-P_c^i)+A_{ji}(\tau))}\right), \quad \forall j\in\mathcal{U}_D \\
& C2: \sum_{i\in\mathcal{B}} \frac{M_V \check{p}_{ji}(\tau)}{\theta(\tilde{P}_i-P_c^i)+A_{ji}(\tau)} \geq \gamma_j, \quad \forall j\in\mathcal{U}_V, \forall\tau\in\Upsilon \\
& C3: \sum_{i\in\mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j\in\mathcal{U}_T, \forall\tau\in\Upsilon \\
& C4: \sum_{j\in\mathcal{U}_D} n_{ji}(\tau) \leq n_D^{(i)}, \quad \forall i\in\mathcal{B}, \forall\tau\in\Upsilon \\
& C5: \sum_{j\in\mathcal{U}_V^{(i)}} \check{p}_{ji}(\tau) + \sum_{k\in\mathcal{U}_D^{(i)}} p_{ki}(\tau) + P_{\text{CPICH}}^i + P_c^i \leq \tilde{P}_i, \quad \forall i\in\mathcal{B}, \forall\tau\in\Upsilon \\
& C6: \tilde{P}_i \leq \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_i(\tau), \quad \forall i\in\mathcal{B} \\
& C7: 0 \leq n_{ji}(\tau) \leq x_{ji}(\tau) n_D^{(i)}, \quad \forall i\in\mathcal{B}, \forall j\in\mathcal{U}_D, \forall\tau\in\Upsilon \\
& C8: 0 \leq \check{p}_{ji}(\tau) \leq x_{ji}(\tau)(\tilde{P}_i - P_{\text{CPICH}}^i - P_c^i), \quad \forall i\in\mathcal{B}, \forall j\in\mathcal{U}_V, \forall\tau\in\Upsilon \\
& C9: x_{ji}(\tau) = 0, \quad \forall i\notin\mathcal{S}_j(\tau), \forall j\in\mathcal{U}_T, \forall\tau\in\Upsilon \\
& C10: x_{ji}(\tau) \geq 0, \quad \forall i\in\mathcal{B}, \forall j\in\mathcal{U}_T, \forall\tau\in\Upsilon \\
& C11: p_{ji}(\tau) \geq 0, \quad \forall i\in\mathcal{B}, \forall j\in\mathcal{U}_D, \forall\tau\in\Upsilon,
\end{aligned}$$

where $|\Upsilon|$ is is the cardinality of set Υ. Variable $\tilde{P}_i$ is a measure of the average long-term power extracted from the battery. Notice that through constraint $C5$ we force the maximum power to be used during a particular epoch to be the same for all epochs. As it will be presented later, we will eventually allow such $\tilde{P}_i$ to have a dynamic behavior by letting it depend on time capturing the time evolution of the energy harvested and, therefore, the maximum power will vary over epochs, i.e., $\tilde{P}_i(\tau)$. In any case, note that in order to have a feasible solution $\tilde{P}_i \geq P_{\text{CPICH}}^i + P_c^i$. In this regard, if at any epoch, $\tilde{P}_i < P_{\text{CPICH}}^i + P_c^i$, then we assume that such BS is not available and, thus, it is turned off, until there is enough energy at the battery to be able to fulfill $\tilde{P}_i \geq P_{\text{CPICH}}^i + P_c^i$ (as we will present later in Section 7.4.3). Constraint $C6$ guarantees that, in the long term, the energy stored in the battery and obtained from harvesting is not lower than that taken from it. Hence, $C6$ is a relaxation of the dynamic battery equation (7.2). However, as it will be presented later in Section 7.4.3, the proposed algorithm will always fulfill (7.2).

Even after the simplifications considered when transforming problem (7.9) into problem (7.46), the remaining problem (7.46) is still quite challenging. First, note that it is time-coupled through constraints $C1$, $C5$, and $C6$. Second, it is not convex in $\{\tilde{P}_i\}$. To deal with the time coupling constraints, we will use stochastic approximation theory [Rib10b]. In order to deal

with the second issue, we will perform a two-step optimization[5], i.e., we first fix all $\{\tilde{P}_i\}$ and solve problem (7.46) and, then, optimize the variables $\{\tilde{P}_i\}$. With this approach, we divide the original optimization problem into two optimization problems that are solved sequentially. The fist optimization problem to be solved, that will be called the *inner problem*, is expressed as

$$\begin{aligned} \underset{\substack{\tilde{\mathbf{R}},\, \mathbf{p}(\tau),\, \check{\mathbf{p}}(\tau),\\ \mathbf{n}(\tau),\, \mathbf{x}(\tau)}}{\text{maximize}} \quad & \sum_{j \in \mathcal{U}_D} U_j\left(\tilde{R}_j\right) \\ \text{subject to} \quad & C1, \ldots, C5, C7, \ldots, C11 \text{ from problem (7.46).} \end{aligned} \tag{7.47}$$

Once we have solved the previous optimization problem and have obtained the optimum variables as a function of $\{\tilde{P}_i\}$, we then substitute the optimum values into the objective function in (7.46) and solve for the optimum $\{\tilde{P}_i\}$. Following this reasoning, the *outer problem* can be defined as

$$\begin{aligned} \underset{\tilde{\mathbf{P}}}{\text{maximize}} \quad & f(\boldsymbol{\Omega}^\star(\tilde{\mathbf{P}}), \tilde{\mathbf{P}}) \\ \text{subject to} \quad & \tilde{P}_i \leq \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau \in \Upsilon} H_i(\tau), \quad \forall i \in \mathcal{B}, \end{aligned} \tag{7.48}$$

where $\boldsymbol{\Omega}^\star(\tilde{\mathbf{P}}) = (\tilde{\mathbf{R}}^\star(\tilde{\mathbf{P}}),\, \mathbf{p}^\star(\tilde{\mathbf{P}}),\, \check{\mathbf{p}}^\star(\tilde{\mathbf{P}}),\, \mathbf{n}^\star(\tilde{\mathbf{P}}),\, \mathbf{x}^\star(\tilde{\mathbf{P}}))$ and $f(\cdot)$ is the objective function of problem (7.46) with the rest of the variables already optimized in the inner problem.

In the following sections, we will describe the mechanisms employed to solve the previous optimization problems (7.47) and (7.48).

7.4.2 Resolution of the Inner Problem (7.47)

Considering the powers $\{\tilde{P}_i\}$ fixed, the resulting optimization problem (7.47) is convex and has strong duality. However, the optimization problem is very difficult to solve as it is time-coupled since the objective is to maximize the average utility of the expected throughputs of the data users. Taking this into account, the idea is to develop a strategy that decouples the problem for each particular epoch. Without loss of optimality, we will first solve for the optimum long-term rates $\tilde{R}_j$ and then find the short-term optimization variables $(x_{ji}(\tau),\, p_{ji}(\tau),\, \check{p}_{ji}(\tau),\, n_{ji}(\tau))$. As the duality gap is zero, we employ a dual method to solve the problem. Let us start by dualizing constraint $C1$ of problem (7.47). For simplicity in the notation, let us use the following notation:

$$\bar{R}_{ji}(\tau) = n_{ji}(\tau)\log_2\left(1 + \frac{M_D p_{ji}(\tau)}{n_{ji}(\tau)(\theta\tilde{P}_i - \theta P_c^i + A_{ji}(\tau))}\right). \tag{7.49}$$

[5]An optimization problem of the form $\min_{x,y} f(x,y)$ can always be solved by solving first for the variable x and then for the variable y, or vice versa, i.e., $\min_{x,y} f(x,y) = \min_y \min_x f(x,y)$ even if $f(\cdot)$ is not jointly convex in x, y.

Additionally, let $\boldsymbol{\lambda} = \{\lambda_j, j \in \mathcal{U}_D\}$ be the vector of dual variables associated with constraint $C1$. Let us define $\bar{\mathbf{R}}(\tau) = \{\bar{R}_{ji}(\tau), j \in \mathcal{U}_D, i \in \mathcal{B}\}$. The partial Lagrangian is denoted as

$$\mathcal{L}_{C1}(\tilde{\mathbf{R}}, \bar{\mathbf{R}}(\tau); \boldsymbol{\lambda}) = -\sum_{j\in\mathcal{U}_D} U_j\left(\tilde{R}_j\right) + \sum_{j\in\mathcal{U}_D} \lambda_j \left(\tilde{R}_j - \frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon}\sum_{i\in\mathcal{B}} \bar{R}_{ji}(\tau)\right). \tag{7.50}$$

Setting the gradient of $\mathcal{L}_{C1}(\tilde{\mathbf{R}}, \bar{\mathbf{R}}(\tau); \boldsymbol{\lambda})$ to zero w.r.t. the variables $\tilde{R}_j$ yields

$$\nabla_{\tilde{R}_j}\mathcal{L}_{C1}(\tilde{\mathbf{R}}, \bar{\mathbf{R}}(\tau); \boldsymbol{\lambda}) = -\dot{U}_j(\tilde{R}_j) + \lambda_j = 0, \tag{7.51}$$

$$\tilde{R}_j^{\star}(\boldsymbol{\lambda}) = (\dot{U}_j)^{-1}(\lambda_j), \tag{7.52}$$

where $\dot{U}_j(\cdot)$ and $(\dot{U}_j)^{-1}(\cdot)$ denote the derivate of $U_j(\cdot)$ and the inverse function of $\dot{U}_j(\cdot)$, respectively. Given the dual variables $\boldsymbol{\lambda}$, the optimum average long-term rates $\tilde{\mathbf{R}}^{\star}(\boldsymbol{\lambda})$ are known and, thus, the maximization of the objective function in (7.47) is equivalent to the maximization of $\sum_{\tau\in\Upsilon}\sum_{j\in\mathcal{U}_D}\sum_{i\in\mathcal{B}} \lambda_j \bar{R}_{ji}(\tau)$ (where we have omitted the terms that do not depend on the remaining optimization variables), which can be clearly decomposed across epochs provided that the dual variables are known. The problem now resides in the computation of the optimum Lagrange multipliers.

If we solve the dual problem of (7.47) [Boy04], i.e., $\sup_{\boldsymbol{\Psi}} \inf_{\boldsymbol{\Xi}} \mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Psi})$ where $\boldsymbol{\Xi}$ contains all primal variables and $\boldsymbol{\Psi}$ all dual variables, and $\mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Psi})$ is the Lagrangian of problem (7.47) (see Appendix 7.A), the optimum $\boldsymbol{\lambda}$ could be found iteratively using a gradient approach as [Boy04]:

$$\lambda_j^{(q+1)} = \left(\lambda_j^{(q)} + \epsilon^{(k)}\left(\tilde{R}_j^{\star}\left(\boldsymbol{\lambda}^{(q)}\right) - \frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon}\sum_{i\in\mathcal{B}} \bar{R}_{ji}(\tau)\right)\right)_0^{\infty}, \quad \forall j \in \mathcal{U}_D, \tag{7.53}$$

where $\epsilon^{(k)}$ is the step size and q denotes the iteration index. Note that it is not possible to compute the value of the dual variables in real time as they depend on the specific values of $\bar{R}_{ji}(\tau)$, which are functions not known a priori (it is the solution of the optimization problem itself for all epochs τ). In this situation, we propose to follow a stochastic approximation strategy [Wan07], [Rib10b] and estimate the multipliers stochastically at each epoch as follows (with a noisy instantaneous unbiased estimate of the gradient)[6]:

$$\lambda_j(\tau+1) = \left(\lambda_j(\tau) + \epsilon^{(k)}\left(\tilde{R}_j^{\star}(\boldsymbol{\lambda}(\tau)) - \sum_{i\in\mathcal{B}} \bar{R}_{ji}(\tau)\right)\right)_0^{\infty}, \quad \forall j \in \mathcal{U}_D. \tag{7.54}$$

Now, we proceed to obtain the rest of the optimization variables. The optimization problem

[6]Note that this philosophy is similar to the instantaneous estimation of the gradient in the LMS algorithm [Hay02].

(7.47) can now be rewritten as

$$\begin{aligned} \underset{\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau)}{\text{maximize}} \quad & \sum_{j\in\mathcal{U}_D}\sum_{i\in\mathcal{B}}\lambda_j(\tau)\bar{R}_{ji}(\tau) \\ \text{subject to} \quad & C2,\ldots,C5,C7,\ldots,C11 \text{ from problem (7.46).} \end{aligned} \tag{7.55}$$

Notice that problem (7.55) is not time-coupled and standard optimization techniques can be employed to find the optimum variables for each epoch. Problem (7.55) is, thus, solved at the beginning of each particular epoch executing the user association procedure independently for each epoch. In Appendix 7.A, we develop a procedure based on a primal-dual approach to obtain the optimum variables. From now on, we will assume that all optimization variables are found as a function of the estimated dual variables $\{\lambda_j(\tau)\}$ and the powers $\{\tilde{P}_i\}$.

7.4.3 Resolution of the Outer Problem (7.48)

Having obtained the solution of the inner problem for a fix set $\{\tilde{P}_i\}$, we now proceed to solve the outer problem (7.48). Unfortunately, function $f(\cdot)$ is very difficult to characterize and no convexity claim can be conveyed. The only information that we know is that function $f(\cdot)$ increases monotonically in each individual argument $\tilde{P}_i$. However, in order to apply a primal-dual approach, the only property that needs to be guaranteed is the nullity of the duality gap. Let us introduce the following result.

Proposition 7.1. *Assuming that the optimum of the Lagrangian and the dual problem of problem* (7.48) *are attained at a point where the derivative of the Lagrangian and the dual function are zero, the duality gap of the outer problem* (7.48) *is zero.*

Proof. See Appendix 7.B. ■

The validity of the null duality gap will be validated through simulations in the simulations section. Based on the previous proposition, we can solve the outer problem by applying a primal-dual approach. Let us introduce the Lagrangian of problem (7.48):

$$\mathcal{L}(\tilde{\mathbf{P}};\boldsymbol{\pi}^1) = -f(\tilde{\mathbf{P}}) + \sum_{i\in\mathcal{B}}\pi_i^1\left(\tilde{P}_i - \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon}H_i(\tau)\right) \tag{7.56}$$

As it was mentioned before, the full characterization of function $f(\cdot)$ is very difficult, but we can still obtain the optimum powers $\tilde{\mathbf{P}}^\star$ thanks to sensitivity theory [Boy04]. The gradient of the Lagrangian must vanish for the optimum values of $\tilde{\mathbf{P}}^\star$ and $\boldsymbol{\pi}^{1\star}$, and, therefore, we can employ a gradient procedure to find the optimum variables [Ber99]. Following a gradient strategy, the

update equation for the power of the i-th BS is given by

$$\tilde{P}_i^{(k+1)} = \tilde{P}_i^{(k)} - \delta^{(k)} \left(-\frac{\partial f(\tilde{\mathbf{P}}^{(k)})}{\partial \tilde{P}_i} + \pi_i^1 \right), \tag{7.57}$$

where $\delta^{(k)}$ is the step size. We can now make use of the extended sensitivity theory presented in Appendix 7.C in order to be able to compute $\frac{\partial f(\tilde{\mathbf{P}})}{\partial \tilde{P}_i}$. Let us present the following result:

Proposition 7.2. *The partial derivative* $-\frac{\partial f(\tilde{\boldsymbol{P}})}{\partial \tilde{P}_i}$ *is given by*

$$\begin{aligned} -\frac{\partial f(\tilde{\boldsymbol{P}})}{\partial \tilde{P}_i} = & \sum_{j \in \mathcal{U}_D} \lambda_j^\star \frac{\partial f_{1j}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right)}{\partial \tilde{P}_i} + \sum_{j \in \mathcal{U}_D} \nu_j^\star \frac{\partial f_{2j}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right)}{\partial \tilde{P}_i} \\ & + \pi_i^{2\star} \frac{\partial f_{5i}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right)}{\partial \tilde{P}_i} + \sum_{j \in \mathcal{U}_D} \xi_{ji}^\star \frac{\partial f_{8ji}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right)}{\partial \tilde{P}_i}, \end{aligned} \tag{7.58}$$

where

$$\begin{aligned} & f_{1j}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right) = \\ & \quad \tilde{R}_j^\star(\tilde{\boldsymbol{P}}) - \frac{1}{|\Upsilon|} \sum_{\tau \in \Upsilon} \sum_{i \in \mathcal{B}} n_{ji}^\star(\tau, \tilde{\boldsymbol{P}}) \log_2 \left(1 + \frac{M_D p_{ji}^\star(\tau, \tilde{\boldsymbol{P}})}{n_{ji}^\star(\tau, \tilde{\boldsymbol{P}})(\theta \tilde{P}_i - \theta P_c^i + A_{ji}(\tau))} \right), \end{aligned} \tag{7.59}$$

$$f_{2j}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right) = \gamma_j - \sum_{i \in \mathcal{B}} \frac{M_V \check{p}_{ji}^\star(\tau, \tilde{\boldsymbol{P}})}{\theta \tilde{P}_i - \theta P_c^i + A_{ji}(\tau)}, \tag{7.60}$$

$$f_{5i}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right) = \sum_{j \in \mathcal{U}_D} p_{ji}^\star(\tau, \tilde{\boldsymbol{P}}) + \sum_{j \in \mathcal{U}_V} \check{p}_{ji}^\star(\tau, \tilde{\boldsymbol{P}}) + P_{\text{CPICH}}^i + P_c^i - \tilde{P}_i, \tag{7.61}$$

$$f_{8ji}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right) = \check{p}_{ji}^\star(\tau, \tilde{\boldsymbol{P}}) - x_{ji}^\star(\tau, \tilde{\boldsymbol{P}})(\tilde{P}_i - P_{\text{CPICH}}^i - P_c^i), \tag{7.62}$$

and

$$\begin{aligned} & \frac{\partial f_{1j}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right)}{\partial \tilde{P}_i} = \\ & \quad \frac{M_D \theta}{|\Upsilon| \log(2)} \sum_{\tau \in \Upsilon} \frac{n_{ji}^\star(\tau, \tilde{\boldsymbol{P}}) p_{ji}^\star(\tau, \tilde{\boldsymbol{P}})}{n_{ji}^\star(\tau, \tilde{\boldsymbol{P}})(\theta \tilde{P}_i - \theta P_c^i + A_{ji}(\tau))^2 + M_D p_{ji}^\star(\tau, \tilde{\boldsymbol{P}})(\theta \tilde{P}_i - \theta P_c^i + A_{ji}(\tau))} \end{aligned} \tag{7.63}$$

$$\frac{\partial f_{2j}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right)}{\partial \tilde{P}_i} = \frac{M_V \check{p}_{ji}^\star(\tau, \tilde{\boldsymbol{P}}) \theta}{(\theta \tilde{P}_i - \theta P_c^i + A_{ji}(\tau))^2}, \tag{7.64}$$

$$\frac{\partial f_{5i}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right)}{\partial \tilde{P}_i} = -1, \tag{7.65}$$

$$\frac{\partial f_{8ji}\left(\boldsymbol{\Omega}^\star(\tilde{\boldsymbol{P}}), \tilde{\boldsymbol{P}}\right)}{\partial \tilde{P}_i} = -x_{ji}^\star(\tau, \tilde{\boldsymbol{P}}), \tag{7.66}$$

being $\{\nu_j\}$, $\{\pi_i^2\}$, $\{\xi_{ji}\}$ *the sets of dual variables associated with constraints* $C2$, $C5$, *and* $C8$ *of problem* (7.47).

Proof. The proof is a direct application of the result derived in Appendix 7.C concerning perturbation analysis in optimization problems. ■

Based on the previous derivation, we can express the update equation $\tilde{P}_i^{(k+1)}$ as

$$\tilde{P}_i^{(k+1)} = \tilde{P}_i^{(k)} + \delta^{(k)} \left(\pi_i^2 - \pi_i^1 - \mathcal{C}(\tilde{P}_i) \right), \tag{7.67}$$

where the term $\mathcal{C}(\tilde{P}_i)$ is defined as $\mathcal{C}(\tilde{P}_i) \triangleq \sum_{j \in \mathcal{U}_D} \lambda_j^\star \frac{\partial f_{1j}(\mathbf{\Omega}^\star(\tilde{\mathbf{P}}), \tilde{\mathbf{P}})}{\partial \tilde{P}_i} + \sum_{j \in \mathcal{U}_D} \nu_j^\star \frac{\partial f_{2j}(\mathbf{\Omega}^\star(\tilde{\mathbf{P}}), \tilde{\mathbf{P}})}{\partial \tilde{P}_i} + \sum_{j \in \mathcal{U}_D} \xi_{ji}^\star \frac{\partial f_{8ji}(\mathbf{\Omega}^\star(\tilde{\mathbf{P}}), \tilde{\mathbf{P}})}{\partial \tilde{P}_i}$.

