Ultra Wideband OFDM Systems: Channel Estimation and Improved dtection Accounting for Estimation Inaccuracies

Seyed Mohammad Sajad Sadough

Université Paris-Sud 11

Faculté des sciences d'Orsay

THÈSE

presentée pour obtenir le grade de :

Docteur en sciences de l'Université Paris-Sud 11

Systèmes OFDM Ultra Large Bande : Estimation de canal et détection améliorée prenant en compte les imprécisions d'estimation

Presentée par:

Seyed Mohammad Sajad SADOUGH

Soutenue le **7 janvier 2008** *devant les membres du jury :*

Prof. Pascal Larzabal	Université Paris-Sud 11	*Président*
Prof. Luc Vandendorpe	Université catholique de Louvain	*Rapporteur*
Prof. Michel Terré	Conservatoire National des Arts et Métiers	*Rapporteur*
Prof. Hamid Aghvami	King's College London	*Examinateur*
Prof. Alain Sibille	École Nationale Supérieure de Tech. Avancées	*Directeur de thèse*
Prof. Pierre Duhamel	Laboratoire des Signaux et Systèmes-CNRS	*Directeur de thèse*

University of Paris-Sud 11

Faculty of Sciences of Orsay

THESIS
presented to obtain the degree of:

Doctor of Sciences of the University of Paris-Sud 11

Ultra Wideband OFDM Systems: Channel Estimation and Improved Detection Accounting for Estimation Inaccuracies

---

Presented by:

Seyed Mohammad Sajad SADOUGH

Defended on **January 7, 2008** *in front of the following jury:*

Prof. Pascal Larzabal	Université Paris-Sud 11	*President*
Prof. Luc Vandendorpe	Université catholique de Louvain	*Reviewer*
Prof. Michel Terré	Conservatoire National des Arts et Métiers	*Reviewer*
Prof. Hamid Aghvami	King's College London	*Examinator*
Prof. Alain Sibille	École Nationale Supérieure de Tech. Avancées	*Thesis advisor*
Prof. Pierre Duhamel	Laboratoire des Signaux et Systèmes-CNRS	*Thesis advisor*

Remerciements

Cette thèse est le fruit de trois années de recherche au sein du Laboratoire Electronique Informatique (UEI) de l'ENSTA et du Laboratoire des Signaux et Systèmes (LSS) du CNRS. Je tiens avant tout à remercier l'UEI pour m'avoir donné les moyens de préparer mon doctorat dans les meilleures conditions.

J'exprime toute ma gratitude envers les membres du jury: Pascal Larzabal professeur à l'Université Paris-Sud 11 qui en a été le président, Luc Vandendorpe professeur à l'Université catholique de Louvain et Michel Terré professeur au Conservatoire National des Arts et Métiers de Paris qui ont bien voulu en être les rapporteurs, Hamid Aghvami professeur à King's College de Londres qui a accepté d'examiner mes travaux.

Je tiens à remercier chaleureusement Alain Sibille pour sa patience, son soutien, ses conseils avisés et surtout pour m'avoir laissé mener mes recherches (et mes enseignements) avec beaucoup d'ouverture d'esprit. Merci aussi pour son incroyable disponibilité et sa bonne humeur lors des discussions d'ordre scientifique ou logistique !

Je tiens ensuite à exprimer ma profonde gratitude à Pierre Duhamel pour m'avoir accueilli dans son groupe au LSS, pour toute sa patience, sa rigueur et sa riche culture scientifique. Pierre a su me faire profiter de ses compétences en m'orientant dans mes travaux de recherche; je n'oublierai pas les longues heures de discussion passées dans son bureau.

Un grand merci à Emmanuel Jaffrot qui m'a accordé sa confiance dès mon stage de DEA et qui m'a donné la chance de réaliser ce travail. Je lui sais gré de son soutien efficace et compréhensif, de son inspiration, de ses compétences techniques et de ses qualités humaines. Je n'oublie pas ses encouragements répétés, lors des moments de doute qui ont alterné avec les moments de progrès.

Je remercie sincèrement Benoît Geller et Mohammad Ali Khalighi qui ont accepté de relire mon manuscrit et de me faire part de leurs remarques, toujours constructives. Ceci s'adresse aussi à Mohammad Ali, pour son aide et ses précieux commentaires sur les détecteurs soft-PIC.

Je remercie particulièrement mes collègues du LSS Mahieddine Ichir et Pablo Piantanida avec qui j'ai collaboré, et dont les contributions ont sans doute permis d'enrichir le contenu de

cette thèse.

Je remercie tous les enseignants-chercheurs, thésards et permanents de l'UEI et du LSS qui ont toujours été disponibles pour discuter ou m'aider sur certains aspects techniques. J'ai une pensée particulière pour Vladimir Koudritch qui nous a hélas quitté au milieu de la thèse, et de la reconnaissance envers Serge Bories pour avoir mis à ma disposition d'utiles mesures de canaux UWB. Merci également à Reza Mohammad-Khani pour nos nombreuses discussions techniques, ainsi qu'à Alexandre Renaux pour ses suggestions concernant la soutenance.

Enfin, je dédie cette thèse à tous les membres de ma famille qui m'ont soutenu et encouragé quotidiennement, et sans qui sans nul doute rien de tout cela n'aurait été possible.

À mon père, à ma mère et à mon épouse.

Paris, Janvier 2008

Seyed Mohammad Sajad SADOUGH

Contents

Résumé étendu en Français

Introduction générale

C'est un double *big bang*[1], l'un technologique, l'autre théorique, qui a ouvert ce que l'on appelle communément aujourd'hui l'ère de la communication. C'est en effet à quelques mois d'intervalle et dans le même établissement[2] que John Bardeen et ses collègues inventent le transistor et que Claude Elwood Shannon établit la théorie de l'information et des communications numériques. On mesure bien aujourd'hui toute l'importance de ces deux découvertes exceptionnelles qui ont permis le formidable essor de l'informatique et des télécommunications, entre autres. Depuis 1948, les fulgurants progrès de l'électronique puis de la micro-électronique, ont apporté aux ingénieurs les moyens de concrétiser les inventions des chercheurs afin de répondre aux défis lancés par Shannon. Un exemple typique en est l'invention, plutôt tardive, des turbocodes et des traitements itératifs dans les récepteurs, qui ne purent être imaginés que parce que les dizaines ou centaines de milliers de transistors requis étaient disponibles[3].

Aujourd'hui, les exigences en terme de communications sont de plus en plus grandes et diverses. Avec l'émergence de la téléphonie portable et de l'Internet, les méthodes de communications ont dû énormément s'adapter puisqu'il faut désormais pouvoir communiquer "n'importe quand, de n'importe où et de plus en plus vite". De plus, les services demandés ne se limitent plus à de la voix ou de simples messages écrits mais la transmission d'images et de vidéos en temps réel. Ainsi les débits de communications doivent être de plus en plus élevés tout en gardant une très bonne qualité de transmission et ce même quand le canal de propagation est très hostile. En effet, la plupart des transmissions radiomobiles s'effectuent sur des canaux d'autant plus sélectifs en fréquence que le débit de transmission croît.

Pour pallier à ces problèmes, l'utilisation des techniques multi-antennes et leurs couplages plus récents avec des modulations à sous porteuses orthogonales (de type OFDM) ont permis à

[1] Cette image est directement empruntée à l'avant-propos de l'ouvrage "Codes et turbocodes" écrit sous la direction de Claude Berrou.

[2] Laboratoires Bell.

[3] Ce premier paragraphe est inspiré de l'avant-propos de l'ouvrage "Codes et turbocodes".

la fois d'augmenter les débits tout en supprimant les interférences entre symboles dues au canal. Durant la dernière décennie, plusieurs schémas originaux ont permis d'améliorer considérablement l'exploitation de la diversité spatiale apportée par les systèmes multi-antennes. En 1998, ce sont les codes espace-temps orthogonaux proposés par Alamouti pour 2 antennes d'émission, puis généralisés par Tarokh à des systèmes à plus de 2 antennes d'émission qui ont permis d'améliorer sensiblement la fiabilité de la transmission. Par ailleurs, le multiplexage spatial (connu aussi sous le nom de V-BLAST) associé aux différentes techniques de détection développées par Foschini en 1996 permettent d'augmenter les débits linéairement avec le minimum du nombre d'antennes d'émission et de réception.

Outre les systèmes multi-antennes, les techniques *ultra large bande* (UWB), utilisées depuis le milieu des années 60 dans des applications de type radar, ont connu un spectaculaire regain d'intérêt ces dernières années. L'UWB qui consiste à utiliser des systèmes de transmissions dans des bandes de fréquences allant de 500 MHz à plusieurs GHz, se présente comme une technologie attractive pour les systèmes de communications radio à très hauts débits sur des distances relativement courtes.

Il est évident que pour la conception et le développement de toute technique de transmission radio, la connaissance du canal de propagation est prépondérante. Dans ce contexte, la quasi-totalité des études sur les systèmes pratiques de transmission radio sont basées sur une *parfaite* connaissance des paramètres du canal à l'émission et à la réception. De plus, la plupart des limites théoriques des communications numériques (par exemple la capacité de Shannon), ont été obtenues en considérant des conditions idéales de parfaite connaissance du canal. Cette hypothèse n'est pourtant pas valable pour les systèmes de communications pratiques où l'estimation du canal est la plupart du temps basée sur l'envoi des symboles d'apprentissage (aussi appelés pilotes). Il est clair que, du fait du nombre limité de pilotes et la présence du bruit additif du canal, l'émetteur et le récepteur n'obtiennent qu'un estimé *imparfait*, et parfois même très bruité, du vrai canal. Face à cette situation, l'utilisation des modulations différentielles permet d'éviter l'estimation du canal. Cependant, ces modulations engendrent une perte d'environ 3 dB sur les performances du récepteur. Le scénario décrit ci-dessus ouvre donc un grand chantier de recherche pour le développement des limites théoriques ainsi que la conception des systèmes de transmission pratiques prenant en compte une connaissance imparfaite du canal. En effet, il est nécessaire de redéfinir le concept fondamental de la capacité en considérant une connaissance imparfaite du canal: **i)** "Quel est le taux maximal avec lequel on peut transmettre sur un canal *imparfaitement* connu ?". Ensuite on peut se poser la question: **ii)** "Comment concevoir un système de transmission qui prenne en compte une estimation imparfaite du canal et qui se rapproche au mieux de cette limite?". D'un point de vue pratique, il

y a un grand intérêt à concevoir de nouveaux schémas de détection pouvant fournir des taux d'erreurs proches de la limite théorique en présence d'une estimation imparfaite du canal. Cette exigence constitue notre feuille de route tout au long de cette thèse.

Principales contributions

Ce rapport de thèse présente de nouvelles méthodes de réception itératives pour les systèmes mono- et multi-antennes associées à la modulation OFDM travaillant sur des canaux UWB. Nous nous plaçons dans un contexte réaliste où le récepteur ne possède que d'un estimé bruité du canal, obtenu à partir d'un nombre *limité* de symboles d'apprentissage. De plus, on suppose que cet estimé n'est pas disponible à l'émetteur. L'objectif principal de cette thèse est d'établir une connexion entre la méthode couramment utilisée d'estimation de canal par pilotes, et la conception de récepteurs itératifs qui prennent en compte dans leurs critères de détection, la présence d'erreurs sur les estimés du canal. De plus, une attention particulière est apportée au développement d'algorithmes qui gardent une complexité de calcul raisonnable.
Les principales contributions de cette thèse sont brièvement résumées ci-dessous:

- Nous commençons par donner un état de l'art des diverses techniques de transmission et de détection proposées pour les systèmes UWB en mettant l'accent sur la solution multi-bandes OFDM (MB-OFDM) qui constitue notre scénario d'application tout au long de cette thèse.

- Le premier objectif de cette thèse consiste à améliorer la qualité de l'estimation du canal sans modifier la structure du récepteur qui pour sa part est conçu en supposant une parfaite connaissance du canal. Nous donnons d'abord un état de l'art sur les techniques d'estimation de canal des systèmes OFDM. Ensuite, on propose une méthode d'estimation de canal semi-aveugle basée sur l'algorithme *Expectation-Maximization* (EM). Dans cette approche, les symboles pilotes fournissent un premier estimé du canal souvent très bruité qui est ensuite amélioré de manière itérative en intégrant l'estimation du canal dans le processus itératif de détection des données. Bien que cette approche soit connue pour améliorer sensiblement les performances des récepteurs itératifs, elle présente une grande complexité. Par conséquent, nous présentons un algorithme capable de réduire le nombre de paramètres à estimer en exploitant la parcimonie de la représentation des canaux UWB dans le domaine des ondelettes.

- Nous cherchons ensuite à améliorer les performances des récepteurs itératifs qui utilisent des estimés imparfaits du canal obtenue à partir d'un nombre limité de symboles pilotes.

Les deux configurations mono- et multi-antennes associées au MB-OFDM sont étudiées. Nous exploitons une propriété intéressante de l'estimation par pilotes, à savoir la disponibilité des statistiques de l'erreur d'estimation. Ces statistiques permettent d'établir un cadre Bayésien pour la conception de détecteurs robustes aux erreurs sur les estimés du canal. En considérant une détection au sens du maximum vraisemblance (ML), nous formulons une nouvelle métrique de détection qu'on appelle *métrique améliorée*, en moyennant la métrique classique (basée sur la connaissance parfaite du canal) sur toutes les erreurs d'estimation du canal. Nous considérons aussi un détecteur ML basé sur une *métrique désadaptée* (à la connaissance imparfaite du canal) qui est obtenue en remplaçant le vrai canal par son estimé dans la métrique de détection. Dans un premier temps, un récepteur maximum *a posteriori* itératif (turbo-MAP) basée sur la métrique ML améliorée, est proposé pour réduire les effets de l'incertitude sur le canal au niveau du décodeur. Ensuite, en utilisant des outils de la théorie d'information, on calcule les limites des taux de coupure atteignables (achievable outage rates) associées aux métriques ML améliorée et désadaptée. Ceci nous permet de déterminer les débits qu'on peut atteindre en pratique avec une qualité de service définie par la probabilité de coupure (outage probability) et ce même lorsque le récepteur ne possède que l'estimé du canal. Nos résultats numériques montrent que la métrique améliorée offre des gains significatifs en termes de taux d'information atteignables et de taux d'erreurs binaires, sans pour autant augmenter la complexité du récepteur. De plus, nos résultats peuvent servir à fixer les paramètres d'un système de transmission (par exemple le nombre et la puissance des symboles pilotes, la probabilité de coupure, etc.) visant une qualité de service en terme de taux d'erreurs binaires et taux de coupure atteignables, en présence d'une estimation imparfaite du canal.

- Finalement, nous présentons un détecteur itératif sous optimal et à complexité réduite pour les modulations à hautes efficacité spectrale transmises sur des canaux à antennes multiples. Ce détecteur est basé sur l'association d'une annulation parallèle d'interférences à entrées souples (soft-PIC) et d'un filtrage MMSE. Notre objectif sera une fois de plus d'apporter des modifications au détecteur afin de le rendre robuste aux incertitudes sur les estimés du canal.

Chapitre 2 :

Les différentes techniques de modulation et de détection UWB

Introduction et motivations

L'Ultra Large Bande ou *Ultra Wideband* (UWB), est une technique de transmission radioélectrique qui consiste à utiliser des signaux s'étalant sur une large bande de fréquences, typiquement de l'ordre de 500 MHz à plusieurs GHz. Une définition aujourd'hui communément admise est que les signaux UWB ont un rapport largeur de bande sur fréquence centrale, ou *fractional bandwidth*, au moins égal à 0,25 ou bien une largeur de bande supérieure à 500 MHz. Historiquement, la première forme de modulation proposée pour l'UWB a été la radio impulsionnelle, ou *Impulse Radio*. Elle se caractérise par des impulsions très brèves qui occupent instantanément toute la bande de fréquences disponible. C'est pourquoi ces modulations sont aussi connues sous le nom de modulations mono-bande. Cependant, cette approche permet peu de flexibilité dans l'utilisation du spectre radio, et nécessite des solutions de composants RF très performantes. Une autre solution consiste à diviser le spectre alloué à l'UWB en plusieurs sous-bandes de largeur minimale de l'ordre de 500 MHz: c'est l'approche multi-bandes. Cette solution présente une très grande flexibilité pour la gestion du spectre radio et permet d'utiliser des technologies de circuits intégrés moins onéreuses. En 2002, la *Federal Communications Commission* (FCC) aux Etats-Unis, a élargi la notion d'UWB à d'autres schémas de modulations que les transmissions par impulsions. Au niveau de la normalisation, l'institut américain IEEE a travaillé durant ces dernières années à la définition d'un système de communication haut débit utilisant le spectre radio UWB sans pouvoir arriver à un consensus sur le type de modulation. Le débat pour une solution unique s'articulait autour de deux propositions qui ont divisé les participants du groupe : l'étalement de spectre UWB (DS-UWB) et la modulation OFDM à bandes multiples (MB-OFDM). Aujourd'hui, faute d'un accord général sur une solution unique, d'importants groupes industriels, comme *UWB Forum* et *Wimedia Alliance* se sont engagés dans la conception d'équipements basés sur la technologie UWB en adoptant respectivement la soltution DS-UWB et MB-OFDM.

Les approches mono-bande

Techniques de modulations et d'accès multiples

Le principe des modulations mono-bande repose sur l'émission d'impulsions de très courtes durée $w_{\text{tr}}(t)$. Ils se distinguent selon le type de codage utilisé pour coder l'information binaire à transmettre. Ainsi, pour coder l'information à transmettre, la modulation par position

d'impulsion (PPM) utilise la position de l'impulsion par rapport à une position nominale; la modulation par amplitude d'impulsion (PAM) utilise des impulsions identiques avec une différence d'amplitude et la modulation par forme d'impulsion (PSM) utilise des signaux avec des formes différentes. Afin d'assurer l'accès multiple dans un contexte multi-utilisateurs, les modulations mono-bandes utilisent des codes à sauts temporels ou *time-hopping* (TH) ou une méthode basée sur l'étalement de spectre à séquence directe (DS-UWB).

L'approche TH consiste à répéter la même impulsion dans plusieurs trames successives. L'expression du signal transmis pour l'utilisateur j s'écrit [1,2] :
pour la modulation PPM :

$$s_{\mathrm{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\mathrm{tr}}\Big(t - kT_s - lT_f - c_l^{(j)} T_c - d_k^{(j)} \delta\Big), \tag{1}$$

pour la modulation PAM :

$$s_{\mathrm{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\mathrm{tr}}\Big(t - kT_s - lT_f - c_l^{(j)} T_c\Big) d_k^{(j)}, \tag{2}$$

et pour la modulation PSM :

$$s_{\mathrm{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\mathrm{tr}}^{d_k^{(j)}}\Big(t - kT_s - lT_f - c_l^{(j)} T_c\Big). \tag{3}$$

Dans cette approche, la durée T_s de transmission d'un symbole est divisée en N_s trames de durée T_f qui elles mêmes sont divisées en chips de durée T_c. Chaque trame contient une impulsion avec la position déterminée par la séquence de code $\{c_l^{(j)}\}$ pseudo aléatoires et unique pour l'utilisateur j.

Comme la modulation PPM, la modulation DS-UWB utilise une trame divisée en chips. Cependant, une impulsion UWB peut être émise dans chaque chip de la trame. En conséquence, le signal est transmis de façon continue. Les symboles transmis sont représentés par des codes d'étalement ternaires (i.e., composés de 1, 0 et -1) de la longueur de la trame. L'expression du signal DS-UWB transmis pour l'utilisateur j est donnée par

$$s_{\mathrm{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\mathrm{tr}}\Big(t - kT_s - lT_f\Big)\, c_l^{(j)}\, d_k^{(j)}. \tag{4}$$

Techniques de détection

En radio impulsionnelle, la réception des signaux se fait par corrélation. Les deux techniques utilisées sont basées sur les récepteurs corrélateurs (*correlation receiver*) et les récepteurs Rake. Le principe de base repose sur la corrélation du signal reçu avec un signal modèle ou *template*. Une décision binaire à partir du signe de la corrélation permet ensuite de démoduler les données

transmises. Par exemple, le critère de décision pour décider entre les hypothèse $\mathcal{H}_0$ (bit "0") et $\mathcal{H}_1$ (bit "1") dans le cas d'une modulation PPM s'écrit

$$(\text{décide } d^{(1)} = \text{“0”}) \Leftrightarrow \sum_{l=0}^{N_s-1} \int_{\tau_1+lT_f}^{\tau_1+(l+1)T_f} r(t)\, v\big(t-\tau_1-lT_f-c_l^{(1)}T_c\big)\,\mathrm{d}t > 0 \qquad (5)$$

où $r(t)$ est le signal reçu et $v(t)$ et le signal modèle utilisé pour la corrélation.

Dans un canal à trajets multiples, les techniques mono-bandes utilisent un récepteur Rake afin de combiner de façon constructive les signaux issus des différents trajets. Le signal reçu à l'entrée du récepteur Rake est corrélé avec des versions décalées d'un signal modèle, échantillonné, pondérées par des poids fixés par les paramètres du canal et enfin combinées de façon linéaire. Le nombre de corrélateurs (aussi appelé doigts) est fixé par le nombre de trajets qui caractérisent le canal de propagation. Cependant, la complexité d'un récepteur Rake augmente linéairement avec le nombre de ses doigts.

Les approches multi-bandes

La solution multi-bandes impulsionnelle

L'approche multi-bandes impulsionnelle consiste à diviser la bande de fréquence UWB en sous-bandes de largeurs proches de 500 MHz et à utiliser à l'intérieur de chaque sous-bande, l'une des modulations mono-bande décrite ci-dessus. Afin éviter les interférences inter-symboles, cette technique adopte une période de répétition des impulsions, sur chaque sous-bande, supérieure à l'étalement des retards du canal. Pour atteindre des débits importants, la solution repose sur une utilisation séquentielle des sous bandes par l'intermédiaire d'un saut de fréquence. L'avantage de cette approche est qu'un récepteur Rake avec moins de doigts est nécessaire pour démoduler le signal transmis sur chaque sous-bande.

La solution multi-bandes OFDM

La solution *Multiband Orthogonal Frequency Division Multiplexing* (MB-OFDM) est une approche à bandes multiples, où le spectre défini par la FCC est divisé en 14 sous-bandes de largeur 528 MHz chacune. Ces 14 sous-bandes sont subdivisé en 5 groupes de bandes partielles comportant chacun 2 ou 3 sous-bandes. Dans un premier temps, seul le groupe I (3,1 GHz - 4,9 GHz) sera exploité, les autres groupes étant utilisés à mesure du développement des solutions pour les composants RF. Dans chaque sous-bande, une modulation OFDM associée à une modulation codée à bits entrelacés (BICM) est appliquée, le signal étant réparti sur 128 porteuses à bande étroite. La modulation en bande de base pour chaque porteuse est de type *Quaternary Phase Shift Keying* (QPSK). Cette configuration permet une gestion très souple

du spectre radio. En effet, pour éviter de brouiller une bande de fréquences particulière, il suffit d'interdire une série de porteuses, voire la totalité d'une bande partielle. La gestion des utilisateurs multiples d'un même groupe de bandes partielles est opérée par une technique de codes temps-fréquence (TFC). Dans un groupe de bandes partielles, le signal d'un utilisateur passe régulièrement d'une sous-bande à une autre selon un ordre qui est déterminé par le TFC. Les débits offerts par cette technique s'étendent de 53,3 Mbps à 480 Mbps. Les références [3, 4] fournissent tous les détails nécessaires à l'implémentation d'un système MB-OFDM.

Les avantages de la solution MB-OFDM résident principalement dans sa faible complexité technique, la modulation OFDM présentant un grand degré de maturité et étant dèja adoptée par plusieurs standards (e.g., ADSL, DVB, 802.11a, 802.16a, etc.). La restriction de la bande de fréquences utilisée au premier groupe de bandes partielles permet également de profiter de systèmes et composants RF existants.

Chapitre 3 :
Estimation semi-aveugle de canal basée sur une représentation en ondelettes pour les systèmes multi-bandes OFDM

Introduction et motivations

Le schéma de réception proposé pour les systèmes MB-OFDM dans [3, 4] préconise une estimation de canal basée *uniquement* sur l'envoi de quelques symboles pilotes au début de la trame d'information. Il est bien connu qu'une détection fiable des données nécessite une estimation précise des paramètres du canal au niveau du récepteur. Cependant, l'obtention d'un tel estimé précis, uniquement par l'intermédiaire des pilotes, exigerait l'envoi de multiples symboles d'apprentissage, ce qui aurait pour conséquence une réduction significative de l'efficacité spectrale.

Des travaux récents ont montré des résultats prometteurs sur la combinaison de l'estimation de canal et le décodage de données. En particulier, un traitement itératif (à la *turbo*) qui inclut l'estimation du canal dans le processus de détection et de décodage conjoint des données, est effectué dans [5, 6] afin d'atteindre l'objectif d'excellentes performances avec un nombre très faible de symboles pilotes. Dans le même contexte, plusieurs travaux ont utilisé l'algorithme EM [7] pour estimer le canal et détecter les données de manière conjointe [8–10]. Bien que ces méthodes soient capables d'atteindre des performances très proches de celles obtenues avec une connaissance parfaite du canal, elles présentent toutefois une complexité croissante avec le nombre de paramètres qu'il faut estimer pour mettre à jour le canal à chaque itération.

L'augmentation de la largeur de bande en communication UWB a comme conséquence une meilleure résolution temporelle du canal. Cela implique une représentation parcimonieuse du canal (i.e., avec peu de coefficients significatifs) dans le domaine des ondelettes. Cette propriété est exploitée dans ce travail pour proposer un algorithme efficace estimant conjointement le canal et les symboles transmis. Plus précisément, ce chapitre propose une technique semi-aveugle basée sur l'algorithme EM pour estimer le canal dans le domaine des ondelettes. D'abord, une distribution *a priori* qui modélise bien la parcimonie du canal est imposée sur les coefficients d'ondelettes de la réponse impulsionnelle (RI) du canal. Cela rend l'estimation du canal au sens du MAP équivalente à un seuillage "dur" qui est exploité pour réduire itérativement le nombre de paramètres à estimer au sein de l'algorithme EM. Par ailleurs, puisque la probabilité des données codées est impliquée dans le calcul des paramètres du canal, nous combinons naturellement le processus itératif de l'estimation de canal avec l'opération de décodage des données codées. Nos résultats montrent que l'algorithme proposé réduit considérablement le nombre de paramètres à estimer tout en améliorant les performances du récepteur, comparé aux techniques semi-aveugles

classiques et aux méthodes basées uniquement sur les pilotes.

Formulation du problème

On considère une transmission MB-OFDM employant uniquement les trois premières sous-bandes. Chaque symbole OFDM contient N sous porteuses. A la réception, en supposant un préfixe cyclique plus long que l'étalement des retards du canal, la transmission OFDM convertit le canal multi trajet en N sous canaux parallèles avec un évanouissement plat. Ainsi, la réception du n-ième symbole OFDM émis sur la i-ième sous bande s'écrit

$$\mathbf{y}_{i,n} = \mathcal{D}_{\mathbf{s}_{i,n}} \bar{\mathbf{H}}_{i,n} + \mathbf{z}_{i,n} \quad i \in \{1,2,3\}, \quad n = 1, \ldots, N_{\mathrm{SYM}} \tag{6}$$

où $\mathcal{D}_{\mathbf{s}_{i,n}} \triangleq \mathrm{diag}(\mathbf{s}_{i,n})$, les vecteurs de taille $(1 \times N)$ $\mathbf{y}_{i,n}$, $\mathbf{s}_{i,n}$ and $\bar{\mathbf{H}}_{i,n}$ sont respectivement les symboles reçus et transmis, et la réponse fréquentielle du canal; le vecteur $\mathbf{z}_{i,n}$ est un bruit blanc Gaussien distribué suivant la loi $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbb{I}_N)$.

Bien qu'à l'émission le canal soit utilisé sur des sous bandes de 528 MHz, à la réception on concatène trois symboles OFDM reçus pour estimer le canal sur une bande de 1.584 GHz (correspondant à trois sous bandes). Cette approche est motivée par le fait qu'une largeur de bande plus importante implique des canaux plus épars dans le domaine des ondelettes. Cette opération nous amène à utiliser le modèle suivant

$$\mathbf{Y}_m = \mathcal{D}_{\mathbf{s}_m} \mathbf{H}_m + \mathbf{Z}_m \quad m = 1, \ldots, M_{\mathrm{SYM}} \tag{7}$$

où $\mathcal{D}_{\mathbf{s}_m} \triangleq \mathrm{diag}(\mathbf{S}_m)$, $\mathbf{Y}_m = [\mathbf{y}_{1,n}, \mathbf{y}_{2,n}, \mathbf{y}_{3,n}]^T$, $\mathbf{S}_m = [\mathbf{s}_{1,n}, \mathbf{s}_{2,n}, \mathbf{s}_{3,n}]^T$, $\mathbf{H}_m = [\bar{\mathbf{H}}_{1,n}, \bar{\mathbf{H}}_{2,n}, \bar{\mathbf{H}}_{3,n}]^T$ et $\mathbf{Z}_m = [\mathbf{z}_{1,n}, \mathbf{z}_{2,n}, \mathbf{z}_{3,n}]^T$ sont des vecteurs de taille $(M \times 1)$, avec $M = 3N$ et $M_{\mathrm{SYM}} = N_{\mathrm{SYM}}/3$. Dans ce qui suit on omet l'index temporel m.

Soit $\mathbf{h} = [h_1, ..., h_L]^T$ la réponse impulsionnelle (RI) du canal sur les trois sous-bandes et $\mathbf{g} = [g_1, ..., g_L]^T$ le vecteur de la transformée en ondelettes (TO) de $\mathbf{h}$. On définit $\mathbf{F}_{M,L}$ comme la matrice tronquée de FFT et $\mathbf{W}$ comme la matrice de la transformée en ondelettes orthonormales. Le modèle (7) s'écrit

$$\mathbf{Y} = \mathcal{D}_{\mathbf{s}} \mathbf{T}\, \mathbf{g} + \mathbf{Z} \tag{8}$$

où $\mathbf{T} = \mathbf{F}_{M,L} \mathbf{W}^{\dagger}$. Notre objectif est d'estimer le vecteur des coefficients d'ondelettes $\mathbf{g}$ à partir du modèle (8).

Estimation de canal par l'algorithme EM dans le domaine des ondelettes

Choix de la loi *a priori* **:** La parcimonie des coefficients d'ondelettes est modélisée par la supposition *a priori* suivante : chaque coefficient d'ondelette est supposé être égal à zéro avec

une probabilité λ ou être une variable aléatoire gaussienne $\mathcal{CN}_{g_j}(0,\tau^2)$ avec une probabilité $1-\lambda$. Cela correspond à une loi i.i.d. Bernoulli-Gaussien qui s'écrit

$$\pi(g_j) = \lambda\,\delta(g_j) + (1-\lambda)\,\mathcal{CN}_{g_j}(0,\tau^2) \qquad j = 1,\ldots,L \tag{9}$$

où λ et τ sont des hyper-paramètres qui seront estimés par les observations.

Un modèle équivalent et le principe de l'algorithme EM : Dans l'objectif d'avoir un modèle où les coefficients d'ondelettes sont directement bruités par un bruit additif, on propose de décomposer le bruit $\mathbf{Z}$ de la manière suivante

$$\mathbf{Z} = \mathcal{D}_{\mathbf{S}}\mathbf{Z}_1 + \mathbf{Z}_2 \tag{10}$$

où $\mathbf{Z}_1 \sim \mathcal{CN}(\mathbf{0}, \alpha^2\mathbb{I}_M)$ et $\mathbf{Z}_2 \sim \mathcal{CN}(\mathbf{0}, (\sigma^2-\alpha^2)\mathbb{I}_M)$. On définit $\rho \triangleq \alpha^2/\sigma^2$ comme la proportion de bruit attribué à $\mathbf{Z}_2$. Ainsi nous pouvons récrire le modèle (8) de manière équivalente comme suit

$$\left\{\begin{array}{ll} \widetilde{\mathbf{H}} & = \mathbf{T}\,\mathbf{g} + \mathbf{Z}_1 \\ \mathbf{Y} & = \mathcal{D}_{\mathbf{S}}\,\widetilde{\mathbf{H}} + \mathbf{Z}_2. \end{array}\right. \tag{11}$$

Ce nouveau modèle attribue implicitement une partie du bruit aux coefficients d'ondelettes et le reste aux observations.

Après avoir initialisé le vecteur des paramètres $\mathbf{g}$ par quelques symboles d'apprentissages, l'algorithme EM alterne (jusqu'à ce qu'un critère de convergence soit atteint) entre les étapes E et M et fournit la séquence des estimés $\{\mathbf{g}^{(t)},\, t = 0,1,\ldots,t_{\max}\}$.

- **Etape d'espérance** :

$$Q\big(\mathbf{g},\mathbf{g}^{(t)}\big) = \mathbb{E}_{\mathbf{S},\widetilde{\mathbf{H}}}\Big[\log p(\mathbf{Y},\mathbf{S},\widetilde{\mathbf{H}}|\mathbf{g})\Big|\mathbf{Y},\mathbf{g}^{(t)}\Big] \tag{12}$$

- **Etape de maximisation** :

$$\mathbf{g}^{(t+1)} = \arg\max_{\mathbf{g}}\Big\{Q\big(\mathbf{g},\mathbf{g}^{(t)}\big) + \log\pi(\mathbf{g})\Big\}. \tag{13}$$

En appliquant les étapes E et M ci-dessus au modèle (11), l'estimé au sens du MAP des coefficients d'ondelettes du canal est donné par une simple règle de seuillage dur :

$$\boxed{g_j^{(t+1)} = \left\{\begin{array}{ll} 0, & \text{if } \beta_j^{(t+1)} = 0 \\ \dfrac{\tau^2}{\alpha^2+\tau^2}\,\widetilde{g}_j^{(t+1)}, & \text{if } \beta_j^{(t+1)} = 1 \end{array}\right.} \tag{14}$$

où $\widetilde{\mathbf{g}}^{(t)} = (1-\rho)\,\mathbf{g}^{(t)} + \rho\,(\overline{\mathcal{D}}_{\mathbf{S}}\mathbf{T})^{\dagger}\mathbf{Y}$, $\overline{\mathcal{D}}_{\mathbf{S}} = \sum_{\mathbf{s}\in\mathscr{C}}\mathcal{D}_{\mathbf{S}}\,p(\mathbf{S}|\mathbf{Y},\mathbf{g}^{(t)})$ et $\beta_j^{(t+1)}$ est une variable indicatrice (cf Appendice A).

Réduction du nombre des paramètres estimés

L'expression de l'estimateur (14) étant équivalente à un seuillage dur, on peut facilement réduire le nombre de coefficients à estimer. Cela se fait à chaque itération, en ne gardant dans le cycle itératif d'estimation, que les coefficients qui n'ont pas précédemment été remplacés par zéro. Ainsi le nombre de paramètres à estimer se réduit au fil des itérations. Cet opération est modélisée par

$$\mathbf{g}_{\text{tr}}^{(t+1)} = \Theta\big(\,\mathbf{g}^{(t+1)}\,\big), \quad \mathbf{T}_{\text{tr}} = \Xi\big(\,\mathbf{T}\,\big) \tag{15}$$

où l'opérateur $\Theta(.)$ rassemble dans $\mathbf{g}_{\text{tr}}^{(t+1)}$ les composants de $\mathbf{g}^{(t+1)}$ qui restent dans le processus d'estimation et l'opérateur $\Xi(.)$ construit la matrice tronquée $\mathbf{T}_{\text{tr}}$ à partir de la matrice $\mathbf{T}$.

Détection et décodage conjoint

Au récepteur, la détection et le décodage se fait conjointement et de manière itérative. Le récepteur est composé d'un détecteur qui fournit les probabilités sur les symboles émis et d'un décodeur à entrées et sorties souples (SISO). Chaque partie tire avantage de la probabilité *a posteriori* fournie par l'autre partie comme une information supplémentaire. Ici le décodeur employé est le décodeur BCJR [11]. Il est clair d'après l'expression (14) que la probabilité des symboles émis $P\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)$ est nécessaire pour mettre à jour le canal à chaque itération. Par ailleurs, le décodeur SISO a besoin d'un estimé du canal afin de calculer la probabilité des bits codés. Par conséquent, l'algorithme proposé se combine naturellement avec le processus itératif de décodage des données. La probabilité $P\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)$ est calculée par le décodeur comme suit

$$P\big(S_k|Y_k, \widehat{H}_k^{(t)}\big) = \prod_{\substack{j=1 \\ j\neq i}}^{B} P_{\text{dec}}(c_{k,j}) \tag{16}$$

où $P_{\text{dec}}(c_{k,j})$ est la probabilité *a priori* des bits codés et entrelacés $c_{k,j}$ qui est fournie par le décodeur SISO.

Conclusion

Dans ce chapitre, nous avons proposé un algorithme semi aveugle d'estimation de canal basé sur une représentation en ondelettes de la RI du canal. En choisissant une distribution *a priori* qui modélise bien la parcimonie du canal dans le domaine des ondelettes, nous avons pu rendre l'estimateur au sens du MAP équivalent à un seuillage dur des coefficients d'ondelettes. Ceci a été utilisé pour réduire itérativement le nombre de paramètres à estimer. Nous avons observé que quand le canal a une représentation parcimonieuse, le model a priori choisi est capable d'apporter cette information supplémentaire à l'estimateur du canal. De plus, nous avons montré que dans

ce cas, l'algorithme proposé réduit considérablement le nombre de paramètres à estimer et présente de meilleurs performances que les méthodes semi aveugles classiques et les méthodes uniquement basées sur les pilotes.

Chapitre 4 :

Multi-bandes MIMO-OFDM: détection améliorée et taux de transmission atteignables avec une estimation imparfaite du canal

Introduction et motivations

Ce chapitre vise à étudier un schéma itératif de détection et de décodage conjoint dans un système de transmission radio pratique, où le récepteur ne dispose que de l'estimé *imparfait* (et parfois même de qualité très médiocre) du vrai canal. Pour obtenir les paramètres du canal, on adopte une estimation *uniquement* basée sur l'envoi de quelques symboles pilotes. Dans ce contexte, l'obtention d'un estimé précis du canal, uniquement par l'intermédiaire des pilotes, exigerait l'envoi de multiples symboles d'apprentissage, ce qui aura pour conséquence une diminution des taux d'information atteignables. Pour pallier à ce problème, plutôt que d'adopter une technique itérative d'estimation de canal comme au chapitre précédant, on propose dans ce chapitre de modifier la structure du récepteur afin de prendre en compte la présence d'erreurs sur les estimés du canal. Nous utilisons pour cela les statistiques de l'erreur d'estimation du canal. En effet, ces statistiques sont disponible dans le cas où on estime le canal par des symboles pilotes. Nous dérivons une *métrique améliorée* pour la détection ML en moyennant la métrique classique (basée sur une connaissance parfaite du canal) sur les erreurs d'estimation de canal. Cette approche est une alternative à une détection dite *désadaptée* qui remplace le vrai canal par son estimé dans la métrique de détection classique.

Le scénario décrit ci-dessus suscite deux questions importantes: **i)** "Quel récepteur pratique peut améliorer les performances du système avec une estimation imparfaite du canal ?", **ii)** "Quelles sont les limites des taux de transmission fiables associés aux détecteurs ML en présence d'erreurs sur les estimés du canal ?" L'objectif de ce chapitre est d'étudier ces deux questions pour les systèmes MB-OFDM mono- et multi-antennes. Dans un premier temps, et afin de prendre en compte l'incertitude sur le canal, on propose un récepteur itératif basé sur une métrique ML adaptée à l'estimé du canal. Nous cherchons ensuite à calculer les taux atteignables associés aux métriques améliorée et désadaptée en terme de taux de coupure atteignables. Nous verrons que la métrique désadaptée est largement sous optimale quand un nombre limité de pilotes est dédié à l'estimation du canal et que des gains considérables sont apportés par la métrique améliorée sans augmentation de la complexité du récepteur.

Modèle de transmission et estimation de canal

MB-OFDM avec multiplexage spatial : On considère un système MB-OFDM avec M_T antennes d'émission, M_R antennes de réception, M sous porteuses dans chaque sous bande,

et employant une modulation 16-QAM. Le modèle à temps discret et en bande de base de la transmission s'écrit

$$\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{s}_k + \boldsymbol{z}_k, \qquad k = 0, ..., M-1 \tag{17}$$

où $\boldsymbol{y}_k$ est le vecteur des symboles reçus, $\boldsymbol{s}_k$ est le vecteur des symboles transmis avec une puissance $E_s = \frac{1}{M_T}\mathbb{E}[\,\mathrm{tr}(\boldsymbol{s}_k \boldsymbol{s}_k^\dagger)\,]$. Le canal $\mathbf{H}_k$ est une matrice dont chaque entrée H_{ij} est donnée par

$$H_{ij}(k) = \sum_{l=0}^{L-1} \alpha_{ij}(l)\, \exp\{-\mathbf{j}\, 2\pi k \Delta_f \tau_l\}$$

où $\alpha_{ij}(j)$ et τ_l représentent respectivement le gain et le retard du l-ième composant multi trajet de la RI du canal, et Δ_f est la bande fréquentielle occupée par une sous porteuse. Le vecteur $\boldsymbol{z}_k$ est un bruit AWGN avec une matrice de covariance $\mathbf{\Sigma}_{z,k} = \mathbb{E}[\,\boldsymbol{z}_k \boldsymbol{z}_k^\dagger\,] = \sigma_z^2\, \mathbb{I}_{M_R}$, k représente l'index de la sous porteuse.

Estimation de canal : On consacre N_P utilisations de canal pour transmettre la séquence d'apprentissage $\mathbf{S}_{P,k}$ avec une puissance moyenne $E_P = \frac{1}{N_P M_T}\mathrm{tr}\left(\mathbf{S}_{P,k}\mathbf{S}_{P,k}^\dagger\right)$. On estime ensuite le canal $\mathbf{H}_k$ au sens du maximum vraisemblance à partir des observations. Cela conduit à $\widehat{\mathbf{H}}_k^{\mathrm{ML}} = \mathbf{H}_k + \boldsymbol{\mathcal{E}}$ où $\boldsymbol{\mathcal{E}}$ est la matrice d'erreur d'estimation. La matrice de covariance de chaque colonne de $\boldsymbol{\mathcal{E}}$ est égale à $\mathbf{\Sigma}_{\mathcal{E}} = \sigma_{\mathcal{E}}^2\, \mathbb{I}_{M_R}$. A présent, en choisissant une distribution *a priori* pour la matrice $\mathbf{H}_k$ suivant la loi $\mathcal{CN}\big(\mathbf{0}, \mathbb{I}_{M_T} \otimes \mathbf{\Sigma}_{H_k}\big)$ et en utilisant le modèle linéaire ci-dessus pour l'estimé $\widehat{\mathbf{H}}_k^{\mathrm{ML}}$, on peut dériver la distribution *a posteriori* du vrai canal sachant son estimé (voir l'Appendice B) :

$$p\left(\mathbf{H}_k | \widehat{\mathbf{H}}_k^{\mathrm{ML}}\right) = \mathcal{CN}\left(\mathbf{\Sigma}_\Delta \widehat{\mathbf{H}}_k^{\mathrm{ML}},\, \mathbb{I}_{M_T} \otimes \mathbf{\Sigma}_\Delta \mathbf{\Sigma}_{\mathcal{E}}\right) \tag{18}$$

où

$$\mathbf{\Sigma}_\Delta = \mathbf{\Sigma}_{H_k}\big(\mathbf{\Sigma}_{\mathcal{E}} + \mathbf{\Sigma}_{H_k}\big)^{-1} = \delta\, \mathbb{I}_{M_R}.$$

Détection améliorée avec estimation imparfaite du canal

Soit $f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k)$ le critère de décision dans un détecteur donné. Suivant le critère de décision, la fonction $f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k)$ peut être la loi a posteriori $p(\boldsymbol{s}_k | \boldsymbol{y}_k, \mathbf{H}_k)$, le logarithme de la fonction de vraisemblance $W(\boldsymbol{y}_k | \mathbf{H}_k, \boldsymbol{s}_k)$, ou encore l'erreur quadratique moyenne. Comme le canal inconnu $\mathbf{H}_k$ est impliqué dans la détection, on propose une détection basée sur la fonction de coût suivante:

$$\begin{aligned} \widetilde{f}(\boldsymbol{y}_k, \boldsymbol{s}_k, \widehat{\mathbf{H}}_k) &= \int_{\mathbf{H}_k} f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k)\, p(\mathbf{H}_k | \widehat{\mathbf{H}}_k)\, \mathrm{d}\mathbf{H}_k \\ &= \mathbb{E}_{\mathbf{H}_k | \widehat{\mathbf{H}}_k}\left[f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k) \middle| \widehat{\mathbf{H}}_k \right] \end{aligned} \tag{19}$$

où $p(\mathbf{H}_k | \widehat{\mathbf{H}}_k)$ est la distribution donnée par (18).

Le détecteur ML désadapté : Cette approche *sous optimale* repose sur une fonction de vraisemblance dans laquelle on remplace le vrai canal par son estimé. Cela conduit à la distance Euclidienne suivante comme critère de détection:

$$\mathcal{D}_{\text{MM}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k) = -\log W(\boldsymbol{y}_k|\boldsymbol{s}_k, \widehat{\mathbf{H}}_k) \propto \|\boldsymbol{y}_k - \widehat{\mathbf{H}}_k \boldsymbol{s}_k\|^2. \tag{20}$$

Le détecteur ML amélioré : En utilisant l'expression générale (19), on peut évaluer la métrique de détection améliorée au sens ML comme suit

$$\begin{aligned}\widetilde{W}(\boldsymbol{y}_k|\widehat{\mathbf{H}}_k, \boldsymbol{s}_k) &= \int_{\mathbf{H}_k \in \mathbb{C}^{M_R \times M_T}} W(\boldsymbol{y}_k|\mathbf{H}_k, \boldsymbol{s}_k)\, p(\mathbf{H}_k|\widehat{\mathbf{H}}_k) \quad \mathrm{d}\mathbf{H}_k \\ &= \mathbb{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\Big[W(\boldsymbol{y}_k|\mathbf{H}_k, \boldsymbol{s}_k)\big|\, \widehat{\mathbf{H}}_k\Big]. \end{aligned} \tag{21}$$

La distance Euclidienne associée à la métrique améliorée est donnée par (voir Appendice B)

$$\begin{aligned}\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k) &= -\log \widetilde{W}(\boldsymbol{y}_k|\widehat{\mathbf{H}}_k, \boldsymbol{s}_k) \\ &= M_R \log\, \pi\big(\sigma_z^2 + \delta\, \sigma_{\mathcal{E}}^2\, \|\boldsymbol{s}_k\|^2\big) + \frac{\|\boldsymbol{y}_k - \delta\, \widehat{\mathbf{H}}_k\, \boldsymbol{s}_k\|^2}{\sigma_z^2 + \delta\, \sigma_{\mathcal{E}}^2\, \|\boldsymbol{s}_k\|^2}. \end{aligned} \tag{22}$$

Détection et décodage itératif conjoint

A la réception, on adopte un schéma de détection et de décodage itératif conjoint composé d'un détecteur souple et d'un décodeur SISO. Afin de prendre en compte l'erreur d'estimation du canal, on propose d'utiliser la métrique ML améliorée ci-dessus dans le détecteur. Soit $d_k^{j,m}$ le m-ième ($m = 1, ..., BM_T$) bit codé et entrelacé du vecteur $\boldsymbol{s}_k$, transmis à partir de l'antenne d'émission j et sur la sous porteuse d'indice k. Soit $\mathcal{S}$ l'ensemble de tous les symboles transmis possibles $\boldsymbol{s}_k$. On partitionne $\mathcal{S}$ en deux sous ensembles $\mathcal{S}_0^m$ et $\mathcal{S}_1^m$, pour lesquels le m-ième bit de $\boldsymbol{s}_k$ vaut respectivement "0" ou "1". On dénote par $L(d_k^{j,m})$ le logarithme du rapport de vraisemblance (LRV) du bit $d_k^{j,m}$ en sortie du détecteur.
Le LRV du détecteur proposé est donné par

$$L(d_k^{j,m}) = \log \frac{\displaystyle\sum_{\boldsymbol{s}_k \in \mathcal{S}_1^m} e^{-\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k)} \prod_{\substack{n=1 \\ n \neq m}}^{B} P_{\text{dec}}^1\big(d_k^{j,n}\big)}{\displaystyle\sum_{\boldsymbol{s}_k \in \mathcal{S}_0^m} e^{-\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k)} \prod_{\substack{n=1 \\ n \neq m}}^{B} P_{\text{dec}}^0\big(d_k^{j,n}\big)}, \tag{23}$$

où $P_{\text{dec}}^1(d_k^{j,n})$ et $P_{\text{dec}}^0(d_k^{j,n})$ sont les probabilités *a priori* fournies par le décodeur SISO.

Taux de transmission atteignables associés à la métrique ML améliorée

Taux atteignables instantanés : En utilisant des outils de la théorie d'information, les taux atteignables $C_{\mathcal{M}}$ associés à la métrique $\mathcal{D}_{\mathcal{M}}$ (22) sont donnés par

$$\boxed{C_{\mathcal{M}}(\mathbf{H},\widehat{\mathbf{H}}) = \frac{1}{M}\sum_{k=0}^{M-1}\log_2\det\left\{\mathbb{I}_{M_R} + \frac{\bar{P}\boldsymbol{\Upsilon}_{\text{opt},k}\boldsymbol{\Upsilon}_{\text{opt},k}^{\dagger}}{\sigma_k^2\big(\underline{\mu}_{\mathcal{M},k}^{\text{opt}}\big)}\right\}} \tag{24}$$

où $\mathbf{H}_k = \mathbf{U}_k\,\text{diag}(\underline{\lambda}_k)\mathbf{V}_k^{\dagger}$; nous définissons $\widetilde{\mathbf{H}}_k^{\dagger} = \mathbf{V}^{\dagger}\widehat{\mathbf{H}}_k^{\dagger}\mathbf{U}$ et le vecteur $\widetilde{\mathbf{h}}_k = \text{diag}(\widetilde{\mathbf{H}}_k^{\dagger})$ resultant de sa diagonale.

La solution optimale est:

$$\boldsymbol{\Upsilon}_{\text{opt},k} = \mathbf{U}_k\,\text{diag}\left(\underline{\mu}_{\mathcal{M},k}^{\text{opt}}\right)\mathbf{V}_k^{\dagger}$$

avec $\sigma_k^2\big(\underline{\mu}_{\mathcal{M},k}^{\text{opt}}\big) = \frac{\bar{P}}{M_R}\left(\|\underline{\lambda}_k\|^2 - \|\underline{\mu}_{\mathcal{M},k}^{\text{opt}}\|^2\right) + \sigma_z^2$ et

$$\underline{\mu}_{\mathcal{M},k}^{\text{opt}} = \begin{cases} \left(\dfrac{\sqrt{b_{\mathcal{M},k}}}{\|\widetilde{\mathbf{h}}_k\|} - |a_{\mathcal{M}}|\right)\widetilde{\mathbf{h}}_k & \text{si } b_{\mathcal{M},k} \geq 0 \\ \mathbf{0} & \text{sinon} \end{cases} \tag{25}$$

où $a_{\mathcal{M}}$ et $b_{\mathcal{M},k}$ sont des constantes définies dans le chapitre 4.

Taux de coupure moyennes : La notion de taux de coupure ou *outage rates* est plus appropriée dans les scénarios où le canal reste invariant sur toute la durée de la transmission, ce qui est le cas en MB-OFDM. Pour un taux de coupure R et une réalisation du canal $\mathbf{H}$, nous utilisons la distribution (18) pour définir la probabilité de coupure (*outage probability*) comme

$$P_{\mathcal{M}}^{\text{out}}(R,\widehat{\mathbf{H}}) = \int_{\{\mathbf{H}\in\Lambda_{\mathcal{M}}(R,\widehat{\mathbf{H}})\}} p(\mathbf{H}|\widehat{\mathbf{H}})\,\mathrm{d}\mathbf{H}, \tag{26}$$

avec l'ensemble $\Lambda_{\mathcal{M}}(R,\widehat{\mathbf{H}}) = \big\{\mathbf{H}\in\mathbb{C}^{MM_R\times MM_T} : C_{\mathcal{M}}(\mathbf{H},\widehat{\mathbf{H}}) < R\big\}$ où la matrice bloc diagonale $\mathbf{H}$ est donnée par

$$\mathbf{H} = \text{diag}\Big(\big[\;\mathbf{H}_0 \quad \mathbf{H}_1 \quad \cdots \quad \mathbf{H}_{M-1}\;\big]\Big).$$

Les taux de coupure pour une probabilité de coupure γ sont donnés par

$$C_{\mathcal{M}}^{\text{out}}(\gamma,\widehat{\mathbf{H}}) = \sup_R\big\{R\geq 0 : P_{\mathcal{M}}^{\text{out}}(R,\widehat{\mathbf{H}}) \leq \gamma\big\}. \tag{27}$$

Comme les taux de coupure donnés par (27) dépendent de l'estimé $\widehat{\mathbf{H}}$, on moyenne les taux de coupure sur les estimés du canal:

$$\overline{C}_{\mathcal{M}}^{\text{out}}(\gamma) = \mathbb{E}_{\widehat{\mathbf{H}}}\big[C_{\mathcal{M}}^{\text{out}}(\gamma,\widehat{\mathbf{H}})\big]. \tag{28}$$

En suivant une démarche similaire à ce qui a été présenté ci-dessus, nous pouvons calculer les taux de coupures moyennes associés au détecteur ML désadapté (20).

Conclusion

En utilisant les statistiques de l'erreur d'estimation de canal, nous avons proposé une nouvelle approche pour concevoir des détecteurs prenant en compte les imprécisions d'estimation. Dans le cas d'une détection au sens du maximum de vraisemblance, notre approche conduit à une métrique améliorée que nous avons utilisé dans la formulation d'un détecteur MAP itératif. De plus, nous avons calculé les taux de coupure atteignables associés aux métriques ML améliorées et désadaptées. Nos résultats de simulation ont montrés que l'approche désadaptée est largement sous optimale en terme de taux d'erreurs binaires et de taux de coupure atteignables surtout quand la séquence d'apprentissage du canal est courte. Ils ont aussi confirmé le fait que le détecteur amélioré est plus adapté à la présence d'erreurs d'estimation de canal. Notons que le gain de performance du détecteur proposé a été obtenu sans nécessiter l'augmentation de la complexité du récepteur.

Chapitre 5 :
Détection itérative à complexité réduite prenant en compte les erreurs d'estimation de canal

Introduction et motivations

Au chapitre précédent, nous avons proposé un récepteur itératif basé sur une détection améliorée au sens du MAP. Bien que la détection MAP soit connue pour être la meilleure stratégie pour minimiser le taux d'erreur binaire, elle présente toutefois une complexité calculatoire qui s'accroît avec le nombre de bits transmis à chaque utilisation du canal (i.e., avec le nombre d'antennes en réception et la taille de la constellation). Par conséquent, il est d'un grand intérêt pratique de considérer des détecteurs à faible complexité pouvant fournir des performances proche de la solution optimale.

Les solutions à complexités réduites proposées dans la littérature se divisent en deux catégories. La première catégorie regroupe les solutions qui reposent sur une marginalisation partielle (i.e., sur un ensemble réduit) plutôt que sur une marginalisation complète sur l'ensemble de tous les vecteurs possibles. Dans cette catégorie on peut nommer le décodage par sphères et les techniques de relaxation semi définies [12–14] ou encore le décodage par liste [15, 16]. Les solutions de la deuxième catégorie sont basées sur des techniques de filtrages linéaires.

Dans ce travail, on considère une approche basée sur un filtrage linéaire MMSE et une annulation parallèle d'interférence à entrées souples (soft-PIC). Cette approche a été proposée pour la première fois par Wang et Poor [17] pour les systèmes CDMA multi utilisateurs et a par la suite été appliquée et étendue par Tuchler [18], Dejonghe et Vandendorpe [19, 20] dans le contexte de l'égalisation itérative. En considérant le multiplexage spatial, Sellathurai et Haykin ont proposé une réception itérative basée sur le filtrage MMSE et soft-PIC [21]. Cependant, la majeure partie des travaux ont considéré une *parfaite* connaissance du canal au récepteur. L'objectif de notre travail est de proposer un détecteur MMSE pour les systèmes MIMO, qui soit robuste aux incertitudes sur le canal, dues à une estimation basée uniquement sur les pilotes. Pour cela, on adopte le même cadre probabiliste introduit au chapitre précédant pour calculer la métrique ML améliorée. Nous dérivons un récepteur à complexité réduite qui prend en compte les erreurs d'estimation de canal dans la formulation du filtre MMSE mais aussi au niveau de l'annulation des interférences. Une méthode simplifiée pour calculer les coefficients du filtre MMSE est aussi présentée. Nous montrons finalement plusieurs scénarios de simulation dans le cas du multiplexage spatial et les codes espace-temps. Nous observons que dans le cas des codes espace-temps, le récepteur est très sensible aux imperfections sur les estimés du canal.

Modèle de transmission et estimation de canal

Canal MIMO : On considère une transmission à antennes multiples avec M_T antennes d'émission et M_R antennes de réception à travers un canal non sélectif en fréquence. La séquence binaire est codée par un code convolutif avant d'être entrelacée de manière pseudo-aléatoire. Ensuite, la séquence codée est transformée en symboles QPSK ou 16-QAM qui sont envoyés vers un encodeur espace-temps. Dans ce qui suit, on considère le multiplexage spatial. La généralisation au cas des codes espace-temps est présentée dans l'Appendice D. On suppose une trame de symboles correspondant à L utilisations de canal, transmis à travers un canal représenté par la matrice $\mathbf{H}$ de taille $(M_R \times M_T)$. Le vecteur de symboles reçus $\boldsymbol{y}_k$ est donné par la relation

$$\boldsymbol{y}_k = \mathbf{H}\,\boldsymbol{s}_k + \boldsymbol{z}_k \qquad k = 1, ..., L, \tag{29}$$

où $\boldsymbol{s}_k$ représente le vecteur de taille $(M_T \times 1)$ des symboles transmis avec une puissance $E_s = \frac{1}{M_T}\mathbb{E}[\,\mathrm{tr}(\boldsymbol{s}_k \boldsymbol{s}_k^\dagger)\,]$. On suppose que le canal est distribué selon la loi $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbb{I}_{M_T} \otimes \mathbf{\Sigma}_H)$ avec $\mathbf{\Sigma}_H = \sigma_h^2 \mathbb{I}_{M_R}$. Le vecteur $\boldsymbol{z}_k$ est un bruit AWGN avec une matrice de covariance $\mathbf{\Sigma}_z = \sigma_z^2\ \mathbb{I}_{M_R}$. De plus, on considère un canal avec des évanouissements par bloc, où à chaque trame correspond N_c blocs d'évanouissement indépendants.

Estimation de canal : Pour estimer le canal correspondant à chaque bloc d'évanouissement, on envoie N_P symboles pilotes en plus des données. Dans ce cas, en supposant des séquences d'apprentissage orthogonales, l'estimé au sens du maximum vraisemblance du canal conduit à $\widehat{\mathbf{H}}_{\mathrm{ML}} = \mathbf{H} + \boldsymbol{\mathcal{E}}$. Cette estimation peut être caractérisée par la loi *a posteriori* suivante (voir Appendice B):

$$p\,(\mathbf{H}|\widehat{\mathbf{H}}_{\mathrm{ML}}) = \mathcal{CN}\Big(\delta\,\widehat{\mathbf{H}}_{\mathrm{ML}},\, \mathbb{I}_{M_T} \otimes \delta\sigma_{\mathcal{E}}^2\,\mathbb{I}_{M_R}\Big). \tag{30}$$

Formulation générale du détecteur soft-PIC

Dans cette approche, afin de détecter le symbole correspondant à une antenne donnée, on utilise les informations souples fournies par un décodeur SISO pour annuler (en réalité réduire) l'interférence causée par les signaux des autres antennes. Ensuite, un filtre MMSE est appliqué afin d'annuler les interférences résiduelles. Notons que cette approche diffère d'un filtrage MMSE classique dans la mesure où le filtre MMSE est évalué en moyennant à la fois sur la distribution du bruit et sur celle des symboles d'information.

Soit $\boldsymbol{s}_k = [s_k^1, ..., s_k^{M_T}]^T$ le vecteur des symboles transmis à l'instant k. Intéressons nous à la détection du i-ième symbole s_k^i. Pour ce faire, on commence par évaluer la moyenne et la

variance des symboles s_k^j pour tout $j \neq i$ à partir du décodeur :

$$\hat{s}_k^j = \mathbb{E}\big[s_k^j\big] = \sum_{j=1}^{2^B} s_k^j \, P[s_k^j]$$

$$\sigma^2_{s_k^j} = \mathbb{E}\big[\, |s_k^j|^2 \,\big] = \sum_{j=1}^{2^B} |s_k^j|^2 \, P[s_k^j] \tag{31}$$

où $P[s_k^j]$ est la probabilité que s_k^j soit transmis qui est fournie par le décodeur.
On introduit les définitions suivantes. $\underline{\mathbf{H}}_i$ et $\underline{\widehat{\mathbf{H}}}_i$ sont des matrices $(M_R \times M_T - 1)$ formées à partir de $\mathbf{H}$ et $\widehat{\mathbf{H}}$ en supprimant leurs i-ième colonne, à savoir $\mathbf{h}_i$ et $\widehat{\mathbf{h}}_i$, respectivement. On définit aussi les vecteurs $\underline{\boldsymbol{s}}_k^i$ et $\underline{\hat{\boldsymbol{s}}}_k^i$ de tailles $((M_T - 1) \times 1)$ comme suit :

$$\underline{\boldsymbol{s}}_k^i \triangleq \big[s_k^1, s_k^2, ..., s_k^{i-1}, s_k^{i+1}, ..., s_k^{M_T}\big]^T$$

et

$$\underline{\hat{\boldsymbol{s}}}_k^i \triangleq \big[\hat{s}_k^1, \hat{s}_k^2, ..., \hat{s}_k^{i-1}, \hat{s}_k^{i+1}, ..., \hat{s}_k^{M_T}\big]^T.$$

A présent, l'annulation d'interférence pour détecter s_k^i s'écrit

$$\begin{aligned} \underline{\boldsymbol{y}}_k^i &= \boldsymbol{y}_k - \underline{\widehat{\mathbf{H}}}_i \, \underline{\hat{\boldsymbol{s}}}_k^i \\ &= \mathbf{h}_i s_k^i + \underline{\mathbf{H}}_i \, \underline{\boldsymbol{s}}_k^i - \underline{\widehat{\mathbf{H}}}_i \, \underline{\hat{\boldsymbol{s}}}_k^i + \boldsymbol{z}_k, \quad \text{pour} \;\; i = 1, ..., M_T. \end{aligned} \tag{32}$$

Comme en pratique $\hat{s}_k^j$ est différent de s_k^j, il existe une interférence résiduelle dans $\underline{\boldsymbol{y}}_k^i$. Pour réduire cette interférence, un filtre MMSE $\mathbf{w}_k^i$ est appliqué sur $\underline{\boldsymbol{y}}_k^i$:

$$r_k^i = \mathbf{w}_k^i \, \underline{\boldsymbol{y}}_k^i \tag{33}$$

où le vecteur de taille $(1 \times M_R)$ $\mathbf{w}_k^i$ est donné par

$$\mathbf{w}_k^i = \underset{\mathbf{w}_k^i \in \mathbb{C}^{1 \times M_R}}{\arg\min} \;\; \mathbb{E}_{\boldsymbol{s}_k, \boldsymbol{z}_k} \left[\left| s_k^i - \mathbf{w}_k^i \, \underline{\boldsymbol{y}}_k^i \right|^2 \right]. \tag{34}$$

Avant de passer à travers le décodeur, la sortie du filtre MMSE (33) doit être convertit en LRV. Cela sera présenté un peu plus loin dans le cas du détecteur soft-PIC amélioré.

Détection soft-PIC améliorée avec estimation imparfaite du canal

Il est clair à partir des équations (32) et (34) que la connaissance du canal $\mathbf{H}$ est nécessaire pour les parties annulation d'interférence et filtrage MMSE. Une approche sous optimale et *désadaptée* (au canal) consiste à utiliser l'estimé $\widehat{\mathbf{H}}$ au lieu du vrai canal $\mathbf{H}$. Dans ce travail, en utilisant la loi *a posteriori* (30), on apporte deux modifications au détecteur présenté ci-dessus.

La première modification concerne le calcul du filtre $\mathbf{w}_k^i$ dans (34). On propose un filtre modifié $\widetilde{\mathbf{w}}_k^i$ calculé selon le critère suivant

$$\begin{aligned}\widetilde{\mathbf{w}}_k^i &= \underset{\widetilde{\mathbf{w}}_k^i \in \mathbb{C}^{1\times M_R}}{\arg\min}\ \mathbb{E}_{\mathbf{H},\boldsymbol{s}_k,\boldsymbol{z}_k}\left[\left|s_k(i) - \widetilde{\mathbf{w}}_k^i\,\underline{\boldsymbol{y}}_k^i\right|^2\middle|\widehat{\mathbf{H}}\right] \\ &= \underset{\widetilde{\mathbf{w}}\in \mathbb{C}^{1\times M_R}}{\arg\min}\ \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left[\mathbb{E}_{\boldsymbol{s}_k,\boldsymbol{z}_k}\left[\left|s_k^i - \widetilde{\mathbf{w}}_k^i\,\underline{\boldsymbol{y}}_k^i\right|^2\right]\right]. \end{aligned} \tag{35}$$

En utilisant le principe d'orthogonalité, on obtient (voir Appendice D)

$$\boxed{\widetilde{\mathbf{w}}_k^i = \overline{\mathbf{R}}_{s_k^i \underline{\boldsymbol{y}}_k^i}\,\overline{\mathbf{R}}_{\underline{\boldsymbol{y}}_k^i}^{-1}} \tag{36}$$

où

$$\overline{\mathbf{R}}_{s_k^i \underline{\boldsymbol{y}}_k^i} = \delta\,\sigma^2_{s_k^i}\,\widehat{\mathbf{h}}_i^\dagger + (\delta-1)\,\mathbf{m}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^\dagger \tag{37}$$

et

$$\begin{aligned}\overline{\mathbf{R}}_{\underline{\boldsymbol{y}}_k^i} &= \delta^2\sigma^2_{s_k^i}\,\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^\dagger + \delta^2\,\underline{\widehat{\mathbf{H}}}_i\,\boldsymbol{\Lambda}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^\dagger + (\delta^2-\delta)\,\widehat{\mathbf{h}}_i\,\mathbf{m}_{k,i}\,\underline{\mathbf{H}}_i^\dagger + (\delta^2-\delta)\,\underline{\widehat{\mathbf{H}}}_i\,\mathbf{m}_{k,i}^\dagger\,\widehat{\mathbf{h}}_i^\dagger \\ &\quad + (1-2\delta)\,\underline{\widehat{\mathbf{H}}}_i\,\widetilde{\boldsymbol{\Lambda}}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^\dagger + \big(\sigma_z^2 + (1-\delta)\,\sigma^2_{s_k^i} + (1-\delta)\,\mathrm{tr}(\boldsymbol{\Lambda}_{k,i})\big)\mathbb{I}_{M_R}. \end{aligned} \tag{38}$$

La deuxième modification concerne la partie annulation d'interférence. En effet, au lieu d'utiliser l'équation (32), on propose d'appliquer le filtre $\widetilde{\mathbf{w}}_k^i$ a un signal modifié $\widetilde{\underline{\boldsymbol{y}}}_k(i)$ qui est donné par

$$\widetilde{\underline{\boldsymbol{y}}}_k(i) = \widetilde{\mathbf{h}}_i s_k^i + \underline{\widetilde{\mathbf{H}}}_i\,\underline{\boldsymbol{s}}_k^i - \underline{\widehat{\mathbf{H}}}_i\,\underline{\hat{\boldsymbol{s}}}_k^i + \boldsymbol{z}_k, \tag{39}$$

où

$$\underline{\widetilde{\mathbf{H}}}_i = \mathbb{E}_{\underline{\mathbf{H}}_i|\underline{\widehat{\mathbf{H}}}_i}[\underline{\mathbf{H}}_i] = \delta\,\underline{\widehat{\mathbf{H}}}_i \quad \text{et} \quad \widetilde{\mathbf{h}}_i = \mathbb{E}_{\mathbf{h}_i|\widehat{\mathbf{h}}_i}[\mathbf{h}_i] = \delta\,\widehat{\mathbf{h}}_i.$$

A présent, on applique le filtre $\widetilde{\mathbf{w}}_k^i$ au signal $\widetilde{\underline{\boldsymbol{y}}}_k(i)$, ce qui nous conduit à

$$\widetilde{r}_k^i = \widetilde{\mathbf{w}}_k^i\,\widetilde{\underline{\boldsymbol{y}}}_k(i) = \underbrace{\delta\,\widetilde{\mathbf{w}}_k^i\,\widehat{\mathbf{h}}_i}_{\mu_{k,i}}\,s_k^i + \underbrace{\delta\,\widetilde{\mathbf{w}}_k^i\,\underline{\widehat{\mathbf{H}}}_i\,\underline{\boldsymbol{s}}_k^i - \widetilde{\mathbf{w}}_k^i\,\underline{\widehat{\mathbf{H}}}_i\,\underline{\hat{\boldsymbol{s}}}_k^i + \widetilde{\mathbf{w}}_k^i\,\boldsymbol{z}_k}_{\eta_{k,i}} \tag{40}$$

où $\eta_{k,i}$ est l'interférence plus le bruit qui affecte la sortie du filtre MMSE $\widetilde{r}_k^i$.
L'équation (40) peut être considérée comme la sortie d'un canal AWGN ayant pour entrée le symbole s_k^i, c'est à dire

$$\widetilde{r}_k^i = \mu_{k,i}\,s_k^i + \eta_{k,i} \tag{41}$$

où $\mu_{k,i}$ et $\eta_{k,i}$ sont calculés à chaque instant k à partir des statistiques que le décodeur SISO fournit sur les symboles. La dérivation exacte de la variance $\sigma^2_{\eta_{k,i}}$ de $\eta_{k,i}$, est présentée dans l'Appendice D.

Calcul des LRV : En utilisant l'équation (41), on peut calculer les LRV correspondant aux bits qui composent le symbole d'intérêt s_k^i. Ces LRV seront utilisés par le décodeur SISO dans un traitement itératif de détection et de décodage conjoint. Par exemple pour le m-ième bit $d_k^{i,m}$, on a

$$L(d_k^{i,m}) = \log \frac{P_{\text{dem}}\big(\, d_k^{i,m} = 1 | \, \widetilde{r}_k^i, \mu_{k,i} \,\big)}{P_{\text{dem}}\big(\, d_k^{i,m} = 0 | \, \widetilde{r}_k^i, \mu_{k,i} \,\big)}$$

$$= \log \frac{\displaystyle\sum_{s_k^i \in \underline{\mathcal{S}}_1^m} \exp\left\{ -\frac{|\widetilde{r}_k^i - \mu_{k,i}\, s_k^i|^2}{\sigma^2_{\eta_{k,i}}} \right\} \prod_{\substack{n=1 \\ n \neq m}}^{B} P_{\text{dec}}^1(d_k^{i,n})}{\displaystyle\sum_{s_k^i \in \underline{\mathcal{S}}_0^m} \exp\left\{ -\frac{|\widetilde{r}_k^i - \mu_{k,i}\, s_k^i|^2}{\sigma^2_{\eta_{k,i}}} \right\} \prod_{\substack{n=1 \\ n \neq m}}^{B} P_{\text{dec}}^0(d_k^{i,n})}. \tag{42}$$

Notons finalement qu'à l'inverse d'un détecteur MAP pour lequel les ensembles $\mathcal{S}_1^m$ et $\mathcal{S}_0^m$ dans (23) ont chacun $2^{M_T B - 1}$ éléments, la cardinalité des ensembles $\underline{\mathcal{S}}_1^m$ et $\underline{\mathcal{S}}_0^m$ est à présent égale à 2^{B-1}, ce qui réduit énormément la complexité du détecteur soft-PIC par rapport à un détecteur MAP.

Conclusion

Dans ce chapitre, nous avons proposé un détecteur itératif amélioré et à complexité réduite qui est basé sur l'association d'une annulation parallèle d'interférences à entrées souples (soft-PIC) et d'un filtrage MMSE. En utilisant une fois de plus les statistiques de l'erreur d'estimation de canal, nous avons apporté des modifications au détecteur soft-PIC afin pour prendre en compte l'estimation imparfaite du canal dans la calcul du filtre MMSE et dans la partie annulation d'interférences. Nos résultas ont montré que le détecteur proposé présente de meilleures performances en terme de taux d'erreur binaire qu'un détecteur soft-PIC qui remplace le vrai canal par son estimé. Dans le cas des codes espace-temps, nous avons observé que récepteur est plus sensible au erreurs d'estimation et que le détecteur proposé offre des gains plus importants.

Abstract

The aim of this thesis is to study the problem of iterative data detection in a realistic wireless communication system, where the receiver disposes only of an *imperfect* (and possibly poor) estimate of the unknown channel parameters. The application scenarios on which we focus are single- and multi-antenna OFDM systems working over ultra wideband (UWB) channels. **First**, we propose an efficient receiver jointly estimating the channel and the transmitted symbols in an iterative manner. This receiver is based on a *wavelet* representation of the unknown channel and exploits the sparsness property of UWB channels in the wavelet domain to reduce the receiver's computational complexity. **Second**, we rely on the statistics characterizing the quality of the channel estimation as a mean to integrate the imperfect channel knowledge into the design of iterative receivers. In this way, we formulate an *improved* maximum likelihood (ML) detection metric taking into account the presence of channel estimation errors. We propose a modified iterative detector based on maximum a posteriori (MAP) which mitigates the effect of channel uncertainty on the detector performance, by an appropriate use of this metric. The results are compared to those obtained by using a classical detector based on a *mismatched* ML metric, which uses the channel estimate as if it was the perfect channel. The influence of the constellation labeling is also experimentally studied. Furthermore, we calculate the achieved throughputs associated to both improved and mismatched ML detectors, in terms of maximal achievable outage rates. Our results may serve to evaluate the trade-off between the required quality of service (in terms of BER and achieved throughputs) and the system parameters (e.g., power allocated to pilot and data symbols, number of pilots per frame, number of decoding iterations, outage probability) in the presence of channel estimation errors. **Finally**, we propose an improved low-complexity iterative detector based on soft parallel interference cancellation and linear minimum mean-square error (MMSE) filtering. This receiver takes into account the presence of channel estimation errors in the formulation of the linear MMSE filter, as well as in the interference cancellation part. The important point is that the performance improvements reported in this thesis are obtained while imposing practically no additional complexity to the receiver.

Index Terms–*UWB channels, OFDM, MIMO, wavelet based channel estimation, channel estimation errors, mismatched detection, improved iterative detection, achievable outage rates.*

List of Figures

List of Tables

List of Notations

Acronyms

ADC	analog-to-digital converter
APP	a posteriori probability
AR	auto-regressive
AWGN	additive white Gaussian noise
BCJR	Bahl-Cocke-Jelinek-Raviv (algorithm)
BER	bit error rate
BG	Bernoulli-Gaussian
BICM	bit interleaved coded modulation
CDMA	code division multiple access
CFR	channel frequency response
CIR	channel impulse response
CP	cyclic prefix
CSIR	channel state information at the receiver
DS	direct-sequence
DFT	discrete Fourier transform
EM	expectation-maximization (algorithm)
EGC	equal gain combining
FCC	federal communication commission
(I) FFT	(inverse) fast Fourier transform
GLD	golden code
i.i.d.	independent and identically distributed
ISI	inter-symbol interference
ISM	industrial, scientific and medical
IR	impulse radio
LD	linear dispersion

LLR	log-likelihood ratio
LS	least squares
MAP	maximum a posteriori
Mbps	mega bits per second
MB-OFDM	multiband-OFDM
MIMO	multiple-input multiple-output
ML	maximum likelihood
MMSE	minimum mean square error
MRC	maximum-ratio combining
MSE	mean-squared error
NRNSC	non-recursive non-systematic convolutional
ODWT	orthogonal discrete wavelet transform
OFDM	orthogonal frequency-division multiplexing
OOK	on-off keying
PAM	pulse amplitude modulation
pdf	probability density function
PIC	parallel interference cancellation
PPM	pulse position modulation
PSD	power spectral density
PSAM	pilot symbol assisted modulation
PSM	pulse shape modulation
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quaternary phase shift keying
RF	radio frequency
SISO	soft-input soft-output
SNR	signal-to-noise ratio
SP	set-partition
STC	space-time code
SVD	singular-value decomposition
TH	time-hopping
TFC	time-frequency code
UNII	unlicensed national information infrastructure
UWB	ultra wideband
V-BLAST	vertical Bell Labs layered space time

WPAN	wireless personal area network
WLAN	wireless local area network
ZF	zero-forcing
ZP	zero-padded
ZMCSCG	zero-mean circularly-symmetric complex Gaussian

Mathematical Notations

$\mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$	complex circularly Gaussian distribution with parameters $\mathbf{m}$ and $\mathbf{\Sigma}$
$\mathrm{diag}(\mathbf{x})$	diagonal matrix having the vector $\mathbf{x}$ on its diagonal
$\mathbb{E}_{\mathbf{x}}[.]$	expectation with respect to the vector $\mathbf{x}$
$\mathbb{I}_N$	$(N \times N)$ identity matrix
$\mathbf{0}_N$	$(N \times N)$ all-zero matrix
$\mathbb{C}^{N \times M}$	$(N \times M)$ matrix of complex numbers
$\otimes$	Kronecker product
$\approx$	is approximated by
$\propto$	is proportional to
$(.)^*$	matrix or vector conjugation
$(.)^T$	matrix or vector transpose
$(.)^\dagger$	matrix or vector Hermitian transpose
$\lvert . \rvert$	absolute value
$\lVert . \rVert$	vector norm
$\lVert . \rVert_F$	Frobenious norm
Card.{.}	set cardinality
det{.}	matrix determinant
exp{.}	exponential
tr(.)	matrix trace
$\mathbb{R}e(.)$	real part
$\mathbb{I}m(.)$	imaginary part

Published and Submitted Papers

Accepted and Under Review Journal Papers

1. S. Sadough and P. Duhamel, "Improved Iterative Detection and Achieved Throughputs of OFDM Systems Under Imperfect Channel Estimation", *To appear at the IEEE Trans. Wireless Commun.*.

2. S. Sadough, M. Ichir, P. Duhamel and E. Jaffrot, "Wavelet Based Semi-blind Channel Estimation For Multiband OFDM", *submitted to IEEE Trans. Veh. Technol.*, July 2007.

3. S. Sadough, M. A. Khalighi and P. Duhamel, "Optimal Iterative MIMO Signal Detection Accounting for Channel Estimation Inaccuracies", *submitted to IEEE Trans. Veh. Technol.*, Jan. 2008.

4. S. Sadough, E. Jaffrot and P. Duhamel, "Multiband-OFDM: A New Approach for High Data Rate Ultra Wideband Communication", *To appear at the Journal of the Iranian Association of Electrical and Electronics Engineers (IAEEE).*

5. P. Piantanida, S. Sadough and P. Duhamel, "On the Outage Capacity of a Practical Decoder Accounting for Channel Estimation Inaccuracies", *Accepted with minor revisions for IEEE Trans. Commun.*, July 2007.

International Conferences

1. S. Sadough and E. Jaffrot, "A Wavelet Packet Based Model for an Ultra-Wide-Band Indoor Propagation Channel", *in Proc. European Conf. Propagation and Systems, (ECPS)*, Brest, France, Mar. 2005.

2. S. Sadough, A. Mahmood, E. Jaffrot and P. Duhamel, "Performance Evaluation of IEEE 802.15.3a Physical Layer Proposal Based on Multiband-OFDM", *in Proc. Int. Symp. Telecommun., (IST)*, Shiraz, Iran, Sep. 2005.

3. S. Sadough, M. Ichir, E. Jaffrot and P. Duhamel, "Multiband OFDM UWB Channel Estimation Via a Wavelet Based EM-MAP Algorithm", *in Proc. IEEE Signal Process. Advances Wireless Commun., (SPAWC)*, Cannes, France, July 2006.

4. S. Sadough, E. Jaffrot and P. Duhamel, "Wavelet Domain Channel Estimation for Multiband OFDM UWB Communications", *in Proc. European Signal Process. Conf., (EUSIPCO)*, Florence, Italy, Sep. 2006.

5. S. Sadough, P. Piantanida and P. Duhamel, "Achievable Outage Rates with Improved Decoding of Multiband OFDM Under Channel Estimation Errors", *in Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, California, Oct.-Nov. 2006.

6. S. Sadough and P. Duhamel, "On Optimal Turbo Decoding of Wideband MIMO-OFDM Systems Under Imperfect Channel Estimation", *Presented at the COST2100 meeting*, Lisbon, Portugal, Feb. 2007.

7. S. Sadough, M. Ichir, E. Jaffrot and P. Duhamel, "Wavelet Based Semi-blind Channel Estimation For Multiband OFDM", *in Proc. European Wireless*, Paris, France, April 2007.

8. S. Sadough, P. Piantanida and P. Duhamel, "MIMO-OFDM Optimal Decoding and Achievable Information Rates under Imperfect Channel Estimation", *in Proc. IEEE Signal Process. Advances Wireless Commun., (SPAWC)*, Helsinki, Finland, June 2007.

9. P. Piantanida, S. Sadough and P. Duhamel, "On the Outage Capacity of a Practical Decoder Using Channel Estimation Accuracy", *in Proc. IEEE Int. Symp. Inform. Theory, (ISIT)*, Cannes, France, June 2007.

10. S. Sadough and M. A. Khalighi, "Optimal Turbo-BLAST Detection of MIMO-OFDM Systems with Imperfect Channel Estimation", *in Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun., (PIMRC)*, Athens, Greece, Sep. 2007.

Chapter 1

General Introduction

Novel wireless communication and multimedia services are being introduced almost daily and the demand for higher data rates and higher quality connectivity continue to grow. Cellular telephony and wireless networking are currently the most obvious signs of the success of wireless communications. This spectacular progress is to a great extent due to continuous progress in electronic and micro-electronic technology. Such advances have also been fostered by major theoretical developments. The synergy between components and signal processing techniques is considered to be the main cornerstone of modern communication systems.

There are actually two theoretical discoveries that have had a considerable impact on communication systems. The first one occurred in 1948 where Claude E. Shannon established the fundamental limits on the transmission rates in digital communication systems and motivated the search for coding techniques to approach the capacity limit [27]. In the second landmark development, Claude Berrou *et al.* invented *turbo* error control coding by which the gap between the Shannon's capacity limit and practically feasible channel utilization is almost closed [28].

The key challenge faced by the emerging fourth-generation (4G) wireless systems is to provide broadband access with high data rates at guaranteed quality of service (QoS) for each user, even with very hostile channel environments. For reaching this goal, the use of multiple antennas has been shown to constitute an efficient mean if *perfect* knowledge of the instantaneous channel fading is available at both ends of the wireless link. With the advent of multiple antenna (MIMO) techniques, several original schemes have been devised over the past decade that benefit particularly well from the added spatial dimension provided by multiple antennas: antenna subset selection and space-time coding [29] increase the reliability of a wireless link, while spatial multiplexing [30] and its corresponding demultiplexing and detection algorithms [31] achieve high spectral efficiencies. As higher bit rates involve wideband communications, wireless channels become usually frequency selective. Multicarrier modulation realized by orthogonal

frequency-division multiplexing (OFDM) is well suited for such broadband applications [32]. Recently, MIMO wireless transmission in combination with OFDM (MIMO-OFDM) has been proposed as an attractive solution for the future generation of wireless communication systems [33].

It is well known that reliable *coherent* data detection is not possible unless an accurate channel estimate is available at the receiver. Although perfect channel knowledge rarely happens in practical wireless systems, a large part of research activity on MIMO and OFDM systems has been carried out under the assumption of perfect channel estimate available at both the transmitter and receiver sides. In order to obtain the channel state information at the receiver (CSIR), a commonly-used approach consists in sending some known training symbols (also called pilots) from the transmitter, based on which the receiver estimates the channel before proceeding to the detection of data symbols. Obviously, due to the finite number of pilot symbols and noise, in practice, the receiver can only obtain an *imperfect* (and possibly very poor) estimate of the channel. In "closed-loop" systems, this imperfect channel estimate is usually sent to the transmitter via a limited bandwidth feedback channel which degrades further the quality of the channel estimate. In this situation, one may resort to differential modulations [34–36] that do not require channel estimation and rely on fully incoherent detection procedures. However, performance loss of about 3 dB is paid by these incoherent systems.

In the described scenario, when we compare the well established results concerning the *potential* capacities and the optimal decoding performance of wireless systems with their *achieved* throughputs and decoding performance in the presence of channel estimation errors, we notice large gaps. This effect arises naturally when, due to imperfect channel estimation, the receiver performs signal detection based on maximum likelihood (ML) by using a wrong channel law, or when the receiver is intentionally designed to perform a suboptimal decoding rule so as to simplify its implementation. As incoherent schemes degrade the detection performance and perfect channel estimation is an utopia for most wireless links, recent efforts tried to exploit *partial* CSIR informations for signal detection. In this regard, the impact of imperfect channel knowledge on the receiver performance and the design of improved transceiver structures, that take into account the effect of channel estimation errors, are of considerable relevance to be investigated, from both a theoretical and a practical point of view.

For a MIMO system using pilot-based channel estimation, Garg *et al.* showed in [37] that for compensating the performance degradation due to imperfect channel estimation, the number of receive antennas should be increased. Obviously, this may not be always possible in practice. In [38], Taricco and Biglieri investigated the effect that imperfect channel estimation has on space-time decoding and showed that the classical ML detector derived for the case

of perfect CSIR, becomes largely suboptimal in the presence of channel estimation errors. As an alternative, they adopted an *improved* ML detection metric that mitigates the impact of imperfect CSIR. A similar investigation was carried out in [39] for trellis-coded modulations in scalar channels. From an information-theoretic point of view, Medard derived in [40] an inner and outer bound of the capacity for AWGN channels by considering MMSE channel estimation at the receiver and no channel information at the transmitter. In [41], Yoo and Goldsmith extended the results of Medard to MIMO fading channels by assuming perfect a feedback between the receiver and transmitter. Similar investigations was carried out in [42] by Hassibi and Hochwald for block-fading MIMO channels estimated by training sequences.

In this thesis, we are particularly interested to iterative detection techniques. In fact, since 1993, the concept of turbo decoding introduced by Berrou has been successfully extended to turbo detection [43], turbo equalization [19] and turbo channel estimation [6] for both single- and multi-antenna OFDM systems. Actually, the optimal detection strategy when channel coding is used consists in jointly detecting the data symbols and decoding the encoded data in the maximum *a posteriori* (MAP) sense, however, this scheme suffers from a prohibitive computational complexity. By adopting turbo detection, the receiver is splitted into a "soft" detector (also called demapper) and a soft-input soft-output (SISO) channel decoder. This scheme approximates joint detection by exchanging estimates of the likelihood of each information bit in the message between the processing elements through several iterations. However, in order to provide reliable informations for the SISO decoder, the soft detector requires a *perfect* knowledge (or a very accurate estimate of it) of the channel parameters. Obtaining such an accurate estimate in wireless channels through the use of pilots, would require inserting too many training symbols per frame, which can result in a considerable reduction of the system throughput due to the pilot overhead. Hence, it is of great interest to develop new reception schemes being able to provide the desired performance without excessive increase in the number of training symbols. This requirement will constitutes our roadmap throughout this thesis.

The objective of this thesis is to propose improved iterative detection schemes for both single- and multi-antenna OFDM systems in the presence of channel estimation errors. More precisely, we study the problem of signal detection in a practical wireless communication system, where the receiver has only access to a noisy estimate of the channel and no information about the channel is available at the transmitter. Particular attention is devoted throughout this thesis to the design of low-complexity receivers taking into account the presence of channel estimation errors. The application scenario on which we focus is multiband OFDM (MB-OFDM), proposed for IEEE802.15 wireless personal area networks (WPANs) based on ultra wideband (UWB) transmission. In fact, MB-OFDM is no more than a conventional OFDM

system combined with bit interleaved coded modulation (BICM) [44] for error prevention and a frequency hopping technique for improved diversity and multiple access [4].

As a first solution to mitigate the impact of channel estimation errors on the decoding performance, we consider a semi-blind method based on the Expectation-Maximization (EM) algorithm. In this scheme, a coarse estimate of the channel is first obtained based on few pilots and then, improved channel estimates are obtained by integrating the estimation of the channel into the iterative process of data detection. Soft information from the iterative decoder is used to improve channel estimates after each iteration. Though it has been shown that such iterative joint channel estimation and data detection schemes outperform receivers using pilot-only based channel estimation, they have higher complexities that may be of a critical concern for their practical implementation [5,9]. This complexity is significantly reduced in our proposed scheme by expressing the unknown channel in terms of its discrete wavelet series, which has been shown to provide a *parsimonious* representation [45].

As a second solution to deal with channel uncertainty, we consider a pilot-only based channel estimation and we rely on the statistics characterizing the quality of the channel estimation process to reduce the number of required pilot symbols. Actually, an interesting feature of pilot assisted channel estimation techniques is the availability of the channel estimation error distribution. In our work, we consider the pdf of the perfect channel conditioned on its estimate as a measure of the channel estimation accuracy. This pdf provides us with a statistical framework which we exploit for the design of *improved* iterative detectors under imperfect channel estimation. In addition to studying the error rate performance, the maximal information rates that can be achieved by practical detectors under imperfect channel estimation are also studied and compared with the rates provided by the best possible decoder in the presence of channel estimation errors. Actually, most of the research activity concerning imperfect CSIR is focused on performance analysis in terms of bit error rate (BER). However, our results may serve for selecting the parameters of a wireless communication system (e.g., training sequence length, training power, transmission power, outage probability, etc.) where a prescribed QoS, in terms of achievable rate and BER, must be guaranteed even under imperfect channel estimation.

Although the turbo detection scheme based on MAP leads to an excellent error rate performance, its complexity increases exponentially with the number of bits transmitted per channel-use (i.e., the number of bits per symbol in the signal constellation and the number of transmit antennas). Thus, there is a great interest in the development of reduced-complexity detectors for MIMO systems, especially. Towards this end, by using the same statistical framework described above, we propose a low-complexity turbo detector based on soft interference cancellation able to cope with imperfect channel estimation.

The main questions motivating our research can be summarized as follows.

1. How can the wavelet domain representation of UWB channels be exploited to reduce the complexity and to enhance the performance of a joint iterative channel estimation and data detection scheme ? (treated in Chapter 3)

2. How to design an improved turbo detection scheme that takes into account the presence of channel estimation errors ? (treated in Chapter 4)

3. How does the uncertainty about the channel affect the performance in terms of reliable information rates and bit error rate in single- and multi-antenna OFDM systems ? (treated in Chapter 4)

4. How to reduce the impact of channel uncertainty on the receiver performance for a low-complexity interference cancelling-based iterative receiver ? (treated in Chapter 5)

In the following, we give an overview on how these questions have been addressed in this thesis.

Thesis Overview and Contributions

This dissertation is organized in six chapters. Notice that although our application scenario is a MB-OFDM system working over a UWB channel, the ideas and principles described in this thesis are applicable to conventional OFDM systems as well. Furthermore, our results concerning multiband MIMO-OFDM systems are also transposable to narrowband single-carrier MIMO systems by setting the number of subcarriers to one.

In **Chapter 2**, we present a state of the art of the UWB technology along with different modulation and detection techniques proposed for UWB systems. The first part of this chapter provides an historical overview of UWB, the key benefits and applications of UWB for wireless transmission as well as some UWB regulatory issues. Next, we present the traditional UWB modulation schemes (called single band modulations). Then, we present the multiband UWB modulation scheme followed by a detailed survey of MB-OFDM systems. We finally perform some simulations to show the sub-optimality of the basic receiver proposed for MB-OFDM systems in [4] and this motivates the contribution of the subsequent chapters.

In **Chapter 3**, we first review the major channel estimation techniques advocated in the literature for OFDM systems. Next, we specifically turn our attention to semi-blind methods based on the EM algorithm. After an overview of the EM algorithm principle, we introduce an EM based semi-blind joint channel estimation and data detection scheme, where the channel is represented in the wavelet domain. A *prior* distribution is chosen for the wavelet coefficients of the unknown channel impulse response in order to model the sparseness property of the wavelet

representation. This prior yields, in the maximum *a posteriori* sense, a thresholding rule within the EM algorithm. We particularly focus on reducing the number of estimated parameters by iteratively discarding "insignificant" wavelet coefficients from the estimation process. Finally, the combination of the channel estimation with the decoding operation is discussed, as well as some implementation issues.

In **Chapter 4**, under the assumption of imperfect channel estimation provided by pilot symbols, we propose an improved turbo-MAP detector for both single- and multi-antenna MB-OFDM systems. We start by reviewing the major contributions investigating the impact of imperfect channel estimation on the performance of communication systems employing multiple antennas. Then, by adopting a Bayesian approach involving the statistics of the channel estimation errors, we formulate an *improved* ML detection metric under imperfect channel estimation. First, we use this metric to propose a modified turbo-MAP detector which reduces the impact of channel uncertainty on the decoder performance. For comparison, we also consider a detector based on a sub-optimal *mismatched* ML metric, which uses the channel estimate in the same way as if it was the perfect channel. Second, by using the tools of information theory, we derive the expression of information rates achieved by the improved and mismatched ML detectors, in terms of maximal achievable outage rates. These outage rates are compared to those provided by a theoretical (but not practical) decoder. Numerical results conducted over realistic UWB channels show that the proposed approach provides significant gains in terms of bit error rate and achievable outage rates, compared to the classically-used mismatched approach, with practically no additional increase in the receiver complexity.