Notice that the optimum value of the dual variables $\{\pi_i^1\}$ could be found with a gradient strategy applied to the dual problem just as we explained in the previous section when dealing with the dual variables $\{\lambda_j\}$:

$$\pi_i^{1(k+1)} = \left(\pi_i^{1(k)} + \epsilon^{(k)} \left(\tilde{P}_i - \frac{1}{T_e} \frac{1}{|\Upsilon|} \sum_{\tau \in \Upsilon} H_i(\tau) \right) \right)_0^\infty, \quad \forall i \in \mathcal{B}. \tag{7.68}$$

Note, however, that we need to know in advance the whole realization of the stochastic harvesting process $H_i(\tau)$ to compute $\{\pi_i^1\}$, and that is not possible. For this reason, also in this case, we propose to estimate the dual variables $\{\pi_i^1\}$ using stochastic approximation theory as

$$\pi_i^1(\tau + 1) = \left(\pi_i^1(\tau) + \epsilon^{(k)} \left(\tilde{P}_i - \frac{1}{T_e} H_i(\tau) \right) \right)_0^\infty, \quad \forall i \in \mathcal{B}. \tag{7.69}$$

Notice that we have substituted the true gradients by the instantaneous noisy gradients as it is done in the classical LMS algorithm. The remaining dual variables $\boldsymbol{\nu}$, $\boldsymbol{\pi}^2$, and $\boldsymbol{\xi}$ are obtained directly from the inner optimization problem (7.47), so there is no need to estimate them. The dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\pi}^1$ that are estimated using $\boldsymbol{\lambda}(\tau)$ and $\boldsymbol{\pi}^1(\tau)$ will converge to values sufficiently close to the optimum ones (see for example [Rib10b]). Because the dual variables $\boldsymbol{\lambda}(\tau)$ and $\boldsymbol{\pi}^1(\tau)$ are estimated at each epoch, the power to be radiated, $\tilde{P}_i$, now depends on the particular epoch τ, i.e., $\tilde{P}_i(\tau)$. Accordingly, $\tilde{P}_i(\tau)$ will be the maximum available power to be used during one particular epoch. In this way, we propose to modify the available power in (7.67) by updating $\tilde{P}_i(\tau)$ at each epoch according to the following rule:

$$\tilde{P}_i(\tau + 1) = \tilde{P}_i(\tau) + \delta^{(k)} \left(\pi_i^2(\tau) - \pi_i^1(\tau) - \mathcal{C}(\tilde{P}_i(\tau)) \right). \tag{7.70}$$

The previous rule has a nice interpretation. For the moment let us not consider the term $\mathcal{C}(\tilde{P}_i(\tau))$. Consider $\pi_i^1(\tau)$ as the price of the long-term power and $\pi_i^2(\tau)$ as the price of the short-term power. If a lot of energy is being captured thanks to harvesting, the price of the long-term power decreases, i.e., $\pi_i^1(\tau)$ decreases (see (7.69)), and as a consequence $\tilde{P}_i(\tau + 1)$ increases (to allow using more power in the current epoch). If, on the other hand, we are using more energy than the one that is being harvested, $\pi_i^1(\tau)$ increases and $\tilde{P}_i(\tau + 1)$ decreases. The

same reasoning can be done for the interpretation of $\pi_i^2(\tau)$. The tighter the constraint $C5$ of problem (7.47) is, the larger the value of the dual variable $\pi_i^2(\tau)$ will be (see (7.88) in Appendix 7.A) and, thus, the the value of $\tilde{P}_i(\tau+1)$ will increase in order to make more power available. Notice that if the harvesting is relatively large compared to the energy spent in a particular epoch, $\pi_i^1(\tau)$ could have a value such that $\tilde{P}_i(\tau+1)$ may be larger than the battery capacity. For this reason, we force $\tilde{P}_i(\tau+1)$ to lie within the interval $[0, P^i_{\max}(\tau)]$, i.e.,

$$\tilde{P}_i(\tau+1) = \left(\tilde{P}_i(\tau) + \delta^{(k)}\left(\pi_i^2(\tau) - \pi_i^1(\tau) - \mathcal{C}(\tilde{P}_i)\right)\right)_{P^i_{\text{CPICH}}+P^i_c}^{P^i_{\max}(\tau+1)}, \tag{7.71}$$

where $P^i_{\max}(\tau) = \min\left\{\frac{1}{T_e}C_i(\tau), P^{\max}_{\text{BS}_i} + P^i_{\text{CPICH}} + P^i_c\right\}$ is the maximum power that can be radiated during epoch τ. Note that if $P^i_{\max}(\tau+1) < P^i_{\text{CPICH}} + P^i_c$ then no feasible value for $\tilde{P}_i(\tau+1)$ can be found. In that case, we shut down the i-th BS until it has enough energy in the battery. It is interesting to note that there is an alternative definition of $\pi_i^1(\tau)$ that is directly related to the current battery level of the BS. First, notice that constraint $C5$ in (7.47) will be tight at the optimum. Thus, $T_e\tilde{P}_i(\tau)$ will be the energy extracted from the battery at the τ-th epoch. As a consequence, we can alternatively define the iteration of $\pi_1^i(\tau)$ as

$$\pi_i^1(\tau+1) = \kappa_1(\tau) - \kappa_2 C_i(\tau+1), \tag{7.72}$$

where $\kappa_1 = \pi_i^1(\tau) + \frac{\delta}{T_e}C_i(\tau)$ and $\kappa_2 = \delta T_e$. The more energy is available at the battery, the smaller the price of the long-term power will be and vice versa. In other words, the stochastic Lagrange multipliers are scaled and biased versions of the batteries, so that when the battery $C_i(\tau)$ decreases, the price $\pi_i^1(\tau)$ increases.

Finally, as $\tilde{P}_i(\tau)$ now depends on τ, the partial derivative of $f_{1j}(\cdot)$ in (7.63) is now expressed as

$$\frac{\partial f_{1j}\left(\boldsymbol{\Omega}^\star(\tilde{\mathbf{P}}), \tilde{\mathbf{P}}\right)}{\partial \tilde{P}_i(\tau)} = \tag{7.73}$$
$$\frac{M_D\theta}{|\Upsilon|\log(2)} \frac{n^\star_{ji}(\tau,\tilde{\mathbf{P}})p^\star_{ji}(\tau,\tilde{\mathbf{P}})}{n^\star_{ji}(\tau,\tilde{\mathbf{P}})(\theta\tilde{P}_i(\tau) - \theta P^i_c + A_{ji}(\tau))^2 + M_D p^\star_{ji}(\tau,\tilde{\mathbf{P}})(\theta\tilde{P}_i(\tau) - \theta P^i_c + A_{ji}(\tau))}.$$

7.4.4 Overall User Association Algorithm

In this section we present the overall user association algorithm implemented to solve the optimization problem in (7.46) based on stochastic approximation theory. In the previous section we found the optimal primal variables based on a two-stage optimization from which we obtained the optimum association variables $\mathbf{x}^\star(\tau)$. Recall, however, that we relaxed the association variables x_{ji} in order to make the optimization problem convex allowing each user to be able to connect to multiple BSs simultaneously, so the solution will provide, in general, a value of $x^\star_{ji}(\tau) \in [0,1]$. At this point, we propose to quantize such association variables as, for the actual implementation

of the user association, we require $x^{\star}_{ji}(\tau) \in \{0,1\}$ with $\sum_{i\in\mathcal{B}} x^{\star}_{ji}(\tau) = 1$, $\forall j \in \mathcal{U}_T$.

The optimal association policy for any user j can be written in closed form using the indicator function as

$$x^{\star}_{ji}(\tau) = \mathbb{1}_{\{i=\arg\max_{i'}\{x_{ji'}(\tau)\}\}}. \tag{7.74}$$

If several BSs provide the same $x^{\star}_{ji}(\tau)$ before quantization for a given user, we select one of the BSs randomly. The overall algorithm that solves the inner and the outer optimization problem is detailed in Algorithm 7.3.

Algorithm 7.3 Algorithm for solving ergodic user association problem (7.46)

repeat (index τ)
initialize $\boldsymbol{\lambda}(0) \succeq 0$, $\boldsymbol{\pi}^1(0) \succeq 0$, $\tilde{R}^{\star}_j(\boldsymbol{\lambda}) = (\dot{U}_j)^{-1}(\lambda_j(0))$, and $\tilde{P}_i(0) = \frac{\mathbb{E}[H_i(\tau)]}{T_e}$
solve optimization problem (7.55) with $\boldsymbol{\lambda}(\tau)$ and $\tilde{\mathbf{P}}(\tau)$, and obtain:
primal variables: $p^{\star}_{ji}(\tau)$, $\check{p}^{\star}_{ji}(\tau)$, $n^{\star}_{ji}(\tau)$, and $x^{\star}_{ji}(\tau)$
dual variables: $\boldsymbol{\nu}(\tau)$, $\boldsymbol{\pi}^2(\tau)$, and $\boldsymbol{\xi}(\tau)$
quantization of $x_{ji}(\tau)$:
$\forall j \in \mathcal{U}_T \longrightarrow i^{\star} = \arg\max_{i\in\mathcal{B}} x_{ji}(\tau)$
$x^{\star}_{ji^{\star}}(\tau) = 1$, $x^{\star}_{ji}(\tau) = 0, \forall i \neq i^{\star}$
update stochastic dual variables:
$\lambda_j(\tau+1) = \left(\lambda_j(\tau) + \epsilon^{(k)}\left(\tilde{R}^{\star}_j(\boldsymbol{\lambda}(\tau)) - \sum_{i\in\mathcal{B}} \bar{R}_{ji}(\tau)\right)\right)_0^{\infty}, \quad \forall j \in \mathcal{U}_D$
$\pi^1_i(\tau+1) = \left(\pi^1_i(\tau) + \epsilon^{(k)}\left(\tilde{P}_i(\tau) - \frac{1}{T_e}H_i(\tau)\right)\right)_0^{\infty}, \quad \forall i \in \mathcal{B}$
update battery and maximum radiated power:
$C_i(\tau+1) = (C_i(\tau) - E_i(\tau) + H_i(\tau))_0^{C^i_{\max}}, \quad \forall i \in \mathcal{B}$
$P^i_{\max}(\tau+1) = \min\left\{\frac{1}{T_e}C_i(\tau+1), P^{\max}_{\mathrm{BS}_i} + P^i_{\mathrm{CPICH}} + P^i_c\right\}, \quad \forall i \in \mathcal{B}$
update primal variables:
$\tilde{R}^{\star}_j(\boldsymbol{\lambda}) = (\dot{U}_j)^{-1}(\lambda_j(\tau+1)), \quad \forall j \in \mathcal{U}_D$
$\tilde{P}_i(\tau+1) = \left(\tilde{P}_i(\tau) + \delta^{(k)}\left(\pi^2_i(\tau) - \pi^1_i(\tau) - \mathcal{C}(\tilde{P}_i(\tau))\right)\right)_{P^i_{\mathrm{CPICH}}+P^i_c}^{P^i_{\max}(\tau+1)}, \quad \forall i \in \mathcal{B}$
for all epochs $\tau \in \Upsilon$
end algorithm

7.5 Numerical Simulations

In this section we evaluate the performance of the proposed strategy. The scenario under consideration is composed of 15 BSs with 4 tiers. The deployment layout is shown in Figure 7.2, where we also show the number of available BSs at each point of the scenario (at a given

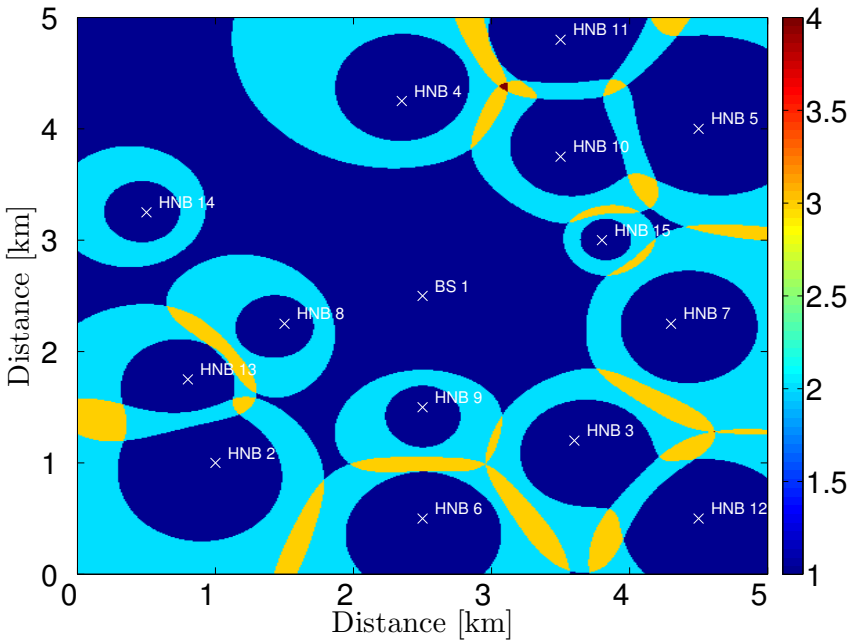
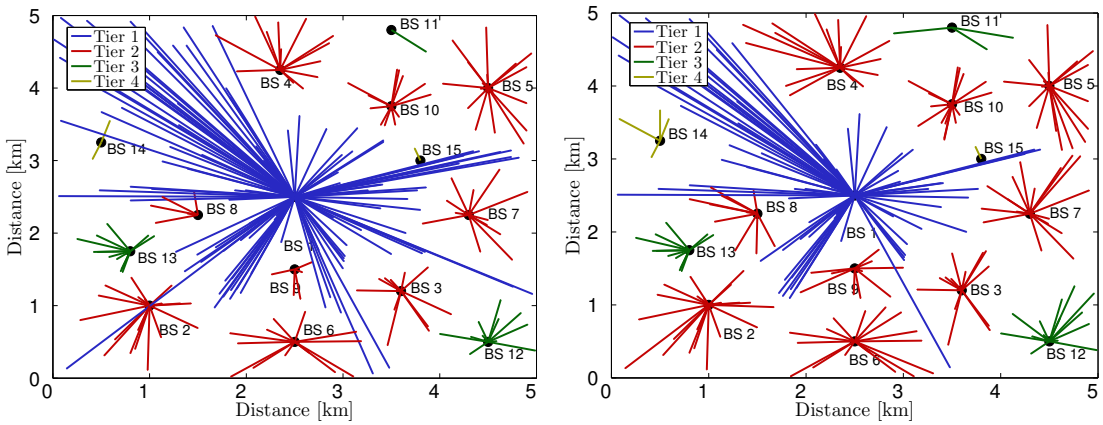


Figure 7.2: Reference scenario and available BSs at a given epoch. The color bar represents the number of BSs available at each point.

epoch). BS 1 belongs to tier 1, BSs $\{2,\ldots,10\}$ belong to tier 2, BSs $\{11,\ldots,13\}$ belong to tier 3, and BSs $\{14,15\}$ belong to tier 4. The maximum radiated power, $P_{\mathrm{BS}_i}^{\max}+P_{\mathrm{CPICH}}^{i}$, is 46 dBm, 24 dBm, 20 dBm, and 13 dBm for each tier. The pilot power, P_{CPICH}^{i}, is 5% of $P_{\mathrm{BS}_i}^{\max}$. The fixed power, P_c^i, is 33 dBm, 17 dBm, 13 dBm, and 6 dBm for each tier. The system contains 85 data users and 15 voice users. All the users in the system are mobile with a speed of 4 km/h. The instantaneous channel, $\tilde{h}_{ji}(t)$, incorporates Rayleigh fading with unitary power and a path loss based on Okumura-Hata for open areas. The orthogonality factor is $\theta = 0.35$ [Awo03]. The code gain of data codes is $M_D = 16$ and the minimum SINR normalized by gain is $\frac{\gamma_j}{M_V} = -13.7$ dB. The minimum SINR for pilot signals is $\gamma_{\mathrm{CPICH}} = -20$ dB. The noise power is $\sigma^2 = -102$ dBm. The battery capacities, $C_{\max}^i$, are $1,200$ J, 30 J, 12 J, and 2.5 J, for each tier and the quantity of energy in an energy harvesting packet is, $e_i = C_{\max}^i/10$ J. The number of epochs considered is 300 and the time between epochs is $T_{\mathrm{e}} = 10$s. The utility function is $U_j(\cdot) = \log(\cdot)$. In the simulations, for each particular user association, PF scheduling with instantaneous channels is run at each particular BS for assigning the instantaneous resources. Thus, in this section, instantaneous data rates will be evaluated and shown in the figures and the batteries will be updated with the actual (instantaneous) powers radiated by the BSs. For a more complete description of the deployment, see [tuc14].

7.5.1 Numerical Simulations of the User Association Strategy from Section 7.3

In the max-SINR strategy, the users are associated with the strongest BS in terms of received pilot power without considering the actual load of the BS nor its current battery level. Figure

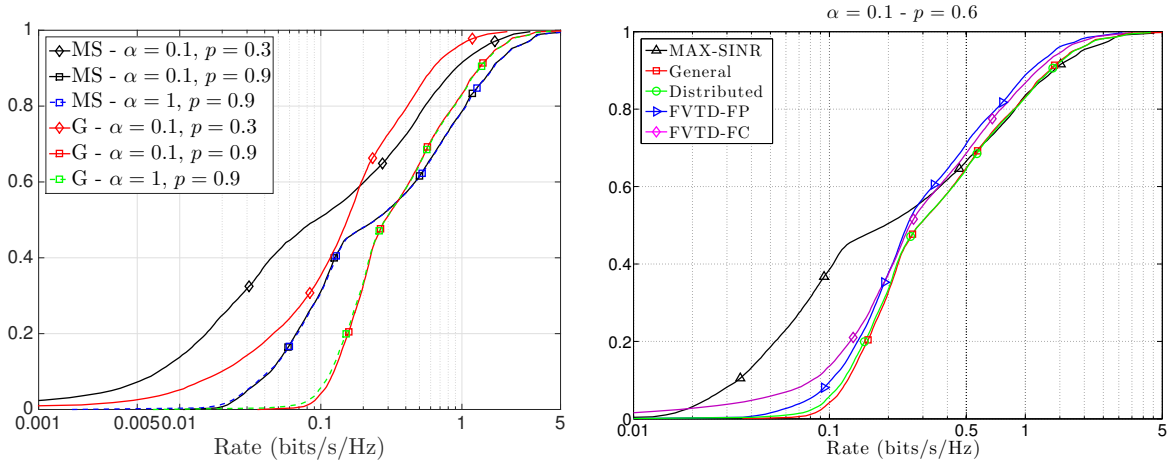


(a) User association policy: max-SINR (b) User association policy: problem in (7.9)

Figure 7.3: Snapshot of the user association for different policies.

7.3 presents the user association for the first epoch, considering that the batteries are full and $\alpha = 0.1$. As we can conclude from the figures, some of the users connected to the BS1 (macro BS) are transferred to smaller cells if a proper load balancing mechanism is considered (see for example BS8, BS11, BS14). If we compare the number of users associated to the macro BS (BS1) for both approaches, i.e., max-SINR and our strategy, we conclude from the figures that, with the max-SINR approach, the macro is serving 40 more users than with the proposed strategy. This makes the macro BS saturate and assign very little resources to the users while the small cells are operating with few users.

Figure 7.4-(a) presents the CDF of the rates for the strategy proposed in this chapter in (7.13), denoted as 'G' (or 'General' in some subsequent figures) in the legend, and the maximum SINR association strategy, denoted as 'MS', for different values of α and harvesting p. Notice that $\alpha = 1$ allows to use all the battery at each particular epoch. As we can see, the proposed scheme provides two to three times higher rates in the low-rate regime compared to the max-SINR strategy and outperforms the traditional approach up to rates in the order of 0.4 bits/s/Hz, which represents around 70% of the user rates. This means that the proposed approach provides a fair load balancing network in terms of user distribution which is translated into an improvement of the overall network throughput. However, this comes with a small reduction in terms of peak rates (lower than 20%). An interesting insight is that for larger energy harvested values, the influence of the value of α is lower. However, if more energy is available, a better load balancing is achieved. This is due to the fact that small cells have more power and the coverage radius are larger offering more possibilities for the users to associate to BSs. Contrarily, for larger values of α we obtain larger peak rates, but a worse balanced network. Thus, from a load-balancing perspective, it is better to control how the energy is being used at different epochs.