Chapter 5 is devoted to a low-complexity turbo receiver based on interference cancellation according to the MMSE criterion. We derive an improved soft parallel interference cancellation (soft-PIC) detector that mitigates the impact of channel uncertainty on the detection performance. A comparison with the optimum turbo-MAP receiver is also provided. The formulation of the improved soft-PIC detector under channel estimation errors is derived for the simple case of V-BLAST scheme. We also provide its generalization to the case of an arbitrary space-time coded MIMO system. Finally, we examine the performance gain that may be achieved with respect to suboptimal soft-PIC detectors for the case of V-BLAST as well as for full-rate full-diversity space-time codes.

Finally in **Chapter 6**, we summarize this thesis and give some concluding remarks as well as suggestions for future research directions.

Chapter 2

Ultra Wideband Modulation and Detection Schemes

2.1 Introduction to Ultra Wideband

2.1.1 Historical Overview

Although, often considered as a recent technology in wireless communications, ultra wideband (UWB) has actually experienced over 40 years of technological developments. In fact, UWB has its origin in the spark-gap transmission design of Marconi and Hertz in the late 1890s [46]. In other words, the first wireless communication system was based on UWB. Owing to technical limitations, narrowband communications were prefered to UWB. In the past 20 years, UWB was used for applications such as radar, sensing, military communication and localization. A substantial change occurred in February 2002, when the Federal Communication Commission (FCC) issued a report [47] allowing the commercial and unlicensed deployment of UWB with a given spectral mask for both indoor and outdoor applications in the USA. This wide frequency allocation initiated a lot of research activities from both industry and academia. In recent years, UWB technology has mostly focused on consumer electronics and wireless communications.

2.1.2 UWB Definition

When UWB technology was proposed for commercial applications, there was no definition for a UWB signal. The first definition for a UWB signal was based on the fractional bandwidth $B_{f,3\text{dB}}$ of the signal. The fractional bandwidth is defined as [48]

$$B_{f,3\text{dB}} = 2\,\frac{f_H - f_L}{f_H + f_L}, \tag{2.1}$$

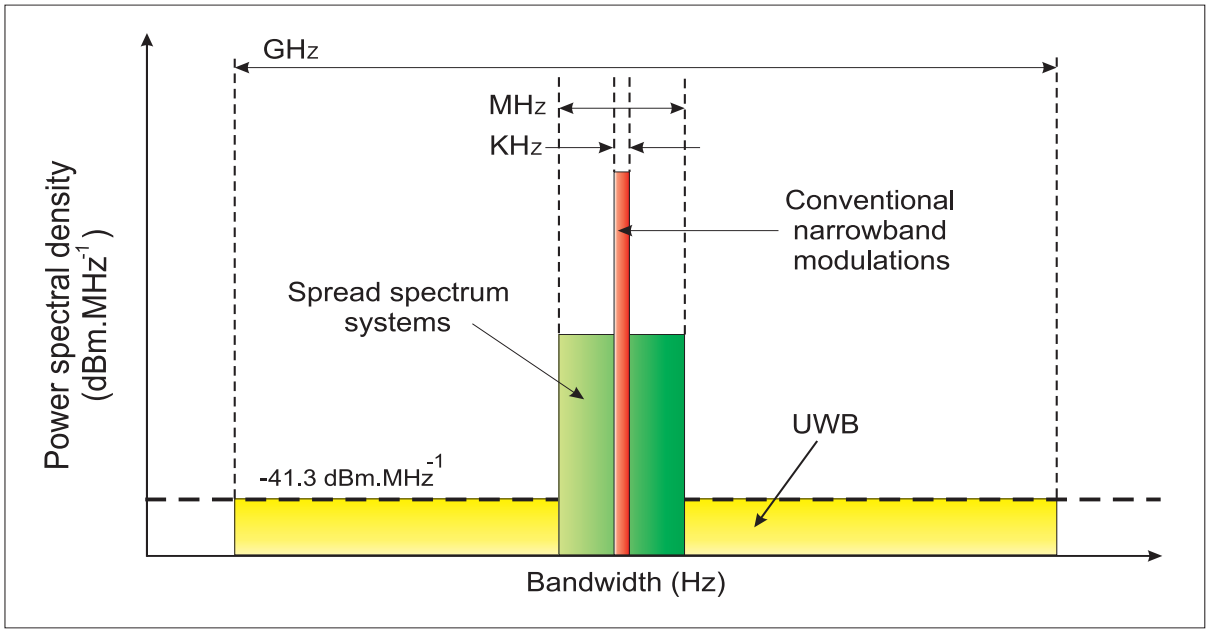

Figure 2.1: Comparison of the spectrum allocation for different wireless radio systems.

where f_L and f_H are respectively the lower and the higher -3 dB point in a spectrum. In this first definition, a signal can be classified as a UWB signal if $B_{f,3\text{dB}}$ is greater than 25 %. In 2002, the FCC approved that any signal having a -10 dB fractional bandwidth larger than 20 %, or a signal bandwidth greater than 500 MHz is considered as UWB. These regulatory rules also specify indoor and outdoor spectral masks, which restrict transmission powers of UWB devices in order to minimize the interference with other narrowband technologies operating in the same frequency band. Figure 2.1 presents a comparative illustration of the UWB spectrum occupation and other existing narrowband systems.

2.1.3 Key Benefits of UWB

UWB has a number of advantages that makes it attractive for consumer communication applications. In particular, UWB systems [49]

- provide high data rates
- have very good time domain resolution allowing for ranging and communication at the same time
- have immunity to multipath and interference
- have potentially low complexity and low equipment cost.

The high rates are perhaps the most compelling aspect from a user's point of view and also from a commercial manufacturer's position. With UWB, transmission rates of over 100 Mbps have been demonstrated, and the potential for higher data rates over short distances is there. The high data rate capability of UWB can be best understood by examining the Shannon's famous capacity equation:

$$C = W \log_2 \left(1 + \frac{S}{N}\right), \tag{2.2}$$

where C is the channel capacity in bits/second, W is the channel bandwidth in Hz, S is the signal power and N is the noise power. This equation tells us that the capacity of a channel grows linearly with the bandwidth W, but only logarithmically with the signal power S. Since the UWB channel has an abundance of bandwidth, it can trade some of the bandwidth against reduced signal power and interference from other sources. Thus, from Shannon's equation we can see that UWB systems have a great potential for high capacity wireless communications.

Thanks to their very large bandwidth, UWB signals have a very high temporal resolution, typically in the order of a nanosecond (ns). Being able to measure the delay of a transmitted signal with a precision of 0.1 to 1 ns, UWB systems provide some information about the position of the transmitter with a precision of 3 to 30 cm. Thus, it is possible to have both precise ranging and high speed data communication in the same wireless terminal providing the possibility for new devices and applications.

The low complexity and low cost of single band UWB systems arises from the ability of UWB systems to directly modulate a pulse onto an antenna. Unlike conventional radio systems, the UWB transmitter produces a very short duration pulse, which is able to propagate without the need for an additional radio frequency (RF) mixing stage. The very wideband nature of the UWB signal means that it spans frequencies commonly used as carrier frequencies. Thus, the signal will propagate well without the need for additional up-conversion and amplification stages.

In single band UWB modulation (described in Section 2.2), the short duration of transmitted pulses provides a fine resolution of reflected pulses at the receiver. In multiband UWB (described in Section 2.4), the spectral flexibility provides robustness against interference by turning-off the interfering frequency bands.

2.1.4 UWB Applications

In recent years, an increasing request appeared for high speed wireless connectivity between a host (e.g., a PC) and associated peripherals such as wireless modem, camcorder, video palayer and so on. This increasing need led to the development of many standards for wireless communication systems over short distances. One can quote Bluetooth, the family of WiFi standards

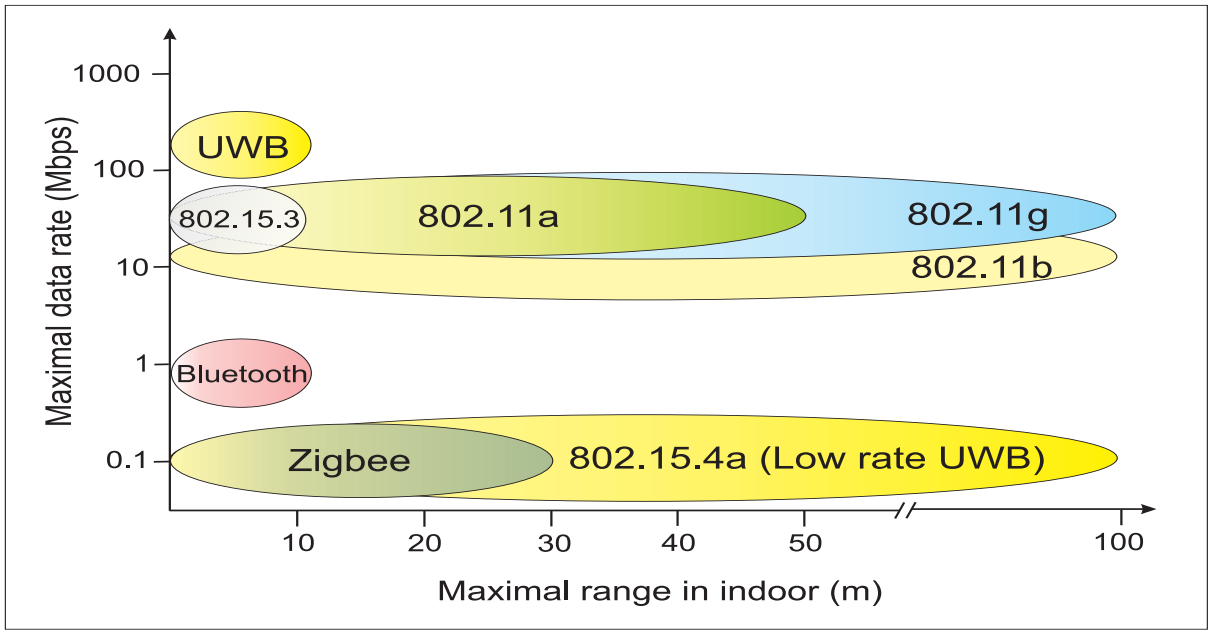

Figure 2.2: Maximal range and data rate of principal WLAN/WPAN standards.

(IEEE802.11), Zigbee (IEEE802.15.4) and the recent standard 802.15.3, which are used for wireless local area networks (WLAN) and wireless personal area networks (WPAN). However, most of these technologies use the ISM and UNII bands with maximum bandwidths about 10 MHz.

An UWB link functions as a "cable replacement" with data rate requirement that ranges from 100 Kbps for a wireless mouse to several hundreds of Mbps for rapid file sharing or download of video files. In summary, UWB is seen as having the potential for applications which to date have not been fulfilled by the aforementioned wireless short range technologies. Figure 2.2 depicts the positioning of the UWB compared to WLAN/WPAN standards in terms of data rate and maximum range. As observed, the potential applications of UWB technology concern two technical areas: very high data rate transmission over short distances (typically 200 Mbps up to 10 m), and low data rate communications with ranges of 100 m with positioning capabilities. It is noticed that in contrast with the WiFi standard, the high data rate mode of UWB belongs to the family of short range WPANs. However, the potential data rate of UWB exceeds the performance of all current WLAN and WPAN standards. In the low data rate mode, the IEEE802.15.4a standard targets UWB systems with centimeter accuracy in ranging as well as with low power and low cost implementation. These features allow a new range of applications, including military applications, medical applications (e.g., monitoring of patients), search-and-rescue applications, logistics (e.g., package tracking), and security applications (e.g.,

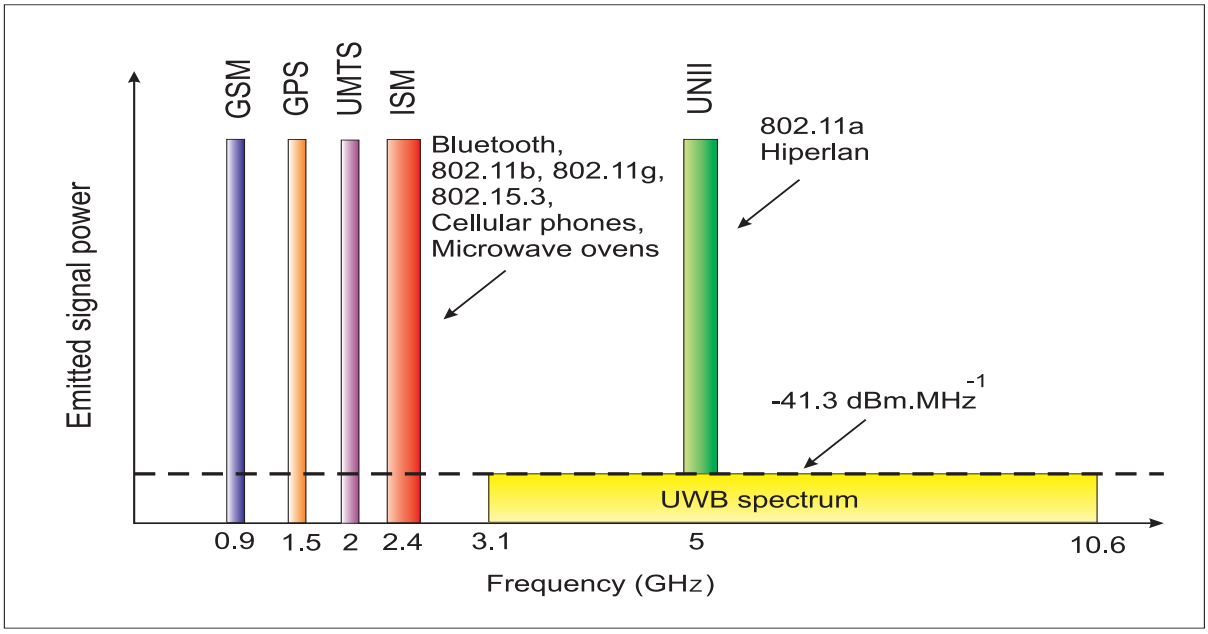

Figure 2.3: Different radio systems in the UHF and SHF band.

localizing authorized persons in high-security areas).

2.1.5 UWB Regulations

Devices utilizing UWB spectrum are subject to more stringent requirements because the UWB spectrum underlays other existing licensed and unlicensed spectrum allocations. In order to optimize spectrum use and to reduce interference to existing systems, regulatory bodies in both Europe and the United States impose very restrictive rulings to UWB devices. Figure 2.3 compares the spectral occupation and emitted power of different radio systems. The essence of these rulings is that the power spectral density (PSD) of the modulated UWB signal must satisfy predefined spectral masks specified by spectrum-regulating agencies.

In the United States, the FCC requires that UWB devices occupy more than 500 MHz of bandwidth in the $3.1 - 10.6$ GHz band, according to the spectrum mask of Fig. 2.4. As observed, the PSD must not exceed -43 dBm per MHz of bandwidth. This limit is low enough not to cause any interference to other services sharing the same bandwidth. Cellular phones, for example, transmit up to $+30$ dBm per MHz, which is equivalent to 10^7 higher PSD than UWB transmitters are permitted.

In Europe, the European Telecommunications Standards Institute (ETSI) works since 2001 to develop a European standard for UWB systems. The studies are carried out in close cooperation with group SE24 of the European Conference of Postal and Telecommunications Adminis-

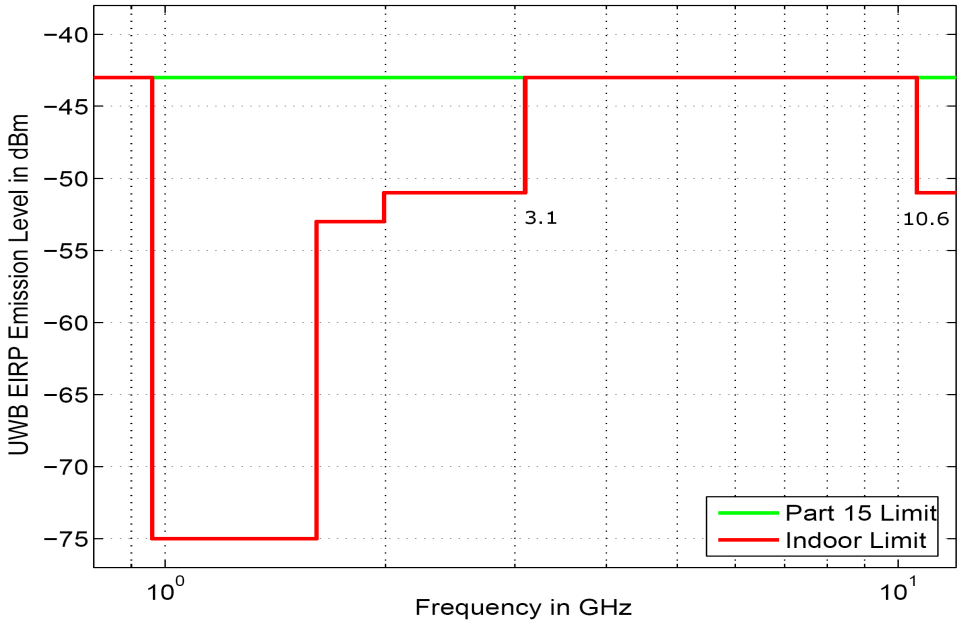

Figure 2.4: FCC spectral mask for indoor UWB transmission [22].

trations (CEPT), which more particularly analyzes the possible impact of UWB on the existing systems [50]. Actually, these European authorities aim at a certain harmony between all the states of the European Union, but the various national regulation authorities remain sovereign in their choice of management of the radio spectrum. Consequently, the regulatory rules for UWB devices have not been finalized in Europe yet. However, it is expected that ETSI/CEPT will follow the FCC's recommendations but will not necessarily adopt the regulations of the FCC [51], due to the more emphasis on the protection of existing services.

2.1.6 Modulation Techniques

Early implementation of UWB communication systems was based on transmission and reception of extremely short duration pulses (typically sub nanosecond), referred to as impulse radio [52]. Each impulse radio has a very wide spectrum, which must adhere to the very low power levels permitted for UWB transmission. These schemes transmit the information data in a carrierless modulation, where no up/down conversion of the transmitted signal is required at the transceiver. A pioneering work in this area is the time hopping pulse position modulation (TH-PPM) introduced in 1993 by Scholtz [53] and better formalized later by Win and Scholtz in [23].

Until February 2002, the term UWB was tied solely to impulse radio modulation. According to the new UWB ruling of FCC from 2002, 7.5 GHz of frequency spectrum (from 3.1 to 10.6 GHz) is allocated for unlicensed applications. Furthermore, *any* communication system that has

a bandwidth larger than 500 MHz is considered as UWB. As a consequence, a variety of well known and more established wireless communication technologies (e.g., OFDM, DS-CDMA) can be used for UWB transmission.

In recent years, UWB system design has experienced a shift from the traditional "single-band" radio that occupies the whole allocated spectrum to a "multiband" design approach [54]. "Multibanding" consists in dividing the available UWB spectrum into several subbands, each one occupying approximately 500 MHz (minimum bandwidth for a UWB system according to the FCC definition). This bandwidth reduction relaxes the requirement on sampling rates of analog-to-digital converters (ADC), consequently enhancing digital processing capability. One example of multiband UWB is multiband orthogonal frequency-division multiplexing (MB-OFDM) [3] proposed by the former IEEE802.15.3a [4] working group on WPAN. In this scheme, high data rate UWB transmission inherits all the strength of OFDM that has already been proven for wireless communications (e.g., DVB, 802.11a, 802.16.a, etc.).

In the sequel, we will begin with the signal model for traditional impulse radio UWB and then move to the multiband UWB systems.

2.2 Single Band UWB Modulations

Single band UWB modulation (also called impulse radio modulation) is based on continuous transmission of very short-time impulse radio which are typically the derivative of Gaussian pulses. Each pulse has an ultra wide spectral occupation in the frequency domain. This type of transmission does not require the use of additional carrier modulation as the pulse will propagate well in the radio channel. The technique is therefore a baseband signal approach.
The most common modulation schemes in this family are depicted in Fig. 2.5. In what follows, we present the signal model for each modulation technique.

2.2.1 Modulations Techniques

2.2.1.1 Pulse Amplitude Modulation

The classical binary pulse amplitude modulation (PAM) is implemented using two antipodal Gaussian pulses as shown in Fig. 2.5. (a). The transmitted binary pulse amplitude modulated signal $s_{\mathrm{tr}}(t)$ can be represented as

$$s_{\mathrm{tr}}(t) = d_k \, w_{\mathrm{tr}}(t), \tag{2.3}$$

where $w_{\mathrm{tr}}(t)$ is the UWB pulse waveform, k represents the transmitted bit ("0" or "1") and

$$d_k = \begin{cases} -1 & \text{if } k = 0 \\ +1 & \text{if } k = 1 \end{cases} \tag{2.4}$$

is used for the antipodal representation of the transmitted bit k. The transmitted pulse is commonly the first derivative of the Gaussian pulse defined as

$$w_{\text{tr}}(t) = -\frac{t}{\sqrt{2\pi}\sigma^3}\, e^{\frac{-t^2}{2\sigma^2}}, \tag{2.5}$$

where σ is related to the pulse length T_p by $\sigma = T_p/2\pi$.

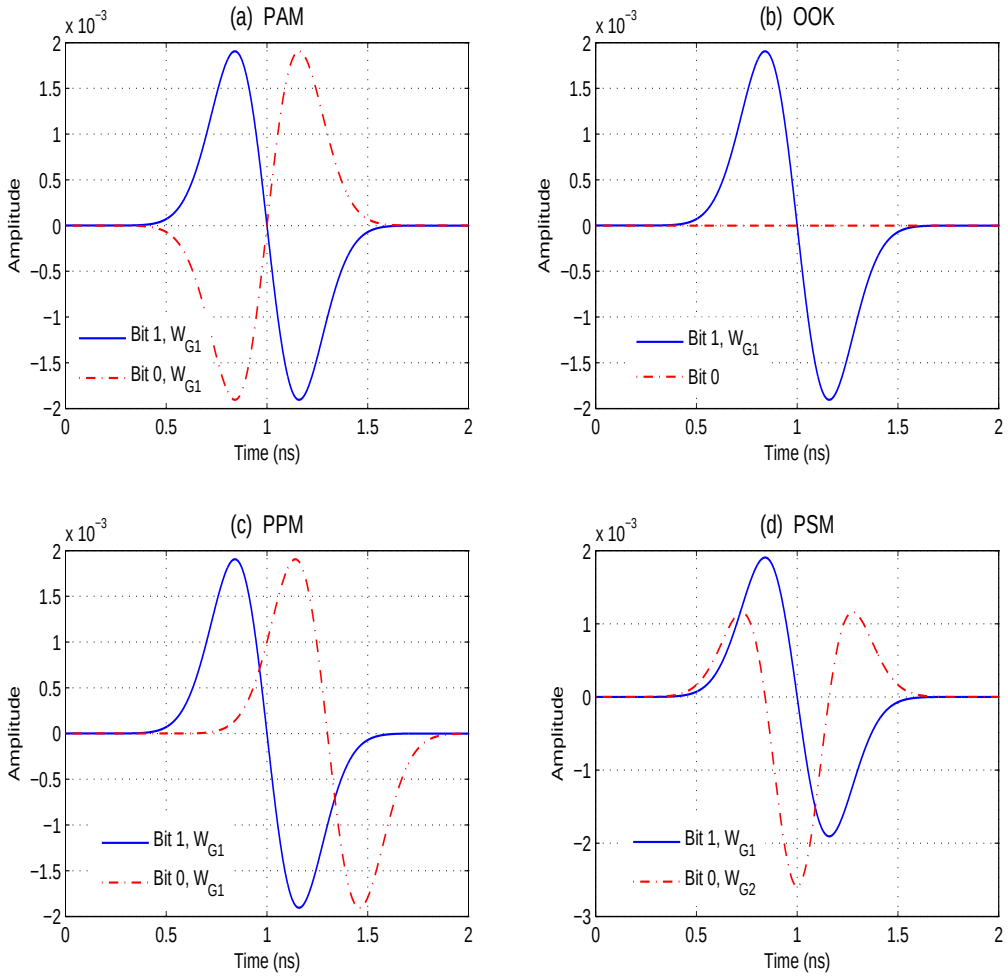

Figure 2.5: Single band (impulse radio) UWB modulation schemes.

2.2.1.2 On-off Keying

The second modulation scheme is the binary on-off keying (OOK) and is depicted in Fig. 2.5 (b). The waveform used for this modulation is defined as in (2.3) with

$$d_k = \begin{cases} 0 & \text{if } k = 0 \\ 1 & \text{if } k = 1. \end{cases} \tag{2.6}$$

The difference between OOK and PAM is that in OOK, no signal is transmitted in the case of bit "0".

2.2.1.3 Pulse Position Modulation

With pulse position modulation (PPM), the information of the data bit to be transmitted is encoded by the position of the transmitted impulse with respect to a nominal position. More precisely, while bit "0" is represented by a pulse originating at the time instant 0, bit "1" is shifted in time by the amount of δ from 0. Let us first assume that a single impulse carry the information corresponding to each symbol. The PPM signal can be represented as

$$s_{\text{tr}}(t) = \sum_{k=-\infty}^{\infty} w_{\text{tr}}(t - kT_s - d_k\delta) \tag{2.7}$$

where $w_{\text{tr}}(t)$ denotes the transmitted impulse radio and δ indicates the time between two states of the PPM modulation. The value of δ may be chosen according to the autocorrelation characteristics of the pulse. For instance, to implement a standard PPM with orthogonal signals, the optimum value of δ (δ_{opt}) which results in zero auto correlation $\rho(\delta_{\text{opt}})$ is such as:

$$\rho(\delta_{\text{opt}}) = \int_{-\infty}^{\infty} w_{\text{tr}}(\tau) w_{\text{tr}}(\delta_{\text{opt}} + \tau) = 0.$$

In a more general case, the symbol is encoded by the integer d_k ($0 \leq d_k \leq M$) where M is the number of states of the modulation. The total duration of the symbol is T_s which is fixed and chosen greater than $M\delta + T_{GI}$ where T_{GI} is a guard interval inserted for inter symbol interference (ISI) mitigation. The binary transmission rate is thus equal to $R = \log_2(M)/T_s$. Figure 2.5 (c) shows a two-state (binary) PPM where a data bit "1" is delayed by a fractional time interval δ whereas a data bit "0" is sent at the nominal time.

2.2.1.4 Pulse Shape Modulation

Pulse shape modulation (PSM) is an alternative to PAM and PPM modulations. As depicted in Fig. 2.5 (d), in PSM the information data is encoded by different pulse shapes. This requires a suitable set of pulses for higher order modulations. Modified Hermite polynomial functions

(MHPF) [55], wavelets [56], and prolate spheroidal wave functions (PSWF) [57] have been proposed in the literature as pulse sets for PSM systems. The orthogonality of signals used in PSM is a desirable property since it permits an easier detection at the receiver. The application of orthogonal signal sets also enables multiple access techniques to be considered. This can be attained by assigning a group of orthogonal pulses to each user, who uses the assigned set for PSM. The transmission will then be mutually orthogonal and different user signals will not interfere with each other.

2.2.2 Enabling Multiple Access in Single Band UWB

Up to now, we assumed that each symbol was transmitted by a single pulse. This continuous pulse transmission can lead to strong lines in the spectrum of the transmitted signal. The regularity of these energy spikes may interfere with other communication systems over short distances. In practical systems, due to the very restrictive UWB power limitations, such a described UWB system shows a high sensitivity to interference from existing systems. On the other hand, the described modulations do not provide multiple access capability.

In order to minimize the potential interference from UWB transmissions and provide multiple access capability, a randomizing technique is applied to the transmitted signal. This makes the spectrum of the UWB signal more noise-like. The two main randomizing techniques used for single band UWB systems are time-hopping (TH) and direct-sequence (DS). The TH technique randomizes the position of the transmitted UWB impulse in time whereas the DS approach is based on continuous transmission of pulses composing a single data bit. The DS-UWB scheme is similar to conventional DS spread-spectrum systems where the chip waveform has a UWB spectrum. A number of other randomizing techniques may be found in [58].

2.2.2.1 Data Modulation with Time-Hopping UWB

As described above, the multiple access and power limit considerations motivate the use of an improved UWB transmission scheme where each data symbol is encoded by the transmission of multiple impulse radios shifted in time. In the TH scheme, the position of each impulse is determined by a pseudo-random (PR) code. In this way, more energy is allocated to a symbol and the range of the transmission is increased. Besides, different users, distinguished by their unique TH code, can transmit at the same time.

A typical TH format for the j-th user is written as follows [1, 2].
For PAM modulation:

$$s_{\mathrm{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\mathrm{tr}}\left(t - kT_s - lT_f - c_l^{(j)} T_c\right) d_k^{(j)}, \tag{2.8}$$

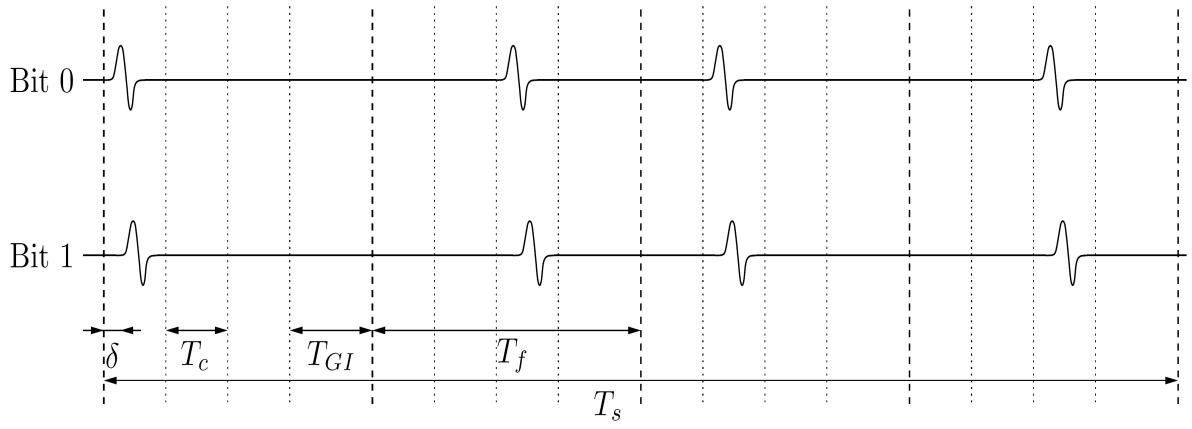

Figure 2.6: Illustration of the TH-PPM binary modulation.

for PPM modulation:

$$s_{\mathrm{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\mathrm{tr}}\Big(t - kT_s - lT_f - c_l^{(j)}T_c - d_k^{(j)}\delta\Big), \tag{2.9}$$

and for PSM modulation:

$$s_{\mathrm{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\mathrm{tr}}^{d_k^{(j)}}\Big(t - kT_s - lT_f - c_l^{(j)}T_c\Big), \tag{2.10}$$

where $d_k^{(j)}$ is the k-th data bit of user j. Here, N_s is the number of impulses transmitted for each information symbol. In this improved scheme, the total symbol transmission time T_s is divided into N_s frames of duration T_f and each frame is itself sub-divided into slots of duration T_c. Each frame contains one impulse in a position determined by the PR TH code sequence $c_l^{(j)}$ (unique for the j-th user) and the symbol to be encoded (see Fig. 2.6). The TH spreading can be combined with PAM, PPM, and PSM. However, OOK cannot take advantage of the TH spreading because of the blank transmission in the case of bit "0".

2.2.2.2 Data Modulation with Direct-sequence UWB

In DS-UWB, the pulse waveform takes the role of the chip in a spread spectrum system [59]. Similar in spirit to spread spectrum techniques, DS-UWB employs sequences of UWB pulses (analogous to "chips"). Each user is distinguished by its specific pseudo random sequence which performs pseudo random inversions of the UWB pulse train. A data bit is then used to modulate these UWB pulses. The resulting signal will then be a continuous transmission of UWB pulses whose number depends on the length of the pulse itself and the bit rate defined by the system.

The DS-UWB scheme is suitable for PAM, OOK and PSM modulations. Since PPM is intrinsically a time-hopping technique, it is not used for DS-UWB transmission. The expression

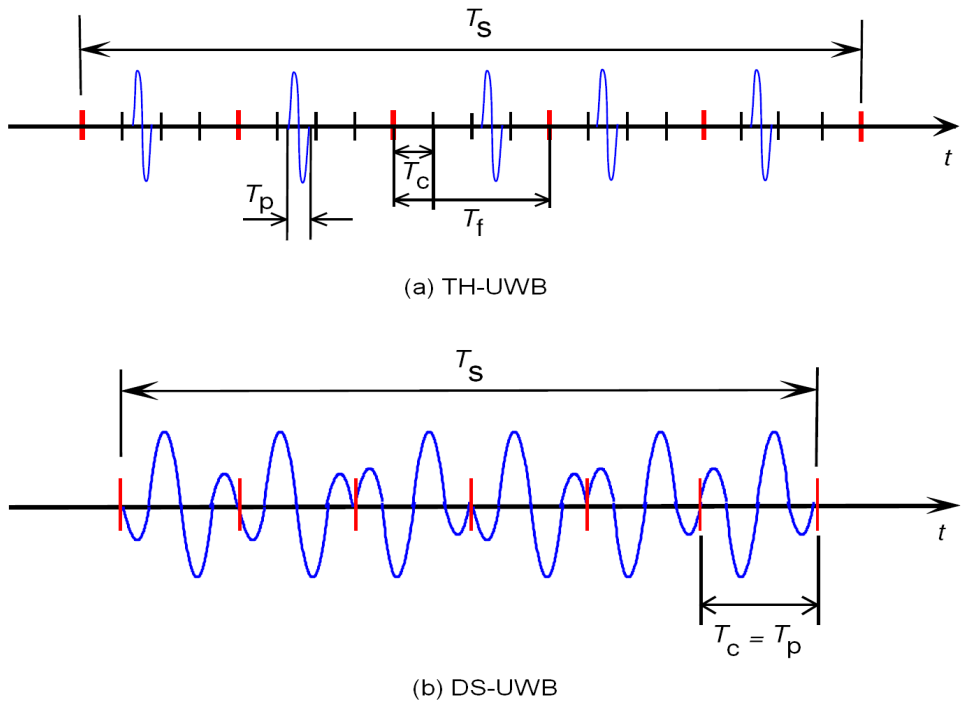

Figure 2.7: Time domain representation of (a) TH-UWB and (b) DS-UWB spreading techniques.

characterizing the DS spreading approach in the case of PAM and OOK modulations for user j is given by [2]

$$s_{\text{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\text{tr}}\Big(t - kT_s - lT_c\Big)\, c_l^{(j)}\, d_k^{(j)} \tag{2.11}$$

where $d_k^{(j)}$ is the k-th data bit, $c_l^{(j)}$ is the l-th chip of the PR code, $w_{\text{tr}}(t)$ is the pulse waveform of duration T_p, T_c is the chip length (equal to T_p), N_s is the number of pulses per data bit, and j stands for the user index. The PR sequence has values in $\{-1, +1\}$ and the bit length is $T_s = N_s T_c$.

For PSM, the signal model for the j-th user is [2]

$$s_{\text{tr}}^{(j)}(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N_s-1} w_{\text{tr}}^{d_k^{(j)}}\Big(t - kT_s - lT_c\, c_l^{(j)}\Big) \tag{2.12}$$

where the bit $d_k^{(j)}$ determines the choice of the UWB pulse waveform to be transmitted.

Figure 2.7 compares the temporal behavior of binary TH-UWB and DS-UWB transmission techniques.

2.2.3 Detection Techniques

In single band UWB systems, two widely used demodulators are correlation receivers and Rake receivers [60]. A brief description of these receivers is presented in the sequel.

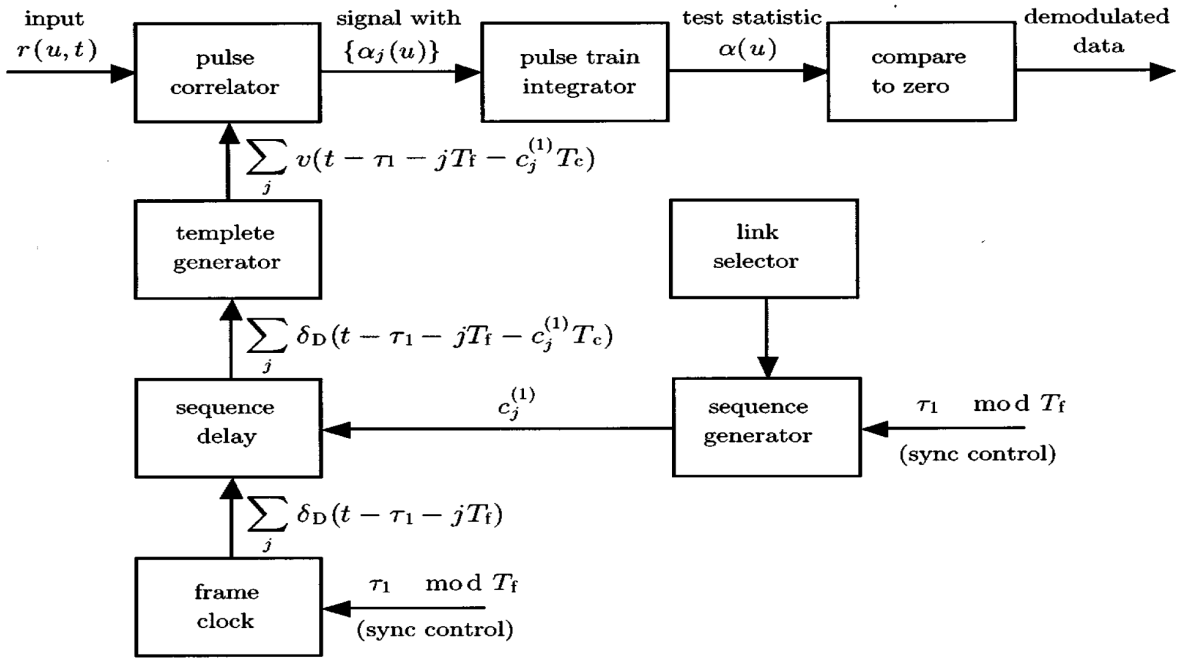

Figure 2.8: Correlation receiver block diagram for the reception of the first user's TH-PPM signal [23].

2.2.3.1 Correlation Receiver

The correlation receiver is the optimum receiver for binary TH-UWB signals in additive white Gaussian noise (AWGN) channels [61]. As the TH format is typically based on PPM and TH-PPM was the first physical layer proposed for UWB communications [23, 53], we present the correlation receiver for the case of a TH-PPM signal.

Let us consider that N_u transmitters are active in the multiple access scheme of the TH-PPM transmitter. The composite received signal $r(t)$ at the receiver is modeled as

$$r(t) = \sum_{j=1}^{N_u} A_j \, s_{\text{rec}}^{(j)}(t - \tau_j) + n(t) \tag{2.13}$$

in which A_j stands for the attenuation over the propagation path of the signal $s_{\text{rec}}^{(j)}(t)$ received from the j-th user (the transmitted signal is given in (2.9)). The random variable τ_j represents the time asynchronism between the clock of the signal received from the transmitter j and the receiver clock, and $n(t)$ represents the additive receiver noise.

The propagation channel modifies the shape of the transmitted impulse $w_{\text{tr}}(t)$ to $w_{\text{rec}}(t)$ and this justify the subscript "rec" in (2.13) for $s_{\text{rec}}^{(j)}(t)$. We consider the detection of the data from the first user, i.e., $d^{(1)}$. For simplicity, we consider a binary transmission.

As depicted in Fig. 2.8, the data detection process is performed by correlating the received

signal with a *template* $v(t)$ defined as[1]

$$v(t) \triangleq w_{\text{rec}}(t) - w_{\text{rec}}(t-\delta)$$

where $w_{\text{rec}}(t)$ and $w_{\text{rec}}(t-\delta)$ represent a symbol with duration T_s encoding "0" and "1", respectively.

According to (2.9), the received signal in a time interval of duration $T_s = N_s T_f$ is given by

$$r(t) = A_1 \sum_{l=0}^{Ns-1} w_{\text{rec}}\big(t - \tau_1 - lT_f - c_l^{(1)} T_c - d^{(1)}\,\delta\big) + n_{\text{tot}}(t) \tag{2.14}$$

where $n_{\text{tot}}(t)$ gathers the multi-user interference and noise. Moreover, it is assumed that the receiver knows the first transmitter's TH sequence $\{c_l^{(1)}\}$ and the delay τ_1.

When the number of users is large, it is classical to approximate the interference-plus noise $n_{\text{tot}}(t)$ as a Gaussian random process [53]. This justifies the optimality of the correlation receiver for TH-PPM signals.

The decision rule at the correlator output for deciding between hypotheses $\mathcal{H}_0$ (bit "0") and $\mathcal{H}_1$ (bit "1") is given by

$$(\text{decide } \; d^{(1)} = \text{“0”}) \Leftrightarrow \sum_{l=0}^{N_s-1} \int_{\tau_1+lT_f}^{\tau_1+(l+1)T_f} r(t)\, v\big(t - \tau_1 - lT_f - c_l^{(1)} T_c\big)\, \mathrm{d}t > 0. \tag{2.15}$$

The sum of integrations in (2.15) corresponds to N_s impulses that carry the information of each data symbol and provides a processing gain which increases linearly with the number of impulses per symbol. Although this constitutes an interesting feature of TH-PPM, we note that the data rate is reduced by a factor of N_s. The other disadvantage of this approach is the severe modification introduced by the UWB channel on the shape of the transmitted signal. Thus, the receiver has to construct a template by using the shape of the received signal. The construction of an optimal template is an important concern for practical PPM based systems. Besides, due to extremely short duration pulses employed, timing mismatches between the correlator template and the received signal can result in serious degradation in the performance of TH-PPM systems. For this reason, accurate synchronization is of great importance for UWB systems employing PPM modulation.

2.2.3.2 Rake Receiver

A typical Rake receiver is depicted in Fig. 2.9. It is composed of a bank of correlators followed by a linear combiner. The signal received at the Rake receiver is correlated with the equally delayed versions of the reference pulse, sampled, multiplied by the tap weights $\{\omega_j\}$ and finally

[1] This is the optimal template under the assumption of Gaussian noise [61].

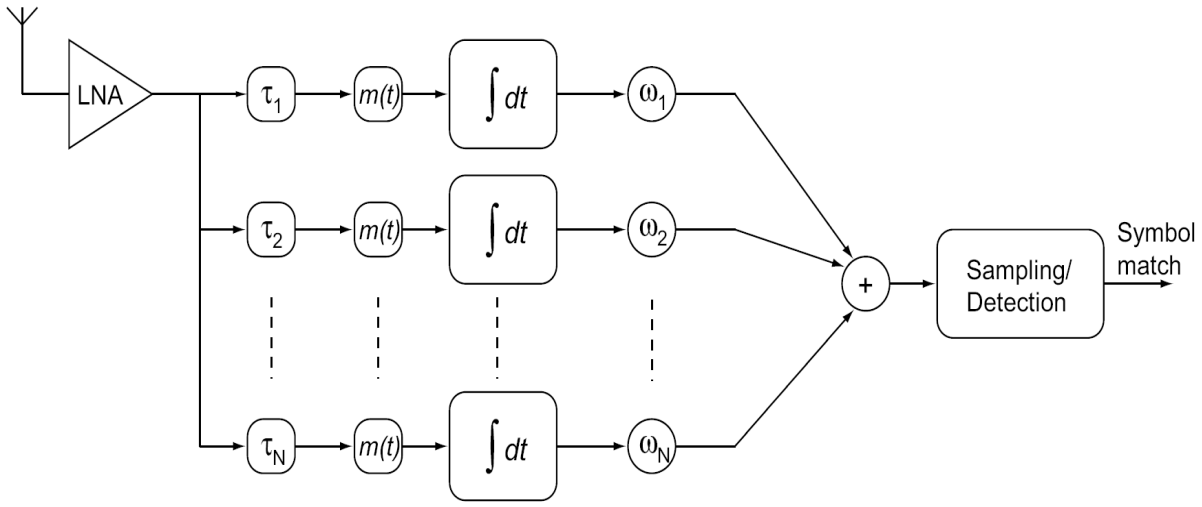

Figure 2.9: Architecture of a Rake receiver with N parallel fingers [24].

linearly combined. The Rake receiver takes advantage of multipath propagation by combining a large number of different and independent replicas of the same transmitted pulse, in order to exploit the multipath diversity of the channel. In general, Rake receivers can support both TH and DS modulated systems, applying soft or hard decision detection. The number of Rake correlators (also called fingers) is selected so as to match the total number of resolvable channel taps. This scheme is referred to as the all-Rake (A-Rake) receiver [51]. However, the major consideration in the design of a UWB Rake receiver is the number of paths to be combined, since the complexity increases with the number of fingers.

2.3 Multiband UWB Modulations

In recent years, there has been a shift in UWB system design away from the traditional single band radio that uses all of the 3.1-10.6 GHz spectrum simultaneously, in favor of a transmission over multiple frequency subbands, which is referred to as *multiband* UWB [3, 25, 54, 62]. In multiband UWB radio, pulses are successively modulated by several analog carriers and transmitted through subbands of approximately 500 MHz bandwidth (see Fig. 2.10). Compared to impulse-based UWB modulations, it is obvious that multiband UWB can make a more efficient use of the spectral resources, minimizes interference to existing narrowband systems by flexible band selection, and facilitates future scalability of the spectrum use. Moreover, a narrower subband bandwidth eases the requirement on ADC sampling rates (compared to a full-band receiver), and consequently, facilitates the digital processing. Nevertheless, the bandwidth of each subband is wide enough to allow different multiple-access and modulation options. This

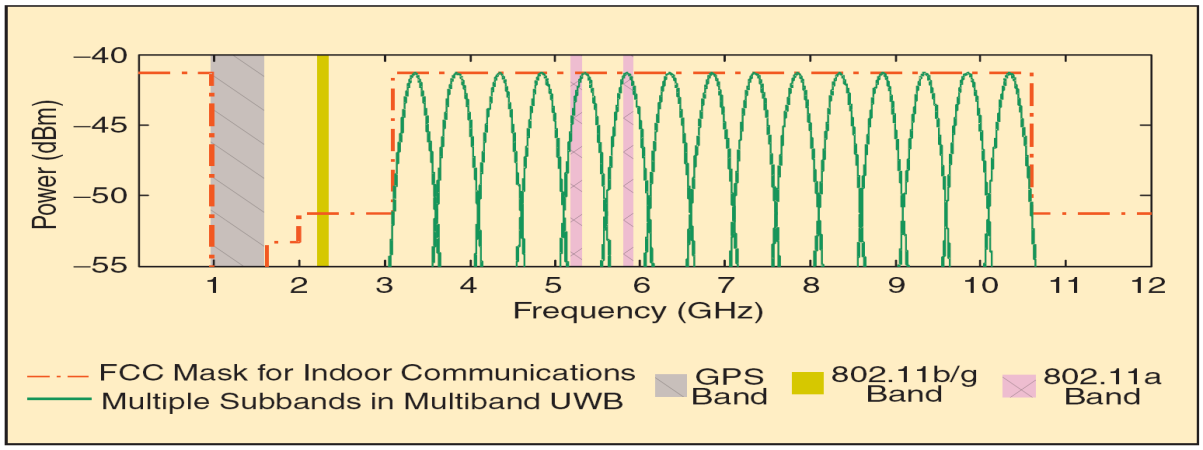

Figure 2.10: Multiple subbands in multiband UWB [25].

scheme allows for tradeoffs between simplified time-domain impulse modulations and frequency-domain modulations/spreading in order to obtain the desired performance in multipath fading and in the presence of interference from other UWB users. Multiband UWB modulation can be classified into multiband impulse radio (MB-IR) and multiband OFDM (MB-OFDM).
In what follows, we start by describing MB-IR systems and then we devote a whole section to MB-OFDM systems.

2.3.1 Multiband Impulse Radio

In this scheme the whole allocated UWB spectrum is divided into smaller non-overlapping subbands of at least 500 MHz bandwidth. The modulation used is one of the single band modulations (PAM, PPM, PSM, etc.) performed over each subband. Each pulse waveform is transmitted with a pulse repetition interval T_{PRI} in order to avoid the ISI. Changing T_{PRI} affects the data rate of the system as well as the robustness to ISI. One of the main advantages of MB-IR UWB is that a lower complexity Rake receiver (i.e., with fewer number of fingers) per subband suffices for energy capture (as compared to a Rake receiver that spans the entire bandwidth). The disadvantage is that one Rake receiver is required per subband, albeit with a small number of fingers.

2.4 Multiband OFDM

Up to now, we presented single band UWB modulations and MB-IR systems. Since this thesis is mainly focused on the multiband OFDM approach, we devote this section to a more detailed description and performance evaluation of MB-OFDM systems at the physical layer (PHY).

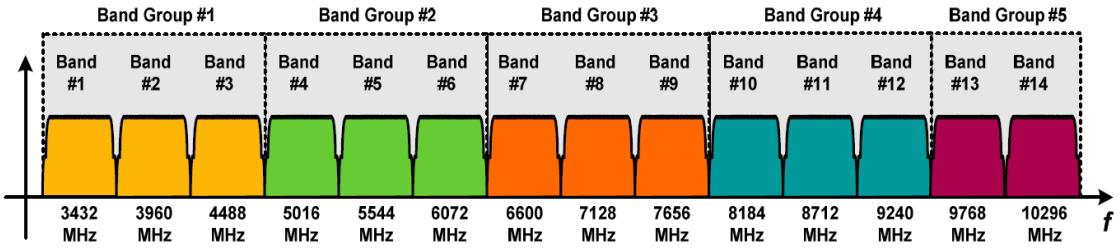

Figure 2.11: Division of the UWB spectrum from 3.1 to 10.6 GHz into band groups containing subbands of 528 MHz in MB-OFDM systems [26].

2.4.1 Introduction

We previously saw that MB-IR UWB systems needed slower time-frequency hopping, i.e., longer contiguous symbol transmission in each subband in order to improve the energy capture. This requirement led naturally to the choice of coded OFDM instead of pure pulse modulation in each subband owing to the former's inherent robustness to multipath. Moreover, for highly dispersive UWB channels, an OFDM based receiver is more efficient at capturing multipath energy than an equivalent single band Rake receiver using the same total bandwidth.[2] OFDM systems possess additional desirable properties, such as high spectral efficiency, inherent resilience to narrowband RF interference and spectral flexibility, which is important because the regulatory rules for UWB devices have not been finalized through the entire world. A brief overview of OFDM is given in the next subsection; for further details, the reader is referred to [32].

In recent years, a group of international companies including Texas Instrument, Alereon, Hewlett-Packard, etc., made an alliance (called MBOA and then WiMedia Alliance) [26] in order to support an OFDM based solution for multiband UWB.

In 2004, Batra *et al.* from Texas Instrument proposed the MB-OFDM scheme to IEEE802.15.3a [3, 4]. The proposed scheme divides the available UWB spectrum into several non-overlapping subbands of 528 MHz bandwidth each. As shown in Fig. 2.11, five band groups are defined within the 3.1-10.6 GHz frequency band. The first four band groups have three subbands each, and the last group has two subbands. Data transmission over the three lowest subbands is called the *mandatory* mode or mode I. This operating mode is reserved for preliminary and low-cost implementation since the degradation due to the RF noise is limited.

Within each subband, information is transmitted using conventional coded OFDM modulation. The main difference between the transmitter architecture of an MB-OFDM system with that of a conventional OFDM system is the presence of a time-frequency code (TFC), which provides

[2]For a complexity comparison between MB-OFDM and DS-UWB using a Rake receiver, the reader is referred to [3].

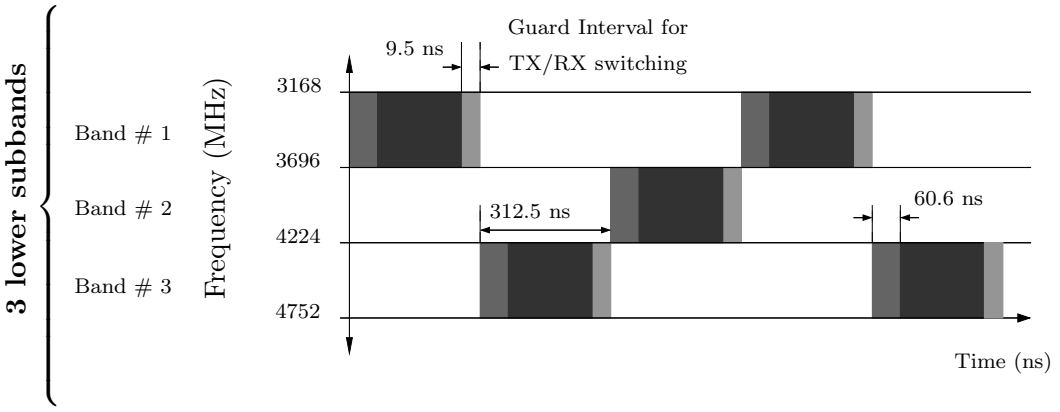

Figure 2.12: Example of time-frequency coding for the multiband OFDM system in mode I, TFC = {1, 3, 2, 1, 3, 2, ...}.

a different carrier frequency at each time-slot, corresponding to one of the center frequencies of different subbands (see Fig. 2.12). The TFC is used not only to provide frequency diversity but also to distinguish multiple users.

2.4.2 MB-OFDM Transmitter Architecture

As depicted in Fig. 2.13, in MB-OFDM, the information is transmitted using coded OFDM modulation over one of the subbands in a particular time-slot. The binary sequence is encoded by a non-recursive non-systematic convolutional (NRNSC) code, before being interleaved. The interleaved bits are gathered in subsequences of B bits $d_k^1, \ldots, d_k^B$ and mapped to complex $\mathrm{M_c}$-QAM ($\mathrm{M_c} = 2^B$) symbols s_k. In the basic proposal of MB-OFDM [4], quaternary phase-shift keying (QPSK) symbols using Gray labeling is employed. We will later extend MB-OFDM to the higher order 16-QAM constellation with Gray or set-partition (SP) labeling.

According to [4], MB-OFDM uses $N_c = 128$ subcarriers per subband, through a frequency selective multipath fading channel with a bandwidth of 528 MHz. This leads to a subcarrier separation of $\Delta f = 4.125$ MHz. At each time slot, the transmitter applies 128 point inverse fast Fourier transform (IFFT) yielding an OFDM symbol of duration $T_{\mathrm{FFT}} = 1/\Delta_f = 242.42$ ns. In order to mitigate the impact of ISI, a cyclic prefix (CP) of length $T_{\mathrm{CP}} = 60.6$ ns is added to the output of the IFFT signal. Besides, an additional guard interval (GI) of duration $T_{\mathrm{GI}} = 9.5$ ns is added to allow the transmitter and receiver to switch from one subband to next. After adding the CP and the GI, the OFDM symbol is passed through a digital-to-analog converter (DAC) resulting to an analog baseband OFDM signal of symbol duration $T_{\mathrm{SYM}} = T_{\mathrm{FFT}} + T_{\mathrm{CP}} + T_{\mathrm{GI}} = 312.5$ ns (see Fig. 2.12). Let s_k^n be the complex symbol to be transmitted over the k-th OFDM subcarrier during the n-th OFDM symbol period. The

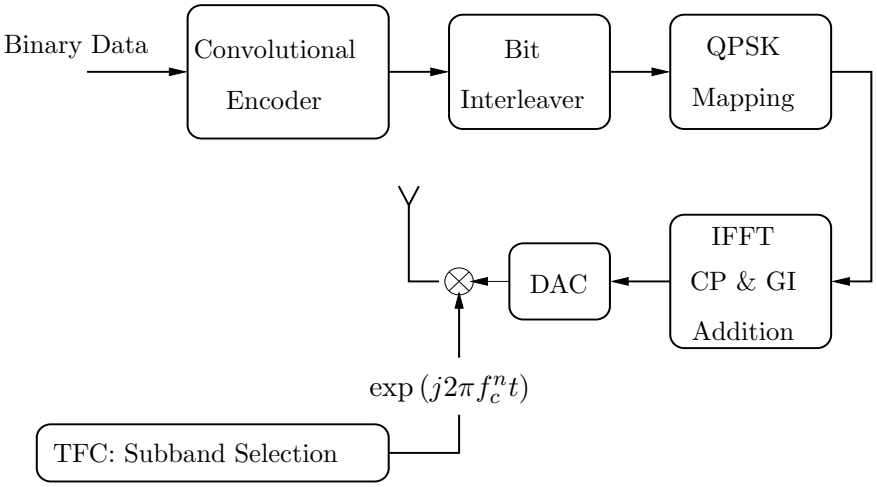

Figure 2.13: Transmitter architecture for the MB-OFDM system.

baseband OFDM signal to be transmitted at the n-th block can be expressed as

$$x_n(t) = \sum_{k=0}^{N_c-1} s_k^n \exp\left\{ j2\pi k\, \Delta f\, (t - T_{\mathrm{CP}}) \right\} \tag{2.16}$$

where $t \in [T_{\mathrm{CP}}, T_{\mathrm{FFT}}+T_{\mathrm{CP}}]$ and $j \triangleq \sqrt{-1}$. In the time interval $[0, T_{\mathrm{CP}}]$, $x_n(t)$ is a copy of the last part of the OFDM symbol, and $x_n(t)$ is zero in the interval $[T_{\mathrm{FFT}} + T_{\mathrm{CP}}, T_{\mathrm{SYM}}]$ corresponding to the GI duration.

The complex baseband signal $x_n(t)$ is filtered, up-converted to an RF signal with a carrier frequency f_c^n, and sent to the transmit antenna.

The transmitted MB-OFDM signal is given by

$$r_{\mathrm{RF}}(t) = \sum_{n=0}^{N_{\mathrm{SYM}}-1} \mathbb{R}e\Big(x_n(t - nT_{\mathrm{SYM}}) \exp\left\{ j2\pi f_c^n t \right\} \Big) \tag{2.17}$$

where N_{SYM} is the total number of OFDM symbols in a transmitted frame (also called packet). The carrier frequency f_c^n specifies the subband over which the n-th OFDM symbol is transmitted, according to the TFC.

In the sequel, we describe each part of the MB-OFDM transmitter in Fig. 2.13.

2.4.2.1 Channel Encoding

In OFDM transmission over multipath channels, symbols sent on different subcarriers may undergo deep fades, which would (with a high probability) lead to erroneous decisions. Thus, uncoded OFDM is in practice unusable on multipath fading channels with deep notches occurring in the frequency spectrum. For this reason, MB-OFDM proposes forward error correction coding with different code rates by using a convolutional code [60] (called mother code). The mother code is a rate $R = 1/3$ NRNSC code of constraint length $K = 7$ defined in octal form

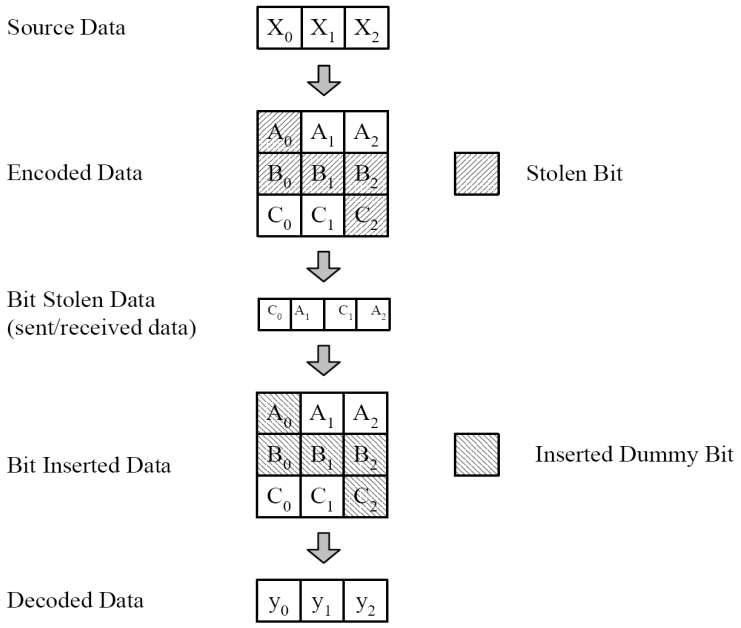

Figure 2.14: An example of bit-stealing and bit-insertion procedure for obtaining $R = 3/4$ from $R = 1/3$ [4].

by the generator polynomials $(133, 145, 175)_8$. Various coding rates ($R = 11/32, 1/2, 5/8, 3/4$) are derived from the rate $R = 1/3$ mother code by employing "puncturing". Puncturing is a procedure for omitting some of the encoded bits in the transmitter (thus reducing the number of transmitted bits and increasing the coding rate) and inserting a "dummy" bit into the convolutional decoder on the receive side in place of the omitted bits. An example of puncturing pattern for deriving the rate $R = 3/4$ code is shown in Fig. 2.14.

2.4.2.2 Bit Interleaving

Several standard proposals such as IEEE802.15.3a that proposed MB-OFDM, employ bit-interleaving combined with convolutional channel coding. This scheme, referred in the literature as bit-interleaved coded modulation (BICM) [44], can provide a high diversity order for transmission over multipath fading channels. In the basic proposal of MB-OFDM, the bit interleaving operation is performed in two stages [4]:

- *Inter-symbol interleaving*, which permutes the bits across 6 consecutive OFDM symbols, enables the PHY to exploit frequency diversity within a band group.
- *Intra-symbol tone interleaving*, which permutes the bits across the data subcarriers within one OFDM symbol, exploits frequency diversity across subcarriers and provides robustness against narrow-band interferers.

For intra-symbol interleaving, the coded bits are first grouped together into blocks of $6N_{\mathrm{CBPS}}$ coded bits (corresponding to six OFDM symbols), where N_{CBPS} is the number of coded bits per OFDM symbol. Each group of coded bits is then permuted using a block interleaver of size $N_{\mathrm{BI}_1} = 6 \times N_{\mathrm{CBPS}}$. Let the sequences $\{U(i)\}$ and $\{S(i)\}$, $(i = 0, ..., 6N_{\mathrm{CBPS}} - 1)$ represent the input and output bits of the symbol interleaver, respectively.
We have [4]

$$S(i) = U\left\{\mathrm{Floor}\left(\frac{i}{N_{\mathrm{CBPS}}}\right) + 6\,\mathrm{Mod}\big(i, N_{\mathrm{CBPS}}\big)\right\}, \tag{2.18}$$

where Floor(.) returns the largest integer value less than or equal to its argument and $\mathrm{Mod}(a, b)$ returns the remainder after division of a by b.
The outputs of the symbol block interleaver are then grouped into blocks of N_{CBPS} bits and permuted using a regular block intra-symbol tone interleaver of size $N_{\mathrm{BI}_2} = 10 \times N_{T\mathrm{int}}$. Let the sequences $\{S(i)\}$ and $\{V(i)\}$, $(i = 0, ..., 6N_{\mathrm{CBPS}} - 1)$ represent the input and output bits of the tone interleaver, respectively. The output of the interleaver is given by the following relation [4]

$$V(i) = S\left\{\mathrm{Floor}\left(\frac{i}{N_{T\mathrm{int}}}\right) + 10\,\mathrm{Mod}\big(i, N_{T\mathrm{int}}\big)\right\}. \tag{2.19}$$

2.4.2.3 Time and Frequency Domain Spreading

In MB-OFDM, two diversity schemes may be be used to obtain further bandwidth expansion, beyond that provided by the forward error correction code. The first one is the *frequency domain spreading* which consists in transmitting twice the same information in a single OFDM symbol. This is performed by introducing conjugate symmetric inputs to the IFFT. Specifically, the data symbols are sent on the first half of the data subcarriers and their conjugate symmetrics are transmitted on the second half of the subcarriers. This introduces a spreading factor of two and results in "intra-subband" frequency diversity.

The second scheme is *time domain spreading* which is achieved by transmitting the same OFDM symbol across two different frequency subbands. This technique results in "inter-subband" diversity and is used to maximize the frequency-diversity and to improve the performance in the presence of other non-coordinated devices.

As listed in Tab. 2.1, MB-OFDM combines different channel code rates with time and/or frequency diversity to provide data rates ranging from 53.3 Mbps to 480 Mbps. For data rates lower than 80 Mbps (low data rate mode), both time and frequency spreading are performed, yielding an overall spreading gain of four. For data rates between 106.7 and 200 Mbps (medium data rate mode) only time domain spreading is used which results in a spreading gain of two. The transmission with data rates higher than 200 Mbps (high data rate mode) exploits neither frequency nor time spreading, and the overall spreading gain is equal to one.

Data Rate (Mbps)	Modulation	Code Rate	Freq. Spread.	Time Spread. Factor
53.3	QPSK	1/3	Yes	2
55	QPSK	11/32	Yes	2
80	QPSK	1/2	Yes	2
106.7	QPSK	1/3	No	2
110	QPSK	11/32	No	2
160	QPSK	1/2	No	2
200	QPSK	5/8	No	2
320	QPSK	1/2	No	1
400	QPSK	5/8	No	1
480	QPSK	3/4	No	1

Table 2.1: Rate-dependent Parameters in multiband OFDM systems.

2.4.2.4 Subcarrier Constellation Mapping

The constellation adopted in [4] is QPSK. The coded and interleaved binary data is divided into groups of two bits and converted into one of the four complex points of the QPSK constellation. The conversion is performed according to the Gray labeling, as illustrated in Fig. 2.15.

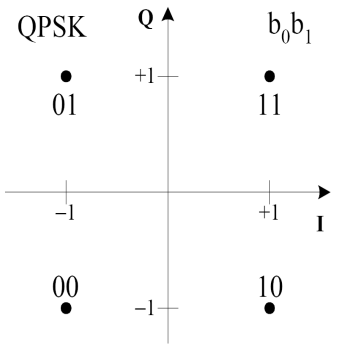

Figure 2.15: QPSK Constellation with Gray Mapping.

2.4.3 MB-OFDM Receiver Architecture

2.4.3.1 System Model

The receiver proposed for MB-OFDM [3] is depicted in Fig. 2.16. As shown, the process of channel estimation and data detection are performed independently. We will later propose an enhanced receiver based on joint channel estimation and iterative data detection.

Let us consider a single-user MB-OFDM transmission with $N_{\text{data}} = 100$ data subcarriers

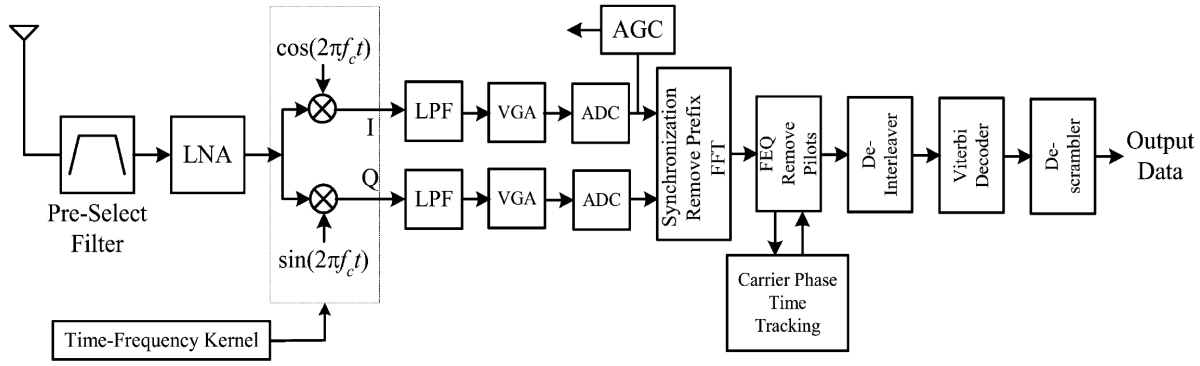

Figure 2.16: The basic receiver architecture proposed for MB-OFDM in [3].

per subband, through a frequency selective multipath fading channel, described in discrete-time baseband equivalent form by the channel impulse response coefficients $\{h_l\}_{l=0}^{L-1}$. Furthermore, we assume that the CP is longer than the maximum delay spread of the channel. After removing the CP and performing FFT at the receiver, the received OFDM symbol over a given subband can be written as

$$\boldsymbol{y} = \mathbf{H}_d\, \boldsymbol{s} + \boldsymbol{z}, \tag{2.20}$$

where $(N_{\text{data}} \times 1)$ vectors $\boldsymbol{y}$ and $\boldsymbol{s}$ denote the received and transmitted symbols, respectively; the noise vector $\boldsymbol{z}$ is assumed to be a zero-mean circularly symmetric complex Gaussian (ZMCSCG) random vector with distribution $\boldsymbol{z} \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2\, \mathbb{I}_{N_{\text{data}}})$; and $\mathbf{H}_d = \text{diag}(\mathbf{H})$ is the $(N_{\text{data}} \times N_{\text{data}})$ diagonal channel matrix with diagonal elements given by the vector $\mathbf{H} = [H_0, \ldots, H_{N_{\text{data}}-1}]^T$, where $H_k = \sum_{l=0}^{L-1} h_l\, e^{-j2\pi kl/N_c}$.

In MB-OFDM, the channel is assumed to be time invariant over the transmission of one frame and changes to new independent values from one frame to the next.

2.4.3.2 Channel Estimation

In order to estimate the channel, a MB-OFDM system sends some OFDM pilot symbols at the beginning of the information frame. Here, we consider the estimation of the channel vector $\mathbf{H}$ with N_P training symbols $\boldsymbol{s}_{P,i}, (i = 1, ..., N_P)$. According to the observation model (2.20), the received signal for a given channel training interval is:

$$\mathbf{Y}_P = \mathbf{H}_d\, \mathbf{S}_P + \mathbf{Z}_P \tag{2.21}$$

where each column of the $(N_{\text{data}} \times N_P)$ matrix $\mathbf{S}_P = [\boldsymbol{s}_{P,1}, ..., \boldsymbol{s}_{P,N_P}]$ contains one OFDM pilot symbol. The entries of the noise matrix $\mathbf{Z}_P$ has the same distribution as those of $\boldsymbol{z}$.

The least-square (LS) estimate of $\mathbf{H}_d$ is obtained by minimizing $\|\mathbf{Y}_P - \mathbf{H}_d\, \mathbf{S}_P\|_F^2$ with respect

to $\mathbf{H}_d$. We have:

$$\widehat{\mathbf{H}}_d^{\mathrm{LS}} = \mathbf{Y}_P \, \mathbf{S}_P^{\dagger} \, (\mathbf{S}_P \mathbf{S}_P^{\dagger})^{-1}. \tag{2.22}$$

2.4.3.3 Frequency Domain Channel Equalization

In order to estimate the transmitted signal vector $\boldsymbol{s}$ from the received signal vector $\boldsymbol{y}$, the effect of the channel must be mitigated. To this end, the MB-OFDM proposal uses a frequency domain channel equalizer, as shown in Fig. 2.16 (FEQ block). It consists of a linear estimator as

$$\widehat{\boldsymbol{s}} = \mathbf{G}^{\dagger} \, \boldsymbol{y}. \tag{2.23}$$

The two design criteria usually considered for the choice of the linear filter $\mathbf{G}$ are:

- *Zero-forcing equalization (ZF):* ZF equalization, uses the inverse of the channel transfer function as the estimation filter. In other words, we have $\mathbf{G}^{\dagger} = \mathbf{H}_d^{-1}$. Since in OFDM systems, under ideal conditions, the channel matrix $\mathbf{H}_d$ is diagonal, the ZF estimate of the transmitted signal is obtained independently on each subcarrier as

$$\widehat{s}_{\mathrm{zf},k} = \frac{1}{H_k} \, y_k \quad k = 0, ..., N_{\mathrm{data}} - 1. \tag{2.24}$$

- *Minimum mean-square error equalization (MMSE):* equalization according to the MMSE criterion, minimizes the mean-squared error $\mathbb{E}\big[\|\boldsymbol{s} - \mathbf{G}^{\dagger} \, \boldsymbol{y}\|_F^2\big]$, between the transmitted signal and the output of the equalizer. Applying the orthogonality principle, it is easy to obtain

$$\mathbf{G}_{\mathrm{mmse}}^{\dagger} = \left(\mathbf{H}_d \mathbf{H}_d^{\dagger} + \sigma_z^2 \mathbb{I}_{N_c}\right)^{-1} \mathbf{H}_d^{\dagger}. \tag{2.25}$$

  Due to the diagonal structure of $\mathbf{H}_d$, equalization can again be done on a subcarrier basis as

$$\widehat{s}_{\mathrm{mmse},k} = \frac{H_k^*}{|H_k|^2 + \sigma_z^2} \, y_k \quad k = 0, ..., N_{\mathrm{data}} - 1. \tag{2.26}$$

The main drawback of the ZF solution is that for small amplitudes of H_k, the equalizer enhances the noise level in such a way that the signal-to-noise ratio (SNR) may go to zero on some subcarriers. The computation of the MMSE equalization matrix requires an estimate of the curent noise level. Notice that when the noise level is significant, the MMSE solution mitigates the noise enhancement problem even when H_k's are close to zero while for high SNR regime, the MMSE equalizer becomes equivalent to the ZF solution.

	CM1	CM2	CM3	CM4
Tx-Rx separation (m)	0-4	0-4	4-10	-
(Non-) line of sight	LOS	NLOS	NLOS	NLOS
Mean excess delay (ns)	5	9.9	15.9	30.1
RMS delay spread (ns)	5	8	15	25

Table 2.2: IEEE802.15.3a UWB channel model parameters in four different scenarios.

2.4.3.4 Channel Decoding

After frequency domain equalization and de-interleaving, the MB-OFDM usually uses a hard or soft Viterbi decoder in order to estimate the transmitted data bits. For a detailed description of the Viterbi algorithm, the reader is referred to [63, 64].

2.4.4 MB-OFDM Performance Analysis in Realistic UWB Channel Environments

In this subsection, we present some simulation results in order to analyze the performance of the receiver described in subsection 2.4.3 over different indoor UWB channel scenarios defined in [65].

We simulated the mode I of the MB-OFDM which employs the first three subbands of 528 MHz (from 3.1 GHz to 4.684 GHz). Eeach realization of the channel model is generated independently and assumed to be time-invariant during the transmission of a frame. In our simulations, we have used the UWB channel models CM1-CM4 specified in the IEEE802.15.3a channel modeling sub-committee report [65]. These channel models are based on the Saleh-Valenzuela model [66], where multipath components arrive in clusters. Table 2.2 shows some of the parameters of the four models CM1-CM4. More details can be found in [65]. Punctured convolutional codes with rate 11/32, 1/2 and 3/4 are combined with time and/or frequency domain spreading, in order to achieve three (55, 160 and 480 Mbps) out of eight data-rates depicted in Tab. 2.1.

In our simulations, when there is no time or frequency redundancy (480 Mbps), a per subcarrier MMSE frequency-domain equalizer is used at the receiver. When time and/or frequency-diversity are exploited in the system, the maximal ratio combining (MRC) technique [60] is used to combine different diversity branches. In any case, a hard Viterbi decoder is used to recover the binary data.

Figures 2.17 and 2.18 depicts the results obtained over the CM1 and CM4 channels, respectively. We observe a similar behavior for different transmission modes over these two channel environments. As shown, the 55 Mbps mode provides the best performance due to the exploita-

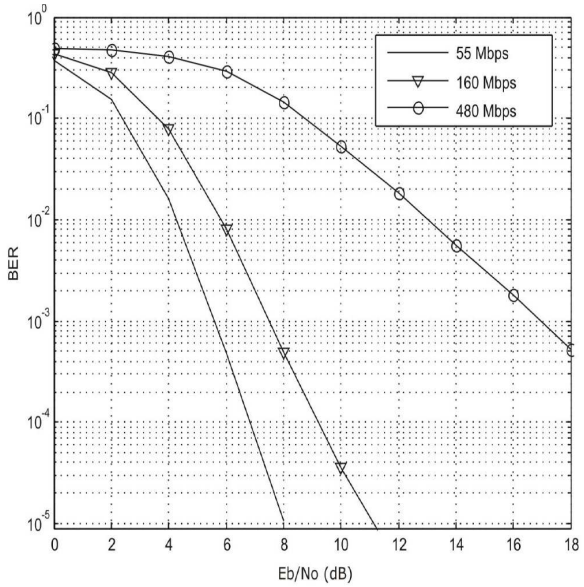

Figure 2.17: BER performance of the MB-OFDM system over the CM1 channel, for data rates of 55, 160 and 480 Mbps.

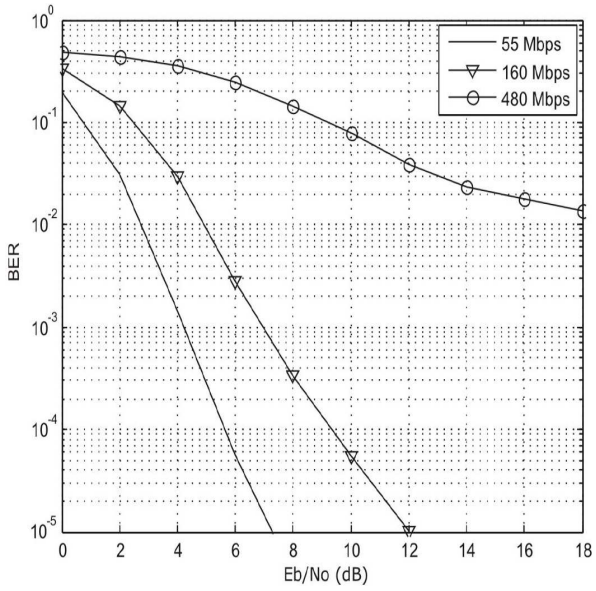

Figure 2.18: BER performance of the MB-OFDM system over the CM4 channel for data rates of 55, 160 and 480 Mbps.

tion of different diversity combining techniques. As observed from Fig. 2.17, at a BER of 10^{-5}, with about 3 dB of SNR degradation compared to the 55 Mbps mode, this mode provides a data rate of almost three times higher than the 55 Mbps mode.

Interesting results are observed from Fig. 2.19 for lowest (55 Mbps) and highest (480 Mbps) data rate modes, in various channel scenarios. As shown, the most robust data rate is 55 Mbps, where channel diversity is fully exploited by employing the MRC technique. We observe that

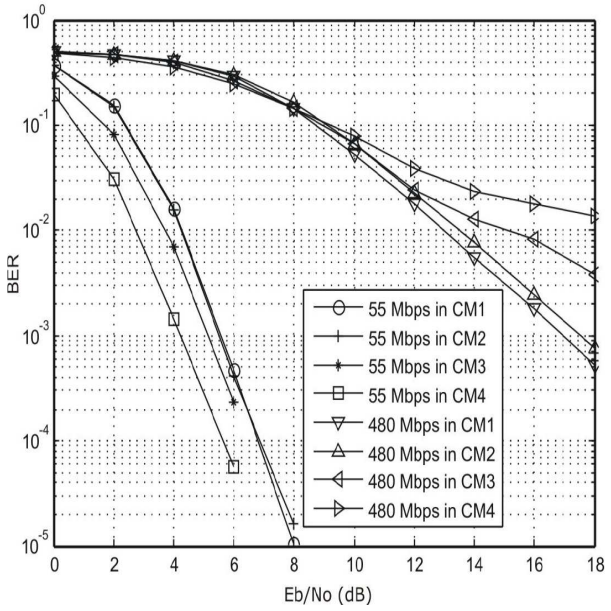

Figure 2.19: BER performance of the MB-OFDM system for data rates of 55 and 480 Mbps over different UWB channel scenarios.

MB-OFDM performs better in the CM4 channel environment than in the CM1 channel thanks to its inherent frequency diversity as shown in Fig. 2.19. In the 480 Mbps mode, we observe that the performance in CM1 is better than that in CM4. This is due to the absence of time and frequency domain spreading and to the high coding rate of 3/4 that prevents the exploitation of channel diversity. This leads to the worst BER for 480 Mbps mode in all channels as shown in Fig. 2.19.

2.5 Conclusion

The 7.5 GHz spectrum allocation by the FCC in 2002, initiated an extremely productive activity related to UWB from industry and academia. Since then, wireless communication experts considered UWB as an available spectrum to be utilized with a variety of transmission techniques, and not specifically related to the generation and detection of short duration impulse radios. UWB systems may be primarily divided into single band (impulse radio systems) and multiband systems.

Single band systems have simple transceiver architecture, and so are potentially lower cost. In addition, they may support many modulation schemes including orthogonal and antipodal schemes. However, this modulation must be combined with some form of spectrum randomization techniques to enhance the detection performance and to enable multiple access capability. Both TH and DS spectrum spreading techniques were presented. The main practical limitation

for impulse based UWB appears in the presence of highly resolved multipath UWB channels. In this situation, Rake receivers with a large number of fingers (ideally equal to the number of channel taps) must be used to capture the multipath energy. Obviously, this would result in significant implementation complexity for the Rake receiver. Another source of complexity in single band UWB systems is the need of high speed ADCs and equalizers working at several GHz.

Multiband UWB systems relax the requirement for high speed ADCs and provide a much more efficient method for capturing multipath energy. The most common multiband UWB modulation is the MB-OFDM which is supported by several key organizations inside the WiMedia Alliance. OFDM already enjoys an outstanding record with other standard organizations such as ADSL, IEEE802.11g, etc. Thus, MB-OFDM systems are potentially good technical solutions for the diverse set of high performance, short range UWB applications. Our simulations showed that in order to achieve a target BER, the basic receiver proposed in [4] for MB-OFDM has to exploit additional time and frequency diversity schemes (in addition to channel coding) which results in a loss of the spectral efficiency. Motivated by these observations, we will propose in subsequent chapters, some enhanced MB-OFDM reception schemes which do no waste the spectral efficiency.

Chapter 3

Wavelet Based Semi-blind Channel Estimation for Multiband OFDM

3.1 Introduction and Motivations

In the previous chapter, we presented the basic MB-OFDM reception scheme that has been proposed for the project of IEEE802.15.3a standard [3, 4]. As stated, the proposed receiver estimates the channel by using known training symbols transmitted at the beginning of the information frame, implicitly assuming a time invariant channel within each frame. Furthermore, the information symbols are detected by a one-tap frequency equalizer followed by a Viterbi decoder. Our numerical results showed that, in order to achieve a low BER, the transmitter may have to introduce time and/or frequency diversity techniques at the expense of significant spectral efficiency loss. Obviously, this is not in agreement with the target of UWB technology which aims at providing data rates of several hundreds of Mbps. Moreover, it is well known that an accurate pilot based channel estimation requires multiple channel-uses per frame for pilot transmission, which can result in a considerable reduction of the system throughput.

Recent works have reported promising results on the combination of the channel estimation and of the data decoding process for OFDM systems. In particular, iterative or "*turbo*" processing that includes the channel estimation into the iterative process of decoding turbo-like codes is addressed in [5, 6], for instance. Similar works on this subject have investigated the use of the Expectation-Maximization (EM) algorithm [7] for joint semi-blind channel estimation and data detection [8, 9].

Though iterative joint channel estimation and data detection outperforms receivers using a pilot assisted channel estimation approach, it has a higher complexity, which may be of critical concern for practical implementations. This complexity is mainly driven by the number of

estimated parameters for channel updating and the decoding algorithm within each iteration.

In this chapter, as an alternative to the basic receiver of [4], we propose a reduced-complexity receiver for MB-OFDM systems that outperforms the basic proposed scheme without scarifying any spectral efficiency. To this end, we consider an iterative semi-blind channel estimation based on the EM algorithm, with the objective of minimizing the number of estimated parameters and enhancing the estimation accuracy. This is achieved by expressing the unknown channel impulse response (CIR) in terms of discrete wavelet series, which has been shown to provide a *parsimonious* representation [45, 67]. The adopted wavelet based channel representation enables us to choose a particular *prior* distribution for the channel wavelet coefficients, that renders the maximum *a posteriori* (MAP) channel estimation equivalent to a hard thresholding rule at each iteration of the EM algorithm. The latter is then exploited to reduce the estimator computational load by discarding "insignificant" wavelet coefficients from the estimation process. We notice that this wavelet prior is not associated to a specific propagation environment since its parameters are learnt from the observed data. Moreover, since the probability of encoded bits are involved in the EM computation, we naturally combine the iterative process of channel estimation with the decoding operation of encoded data.

This chapter is organized as follows. In Section 3.2, we provide the state of the art of different channel estimation techniques proposed for OFDM systems. In order for this thesis to be self-sufficient, we describe in Section 3.3 the EM algorithm and its extension to MAP parameter estimation. Section 3.4 is devoted to our proposed wavelet based channel estimation and data detection algorithm. We will first describe a MAP version of the EM algorithm for channel estimation and then show how the number of estimated parameters can be reduced through the EM iterations. The combination of the channel estimation with the decoding operation is discussed, as well as some implementation issues. We will also illustrate via simulations, the performance of the proposed receiver in different realistic UWB channel environments. Finally, Section 3.5 concludes the chapter.

3.2 State of the Art of Channel Estimation for OFDM Systems

It is well known that in any wireless communication system, reliable coherent data detection is not possible unless an accurate channel estimate is available at the receiver. However, differential modulation can be used to detect the transmitted data without any channel knowledge. An example of this scheme is the differential phase shift keying (DPSK) modulation which is adopted in the European digital audio broadcast (DAB) norm [68]. Although simple, differential modulations entail restrictions on the choice of the constellation and leads to a 3 dB loss over Gaussian channels compared to coherent modulations [60]. As an alternative to DPSK,

Engels *et al.* have proposed in [69] the differential amplitude and phase shift keying (DAPSK) modulation in which the amplitude of the transmitted symbols is also differentially encoded. A coherent modulations allows the use of arbitrary constellations and constitute a good and flexible choice when channel time variations are relatively slow. By using high spectral effficiency constellations, coherent modulation can achieve high data rates as in the digital video braodcast (DVB) norm [70].

Typically, channel estimation techniques for coherent OFDM modulations can be classified into two different families: **i**) techniques based on the transmission of training sequences and **ii**) "blind" channel estimation methods. Blind methods are based on the statistics of the unknown data symbols and the statistical properties of the channel and do not require any pilot symbol. Between these two extremes, there exists "semi-blind" methods (considered in this chapter) which require a small number of pilots, usually used for the algorithm initialization.
There exists a very rich literature on channel estimation techniques for OFDM systems. In what follows, we present some of the main contributions concerning each of the aformentioned channel estimation techniques.

3.2.1 Pilot-only Based Channel Estimation Techniques

We describe here different channel estimation techniques based only on pilot symbols. These techniques which are also called pilot symbol assisted modulation (PSAM), were introduced for single carrier systems by Moher and Lodge [71] and later analyzed by Cavers [72].

3.2.1.1 Techniques Based on the Least-squares Criterion

It is well known that in OFDM, thanks to the CP insertion, each subcarrier experiences a non frequency selective fading channel. Let us consider a frame of N_{SYM} consecutive OFDM symbols with N_c subcarriers per symbol. The transmitted frame can be viewed as a time-frequency grid of dimension $(N_{\mathrm{SYM}} \times N_c)$. The received symbol $y_{k,n}$ at each position of the grid is written as [32]

$$y_{k,n} = H_{k,n}\, s_{k,n} + b_{k,n} \tag{3.1}$$

where $H_{k,n}$ and $s_{k,n}$ are respectively the channel frequency response (CFR) and the transmitted symbol at the (k,n) position of the grid, and $b_{k,n}$ is the AWGN affecting the transmission of data. A simple method is to estimate $H_{k,n}$ so as to minimize $|y_{k,n} - H_{k,n}\, s_{k,n}|^2$. In the presence of independent and identically distributed (i.i.d.) Gaussian noise, the least squares (LS) estimate is given by

$$\widehat{H}_{k,n}^{\mathrm{LS}} = \frac{y_{k,n}}{s_{k,n}} \tag{3.2}$$

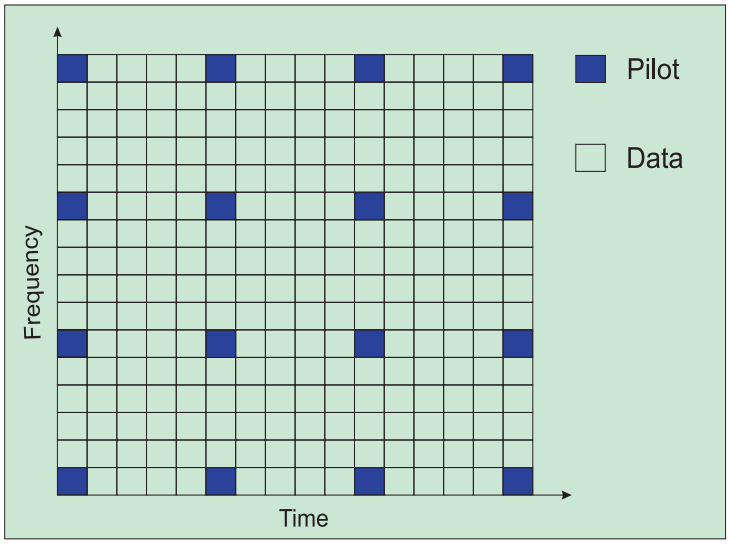

Figure 3.1: An example of two dimensional pilot arrangment in OFDM packet transmission.

where $s_{k,n}$ is assumed to be known at the receiver. In PSAM, some training symbols are inserted in both time and frequency in the two dimensional grid (see Fig. 3.1). The pilot repetition interval depends on the channel coherence time and on the coherence bandwidth. Obviously, in PSAM an estimate of the channel is obtained only at the pilot positions. In order to estimate the channel at all time and frequency positions, different interpolation techniques are proposed in the litterature, that we describe briefly in the following.

Polynomial Interpolation: Interpolation can be performed only in the frequency domain by considering a single OFDM symbol [73] or in both time and frequency domains [74]. For instance, in [73], Rinne and Renfors propose two LS channel estimators. The first technique estimates the channel at the pilot frequencies and assumes that the channel is constant over a frequency bandwidth equal to the frequency spacing between two pilot tones (centered at the pilot tone frequency). The second technique performs a frequency domain interpolation by considering that the channel is changing linearly between two consecutive pilot tones. In [74], Chang and Su proposed a two dimensional interpolation in the sense of LS by considering a paraboloid model for the channel as

$$H_{k,n} = a_1 k^2 + a_2 kn + a_3 n^2 + a_4 k + a_5 n + a_6 \tag{3.3}$$

where the coefficients $a_1, ..., a_6$ are determined by minimizing

$$\sum_{k=0}^{N_c-1} \sum_{n=0}^{N_{\mathrm{SYM}}-1} |\widehat{H}_{k,n}^{\mathrm{LS}} - H_{k,n}|^2. \tag{3.4}$$

Higher order piecewise polynomial interpolations are proposed in [75] where the bandwidth of

an OFDM symbol is divided into different frequency bands. Over each frequency band, the channel is modeled as a polynomial and the whole OFDM symbol is recovered by considering every elementary polynomials.

Interpolation Using FIR Filters: As an alternative to polynomial interpolation, Moon and Choi proposed different two dimensional interpolations in [76]. This technique uses Gaussian or cubic spline filters for interpolation. Moreover, the adopted interpolation filters have a finite impulse response (FIR) with three coefficients and use the estimated channel coefficients provided by three pilot symbols. However, conversely to the approach proposed in [74], the bi-dimensional interpolation is performed by two uni-dimensional FIR filters, one realizing the interpolation in the time direction and the other in the frequency direction. In [77], channel is estimated at the pilot frequencies and then converted to the time domain by an IFFT. Then the signal is interpolated by using a cubic spline filter before being converted to the frequency domain. Ozinawa *et al.* proposed in [78] a method for selecting an FIR interpolation filter among a predefined set of filters.

3.2.1.2 Techniques Based on the Minimization of the Mean-squared Error

The minimum mean square error (MMSE) criterion has been extensively used for OFDM channel estimation. The optimal two dimensional channel estimation in the sense of the MMSE was addressed in [79]. Due to the high complexity of this estimator, different sub-optimal estimators with lower complexities were proposed in [80] [81]. Other works in this area have suggested the use of two cascaded uni-dimensional filters performing respectively an MMSE estimation in time and frequency domains.

In [82], van de Beek *et al.* proposed modifications to the MMSE CFR estimator under the assumption of a finite length impulse response. This work uses the theory of low-rank approximations, based on the truncation of the discrete Fourier transform (DFT) matrices. Inspired by the observations in [82], Edfors *et al.* introduced in [83] a linear MMSE (LMMSE) estimator by using the singular value decomposition (SVD) of the channel frequency response covariance matrix. The complexity of this estimator is reduced by using optimal rank reduction [84]. The rank reduction technique is adopted by Hsieh in [85] where the channel transfer function of pilot tones are estimated by using the low-rank MMSE estimator, and the channel transfer function of data tones are interpolated by a piecewise linear interpolation. In [86], Li *et al.* extended the estimator of [83] and derived an MMSE estimator which makes full use of of the channel frequency response correlation at different times and frequencies. In this method, the correlation of the channel frequency response is separated into the multiplication of time and frequency-

domain correlation functions. A similar method was proposed in [87] where an MMSE filtering is performed separately in time and frequency directions. The proposed algorithm calculates the actual delay spread of the channel and uses this result to improve the frequency domain filtering which is performed by an adaptive filter. In [88], Seshadri *et al.* developed OFDM channel estimation in the case of space-time coding and multiple antennas and derived some bounds for the mean square error (MSE).