(a) CDF of the user rates for different energy harvesting intensities ($p = 0.3$, $p = 0.6$, and $p = 0.9$) and the value of α for the proposed strategy and the max-SINR strategy.

(b) CDF of the user rates for the different proposed strategies and the max-SINR.

Figure 7.4: CDF of the instantaneous user rates for different user association strategies.

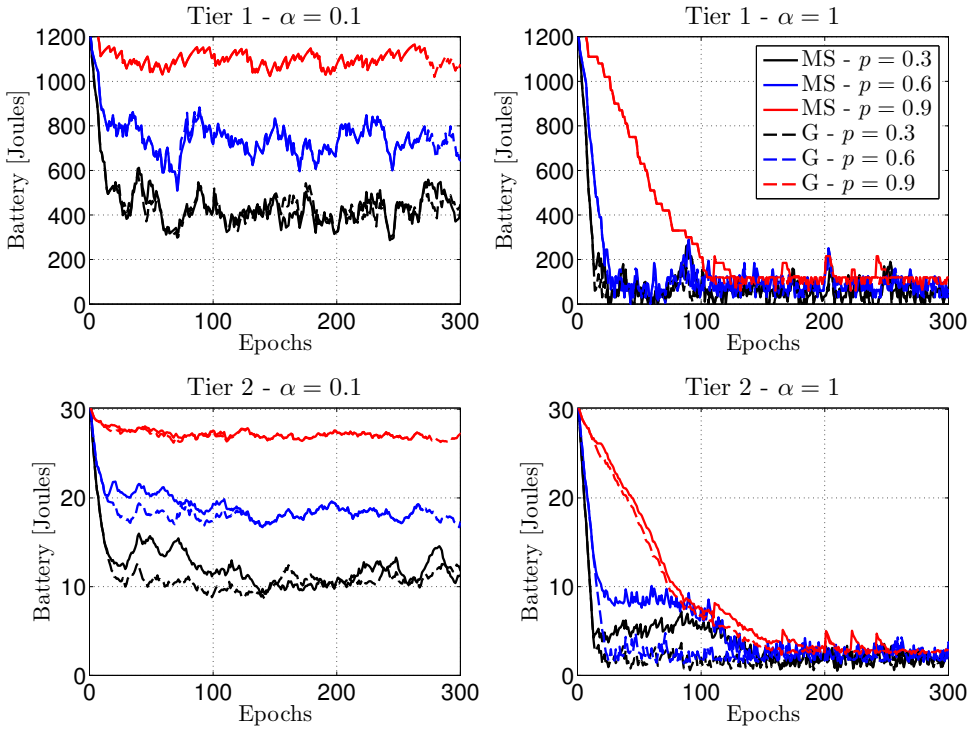


Figure 7.5: Average battery evolution (among BSs) of tier 1 and tier 2.

Figure 7.4-(b) presents the CDF of the rates of the data users with $\alpha = 0.1$ and $p = 0.6$ for the low-complexity approaches derived from (7.34) and (7.35), the distributed solution from (7.36), and the max-SINR. In this case, the distributed solution yields the same performance as the FVTD strategy (even though it is not shown in the figure). This is expected as both problems attempt to solve the same optimization problem. As we can see, even if we use the simple low-complexity solution fixing the power or the codes to be allocated, the performance

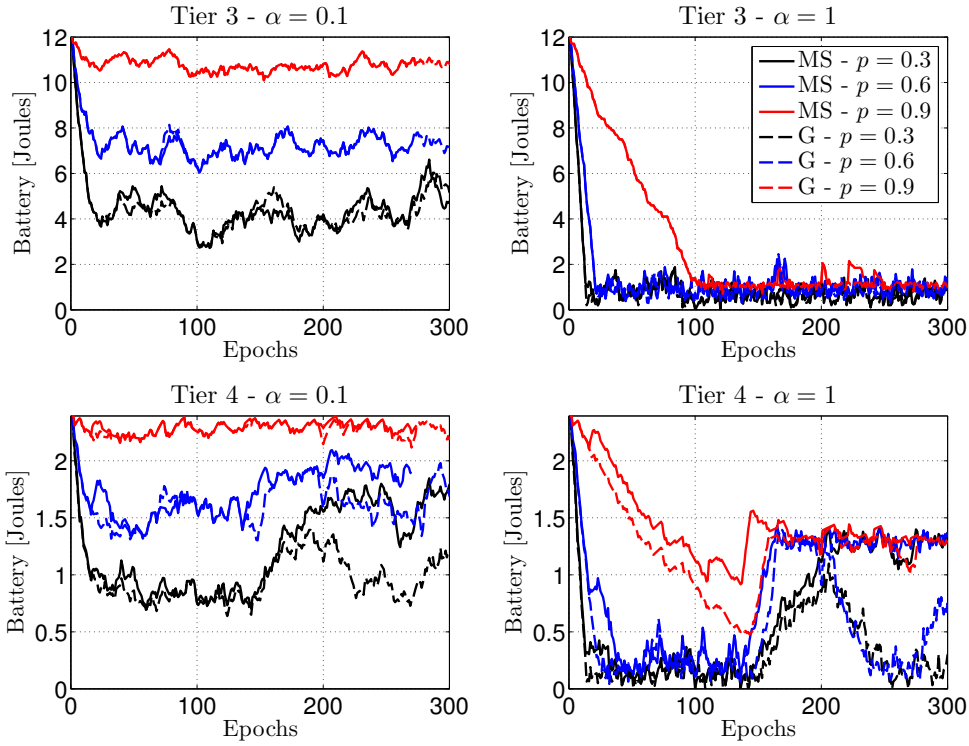


Figure 7.6: Average battery evolution (among BSs) of tier 3 and tier 4.

obtained is quite close to the rest of the strategies proposed in this chapter. All the strategies outperform the max-SINR providing a better balanced network.

Figure 7.5 and Figure 7.6 depict the battery evolution of the BSs in tier 1, tier 2, tier 3, and tier 4 for different values of harvesting p and α considering the traditional max-SINR ('MS') approach and the proposed strategy ('G'). As expected, if the harvesting increases, the expected battery also increases. Also the value of α impacts the battery evolution. For larger values of α we obtain lower residual battery levels at convergence (after the transient period). In fact, it can be shown that if the battery converges in average terms, i.e., $0 < \lim_{\tau\to\infty} \mathbb{E}[C_i(\tau)] < B^i_{\max}$, then, as we mentioned before, $\lim_{\tau\to\infty} \mathbb{E}[C_i(\tau)] = \frac{\mathbb{E}[H_i(\tau)]}{\alpha_i} = \frac{p_i \cdot e_i}{\alpha_i}$. The variance can also be checked to fulfill the expression in (7.45). Notice, however, that this is not always the behavior of tier 4. Recall that, through the development in Section 7.3.4, we assumed that constraint $C4$ was attained with equality. This will only happen if there is at least one data user connected to any BS of the tier. Because there are just two BSs in tier 4, it is very likely that no data users are within the coverage area of these two BSs at a given epoch. In fact, in the figure, the battery evolution experiences an abrupt jump because no users were connected to those BSs, and the harvested energy increased the battery levels.

The evolution of the average number of users associated to a particular tier as time evolves (average among the BSs of the same tier) is shown in Figure 7.7 and Figure 7.8 for different values of α, different harvesting intensities, and different association strategies. Each vertical cut of these plots would yield the snapshot of the user association presented in Figure 7.3 (epoch

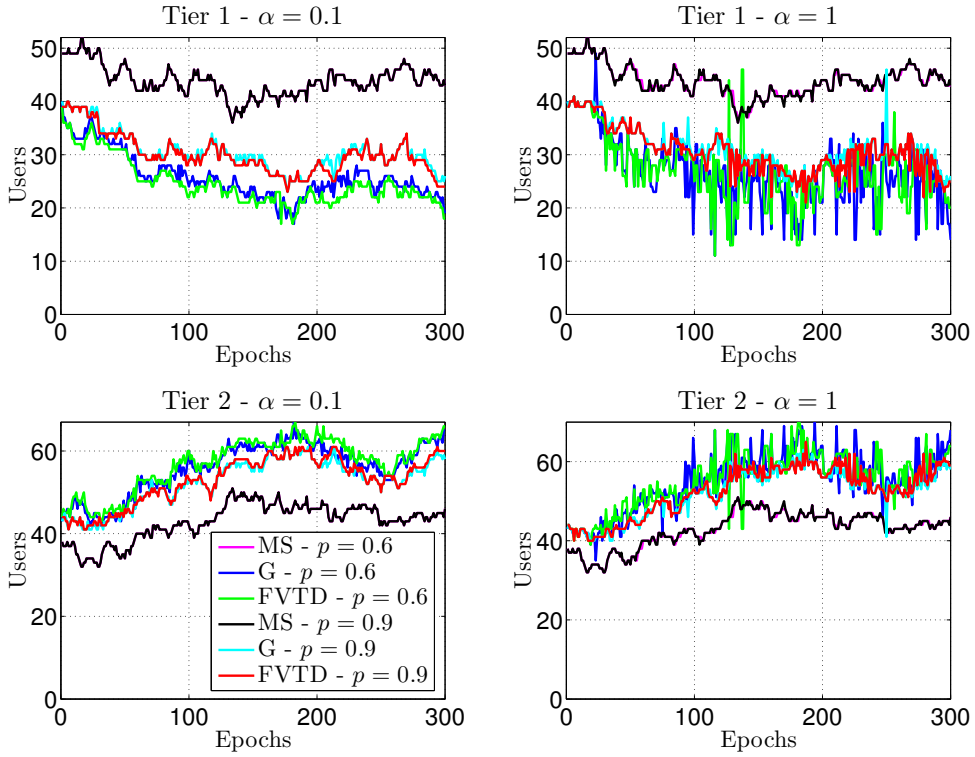


Figure 7.7: Evolution of user association in tier 1 and tier 2.

1 was considered in those previous plots). The main insight that we obtain from these figures is that, the max-SINR strategy tends to associate more users to the macro BS (tier 1) than the proposed strategies. As time evolves, with the max-SINR strategy, the macro BS tends to reduce its battery level (as it is widely used) and the number of users that associate with it decreases a little. This effect is more notable with the proposed strategies. If we have a look at tier 2, for example, the number of users increases as time evolves. In general terms, users associated to the macro BS when the max-SINR strategy is considered are transferred to other tiers if a load balancing technique is implemented. Users associated to tier 4 are almost double in average terms when comparing max-SINR strategy with the proposed one. In terms of the macro BS, there are roughly 20 users less in average, if we compare the proposed approach with the max-SINR strategy. The solutions of the user association for the general case derived in (7.13) and the FVTD strategy are quite similar. Another insight is that for larger values of α we see that the fluctuation of the curves increases. This is because the batteries of the BSs tend to run out of energy and users tend to associate to other BSs more frequently.

Since the voice users require a fixed rate service, the association problem may not be feasible under some circumstances. If the remaining energy in the batteries is low, and the minimum SINR requested is high, the overall problem may not have a feasible solution. In such a case, voice users may be dropped from the system, or the service may be reduced (i.e., the rate may be reduced by means of reducing the minimum target SINR). As the FVTD algorithm solves the voice users first in a greedy manner, following a max-SINR strategy, it will produce more infeasibilities than the general approach where the overall association of the voice users is

considered jointly along with the data users. The number of infeasibilities will, of course, depend on the value of α and the harvesting intensities. Figure 7.9 shows the percentage of feasibilities produced by each strategy for different values of α and harvesting intensity. As we can see from the figure, the best strategy in terms of feasibility avoidance is the general case and the worst is the FVTD strategy. For a moderate value of harvesting intensity, infeasibilities almost never occur, but if the energy collected is not so high or the value of α is too large, then, infeasibilities start occurring too often.

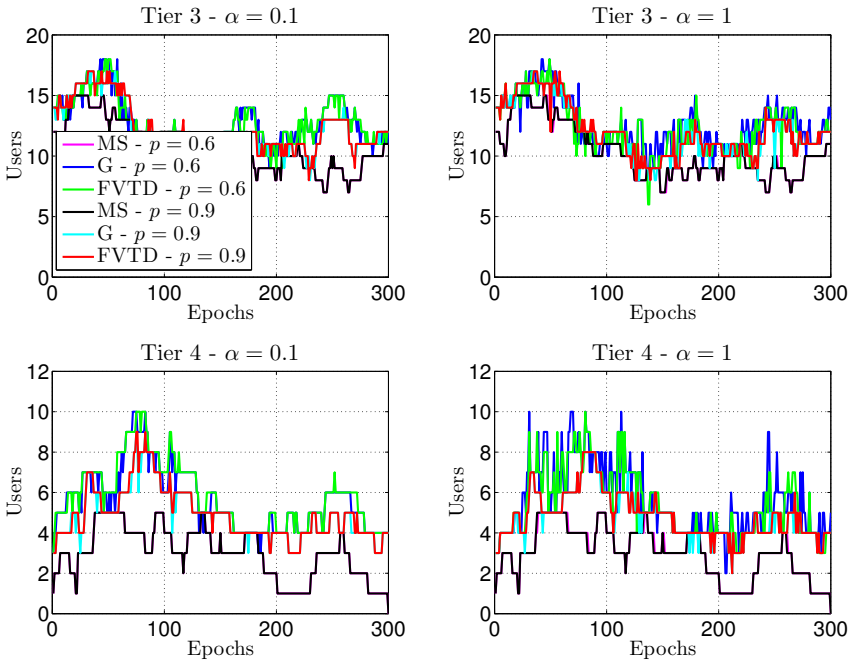


Figure 7.8: Evolution of user association in tier 3 and tier 4.

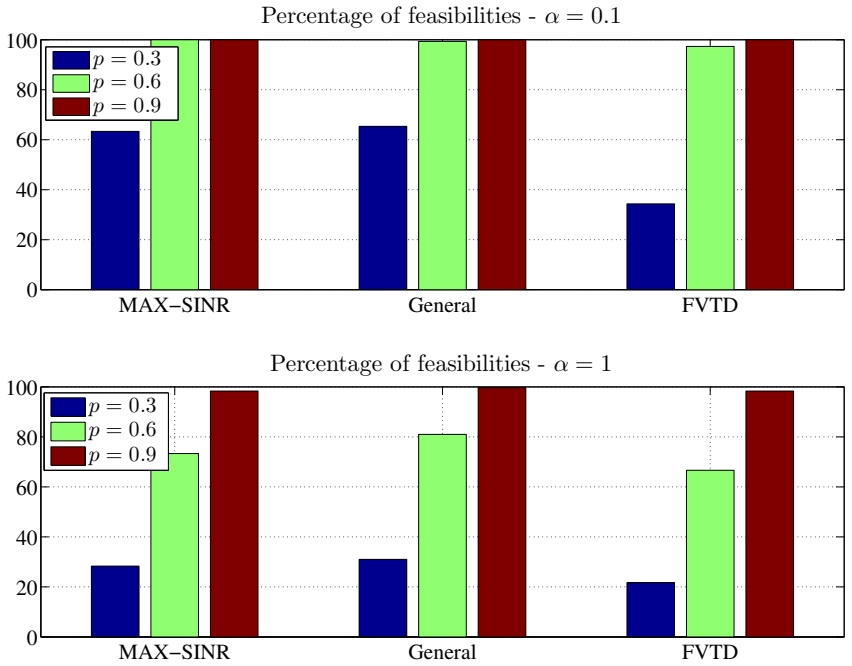


Figure 7.9: Percentage of feasibilities.

7.5.2 Numerical Simulations of the User Association Strategy from Section 7.4

In this section we evaluate the performance of the proposed ergodic strategy. We are going to start by showing the results of a simplified example composed of two BSs and 4 users. This will allow us to understand the behavior and the dynamics of the stochastic algorithm. Later, we will present results of the performance for the proposed scenario presented in the previous section.

7.5.2.1 Results for the Simplified Example

For the simulations presented in the sequel, the specific simulations parameters will be (unless stated otherwise) $e = 120$ J, $p = 0.7$, $P_{\text{BS}_i}^{\max} + P_{\text{CPICH}}^{i} = 46$ dBm for both BSs, and $P_c^i = 33$ dBm for both BSs. The rest of the parameters are the same as the ones considered before. First of all, recall that the optimality of the proposed algorithm was subject to a result in which we claimed that problem (7.48) had duality gap zero (see Proposition 7.1). In order to proof that proposition, we had to make some assumptions in the derivations. However, we also commented that the proposition would be validated through simulations. Note that we know that the optimum $\tilde{P}^\star$ of problem (7.48) is $\tilde{P}^\star = \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_i(\tau)$ as the objective function is an increasing function w.r.t. each $\tilde{P}_i$. However, we still need to prove that $\tilde{P}^\star(\tau;\pi^1(\tau),\pi^2(\tau))$ (from (7.70)) converges to $\frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_i(\tau)$. In fact, as we allowed $\tilde{P}_i$ to depend on τ and used an stochastic approximation method, what needs to converge to $\frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_i(\tau)$ is the expected value of $\tilde{P}^\star(\tau;\pi^1(\tau),\pi^2(\tau))$. Figure 7.10 shows in the top plot the evolution of $\tilde{P}^\star(\tau;\pi^1(\tau),\pi^2(\tau))$ of one of the two BSs and $\hat{H}$ is defined as $\hat{H} = \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_i(\tau)$. In the lower plot, we show the estimation of the expected value of $\tilde{P}^\star(\tau;\pi^1(\tau),\pi^2(\tau))$ computed as $\hat{\tilde{P}}^\star(\tau;\pi^1(\tau),\pi^2(\tau)) = \frac{1}{\tau}\sum_{u=1}^{\tau}\tilde{P}^\star(u;\pi^1(u),\pi^2(u))$. As we see in the lower plot, the claim of null duality gap is proved.

Let us continue by showing the convergence of the multipliers $\lambda_j(\tau)$ associated with the mean data rate $\tilde{R}_j$. Figure 7.11 shows the evolution of the multipliers. As it can be seen, they converge in mean to a given value after 60 epochs, approximately. However, the important thing to consider is the convergence of the data rates, $\mathbb{E}[\sum_{i\in\mathcal{B}}\bar{R}_{ji}(\tau)]$ and their values compared to the convergence of $\tilde{R}_j^\star(\lambda_j(\tau))$. For this reason, let us depict in Figure 7.12 the evolution of $\tilde{R}_j^\star(\lambda_j(\tau))$ for the 4 users in the system.

Finally, in Figure 7.13 we show the individual instantaneous rates and the mean rate achieved at each particular epoch (i.e., at epoch τ the mean rate of user j is computed as $\frac{1}{t}\sum_{u=1}^{t}\sum_{i\in\mathcal{B}}\bar{R}_{ji}(u)$). We see that the individual rates converge to a point and the final expected value is also shown in the legend. From the figure, we see that the expected individual rates almost converge to the same value as the one obtained in Figure 7.12, so the system is behaving well and as expected.

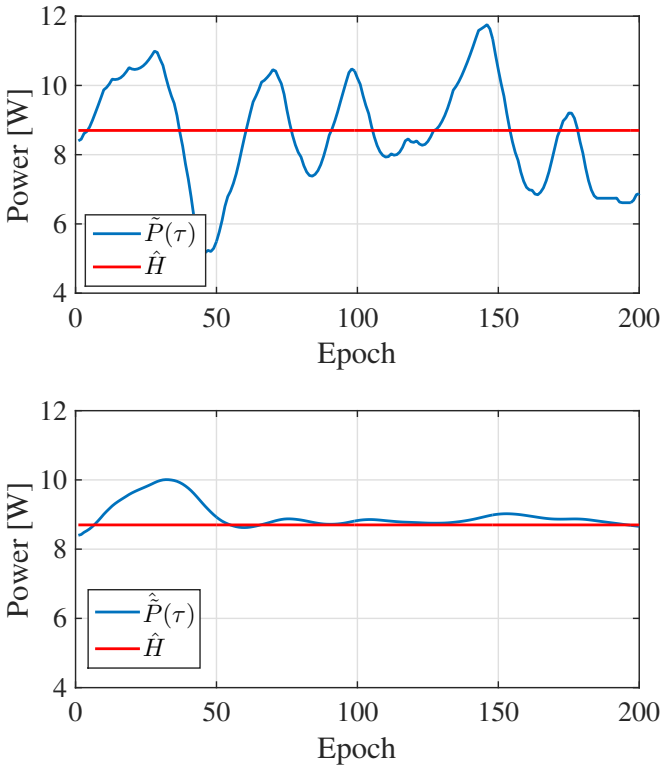


Figure 7.10: Convergence of $\tilde{P}_i$ to proof the null duality gap.

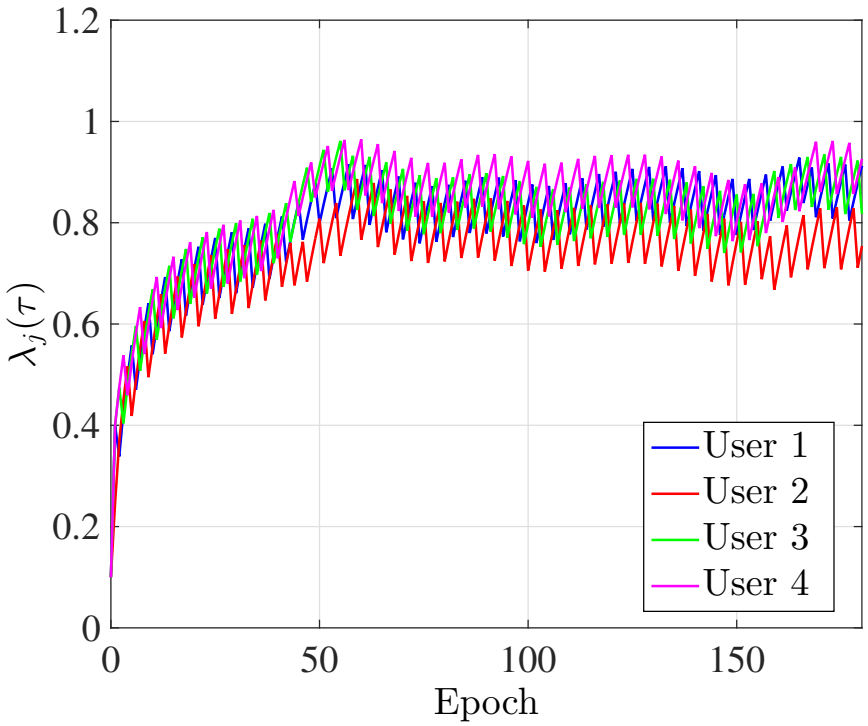


Figure 7.11: Evolution and convergence of the stochastic multipliers $\lambda_j(\tau)$.

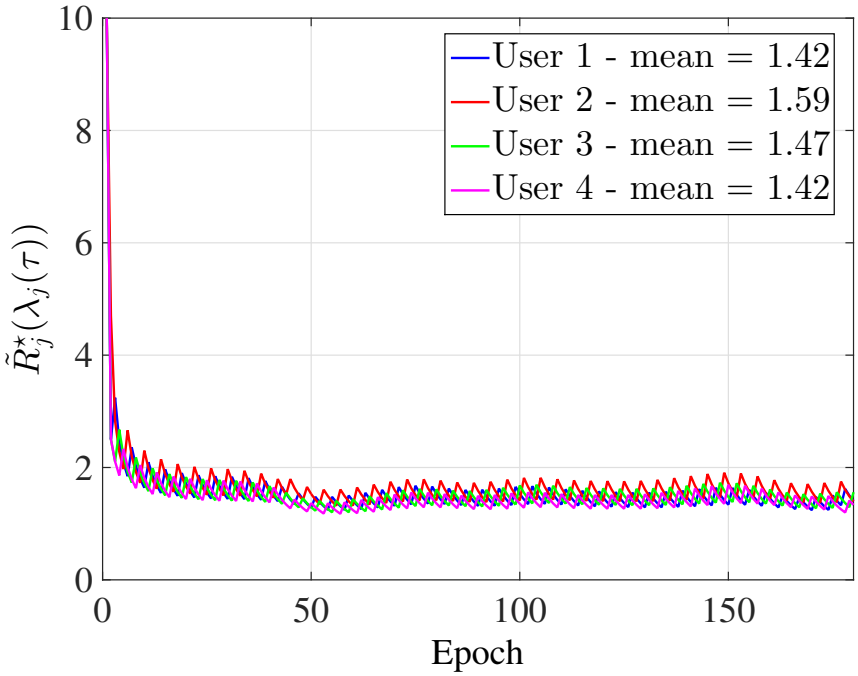


Figure 7.12: Evolution and convergence of the expected throughput of the users $\tilde{R}_j^\star(\lambda_j(\tau))$. We also show the mean value in convergence of $\tilde{R}_j^\star(\lambda_j(\tau))$ in the legend.

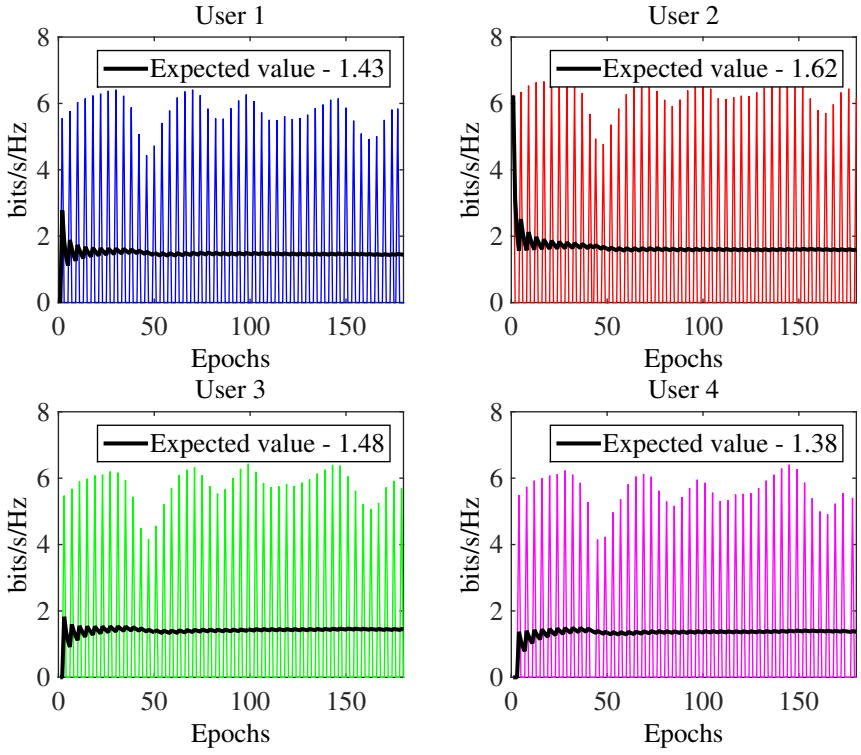


Figure 7.13: Instantaneous data rates $\sum_{i\in\mathcal{B}} \bar{R}_{ji}(\tau)$ and the expected value of them.

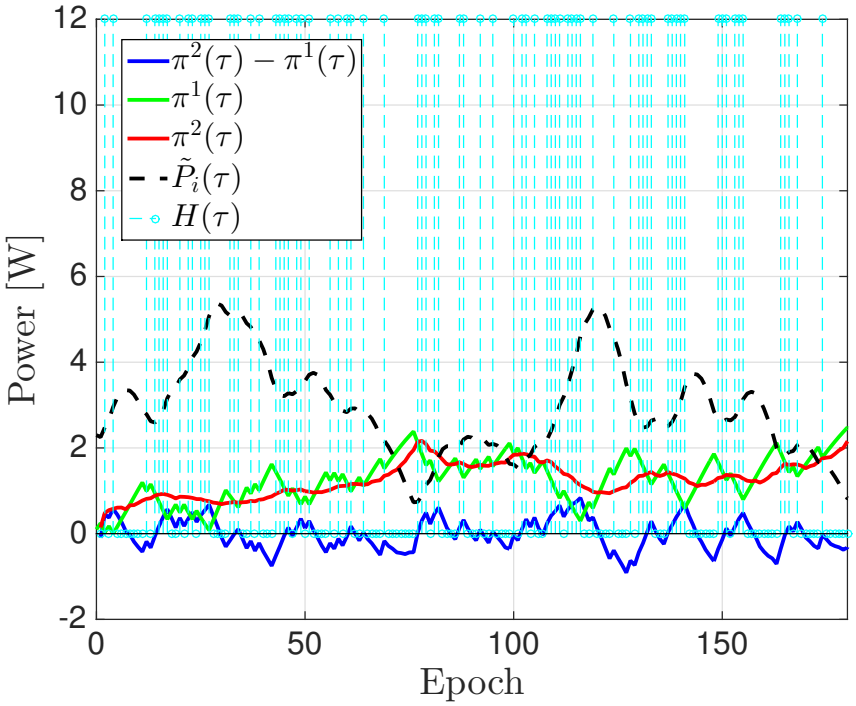


Figure 7.14: Evolution of stochastic variables with harvesting intensity $p = 0.4$.

Now, we show the behavior of the rest of variables, that is, $\tilde{P}_i$ and the rest of Lagrange multipliers. In Figure 7.14 we depict the evolution of some of the dynamic variables involved of one of the two BSs. In the plot, each cyan bar represents that an energy packet has arrived. The probability of receiving an energy packet is $p = 0.4$ in this case. In the plot, we also represent the difference $\pi^2 - \pi^1$ to show that in average $\tilde{P}_i$ converges (it is important to say here that through simulations we have seen that the term $\mathcal{C}(\tilde{P}_i(\tau))$ in (7.70) is negligible, thus, $\pi^{2,\star} = \pi^{1,\star}$ is a stationary condition that vanishes the gradient of the Lagrangian). We clearly see that if harvesting is coming, then the price of the long-term power $\pi^1(\tau)$ decreases and, thus, $\tilde{P}_i(\tau)$ increases, allowing the BS to spend more power now. If the harvesting decreases, then so does $\tilde{P}_i(\tau)$. The same reasoning can be applied to Figure 7.15 in which we have changed the harvesting intensity through the probability parameter p. Now, as a matter of example, in this case $p = 0.7$.

7.5.2.2 Results for the Proposed Scenario

Now that the convergence and the behavior of the stochastic approach have been covered, we will show the performance obtained with this approach evaluated in the same scenario and setup than the previous epoch-by-epoch strategy. The layout under consideration and simulation parameters are the ones introduced in at the beginning of the simulations section.

In order to show the convergence behavior of this complex scenario, Figure 7.16 presents the evolution of the stochastic Lagrange multipliers $\lambda_j(\tau)$. Recall that there is one $\lambda_j(\tau)$ for

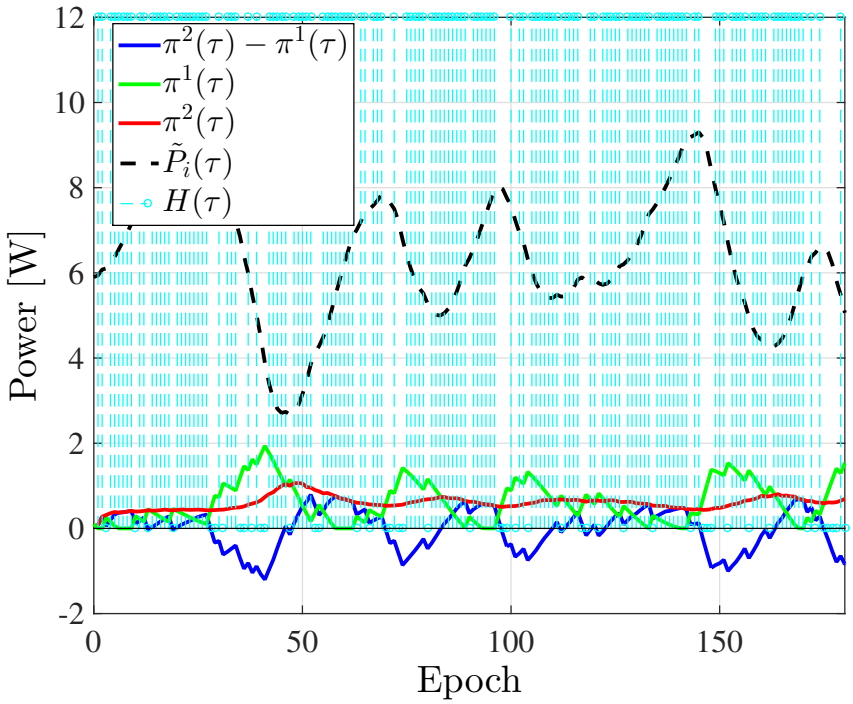


Figure 7.15: Evolution of stochastic variables with harvesting intensity $p = 0.7$.

each data user and there are a total of 85 data users in the system.

The time evolution of the batteries of the different tiers (averaged over the BSs belonging to the same tier) is shown in Figure 7.17 for different values of harvesting. 'STCH' in the legend refers to the stochastic user association approach. We see that in tier 2, tier 3, and in tier 4 the batteries converge (in mean) after some epochs. However, the level of harvesting intensity does not affect the residual battery level in convergence, as happened in the epoch-by-epoch strategy. Now, the harvesting intensity affects the slope of convergence, that is, if the harvesting is low, the battery will reach its stable state before than for high harvesting intensities. In tier 4, however, the battery fluctuates during the transmission. As happened in the epoch-by-epoch case (see Figure 7.6) this is due to the fact that no users were connected to any BSs belonging to tier 4 and the batteries of such BSs kept increasing for a period of 50 epochs approximately.

Figure 7.18 depicts the CDF of the individual user rates of the stochastic approach compared with the max-SINR strategy and the general epoch-by-epoch strategy. The value of α for the max-SINR and the epoch-by-epoch cases is $\alpha = 0.1$ and the value of harvesting is $p = 0.9$ except for the stochastic approach where we also show the CDF with harvesting $p = 0.6$. As we can conclude from the figure, the performance of the stochastic approach is really close to the performance obtained in the epoch-by-epoch strategy. Actually, it yields slightly better results for the same harvesting intensity. When compared with the max-SINR policy, the stochastic approach produces a better load-balanced network. In any case, the step sizes that are used for the stochastic iterations of the Lagrange multipliers and the stochastic variable $\tilde{P}_i(\tau)$ affect the

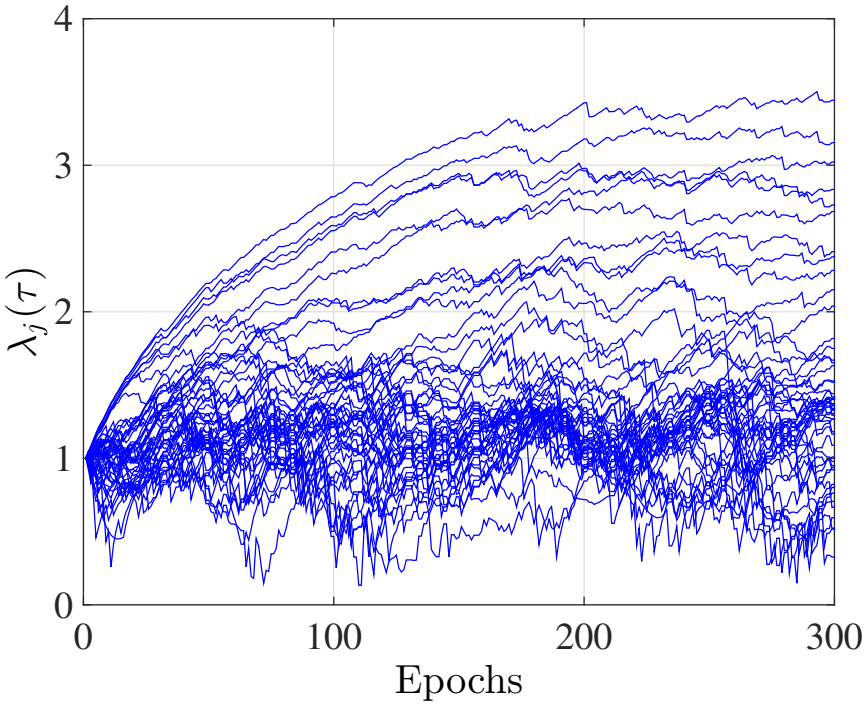


Figure 7.16: Evolution of the stochastic Lagrange multipliers $\lambda_j(\tau)$.

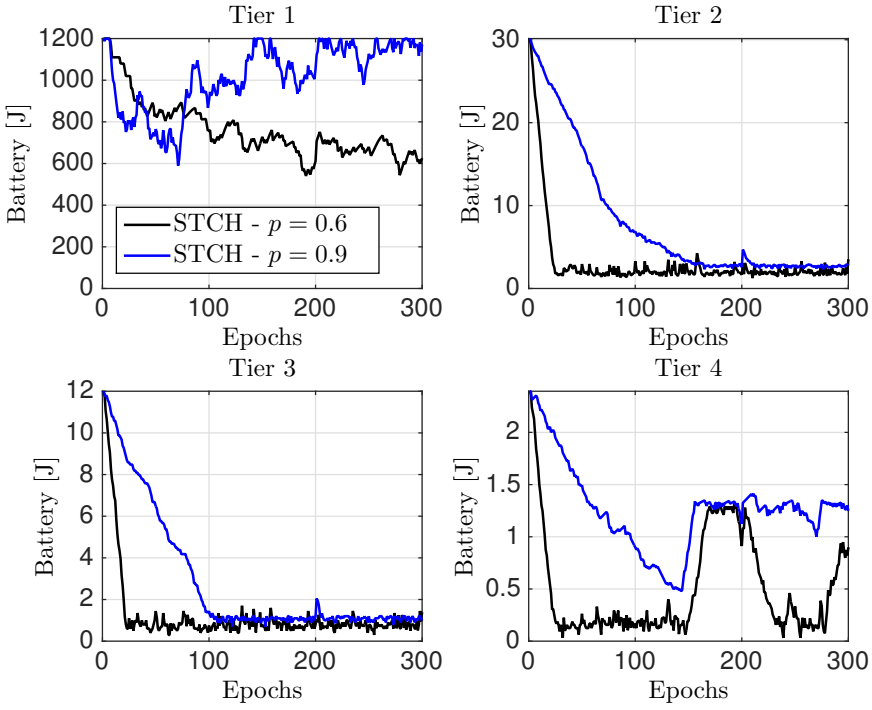


Figure 7.17: Battery evolution of the different tiers (averaged over BSs) with different harvesting intensities.

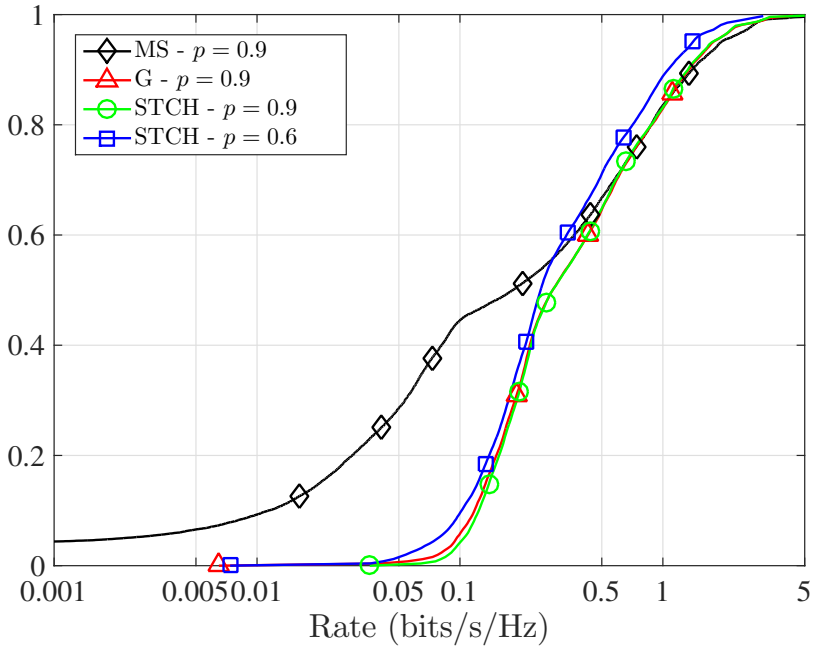


Figure 7.18: Comparison of the CDF of the individual user rates of the stochastic approach with the greedy epoch-by-epoch general approach and the max-SINR approach.

overall performance of the procedure. Therefore, by adjusting such parameters to fit better the scenario, it could be possible to obtain an overall better performance than the one presented in the figure.

Finally, as we also presented before, the user association for the different tiers as a function of time (epochs) is shown in Figure 7.19. As concluded from the CDF curve (see Figure 7.18) the users tend to associate with small cells and avoid saturation of the macro cell (tier 1), thus, achieving a better load-balanced network, similar effects that the ones obtained in the greedy epoch-by-epoch strategy (see Figures 7.7 and 7.8).