Some of the techniques found in the litterature are based on Kalman filtering. For instance, in [89], Tufvesson and Maseng used a Kalman filter to estimate the time-variant channel taps which are modeled by a first order auto-regressive (AR) process.

3.2.2 Decision-directed Techniques

Tha main drawback of pilot-only based channel estimation techniques is the loss of spectral efficiency due to the pilot overhead. The number of transmitted training sequence can be reduced by adopting decision directed methods which require the transmission of a single OFDM pilot symbol at the beginning of the frame. The simple idea of this method is that in the *absence* of transmission errors, one can use the detected symbols as *a posteriori* reference signals for channel estimation instead of pilot symbols. A pioneering work in this area is that initiated by Frenger and Svensson [90] [91] where a decision directed coherent detector for single carrier and multicarrier systems based on an MMSE channel estimation is proposed. However, the latter technique assumes that at each instant, all the previous decisions are correct. Obviously, this is not a realistic assumption in practical situations. In order to mitigate the problem of error propagation, Mignone and Morello [92] proposed a decision directed channel estimator that exploits forward error correction (FEC) codes. In this technique, reliable decoded symbols are used for channel estimation/updating and the problem of error propagation is reduced. However, the channel estimation loop of this method requires long delays which can become a limiting factor for its implementation especially in mobile channel environments. In [93], Bulumulla *et al.* considered a MAP receiver based on decision feedback where the *a posteriori* probability of the symbols are calculated by using a Kalman filter.

Recenty, Akhtman *et al.* [94] proposed a decision directed channel estimation for OFDM systems which employs a fractionally-spaced model for the CIR. The adoption of this fractionally-spaced model is motivated by the fact that in realistic channel environments, one has no control over the delay of CIR taps.

3.2.3 Totally Blind Channel Estimation Techniques

The increasing need for high data rates motivated the search for "blind" channel identification and equalization methods as they save bandwidth by avoiding the use of training sequences. Numerous blind algorithms have been developed in the litterature [95], where several works have focused specifically on multicarrier systems. Existing blind channel estimation methods for OFDM systems usually fall in either the statistical or the deterministic category.

Among statistical blind methods, the inherent cyclostationarity induced by the CP at the transmitter has been exploited in [96] [97]. Specifically, Health and Giannakis [96] proposed a blind method based on the cyclostationarity property of the time-varying correlation of the received data samples due to the CP insertion at the transmitter; however this approach suffers from slow convergence of the estimator. Other methods in this category are based on a subspace decomposition [98] and take advantage of the inherent redundancy introduced by the CP to blindly estimate the channel. These algorithms exploit the orthogonality of the *noise subspace* and *signal subspace* and rely on the autocorrelation matrix of the pre-DFT received data [99] [100] [101]. A subspace based blind channel estimation is presented by Zhou *et al.* [102] for space-time coded OFDM systems employing linear precoding [103]. By using linear precoding, the latter algorithm guarantees channel identifiability, regardless of the underlying channel zero locations. Recently, Lin and Petropulu [104] proposed a nonredundant nonunitary linear precoded OFDM system and used the correlation introduced by the precoding matrix to estimate the channel at the receiver based on cross-correlation operations. Other than CP, virtual subcarriers have also been exploited for purposes of blind channel estimation. In some of the standards like IEEE802.11a, there are some so-called virtual subcarriers that are left unmodulated in order to ease filter implementation at the band edge. A maximum likelihood (ML) joint blind channel and data estimator that exploits the finite alphabet property of modulation and the presence of virtual carriers is presented in [105]. However, each of the above algorithms has its own limitation. For example, subspace based methods typically require a large number of OFDM symbols and thereby introduce a considerable latency into the overall system. A finite alphabet based algorithm can be applied only to a constant modulus signal. Thus, estimators that require few OFDM symbols are preferable, as they can also operate over nonzero Doppler channels without introducing any appreciable delay.

Deterministic methods are usually applied on received OFDM symbols after FFT demodulation and use the finite alphabet property of the received symbols. They usually converge much faster than statistical methods, while involving higher computational complexity at the receiver. This category includes ML based approaches [106] and the exhaustive search approach of [107]. The method of [106] has the advantage of producing a channel estimate from a single

received OFDM symbol. Its principal drawback is the huge computational complexity needed to execute the maximization operation embedded in the algorithm. However, the channel estimate still has a phase ambiguity. In [108], Necker and Stuber modified the basic ML method of [106] for the case of PSK signals and developed a low complexity iterative version of the ML based algorithm that exploits the time domain correlation of the transfer function.

3.2.4 Semi-blind Channel Estimation Techniques

Blind methods can also be used in cooperation with training data in order to better track channel variations and to enable faster convergence. In that case, they are referred to as "semi-blind" methods [109]. Usually, one or two pilot symbols are transmitted at the beginning of each frame for synchronization and initial channel estimation purposes. Most of blind algorithms described above can be extended to a semi-blind method. For instance in [99], a pilot symbol is used to avoid the convergence period of the blind subspace algorithm. Besides, pilot subcarriers are used to overcome the inherent scalar indetermination that is common in most of blind algoritms (see [104] for instance). In [110], the authors introduce a zero-padded OFDM (ZP-OFDM) system in which the null samples inserted between each OFDM modulated block are replaced by a pseudorandom scalar sequence. The observation provided by this way is used to perform a semi-blind channel estimation and tracking.

As stated in the introduction, an efficient and extensively used method for semi-blind channel estimation is that based on the EM algorithm. The EM algorithm is an iterative algorithm that can be used to approximate an ML or MAP solution of the unknown channel when the transmitted symbols are unknown at the channel estimator (blind situation). A brief description of this algorithm is provided in the next section. For a more detailed and general exposition of the EM algorithm, the reader is urged to read [7, 111].
Since the algorithm presented in this chapter is an EM type algorithm, in what follows, we present different categories of channel estimation methods based on the EM algorithm.

3.2.4.1 EM Based Algorithms for ML Channel Estimation

Several papers have addressed ML channel estimation without using any *a priori* information for the unknown channel. Very recently in [112] (see also [113]), the authors proposed several improvements to the EM algorithm for MIMO channel estimation which can also be applied to OFDM systems. In particular, they proposed an unbiased EM channel estimator that outperforms the classical EM estimator. In [114], the authors used the EM algorithm to estimate the channel without any knowledge on the transmitted symbols. Xie *et al.* [115] proposed an EM algorithm for MIMO-OFDM systems employing space-time coding which exhibits fast conver-

gence. However, the proposed algorithm is just used as an iterative way to calculate their LS solution since the transmitted data symbols are assumed to be known for channel estimation. Ma *et al.* [114] presented several EM based algorithms to estimate the CFR or the CIR of an OFDM system that is subject to slow time-varying frequency-selective fading. Although the proposed estimators are simple, they do not consider any error correction code at the transmitter. However, it is well known that uncoded OFDM systems are likely to perform poorly in frequency-selective fading channels. To enhance the OFDM's performance, many architectures utilize channel coding in conjunction with OFDM (COFDM) [116]. As we shall see later in this chapter, the probability of encoded bits are involved in the EM algorithm computation. Mazet *et al.* [117] were among the first to propose an EM based channel estimation that takes into account the probability of encoded data coming from a soft-input soft-output (SISO) decoder. They used the well known forward-backward algorithm [11] to provide directly at the E-step of the EM algorithm the probability of the encoded bits (see the description of the EM algorithm in the next section). In that way, the receiver can perform a joint channel estimation and data decoding. Following a similar idea, Touati *et al.* improved the approach of [117] by introducing the EM-block algorithm [118] which estimates the CFR by assuming that each channel frequency gain is constant over a block composed of several OFDM symbols.

Recent works have reported promising results on the combination of EM based channel estimation and data decoding process (see [5] [10, 119–122]). In practical COFDM systems, the use of a SISO decoder [11] is capable of supplying the probabilities of transmitted symbols, which is exactly what the EM algorithm requires for channel estimation. The availability of the probability information on transmitted symbols also helps to mitigate the error propagation problem in decision directed methods (as discussed previously), for example, by using only "reliable" symbols or using the so-called "soft" symbols. Iterative or "turbo" processing that includes the channel estimation into the iterative process of decoding turbo like codes is addressed in [5,119,122], for instance. However, the limitation of the receiver proposed in [5] is that channel encoding must be performed accross subcarriers belonging only to one OFDM symbol and not over the whole frame. A similar method is proposed in [119] where a two dimensional channel estimation is performed by applying a concatenation of two one-dimensional estimators in the frequency and time domains. Ma *et al.* addressed in [123], the combination of an EM based channel estimation with polynomial fitting. In this work, the algorithm obtains first a near optimal channel estimate by using the observation of a single OFDM symbol. Then to further improve the performance, polynomial fitting is adopted by gathering channel estimates of several consecutive OFDM frames.

3.2.4.2 EM Based Algorithms for MAP Channel Estimation

We refer to MAP channel estimation, every contribution that exploits an additional information, usually called *a priori* information, in the *channel* estimation process. Clearly, in our classification, this *prior* information concerns only the channel parameters and in this sense the algorithms using prior probabilies on data symbols coming from the soft decoder (e.g., reference [117]) are not viewed as a MAP channel estimator.

An EM based channel estimation that exploits the additional observations provided by the CP is presented in [124]. In this algorithm, a normal prior distribution is chosen for the CIR. However, except for theoretical Rayleigh fading where normal priors can be used, this choice is not always justified for OFDM channels. Moreover, the performance improvement (reported in terms of MSE) achieved by using observations provided by the CP is always limited due to the presence of ISI. A similar algorithm is proposed by the same authors in [125] for time-variant channels where the frequency response of the channel is modeled by a first order AR equation. Mazet *et al.* have also adopted in [126] a first order AR modeling of the channel time-variations. In this work, to take the AR model into consideration in the EM algorithm, the channel frequency coefficient at the previous OFDM symbol is considered as an additive observation for the estimation of the current channel parameter. In order to further improve the performance, the latter algorithm estimates the channel frequency coefficients once in the forward order and then in the backward order.

Jaffrot and Siala [6] proposed a turbo channel estimation method based on a Karhunen-Loève (KL) expansion of the unknown channel frequency coefficients. This method considers a frequency and time selective fading channel characterized by its spaced-frequency spaced-time correlation matrix [60]. This two-dimensional channel correlation matrix is theoretically evaluated in advance and assumed to be known at the receiver. Using this, the unknown channel is represented as the weighted sum of the eigen vectors of its covariance matrix. In order to reduce the estimation complexity, the algorithm estimates only a predefined subset (much smaller than the number of OFDM subcarriers) of parameters corresponding to the largest eigen values of the channel covariance matrix. The probability of encoded bits needed in the EM algorithm are provided by the BCJR [11] algorithm. Although the latter method achieves BER performance near the theoretical receiver with perfect channel knowledge, its implementation requires the channel second order statistics as an *a priori* information which cannot be available (at least in advance) in realistic situations. Recently in [127], the idea presented by Jaffrot *et al.* has been extended to OFDM systems with space-frequency transmit diversity. However, unlike the approach treated in [6], the authors have adopted a two-step procedure where the channel estimation and data detection are performed separately.

3.3 Brief Description of the EM Algorithm

3.3.1 Introduction

The EM algorithm is a broadly applicable approach to the iterative computation of ML estimates, useful in a variety of *incomplete-data* problems, where other iterative algorithms may turn out to be more complicated. At each iteration of the EM algorithm, there are two processing steps called the *expectation step* (or the E-step) and the *maximization step* or (the M-step). That is why the algorithm is called EM. This name was given by Dempster, Laird, and Rubin (1977), referred usually as DLR, in their fundamental paper [111]. However, the EM algorithm was discovered and employed independently by several different researchers until DLR brought their ideas together, proved its convergence and coined the term "EM algorithm". The idea behind the EM algorithm being intuitive and natural, algorithms like EM had already been formulated and applied to a variety of problems.

In signal processing applications, the largest area of interest for the EM algorithm is in ML estimation/detection problems with incomplete-data, where there are missing data, truncated distributions, censored and grouped observations which result in complicated likelihood functions. However, the EM principle can be applied to a variety of situations where the incompleteness of data is not so natural or evident. These include statistical models such as random effects, mixtures, convolutions, log linear models, etc. A large list of references is found in [128].

The basic idea behind the EM algorithm is to associate with the given *incomplete-data problem*, a *complete-data problem* for which ML estimation is computationally more tractable. The methodology of the EM algorithm then consists in reformulating the problem in terms of this more easily solved complete-data problem. The E-step consists in manufacturing data for the complete-data problem using the incomplete observed data set and the current value of the unknown parameters, so that a simpler M-step computation can be applied to this "completed" data set. More precisely, it is the log-likelihood of the complete-data problem that is computed in the E-step. As it is partly based on unobservable (or hidden) data, it is replaced by its conditional expectation given the observed data, where this E-step is affected using the current estimate of the unknown parameters. Starting from suitable initial parameter values, the E- and M-steps are repeated until convergence.

3.3.2 General Statement of the EM Algorithm

3.3.2.1 Mathematical Formulation

Let $\mathcal{Y}$ denote the sample space of the observations, and let $\boldsymbol{y} \in \mathbb{R}^m$ denote an observation from $\mathcal{Y}$ of size m. Let $\mathcal{X}$ denote an underlying space and let $\boldsymbol{x} \in \mathbb{R}^n$ be an outcome from $\mathcal{X}$ with $m < n$.

The data vector $\boldsymbol{x}$ is referred to as the *complete-data*. The complete data $\boldsymbol{x}$ is not observed directly but only by means of $\boldsymbol{y}$ where $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x})$, and $\boldsymbol{y}(\boldsymbol{x})$ is a many-to-one mapping from $\mathcal{X}$ to $\mathcal{Y}$. As shown in Fig. 3.2, an observation $\boldsymbol{y}$ determines a subset of $\mathcal{X}$, which is denoted as $\mathcal{X}(\boldsymbol{y})$. The pdf of the complete-data vector is $f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^r$ is the set of unknown

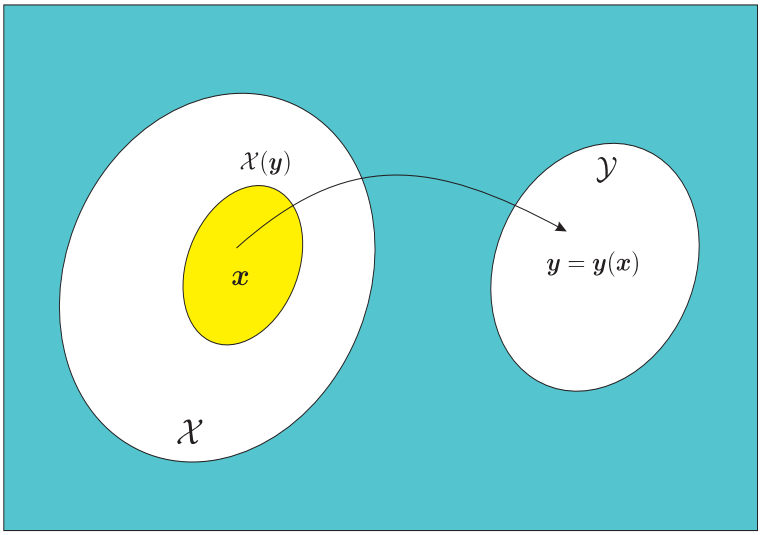

Figure 3.2: An illustration of the complete- and incomplete-data sets of the EM algorithm.

parameters that we have to estimate. (We will refer to the *density* of the random variables for convenience, even for discrete random variables for which probability mass functions (pmf) would be appropriate). Moreover, the pdf f is assumed to be a continuous function of $\boldsymbol{\theta}$ and appropriately differentiable. The ML estimate of $\boldsymbol{\theta}$ is assumed to lie within the region $\boldsymbol{\Theta}$. The pdf of the incomplete-data is

$$g(\boldsymbol{y}|\boldsymbol{\theta}) = \int_{\mathcal{X}(\boldsymbol{y})} f(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x} \tag{3.5}$$

and denotes the incomplete-data likelihood function. Let $L_i(\boldsymbol{\theta}) = \log g(\boldsymbol{y}|\boldsymbol{\theta})$ and $L_c(\boldsymbol{\theta}) = \log f(\boldsymbol{x}|\boldsymbol{\theta})$ denote respectively the incomplete- and complete-data log-likelihood. The integral operation in (3.5) may render very difficult the estimation of the parameter $\boldsymbol{\theta}$ which maximizes the likelihood function $g(\boldsymbol{y}|\boldsymbol{\theta})$, even if the function $\log f(\boldsymbol{x}|\boldsymbol{\theta})$ is easy to maximize. This remark justifies the idea of the EM algorithm.

As stated before, the basic idea behind the EM algorithm is that we would like to find $\boldsymbol{\theta}$ to maximize $L_c(\boldsymbol{\theta}) = \log f(\boldsymbol{x}|\boldsymbol{\theta})$, but we do not have the data $\boldsymbol{x}$ to compute the log-likelihood. So instead, we maximize the expectation of $\log f(\boldsymbol{x}|\boldsymbol{\theta})$ given the data $\boldsymbol{y}$ and our current estimate of $\boldsymbol{\theta}$. This can be expressed in two steps.

More specifically, let $\boldsymbol{\theta}^{(0)}$ be some initial value for $\boldsymbol{\theta}$. Then at the first iteration, the E-step

requires the calculation of

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(0)}) = \mathbb{E}\big[\log f(\boldsymbol{x}|\boldsymbol{\theta})\big|\boldsymbol{y}, \boldsymbol{\theta}^{(0)}\big] \tag{3.6}$$

where $Q(.,.)$ is called the *auxiliary function*. It is important to distinguish between the first and the second arguments of the auxiliary function. The second argument is a conditioning argument to the expectation and is regarded as fixed and known at every E-step. The first argument conditions the likelihood of the complete-data.

The M-step requires the maximization of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(0)})$ with respect to $\boldsymbol{\theta}$ over the parameter space $\boldsymbol{\Theta}$. That is, we choose $\boldsymbol{\theta}^{(1)}$ such that

$$Q(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(0)}) \geq Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(0)}) \tag{3.7}$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The E- and M-steps are then carried out again, but this time with $\boldsymbol{\theta}^{(0)}$ replaced by the current estimate $\boldsymbol{\theta}^{(1)}$. On the $(t+1)$-th iteration, the E- and M-steps are defined as follows.

- **E-step:** Calculate $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ where

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) = \mathbb{E}\big[\log f(\boldsymbol{x}|\boldsymbol{\theta})\big|\boldsymbol{y}, \boldsymbol{\theta}^{(t)}\big] \tag{3.8}$$

- **M-step:** Choose $\boldsymbol{\theta}^{(t+1)}$ to be any value of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ that maximizes $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ as

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \; Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}). \tag{3.9}$$

We mention that the expectation in the E-step is with respect to all unobserved (or hidden) variables in the complete-data set $\mathcal{X}$. We also note that the maximization in the M-step is with respect to the first argument of the Q function, i.e., the conditioner of the complete-data likelihood.

After initialization, the E- and M-steps are alternated repeatedly until convergence. Convergence may be determined by examining when the parameters remain almost unchanges, i.e., stop when $\|\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^{(t-1)}\| < \epsilon$ or $L_i(\boldsymbol{\theta}^{(t)}) - L_i(\boldsymbol{\theta}^{(t-1)}) < \epsilon$, for some small value of ϵ and some appropriate distance measure $\|.\|$.

3.3.2.2 Monotonicity of the EM Algorithm

DLR showed that the incomplete-data likelihood function $g(\boldsymbol{y}|\boldsymbol{\theta})$ is not decreased after an EM iteration. This is formulated in the following theorem proved in [111].

Theorem 3.3.1.

$$Q(\boldsymbol{\theta}^{(t+1)}, \boldsymbol{\theta}^{(t)}) \geq Q(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t)}) \Longrightarrow g(\boldsymbol{y}|\boldsymbol{\theta}^{(t+1)}) \geq g(\boldsymbol{y}|\boldsymbol{\theta}^{(t)}) \quad \text{for all } \; t. \tag{3.10}$$

3.3.2.3 Convergence to a Stationary Value

As shown in the last section, for a sequence of likelihood values $\{g(\boldsymbol{y}|\boldsymbol{\theta}^{(t)})\}$, $g(\boldsymbol{y}|\boldsymbol{\theta}^{(t)})$ converges monotically to some stationary value g^*. The stationary point may be a local maximum or a saddle point of the likelihood function. In general, if $g(\boldsymbol{y}|\boldsymbol{\theta})$ has several stationary points, convergence of the EM sequence to either type (global or local maximum, saddle points) depends on the choice of the starting point $\boldsymbol{\theta}^{(0)}$. Obviously, when the likelihood function is unimodal in $\boldsymbol{\theta}$ (and a certain differentiability is satisfied), any EM sequence converges to the unique global maximum irrespective of its starting. In what follows, we state without proof the main convergence theorem given by Wu in [129].

Theorem 3.3.2. *Let $\{\boldsymbol{\theta}^{(t)}\}$ be a sequence of parameters obtained from successive maximization of the auxiliary function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ at the M-step. Then all the limit points of $\{\boldsymbol{\theta}^{(t)}\}$ are stationary points of $g(\boldsymbol{y}|\boldsymbol{\theta}^{(t)})$ and $g(\boldsymbol{y}|\boldsymbol{\theta}^{(t)})$ converges monotonically to $g^* = g(\boldsymbol{y}|\boldsymbol{\theta}^*)$ for some stationary point $\boldsymbol{\theta}^*$.*

3.3.3 Extension of the EM Algorithm to MAP Parameter Estimation

Up to now, we addressed the EM algorithm for ML estimation. Let us now consider a MAP criterion for the estimation of the unknown parameter $\boldsymbol{\theta}$ of which ML estimation is a particular case. Considering some *prior* distribution $\pi(\boldsymbol{\theta})$ for the unknown parameter, the MAP estimate is given by

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \big\{ \log g(\boldsymbol{y}|\boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta}) \big\}. \tag{3.11}$$

When the likelihood function $g(\boldsymbol{y}|\boldsymbol{\theta})$ is hard to maximize, the EM algorithm is a mean for obtaining MAP estimates of a parameter $\boldsymbol{\theta}$.
The EM algorithm for MAP estimation can be summarized as follows.

- **E-step:** Calculate $Q_{\text{map}}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ where

$$\begin{aligned} Q_{\text{map}}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) &= \mathbb{E}\big[\log f(\boldsymbol{x}|\boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta}) \big| \boldsymbol{y}, \boldsymbol{\theta}^{(t)} \big] \\ &= Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) + \log \pi(\boldsymbol{\theta}) \end{aligned} \tag{3.12}$$

- **M-step:** Choose $\boldsymbol{\theta}^{(t+1)}$ to be any value of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ that maximizes $Q_{\text{map}}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ as

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \; Q_{\text{map}}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}). \tag{3.13}$$

We note that the E-step of MAP estimation differs from the E-step of ML estimation by the additive term $\log \pi(\boldsymbol{\theta})$. The presence of the term $\log \pi(\boldsymbol{\theta})$ can also be exploited to render the auxiliary function concave. The M-step is also different since the maximization is performed

over a modified auxiliary function.

The aforementioned convergence properties of the ML based EM are also valid for MAP estimation [111] [7]. Thus, each iteration of the EM algorithm is guaranteed to increase the logarithm of the incomplete-data *a posteriori* probability, that is

$$\log g\big(\boldsymbol{y}|\boldsymbol{\theta}^{(t+1)}\big) + \log \pi\big(\boldsymbol{\theta}^{(t+1)}\big) \geq \log g\big(\boldsymbol{y}|\boldsymbol{\theta}^{(t)}\big) + \log \pi\big(\boldsymbol{\theta}^{(t)}\big). \tag{3.14}$$

3.4 MB-OFDM Wavelet Domain Channel Estimation and Data Detection

In this section, we present our semi-blind channel estimation algorithm for MB-OFDM systems which is based on a wavelet decomposition of the unknown CIR. We start by introducing our system model and then provide a brief description of the UWB channel model. This leads us to present our motivations for a wavelet domain channel etimation. Then we present our specific contribution.

3.4.1 System Model for MB-OFDM Transmission

As stated in chapter 2, MB-OFDM divides the spectrum between 3.1 and 10.6 GHz into several non-overlapping subbands each one occupying 528 MHz of bandwidth [3]. Information is transmitted using OFDM modulation over one of the subbands in a particular time-slot. The MB-OFDM system uses a time-frequency code (TFC) to select the center frequency of different subbands which is used not only to provide frequency diversity but also to distinguish between multiple users (see Fig. 2.12). As shown in Fig. 3.3 (repeated for convenience), after channel coding, a block of bits is interleaved and mapped to QPSK symbols. Here, we consider MB-OFDM in its basic mode, i.e., employing the first three subbands with N data subcarriers over each subband. At the receiver, assuming a CP longer than the channel maximum delay spread and perfect carrier synchronization, OFDM converts a frequency-selective channel into N parallel flat fading subchannels [32]. Under these conditions, the transmission of the n-th OFDM symbol (inside a frame of size N_{SYM}) over the i-th subband can be written as

$$\mathbf{y}_{i,n} = \mathcal{D}_{\mathbf{s}_{i,n}}\, \bar{\mathbf{H}}_{i,n} + \mathbf{z}_{i,n} \quad i \in \{1,2,3\}, \quad n = 1, \ldots, N_{\mathrm{SYM}} \tag{3.15}$$

where $\mathcal{D}_{\mathbf{s}_{i,n}} \triangleq \mathrm{diag}(\mathbf{s}_{i,n})$, $(1 \times N)$ vectors $\mathbf{y}_{i,n}$, $\mathbf{s}_{i,n}$ and $\bar{\mathbf{H}}_{i,n}$ denote respectively the received and transmitted symbols, and the channel frequency response, the noise vector $\mathbf{z}_{i,n}$ is assumed to be a zero-mean circularly symmetric complex Gaussian (ZMCSCG) random vector with distribution $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbb{I}_N)$.

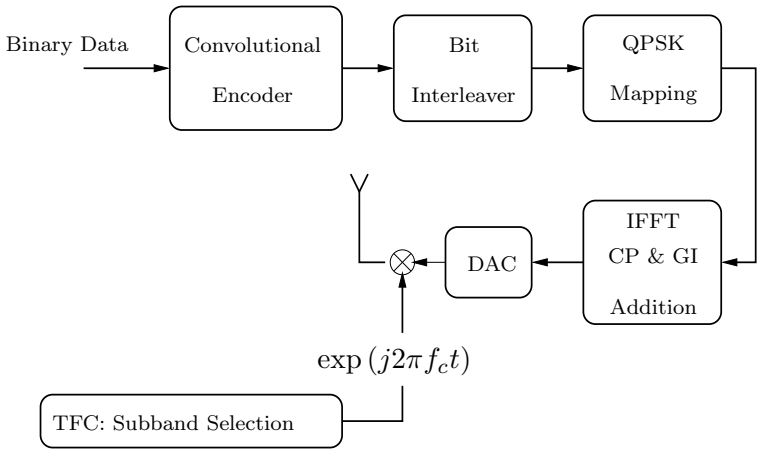

Figure 3.3: TX architecture of the multiband OFDM system.

In what follows, we group the data and observations corresponding to three subbands of the MB-OFDM system in a single vector (we call it a *compound* OFDM symbol). As described above, the data are sent in each subband in sequence. This ensures that data as well as observations are available within each subband at each time-slot. The reason behind this operation is provided in Section 3.4.3. This operation is depicted in Fig. 3.4 and can be written in the frequency domain as [1]

$$\mathbf{Y}_m = \mathcal{D}_{\mathbf{S}_m}\,\mathbf{H}_m + \mathbf{Z}_m \quad m = 1, \ldots, M_{\mathrm{SYM}} \tag{3.16}$$

where $\mathcal{D}_{\mathbf{S}_m} \triangleq \mathrm{diag}(\mathbf{S}_m)$, $\mathbf{Y}_m = [\mathbf{y}_{1,n}, \mathbf{y}_{2,n+1}, \mathbf{y}_{3,n+2}]^T$, $\mathbf{S}_m = [\mathbf{s}_{1,n}, \mathbf{s}_{2,n+1}, \mathbf{s}_{3,n+2}]^T$, $\mathbf{H}_m = [\bar{\mathbf{H}}_{1,n}, \bar{\mathbf{H}}_{2,n+1}, \bar{\mathbf{H}}_{3,n+2}]^T$ and $\mathbf{Z}_m = [\mathbf{z}_{1,n}, \mathbf{z}_{2,n+1}, \mathbf{z}_{3,n+2}]^T$ are $(M \times 1)$ vectors, with $M = 3N$ and $M_{\mathrm{SYM}} = N_{\mathrm{SYM}}/3$. In the remainder, unless otherwise mentioned, we will not write the time index m for notational convenience. Our aim is to use the data model (3.16) to estimate the unknown channel frequency coefficients $\mathbf{H} = [H_1, ..., H_M]^T$. Let $\mathbf{h} = [h_1, ..., h_L]^T$ be the vector of UWB CIR over the first three subbands, that is, we have

$$\mathbf{H} = \mathbf{F}_{M,L}\,\mathbf{h},$$

where $\mathbf{F}_{M,L}$ is the truncated FFT matrix constructed from the $(M \times M)$ FFT matrix by keeping the first L columns, where L is the length of the CIR over a group of three subbands. Note that using directy (3.16) would require the estimation of M unknown parameters[2] at each iteration of the EM algorithm. The purpose of our method is to reduce the number of estimated parameters and to enhance at the same time the estimation accuracy by exploiting the sparsness property of UWB channels in the wavelet domain.

Thus, before going into the detail of our proposed channel estimation method, we propose to

[1] For the sake of notational brevity in equation (3.16), we have assumed that the TFC is equal to {1,2,3,...}.

[2] For non quasi-static channels, the number of estimated parameters is even greater than M.

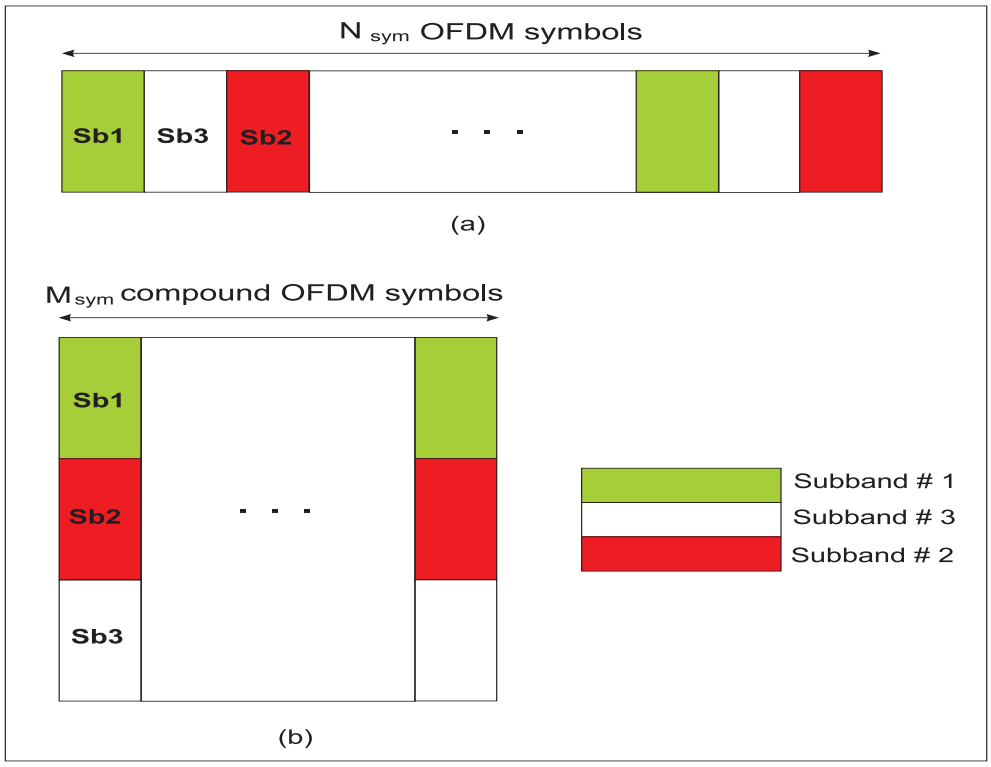

Figure 3.4: (a) MB-OFDM packet transmission corresponding to a TFC={1,3,2,...}; (b) Proposed arrangment of received OFDM symbols at the receiver.

have a closer look at the UWB CIR **h** and its wavelet domain representation.

3.4.2 UWB Channel Model in the Wavelet Domain

3.4.2.1 UWB Channel Model

Due to the very large bandwidth of UWB waveforms, a model different from that of classical narrowband channels should be considered to characterize the discrete arrivals of multipath rays. For narrowband systems, these multipath components would not be resolvable by the receiver when the system bandwidth is less than the coherence bandwidth of the channel. In contrast, the large bandwidth of UWB signals significantly increases the number of resolvable multipath components. In order to provide a model for system performance evaluation, the IEEE802.15.3a channel modeling task group analyzed various contributions describing UWB channel characteristics from measurements and finally adopted [65], a slightly modified version of the Saleh-Valenzuela model [66].

In this model, the rays are grouped into "clusters" and the discrete time CIR sampled at T_s is represented as

$$h_n^r = \sum_{c=1}^{C} \sum_{l=1}^{L} \alpha_{c,l}^r \, \delta(nT_s - T_c^r - \tau_{c,l}^r) \qquad n = 1, \ldots, L \tag{3.17}$$

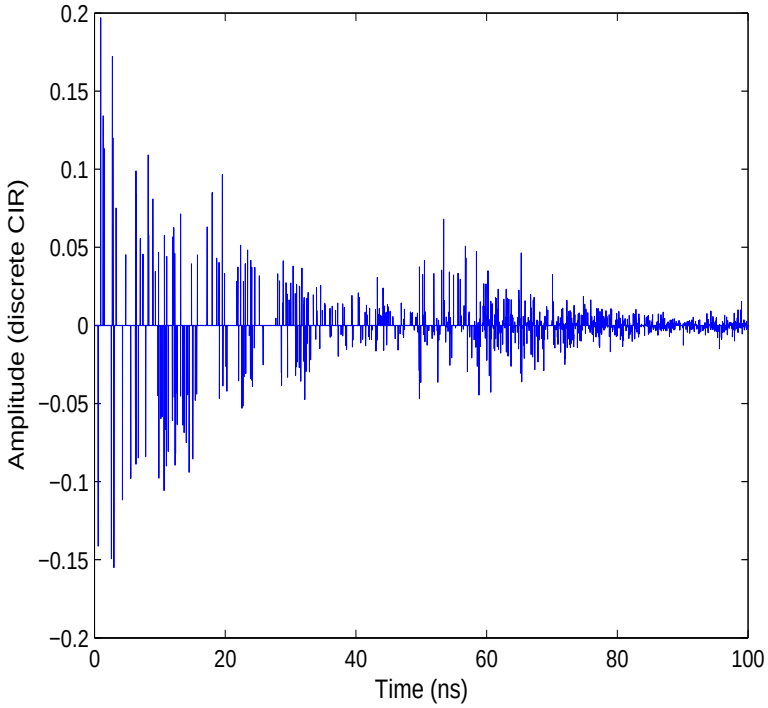

Figure 3.5: Example of a discrete UWB channel impulse response realization with tap spacing of 631 ps, equivalent to a bandwidth of 1.584 GHz, CM3 UWB channel model.

where $\delta(.)$ is the Dirac delta function, $\{\alpha_{c,l}^r\}$ are the multipath gain coefficients, r refering to the impulse response realization, $\{T_c^r\}$ is the delay of the c-th cluster for the r-th channel realization, and $\{\tau_{c,l}^r\}$ is the delay of the l-th multipath component within the c-th cluster for the r-th channel realization. Fig. 3.5 shows an example of such a modeled UWB CIR and highlights the challenges that the multipath model poses to UWB receivers. In particular, it is obvious that a significant amount of energy may exist in the multipath components. Therefore, the channel estimation part has to estimate a large number of coefficients in order to ensure an accurate channel acquisition.

3.4.2.2 Wavelet Representation of UWB Channels

In order to reduce the number of estimated channel coefficients at the receiver, we consider an orthogonal wavelet expansion of the unknown complex baseband channel vector $\mathbf{h} = [h_1, \ldots, h_L]^T$, with $J_{\max}$ level of decomposition as [130]

$$h_n = \sum_{j=1}^{J_{\max}} \sum_{k=1}^{2^j} g_{j,k}\, \psi_{j,k}[n] \qquad n = 1, \ldots, L \tag{3.18}$$

where the set of coefficients $\{g_{j,k}\} = \langle h_n, \psi_{j,k}[n] \rangle$ ($\langle ., . \rangle$ denotes scalar product) is the orthogonal discrete wavelet transform (ODWT) of $\mathbf{h}$ and $\{\psi_{j,k}[n]\}$ are the basis functions of the ODWT.

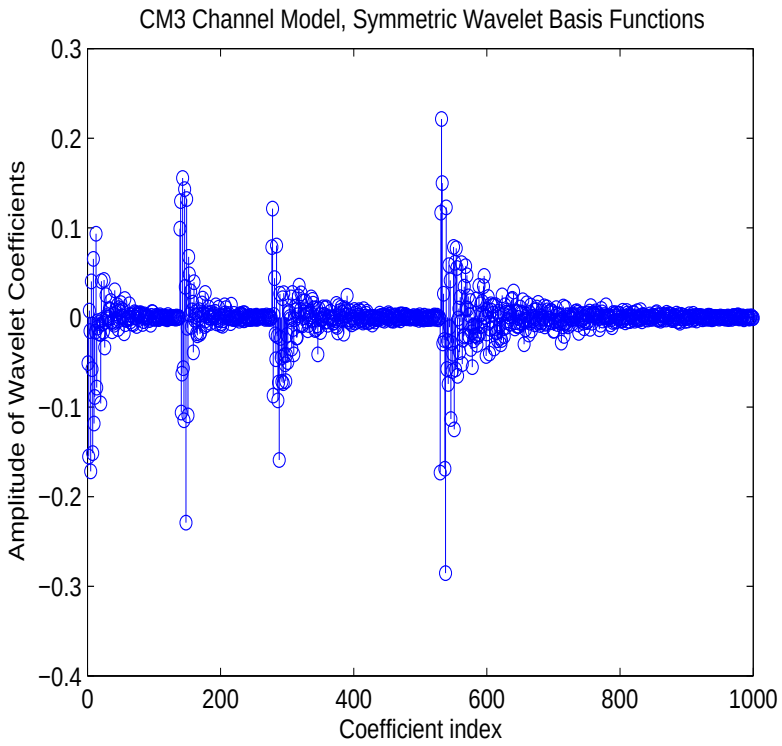

Figure 3.6: Illustration of the parsimonious representation of UWB channels in the wavelet domain, CM3 UWB channel model with 1.584 GHz of bandwidth, "Symmetric" wavelets with 3 levels of decomposition.

As shown in Fig. 3.6, the advantage of the wavelet representation is that a large set of signals have a parsimonious representation in wavelet series [67].

In the presence of AWGN, estimating the wavelet coefficients $\widetilde{g}_{j,k}$ can be written as

$$\widetilde{g}_{j,k} = g_{j,k} + \epsilon_{j,k} \quad j = 1, \ldots, J, \; k = 1, \ldots, 2^j, \tag{3.19}$$

where $\epsilon_{j,k} \sim \mathcal{CN}(0, \alpha^2)$ denotes the estimation error. If the channel has a sparse wavelet representation, one can assume that only a few large $\widetilde{g}_{j,k}$ really contain the most part of the channel energy while "insignificant" coefficients are attributed to the noise. The extraction of those "significant" coefficients can be naturally done by hard or soft thresholding [131]. However, the choice of the threshold value is crucial: an adequate choice of the threshold necessitates a large number of observations which are not always available in wireless communication applications.

In this work, we consider a Bayesian framework which involves a prior distribution of the unknown channel wavelet coefficients. This prior model is adopted so as to model the possibly sparseness of UWB CIR wavelet expansion which is especially important for large channel bandwidths [45]. Section 3.4.4 provides a specific prior distribution which has the nice property that, when used with a Bayesian estimation, it behaves as a hard "keep" or "kill" thresholding rule without any need to define an explicit threshold value. In the case of a sparse channel, this property has two interesting features: **i**) the precision of the estimator is improved due to the

adequacy of the prior distribution, and **ii**) "insignificant" wavelet coefficients are discarded from the estimation process, thus reducing the estimation computational load.

In our choice, the parsimonious characteristic of wavelet bases is imposed through the following prior model: wavelet coefficients are assumed to have a probability λ to be zero and a probability $1-\lambda$ to be Gaussian distributed as $\mathcal{CN}_{g_{j,k}}(0,\tau^2)$. This corresponds to an i.i.d. Bernoulli-Gaussian [67] prior model for the probability density of $g_{j,k}$ as

$$\pi(g_{j,k}) = \lambda\,\delta(g_{j,k}) + (1-\lambda)\,\mathcal{CN}_{g_{j,k}}(0,\tau^2) \tag{3.20}$$

for $j = 1,\ldots,J_{\max}$ and $k = 1,\ldots,2^j$. The parameters λ and τ (hyperparameters in the Bayesian terminology) are estimated from the observed data (see Section 3.4.4). In other words, we do not assume the percentage of null coefficients to be known.

In what follows, we make use of a compact notation for the wavelet transform, based on a $(L \times L)$ orthonormal wavelet transform matrix $\mathbf{W}$, where L is the length of the channel impulse response.

3.4.3 Wavelet Domain Problem Formulation

In order to take advantage of the wavelet based estimation, the channel impulse response $\mathbf{h}$ is expressed in terms of its orthogonal discrete wavelet coefficients as

$$\mathbf{h} = \mathbf{W}^{\dagger}\mathbf{g},$$

and the channel frequency response as

$$\mathbf{H} = \mathbf{F}_{M,L}\mathbf{h} = \mathbf{F}_{M,L}\mathbf{W}^{\dagger}\mathbf{g},$$

where $\mathbf{g}$ is a $(L \times 1)$ vector of the CIR wavelet coefficients.
The frequency domain observation model (3.16) can be rewritten as

$$\mathbf{Y} = \mathcal{D}_{\mathbf{S}}\,\mathbf{T}\,\mathbf{g} + \mathbf{Z} \tag{3.21}$$

where $\mathbf{T} = \mathbf{F}_{M,L}\mathbf{W}^{\dagger}$.

In this model, although the channel is practically used (by the transmitter) by slices of 528 MHz bandwidth (corresponding to a single subband), on the receiver side, three received OFDM symbols are grouped for estimating the wavelet coefficients of the CIR, taken over all three subbands (1.584 GHz bandwidth, see Fig.3.4). This is motivated by the fact that estimating the channel over a wider bandwidth leads to a sparser representation in the wavelet domain. Moreover, this approach simplifies the receiver architecture since there is no need to change the central frequency for down converting different subbands.

3.4.4 The EM-MAP Algorithm for Wavelet Domain Channel Estimation

The EM algorithm proposed in this section is able to integrate the advantages of wavelet based estimation via the prior chosen for channel wavelet coefficients. We also emphasize that the MAP estimator corresponds to a thresholding procedure which is used to reduce the number of estimated coefficients at each iteration of the EM algorithm.

3.4.4.1 An Equivalent Model and the EM Principle

Most of wavelet based estimation algorithms, rely on an observation model in which the unknown wavelet coefficients are corrupted by AWGN. Since the model (3.21) does not provide such a framework, our first step consists in enforcing this property. In order to do so, the AWGN in (3.21) is split into two independent Gaussian terms as

$$\mathbf{Z} = \mathcal{D}_{\mathbf{S}}\mathbf{Z}_1 + \mathbf{Z}_2 \tag{3.22}$$

where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are $(M\times 1)$ independent Gaussian noise vectors such that $p(\mathbf{Z}_1) = \mathcal{CN}(\mathbf{0}, \alpha^2\mathbb{I}_M)$ and $p(\mathbf{Z}_2) = \mathcal{CN}(\mathbf{0}, \sigma^2\mathbb{I}_M - \alpha^2\mathcal{D}_{\mathbf{S}}\mathcal{D}_{\mathbf{S}}^{\dagger})$. Since we are using power normalized QPSK symbols, $\mathcal{D}_{\mathbf{S}}\mathcal{D}_{\mathbf{S}}^{\dagger} = \mathbb{I}_M$ and the covariance matrix $\mathbf{\Sigma}_{\mathbf{Z}_2}$ of $\mathbf{Z}_2$ reduces to $\mathbf{\Sigma}_{\mathbf{Z}_2} = (\sigma^2 - \alpha^2)\mathbb{I}_M$. We define the positive design parameter $\rho \triangleq \alpha^2/\sigma^2$, $(0 \leq \rho < 1)$ as the proportion of noise that is assigned to $\mathbf{Z}_2$. Note that setting $\rho = 0$ leads to $\mathbf{Z}_1 = \mathbf{0}$ and is equivalent to working with the initial model (3.21). However, for $0 < \rho < 1$, the above noise decomposition allows the introduction of a hidden channel vector $\widetilde{\mathbf{H}}$ defined as

$$\begin{cases} \widetilde{\mathbf{H}} &= \mathbf{T}\,\mathbf{g} + \mathbf{Z}_1 \\ \mathbf{Y} &= \mathcal{D}_{\mathbf{S}}\,\widetilde{\mathbf{H}} + \mathbf{Z}_2. \end{cases} \tag{3.23}$$

This procedure implicitly assigns part of the noise to the wavelet coefficients, and the rest to the measurement. This introduces a hidden vector $\widetilde{\mathbf{H}}$ which provides us with a direct relation [3] between the true and the estimated wavelet coefficients corrupted by an AWGN, even if the two-stage observation model (3.23) is equivalent to (3.21). However, the difference with a standard denoising problem like (3.19) is that $\mathbf{S}$ and $\widetilde{\mathbf{H}}$ are unknown. Hence, the observation model has missing data and hidden variables and the MAP solution of $\mathbf{g}$ has no closed form. In such situations, the EM algorithm [7] is often used to maximize the expectation of the posterior distribution over all possible missing and hidden variables. Obviously, the value assigned arbitrarily to parameter ρ will influence the performance of the estimate, and will have to be tuned.

Let $\mathbf{X} = \{\mathbf{Y}, \mathbf{S}, \widetilde{\mathbf{H}}\}$ be the *complete-data set* in the EM algorithm terminology. Note that the observation set $\mathbf{Y}$ determines only a subset of the space $\mathscr{X}$ of which $\mathbf{X}$ is an outcome.

[3] Up to a left-multiplication by the matrix $\mathbf{T}^{\dagger}$.

We search $\mathbf{g}$ that maximizes $\log p(\mathbf{g}|\mathbf{X})$. After initialization by a short pilot sequence at the beginning of the frame, the EM algorithm alternates between the following two steps (until some stopping criterion) to produce a sequence of estimates $\{\mathbf{g}^{(t)},\, t = 0, 1, \ldots, t_{\max}\}$.

- **Expectation Step** (E-step): The conditional expectation of the complete log-likelihood given the observed vector and the current estimate $\mathbf{g}^{(t)}$ is computed. This quantity is called the *auxiliary* or Q-function

$$Q\big(\mathbf{g}, \mathbf{g}^{(t)}\big) = \mathbb{E}_{\mathbf{S},\widetilde{\mathbf{H}}}\Big[\log p(\mathbf{Y}, \mathbf{S}, \widetilde{\mathbf{H}}|\mathbf{g})\Big|\mathbf{Y}, \mathbf{g}^{(t)}\Big] \tag{3.24}$$

- **Maximization Step** (M-step): The estimated parameter is updated according to

$$\mathbf{g}^{(t+1)} = \arg\max_{\mathbf{g}} \Big\{Q\big(\mathbf{g}, \mathbf{g}^{(t)}\big) + \log \pi(\mathbf{g})\Big\} \tag{3.25}$$

where $\pi(\mathbf{g})$ is the *prior* distribution for the wavelet coefficients introduced in section 3.4.2, which ensures a certain percentage of coefficients to be set to zero. When applied to (3.23), each step can be written as follows.

E-step: Computation of the Q-function

The complete likelihood is

$$p(\mathbf{Y}, \mathbf{S}, \widetilde{\mathbf{H}}|\mathbf{g}) = p(\mathbf{Y}|\mathbf{S}, \widetilde{\mathbf{H}}, \mathbf{g})\, p(\mathbf{S}|\widetilde{\mathbf{H}}, \mathbf{g})\, p(\widetilde{\mathbf{H}}|\mathbf{g}).$$

According to (3.23), $p(\mathbf{Y}|\mathbf{S}, \widetilde{\mathbf{H}}, \mathbf{g})$ is a pdf not depending on $\mathbf{g}$. Furthermore, $\mathbf{S}$ which results from coding and interleaving of bit sequence is independent of $\widetilde{\mathbf{H}}$ and $\mathbf{g}$. Since $\mathbf{Z}_1$ is a complex white Gaussian noise, the complete log-likelihood can be simplified to

$$\begin{aligned}
\log p(\mathbf{Y}, \mathbf{S}, \widetilde{\mathbf{H}}|\mathbf{g}) &= \log\big[p(\mathbf{Y}|\mathbf{S}, \widetilde{\mathbf{H}})\; p(\mathbf{S})\; p(\widetilde{\mathbf{H}}|\mathbf{g})\big] \\
&= \log p(\widetilde{\mathbf{H}}|\mathbf{g}) \;+\; \text{cst.1} \\
&= -\frac{\|\,\widetilde{\mathbf{H}} - \mathbf{T}\mathbf{g}\,\|^2}{\alpha^2} \;+\; \text{cst.2} \\
&= -\frac{\mathbf{g}^\dagger\mathbf{T}^\dagger\mathbf{T}\mathbf{g} - \mathbf{g}^\dagger\mathbf{T}^\dagger\widetilde{\mathbf{H}} - \widetilde{\mathbf{H}}^\dagger\mathbf{T}\mathbf{g}}{\alpha^2} \;+\; \text{cst.3}
\end{aligned} \tag{3.26}$$

where cst.1, cst.2 and cst.3 are constant terms that do not depend on $\mathbf{g}$.
According to (3.24) we have

$$\begin{aligned}
Q\big(\mathbf{g}, \mathbf{g}^{(t)}\big) &= \mathbb{E}_{\mathbf{S},\widetilde{\mathbf{H}}}\Big[-\frac{\mathbf{g}^\dagger\mathbf{T}^\dagger\mathbf{T}\mathbf{g} - 2\,\mathbb{R}e\big(\mathbf{g}^\dagger\mathbf{T}^\dagger\widetilde{\mathbf{H}}\big)}{\alpha^2} + \text{cst.}\,\Big|\mathbf{Y}, \mathbf{g}^{(t)}\Big] \\
&= -\frac{\mathbf{g}^\dagger\mathbf{T}^\dagger\mathbf{T}\mathbf{g} - 2\,\mathbb{R}e\big(\mathbf{g}^\dagger\mathbf{T}^\dagger\,\mathbb{E}_{\mathbf{S},\widetilde{\mathbf{H}}}[\widetilde{\mathbf{H}}|\mathbf{Y}, \mathbf{g}^{(t)}]\big)}{\alpha^2} + \text{cst.} \\
&= -\frac{\|\,\langle\widetilde{\mathbf{H}}^{(t)}\rangle - \mathbf{T}\mathbf{g}\,\|^2}{\alpha^2} + \text{cst.}
\end{aligned} \tag{3.27}$$

where $\langle\widetilde{\mathbf{H}}^{(t)}\rangle \triangleq \mathbb{E}_{\mathbf{S},\widetilde{\mathbf{H}}}[\widetilde{\mathbf{H}}\big|\mathbf{Y},\mathbf{g}^{(t)}]$ and cst. represents a constant term.
From (3.27), it is obvious that the E-step involves the computation of $\langle\widetilde{\mathbf{H}}^{(t)}\rangle$, as follows:

$$\begin{aligned}\langle\widetilde{\mathbf{H}}^{(t)}\rangle &= \sum_{\mathbf{S}\in\mathscr{C}}\int_{\widetilde{\mathbf{H}}\in\mathscr{H}} \widetilde{\mathbf{H}} \quad p(\widetilde{\mathbf{H}},\mathbf{S}\big|\mathbf{Y},\mathbf{g}^{(t)})\,\mathrm{d}\widetilde{\mathbf{H}} \\ &= \sum_{\mathbf{S}\in\mathscr{C}}\Big(\int_{\widetilde{\mathbf{H}}\in\mathscr{H}} \widetilde{\mathbf{H}} \quad p(\widetilde{\mathbf{H}}\big|\mathbf{Y},\mathbf{g}^{(t)})\,\mathrm{d}\widetilde{\mathbf{H}}\Big)p(\mathbf{S}\big|\mathbf{Y},\mathbf{g}^{(t)}) \end{aligned} \tag{3.28}$$

where the last equation results from the independence between $\mathbf{S}$ and $\widetilde{\mathbf{H}}$ belonging respectively to the sets $\mathscr{C}$ and $\mathscr{H}$. Note that each entry of $\mathbf{S}$ takes one (unknown) discrete value inside the QPSK constellation whereas components of $\widetilde{\mathbf{H}}$ are continuous variables.

In order to evaluate $\langle\widetilde{\mathbf{H}}^{(t)}\rangle$ in (3.28), we first have to evaluate the conditional mean $\boldsymbol{\mu}_{\widetilde{\mathbf{H}}}^{(t)}$ of $\widetilde{\mathbf{H}}$ as

$$\boldsymbol{\mu}_{\widetilde{\mathbf{H}}}^{(t)} = \mathbb{E}_{\widetilde{\mathbf{H}}}[\widetilde{\mathbf{H}}\big|\mathbf{Y},\mathbf{g}^{(t)}] = \int_{\widetilde{\mathbf{H}}\in\mathscr{H}} \widetilde{\mathbf{H}} \quad p(\widetilde{\mathbf{H}}|\mathbf{Y},\mathbf{g}^{(t)})\,\mathrm{d}\widetilde{\mathbf{H}}. \tag{3.29}$$

In order to evaluate the latter expectation, we write $p(\widetilde{\mathbf{H}}|\mathbf{Y},\mathbf{g}^{(t)}) \propto p(\mathbf{Y}|\widetilde{\mathbf{H}})\; p(\widetilde{\mathbf{H}}|\mathbf{g}^{(t)})$. Since both $p(\mathbf{Y}|\widetilde{\mathbf{H}})$ and $p(\widetilde{\mathbf{H}}|\mathbf{g}^{(t)})$ are Gaussian densities, it is well known that their product remains Gaussian. We have to evaluate the mean of this Gaussian density. To this end we use the following known result in the form of a lemma (see, e.g., [132]).

Lemma 3.4.1. *Let $\mathcal{CN}(\mathbf{m}_1,\boldsymbol{\Sigma}_1)$ and $\mathcal{CN}(\mathbf{m}_2,\boldsymbol{\Sigma}_2)$ be two complex Gaussian densities of $\mathbf{x}$, then*

$$\mathcal{CN}(\mathbf{m}_1,\boldsymbol{\Sigma}_1)\,.\,\mathcal{CN}(\mathbf{m}_2,\boldsymbol{\Sigma}_2) = K_c\,\mathcal{CN}(\mathbf{m}_c,\boldsymbol{\Sigma}_c),$$

where K_c is a normalization factor and

$$\begin{aligned}\mathbf{m}_c &= \big(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1}\big)^{-1}\big(\boldsymbol{\Sigma}_1^{-1}\,\mathbf{m}_1+\boldsymbol{\Sigma}_2^{-1}\mathbf{m}_2\big),\\ \boldsymbol{\Sigma}_c &= \big(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1}\big)^{-1}.\end{aligned}$$

Noting from (3.23) that $\mathbf{Y}|\widetilde{\mathbf{H}} \sim \mathcal{CN}\big(\mathcal{D}_{\mathbf{S}}\widetilde{\mathbf{H}},(\sigma^2-\alpha^2)\mathbb{I}_M\big)$ and $\widetilde{\mathbf{H}}|\mathbf{g}^{(t)} \sim \mathcal{CN}\big(\mathbf{T}\mathbf{g}^{(t)},\alpha^2\mathbb{I}_M\big)$, and using Lemma 3.4.1, we get

$$\boldsymbol{\mu}_{\widetilde{\mathbf{H}}}^{(t)} = \mathbf{T}\mathbf{g}^{(t)} + \rho\,\mathcal{D}_{\mathbf{S}}^{\dagger}\Big(\mathbf{Y}-\mathcal{D}_{\mathbf{S}}\mathbf{T}\mathbf{g}^{(t)}\Big). \tag{3.30}$$

By introducing (3.30) in (3.28) we obtain

$$\langle\widetilde{\mathbf{H}}^{(t)}\rangle = (1-\rho)\,\mathbf{T}\mathbf{g}^{(t)} + \rho\,\overline{\mathcal{D}}_{\mathbf{S}}^{\dagger}\mathbf{Y} \tag{3.31}$$

where $\overline{\mathcal{D}}_{\mathbf{S}} = \sum_{\mathbf{s}\in\mathscr{C}}\mathcal{D}_{\mathbf{S}}\,p(\mathbf{S}|\mathbf{Y},\mathbf{g}^{(t)})$. Here, we assume that a part of the receiver called soft-input soft-output (SISO) decoder, is able to provide the probabilities $p(\mathbf{S}|\mathbf{Y},\mathbf{g}^{(t)})$ required in (3.31) (see Subsection 3.4.5 for more details). The E-step is then completed by inserting $\langle\widetilde{\mathbf{H}}^{(t)}\rangle$ into $Q(\mathbf{g},\mathbf{g}^{(t)})$ of equation (3.27).

M-step: Wavelet Based MAP Estimation

In this step the estimate of the parameter $\mathbf{g}$ is updated as given in (3.25) where $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ is given by (3.27). We have

$$\mathbf{g}^{(t+1)} = \arg\max_{\mathbf{g}} \left\{ -\frac{\| \langle \widetilde{\mathbf{H}}^{(t)} \rangle - \mathbf{T}\mathbf{g} \|^2}{\alpha^2} + \log \pi(\mathbf{g}) \right\}. \tag{3.32}$$

It is shown in Appendix A that the expression (3.32) is equivalent to

$$\mathbf{g}^{(t+1)} = \arg\max_{\mathbf{g}} \left\{ -\frac{\| \widetilde{\mathbf{g}}^{(t)} - \mathbf{g} \|^2}{\alpha^2} + \log \pi(\mathbf{g}) \right\} \tag{3.33}$$

where

$$\boxed{\widetilde{\mathbf{g}}^{(t)} = \mathbf{T}^{\dagger} \langle \widetilde{\mathbf{H}}^{(t)} \rangle = (1-\rho)\, \mathbf{g}^{(t)} + \rho\, (\overline{\mathcal{D}}_{\mathbf{s}} \mathbf{T})^{\dagger} \mathbf{Y}.} \tag{3.34}$$

In fact, $\mathbf{g}^{(t+1)}$ in (3.33) is no more than the MAP estimate of $\mathbf{g}$ from the observation model:

$$\widetilde{\mathbf{g}}^{(t)} = \mathbf{g} + \mathbf{Z}'_1 \tag{3.35}$$

where $\mathbf{Z}'_1 \sim \mathcal{CN}(\mathbf{0}, \alpha^2 \mathbf{I}_L)$.

Note that equation (3.35) is very important since it shows that the initial estimation problem reduces to an observation model which involves a direct relation between the unknown wavelet coefficient $\mathbf{g}$ and its estimate $\widetilde{\mathbf{g}}^{(t)}$, and this direct relation is corrupted by an AWGN (similar to equation (3.19)). This is the reason of using the two level observation model in (3.23). Starting from equation (3.35), our channel estimation problem can be viewed as a standard wavelet domain denoising problem. A rich litterature exists on the latter topic especially in the image processing community. In what follows, we derive the update formula of our wavelet domain channel estimator.

From the Bayes theorem, the posterior distribution of $\mathbf{g}$ is given by

$$p\left(\mathbf{g}|\widetilde{\mathbf{g}}^{(t)}\right) \;\propto\; p\left(\widetilde{\mathbf{g}}^{(t)}|\mathbf{g}\right) \pi\left(\mathbf{g}\right) \tag{3.36}$$

where from (3.35), $p(\widetilde{\mathbf{g}}^{(t)}|\mathbf{g})$ is the Gaussian likelihood, $\widetilde{\mathbf{g}}^{(t)} \sim \mathcal{CN}(\mathbf{g}, \alpha^2 \mathbf{I}_L)$. In this approach, $\pi\left(\mathbf{g}\right)$ is a prior distribution, chosen for the wavelet coefficients $\mathbf{g}$ of the unknown CIR.

- **Uniform Prior Model:** At first, we consider a *non-informative* uniform i.i.d. prior model for the wavelet coefficients. Obviously, in this case, the MAP estimate coincides with the ML estimate and is given by

$$\boxed{g_j^{(t+1)} = \widetilde{g}_j^{(t)}} \quad \text{for } j = 1, \ldots, L \tag{3.37}$$

  where $\widetilde{g}_j^{(t+1)}$ is calculated in (3.34). The above estimate will be used later to study the behavior of the MAP approach in the case where the channel fails to satisfy sparse wavelet domain assumptions.

- **Bernoulli-Gaussian Prior Model:** As previously discussed in section 3.4.2, when the channel is sparse, we adopt the i.i.d Bernoulli-Gaussian model described by

$$\pi\left(g_j\right) = \lambda\, \delta(g_j) + (1-\lambda)\, \mathcal{CN}_{g_j}\left(0, \tau^2\right) \tag{3.38}$$

for $j = 1, \ldots, L$, which allows us to model a sparsness property of UWB channels in the wavelet domain. In order to deal with that particular model, we introduce an additional state variable (or indicator) $\beta_j \in \{0, 1\}$ such that we can express this prior conditionally as

$$\begin{cases} (g_j|\beta_j = 0) & = & \delta(g_j) & \text{with probability } \lambda, \\ & & & \\ (g_j|\beta_j = 1) & \sim & \mathcal{CN}_{g_j}\left(0, \tau^2\right) & \text{with probability } 1-\lambda. \end{cases} \tag{3.39}$$

This prior model, conditionally on the state variable β_j, leads to a Gaussian posterior for g_j which makes the estimation explicit; from the direct observation model $\widetilde{g}_j^{(t)} = g_j + Z'_{1,j}$ (equation (3.35)), we can express these posterior probabilities of β_j as (see Appendix A)

$$\begin{array}{rcl} p\left(\beta_j = 0|\widetilde{g}_j^{(t)}\right) & = & \lambda\, \mathcal{N}\left(0, \alpha^2\right)/c \\ p\left(\beta_j = 1|\widetilde{g}_j^{(t)}\right) & = & (1-\lambda)\, \mathcal{N}\left(0, \alpha^2 + \tau^2\right)/c \end{array} \tag{3.40}$$

where the parameter $c = \lambda \mathcal{N}\left(0, \alpha^2\right) + (1-\lambda)\, \mathcal{N}\left(0, \alpha^2 + \tau^2\right)$. From this set of equations, we notice that the indicator variable β_j allows us to discriminate between the noise coefficients (for $\beta_j = 0$) and the effective channel wavelet coefficients (for $\beta_j = 1$), eventually corrupted by noise. The indicator variables β_j are estimated, in the MAP sense by

$$\beta_j^{(t+1)} = \begin{cases} 0, & \text{if } \; p\left(\beta_j = 0|\widetilde{g}_j^{(t)}\right) \geq 0.5 \\ & \\ 1, & \text{elsewhere.} \end{cases} \tag{3.41}$$

Therefore, the MAP estimates of the channel wavelet coefficients are obtained by a simple denoising/thresholding rule as (see Appendix A)

$$\boxed{g_j^{(t+1)} = \begin{cases} 0, & \text{if } \; \beta_j^{(t+1)} = 0 \\ \dfrac{\tau^2}{\alpha^2 + \tau^2}\, \widetilde{g}_j^{(t)}, & \text{if } \; \beta_j^{(t+1)} = 1. \end{cases}} \tag{3.42}$$

3.4.4.2 Updating the Prior Parameters τ and λ

The prior parameters τ and λ stand respectively for the (significant)-wavelet coefficients (variance) power and insignificant coefficients probability. The update rules of these two parameters are *maximum a posteriori* based rules, derived from assigning conjugate priors to these parameters [133]. A Chi-square (χ^2) prior with parameters r_0 and η_0 is chosen for the inverse square

value of τ, $\xi = \tau^{-2}$

$$\pi(\xi|r_0, \eta_0) \propto \xi^{\frac{r_0-2}{2}} e^{-\xi\eta_0/2}, \tag{3.43}$$

while a Dirichlet prior with parameter u_0 is chosen for the probability λ

$$\pi(\lambda|u_0) \propto \lambda^{u_0-1}(1-\lambda)^{-u_0}. \tag{3.44}$$

We point out that these parametric priors tend respectively to a *non-informative Jeffrey's* [134] prior for ξ as $\eta_0 = 0$ and $r_0 = 0$, and to a *uniform prior* for λ as $u_0 = 1/2$. From the Bayes rule, the posterior distributions of these two parameters are respectively given by

$$\begin{aligned} p\left(\xi|g_j, r_0, \eta_0\right) &\propto \xi^{\frac{r-2}{2}} e^{-\xi\eta/2} \\ p\left(\lambda|g_j, u_0\right) &\propto \lambda^{u-1}(1-\lambda)^{-u} \end{aligned} \tag{3.45}$$

where

$$\begin{aligned} \eta &= \eta_0 + \sum_{\beta_j=1} g_j^2, \quad r = r_0 + L - \widetilde{L}, \\ u &= u_0 + \widetilde{L} \end{aligned} \tag{3.46}$$

with

$$\widetilde{L} = \text{Card.}\{j :\ \beta_j = 0\}. \tag{3.47}$$

These distributions have the advantage to be tractable and the maximization steps are straight forward. With simple manipulations, these maximization yield

$$\begin{aligned} \hat{\tau}^2 &= \frac{1}{\hat{\xi}} = \frac{\eta}{r-2}, \\ \hat{\lambda} &= \frac{u-1}{L-1}. \end{aligned} \tag{3.48}$$

Note also that when the channel does not satisfy the sparsness property, these update equations still hold: while the sparsity is modeled by the Bernoulli-Gaussian equation (3.38), a non sparse channel is translated by a value of λ that tends to 0. This is easily verified since the sparser the channel, the lower the value of $\widetilde{L}$ (which represents the number of "null" coefficients), and in this case the value of u may go down to u_0. The value of the probability λ in this case tends to

$$\hat{\lambda} = \frac{u_0 - 1}{L-1}.$$

In addition, for such a non sparse channel, τ^2 which represents the power of significant wavelet coefficients tends to

$$\hat{\tau}^2 = \frac{\eta_0 + \sum_{\beta_j=1} g_j^2}{r_0 + L - 2} \underset{L \gg r_0 - 2}{\overset{\eta_0 \to 0}{\longrightarrow}} \frac{\sum_j g_j^2}{L}$$

which is nothing but the ML estimate of τ^2.

3.4.4.3 Reduction of the Number of Estimated Parameters

The thresholding procedure derived in this section, provides an automatic framework for reducing the number of estimated coefficients. This can be achieved by discarding at each iteration, the elements of $\mathbf{g}^{(t+1)}$ that are replaced by zero in (3.42). The underlying assumption is as follows: whenever the estimator assimilates an unknown wavelet coefficient to noise (replaces it by zero), this coefficient will always be considered as noise, hence it will not be estimated in future iterations. We verified by simulations that incorporating this scheme into the EM algorithm reduces the number of estimated parameters without any significant performance degradation. This operation is shown on Fig. 3.7 and can be modeled as

$$\mathbf{g}_{\mathrm{tr}}^{(t+1)} = \mathbf{\Theta}\big(\,\mathbf{g}^{(t+1)}\,\big), \quad \mathbf{T}_{\mathrm{tr}} = \Xi\big(\,\mathbf{T}\,\big) \tag{3.49}$$

where the truncation operator $\mathbf{\Theta}(.)$ gathers in $\mathbf{g}_{\mathrm{tr}}^{(t+1)}$ the components of $\mathbf{g}^{(t+1)}$ that must be kept and the operator $\Xi(.)$ constructs $\mathbf{T}_{\mathrm{tr}}$ from $\mathbf{T}$ by keeping the rows corresponding to the kept indexes. During the first iteration ($t = 0$), the algorithm does not perform any truncation and the EM algorithm estimates all the coefficients. However, after each M-step, the number of unknown parameters to be estimated in the next iteration is reduced according to (3.49) by using $\mathbf{g}_{\mathrm{tr}}^{(t+1)}$ and $\mathbf{T}_{\mathrm{tr}}$ in the updating formula of the E-step (3.31).

3.4.4.4 Extension to Unknown Noise Variance

Up to now, we have assumed that the noise variances α^2 and σ^2 in (3.23) are known at the receiver. However, in practical systems, the noise level is unknown and must be estimated from the observations. We now present an extension of the proposed method where these variances and the channel wavelet coefficients are estimated together.

To this end, we use the indicator variables $\beta_{j,m}$ defined in the M-step section for discriminating between the wavelet coefficients and the noise samples, where the index m denotes the m-th (compound) OFDM symbol inside the current frame. Let us define the matrix $\mathbf{G}$ from (3.35) as $\mathbf{G} = \big[\widetilde{\mathbf{g}}_1, \ldots, \widetilde{\mathbf{g}}_{M_{\mathrm{SYM}}}\big]$ and the set Λ from (3.41) as

$$\Lambda = \{\,(j,m) :\, \beta_{j,m} = 0\} \tag{3.50}$$

where the iteration index has been omitted for notational brevity. It is clear that, using Λ at each iteration, we can derive a subset $\mathcal{V} \subset \mathbf{G}$ as

$$\mathcal{V} = \mathbf{G}(\Lambda) = \{\mathbf{G}(j,m) : \beta_{j,m} = 0\} \tag{3.51}$$

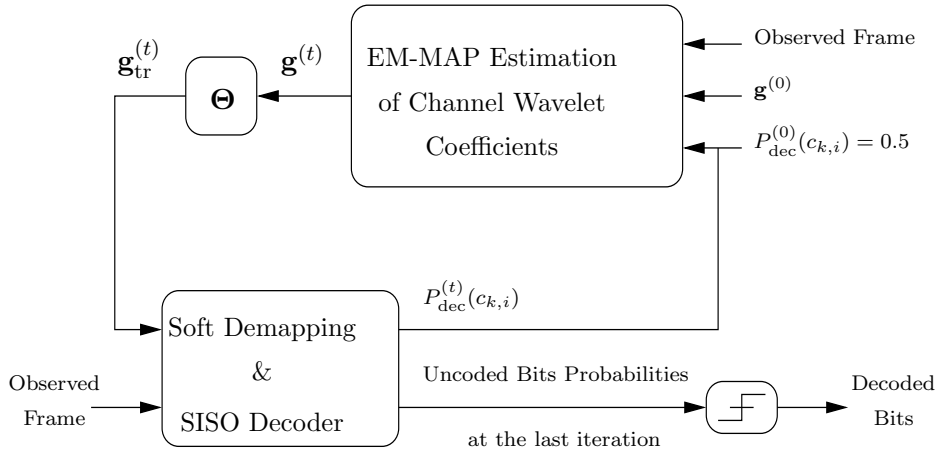

Figure 3.7: EM-MAP channel estimation combined with the decoding process.

which contains samples of the noise vector $\mathbf{Z'}_1 \sim \mathcal{CN}(\mathbf{0}, \alpha^2 \mathbf{I}_L)$ according to the model (3.35). At iteration t, the variance α^2 may be empirically estimated as

$$(\hat{\alpha}^2)^{(t)} = \frac{1}{\widetilde{L}\, M_{\text{SYM}}} \sum_{l=1}^{M_{\text{SYM}}} \sum_{j} |\mathcal{V}(j,m)|^2, \tag{3.52}$$

where $\widetilde{L}$ is given in (3.47).

According to the definition of the parameter ρ in (3.23), the variance σ^2 is obtained as

$$(\hat{\sigma}^2)^{(t)} = \frac{(\hat{\alpha}^2)^{(t)}}{\rho}. \tag{3.53}$$

3.4.5 Decoding Method and Implementation Issues

3.4.5.1 Iterative Demapping and Decoding

At the receiver, we perform MAP symbol detection and channel decoding in an iterative manner. The block diagram of the receiver is shown in Fig. 3.7. Besides the channel estimation part, the rest of the receiver principally consists of the combination of two sub-blocks that exchange soft informations with each other. The first sub-block, refered to as soft demapper (also called detector), produces bit metrics (probabilities) from the input symbols and the second one is a SISO decoder. Each sub-block can take advantage of the soft information provided by the other sub-block as an additional information. Here, SISO decoding is performed using the well known forward-backward algorithm [11]. We present in the following, the formulation of the MAP detector part, assuming that at iteration t, the receiver has an estimate $\widehat{\mathbf{H}}^{(t)}$ of the channel (given by $\widehat{\mathbf{H}}^{(t)} = \mathbf{T}_{\text{tr}}\, \mathbf{g}_{\text{tr}}^{(t)}$) and of the noise variance $(\hat{\sigma}^2)^{(t)}$. Moreover, we consider the observation model $\mathbf{Y} = \mathcal{D}_{\mathbf{S}}\mathbf{H} + \mathbf{Z}$ of equation (3.16).

Let $c_{k,i}$ $(i = 1, \ldots, B)$ be the i-th[4] coded and interleaved bit corresponding to the k-th constellation symbol S_k and let Y_k be the corresponding received symbol $(k = 1, \ldots, MM_{\mathrm{SYM}})$. We denote by $L(c_{k,i})$ the log-likelihood ratio (LLR) of the bit $c_{k,i}$ at the output of the detector. Conditioned on the channel coefficient estimated at the t-th iteration $\widehat{H}_k^{(t)}$, $L(c_{k,i})$ is given by

$$L(c_{k,i}) = \log \frac{P_{\mathrm{dem}}\big(c_{k,i} = 1 \big|\, Y_k, \widehat{H}_k^{(t)}\big)}{P_{\mathrm{dem}}\big(c_{k,i} = 0 \big|\, Y_k, \widehat{H}_k^{(t)}\big)}, \tag{3.54}$$

where $P_{\mathrm{dem}}(c_{k,i}\big|Y_k, \widehat{H}_k^{(t)})$ is the probability of transmission of $c_{k,i}$ at the demapper output. Let $\mathcal{S}$ be the set of all possibly-transmitted symbols corresponding to S_k. We partition $\mathcal{S}$ into two sets $\mathcal{S}_0^i$ and $\mathcal{S}_1^i$, for which the i-th bit of S_k equals "0" or "1", respectively. We have

$$L(c_{k,i}) = \log \frac{\sum\limits_{S_k \in \mathcal{S}_1^i} e^{-\mathcal{D}_{\mathrm{ML}}\left(S_k, Y_k, \widehat{H}_k^{(t)}\right)}\, P_1\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)}{\sum\limits_{S_k \in \mathcal{S}_0^i} e^{-\mathcal{D}_{\mathrm{ML}}\left(S_k, Y_k, \widehat{H}_k^{(t)}\right)}\, P_0\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)}, \tag{3.55}$$

where $P_0\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)$ and $P_1\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)$ denote the probability that S_k belongs to $\mathcal{S}_0^i$ and $\mathcal{S}_1^i$, respectively, and $\mathcal{D}_{\mathrm{ML}}\big(S_k, Y_k, \widehat{H}_k^{(t)}\big) = |Y_k - \widehat{H}_k^{(t)} S_k|^2/(\hat{\sigma}^2)^{(t)}$.

Actually, according to (3.28), we have to use the information on transmitted symbols, obtained from the SISO decoder through the probability $P_m\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)$ $(m \in \{0, 1\})$, to update the channel estimate at each iteration. Furthermore, the soft demapper requires an estimate of the channel in order to provide the probability of encoded bits (see (3.55)). Hence, the proposed semi-blind channel estimation algorithm is naturally combined with the process of data decoding. The probability $P_m\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)$ involved in (3.55), is calculated using the *a posteriori* probabilities provided by the SISO decoder at the end of the t-th iteration as

$$P_m\big(S_k|Y_k, \widehat{H}_k^{(t)}\big) = \prod_{\substack{j=1 \\ j \neq i}}^{B} P_{\mathrm{dec}}^m(c_{k,j}) \quad m \in \{0, 1\}, \tag{3.56}$$

where $P_{\mathrm{dec}}^0(c_{k,j})$ and $P_{\mathrm{dec}}^1(c_{k,j})$ are *prior* probabilities coming from the SISO decoder.

Note that in (3.56), the a priori probability of the bit $c_{k,i}$ itself has been excluded, so as to let the exchange of *extrinsic* information between the channel decoder and the soft detector [135]. Also, note that this term assumes independent coded bits $c_{k,j}$, which holds for large size random interleaving. At the first iteration, where no a priori information is available on bits $c_{k,j}$, the probabilities $P_{\mathrm{dec}}^0(c_{k,j})$ and $P_{\mathrm{dec}}^1(c_{k,j})$ are set to 0.5.

3.4.5.2 Global Procedure for Joint Channel Estimation and Decoding

There are several possible ways to practically implement a joint channel estimation and decoding receiver. In fact, inside each EM iteration t, the receiver should perform several decoding

[4]Here $B = 2$ since we are using QPSK.

iterations, keeping the channel estimate $\widehat{\mathbf{H}}^{(t)}$. For complexity issues, the considered receiver performs only one pass through the decoder inside each EM iteration. The main steps of the iterative MAP channel parameter estimation are summarized as follows.

- **Initialization** $(t = 0)$
  - Set all probabilities of coded bits $P^m_{\text{dec}}(c_{k,i})$ to 0.5 and derive $P_m\big(S_k|Y_k, \widehat{H}_k^{(0)}\big)$ for all k according to (3.56) and then derive $p(\mathbf{S}|\mathbf{Y}, \mathbf{H}^{(0)}) = p(\mathbf{S}|\mathbf{Y}, \mathbf{g}^{(0)})$.
  - Initialize the unknown vector $\mathbf{g}$ by $\mathbf{g}^{(0)}$ obtained from pilot symbols.
- **for**$\{t = 1, \ldots, t_{\max}\}$
  - Use the previous estimate $\mathbf{g}^{(t-1)}$ and $p(\mathbf{S}|\mathbf{Y}, \mathbf{g}^{(t-1)})$ to calculate $\widetilde{\mathbf{g}}^{(t-1)}$ according to (3.34).
  - Use $\widetilde{\mathbf{g}}^{(t-1)}$ to obtain the updated channel parameters $\mathbf{g}^{(t)}$ by using (3.42).
  - Discard the wavelet coefficients that are replaced by zero in $\mathbf{g}^{(t)}$ by evaluating $\mathbf{g}^{(t)}_{\text{tr}}$ and $\mathbf{T}_{\text{tr}}$ from (3.49).

    if $\{t \neq t_{\max}\}$
    Use the current estimate $\mathbf{g}^{(t)}_{\text{tr}}$ to update the probability of encoded bits $P^m_{\text{dec}}(c_{k,i})$ and derive $P_m\big(S_k|Y_k, \widehat{H}_k^{(t)}\big)$ from (3.56) and then derive $p(\mathbf{S}|\mathbf{Y}, \mathbf{g}^{(t)})$.

    else
    Decode the information data by thresholding the uncoded bit probabilities.
- **end**

3.4.6 Simulation Results and Discussions

In this section we present a comparative performance study of the proposed EM-MAP algorithm according to the parameters described in Section 3.4.1. The performance comparison is made in terms of MSE for channel estimation and bit error rate (BER) for the combination of channel estimation with the decoding process. The MSE is defined as the averaged square norm of the difference between vectors $\mathbf{g}$ and $\mathbf{g}^{(t_{\max})}$, representing the true and the estimated channel wavelet coefficients at the last iteration, respectively. Besides, we consider the number of parameters that must be estimated at each iteration as a measure of the algorithm complexity.

The binary information data are encoded by a non-recursive non-systematic convolutional encoder with rate $R = 1/2$ and constraint length 3, defined in octal form by $(5, 7)_8$. Throughout the simulations, each frame is composed of $N_{\mathrm{SYM}} = 9$ OFDM symbols with $N = 128$ subcarriers each. Channel coefficients are kept constant during each fading block and changed to new independent realizations (measures) from one frame to the next.

In order to initialize the EM algorithm, we devote $N_P = 3$ channel-uses to the transmission of OFDM pilot symbols (one for each subband). Data and pilot symbols belong to the QPSK constellation with Gray labeling. Corresponding to the pilot symbols, we employ a LS estimate to obtain the initial channel parameters.

The interleaver is pseudo-random, operating over the entire frame of size $N_I = N_{\mathrm{SYM}} NB$ bits (excluding pilots, obviously). Among different wavelet families, "symmetric" wavelet basis functions [130] providing a sparser representation [45] have been considered. Unless otherwise mentioned, the BER and MSE curves correspond to the fourth iteration of the algorithm. Moreover, the signal-to-noise ratio (SNR) is considered in the form of E_b/N_0.

Different propagation environments are considered for performance evaluation. The characteristics of these channels are listed in Tab. 3.1. For all environments, the bandwidth of the CFR over three subbands is 1584 MHz with 384 coefficients and the CIR has a total number of 96 taps. The first channel considered is called *theoretical sparse* channel. This channel is manufactured from a random model that generates a vector of size (96×1) in which only 20 wavelet coefficients out of total 96 have non zero values. A realization of this channel is depicted in Fig. 3.8. Note that although not realistic, this theoretical model provides the best adequacy between the prior assumption of sparseness and the actual propagation environment and allows to see the asymptotic behavior of our algorithm in the extreme case of a very sparse channel representation. The second channel is the non-line of sight (NLOS) channel model CM2 specified by the IEEE802.15.3a channel modeling subcommittee [65] for UWB sytem performance analysis.

We have also considered two sets of channel meausures issued from realistic UWB indoor channel measurements performed in our laboratory [136], in the context of the European Ultrawaves project. These channels are called *Corridor* and *Multifloor*, respectively. In the Corridor environment, the transmit and receive antennas are located in a corridor within the line of sight (LOS) of each other, whereas in the Multifloor scenario, the antennas are located in two different floors and separated by 3.7 meters (see Figs. 3.9 and 3.10).

Channel name	Theoretical sparse	CM2	Corridor	Multifloor
Issued from	Model	Model (IEEE)	Measures	Measures
BW per subband (MHz)	528	528	528	528
Tx-Rx separation (m)	-	0-4	17.5	3.7
(Non-) line of sight	-	NLOS	LOS	NLOS

Table 3.1: Characteristics of different channel scenarios used for performance evaluation.

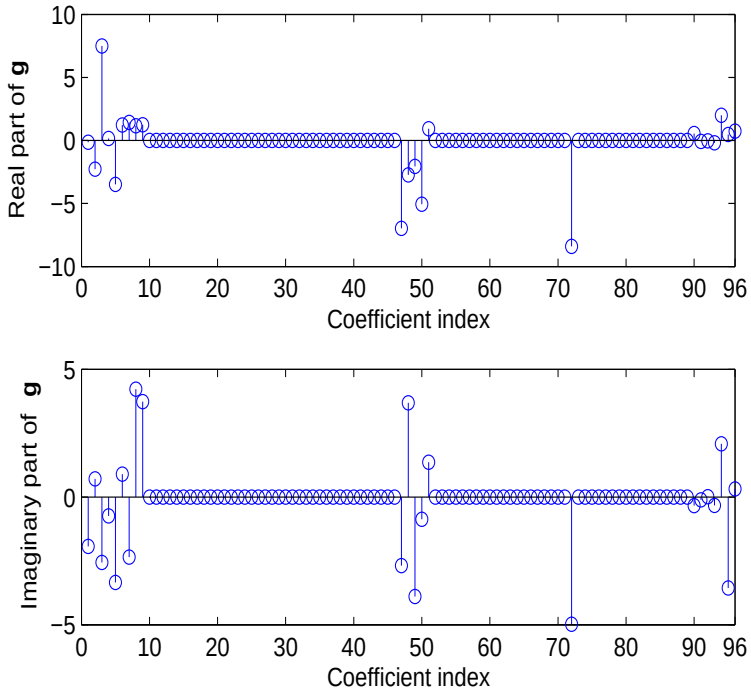

Figure 3.8: Real and imaginary part corresponding to a random vector realization of the *Theoretical sparse* channel wavelet coefficients $\mathbf{g}$.

3.4.6.1 A Proper Choice of Parameter ρ

In Section 3.4.4, we have defined the parameter $\rho \triangleq \alpha^2/\sigma^2$. Of course, the choice of an appropriate value for ρ is important since it affects the variance of the noise vector $\mathbf{Z}_1'$ in (3.35) which is involved in the EM-MAP channel update formula of (3.49). One simple way is to choose ρ according to a specific performance criterion. Here, we consider the minimization of the MSE between the perfect and the estimated channel.

Figure 3.11 shows the MSE between the perfect and the estimated channel from the EM-MAP algorithm, as a function of ρ, obtained in the case of CM2 channel for different values of

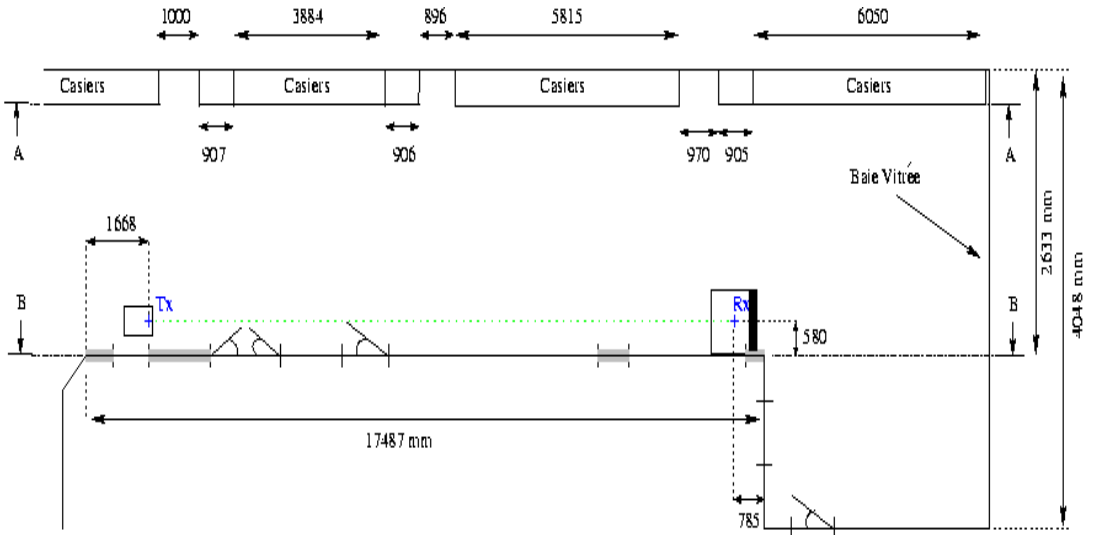

Figure 3.9: Measurement setup of the *Corridor* channel in third floor of ENSTA.

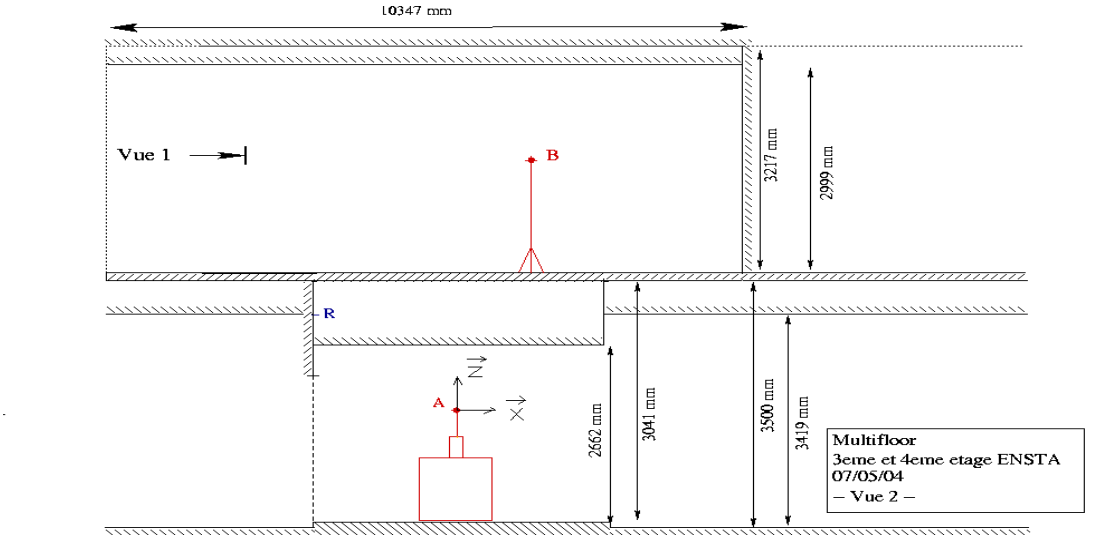

Figure 3.10: Measurement setup of the *Multifloor* channels in third and fourth floor of ENSTA.

E_b/N_0. It can be observed that the MSE is minimized at $\rho = 0.4$ for low SNR and at $\rho = 0.3$ for high SNR values. Further simulations over different propagation environments issued from both IEEE channel models and measurements tend to show that the the interval $0.3 \leq \rho \leq 0.5$ contain the minimum value of the MSE. In the following simulation results, ρ is set to 0.4 for all propagation environments.

3.4.6.2 Performance Evaluation of the EM-MAP Algorithm

For the sake of performance comparison, we consider two pilot-only based approaches using ML and MMSE channel estimation, referred to as pilot-ML and pilot-MMSE. We compare our

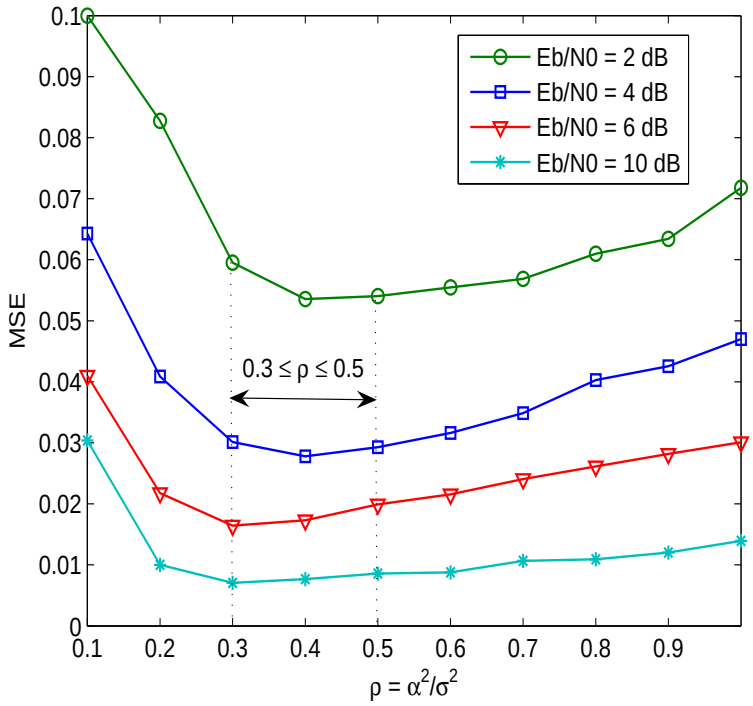

Figure 3.11: Mean square error between the true and the estimated coefficients as a function of ρ.

proposed algorithm with two semi-blind channel estimation methods based on the EM algorithm. The first approach called EM-Freq, estimates all the 384 frequency coefficients corresponding to the channel over the first three subbands using the model (3.21), and is similar to the method proposed in [8]. The second semi-blind approach, called EM-Wav, is a wavelet domain EM based channel estimation where the prior distribution is uniform. This estimate is given by equation (3.37).

Let us first study the case of the Theoretical sparse channel. Figure 3.12 depicts the MSE as a function of E_b/N_0. It can be noticed that, although the pilot-MMSE approach improves the estimation accuracy for low SNR values as compared to pilot-ML, the performance of both pilot-only based channel estimation methods is very far from those of semi-blind methods. Comparing the wavelet domain semi-blind approach (EM-Wav) and the frequency domain approach (EM-freq), we see that a significant gain is achieved by the former. This is due to the inherent averaging present in the estimation formula (via matrix $\mathbf{T}$), since a weighted sum of all elements of the observation vector $\mathbf{Y}$ is combined to estimate a given wavelet coefficient (see equations (3.31) and (3.35)). As shown, the best performance is achieved by the EM-MAP method. We see that by using EM-MAP, a gain of almost 4 dB in SNR is achieved at a MSE of 2×10^{-3}, compared to the EM-Wav method. This clearly shows the adequacy of the EM-MAP method for the case where the unknown channel has few non-zero wavelet coefficients, which is in perfect

agreement with the prior model.

Figure 3.13 shows the BER results along with the BER for the case of perfect channel state information (CSI). It can be seen that at a BER of 10^{-3}, the pilot-ML and the EM-Freq approaches are respectively 3.9 and 2 dB away from the BER obtained with the perfect channel. Furthermore, the performance of the Pilot-MMSE approach is not displayed since it was very close to that of Pilot-ML. Also, we observe that wavelet based semi-blind methods perform closely to the perfect CSI case. For example, at a BER of 10^{-4}, the EM-MAP and EM-Wav method have respectively about 0.2 dB and 0.5 dB of SNR degradation from the performance obtained with perfect CSI.