7.6 Chapter Summary and Conclusions

In this chapter we have proposed several user association strategies for a DL scenario composed of several BSs to achieve load balancing in heterogeneous networks where the BSs were solely powered with finite batteries and energy harvesting sources. The energy harvesting sources allow the BSs to recharge their batteries.

In the first part of the chapter, we developed user association techniques based on the greedy epoch-by-epoch approach followed in previous chapters of this dissertation. For this approach, different association techniques with different computational complexities were developed, aiming at having techniques suitable for scenarios with different user mobility requirements. Additionally, we proposed a distributed solution that is to be solved by each user and by the BSs

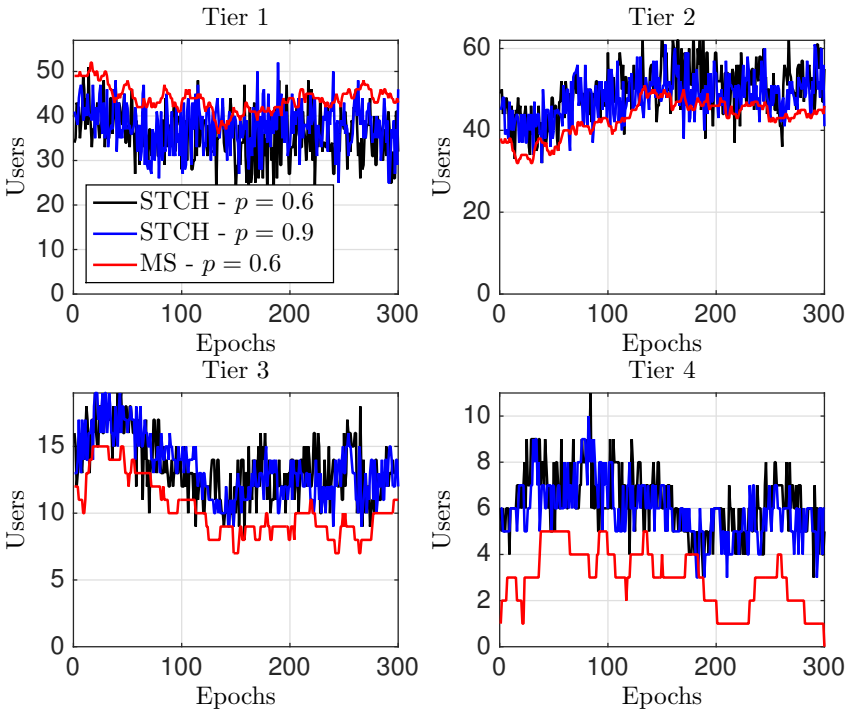


Figure 7.19: Comparison of the user association across time in different tiers of the max-SINR strategy with the proposed stochastic approach.

independently and that requires only local information, i.e., information of the local user channels and local battery information. In the simulations section, we have compared the proposed strategies with the classical max-SINR approach and showed that a clear improvement in terms of load-balancing is possible if a proper balancing technique is designed and the information of the battery status is considered in the user association procedure.

In the second part of the chapter, we proposed a technique based on ergodic stochastic optimization theory. In this case, we introduced time coupling that was handled (and decoupled) thanks to estimating the Lagrange multipliers at each epoch sequentially. We also allowed for some coupling in the energy harvesting/battery constraints and decoupling was also possible in this case by invoking stochastic approximation theory. Thanks to the previous coupling behavior, the association procedure was able to control the amount of energy to spend in a given epoch by considering the past, current, and future impact of the energy that is currently being harvested/spent on the system performance. In the simulations section, we developed a simplified example based on a simple scenario in order to clearly explain the dynamic behavior of this approach. Then, we presented complete simulations in which we compared the stochastic approach with the previous greedy strategy and with the max-SINR approach. The results showed that the stochastic approximation approach is able to produce similar or slightly better performance results than the greedy approach and much better performance than the max-SINR strategy.

7.A Approach to Obtain the Optimum Short-Term Variables

In this appendix, we present the technique to solve (7.55) assuming that the dual variables $\boldsymbol{\lambda}(\tau)$ are known. Let us start by formulating the inner optimization problem:

$$\begin{aligned}
\underset{\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau)}{\text{maximize}} \quad & \sum_{j\in\mathcal{U}_D}\sum_{i\in\mathcal{B}} \lambda_j(\tau) n_{ji}(\tau)\log_2\left(1+\frac{M_D p_{ji}(\tau)}{n_{ji}(\tau)(\theta\tilde{P}_i-\theta P_c^i+A_{ji}(\tau))}\right) \qquad (7.75)\\
\text{subject to} \quad & C2: \sum_{i\in\mathcal{B}} \frac{M_V \check{p}_{ji}(\tau)}{\theta\tilde{P}_i-\theta P_c^i+A_{ji}(\tau)} \geq \gamma_j, \quad \forall j\in\mathcal{U}_V\\
& C3: \sum_{i\in\mathcal{B}} x_{ji}(\tau)=1, \quad \forall j\in\mathcal{U}_T\\
& C4: \sum_{j\in\mathcal{U}_D} n_{ji}(\tau)\leq n_D^{(i)}, \quad \forall i\in\mathcal{B}\\
& C5: \sum_{j\in\mathcal{U}_D} p_{ji}(\tau)+\sum_{j\in\mathcal{U}_V}\check{p}_{ji}(\tau)+P^i_{\text{CPICH}}+P_c^i\leq\tilde{P}_i, \quad \forall i\in\mathcal{B}\\
& C7: 0\leq n_{ji}(\tau)\leq x_{ji}(\tau)n_D^{(i)}, \quad \forall i\in\mathcal{B},\forall j\in\mathcal{U}_D\\
& C8: 0\leq \check{p}_{ji}(\tau)\leq x_{ji}(\tau)(\tilde{P}_i-P^i_{\text{CPICH}}-P_c^i), \quad \forall i\in\mathcal{B},\forall j\in\mathcal{U}_V\\
& C9: x_{ji}(\tau)=0, \quad \forall i\notin\mathcal{S}(j,t),\forall j\in\mathcal{U}_T\\
& C10: x_{ji}(\tau)\geq 0, \quad \forall i\in\mathcal{B},\forall j\in\mathcal{U}_T\\
& C11: p_{ji}(\tau)\geq 0, \quad \forall i\in\mathcal{B},\forall j\in\mathcal{U}_D.
\end{aligned}$$

The approach proposed follows two steps. First, the optimal solution will be presented as a function of the dual variables. Such optimal primal solutions are the ones that minimize the Lagrangian. Then, a gradient-type strategy will be developed for computing the optimum dual variables by maximizing the dual function. Let $\boldsymbol{\nu}=\{\nu_j,\, j\in\mathcal{U}_V\}$, $\boldsymbol{\beta}=\{\beta_j,\, j\in\mathcal{U}_T\}$, $\boldsymbol{\mu}=\{\mu_j,\, j\in\mathcal{U}_D\}$, $\boldsymbol{\pi}^2=\{\pi_i^2,\, i\in\mathcal{B}\}$, $\boldsymbol{\zeta}=\{\zeta_{ji},\, j\in\mathcal{U}_D, i\in\mathcal{B}\}$, and $\boldsymbol{\xi}=\{\xi_{ji},\, j\in\mathcal{U}_V, i\in\mathcal{B}\}$ denote the vectors of dual variables associated to constraints $C2$, $C3$, $C4$, $C5$, $C7$, and $C8$ in problem (7.75). We collect all the dual variables in $\boldsymbol{\Psi}=\{\boldsymbol{\nu},\,\boldsymbol{\beta},\,\boldsymbol{\mu},\,\boldsymbol{\pi}^2,\,\boldsymbol{\zeta},\,\boldsymbol{\xi}\}$. There is no need to dualize constraints $C9$, $C10$, and $C11$ because the solution will turn out to automatically satisfy them as it will be shown later. Then, the Lagrangian of the optimization problem in (7.75) is

$$\begin{aligned}
&\mathcal{L}(\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau);\boldsymbol{\Psi})=\\
&-\sum_{j\in\mathcal{U}_D}\sum_{i\in\mathcal{B}}\lambda_j(\tau)n_{ji}(\tau)\log_2\left(1+\frac{M_D p_{ji}(\tau)}{n_{ji}(\tau)(\theta\tilde{P}_i-\theta P_c^i+A_{ji}(\tau))}\right)\\
&+\sum_{j\in\mathcal{U}_V}\nu_j\left(\gamma_j-\sum_{i\in\mathcal{B}}\frac{M_V\check{p}_{ji}(\tau)}{\theta\tilde{P}_i-\theta P_c^i+A_{ji}(\tau)}\right)+\sum_{j\in\mathcal{U}_T}\beta_j\left(\sum_{i\in\mathcal{B}}x_{ji}(\tau)-1\right)\\
&+\sum_{i\in\mathcal{B}}\mu_i\left(\sum_{j\in\mathcal{U}_D}n_{ji}(\tau)-n_D^{(i)}\right)+\sum_{j\in\mathcal{U}_D}\sum_{i\in\mathcal{B}}\zeta_{ji}\left(n_{ji}(\tau)-x_{ji}(\tau)n_D^{(i)}\right)
\end{aligned}$$

$$+\sum_{i\in\mathcal{B}}\pi_i^2\left(\sum_{j\in\mathcal{U}_D}p_{ji}(\tau)+\sum_{j\in\mathcal{U}_V}\check{p}_{ji}(\tau)+P^i_{\text{CPICH}}+P^i_c-\tilde{P}_i\right)$$
$$+\sum_{j\in\mathcal{U}_D}\sum_{i\in\mathcal{B}}\xi_{ji}\left(\check{p}_{ji}(\tau)-x_{ji}(\tau)(\tilde{P}_i-P^i_{\text{CPICH}}-P^i_c)\right). \tag{7.76}$$

The dual function is defined as the minimum of the Lagrangian w.r.t. the primal variables, i.e.,

$$D(\boldsymbol{\Psi})=\inf_{\substack{\mathbf{p}(\tau)\succeq 0,\,\check{\mathbf{p}}(\tau)\succeq 0,\\ \mathbf{n}(\tau)\succeq 0,\,\mathbf{x}(\tau)\succeq 0}}\mathcal{L}(\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau);\boldsymbol{\Psi}), \tag{7.77}$$

where $\succeq$ means the element-wise inequality and the dual problem is the following optimization problem

$$\underset{\substack{\boldsymbol{\nu}\succeq 0,\,\boldsymbol{\mu}\succeq 0,\,\boldsymbol{\pi}^2\succeq 0,\\ \boldsymbol{\zeta}\succeq 0,\,\boldsymbol{\xi}\succeq 0}}{\text{maximize}} \quad D(\boldsymbol{\Psi}). \tag{7.78}$$

Recall that the dual problem is always convex w.r.t. the dual variables and, thus, can be efficiently solved with a projected gradient method (if $D(\boldsymbol{\Psi})$ is differentiable) or a projected supergradient method if it is not differentiable [Ber99].

In is easy to show that Slater's condition is fulfilled in the convex optimization problem (7.55) (as the problem is convex, Slater's condition simplifies to show strict feasibility). Therefore, strong duality holds and the solution obtained by solving the primal problem (7.55) and the dual problem (7.78) is the same [Ber99].

For a given set of Lagrange multipliers, $\{\boldsymbol{\nu},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\pi}^2,\boldsymbol{\zeta},\boldsymbol{\xi}\}$, we need to minimize the Lagrangian w.r.t. the primal variables. As it will be shown next, the structure of the Lagrangian $\mathcal{L}(\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau);\boldsymbol{\Psi})$ allows the minimization w.r.t. $\mathbf{p}(\tau)$ and $\mathbf{n}(\tau)$ to be found in closed-form. Because $\mathcal{L}(\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau);\boldsymbol{\Psi})$ is strictly convex and differentiable w.r.t. $\mathbf{p}(\tau)$ and $\mathbf{n}(\tau)$, minimization w.r.t. these variables requires to equating the corresponding partial derivatives of $\mathcal{L}(\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau);\boldsymbol{\Psi})$ to zero. Differentiating the Lagrangian w.r.t. the powers, equating the derivatives to zero, and solving such expression for the powers yields

$$p^{\star}_{k\ell}(\tau,\mathbf{n}(\tau),\mathbf{p}(\tau);\boldsymbol{\Psi})=\left(\frac{\lambda_k(\tau)n_{k\ell}(\tau)}{\ln(2)\pi^2_\ell}-\frac{n_{k\ell}(\tau)(\theta\tilde{P}_\ell-\theta P^\ell_c+A_{k\ell}(\tau))}{M_D}\right)^{\infty}_{0}, \tag{7.79}$$

where the projection guarantees that the allocated powers are nonnegative (constraint $C11$ in (7.75)). Notice that the bisection method [Ber99] can be used to obtain the optimum power allocation (7.79). Proceeding similarly in the calculation of the optimum code allocation, we calculate the partial derivatives of the Lagrangian w.r.t. $n_{ji}(\tau)$, equate them to zero, and solve

such equations for the codes, obtaining:

$$n^{\star}_{k\ell}(\tau,\mathbf{n}(\tau),\mathbf{p}(\tau);\mathbf{\Psi}) = \begin{cases} \left(\frac{\lambda_k(\tau) M_D p_{k\ell}(\tau) \left((\theta \tilde{P}_\ell - \theta P_c^\ell + A_{k\ell}(\tau)) \ln(2) \right)^{-1}}{\lambda_k(\tau) \log_2 \left(1 + \frac{M_D p_{k\ell}(\tau)}{n_{k\ell}(\tau)(\theta \tilde{P}_\ell - \theta P_c^\ell + A_{k\ell}(\tau))} \right) - \mu_\ell - \zeta_{k\ell}} - \frac{M_D p_{k\ell}(\tau)}{\theta \tilde{P}_\ell - \theta P_c^\ell + A_{k\ell}(\tau)} \right)_0^\infty, & \text{if} \quad \ell \in \mathcal{S}_k(\tau) \\ 0, & \text{if} \quad \ell \notin \mathcal{S}_k(\tau), \end{cases} \tag{7.80}$$

where the projection guarantees that the codes satisfy the nonnegative constraints $C7$. Notice that the bisection method can also be used to compute the optimum code allocation, $n_{k\ell}(\tau)$.

Unfortunately, the minimization of the Lagrangian w.r.t. the association variables $x_{ji}(\tau)$ and the voice powers $\check{p}_{ji}(\tau)$ cannot be obtained by differentiating the Lagrangian as they appear through linear terms. In order to obtain the optimum power allocation for voice users and the association variables, we employ an iterative projected gradient approach to minimize the Lagrangian. The gradient for the k-th voice user power connected to the ℓ-th BS is given by

$$\nabla_{\check{p}_{k\ell}(\tau)} \mathcal{L}(\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau);\mathbf{\Psi}) \triangleq s_{k\ell}(\tau) = \frac{\nu_k M_D}{\theta \tilde{P}_\ell - \theta P_c^\ell + A_{k\ell}(\tau)} + \pi_\ell^2 + \xi_{k\ell}, \quad \forall k \in \mathcal{U}_V, \forall \ell \in \mathcal{B}, \tag{7.81}$$

and the update equation of the projected gradient method is given by

$$\check{p}_{k\ell}^{(z+1)}(\tau) = \left(\check{p}_{k\ell}^{(z)}(\tau) - \delta^{(z)} s_{k\ell}(\tau) \right)_0^\infty, \quad \forall k \in \mathcal{U}_V, \forall \ell \in \mathcal{B}, \tag{7.82}$$

where z is the iteration index, the projection guarantees the nonnegativity constraints in (7.75), and $\delta^{(z)} = \frac{Z}{\sqrt{z}\|\nabla \mathcal{L}\|_2}$ is the step size chosen such that the diminishing conditions are fulfilled, i.e., $\lim_{z\to\infty} \delta^{(z)} = 0$, $\sum_{z=1}^{\infty} \delta^{(z)} = \infty$, being $\nabla\mathcal{L}$ the overall gradient of the Lagrangian (w.r.t. all variables) [Ber99]. For the association variables, the gradient is given by

$$\nabla_{x_{k\ell}(\tau)} \mathcal{L}(\mathbf{p}(\tau),\check{\mathbf{p}}(\tau),\mathbf{n}(\tau),\mathbf{x}(\tau);\mathbf{\Psi}) \triangleq t_{k\ell}(\tau) = \beta_k - \zeta_{k\ell} n_D^{(\ell)} - \xi_{k\ell}(\tilde{P}_\ell - P_{\text{CPICH}}^\ell - P_c^\ell), \quad \forall k \in \mathcal{U}_T, \forall \ell \in \mathcal{B}, \tag{7.83}$$

and the update equation of the projected method is

$$x_{k\ell}^{(z+1)}(\tau) = \left(x_{k\ell}^{(z)}(\tau) - \delta^{(z)} t_{k\ell}(\tau) \right)_0^\infty, \quad \forall k \in \mathcal{U}_T, \forall \ell \in \mathcal{B}. \tag{7.84}$$

At this point, we have the expressions of the optimum primal variables, either in closed-form in (7.79) and (7.80) or iteratively in (7.82) and (7.84), for given dual variables. If the dual variables were optimum, then the expressions for the primal variables would yield the optimum values. The optimum Lagrange multipliers can be obtained from the dual problem (by maximizing the dual function) in (7.78). As the dual function is concave, we can apply any gradient-type algorithm to find the optimum solution. As the dual function is not differentiable, we propose the projected subgradient method to find the optimum dual variables. A valid subgradient for each particular dual variable is given by the constraint it is associated with [Ber99]. The update

equations are given by

$$\nu_j^{(q+1)} = \left(\nu_j^{(q)} + \kappa^{(q)}\left(\gamma_j - \sum_{i\in\mathcal{B}} \frac{\check{p}_{ji}(\tau)M_V}{\theta\tilde{P}_i - \theta P_c^i + A_{ji}(\tau)}\right)\right)_0^{\infty}, \forall j \in \mathcal{U}_V, \tag{7.85}$$

$$\beta_j^{(q+1)} = \beta_j^{(q)} + \kappa^{(q)}\left(\sum_{i\in\mathcal{B}} x_{ji}(\tau) - 1\right), \quad \forall j \in \mathcal{U}_T, \tag{7.86}$$

$$\mu_i^{(q+1)} = \left(\mu_i^{(q)} + \kappa^{(q)}\left(\sum_{j\in\mathcal{U}_D} n_{ji}(\tau) - n_D^{(i)}\right)\right)_0^{\infty}, \quad \forall i \in \mathcal{B}, \tag{7.87}$$

$$\pi_i^{2(q+1)} = \left(\pi_i^{2(q)} + \kappa^{(q)}\left(\sum_{j\in\mathcal{U}_D} p_{ji}(\tau) + \sum_{j\in\mathcal{U}_V} \check{p}_{ji}(\tau) + P_{\text{CPICH}}^i + P_c^i - \tilde{P}_i\right)\right)_0^{\infty} \forall i \in \mathcal{B}, \tag{7.88}$$

$$\zeta_{ji}^{(q+1)} = \left(\zeta_{ji}^{(q)} + \kappa^{(q)}\left(n_{ji}(\tau) - x_{ji}(\tau)n_D^{(i)}\right)\right)_0^{\infty}, \quad \forall j \in \mathcal{U}_D, \forall i \in \mathcal{B}, \tag{7.89}$$

$$\xi_{ji}^{(q+1)} = \left(\xi_{ji}^{(q)} + \kappa^{(q)}\left(\check{p}_{ji}(\tau) - x_{ji}(\tau)(\tilde{P}_i - P_{\text{CPICH}}^i - P_c^i)\right)\right)_0^{\infty}, \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}, \tag{7.90}$$

where the projections guarantee the nonnegativity constraints of the dual variables in (7.78), q indicates the iteration index, and the step size is given by $\kappa^{(q)} = \frac{Q}{\sqrt{q}\|\nabla D\|_2}$, so that the diminishing conditions for the step size assures convergence (in the previous expression, ∇D denotes the supergradient of the dual function w.r.t. all variables).

The global proposed iterative algorithm is based on the primal-dual block coordinate descent method for the update of the primal variables $p_{ji}(\tau)$ and $n_{ji}(\tau)$ (see Algorithm 7.4).

7.B Proof of Proposition 7.1

Let us start the proof by writing the outer optimization problem (7.48):

$$\begin{aligned} \underset{\tilde{\mathbf{P}}}{\text{maximize}} \quad & f(\tilde{\mathbf{P}}) \\ \text{subject to} \quad & \tilde{P}_i \leq \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_i(\tau), \quad \forall i \in \mathcal{B}. \end{aligned} \tag{7.91}$$

Then, since function $f(\cdot)$ is an increasing function on each $\tilde{P}_i$, we know that the optimum primal solution is $\tilde{P}_i^{\star} = \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_i(\tau)$. However, we still need to prove that problem (7.91) has zero duality gap in order to be able to implement a dual approach (and, thus, use the dual variables as message passing between the different optimization problems) in the resolution of the overall problem (7.46).

Let us define $\boldsymbol{\pi}^1 = \{\pi_i^1, i \in \mathcal{B}\}$ as the set of dual variables associated with the constraints.

Then, the Lagrangian of (7.91) is formulated as

$$\mathcal{L}(\tilde{\mathbf{P}}, \boldsymbol{\pi}^1) = -f(\tilde{\mathbf{P}}) + \sum_{i \in \mathcal{B}} \pi_i^1 \left(\tilde{P}_i - \frac{1}{T_e} \frac{1}{|\Upsilon|} \sum_{\tau \in \Upsilon} H_i(\tau) \right). \tag{7.92}$$

Algorithm 7.4 Primal-dual coordinate iterative algorithm for solving (7.75)

initialize $\boldsymbol{\nu} \succeq 0, \boldsymbol{\beta}, \boldsymbol{\mu} \succeq 0, \boldsymbol{\pi}^2 \succeq 0, \boldsymbol{\zeta} \succeq 0, \boldsymbol{\xi} \succeq 0$
repeat (index q)
  initialize $\mathbf{n} \succeq 0$
  repeat (index z)
    $p_{ji}^{(q,z+1)}(\tau) = p_{ji}^{\star}\left(\tau; \mathbf{n}^{(q,z)}(\tau), \mathbf{p}^{(q,z)}(\tau), \boldsymbol{\Psi}^{(q)}\right)$ using (7.79) $\quad \forall j, i$
    $n_{ji}^{(q,z+1)}(\tau) = n_{ji}^{\star}\left(\tau; \mathbf{n}^{(q,z)}(\tau), \mathbf{p}^{(q,z+1)}(\tau), \boldsymbol{\Psi}^{(q)}\right)$ using (7.80) $\quad \forall j, i$
  until $p_{ji}^{(q,z+1)}(\tau)$ and $n_{ji}^{(q,z+1)}(\tau)$ converge
  repeat (index z)
    $\check{p}_{ji}^{(q,z+1)}(\tau) = \left(\check{p}_{ji}^{(q,z)}(\tau) - \delta^{(k)} s_{ji}^{(q)}(\tau)\right)_0^{\infty} \quad \forall j, i$
  until $\check{p}_{ji}^{(q,z+1)}(\tau)$ converges
  repeat (index z)
    $x_{ji}^{(q,z+1)}(\tau) = \left(x_{ji}^{(q,z)}(\tau) - \delta^{(z)} t_{ji}^{(q)}(\tau)\right)_0^{\infty} \quad \forall j, i$
  until $x_{ji}^{(q,z+1)}(\tau)$ converges
  update the dual variables using $p_{ji}^{(q)}(\tau)$, $\check{p}_{ji}^{(q)}(\tau)$, $n_{ji}^{(q)}(\tau)$, and $x_{ji}^{(q)}(\tau)$ with (7.85), (7.86), (7.87), (7.88), (7.89), and (7.90)
until $\boldsymbol{\nu}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\pi}^2, \boldsymbol{\zeta}$, and $\boldsymbol{\xi}$ converge
end algorithm

Finally, the dual function is, then, defined as

$$g(\boldsymbol{\pi}^1) = \inf_{\tilde{\mathbf{P}}} \; -f(\tilde{\mathbf{P}}) + \sum_{i \in \mathcal{B}} \pi_i^1 \left(\tilde{P}_i - \frac{1}{T_e} \frac{1}{|\Upsilon|} \sum_{\tau \in \Upsilon} H_i(\tau) \right). \tag{7.93}$$

The minimum of the previous expression is attained at $\tilde{P}_i^{\star}(\boldsymbol{\pi}^1)$ and, thus, $g(\boldsymbol{\pi}^1) = -f(\tilde{\mathbf{P}}^{\star}(\boldsymbol{\pi}^1)) + \sum_{i \in \mathcal{B}} \pi_i^1 \left(\tilde{P}_i^{\star}(\boldsymbol{\pi}^1) - \frac{1}{T_e} \frac{1}{|\Upsilon|} \sum_{\tau \in \Upsilon} H_i(\tau) \right)$. As introduced in the statement of the proposition, in what follows, we will assume that the derivative of the Lagrangian in (7.92) is equal to zero at $\tilde{P}_i = \tilde{P}_i^{\star}(\boldsymbol{\pi}^1)$, $\forall i \in \mathcal{B}$, i.e., $-\frac{\partial f(\tilde{\mathbf{P}})}{\partial \tilde{P}_i}\Big|_{\tilde{\mathbf{P}} = \tilde{\mathbf{P}}^{\star}(\boldsymbol{\pi}^1)} + \pi_i^1 = 0$, $\forall i \in \mathcal{B}$.

Now, we need to formulate the dual problem which is the maximization of the dual function:

$$\begin{aligned} \underset{\boldsymbol{\pi}^1}{\text{maximize}} \quad & g(\boldsymbol{\pi}^1) \\ \text{subject to} \quad & \pi_i^1 \geq 0, \quad \forall i \in \mathcal{B}. \end{aligned} \tag{7.94}$$

Assuming that the dual function is differentiable, its gradient is equal to the subgradient, whose components are directly the evaluation of the constraints of problem (7.91) at $\tilde{P}_i = \tilde{P}_i^\star(\boldsymbol{\pi}^1)$ [Ber99], i.e.,

$$\frac{\partial g(\boldsymbol{\pi}^1)}{\partial \pi_k^1} = \tilde{P}_k^\star(\boldsymbol{\pi}^1) - \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_k(\tau), \quad \forall k \in \mathcal{B}. \tag{7.95}$$

Now, since we assume that the optimum of the dual problem $g(\boldsymbol{\pi}^1)$ is attained when its gradients are equal to zero (not at the extreme points of the function), $\frac{\partial g(\boldsymbol{\pi}^1)}{\partial \pi_k^1}\Big|_{\boldsymbol{\pi}^1=\boldsymbol{\pi}^{1\star}} = 0$, this implies that:

$$\tilde{P}_k^\star(\boldsymbol{\pi}^{1\star}) = \frac{1}{T_e}\frac{1}{|\Upsilon|}\sum_{\tau\in\Upsilon} H_k(\tau), \quad \forall k \in \mathcal{B}. \tag{7.96}$$

Therefore, since $\tilde{P}^\star = \tilde{P}_k^\star(\boldsymbol{\pi}^{1\star})$, $\forall i \in \mathcal{B}$, we conclude that the duality gap is zero.

7.C Generalization of the Sensitivity Analysis in Convex Analysis

In this appendix we generalize the development of sensitivity analysis derived in [Boy04]. The development presented in [Boy04] assumes that the perturbation is performed at the constraint space by tightening or loosing a given constraint. In this appendix, we generalize that idea by allowing the perturbation to be performed in the constraint function. We will first analyze the case with a perturbation in a single constraint. Then, we will generalize it by considering that multiple constraints are perturbed simultaneously. Finally, we will allow to have a vector of perturbations affecting all the constraints at the same time.

Let us introduce the following convex optimization problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_1(\mathbf{x}; 0) \leq 0, \end{aligned} \tag{7.97}$$

and the corresponding perturbed problem as follows:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_1(\mathbf{x}; u) \leq 0. \end{aligned} \tag{7.98}$$

Let $p^\star(0)$ and $p^\star(u)$ denote the optimum primal values of problems (7.97) and (7.98), respectively. Let $\mathbf{x}$ be any feasible point of the perturbed problem (7.98). Then, we know that

$$p^\star(0) = g(\lambda^\star) \leq f_0(\mathbf{x}) + \lambda^\star f_1(\mathbf{x}; 0), \tag{7.99}$$

where $g(\cdot)$ is the dual function of problem (7.97) and λ is the dual variable associated with the constraint in (7.97) [Boy04]. Since $\mathbf{x}$ is a feasible point of (7.98), then $f_1(\mathbf{x}; u) \leq 0 \Longrightarrow f_1(\mathrm{x}; 0) \leq f_1(\mathrm{x}; 0) - f_1(\mathrm{x}; u)$. Combining this with (7.99), we have

$$p^\star(0) = g(\lambda^\star) \leq f_0(\mathbf{x}) + \lambda^\star \left(f_1(\mathbf{x}; 0) - f_1(\mathbf{x}; u)\right), \quad \forall \mathbf{x} \text{ feasible of (7.98)}, \tag{7.100}$$

and thus,

$$f_0(\mathbf{x}) \geq p^\star(0) - \lambda^\star \left(f_1(\mathbf{x}; 0) - f_1(\mathbf{x}; u)\right), \quad \forall \mathbf{x} \text{ feasible of (7.98)}. \tag{7.101}$$

Now, let $\mathbf{x}^\star(u)$ be the optimum value of the perturbed problem. Then, we have

$$p^\star(u) = f_0(\mathbf{x}^\star(u)) \geq p^\star(0) - \lambda^\star \left(f_1(\mathbf{x}^\star(u); 0) - f_1(\mathbf{x}^\star(u); u)\right). \tag{7.102}$$

Rearranging terms, dividing both sides by u (considering that $u \geq 0$), and taking the limit at both sides, we end up with:

$$\lim_{u \to 0^+} \frac{p^\star(u) - p^\star(0)}{u} \geq \lim_{u \to 0^+} \frac{\lambda^\star \left(f_1(\mathbf{x}^\star(u); u) - f_1(\mathbf{x}^\star(u); 0)\right)}{u}, \tag{7.103}$$

and, thus, we finally obtain

$$\left.\frac{\partial p^\star(u)}{\partial u}\right|_{u=0} \geq \lambda^\star \frac{\partial f_1(\mathbf{x}^\star(0); 0)}{\partial u}. \tag{7.104}$$

Now, considering the case where $u \leq 0$ and following the previous reasoning, we can obtain

$$\lim_{u \to 0^-} \frac{p^\star(u) - p^\star(0)}{u} \leq \lim_{u \to 0^-} \frac{\lambda^\star \left(f_1(\mathbf{x}^\star(u); u) - f_1(\mathbf{x}^\star(u); 0)\right)}{u}, \tag{7.105}$$

and, thus,

$$\left.\frac{\partial p^\star(u)}{\partial u}\right|_{u=0} \leq \lambda^\star \frac{\partial f_1(\mathbf{x}^\star(0); 0)}{\partial u}. \tag{7.106}$$

Finally, combining (7.104) and (7.106) we conclude that:

$$\left.\frac{\partial p^\star(u)}{\partial u}\right|_{u=0} = \lambda^\star \frac{\partial f_1(\mathbf{x}^\star(0); 0)}{\partial u}. \tag{7.107}$$

Let us now consider the case where we have multiple constraints, all of them perturbed

simultaneously by constant u. Let us introduce the following convex optimization problems

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x};0) \leq 0, \quad i=1,\ldots,N \\ & h_j(\mathbf{x};0) = 0, \quad j=1,\ldots,K, \end{aligned} \tag{7.108}$$

and the corresponding perturbed problem as follows:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x};u) \leq 0, \quad i=1,\ldots,N \\ & h_j(\mathbf{x};u) = 0, \quad j=1,\ldots,K. \end{aligned} \tag{7.109}$$

We follow the same reasoning presented before. Let $\mathbf{x}$ be any feasible point of the perturbed problem (7.109). Then, we have that

$$p^\star(0) = g(\{\lambda_i^\star\},\{\mu_j^\star\}) \leq f_0(\mathbf{x}) + \sum_{i=1}^{N} \lambda_i^\star f_i(\mathbf{x};0) + \sum_{j=1}^{K} \mu_j^\star h_j(\mathbf{x};0), \tag{7.110}$$

where $p^\star(0)$ is the optimum value of the non-perturbed problem (7.108), $g(\{\lambda_i^\star\},\{\mu_j^\star\})$ is the dual function of such problem, and λ_i and μ_j are the dual variables associated to the i-th inequality and j-th equality constraints in (7.108), respectively. Since $\mathbf{x}$ is a feasible point of the perturbed problem, then $f_i(\mathbf{x};u) \leq 0, h_j(\mathbf{x};u) = 0 \implies p^\star(0) = g(\{\lambda_i^\star\},\{\mu_j^\star\}) \leq f_0(\mathbf{x}) + \sum_{i=1}^{N} \lambda_i^\star \left(f_i(\mathbf{x};0) - f_i(\mathbf{x};u)\right) + \sum_{j=1}^{K} \mu_j^\star \left(h_j(\mathbf{x};0) - h_j(\mathbf{x};u)\right)$. In particular, if we take the optimum point of the perturbed problem, $\mathbf{x}^\star(u)$, we obtain

$$\begin{aligned} f_0(\mathbf{x}^\star(u)) = p^\star(u) \geq\ & p^\star(0) - \sum_{i=1}^{N} \lambda_i^\star \left(f_i(\mathbf{x}^\star(u);0) - f_i(\mathbf{x}^\star(u);u)\right) \\ & - \sum_{j=1}^{K} \mu_j^\star \left(h_j(\mathbf{x}^\star(u);0) - h_j(\mathbf{x}^\star(u);u)\right). \end{aligned} \tag{7.111}$$

By rearranging terms and taking the limits at both sides as presented before, we finally end up with:

$$\left.\frac{\partial p^\star(u)}{\partial u}\right|_{u=0} = \sum_{i=1}^{N} \lambda_i^\star \frac{\partial f_i(\mathbf{x}^\star(0);0)}{\partial u} + \sum_{j=1}^{K} \mu_j^\star \frac{\partial h_j(\mathbf{x}^\star(0);0)}{\partial u}. \tag{7.112}$$

The last step generalizes the previous development for the case where the perturbation is a vector, i.e., $\mathbf{u} = (u_1,\ldots,u_P)$, affecting all the constraints. Being this the case, we can straightforwardly follow the procedures presented before to obtain

$$\left.\frac{\partial p^\star(\mathbf{u})}{\partial u_\ell}\right|_{u_\ell=0} = \sum_{i=1}^{N} \lambda_i^\star \frac{\partial f_i(\mathbf{x}^\star(0);0)}{\partial u_\ell} + \sum_{j=1}^{K} \mu_j^\star \frac{\partial h_j(\mathbf{x}^\star(0);0)}{\partial u_\ell}. \tag{7.113}$$

Part IV

Conclusions

Chapter 8

Conclusions and Future Work

This dissertation has focused on the design of resource allocation strategies and network management functions, considering that the network devices are powered with finite batteries and provided with energy harvesting sources that allow them to recharge their batteries. In particular, this dissertation has put emphasis on proposing online transmission solutions, that is, solutions that consider only past and current information of the channel, the battery states, and the harvesting dynamics. Additionally, this dissertation has modeled and introduced other sinks of power consumption in the proposed designs that were not considered so far, such as decoding consumption. As a result, the proposed resource allocation strategies produces longer battery lifetimes and larger aggregate network performance in terms of throughput than classical resource allocation strategies.

8.1 Conclusions

The general motivation of this dissertation as well as a review of the current state of the art has been presented in Chapter 1, jointly with an outline of the work developed in this thesis and the research contributions in terms of publications.

Chapter 2 has presented a brief description of the mathematical tools that have been used extensively throughout this dissertation. In particular, a review of convex optimization theory together with the description of duality theory has been presented; a brief summary of the theory based on ergodic optimization has been developed; finally, the mathematical tool known as MM has been introduced.

Chapter 3 has been devoted to the precoder design of a multiuser MIMO network. The mobile terminals have been considered to be battery-powered devices provided with energy harvesting sources. The key point was that the battery status of the users have been taken into account explicitly in the precoder and resource allocation design, increasing the lifetime of the receivers. We have also considered the case of imperfect CSI at the transmitter and a robust precoder design has been derived for that scenario. Then, we have addressed the problem of

battery quantization due to the use of rate-limited feedback channels. The transmitter designs the precoders taking into account the battery knowledge imperfections explicitly, yielding again to a robust approach. Simulations have shown that if the battery status of the users is incorporated in the design, the nodes improve their battery lifetime, whereas the average sum rate is enhanced at the same time. Thus, classical maximum sum rate techniques that do not take into account the status of the batteries of the users that have to be served are inefficient from the point of view of energy efficiency. The asymptotic behavior of the battery levels and the data rates of the proposed resource allocation strategy has been studied and characterized. Finally, we have addressed the problem of designing a user selection algorithm based on the PF in which the selection strategy exploited not only the channel dynamics but also the battery fluctuations.

Chapter 4 has studied the transmit covariance design that arises in multiuser multi-stream broadcast MIMO SWIPT networks. In the first part of the chapter, we used the BD strategy in which interference among users is pre-canceled at the transmitter so that the resulting transmit covariance problem is convex. For this scenario, we have derived the particular structure of the optimal transmit covariance matrices presented some user grouping techniques that allow the BS to select which users are better suited for information and which ones for battery replenishment in each particular frame. Simulation results have showed that the aggregated throughput can be considerably improved if the proposed grouping strategy is implemented when the results are compared with traditional scheduling approaches. Later, we have proposed different strategies for managing the minimum energy to be harvested. The procedures have been derived from the sensitivity analysis of duality theory where we considered the effect on the system performance increase or decrease when adjusting the harvesting constraints. In the second part of the chapter, we have presented a method to solve the difficult nonconvex problem that arises in multiuser multi-stream broadcast MIMO SWIPT networks when BD is not forced. To obtain local optimal solutions, the nonconvex problems have been solved based on the MM approach in which the solution of a nonconvex problem is obtained by solving a sequence of convex problems. Simulation results showed that the proposed methods outperform the classical BD approach in terms of both system sum rate and power collected by users by a factor of approximately 50%.

Chapter 5 has considered the design of a methodology for dimensioning the energy units, e.g., solar panels and batteries for powering BSs. The scenario under consideration has been based on two BSs placed at the same site, with fully overlapped coverage areas and using two different frequencies. Because the daily traffic profile is not constant, we have also provided a methodology for switching on and off one of the BSs in order to reduce the energy consumption and, thus, deploy smaller solar panels and fewer number of batteries. In this context, we have proposed a decision strategy where we have perfect knowledge of the traffic profile and a robust Bayesian strategy that accounted for possible error modeling in the traffic profile information. Simulations have been performed with real data for a real network deployment and the results have showed that proposed solution can be a sustainable and economical solution to provide

cellular services, for example, in outdoor isolated rural scenarios.

Chapter 6 has been devoted to development of a resource allocation strategy for the DL and the UL scenario based on the maximization of the minimum average data rate with backhaul capacity constraints. By the use of stochastic optimization tools, we have been able to deal with expected rates both in the objective function and in the backhaul capacity constraint, allowing the access network to offer higher rates by taking advantage of good instantaneous wireless channel conditions. We have assumed that the BS was powered with a finite battery that was able to be recharged by means of an energy harvesting source. The dynamics of the energy harvesting, the energy spending, and the battery have also been taken into account explicitly in the proposed resource allocation problem. Simulations results have showed that the proposed approach achieves more fairness among the users when compared to the traditional PF strategy, and provides greater worst-user rate and sum-rate if an average backhaul constraint is considered instead of an instantaneous constraint.

Finally, Chapter 7 has focused on a multi-tier multi-cell scenario. In particular, we have proposed some user association strategies to achieve load balancing in terms of aggregate throughput where the BSs were solely powered with finite batteries and energy harvesting sources. In the first part of the chapter, we have developed user association techniques based on the greedy epoch-by-epoch approach followed in previous chapters of this dissertation. In this setup, we have proposed some centralized solutions and a distributed one that requires only local information, i.e., information of the local user channels and local battery information. Simulation results have compared the proposed strategies with the classical max-SINR approach and have showed that improvement in terms of load-balancing is possible if a proper balancing technique is designed and the information of the battery status is considered in the user association procedure. In the second part of the chapter, we have proposed a user association technique based on ergodic stochastic optimization theory. In this case, we have introduced time coupling in some optimization variables that were handled using a stochastic approximation approach. Thanks to this coupling, the association procedure controlled the amount of energy to spend in a given epoch by considering the past, current, and future impact in the system performance. Simulations results have compared the stochastic approach with the previous greedy strategy and with the max-SINR approach and have showed that the stochastic approximation approach is able to produce similar or slightly better performance results than the greedy approach and much better performance than the max-SINR strategy.