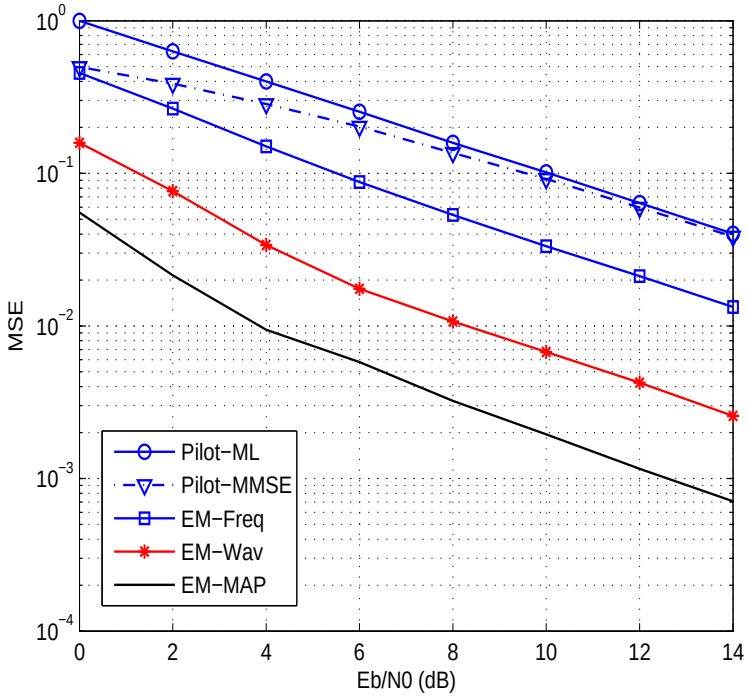

Figure 3.12: Mean square error between the true and estimated coefficients for the sparse channel model.

We now evaluate the performance of the EM-MAP algorithm for the case where the channel does not necessarily have a very sparse representation by considering the Corridor and CM2 channels. Figures 3.14, 3.15 and 3.16 show that wavelet based methods again outperform pilot based and EM-Freq methods in terms of MSE and BER. Furthermore, we observe that the performance of the EM-MAP method is now comparable to that of the EM-Wav method. To understand and explain this result, we analyze the estimation of the prior model parameter λ. Remember from Section 3.4.2 that λ is the probability for a channel wavelet coefficient to be zero, and hence indicates the probability for a channel to have a parsimonious representation in the wavelet domain. This parameter is estimated in the M-step from the estimated wavelet

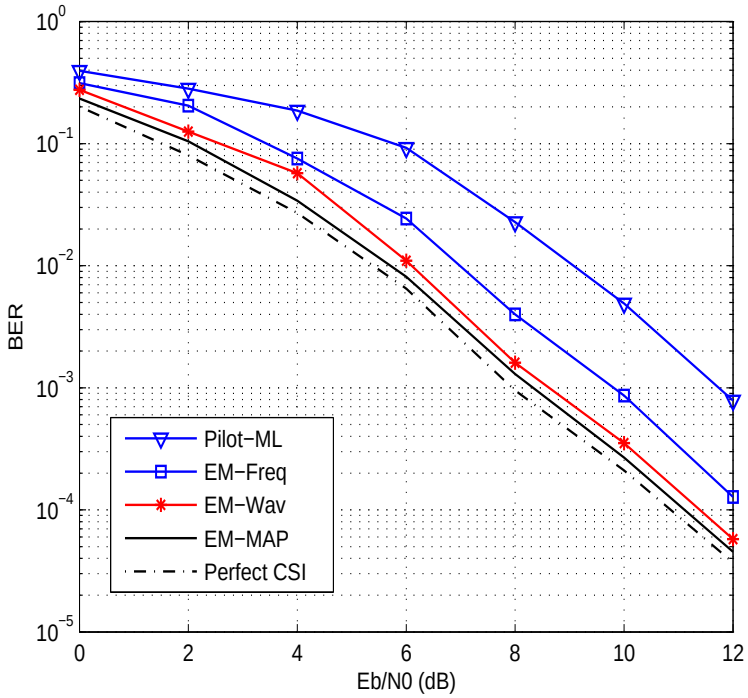

Figure 3.13: BER performance of different channel estimation methods over the sparse channel model.

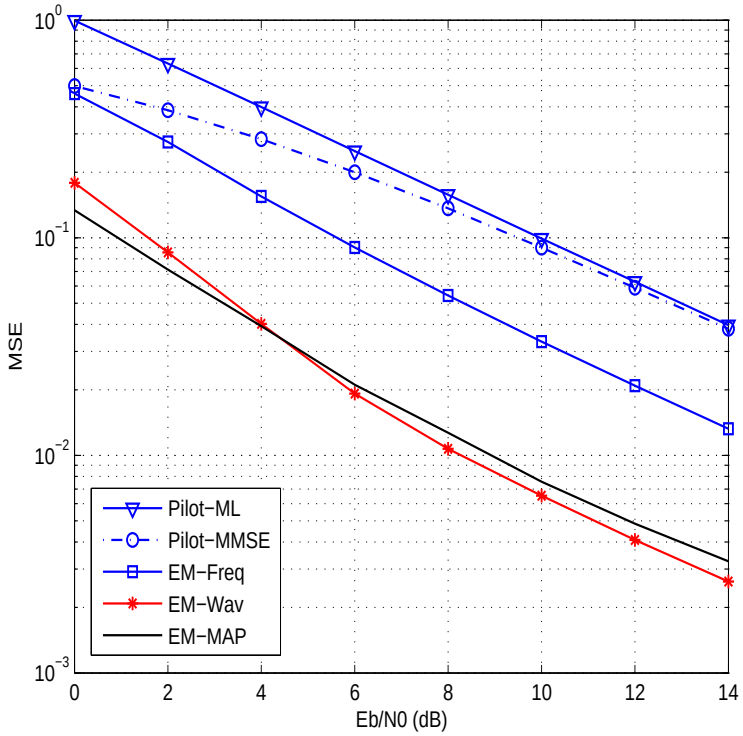

Figure 3.14: Mean square error between the true and estimated coefficients over the Corridor channel.

coefficients, as explained before.

In order to have a measure of channel's sparsness in the wavelet domain, we consider a bank of

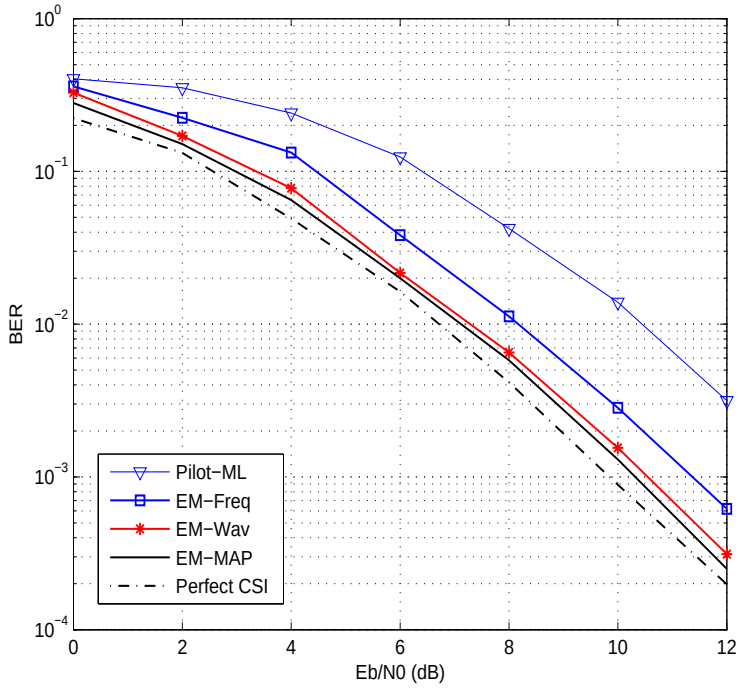

Figure 3.15: BER performance of different channel estimation methods over the Corridor channel.

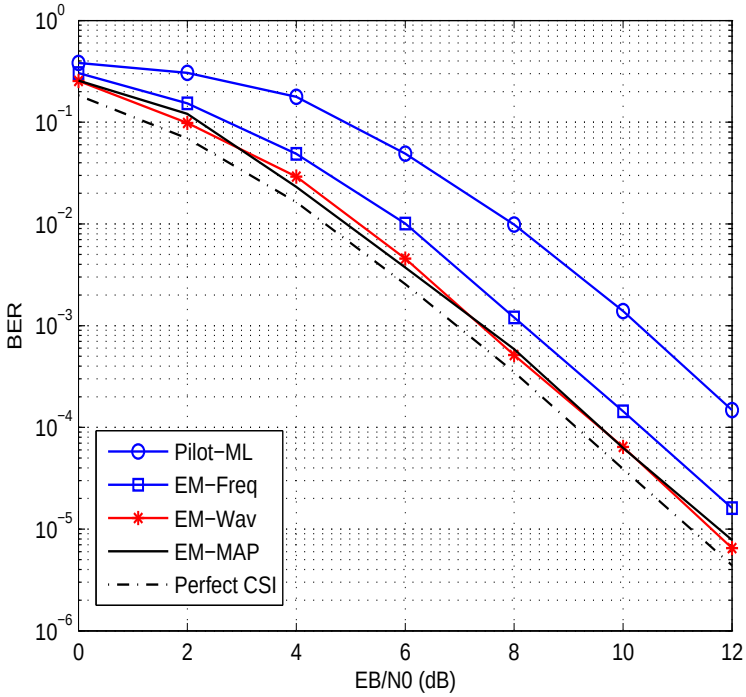

Figure 3.16: BER performance of different channel estimation methods over the CM2 channel.

channels and define the sparsness factor γ as the ratio of zero channel wavelet coefficients to the total number of coefficients. To see the impact of channel's sparsness on the EM-MAP method

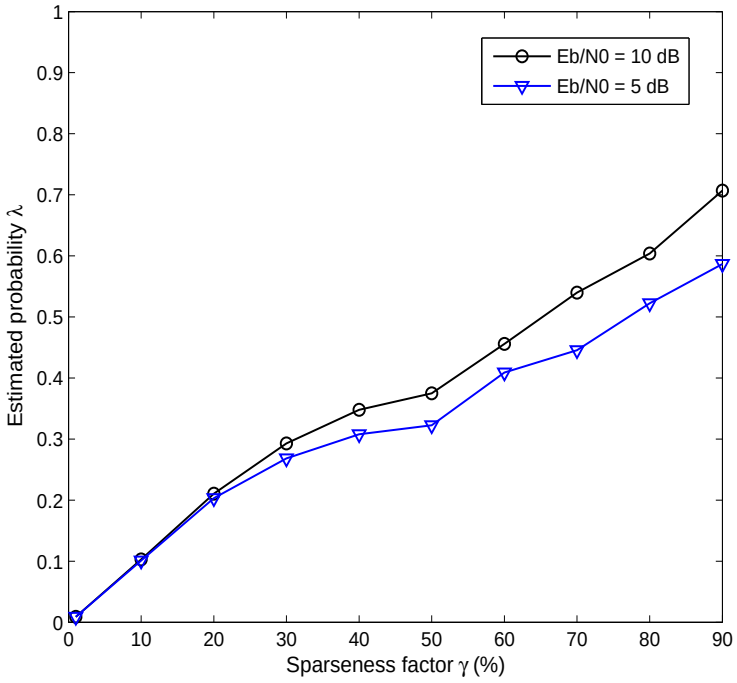

Figure 3.17: Estimation of the a priori model parameter λ versus the sparsness factor.

performance, we have shown in Fig. 3.17 the estimated probability $\hat{\lambda}$ of the Bernoulli-Gaussian prior model (3.38) as a function of γ. It can be observed that $\hat{\lambda}$ increases when the channel tends to become sparser (for large γ values). High values of $\hat{\lambda}$ lead to a Bernoulli-Gaussian prior model with an attenuated Gaussian component. In this case, the EM-MAP algorithm makes use of an adequate prior information and outperforms the EM-Wav approach that does not have access to any prior information (see Figs. 3.12 and 3.13). Also, we observe that when the channel is not very sparse (for low γ values), the algorithm assigns small values to λ. This leads to a Gaussian prior model with a large variance compared to the noise variance, which can be approximated to a uniform prior. As a result, the prior becomes "less informative" and the EM-MAP performs close to EM-Wav, as shown in Figs. 3.14 and 3.15. Thus, the EM-MAP algorithm proposed here is globally able to adapt its prior model parameters to any propagation environment.

Figures 3.18 and 3.19 depict the performance obtained over the Multifloor channel. Obviously, due to the presence of a large number of deep fades in the CFR of this scenario, the overall BER performance is degraded as compared to other scenarios. However, we observe that even in this severe environment, wavelet based channel estimation methods outperform the classical EM-Freq method.

Although we observed that the performance of EM-MAP and EM-Wav methods are close

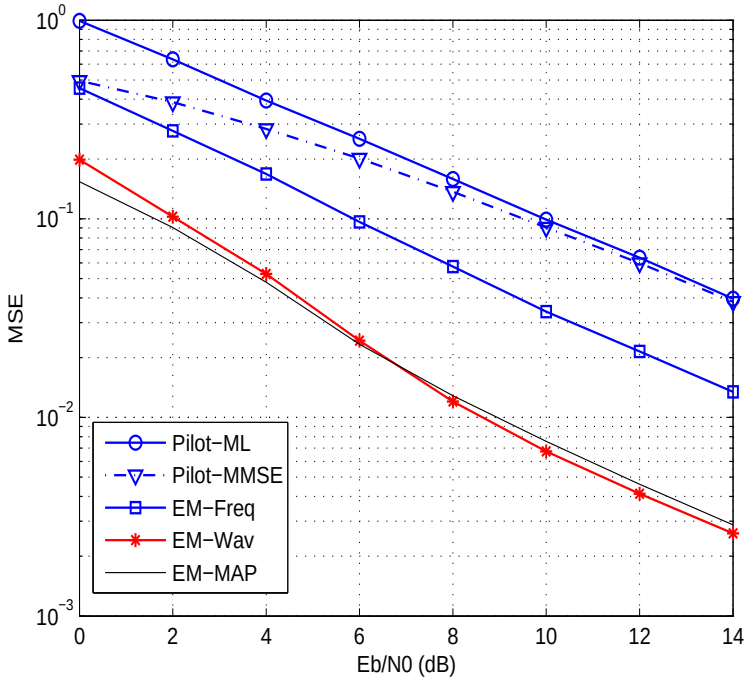

Figure 3.18: Mean square error between the true and estimated coefficients over the Multifloor channel.

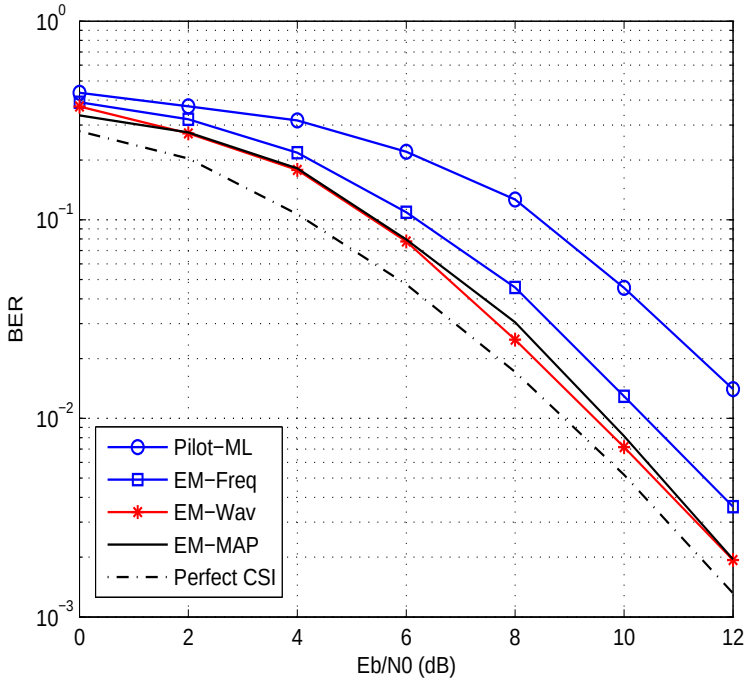

Figure 3.19: BER performance of the EM-MAP method over the Multifloor channel.

over non very sparse channels, it is important to notice that the EM-MAP method takes yet the advantage over the EM-Wav method due to its lower computational complexity. This is explained in the following.

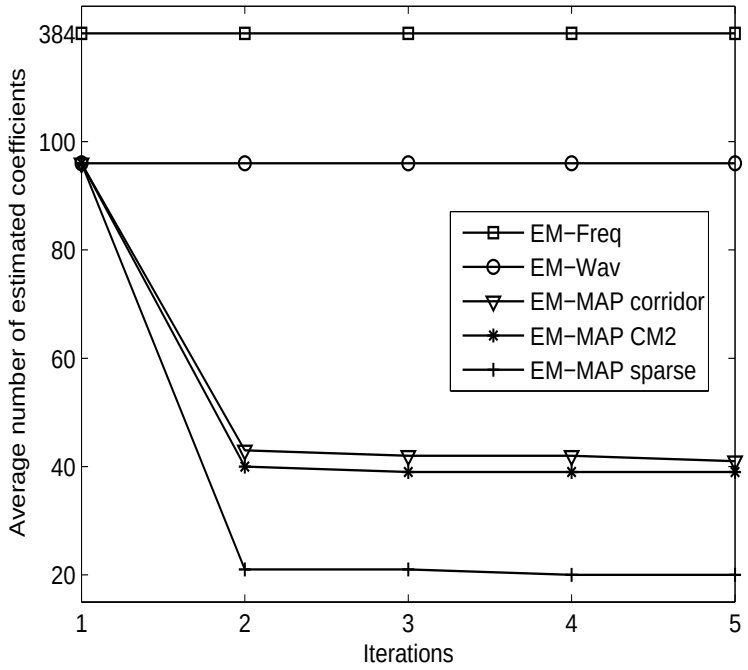

Figure 3.20: Reduction of the number of estimated parameters through iterations, $E_b/N_0 = 8$ dB.

3.4.6.3 Average Number of Estimated Parameters

We now present an important result. We compare the above semi-blind algorithms with respect to the average number of estimated parameters at each iteration of the EM algorithm. This is shown in Fig. 3.20 for different channel scenarios. First, recall that the EM-Freq and EM-Wav methods have to continuously estimate 384 and 96 coefficients at each iteration, respectively. As explained in Subsection 3.4.4.3, by discarding the coefficients that are replaced by zero, the EM-MAP method tends to reduce significantly the number of estimated parameters, specially for sparse channels. This can be seen for the sparse channel where the number of estimated parameters is reduced from 96 down to 20 parameters after the second iteration. Furthermore, for the non-sparse Corridor and CM2 channels, we observe that the EM-MAP method is to be preferred to the EM-Wav, due to its lower computational load, since it estimates about 40 coefficients after the second iteration. Although we observed that in this case these two methods exhibit close performance, the EM-MAP algorithm brings a reduction of about 60% on the number of estimated parameters when compared to the EM-Wav approach.

3.4.6.4 Convergence of the EM-MAP Algorithm

Finally, we analyze the number of iterations that the EM-MAP algorithm requires for convergence. On Fig. 3.21, the MSE performance of the EM-MAP algorithm is presented as a function

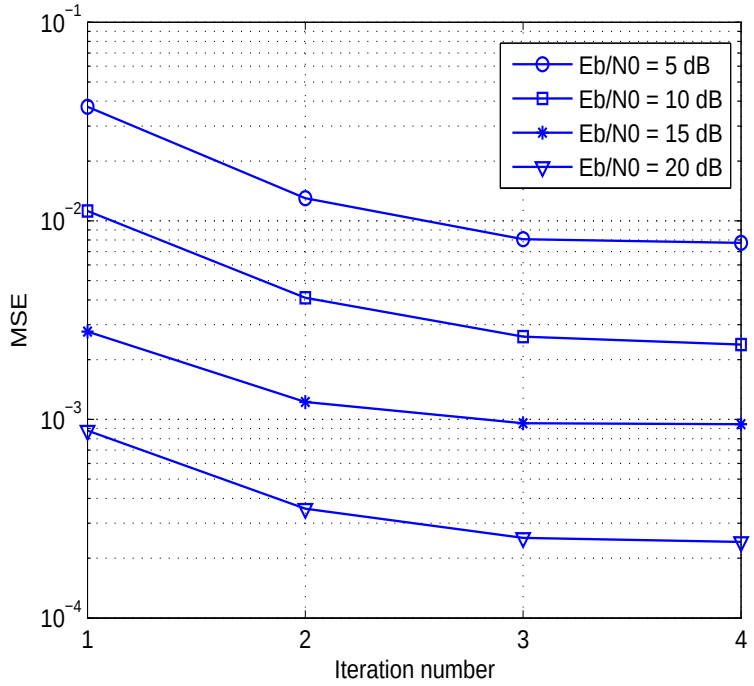

Figure 3.21: Convergence of the MSE with respect to the number of iterations over the CM2 channel.

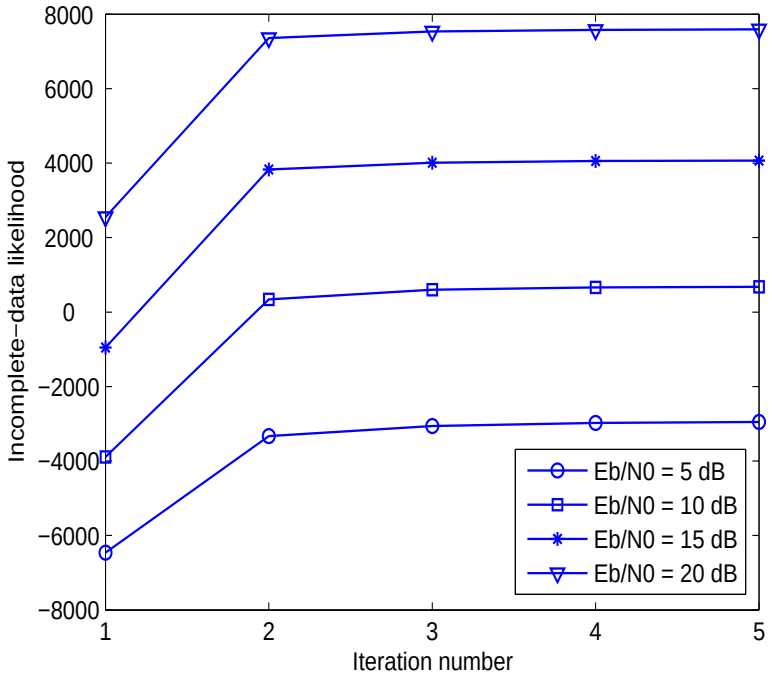

Figure 3.22: Increase of the incomplete-data likelihood with the number of iterations, CM2 channel.

of the number of iterations. It is obvious from these curves that the MSE performance of the proposed algorithm converges within 2 to 4 iterations, depending on the average SNR. This justifies our choice of 4 iterations in the results presented above.

In order to support the theoretical analysis of Section 3.3.2.2 concerning the monotonicity and convergence of the EM algorithm, we have shown in Fig. 3.22 the incomplete-data likelihood as a function of the number of iterations. It is observed that the incomplete-data likelihood increases through iterations and becomes stable within only 3-4 iterations.

3.5 Conclusion

This chapter was devoted to our proposed channel estimation method for MB-OFDM systems. We started by a review of different channel estimation techniques for OFDM systems. First, we presented pilot-only based methods. Although simple, we showed that these techniques require multiple training symbols to provide a reliable channel estimate which can result in a significant loss of the spectral efficiency. Then, we presented different totally-blind channel estimation techniques that estimate the channel without any need for training sequences. As discussed, most of totally-blind algorithms are based on the knowledge of data symbol statistics that must be acquired empirically after receiving a large number of observations. We saw that blind algorithms can be extended to semi-blind methods in which a few number of pilot symbols are required for the algorithm initialization. Among the family of semi-blind methods, we were especially interested by iterative EM based algorithms. We explained two approaches for ML- or MAP-based channel estimation using the EM algorithm.

We presented an EM based semi-blind algorithm, able to integrate the advantages of wavelet based estimation. By expressing the unknown UWB channel in terms of its discrete wavelet coefficients, we choose a prior distribution that captures the possibly sparsness property of UWB channels in the wavelet domain. This led to a MAP estimator equivalent to a hard thresholding procedure at each iteration of the EM algorithm, which we used to reduce the number of estimated coefficients at each iteration. For performance evaluation, we considered different UWB channel environments issued from the IEEE channel models and from realistic indoor measurements. It was observed that when the channel has a sparse wavelet expansion, the prior model parameters which are estimated from the observed data, carry this sparsness information to the EM-MAP algorithm. Moreover, we showed that in this case, the EM-MAP method provides a significant reduction in the number of estimated parameters and outperforms all considered pilot based and semi-blind methods. For non-sparse channels, although both the EM-MAP and EM-Wav methods perform closely, the EM-MAP takes the advantage over the EM-Freq and EM-Wav schemes due to its lower computational complexity.

Finally, we note that for MB-OFDM systems working over a wider channel bandwidth (more than the first three subbands), the EM-MAP method will be even more interesting in terms of complexity reduction and channel estimation accuracy. This is due to the fact that

with the increase of bandwidth, the channel tends to have a more sparse wavelet representation, and consequently, the prior information becomes “more informative”.

Chapter 4

Multiband MIMO-OFDM: Improved Detection and Achieved Throughputs Under Channel Estimation Errors

4.1 Introduction and Motivations

In the previous chapter, we proposed an enhanced receiver for MB-OFDM systems in which the channel is estimated during the iterative process of data detection by using a semi-blind algorithm. Our simulation results showed that the performance of the pilot-only based channel estimation method is far from that provided by our iterative algorithm. However, one must recognize that due to its simplicity, pilot-only based channel estimation is often preferred to iterative methods for obtaining channel state information at the receiver (CSIR). Furthermore, there is a growing demand to increase the data rates of MB-OFDM further beyond 480 Mbps (initially provided by the scheme proposed in [3]), up to 1 Gbps and even higher. The main objective of this chapter is to propose a new transceiver structure being able to increase the data rate and to improve the detection performance of MB-OFDM systems under imperfect channel estimates obtained by a pilot-only technique.

One way to increase the data rate of the current MB-OFDM system is to use higher order modulations. To further enhance the data rates and the coverage ranges, the employment of multiple-input multiple-output (MIMO) scheme to UWB has gained interest recently. Both of these issues are exploited in this chapter. In fact, most UWB systems operate in rich scattering indoor environments, which provide an ideal transmission scenario for MIMO implementation.

In addition, the GHz center frequency of UWB radio relaxes the requirements on the spacing between antenna array elements. Consequently, the combination of UWB and MIMO technology is a good candidate to become a viable and cost-efficient method for very high data rate requirement of future short range wireless applications. However, in UWB communications, the underlying MIMO channel exhibits strong frequency selectivity. By using OFDM and applying a proper cyclic prefix (CP), the frequency selective channel is transformed into a set of frequency-flat subchannels [32]. These considerations motivate the combination of MIMO and OFDM, referred to as MIMO-OFDM.

To this date, multi-antenna UWB technology has been well documented for the traditional single band UWB system [137,138]. On the other hand, research about multi-antenna multiband UWB systems is still largely unexplored. Recently in [139], the authors have proposed the combination of space-time-frequency coding and multiband OFDM and quantified the performance merit of this scheme by considering perfect channel knowledge at the receiver.

In this chapter, we consider a multi-antenna transceiver for MB-OFDM systems and refer to it as MB-MIMO-OFDM. Furthermore, we assume that the channel time-variations are slow enough so that we can consider a constant channel during the transmission of a frame. To obtain the CSIR in the above scenario, we transmit a limited number of known training (pilot) symbols before proceeding to the detection of data symbols. Obviously, due to both the finite number of pilot symbols and to noise, the receiver can only obtain an *imperfect* (and possibly poor) estimate of the channel. It is important to understand the origin of the performance degradation that we have observed in Chapter 3 for MB-OFDM iterative detection using a pilot-only based channel estimation. This can be intuitively explained by the following fact. As the detector decision formula that we have used intrinsically assumes perfect CSIR, the unknown channel is replaced by its imperfect estimate in the detection metric. Regarding this scenario, two important questions will be investigated throughout this chapter. **i**) What type of practical detector can improve the overall system performance under imperfect CSIR ?, **ii**) What are the maximal achievable information rates of practical detectors and how close do they perform with respect to the rates provided by the best possible detector in the presence of channel estimation errors ?

Previous Works: A rich literature exists on the impact of imperfect channel estimation on the performance of communication systems employing multiple antennas. For a MIMO system using pilots for channel estimation, Garg *et al.* showed in [37] that for compensating the performance degradation due to imperfect channel estimation, the number of receive antennas should be increased. Obviously, this may not be always possible in practice. In [140], the authors quantified the performance loss of the V-BLAST scheme [31] due to the presence of

channel estimation errors and derived a tight error floor, whereas in [141] it is shown that a significant performance gain can be achieved by incorporating the imperfect channel into the design of space-time codes. Reference [38] investigated the effect that imperfect channel estimation has on space-time decoding and showed that the classical maximum likelihood (ML) detector, derived for the case of perfect CSIR, becomes largely suboptimal in the presence of channel estimation errors. A similar investigation was carried out in [142, 143] in the case of single-antenna multi-carrier systems based on OFDM.

In order to deal with imperfect channel estimation, one suboptimal approach, known as *mismatched* detection, consists in using the channel estimate for detection, in the same way as if it was a perfect estimate. It is shown, for instance, in [144] that this scheme greatly degrades the detection performance in the presence of channel estimation errors. Furthermore, in [145], the authors showed that under imperfect channel estimation, the rates achieved by the mismatched detector are significantly lower than the limit of the channel capacity.
As an alternative to the aforementioned mismatched approach, Tarokh *et al.* in [144] and recently Taricco and Biglieri in [38], proposed an *improved* ML detection metric under imperfect CSIR and used it with the standard Viterbi algorithm.

Contributions: Recently in [146], we showed that, compared to the mismatched ML metric, the improved ML metric can increase the achievable outage rates of MIMO-OFDM systems, especially when few training symbols are devoted for channel estimation. We have reported similar observations in [147] for the case of single-antenna OFDM systems.
Our aim in this work is to propose an improved iterative detector, that takes into account the imperfect channel estimation obtained via training sequences. To this end, we propose a Bayesian framework based on the *a posteriori* probability density function (pdf) of the perfect channel, conditioned on its estimate. This general framework enables us to formulate any detector by considering the average, over the channel uncertainty, of the detector's cost function that would be applied in the case of perfect channel knowledge. As we shall see, the improved ML metric of [38] becomes a special case of the general framework considered in this work.

At first, we use the improved ML metric for the derivation of an improved turbo-MAP detector. By modifying properly the soft-values at the output of the MAP detector, we reduce the impact of channel uncertainty on the SISO decoder performance. In a second step, in order to answer the question **ii**), we compute the achievable rates associated to the improved and mismatched ML metrics using Gaussian input symbols. This allows us to evaluate the limits of reliable information rates in terms of outage rates, which is an appropriate performance measure for the quasi-static channels [148] considered in this work. Actually, most of the research activity concerning imperfect CSIR is focused on performance evaluation of mismatched detectors

in terms of bit error rate (BER). Our results may serve to evaluate the trade-off between the required quality of service (in terms of BER and achieved throughputs) and the system parameters (e.g., training power, transmission power, period of training, outage probability, etc.) in the presence of channel estimation errors.

Organization of the Chapter: Here, we consider iterative (turbo) detection at the receiver which is an efficient technique when channel coding is used. This scheme has been employed, for instance, in [43, 149] for coded MIMO systems. It is essentially composed of a detector (also called demapper) and of a soft-input soft-output (SISO) channel decoder, exchanging soft information with each other through several iterations. Furthermore, we consider the simple spatial multiplexing [150] (also known under the name of the V-BLAST scheme [30]), which has the potential to drastically increase the capacity of wireless radio links with no additional power or bandwidth consumption [151]. We will also present some simulation results for the case of single-antenna MB-OFDM systems.

The rest of this chapter is organized as follows. In Section 4.2, we describe our MB-MIMO-OFDM channel model and our main assumptions concerning data transmission and channel estimation. In Section 4.3, we introduce a general Bayesian framework for improved detection under imperfect channel estimation and then formulate the improved ML detection metric. In Section 4.4, we provide the formulation of the MAP detector in the case of imperfect CSIR. In Section 4.5, we review the mutual information and capacity of OFDM-based spatial multiplexing systems. Using this and further tools from information theory in Section 4.6, we calculate the achieved throughputs in the sense of outage rates, associated to a receiver using the improved and mismatched ML metric. Section 4.7 illustrates via simulations over realistic UWB channels, the performance of the proposed receiver in terms of BER and achievable outage rates for both single- and multi-antenna MB-OFDM systems. Finally, Section 4.8 concludes the chapter.

4.2 Transmission Model and Channel Estimation

4.2.1 MB-OFDM-Based Spatial Multiplexing

We consider a single-user peer-to-peer MB-OFDM system with M_T transmit and M_R receive antennas ($M_R \geq M_T$). Figure 4.1 shows the block diagram of the transmitter that employs the bit interleaved coded modulation (BICM) scheme which is known to be a simple and efficient method for exploiting channel time-selectivity [44]. The binary data sequence $\boldsymbol{b}$ is encoded by a non-recursive non-systematic convolutional (NRNSC) code before being interleaved by a quasi-random interleaver. The output bits $\boldsymbol{d}$ are multiplexed to M_T sub-streams. Each sub-

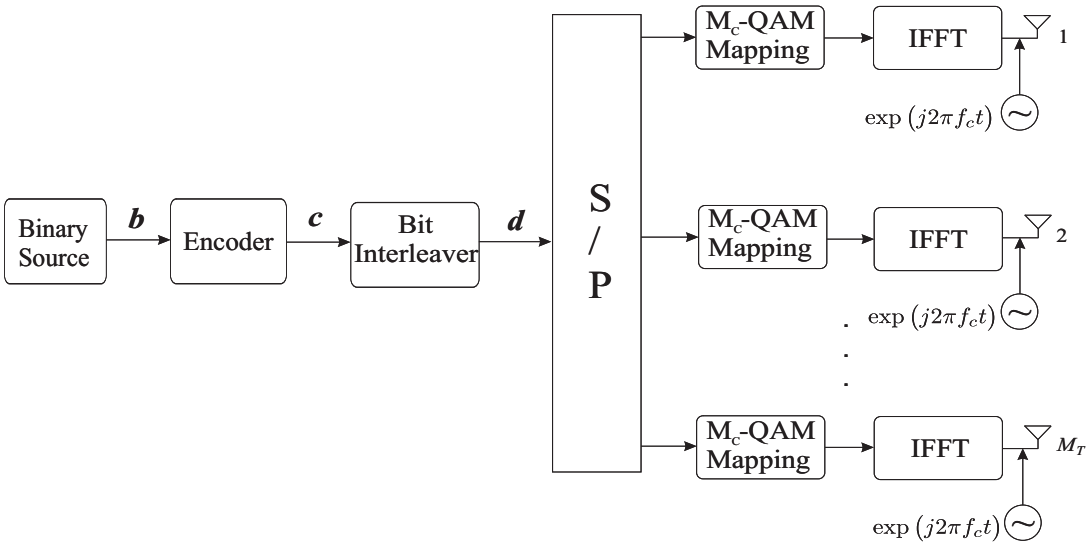

Figure 4.1: Transmitter architecture of multiband MIMO-OFDM transmission.

stream is then mapped to complex $\mathrm{M_c}$-QAM symbols ($\mathrm{M_c} = 2^B$) that are gathered by vectors of length M before being modulated by the OFDM modulator for transmission through the M_T antennas. Note that here, the carrier frequency f_c specifies the subband, over which the signal is transmitted during each OFDM symbol period. The carrier frequency can be changed from one OFDM block to another, so as to enable frequency diversity while minimizing the multiple access interference. Since the signals from all transmit antennas share the same subband, f_c is identical for every transmit antennas. Moreover, the transmission from all of the M_T transmit antennas are simultaneous and synchronous.

Let us organize the transmitted data symbols into $(M_T \times 1)$ frequency vectors $\boldsymbol{s}_k = [s_k(0), ..., s_k(M_T - 1)]^T$ where $k = 0, ..., M-1$ and $s_k(i)$ denoting the data symbol transmitted from the i-th antenna on the k-th subcarrier. The average power of the transmitted symbols is equal to $E_s = \frac{1}{M_T}\mathbb{E}[\,\mathrm{tr}(\boldsymbol{s}_k \boldsymbol{s}_k^\dagger)\,]$ and is common to all the subcarriers. Throughout the chapter, we assume that the length of the CP in the MB-OFDM system is greater than the length of the discrete-time baseband channel impulse response. This assumption guarantees that the frequency-selective fading channel decouples into a set of M parallel frequency-nonselective fading channels, whose fading coefficients are equal to the channel frequency response at the center frequency of the subcarriers.

One then has

$$\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{s}_k + \boldsymbol{z}_k, \qquad k = 0, ..., M-1 \tag{4.1}$$

where $\boldsymbol{y}_k$ denotes the reconstructed data vector for the k-th subcarrier, $\boldsymbol{z}_k$ is a zero-mean circularly symmetric complex Gaussian (ZMCSCG) additive noise satisfying

$$\boldsymbol{\Sigma}_{z,k} = \mathbb{E}[\,\boldsymbol{z}_k \boldsymbol{z}_k^\dagger\,] = \sigma_z^2\, \mathbb{I}_{M_R}, \tag{4.2}$$

and the $(M_R \times M_T)$ channel matrix $\mathbf{H}_k$ has the following structure

$$\mathbf{H}_k = \begin{bmatrix} H_{11}(k) & H_{12}(k) & \cdots & H_{1M_R}(k) \\ H_{21}(k) & H_{22}(k) & \cdots & H_{2M_R}(k) \\ \vdots & \vdots & \ddots & \vdots \\ H_{M_T1}(k) & H_{M_T2}(k) & \cdots & H_{M_TM_R}(k) \end{bmatrix}$$

where $H_{ij}(k)$ denotes the frequency gain from the transmit antenna i to the receive antenna j at the k-th subcarrier.

Channel Model: We consider a tap-delay line fading channel model with L taps where the larger delay is smaller than the CP length. At a given OFDM block, the channel impulse response from the i-th transmit antenna to the j-th receive antenna can be expressed as

$$h_{ij}(t) = \sum_{l=0}^{L-1} \alpha_{ij}(l)\delta(t - \tau_l) \tag{4.3}$$

where $\alpha_{ij}(l)$ and τ_l represent the gain coefficient and the path delay of the l-th multipath component, respectively. From (4.3), the channel frequency response is given by

$$H_{ij}(k) = \sum_{l=0}^{L-1} \alpha_{ij}(l)\ \exp\{-\mathbf{j}2\pi k\Delta_f \tau_l\} \tag{4.4}$$

where Δ_f is the frequency separation between two adjacent subcarriers. We assume that the coefficients $H_{ij}(k)$ are spatially (among antennas) uncorrelated but correlated in frequency (among subcarriers).

For mutual information analysis provided in subsequent sections, we provide an equivalent matrix representation of (4.1) as follows.
We stack the vectors $\boldsymbol{y}_k$, $\boldsymbol{s}_k$ and $\boldsymbol{z}_k$ according to

$$\begin{aligned} \mathbf{Y} &= [\ \boldsymbol{y}_0^T \quad \cdots \quad \boldsymbol{y}_k^T \quad \cdots \quad \boldsymbol{y}_{M-1}^T\ ]^T \\ \mathbf{S} &= [\ \boldsymbol{s}_0^T \quad \cdots \quad \boldsymbol{s}_k^T \quad \cdots \quad \boldsymbol{s}_{M-1}^T\ ]^T \\ \mathbf{Z} &= [\ \boldsymbol{z}_0^T \quad \cdots \quad \boldsymbol{z}_k^T \quad \cdots \quad \boldsymbol{z}_{M-1}^T\ ]^T \end{aligned}$$

where $\mathbf{Y}$ and $\mathbf{Z}$ are $(MM_R \times 1)$ vectors and $\mathbf{S}$ is an $(MM_T \times 1)$ vector. Note that (4.2) implies that the noise vector $\mathbf{Z}$ is white, i.e.,

$$\mathbf{\Sigma}_z = \mathbb{E}[\,\mathbf{Z}\mathbf{Z}^\dagger\,] = \sigma_z^2\,\mathbb{I}_{MM_R}.$$

We furthermore define the $(MM_R \times MM_T)$ block-diagonal matrix

$$\mathbf{H} = \mathrm{diag}\Big(\big[\ \mathbf{H}_0 \quad \cdots \quad \mathbf{H}_k \quad \cdots \quad \mathbf{H}_{M-1}\ \big]\Big). \tag{4.5}$$

With these definitions, (4.1) can be written as

$$\mathbf{Y} = \mathbf{HS} + \mathbf{Z}. \tag{4.6}$$

In the following, we assume that for each frame (corresponding to at least one OFDM symbol), an independent realization of the random channel frequency response matrix $\mathbf{H}$ is drawn and that the channel remains constant within the frame of OFDM symbols.

4.2.2 Pilot-based Channel Estimation

In the following, we aim at estimating the channel matrix $\mathbf{H}_k$. To this end, we send a number of pilot symbols in addition to data symbols. We devote a number of N_P channel-uses to the transmission of $(M_T \times 1)$ pilot vectors $\boldsymbol{s}^i_{P,k}$, $(i = 0, ..., N_P - 1)$ for each subcarrier k. Considering subcarrier k, let us constitute the $(M_T \times N_P)$ matrix $\mathbf{S}_{P,k}$ by stacking in its columns the pilot vectors, i.e., $\mathbf{S}_{P,k} = [\boldsymbol{s}^0_{P,k} ... \boldsymbol{s}^{N_P-1}_{P,k}]$. According to (4.1), during a given channel training interval, we receive

$$\mathbf{Y}_{P,k} = \mathbf{H}_k \, \mathbf{S}_{P,k} + \mathbf{Z}_{P,k} \quad k = 0, \cdots, M-1. \tag{4.7}$$

The definitions of $\mathbf{Y}_{P,k}$ and $\mathbf{Z}_{P,k}$ are similar to that of $\mathbf{S}_{P,k}$. We denote by E_P the average power of the training symbols on any subcarrier as

$$E_P \triangleq \frac{1}{N_P M_T} \mathrm{tr}\left(\mathbf{S}_{P,k} \mathbf{S}^\dagger_{P,k}\right). \tag{4.8}$$

The least-squares estimate of $\mathbf{H}_k$ is obtained by minimizing $\|\mathbf{Y}_{P,k} - \mathbf{H}_k \, \mathbf{S}_{P,k}\|^2_F$ with respect to $\mathbf{H}_k$ and coincides here with the ML estimate. We have

$$\widehat{\mathbf{H}}^{\mathrm{ML}}_k = \mathbf{Y}_{P,k} \, \mathbf{S}^\dagger_{P,k} \left(\mathbf{S}_{P,k} \mathbf{S}^\dagger_{P,k}\right)^{-1}. \tag{4.9}$$

Let us denote by $\boldsymbol{\mathcal{E}}_k$ the matrix of estimation errors. From (4.7) and (4.9), it is easy to see that

$$\widehat{\mathbf{H}}^{\mathrm{ML}}_k = \mathbf{H}_k + \boldsymbol{\mathcal{E}}_k, \quad \text{with} \quad \boldsymbol{\mathcal{E}}_k = \mathbf{Z}_{P,k} \, \mathbf{S}^\dagger_{P,k} \left(\mathbf{S}_{P,k} \mathbf{S}^\dagger_{P,k}\right)^{-1}. \tag{4.10}$$

It is known that the best channel estimate is obtained with mutually orthogonal training sequences, which result in uncorrelated estimation errors. In other words, we should choose $\mathbf{S}_{P,k}$ with orthogonal rows such that

$$\mathbf{S}_{P,k} \mathbf{S}^\dagger_{P,k} = N_P E_P \, \mathbb{I}_{M_T}. \tag{4.11}$$

Then, the j-th column $\boldsymbol{\mathcal{E}}_{k,j}$ of $\boldsymbol{\mathcal{E}}_k$ has the covariance matrix $\boldsymbol{\Sigma}_{\mathcal{E}}$ given by

$$\boldsymbol{\Sigma}_{\mathcal{E}} = \mathbb{E}\left[\boldsymbol{\mathcal{E}}_{k,j} \; \boldsymbol{\mathcal{E}}^\dagger_{k,j}\right] = \sigma^2_{\mathcal{E}} \, \mathbb{I}_{M_R} \;, \quad \text{where } \sigma^2_{\mathcal{E}} = \frac{\sigma^2_z}{N_P E_P} \;. \tag{4.12}$$

Thus from (4.10), the conditional pdf of $\widehat{\mathbf{H}}_k^{\mathrm{ML}}$ given $\mathbf{H}_k$ can be easily expressed as

$$p(\widehat{\mathbf{H}}_k^{\mathrm{ML}}|\mathbf{H}_k) = \mathcal{CN}\big(\mathbf{H}_k, \mathbb{I}_{M_T} \otimes \mathbf{\Sigma}_{\mathcal{E}}\big). \tag{4.13}$$

Furthermore, we assume that the channel matrix $\mathbf{H}_k$ has a normal *prior* distribution as

$$\mathbf{H}_k \sim \mathcal{CN}\big(\mathbf{0}, \mathbb{I}_{M_T} \otimes \mathbf{\Sigma}_{H_k}\big) = \frac{1}{\pi^{M_R M_T} \det\{\mathbf{\Sigma}_{H_k}\}^{M_T}} \exp\left\{ -\operatorname{tr}\left(\mathbf{H}_k \mathbf{\Sigma}_{H_k}^{-1} \mathbf{H}_k^{\dagger}\right)\right\} \tag{4.14}$$

where $\mathbf{\Sigma}_{H_k}$ is the $(M_R \times M_R)$ covariance matrix of the columns of $\mathbf{H}_k$. We assume that the entries of $\mathbf{H}_k$ are i.i.d. and thus $\mathbf{\Sigma}_{H_k}$ is a diagonal matrix with equal diagonal entries $\sigma_{h,k}^2$.
By using the prior pdf of $\mathbf{H}_k$ (4.14) and the pdf of $(\widehat{\mathbf{H}}_k^{\mathrm{ML}}|\mathbf{H}_k)$ (4.13), we can derive the *posterior* distribution of the perfect channel matrix, conditioned on its ML estimate, as follows (see Appendix B).

$$p\left(\mathbf{H}_k|\widehat{\mathbf{H}}_k^{\mathrm{ML}}\right) = \mathcal{CN}\Big(\mathbf{\Sigma}_{\Delta}\widehat{\mathbf{H}}_k^{\mathrm{ML}},\, \mathbb{I}_{M_T} \otimes \mathbf{\Sigma}_{\Delta}\mathbf{\Sigma}_{\mathcal{E}}\Big), \tag{4.15}$$

where

$$\mathbf{\Sigma}_{\Delta} = \mathbf{\Sigma}_{H_k}\big(\mathbf{\Sigma}_{\mathcal{E}} + \mathbf{\Sigma}_{H_k}\big)^{-1}. \tag{4.16}$$

Under the above-mentioned assumptions, we have

$$\mathbf{\Sigma}_{\Delta} = \delta \mathbb{I}_{M_R} \tag{4.17}$$

where

$$\delta = \frac{\sigma_{h,k}^2}{\sigma_{h,k}^2 + \sigma_{\mathcal{E}}^2}\,. \tag{4.18}$$

In particular, when the number of pilot symbols tends to infinity, it is not difficult to see that $\delta \to 1$ and $\delta\,\sigma_{\mathcal{E}}^2 \to 0$ and consequently $p\left(\mathbf{H}_k|\widehat{\mathbf{H}}_k^{\mathrm{ML}}\right)$ tends to a Dirac delta function. The availability of the estimation error distribution is an interesting feature of pilot assisted channel estimation that we used to derive the posterior distribution (4.15). This distribution constitutes a Bayesian framework which is exploited in the following for the design of appropriate detectors under imperfect channel estimation.
For the sake of simplicity, we will not specify hereafter the superscript ML for $\widehat{\mathbf{H}}_k$.