8.2 Future Work

There are many possible research directions that can be considered to extend the results presented in this dissertation. In the sequel, we provide some ideas that, to the author's perspective, may be of general interest.

First of all, the research community responsible for developing electronic equipment should obtain more accurate energy consumption models of all components and stages of the transceiver. In the last years, there has been a lot of attention in modeling the power consumption of the RF chains but just a few results can be found in terms of the modeling of the energy consumption of the baseband processing stages. In this thesis, we have proposed to use some models of decoding consumption and the parameters of the models were set accordingly to obtain the desired energy consumption dynamics. However, it would be interesting to obtain some representative values for those parameters and perform system simulations to validate the performance of the proposed algorithms.

Throughout this dissertation we have proposed online transmission strategies. However, the proposed strategies were based on a suboptimal greedy approach or a stochastic approximation method. In this regard, it is known that the optimal formulation when energy harvesting nodes are present in a system must be solved with DP techniques (as explained in the thesis). These techniques are usually difficult to implement and require a lot of computational resources. This is the main reason as to why they have not been considered in this thesis. However, in some simplified scenarios (for example, a SISO scenario with a single design variable that represents if data transmission is possible), the optimal solutions may be somewhat easy to be found, and understanding the insights of the optimum solution could be valuable and useful to design other online suboptimal but near-optimal solutions.

In terms of available techniques for online solutions, there are a few approaches that could yield potential results. In this dissertation, we have proposed a learning technique based on the stochastic approximation theory but there are many more learning techniques that could be applied. This is the case of, for example, the mathematical tool known as online convex optimization, the goal of which is to produce a sequence of accurate predictions given knowledge of the correct answer to previous predictions. In this sense, if the optimization problem is modeled with a time coupling in a way that the optimum value of the optimization variables depend upon the network dynamics over time, then learning techniques should be exploited. Additionally, if the decisions to configure the network parameters depend on many state variables, such as daily traffic patterns, energy harvested patterns, concentration of users, backhaul state, network congestion and so on, then, in this case, learning methods that extract and exploit these patterns could be used to make current decisions on the network configuration. In this regard, the concept of self organizing networks evolved a few years ago but the reality is that the applicability of this technique to real network configuration is still very limited. Hence, extra effort should be devoted towards this research area.

Finally, a new research paradigm has been defined (although it is still vaguely defined) known as Internet of Things (IoT), which opens new research lines to explore. The network topology envisioned for IoT applications is still being developed but it will be a kind of an ad-hoc network, extremely densified. These IoT networks will be composed of many different

nodes but the majority of them will be sensors. Giving this particular network configuration, we can define a research line on how to provide a sustainable and efficient solution from an energy perspective. The concept of SWIPT presented in this thesis could be a potential solution to recharge the sensors but the results presented in this dissertation should be extended to ah-doc dense networks. In this sense, network protocols that are energy efficient, such as efficient sleep modes, should also be developed, having in mind the type of network architecture and the fact that the cost associated with the battery replacement of the sensors may be extremely high.

Bibliography

[Abr72] M. Abramowitz, and I. A. Stegun, *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, New York: Dover, 1972.

[And01] M. Andrews, *et al.*, "Providing quality of service over a shared wireless link", *IEEE Comm. Magazine*, vol. 39, no. 2, pp. 150–154, Feb. 2001.

[Ant11] M.A. Antepli, and H.E.E. Uysal-Biyikoglu, "Optimal packet scheduling on an energy harvesting broadcast link", *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1712–1731, Sep. 2011.

[Ara14] A. Arafa, and S. Ulukus, "Single-user and multiple access channels with energy harvesting transmitters and receivers", *IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, Dec. 2014.

[Atz10] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey", *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[Aue11] G. Auer, *et al.*, "How much energy is needed to run a wireless network?", *IEEE Trans. on Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[Awo03] O. Awoniyi, N.B. Mehta, and L.J. Greenstein, "Characterizing the orthogonality factor in WCDMA downlinks", *IEEE Trans. Wireless Comm.*, vol. 4, no. 2, pp. 621–625, Jul. 2003.

[Bai13] Q. Bai, A. Mezghani, and J. A. Nossek, "Throughput maximization for energy harvesting receivers", *Int. ITG Workshop on Smart Antennas (WSA)*, Mar. 2013.

[Bee06] S. P. Beeby, M. J. Tudor, and N. M. White, "Energy harvesting vibration sources for microsystems applications", *Measurement Science and Technology*, vol. 17, no. 12, Oct. 2006.

[Bel10] E. Belmega, S. Lasaulce, and M. Debbah, "A survey on energy-efficient communications", *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications*, pp. 289–294, Sep. 2010.

[Ber99] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, 2$^{\underline{\text{nd}}}$ ed., 1999.

[Ber05] D. P. Bertsekas, *Dynamic programming and optimal control*, Athena Scientific, thrid ed., 2005.

[Bha87] R. Bhatia, *Perturbation Bounds for Matrix Eigenvalues*, Longman Scientific & Technical, 1987.

[Bha06] S. A. Bhalerao, A. V. Chaudhary, R. B. Deshmukh, and R. M. Patrikar, "Powering wireless sensor nodes using ambient RF energy", *IEEE Int. Conf. on Systems, Man and Cybernetics*, Oct. 2006.

[Bi15] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: opportunities and challenges", *IEEE Comm. Magazine*, vol. 53, no. 4, pp. 117–125, Apr. 2015.

[Bla81] R. G. Bland, D. Goldfarb, and M. J. Todd, "The ellipsoid method: a survey", *Operations Research*, vol. 29, no. 6, pp. 1039–1091, 1981.

[Bla12] P. Blasco, D. Gunduz, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization", *IEEE Global Communications Conf. (GLOBECOM)*, Dec, 2012.

[Bla13] P. Blasco, D. Gündüz, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization", *IEEE Trans. Wireless Comm.*, vol. 12, no. 4, pp. 1872–1882, Apr. 2013.

[Bou12a] A. Bousia, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Green distance-aware base station sleeping algorithm in LTE-advanced", *IEEE Int. Conf. on Communications (ICC)*, Jun. 2012.

[Bou12b] A. Bousia, E. Kartsakli, L. Alonso, and C. Verikoukis, "Dynamic energy efficient distance-aware base station on/off scheme for LTE-advanced", *IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2012.

[Bou13] A. Bousia, *et al.*, "Game theoretic approach for switching off base stations in multi-operator environments", *IEEE Int. Conf. on Communications (ICC)*, Jun. 2013.

[Boy04] S. Boyd, and L. Vandenbergue, *Convex optimization*, Cambridge, 2004.

[Boy08] S. Boyd, and A. Mutapcic, "Subgradient methods", `https://see.stanford.edu/materials/lsocoee364b/02-subgrad_method_notes.pdf`, Apr. 2008.

[Boy10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers", *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, Nov. 2010.

[Bro13] S. A. Broda, "Tail probabilities and partial moments for quadratic forms in multivariate generalized hyperbolic random vectors", *Tinbergen Institute Discussion Paper 13-001/III. Available at SSRN: http://ssrn.com/abstract=219778*, 2013.

[Ceb] "Cebekit: Windlab junior c-0200 technical characteristics", Available: http://fadisel.cat/docs/c-0200-ing.pdf.

[Cha08] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey", *IEEE Comm. Magazine*, vol. 46, pp. 59–67, Sep. 2008.

[Cha11] B. K. Chalise, Y. D. Zhang, and M. G. Amin, "Simultaneous transfer of energy and information for MIMO-OFDM relay system", *IEEE Int. Conf. on Communications in China*, 2011.

[Che15] H. Chen, Y. Li, Y. Jiang, Y. Ma, and B. Vucetic, "Distributed power splitting for SWIPT in relay interference channels using game theory", *IEEE Trans. Wireless Comm.*, vol. 14, no. 1, pp. 410–420, Jan. 2015.

[Chi03] M. Chiani, D. Dardari, and M.K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels", *IEEE Trans. Wireless Comm.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.

[Cho11] A. Chowdhery, Wei Yu, and J.M. Cioffi, "Cooperative wireless multicell OFDMA network with backhaul capacity constraints", *IEEE Int. Conf. on Comm. (ICC))*, Jun. 2011.

[Cis] "Cisco visual networking index: global mobile data traffic forecast update, 2011 - 2016, tech. report 2012, www.cisco.com", .

[Cov06] T. M. Cover, and J. A. Thomas, *Elements of information theory*, Wiley, 2006.

[Cui03] S. Cui, A. Goldsmith, and A. Bahai, "Power estimation for Viterbi decoders", *Wireless Systems Lab, Stanford Univ., CA, Tech. Rep. Available: http://wsl.stanford.edu/Publications.html*, 2003.

[Cui04] S. Cui, A. Goldsmith, and A. Bahai, "Joint modulation and multiple access optimization under energy constraints", *IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2004.

[Cui05] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization", *IEEE Trans. Wireless Comm.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[Cui07] S. Cui, A. Goldsmith, and S. Lall, "Cross-layer energy and delay optimization in small-scale sensor networks", *IEEE Trans. Wireless Comm.*, vol. 6, no. 10, pp. 3688–3699, Oct. 2007.

[Cui14] Z. Cui, and R. Adve, "Joint user association and resource allocation in small cell networks with backhaul constraints", *IEEE Conf. on Information Sciences and Systems (CISS)*, Mar. 2014.

[Cur03] S. A. Curtis, "The classification of greedy algorithms", *Science of Computer Programming, ELSEVIER*, vol. 49, no. 1, pp. 125–157, Dec. 2003.

[Dai14] B. Dai, and W. Yu, "Sparse beamforming design for network MIMO system with per-base-station backhaul constraints", *IEEE Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2014.

[Dam97] A. Damnjanovic, *et al.*, "Multitier cell design", *IEEE Comm. Magazine*, vol. 35, no. 8, pp. 60–64, Aug. 1997.

[Dam11] A. Damnjanovic, *et al.*, "A survey on 3GPP heterogeneous networks", *IEEE Wireless Comm. Magazine*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[Dan14] L. Dantong, *et al.*, "Adaptive user association in hetnets with renewable energy powered base stations", *Int. Conf. on Telecommunications (ICT)*, 2014.

[Dan15] L. Dantong, *et al.*, "Two dimensional optimization on user association and green energy allocation for HetNets with hybrid energy sources", *IEEE Trans. Comm.*, vol. 63, pp. 4111–4124, Nov. 2015.

[Das97] S. K. Das, S. K. Sen, and R. Jayaram, "A dynamic load balancing strategy for channel assignment using selective borrowing in celular mobile environment", *Wireless Networks*, vol. 3, no. 5, pp. 333–347, Oct. 1997.

[Deb11] B. Debaillie, *et al.*, "Opportunities for energy savings in pico/femto-cell base-stations", *Future Network & Mobile Summit (FutureNetw)*, Jun. 2011.

[Dev12a] Yole Developpement, "Emerging energy harvesting devices", Nov. 2012.

[Dev12b] B. Devillers, and D. Gündüz, "A general framework for the optimization of energy harvesting communication systems with battery imperfections", *Journal of Communications and Networks*, vol. 14, no. 2, pp. 130–139, Apr. 2012.

[Dhr05] P. J. Dhrymes, "Moments of truncated normal distributions", *Tech. Rep. Available: http://www.columbia.edu/ pjd1/*, 2005.

[Dim05] G. Dimic, and N. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm", *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3857–3868, Oct. 2005.

[Din14] Z. Ding, and H.V. Poor, "User scheduling in wireless information and power transfer networks", *IEEE Int. Conf. on Communication Systems (ICCS)*, Nov 2014.

[EC08] "Addressing the challenge of energy efficiency through information and communication technologies, european comisssion", May 2008.

[Ehr05] M. Ehrgott, *Multicriteria optimization*, Springer Verlag, 2005.

[Eic09] R. Eickhoff, R. Kraemer, I. Santamaría, and L. González, "Developing energy-efficient MIMO radios", *IEEE Veh. Technol. Magazine*, vol. 4, no. 1, pp. 34–41, Mar. 2009.

[FB13] J. Fernandez-Bes, A. G. Marques, and J. Cid-Sueiro, "Battery-aware selective communications in energy-harvesting sensor networks: Optimal solution and stochastic dual approximation", *Intl. Symp. on Wireless Commun. Systems*, Aug. 2013.

[Feg13] M. Feghhi, A. Abbasfar, and M. Mirmohseni, "Optimal power and rate allocation in the degraded gaussian relay channel with energy harvesting nodes", *IEEE Workshop on Communication and Inf. Theory*, May 2013.

[Feh11] A. Fehske, G. Gettweis, J. Malmodin, and G. Biczók, "The global footprint of mobile communications : The ecological and economic perspective", *IEEE Comm. Magazine*, vol. 49, no. 8, pp. 55–62, Aug. 2011.

[Fou12] A. M. Fouladgar, and O. Simeone, "On the transfer of information and energy in multi-user systems", *IEEE Communications Letters*, vol. 16, no. 11, pp. 1733–1736, Nov. 2012.

[Gat09] N. Gatsis, A. Ribeiro, and G. Giannakis, "Cross-layer optimization of wireless fading ad-hoc networks", *IEEE Conf. Acoustics Speech Signal Process.*, Apr. 2009.

[Gat10] N. Gatsis, A. Ribeiro, and G. Giannakis, "A class of convergent algorithms for resource allocation in wireless fading networks", *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1808–1823, May 2010.

[Gat14] N. Gatsis, and A. G. Marques, "A stochastic approximation approach to load shedding in power networks", *IEEE Intl. Conf. on Acoustics, Speech and Signal Process.*, May. 2014.

[Gol96] G. H. Golub, and C. F. van Van Loan, *Matrix computations*, Johns Hopkins University Press, 1996.

[Gol02] A. J. Goldsmith, and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks", *IEEE Wireless Comm.*, vol. 9, no. 4, pp. 8–27, Aug. 2002.

[Gol03] A. Goldsmith, *et al.*, "Capacity limits of MIMO channels", *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[Gol05] A. Goldsmith, *Wireless communications*, Cambridge University Press, 2005.

[Gor] M. Gorlatova, J. Sarik, M. Cong, I. Kymissis, and G. Zussman, "Movers and shakers: kinetic energy harvesting for the internet of things", *To be published.* Available at arXiv preprint:1307.0044, 2013.

[Gra13] M. Grant, and S. Boyd, "CVX: Matlab software for disciplined convex programming", `http://cvxr.com/cvx`, Sep. 2013, version 2.0 beta.

[Gre13a] M. Gregori, and M. Payaró, "Energy-efficient transmission for wireless energy harvesting nodes", *IEEE Trans. Wireless Comm.*, vol. 12, no. 3, pp. 1244–1254, Mar. 2013.

[Gre13b] M. Gregori, and M. Payaró, "Mutual information maximization for a wireless energy harvesting node considering the circuitry power consumption", *IEEE Wireless Communications and Networking Conf. (WCNC)*, Apr. 2013.

[Gre13c] M. Gregori, and M. Payaró, "On the precoder design of a wireless energy harvesting node in linear vector gaussian channels with arbitrary input distribution", *IEEE Trans. Comm.*, vol. 61, no. 5, pp. 1868–1879, May 2013.

[Gre14a] M. Gregori, *Transmission strategies for wireless energy harvesting nodes*, PhD Thesis, Universitat Politècnica de Catalunya, 2014, phD thesis.

[Gre14b] M. Gregori, and M. Payaró, "On the optimal resource allocation for a wireless energy harvesting node considering the circuitry power consumption", *IEEE Trans. Wireless Comm.*, vol. 13, no. 11, pp. 5968–5984, Nov. 2014.

[Gro10] P. Grover, and A. Sahai, "Shannon meets Tesla: wireless information and power transfer", *Int. Symp. on Inf. Theory*, Jun. 2010.

[Gro11] P. Grover, K. Woyach, and A. Sahai, "Towards a communication-theoretic understanding of system-level power consumption", *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1744–1755, Sep. 2011.

[Guo12] W. Guo, and T. O'Farrell, "Dynamic cell expansion: Traffic aware low energy cellular network", *IEEE Vehicular Technology Conf. (VTC)*, Sep. 2012.

[Gur12] B. Gurakan, O. Ozel, J. Yang, and S. Ulukus, "Energy cooperation in energy harvesting wireless communications", *IEEE Int. Symp. on Inf. Theory*, Jul. 2012.

[Gü11] D. Gündüz, and B. Devillers, "Multi-hop communication with energy harvesting", *IEEE Int'l Workshop Comput. Adv. in Multi-Sensor Adaptive Proc. (CAMSAP)*, Dec. 2011.

[Hai67] Frank A. Haight, *Handbook of the Poisson Distribution*, John Wiley & Sons, 1967.

[Has03] B. Hassibi, and B.M. Hochwald, "How much training is needed in multiple-antenna wireless links?", *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[Has11] Z. Hasan, H. Boostanimehr, and V. Bhargava, "Green cellular networks: a survey, some research issues and challenges", *IEEE Comm. Surveys and Tutorials*, vol. 13, no. 4, pp. 524–540, Nov. 2011.

[Hay02] S. Haykin, *Adaptive filter theory*, Englewood Cliffs, NJ: Prentice-Hall, 2002.

[Hen11] D. Henrion, and J. Malick, "Projection methods in conic optimization", *Int. Series in Operations Research and Management Science*, vol. 166, pp. 565–600, Sep. 2011.

[Hjo11] Are Hjorungnes, *Complex-Valued Matrix Derivatives*, Cambridge University Press, 2011.

[Ho10] C. K. Ho, P. D. Khoa, and P. C. Ming, "Markovian models for harvested energy in wireless communications", *IEEE Int. Conf. on Communications Systems (ICCS)*, Nov. 2010.

[Ho12a] C. Ho, and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints", *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4808–4818, Sep. 2012.

[Ho12b] C. K. Ho, and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints", *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4808–4818, Sep. 2012.

[Hon12] M. Hong, and A. Garcia, "Mechanisms design for base station association and resource allocation in downlink OFDMA network", *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2238–2250, Dec. 2012.

[Hon16] M. Hong, Q. Li, and Y.-F. Liu, "Decomposition by successive convex approximation: A unifying approach for linear transceiver design in heterogeneous networks", *IEEE Trans. on Wireless Comm.*, vol. 15, no. 2, pp. 1377–1392, Feb. 2016.

[Hu10] Y. Hu, and A. Ribeiro, "Adaptive distributed algorithms for optimal random access channels", *Allerton Conf. on Commun. Control Computing*, Oct. 2010.

[Hu11] Y. Hu, and A. Ribeiro, "Optimal wireless networks based on local channel state information", *IEEE Conf. Acoustics Speech Signal Process.*, May 2011.

[Hu12] Y. Hu, and A. Ribeiro, "Optimal wireless networks based on local channel state information", *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4913–4929, Sep. 2012.

[Hu13] Y. Hu, and A. Ribeiro, "Optimal wireless communications with imperfect channel state information", *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2751–2766, Jun. 2013.

[Hua13] C. Huang, R. Zhang, and S. Cui, "Throughput maximization for the gaussian relay channel with energy harvesting constraints", *IEEE J. Sel. Areas Commun.*, vol. 31, no. 8, pp. 1469–1479, Aug. 2013.

[Hun04] D. R. Hunter, and K. Lange, "A tutorial on MM algorithms", *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[Jal00] A. Jalali, R. Padovani, and R. Pankaj, "Proportional fair scheduling for multi-cell multi-user MIMO systems", *IEEE Vehicular Technology Conf. (VTC) Spring*, May. 2000.

[Jen12] A. R. Jensen, *et al.*, "LTE UE power consumption model: for system level energy and performance optimization", *IEEE Vehicular Technology Conf. (VTC Fall)*, Sep. 2012.

[Jha13] S. C. Jha, A. T. Koc, R. Vannithamby, and M. Torlak, "Adaptive DRX configuration to optimize device power saving and latency of mobile applications over LTE advanced network", *IEEE Int. Conf. on Communications (ICC)*, Jun. 2013.

[Jio14] J. Jiong, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through internet of things", *IEEE Journal Internet of Things*, vol. 1, no. 2, pp. 112–121, Apr. 2014.

[Kan07] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks", *ACM Trans. on Embedded Computing Systems*, vol. 6, no. 4, pp. 32, Sep. 2007.

[Kay93] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice Hall, $1^{\underline{\text{st}}}$ ed., 1993.

[Kel97] F. P. Kelly, "Charging and rate control for elastic traffic", *European Trans. on Telecommunications*, pp. 33–37, 1997.

[Kha01] A. Khandekar, and R.J. McEliece, "On the complexity of reliable communication on the erasure channel", *IEEE Internation Symp. on Inf. Theory (ISIT)*, June 2001.

[Kim06] I.Y. Kim, and O. de Weck, "Adaptive weighted sum method for multi-objective optimization: a new method for Pareto front generation", *Structural and Multidisciplinary Optimization*, vol. 31, no. 2, pp. 105–116, Feb. 2006.

[Lai06] J. Laiho, A. Wacker, and T. Novosad, *Radio Network Planning and Optimisation for UMTS*, John Wiley & Sons, LTD, 2006.

[Lan05] J. Landt, "The history of RFID", *IEEE Potentials*, vol. 24, no. 4, pp. 8–11, Oct. 2005.

[Lee06] J.-W. Lee, R. R. Mazumbar, and N. B. Shroff, "Joint resource allocation and base station assignment for the downlink in CDMA networks", *IEEE/ACM Trans. on Networking*, vol. 14, no. 1, pp. 1–14, Feb. 2006.

[Li11a] G. Li, *et al.*, "Energy-efficient wireless communications: tutorial, survey and open issues", *IEEE Wireless Comm.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[Li11b] G. Ye Li, *et al.*, "Energy-efficient wireless communications: Tutorial, survey, and open issues", *IEEE Wireless Comm.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[Liu04] Lingjia Liu, Young-Han Nam, and Jianzhong Zhang, "Streaming applications over HSDPA in mixed service scenarios", *IEEE Vehicular Technology Conf. (VTC Fall)*, Sep. 2004.

[Liu10] Lingjia Liu, Young-Han Nam, and Jianzhong Zhang, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system", *Annual Conf. on Information Sciences and Systems (CISS)*, Mar. 2010.

[Liu13a] L. Liu, R. Zhang, and K. C. Chua, "Wireless information and power transfer: a dynamic power splitting approach", *IEEE Trans. Comm.*, vol. 61, no. 9, pp. 3990–4001, Sep. 2013.

[Liu13b] L. Liu, R. Zhang, and K.-C. Chua, "Wireless information transfer with opportunistic energy harvesting", *IEEE Trans. Wireless Comm.*, vol. 12, no. 1, pp. 288–300, Jan. 2013.

[Loz06] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel gaussian channels with arbitrary input distributions", *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[LR14] L. M. Lopez-Ramos, A. G. Marques, and J. Ramos, "Jointly optimal sensing and resource allocation for multiuser interweave cognitive radios", *IEEE Trans. Wireless Comm.*, vol. 13, no. 11, pp. 5854–5967, Nov. 2014.

[Lu15] X. Lu, *et al.*, "Wireless networks with RF energy harvesting: A contemporary survey", *IEEE Communications Surveys and Tutorials*, vol. 17, no. 2, pp. 757–789, Secondquarter 2015.

[Luo13] Y. Luo, J. Zhang, and K. B. Letaief, "Optimal scheduling and power allocation for twohop energy harvesting communication systems", *IEEE Trans. Wireless Comm.*, vol. 12, no. 9, pp. 4729–4741, Sep. 2013.

[Mad10] R. Madan, *et al.*, "Cell association and interference coordination in heterogeneous LTE-A cellular networks", *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479–1489, Dec. 2010.

[Mag88] J. R. Magnus, and H. Neudecker, *Matrix Differential Calculus with Application in Statistics and Econometrics*, John Wiley & Sons, Inc., Essex, UK, 1988.

[Mal13] J. Malmodin, P. Bergmark, and D. Lunden, "The future carbon footprint of the ICT and E&M sectors", *Int. Conf. on Information and Communication Technologies for Sustainability*, Feb. 2013.

[Mar09] A. G. Marques, X. Wang, and G. B. Giannakis, "Dynamic resource management for cognitive radios using limited-rate feedback", *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3651–3666, Sep. 2009.

[Mar11a] I. Marić, B. Bostjancic, and A. Goldsmith, "Resource allocation for constrained backhaul in picocell networks", *IEEE Inf. Theory and Applications Workshop (ITA)*, Feb. 2011.

[Mar11b] A. G. Marques, G. B. Giannakis, L. M. Lopez-Ramos, and J. Ramos, "Stochastic resource allocation for cognitive radio networks based on imperfect state information", *IEEE Intl. Conf. on Acoustics, Speech and Signal Process.*, May. 2011.

[Mar12] A. G. Marques, L. M. Lopez-Ramos, G. B. Giannakis, J. Ramos, and A. Caamano, "Optimal cross-layer resource allocation in cellular networks using channel and queue state information", *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2789–2807, Jul. 2012.

[Mar13] A. G. Marques, C. Figuera, C. Rey-Moreno, and J. Simo-Reigadas, "Optimal cross-layer schemes for relay networks with short-term and long-term constraints", *IEEE Trans. Wireless Comm.*, vol. 12, no. 1, pp. 333–345, Jan. 2013.

[Mat06] L. Mateu, C. Codrea, N. Lucas, M. Pollak, and P. Spies, "Energy harvesting for wireless communication systems using thermogenerators", *Conf. on Design of Circuits and Integrated Systems*, Nov. 2006.

[MD13] H. Mahdavi-Doost, and R. D. Yates, "Energy harvesting receivers: finite battery capacity", *IEEE Internation Symp. on Inf. Theory (ISIT)*, Jul. 2013.

[MD14] H. Mahdavi-Doost, and R. D. Yates, "Fading channels in energy-harvesting receivers", *Annual Conf. on Information Sciences and Systems (CISS)*, Mar. 2014.

[Med00] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel", *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[Med10] B. Medepally, and N. B. Mehta, "Voluntary energy harvesting relays and selection in cooperative wireless networks", *IEEE Trans. Wireless Comm.*, vol. 9, no. 11, pp. 3543–3553, Nov. 2010.

[Mia10] G. Miao, N. Himayat, and Y. Li, "Energy-efficient link adaptation in frequency-selective channels", *IEEE Trans. Comm.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[Mio14] Marco Miozzo, Davide Zordan, Paolo Dini, and Michele Rossi, "Solarstat: Modeling photovoltaic sources through stochastic markov processes", *IEEE Energy Conf.*, May 2014.

[Mor15] R. Morsi, D.S. Michalopoulos, and R. Schober, "Multiuser scheduling schemes for simultaneous wireless information and power transfer over fading channels", *IEEE Trans. on Wireless Comm.*, vol. 14, no. 4, pp. 1967–1982, Apr 2015.

[Oh10] E. Oh, and B. Krishnamachari, "Energy savings through dynamic base station switching in cellular wireless access networks", *IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2010.

[Orh12a] O. Orhan, and E. Erkip, "Optimal transmission policies for energy harvesting two-hop networks", *IEEE Annual Conf. on Information Sciences and Systems*, Mar. 2012.

[Orh12b] O. Orhan, D. Gunduz, and E. Erkip, "Throughput maximization for an energy harvesting communication system with processing cost", *IEEE Inf. Theory Workshop*, Sep. 2012.

[Oze10] O. Ozel, and S. Ulukus, "Information-theoretic analysis of an energy harvesting communication system", *IEEE Int. Symp. on Personal, Indoor, and Mobile Radio Communications*, Sep. 2010.

[Oze11] O. Ozel, J. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: optimal policies", *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, 2011.

[Oze12] O. Ozel, J. Yang, and S. Ulukus, "Optimal broadcast scheduling for an energy harvesting rechargeable transmitter with a finite capacity battery", *IEEE Trans. Wireless Comm.*, vol. 11, no. 6, pp. 2193–2203, 2012.

[Pal03] D. P. Palomar, *A unified framework for communications through MIMO channels*, PhD Thesis, Universitat Politècnica de Catalunya, May 2003, phD thesis.

[Pal05] D. Palomar, and S. Barbarossa, "Designing MIMO communication systems: constellation choice and linear transceiver design", *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3804–3818, Oct. 2005.

[Pal06] D. Palomar, and M. Chiang, "A tutorial on decomposition methods for network utility maximization", *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[Par05] J.A. Paradiso, and T. Starner, "Energy scavenging for mobile wireless electronics", *IEEE Computing Pervasive*, vol. 4, no. 1, pp. 18–27, Jan. 2005.

[Par13] J. Park, and B. Clerckx, "Joint wireless information and energy transfer in a two-user MIMO interference channel", *IEEE Trans. on Wireless Comm.*, vol. 12, no. 8, pp. 4210–4221, Aug. 2013.

[Par14] J. Park, and B. Clerckx, "Joint wireless information and energy transfer in a K-user MIMO interference channel.", *IEEE Trans. on Wireless Comm.*, vol. 13, no. 10, pp. 5781–5796, Oct. 2014.

[Per14] C. Perera, C. H. Liu, and S. Jayawardena C. Min, "A survey on internet of things from industrial market perspective", *IEEE Journals and Magazines: open access*, vol. 2, pp. 1660–1679, Jan. 2014.

[Ren11] A. Rendón, P. Jeanneth, and A. Martínez, *Tecnologías de la Información y las Comunicaciones para zonas rurales. Aplicacioón a la atención de salud en países en desarrollo*, CYTED, 1st ed., 2011.

[Rib08a] A. Ribeiro, and G. Giannakis, "Optimal layered architectures of wireless networks", *Asilomar Conf. on Signals Systems Computers*, Nov. 2008.

[Rib08b] A. Ribeiro, and G. Giannakis, "Robust stochastic routing and scheduling for wireless ad-hoc networks", *Wireless Commun. Mobile Computing Conf.*, Aug. 2008.

[Rib10a] A. Ribeiro, "Ergodic stochastic optimization algorithms for wireless communication and networking", *IEEE Int. Conf. Acoustics Speech Signal Process.*, Mar. 2010.

[Rib10b] A. Ribeiro, "Ergodic stochastic optimization algorithms for wireless communication and networking", *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6369–6386, Nov. 2010.

[Rib10c] A. Ribeiro, "Stochastic learning algorithms for optimal design of wireless fading networks", *IEEE Workshop on Signal Process. Advances in Wireless Commun.*, Jun. 2010.

[Rib10d] A. Ribeiro, and G. Giannakis, "Separation principles in wireless networking", *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4488–4505, Sep. 2010.

[Rib12] A. Ribeiro, "Optimal resource allocation in wireless communication and networking", *EURASIP J. Wireless Commun.*, vol. 2012, no. 272, Aug. 2012.

[Ric03] T. Richardson, and R. Urbanke, "The renaissance of gallager's low density parity-check codes", *IEEE Comm. Magazine*, vol. 41, no. 8, pp. 126–131, Aug. 2003.

[Roc97] R. Tyrrell Rockafellar, *Convex Analysis*, Princeton University Press, 1997.

[Ros10] P. Rost, and G. Fettweis, "On the transmission-computation-energy tradeoff in wireless and fixed networks", *IEEE Global Communications Conf. (GLOBECOM), workshop on Green Communications*, pp. 1934–1939, Dec. 2010.

[Rou03] S. J. Round, *Energy scavenging for wireless sensor nodes with a focus on vibration to electricity conversion*, PhD Thesis, University of California, 2003, phD thesis.

[Rub13] J. Rubio, and A. Pascual-Iserte, "Simultaneous wireless information and power transfer in multiuser MIMO systems", *IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2013.

[San08] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Dynamic association for load balancing and interference avoidance in multi-cell networks", *Wireless Networks*, vol. 14, pp. 103–120, Jan. 2008.

[Scu14] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems", *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 641–656, Feb. 2014.

[Sey08] A. Seyedi, and B. Sikdar, "Modeling and analysis of energy harvesting nodes in wireless sensor networks", *Annual Allerton Conf. on Communication, Control, and Computing*, Sep. 2008.

[She13] K. Shen, and W. Yu, "Downlink cell association optimization for heterogeneous networks via dual coordinate descents", *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May 2013.

[Shi14] Q. Shi, L. Liu, W. Xu, and R. Zhang, "Joint transmit beamforming and receive power splitting for MISO SWIPT systems", *IEEE Trans. Wireless Comm.*, vol. 13, no. 6, pp. 3269–3280, Apr. 2014.

[Sho85] N. Z. Shor, *Minimization methods for non-differentiable functions*, Springer-Verlag, 1985.

[Sig09] S. Sigdel, and W. A. Krzymien, "Simplified fair scheduling and antenna selection algorithms for multiuser MIMO orthogonal space-division mutiplexing downlink", *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1329–1344, Mar. 2009.

[Soh13] Y. S. Soh, T. Q. S. Quek, and M. Kountouris, "Dynamic sleep mode strategies in energy efficient cellular networks", *IEEE Int. Conf. on Communications (ICC)*, Jun. 2013.

[Sol11] H. M. Soliman, *et al.*, "Joint power and backhaul bits allocation for coordinated multipoint transmission", *Int. Symp. on Wireless Communication Systems (ISWCS)*, Nov. 2011.

[Son09] K. Son, S. Chong, and G. Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks", *IEEE Trans. Wireless Comm.*, vol. 8, no. 7, pp. 3566–3576, Jul. 2009.

[Spe04] Q. H. Spencer, *et al.*, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels", *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[Stu] J. F. Sturm, "Sedumi software", `http://sedumi.ie.lehigh.edu`.

[Sud11] S. Sudevalayam, and P. Kulkarni, "Energy harvesting sensor nodes: survey and implications", *IEEE Comm. Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, 2011.

[Tel95] I. Telatar, "Capacity of multi-antenna Gaussian channels", *AT&T Technical Memorandum*, Jun. 1995.

[tuc13] "ICT-601102 STP TUCAN3G, UMTS/HSPA network dimensioning", *Deliverable D41, available online: http://www.ict-tucan3g.eu*, Nov. 2013.

[tuc14] "ICT-601102 STP TUCAN3G: optimization and monitoring of HNB network", *Deliverable D42, available online: http://www.ict-tucan3g.eu*, Nov. 2014.

[Tut12] J. Tutuncuoglu., and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes", *IEEE Trans. Wireless Comm.*, vol. 11, no. 3, pp. 1180–1189, Mar. 2012.

[Var08] L. R. Varshney, "Transporting information and energy simultaneously", *Int. Symp. on Inf. Theory*, Jul. 2008.

[Ver90] S. Verdú, "On channel capacity per unit cost", *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[Ver10] W. Vereecken, W. V. Heddeghem, D. Colle, M. Pickavet, , and P. Demeester, "Overall ICT footprint and green communication technologies", *Proceedings of the 4th Int. Symp. on Communications, Control and Signal Processing*, Mar. 2010.

[Vis08] H. J. Visser, A. C. Reniers, and J. A. Theeuwes, "Ambient RF energy scavenging: GSM and WLAN power density measurements", *IEEE European Microwave Conf.*, Oct. 2008.

[Vul10] R. J. Vullers, R. Schaijk, H. J. Visser, J. Penders, and C. V. Hoof, "Energy harvesting for autonomous wireless sensor networks", *IEEE J. Solid-State Circuits*, vol. 2, no. 2, pp. 29–38, Spring 2010.

[Wan92] B. Wang, and F. Zhang, "Some inequalities for the eigenvalues of the product of positive semidefinite Hermitian matrices", *Structural and Multidisciplinary Optimization*, vol. 160, pp. 113–118, Jan. 1992.

[Wan07] X. Wang, G. B. Giannakis, and A. G. Marques, "A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks", *Proceedings of IEEE*, vol. 95, no. 12, pp. 2410–2431, Dec. 2007.

[Wei13] L. Wei, Y. Xu, R. Q. Hu, and Y. Qian, "An algebraic framework for mobile association in wireless heterogeneous networks", *IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2013.

[Wig09] J. Wigard, T. Kolding, L. Dalsgaard, and C. Coletti, "On the user performance of LTE UE power savings schemes with discontinuous reception in LTE", *IEEE Int. Conf. on Communications (ICC)*, Jun. 2009.

[Wu04] X. Wu, and B. Murherjee andD. Ghosal, "Hierarchical architectures in the third-generation cellular network", *IEEE Wireless Comm.*, vol. 11, no. 3, pp. 62–71, Jun. 2004.

[Xia12] Z. Xiang, and M. Tao, "Robust beamforming for wireless information and power transmission", *IEEE Wireless Comm. Letters*, vol. 1, no. 4, pp. 372–375, Aug. 2012.

[Xu13] H. Xu, and P. Ren, "Joint user scheduling and power control for cell-edge performance improvement in backhaul-constrained network MIMO", *IEEE Int. Symp. on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2013.

[Yan04] E. Yanmaz, and O. K. Tonguz, "Dynamic load balancing and sharing performance of integrated wireless networks", *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 862–872, Jun. 2004.

[Yan11] J. Yang, and S. Ulukus, "Optimal packet scheduling in a multiple access channel with rechargable nodes", *IEEE Int. Conf. on Communications (ICC)*, Jun. 2011.

[Yan12a] J. Yang, O. Ozel, and S. Ulukus, "Broadcasting with an energy harvesting rechargeable transmitter", *IEEE Trans. Wireless Comm.*, vol. 11, no. 2, pp. 571–583, Feb. 2012.

[Yan12b] J. Yang, and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system", *IEEE Trans. Comm.*, vol. 60, no. 1, pp. 220–230, Jan. 2012.

[Yan12c] J. Yang, and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system", *IEEE Trans. Comm.*, vol. 60, no. 1, pp. 220–230, Feb. 2012.

[Yat15] R. D. Yates, and H. Mahdavi-Doost, "Energy harvesting receivers: Packet sampling and decoding policies", *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 558–570, Jan. 2015.

[Ye13a] Q. Ye, M. Al-Shalash, C Caramanis, and J. G. Andrews, "On/Off macrocells and load balancing in heterogeneous cellular networks", *IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2013.

[Ye13b] Q. Ye, *et al.*, "User association for load balancing in heterogeneous cellular networks", *IEEE Trans. Wireless Comm.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[Yi11] X. Yi, and E. K. S. Au, "User scheduling for heterogeneous multiuser MIMO systems: a subspace viewpoint", *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 4004–4013, Oct. 2011.

[Yoo06] T. Yoo, and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error", *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[You14] S. You, L. Chen, and Y. E. Liu, "Convex-concave procedure for weighted sum-rate maximization in a MIMO interference network", *IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2014.

[Yu14] J. Yu, *et al.*, "Power allocation for CoMP system with backhaul limitation", *IEEE Int. Conf. on Comm. (ICC))*, Jun. 2014.

[Zaf07] M. Zafer, and E. Modiano, "Delay-constrained energy efficient data transmission over a wireless fading channel", *IEEE Inf. Theory and Applications Workshop*, Jan. 2007.

[Zaf09] M. Zafer, and E. Modiano, "A calculus approach to energy-efficient data transmission with quality-of-service constraints", *IEEE/ACM Trans. on Networks*, vol. 17, no. 3, pp. 898–911, Jun. 2009.

[Zen15] Y. Zeng, and R. Zhang, "Optimized training design for wireless energy transfer", *IEEE Trans. Comm.*, vol. 63, no. 2, pp. 536–550, Feb. 2015.

[Zha10] R. Zhang, Y.-C. Liang, and S. Cui, "Dynamic resource allocation in cognitive radio networks", *IEEE Signal Process. Magazine*, vol. 27, no. 3, pp. 102–114, May 2010.

[Zha11] R. Zhang, and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer", *IEEE Global Communications Conf. (Globecom)*, 2011.

[Zha13] R. Zhang, and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer", *IEEE Trans. Wireless Comm.*, vol. 12, no. 5, pp. 1989–2001, May 2013.

[Zha15] T. Zhang, *et al.*, "User association for energy-load tradeoffs in HetNets with renewable energy supply", *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2214–2217, Dec. 2015.

[Zho08] L. Zhou, *et al.*, "Performance analysis of power saving mechanism with adjustable DRX cycles in 3GPP LTE", *IEEE Vehicular Technology Conf. (VTC Fall)*, Sep. 2008.

[Zho13a] L. Zhou, and W. Yu, "Uplink multicell processing with limited backhaul via per-base-station successive interference cancellation", *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 1981–1993, Oct. 2013.

[Zho13b] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: architecture design and rate-energy tradeoff", *IEEE Trans. Comm.*, vol. 61, no. 11, pp. 4757–4767, Nov. 2013.

[Zon16] Z. Zong, *et al.*, "Optimal transceiver design for SWIPT in K-user MIMO interference channels", *IEEE Trans. on Wireless Comm.*, vol. 15, no. 1, pp. 430–445, Jan. 2016.