4.3 Detector Design in the Presence of Channel Estimation Errors

We now provide the formulation of a modified detection rule that takes into account the available imperfect CSIR. In our approach, we consider a per subcarrier detection scheme. Thus, we consider the model (4.1) and denote by $f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k)$ the cost function that would let us to decide in favor of a particular $\boldsymbol{s}_k$ at the receiver if the channel was perfectly known. We note

that, depending on the detection criterion, $f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k)$ can be for instance the posterior pdf $p(\boldsymbol{s}_k|\boldsymbol{y}_k, \mathbf{H}_k)$, the logarithm of the likelihood function $W(\boldsymbol{y}_k|\mathbf{H}_k, \boldsymbol{s}_k)$ or the mean-square error. Under a pilot-based channel estimation characterized by the posterior pdf (4.15), we propose a detector based on the minimization of a new cost function defined as

$$\boxed{\widetilde{f}(\boldsymbol{y}_k, \boldsymbol{s}_k, \widehat{\mathbf{H}}_k) = \int_{\mathbf{H}} f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k) p(\mathbf{H}_k|\widehat{\mathbf{H}}_k)\, \mathrm{d}\mathbf{H}_k = \mathbb{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\big[f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k)\big|\widehat{\mathbf{H}}_k\big]} \tag{4.19}$$

where we have averaged the cost function f over all possible realizations of the unknown channel $\mathbf{H}_k$ conditioned on its available estimate $\widehat{\mathbf{H}}_k$ by using the distribution (4.15).

We note that the detector minimizing (4.19) is an alternative to the classical and sub-optimal *mismatched* detector; this detector is based on the minimization of the cost function $f(\boldsymbol{y}_k, \boldsymbol{s}_k, \widehat{\mathbf{H}}_k)$, where one uses the estimated channel $\widehat{\mathbf{H}}_k$ in the same metric that would be applied if the channel was perfectly known, i.e., $f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k)$. The proposed approach in (4.19), differs from the mismatched detection on the conditional expectation $\mathbb{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}[.]$ and provides a robust design by averaging the cost function $f(\boldsymbol{y}_k, \boldsymbol{s}_k, \mathbf{H}_k)$ over all possible realizations of the channel estimation errors.

4.3.1 Application to ML detection

The general principle of ML detection is used in a large variety of pratical decoders such as soft and hard Viterbi decoding [63], iterative detection and decoding using the BCJR algorithm [11], space-time decoding, etc. Since we shall use the idea presented above in the case of iterative MAP detection, we specify in what follows the expression (4.19) for ML detection.

To this end, we consider the problem of detecting the symbol vector $\boldsymbol{s}_k$ from the observation model (4.1) in the ML sense so as to maximize the likelihood $W(\boldsymbol{y}_k|\mathbf{H}_k, \boldsymbol{s}_k)$. It is well known that under the i.i.d. Gaussian noise assumption, the ML detection of $\boldsymbol{s}_k$ leads to minimizing the Euclidean distance metric $\mathcal{D}_{\text{ML}}$:

$$\hat{\boldsymbol{s}_k}^{\text{ML}}(\mathbf{H}_k) = \underset{\boldsymbol{s}_k \in \mathbb{C}^{M_T \times 1}}{\arg\min} \big\{ \mathcal{D}_{\text{ML}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \mathbf{H}_k) \big\}, \tag{4.20}$$

with

$$\mathcal{D}_{\text{ML}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \mathbf{H}_k) \triangleq -\log W(\boldsymbol{y}_k|\boldsymbol{s}_k, \mathbf{H}_k) \quad \propto \quad \|\boldsymbol{y}_k - \mathbf{H}_k \boldsymbol{s}_k\|^2, \tag{4.21}$$

where $\mathbb{C}$ is the set of complex numbers and $\propto$ means "is proportional to".

Under imperfect channel estimation, the classical mismatched detector consists in replacing the exact channel by its estimate as

$$\hat{\boldsymbol{s}_k}^{\text{MM}}(\widehat{\mathbf{H}}_k) = \underset{\boldsymbol{s}_k \in \mathbb{C}^{M_T \times 1}}{\arg\min} \big\{ \mathcal{D}_{\text{MM}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k) \big\},$$

with

$$\mathcal{D}_{\text{MM}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k) \propto \|\boldsymbol{y}_k - \widehat{\mathbf{H}}_k \boldsymbol{s}_k\|^2. \tag{4.22}$$

Now, as an alternative to this mismatched detection, we derive a modified likelihood criterion $\widetilde{W}(\boldsymbol{y}_k|\widehat{\mathbf{H}}_k, \boldsymbol{s}_k)$ by evaluating the posterior mean of $W(\boldsymbol{y}_k|\mathbf{H}_k, \boldsymbol{s}_k)$ with respect to $\mathbf{H}_k$. This is done by using the pdf (4.15) and the Bayesian formulation (4.19), as follows:

$$\begin{aligned}\widetilde{W}(\boldsymbol{y}_k|\widehat{\mathbf{H}}_k, \boldsymbol{s}_k) &= \int_{\mathbf{H}_k \in \mathbb{C}^{M_R \times M_T}} W(\boldsymbol{y}_k|\mathbf{H}_k, \boldsymbol{s}_k)\, p(\mathbf{H}_k|\widehat{\mathbf{H}}_k) \quad \mathrm{d}\mathbf{H}_k \\ &= \mathbb{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\Big[W(\boldsymbol{y}_k|\mathbf{H}_k, \boldsymbol{s}_k)\big|\, \widehat{\mathbf{H}}_k\Big]. \end{aligned} \tag{4.23}$$

After some algebraic manipulations developed in Appendix B, we obtain

$$\widetilde{W}(\boldsymbol{y}_k|\widehat{\mathbf{H}}_k, \boldsymbol{s}_k) = \mathcal{CN}(\mathbf{m}_{\mathcal{M}}, \mathbf{\Sigma}_{\mathcal{M}}), \tag{4.24}$$

where

$$\mathbf{m}_{\mathcal{M}} = \delta\, \widehat{\mathbf{H}}_k\, \boldsymbol{s}_k \quad \text{and} \quad \mathbf{\Sigma}_{\mathcal{M}} = \mathbf{\Sigma}_z + \delta\, \mathbf{\Sigma}_{\mathcal{E}}\, \|\boldsymbol{s}_k\|^2. \tag{4.25}$$

Compared to (4.22), we define the improved ML decision metric in the presence of imperfect channel estimation is

$$\begin{aligned}\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k) &\triangleq -\log \widetilde{W}(\boldsymbol{y}_k|\widehat{\mathbf{H}}_k, \boldsymbol{s}_k) \\ &= M_R \log\, \pi\big(\sigma_z^2 + \delta\, \sigma_{\mathcal{E}}^2\, \|\boldsymbol{s}_k\|^2\big) + \frac{\|\boldsymbol{y}_k - \delta\, \widehat{\mathbf{H}}_k\, \boldsymbol{s}_k\|^2}{\sigma_z^2 + \delta\, \sigma_{\mathcal{E}}^2\, \|\boldsymbol{s}_k\|^2}. \end{aligned} \tag{4.26}$$

We note that under near perfect CSIR, obtained when the number of pilots N_P tends to infinity, we have

$$\lim_{N_P \to \infty} \frac{\mathcal{D}_{\mathcal{M}}\big(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k\big)}{\mathcal{D}_{\text{MM}}\big(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k\big)} = 1. \tag{4.27}$$

We see that logically, the improved metric of (4.24) becomes equivalent to the mismatched metric, for negligible estimation errors, i.e., for $\sigma_{\mathcal{E}}^2 \to 0$.

Note also that the metric proposed independently in [38] and [144] for the case of ML detection, can in fact be considered as a special case of the general framework (4.19) that we proposed.

4.4 Iterative MAP Detection of MB-MIMO-OFDM

At the receiver, we perform iterative symbol detection and channel decoding. As shown in Fig. 4.2, the receiver principally consists of a set of MIMO detectors (one per subcarrier) and a SISO channel decoder that exchange *extrinsic* soft information with each other. Here, we consider this soft information in the form of log-likelihood ratio (LLR).

The SISO decoder is based on the forward-backward (BCJR) algorithm, described in [11, 152].

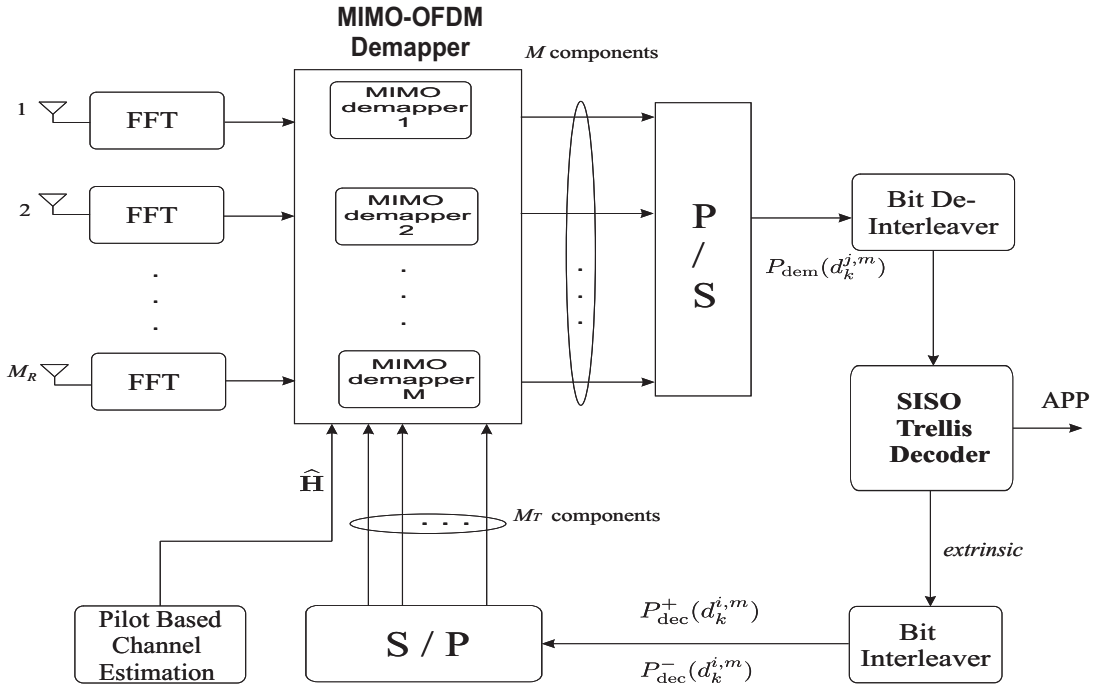

Figure 4.2: Block diagram of the MAP receiver for multiband MIMO-OFDM with pilot based channel estimation.

We briefly present in the following the MIMO detector part, based on the MAP criterion, assuming perfect CSIR available at the receiver. Then, we modify the detector for the case of imperfect CSIR.

4.4.1 MAP Detection Under Perfect CSIR

Let $d_k^{i,m}$ be the m-th ($m = 1, 2, ..., BM_T$) bit corresponding to the symbol vector $\boldsymbol{s}_k$, transmitted from the i-th antenna and the k-th subcarrier. Furthermore, for the case of formulation simplicity, we assume that each frame is only composed of one OFDM symbol and hence the time index does not appear in our formulation. We denote by $L(d_k^{i,m})$ the LLR of the bit $d_k^{i,m}$ at the output of the MIMO detector. Conditioned on perfect CSIR $\mathbf{H}_k$, $L(d_k^{i,m})$ is given by

$$L(d_k^{i,m}) = \log \frac{P_{\rm dem}(d_k^{i,m} = 1 | \boldsymbol{y}_k, \mathbf{H}_k)}{P_{\rm dem}(d_k^{i,m} = 0 | \boldsymbol{y}_k, \mathbf{H}_k)}, \quad k = 0, ..., M-1, \tag{4.28}$$

where $P_{\rm dem}(d_k^{i,m} | \boldsymbol{y}_k, \mathbf{H}_k)$ denotes the probability of transmission of $d_k^{i,m}$, calculated at the MIMO demapper. Let $\mathcal{S}$ be the set of all possibly-transmitted symbol vectors $\boldsymbol{s}_k$. We partition $\mathcal{S}$ into

$\mathcal{S}_0^m$ and $\mathcal{S}_1^m$, for which the m-th bit of $\boldsymbol{s}_k$ equals "0" or "1", respectively. We have

$$L(d_k^{i,m}) = \log \frac{\displaystyle\sum_{\boldsymbol{s}_k \in \mathcal{S}_1^m} e^{-\mathcal{D}_{\text{ML}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \mathbf{H}_k)} \prod_{\substack{n=1 \\ n \neq m}}^{BM_T} P_{\text{dec}}^1(d_k^{i,n})}{\displaystyle\sum_{\boldsymbol{s}_k \in \mathcal{S}_0^m} e^{-\mathcal{D}_{\text{ML}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \mathbf{H}_k)} \prod_{\substack{n=1 \\ n \neq m}}^{BM_T} P_{\text{dec}}^0(d_k^{i,n})}, \tag{4.29}$$

where $P_{\text{dec}}^1(d_k^{i,n})$ and $P_{\text{dec}}^0(d_k^{i,n})$ are *a priori* information coming from the SISO decoder. The summation in (4.29) is taken over the product of the likelihood $W(\boldsymbol{y}_k|\boldsymbol{s}_k, \mathbf{H}_k) = \exp\{-\mathcal{D}_{\text{ML}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \mathbf{H}_k)\}$, by the *a priori* probability on this symbol (the term $\prod P_{\text{dec}}$), fed back from the SISO decoder at the previous iteration. In this latter term, the *a priori* probability of the bit $d_k^{i,m}$ itself has been excluded in order to respect the exchange of *extrinsic* information between the channel decoder and the demapper. Also, note that this term assumes independent coded bits $d_k^{i,n}$, which is likely to hold for random interleaving of large size. At the first iteration, where no *a priori* information is available on bits $d_k^{i,n}$, the probabilities $P_{\text{dec}}^0(d_k^{i,n})$ and $P_{\text{dec}}^1(d_k^{i,n})$ are set to 0.5.

4.4.2 Improved MAP Detection Under Imperfect CSIR

We propose here some modifications to the above MIMO detector for the case of imperfect channel estimation at the receiver. We notice that the metric $\mathcal{D}_{\text{ML}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \mathbf{H}_k)$ involved in (4.29) requires the knowledge of the perfect channel matrix $\mathbf{H}_k$, of which the receiver has *solely* an imperfect estimate $\widehat{\mathbf{H}}_k$. We propose to use the decoding metric $\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k)$ of (4.26) for the evaluation of the LLRs in (4.29). The new LLR under imperfect CSIR is chosen as

$$\widetilde{L}(d_k^{i,m}) = \log \frac{\displaystyle\sum_{\boldsymbol{s}_k \in \mathcal{S}_1^m} e^{-\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k)} \prod_{\substack{n=1 \\ n \neq m}}^{BM_T} P_{\text{dec}}^1(d_k^{i,n})}{\displaystyle\sum_{\boldsymbol{s}_k \in \mathcal{S}_0^m} e^{-\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k)} \prod_{\substack{n=1 \\ n \neq m}}^{BM_T} P_{\text{dec}}^0(d_k^{i,n})}. \tag{4.30}$$

For instance, the nominator of (4.30) is calculated as follows:

$$\begin{aligned} P(d_k^{i,m} = 1|\boldsymbol{y}_k, \widehat{\mathbf{H}}_k) &= \sum_{\boldsymbol{s}_k \in \mathcal{S}_1^m} \exp\big\{ -\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k)\big\} \prod_{\substack{n=1 \\ n \neq m}}^{BM_T} P_{\text{dec}}^1(d_k^{i,n}) \\ &= \frac{1}{\pi\, e^{M_R}} \sum_{\boldsymbol{s}_k \in \mathcal{S}_1^m} \frac{1}{\sigma_z^2 + \delta\sigma_{\mathcal{E}}^2 \|\boldsymbol{s}_k\|^2} \exp\left\{ \frac{\|\boldsymbol{y}_k - \delta\widehat{\mathbf{H}}_k \boldsymbol{s}_k\|^2}{\sigma_z^2 + \delta\sigma_{\mathcal{E}}^2 \|\boldsymbol{s}_k\|^2} \right\} \prod_{\substack{n=1 \\ n \neq m}}^{BM_T} P_{\text{dec}}^1(d_k^{i,n}). \end{aligned} \tag{4.31}$$

4.5 Mutual Information and Capacity of OFDM-based Spatial Multiplexing Systems

In this section, we provide the expression of the mutual information and different capacity definitions of OFDM-based spatial multiplexing systems. These expressions are used in the next section to calculate the achievable rates and to study the outage properties of mismatched and improved ML detectors.

4.5.1 Mutual Information

We use the overall observation model (4.6) and assume that the transmitter disposes of a *deterministically* perfect knowledge of the block-diagonal channel matrix $\mathbf{H} = \mathbf{H}_{\text{known}}$. For simplicity, we consider only one subband of MB-MIMO-OFDM transmission. The mutual information (in bit/s/Hz) of the OFDM-based spatial multiplexing system under an average transmit power constraint $\text{tr}(\mathbf{Q}) \leq \bar{P}$ is given by [151, 153]

$$I = \frac{1}{M} \log_2 \det \left\{ \mathbb{I}_{M_R M} + \frac{\mathbf{H}_{\text{known}} \, \mathbf{Q} \, \mathbf{H}_{\text{known}}^{\dagger}}{\sigma_z^2} \right\} \tag{4.32}$$

where $\mathbf{Q}$ is the covariance matrix of the Gaussian input vector $\mathbf{S}$ in (4.6). Note that the mutual information is normalized by M, since M data symbols are transmitted in one OFDM symbol. Also note that we have ignored the loss in spectral efficiency due to the presence of the CP. Let us consider in more detail the matrix $\mathbf{Q}$. In fact, $\mathbf{Q}$ is a $(MM_T \times MM_T)$ block-diagonal matrix given by

$$\mathbf{Q} = \text{diag}\Big(\big[\, \mathbf{Q}_0 \quad \cdots \quad \mathbf{Q}_k \quad \cdots \quad \mathbf{Q}_{M-1} \, \big] \Big)$$

where the the $(M_T \times M_T)$ matrices $\mathbf{Q}_k$ are the covariance matrices of the Gaussian vectors $\boldsymbol{s}_k$ and determine the power allocation across the transmit antennas and across the OFDM subcarriers. If the channel is perfectly known at the transmitter, the optimum power allocation is obtained by distributing the total available power $\bar{P}$ according to the water-filling solution [154]. In OFDM-based spatial multiplexing systems, statistically independent data symbols are transmitted from different antennas and different subcarriers and the total available power is allocated uniformly across all space-frequency subchannels [154]. In the following, we set $\mathbf{Q}_k = \bar{P}/(MM_T)\,\mathbb{I}_{M_T}$ $(k = 0, ..., M-1)$, which can be easily verified to result in $\text{tr}(\mathbf{Q}) = \bar{P}$. Using (4.32), we therefore obtain

$$I = \frac{1}{M} \sum_{k=0}^{M-1} I_k = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \det \left\{ \mathbb{I}_{M_R} + \rho \, \mathbf{H}_k \mathbf{H}_k^{\dagger} \right\} \tag{4.33}$$

where $\rho = \bar{P}/(MM_T\sigma_z^2)$ and the quantity I_k is the mutual information of the k-th MIMO-OFDM subchannel.

4.5.2 Ergodic Capacity

The basic assumption here is that the channel is perfectly known at the receiver but not at the transmitter. Furthermore, the transmission time is assumed to be long enough to reveal the long-term ergodic properties of the channel. That is, the number of fading blocks spanned by a codeword goes to infinity whereas the size of each fading block (which equals the number of subcarriers in the OFDM system multiplied by the number of OFDM symbols spanning one channel-use) remains constant and finite. Since the CSIR is not available at the transmitter, the total power is allocated uniformly across all space-frequency subchannels that is $\mathbf{Q}_k = \bar{P}/(MM_T)\,\mathbb{I}_{M_T}$ $(k = 0, ..., M-1)$. In this case, a capacity in the sense defined by Shannon exists and is expressed as [151]

$$C_{\text{erg}} = \mathbb{E}_{\{\mathbf{H}_0,\ldots,\mathbf{H}_{M-1}\}}\left[\frac{1}{M}\sum_{k=0}^{M-1}\log_2\det\left\{\mathbb{I}_{M_R} + \rho\ \mathbf{H}_k\mathbf{H}_k^{\dagger}\right\}\right]. \tag{4.34}$$

In fact, the ergodic capacity C_{erg} can *only* be achieved when one transmit a codeword over a very large number of independent fading blocks. Obviously, this is not the case in MB-OFDM transmission considered in this thesis.

4.5.3 Outage Capacity

In several applications in wireless communications of which MB-OFDM is a special case, the channel is chosen randomly at the beginning of the transmission and is held fixed during the whole transmission session. More precisely, the transmission time is not longer than the channel coherence time and consequently the mean of the maximum mutual information in (4.34) is not equal to the channel capacity. On the other hand, the capacity in the Shannon sense does not exist since there is a non-zero probability that the realized $\mathbf{H}$ and its estimate $\widehat{\mathbf{H}}$, are not capable of supporting even a very small rate (it is a matter of luck...). In this case, the capacity is a random entity, as it depends on the instantaneous realization of the channel matrix $\mathbf{H}$. Thus, the concept of *capacity-versus-outage* [148] has to be invoked. That is, with any given rate R and channel realization $\mathbf{H}$, we associate a set of channels $\Lambda(R)$. This set is the largest possible set for which $C(\mathbf{H})$, i.e., the achieved rate for a realization of $\mathbf{H} \in \Lambda(R)$ satisfies $C(\mathbf{H}) < R$. Formally, we can write

$$\Lambda(R) \triangleq \{\mathbf{H} \in \mathbb{C}^{MM_R \times MM_T} : C(\mathbf{H}) < R\}. \tag{4.35}$$

The outage (or failure) probability P^{out} is then determined by

$$P^{\text{out}} = Prob\big(\mathbf{H} \in \Lambda(R)\big) = \int_{\Lambda(R)} p(\mathbf{H})\,\mathrm{d}\mathbf{H}, \tag{4.36}$$

where *Prob* denotes the probability of an event.

4.6 Achievable Outage Information Rates Associated to the Proposed Detector

In this section, we derive the achievable information rates in the sense of outage rates, associated to a receiver using the improved and the mismatched ML detection rules given respectively by (4.26) and (4.22). We start by deriving the instantaneous information achievable rates over MIMO-OFDM channels and then we derive the outage achievable rates, according to the above definition. The following results are a generalization of those we derived in [147] for the case of single-antenna MB-OFDM systems which are annexed in Appendix C and which can be easier to follow at a first time.

4.6.1 Instantaneous Achievable Rates of MB-MIMO-OFDM

We consider a MB-MIMO-OFDM transmission over the block-diagonal channel $\mathbf{H}$; the receiver knows an imperfect estimate $\widehat{\mathbf{H}}$ provided by some training symbols. This transmission is characterized by the likelihood function $W(\mathbf{Y}|\mathbf{S},\mathbf{H}) = \mathcal{CN}(\mathbf{HS},\mathbf{\Sigma}_{\mathrm{z}})$ (see (4.6)), which defines a mapping from the input symbols $\mathbf{S} \in \underline{\mathcal{S}}$ with distribution P_S, to the set of probability measures on the output alphabet $\mathbf{Y} \in \mathcal{Y}$. Since in Section 4.4 we used a per subcarrier detection rule, we consider the model (4.1) in our subsequent developments. The following theorem is a direct generalization to a MIMO-OFDM transmission of the theorem in [155] and provides the general procedure to compute the maximal achievable rates of a receiver using a given suboptimal detection rule in our transmission scenario.

Theorem 4.6.1. *Given a general detection metric $\mathcal{D}_{\mathcal{G}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \widehat{\mathbf{H}}_k)$ and a pair of channel and its estimate $(\mathbf{H}_k, \widehat{\mathbf{H}}_k)$, a MIMO-OFDM detector using the metric $\mathcal{D}_{\mathcal{G}}$ achieves the maximal information rate*

$$C_{\mathcal{D}_{\mathcal{G}}} = \max_{P_{\mathbf{S}}} \min_{V \in \mathcal{F}} I(\mathbf{S},\mathbf{Y}|\mathbf{\Upsilon}), \tag{4.37}$$

where the maximization is over all probability distributions P_S on $\underline{\mathcal{S}}$, and

$$I(\mathbf{S},\mathbf{Y}|\mathbf{\Upsilon}) = \int\int P_S(\mathbf{S})\, V_{Y|S}(\mathbf{Y}|\mathbf{S},\mathbf{\Upsilon}) \log_2 \frac{V_{Y|S}(\mathbf{Y}|\mathbf{S},\mathbf{\Upsilon})}{\int P_S(\mathbf{S}')\, V_{Y|S}(\mathbf{Y}|\mathbf{S}',\mathbf{\Upsilon})\, \mathrm{d}\mathbf{S}'}\, \mathrm{d}\mathbf{Y}\, \mathrm{d}\mathbf{S}, \tag{4.38}$$

denotes the mutual information functional [153]; in (4.37), the set $\mathcal{F}$ denotes the set of all likelihood functions $V(\boldsymbol{y}_k|\boldsymbol{s}_k,\mathbf{\Upsilon}_k) = \mathcal{CN}(\mathbf{\Upsilon}_k\boldsymbol{s}_k, \mathbf{\Sigma}_k)$ on $\mathcal{S}\times\mathcal{Y}$ that satisfy the following constraints[1]

$$(\mathrm{c}_{1,\mathrm{k}}):\ \mathrm{tr}\Big(\mathbb{E}_{P_{\mathbf{S}}}\Big[\mathbb{E}_V\big[\boldsymbol{y}_k\boldsymbol{y}_k^{\dagger}\big]\Big]\Big) = \mathrm{tr}\Big(\mathbb{E}_{P_{\mathbf{S}}}\Big[\mathbb{E}_W\big[\boldsymbol{y}_k\boldsymbol{y}_k^{\dagger}\big]\Big]\Big), \tag{4.39}$$

$$(\mathrm{c}_{2,\mathrm{k}}):\ \mathbb{E}_{P_{\mathbf{S}}}\Big[\mathbb{E}_V\big[\mathcal{D}_{\mathcal{G}}\big(\boldsymbol{s}_k,\boldsymbol{y}_k,\widehat{\mathbf{H}}_k\big)\big]\Big] \leq \mathbb{E}_{P_{\mathbf{S}}}\Big[\mathbb{E}_W\big[\mathcal{D}_{\mathcal{G}}\big(\boldsymbol{s}_k,\boldsymbol{y}_k,\widehat{\mathbf{H}}_k\big)\big]\Big], \tag{4.40}$$

[1] Our constraint $(\mathrm{c}_{1,\mathrm{k}})$ is different from that provided in [155], since here the channel noise is i.i.d.; consequently, we can only satisfy the equality of the matrix traces and not that of the covariance matrices.

for $k = 0, ..., M-1$, where the $(MM_R \times MM_T)$ block-diagonal matrix $\boldsymbol{\Upsilon}$ has the same structure than $\mathbf{H}$ in (4.5), i.e.,

$$\boldsymbol{\Upsilon} = \mathrm{diag}\Big(\begin{bmatrix} \boldsymbol{\Upsilon}_0 & \boldsymbol{\Upsilon}_1 & \cdots & \boldsymbol{\Upsilon}_{M-1} \end{bmatrix} \Big).$$

4.6.1.1 Case of Improved ML Metric

In order to solve the above constrained minimization problem for the metric $\mathcal{D}_{\mathcal{G}} = \mathcal{D}_{\mathcal{M}}$ of (4.26), we must find the M channel matrices $\boldsymbol{\Upsilon}_k \in \mathbb{C}^{M_R \times M_T}$ and the covariance matrices $\boldsymbol{\Sigma}_k = \sigma^2 \mathbb{I}_{M_R}$ defining the likelihood $V(\boldsymbol{y}_k | \boldsymbol{s}_k, \boldsymbol{\Upsilon}_k)$ that minimizes the relative entropy in (4.38). On the other hand, throughout this chapter we assume that the transmitter has not at his disposal the channel estimates, and consequently uniform power allocation is performed over subcarriers. Thus, we choose as the input distribution for all $\boldsymbol{s}_k$, $P_S = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_P)$ with $\boldsymbol{\Sigma}_P = \bar{P}\mathbb{I}_{M_T}$. First, we find the constraint set $\mathcal{F}$, given by ($\mathrm{c}_{1,\mathrm{k}}$) and ($\mathrm{c}_{2,\mathrm{k}}$) for $k = 0, ..., M-1$, and then we factorize the matrices $\mathbf{H}_k$ to solve the minimization problem. Before this, to find the expression of the constraint ($\mathrm{c}_{2,\mathrm{k}}$), we need the following result.

Lemma 4.6.2. *Let $\mathbf{A} \in \mathbb{C}^{M_R \times M_T}$ be an arbitrary matrix and $\mathbf{x} \in \mathbb{C}^{M_T \times 1}$ be a random vector distributed as $\mathcal{CN}(\mathbf{0}, \bar{P}\mathbb{I}_{M_T})$. For every real positive constants K_1, K_2, the following equality holds:*

$$\mathbb{E}_{\mathbf{x}}\left[\frac{\|\mathbf{Ax}\|^2 + K_1}{\|\mathbf{x}\|^2 + K_2}\right] = \frac{\|\mathbf{A}\|_F^2}{M_T} + \left(\frac{K_1}{K_2} - \frac{\|\mathbf{A}\|_F^2}{M_T}\right)\left(\frac{K_2}{\bar{P}}\right)^{M_T} \exp\left\{\frac{K_2}{\bar{P}}\right\} \Gamma\left(1 - M_T, K_2/\bar{P}\right), \quad (4.41)$$

where $\Gamma(a,x) = \displaystyle\int_x^{+\infty} u^{a-1} e^{-u}\, \mathrm{d}u$ denotes the upper incomplete gamma function.

Proof. See Appendix B. □

From Lemma 4.6.2 and after some algebra, one sees that the constraints (4.39) and (4.40) become equivalent to

$$(\mathrm{c}_{1,\mathrm{k}}): \quad \mathrm{tr}\Big(\boldsymbol{\Upsilon}_k \boldsymbol{\Sigma}_P \boldsymbol{\Upsilon}_k^\dagger + \boldsymbol{\Sigma}_k\Big) = \mathrm{tr}\Big(\mathbf{H}_k \boldsymbol{\Sigma}_P \mathbf{H}_k^\dagger + \boldsymbol{\Sigma}_{z,k}\Big), \quad (4.42)$$

$$(\mathrm{c}_{2,\mathrm{k}}): \quad \|\boldsymbol{\Upsilon}_k + a_{\mathcal{M}} \widehat{\mathbf{H}}_k\|_F^2 \leq \|\mathbf{H}_k + a_{\mathcal{M}} \widehat{\mathbf{H}}_k\|_F^2 + \mathrm{C}_k, \quad (4.43)$$

where

$$\begin{aligned} a_{\mathcal{M}} &= \delta\left(\delta\sigma_{\mathcal{E}}^2 \bar{P} - \lambda\sigma_z^2\right)\left[M_T \delta\sigma_{\mathcal{E}}^2 \lambda \bar{P} + \lambda\sigma_z^2 - \delta\sigma_{\mathcal{E}}^2 \bar{P}\right]^{-1}, \\ \mathrm{C}_k &= M_T \lambda\left[\|\mathbf{H}_k\|_F^2 - \|\boldsymbol{\Upsilon}_k\|_F^2 + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{z,k} - \boldsymbol{\Sigma}_k)}{\bar{P}}\right]\left[1 - \frac{\sigma_z^2}{\delta\bar{P}\sigma_{\mathcal{E}}^2}\lambda - M_T\lambda\right]^{-1}, \\ \lambda &= \left(\frac{\sigma_z^2}{\delta\bar{P}\sigma_{\mathcal{E}}^2}\right)^{M_T - 1} \exp\left\{\frac{\sigma_z^2}{\delta\bar{P}\sigma_{\mathcal{E}}^2}\right\} \Gamma\left(1 - M_T, \frac{\sigma_z^2}{\delta\bar{P}\sigma_{\mathcal{E}}^2}\right). \end{aligned}$$

By expressing the MIMO-OFDM mutual information of (4.33) and using expressions (4.42) and (4.43), the minimization (4.37) becomes

$$C_{\mathcal{M}}(\mathbf{H},\widehat{\mathbf{H}}) = \begin{cases} \min\limits_{\boldsymbol{\Upsilon}_0,\ldots,\boldsymbol{\Upsilon}_{M-1}} & \dfrac{1}{M}\displaystyle\sum_{k=0}^{M-1}\log_2\det\left\{\mathbb{I}_{M_R}+\bar{P}\boldsymbol{\Upsilon}_k\boldsymbol{\Upsilon}_k^{\dagger}\boldsymbol{\Sigma}_k^{-1}\right\} \\ \text{subject to} & \|\boldsymbol{\Upsilon}_k + a_{\mathcal{M}}\widehat{\mathbf{H}}_k\|_F^2 \leq \|\mathbf{H}_k + a_{\mathcal{M}}\widehat{\mathbf{H}}_k\|_F^2 + \mathrm{C}_k \end{cases} \tag{4.44}$$

where $\boldsymbol{\Sigma}_k$ must be chosen such that $\mathrm{tr}\big(\bar{P}\boldsymbol{\Upsilon}_k\boldsymbol{\Upsilon}_k^{\dagger}+\boldsymbol{\Sigma}_k\big) = \mathrm{tr}\big(\bar{P}\mathbf{H}_k\mathbf{H}_k^{\dagger}+\boldsymbol{\Sigma}_{z,k}\big)$.

In order to obtain a simpler and even more tractable expression for (4.44), we consider the singular value decomposition (SVD) of the matrix $\mathbf{H}_k = \mathbf{U}_k\,\mathrm{diag}(\underline{\lambda}_k)\mathbf{V}_k^{\dagger}$ with $\underline{\lambda}_k = [\lambda_{k,1},\ldots,\lambda_{k,M_R}]^T$. We also introduce the following definitions.

Let $\mathrm{diag}(\underline{\mu}_k)$ be a diagonal matrix such that $\mathrm{diag}(\underline{\mu}_k) = \mathbf{U}_k^{\dagger}\boldsymbol{\Upsilon}_k\mathbf{V}_k$ with its diagonal elements given by the vector $\underline{\mu}_k = [\mu_{k,1},\ldots,\mu_{k,M_R}]^T$. We define $\widetilde{\mathbf{H}}_k = \mathbf{U}_k^{\dagger}\widehat{\mathbf{H}}_k\mathbf{V}_k$ and $\widetilde{\mathbf{h}}_k = \mathrm{diag}(\widetilde{\mathbf{H}}_k)$ gathers the diagonal elements of $\widetilde{\mathbf{H}}_k$. Finally, let $b_{\mathcal{M},k} = \|\mathbf{H}_k + a_{\mathcal{M}}\widehat{\mathbf{H}}_k\|_F^2 - a_{\mathcal{M}}^2\big(\|\widetilde{\mathbf{H}}_k\|_F^2 - \|\widetilde{\mathbf{h}}_k\|^2\big)$. By using the above definitions and after some algebraic manipulations, the optimization (4.44) can be written as

$$C_{\mathcal{M}}(\mathbf{H},\widehat{\mathbf{H}}) = \begin{cases} \min\limits_{\underline{\mu}_0,\ldots,\underline{\mu}_{M-1}} & \dfrac{1}{M}\displaystyle\sum_{k=0}^{M-1}\sum_{i=0}^{M_R-1}\log_2\left(1+\frac{\bar{P}|\mu_{k,i}|^2}{\sigma_k^2(\underline{\mu}_k)}\right) \\ \text{subject to} & \|\underline{\mu}_k + a_{\mathcal{M}}\widetilde{\mathbf{h}}_k\|^2 \leq b_{\mathcal{M},k} \end{cases} \tag{4.45}$$

with

$$\sigma_k^2(\underline{\mu}_k) = \frac{\bar{P}}{M_R}\left(\|\underline{\lambda}_k\|^2 - \|\underline{\mu}_k\|^2\right) + \sigma_z^2, \quad \text{for } \; k = 0,\ldots,M-1.$$

We note that the constraint sets in the minimization (4.45), which corresponds to the set of vectors $\{\underline{\mu}_k \in \mathbb{C}^{M_T\times 1} : \|\underline{\mu}_k + a_{\mathcal{M}}\widetilde{\mathbf{h}}_k\|^2 \leq b_{\mathcal{M}}\}$ are closed convex polyhedral sets. Thus, the minimum in (4.45) is attainable at the border of the sets given by the equality $\|\underline{\mu}_k + a_{\mathcal{M}}\widetilde{\mathbf{h}}_k\|^2 = b_{\mathcal{M}}$ (cf. [156]). On the other hand, for every vector $\underline{\mu}_k$ such that $\|\underline{\mu}_k\|^2 \leq \|\underline{\lambda}_k\|^2$, we observe that the expression (4.45) is a monotone increasing function of the square norm of $\underline{\mu}_k$. Thus, the problem simplifies to the search of the vector $\underline{\mu}_{\mathcal{M},k}^{\mathrm{opt}}$ resulting from the minimization of $\|\underline{\mu}_k\|^2$ under the constraint sets $\{\underline{\mu}_k \in \mathbb{C}^{M_T\times 1} : \|\underline{\mu}_k + a_{\mathcal{M}}\widetilde{\mathbf{h}}_k\|^2 = b_{\mathcal{M}}\}$. This becomes a classical convex minimization problem that can be solved by using Lagrange multipliers.

Finally, one obtain the information rates achieved by a receiver using the detection rule (4.26) as follows:

$$\boxed{C_{\mathcal{M}}(\mathbf{H},\widehat{\mathbf{H}}) = \frac{1}{M}\sum_{k=0}^{M-1}\log_2\det\left\{\mathbb{I}_{M_R}+\frac{\bar{P}\boldsymbol{\Upsilon}_{\mathrm{opt},k}\boldsymbol{\Upsilon}_{\mathrm{opt},k}^{\dagger}}{\sigma_k^2\big(\underline{\mu}_{\mathcal{M},k}^{\mathrm{opt}}\big)}\right\}} \tag{4.46}$$

where

$$\boldsymbol{\Upsilon}_{\text{opt},k} = \mathbf{U}_k \,\text{diag}\left(\underline{\mu}^{\text{opt}}_{\mathcal{M},k}\right)\mathbf{V}_k^{\dagger}$$

with

$$\underline{\mu}^{\text{opt}}_{\mathcal{M},k} = \begin{cases} \left(\dfrac{\sqrt{b_{\mathcal{M},k}}}{\|\widetilde{\mathbf{h}}_k\|} - |a_{\mathcal{M}}|\right)\widetilde{\mathbf{h}}_k & \text{if } b_{\mathcal{M},k} \geq 0 \\ \mathbf{0} & \text{otherwise} \end{cases}$$

and

$$\sigma_k^2\big(\underline{\mu}^{\text{opt}}_{\mathcal{M},k}\big) = \frac{\bar{P}}{M_R}\left(\|\underline{\lambda}_k\|^2 - \|\underline{\mu}^{\text{opt}}_{\mathcal{M},k}\|^2\right) + \sigma_z^2.$$

4.6.1.2 Case of Mismatched ML Metric

In order to compare the achievable rates (4.46) to those provided by the classical mismatched ML detector, we compute the achievable rates C_{MM} associated to $\mathcal{D}_{\text{MM}}$ of (4.22). Following the same steps as above and under similar assumptions, we obtain

$$C_{\text{MM}}(\mathbf{H},\widehat{\mathbf{H}}) = \begin{cases} \min\limits_{\boldsymbol{\Upsilon}_0,\ldots,\boldsymbol{\Upsilon}_{M-1}} & \dfrac{1}{M}\displaystyle\sum_{k=0}^{M-1} \log_2 \det\left\{\mathbb{I}_{M_R} + \bar{P}\boldsymbol{\Upsilon}_k\boldsymbol{\Upsilon}_k^{\dagger}\boldsymbol{\Sigma}_k^{-1}\right\} \\ \text{subject to} & \mathbb{R}e\Big(\text{tr}(\bar{P}\mathbf{H}_k\widehat{\mathbf{H}}_k^{\dagger})\Big) \leq \mathbb{R}e\Big(\text{tr}(\bar{P}\boldsymbol{\Upsilon}_k\widehat{\mathbf{H}}_k^{\dagger})\Big) \end{cases} \tag{4.47}$$

where $\boldsymbol{\Sigma}_k$ must be chosen such that $\text{tr}\big(\bar{P}\boldsymbol{\Upsilon}_k\boldsymbol{\Upsilon}_k^{\dagger} + \boldsymbol{\Sigma}_k\big) = \text{tr}\big(\bar{P}\mathbf{H}_k\mathbf{H}_k^{\dagger} + \sigma_{z,k}\big)$.
The resulting achievable rates are given by

$$\boxed{C_{\text{MM}}(\mathbf{H},\widehat{\mathbf{H}}) = \frac{1}{M}\sum_{k=0}^{M-1}\log_2\det\left\{\mathbb{I}_{M_R} + \frac{\bar{P}\boldsymbol{\Upsilon}_{\text{opt},k}\boldsymbol{\Upsilon}_{\text{opt},k}^{\dagger}}{\sigma_k^2\big(\underline{\mu}^{\text{opt}}_{\text{MM},k}\big)}\right\}} \tag{4.48}$$

where

$$\boldsymbol{\Upsilon}_{\text{opt},k} = \mathbf{U}\,\text{diag}\left(\underline{\mu}^{\text{opt}}_{\text{MM},k}\right)\mathbf{V}^{\dagger}$$

with

$$\underline{\mu}^{\text{opt}}_{\text{MM},k} = \frac{\mathbb{R}e\Big(\text{tr}\big(\text{diag}(\underline{\lambda}_k)^{\dagger}\widetilde{\mathbf{h}}_k\big)\Big)}{\|\widetilde{\mathbf{h}}_k\|^2}\,\widetilde{\mathbf{h}}_k$$

and

$$\sigma_k^2\big(\underline{\mu}^{\text{opt}}_{\text{MM},k}\big) = \frac{\bar{P}}{M_R}\left(\|\underline{\lambda}_k\|^2 - \|\underline{\mu}^{\text{opt}}_{\text{MM},k}\|^2\right) + \sigma_z^2.$$

4.6.2 Evaluation of Outage Rates Under Imperfect Channel Estimation

In Subsection 4.5.3, we presented the definition of the classical notion of outage capacity, which implicitly assumes that the true channel is available at the decoder. In contrast, when this knowledge is not available at the receiver, i.e., only the channel estimate is available, we have

to invoke the notion of *estimation-induced outage capacity*, which has different properties, as explained in [157]. Following this approach, we make use of the posterior channel distribution of (4.15) that characterizes our channel estimation process, in order to calculate the outage probability $P_{\mathcal{M}}^{\text{out}}$ (associated with the metric $\mathcal{D}_{\mathcal{M}}$), for an outage rate $R \geq 0$ and an estimated channel $\widehat{\mathbf{H}}$ as

$$P_{\mathcal{M}}^{\text{out}}(R, \widehat{\mathbf{H}}) = Prob\big(\mathbf{H} \in \Lambda_{\mathcal{M}} \big| \widehat{\mathbf{H}}\big) = \int_{\{\mathbf{H} \in \Lambda_{\mathcal{M}}(R, \widehat{\mathbf{H}})\}} p(\mathbf{H}|\widehat{\mathbf{H}})\, \mathrm{d}\mathbf{H}, \tag{4.49}$$

with $\Lambda_{\mathcal{M}}(R, \widehat{\mathbf{H}}) = \big\{\mathbf{H} \in \mathbb{C}^{MM_R \times MM_T} : C_{\mathcal{M}}(\mathbf{H}, \widehat{\mathbf{H}}) < R\big\}$. Note that the definition (4.49) differs from the classical definition of equation (4.36) (which assumes perfect CSIR) in that the outage probability is defined by using the a posteriori pdf $p(\mathbf{H}|\widehat{\mathbf{H}})$ instead of the pdf $p(\mathbf{H})$. More precisely, here an outage event is only induced by imperfect channel estimation.
Using this definition, the outage rate for an outage probability of γ is given by

$$C_{\mathcal{M}}^{\text{out}}(\gamma, \widehat{\mathbf{H}}) = \sup_{R} \big\{R \geq 0 : P_{\mathcal{M}}^{\text{out}}(R, \widehat{\mathbf{H}}) \leq \gamma\big\}. \tag{4.50}$$

Since the outage rates in (4.50) still depend on the random channel estimate $\widehat{\mathbf{H}}$, we consider as a performance measure, the expected outage rates over all channel estimates as

$$\overline{C}_{\mathcal{M}}^{\text{out}}(\gamma) = \mathbb{E}_{\widehat{\mathbf{H}}}\big[C_{\mathcal{M}}^{\text{out}}(\gamma, \widehat{\mathbf{H}})\big]. \tag{4.51}$$

By using (4.48) and similar steps as above, we can derive the averaged outage rates $\overline{C}_{\text{MM}}^{\text{out}}(\gamma)$ associated to the mismatched metric (4.22).

The achievable outage rates of the improved and mismatched ML receiver are upper bounded by the outage rates provided by a *theoretical* decoder (i.e., the best decoder in the presence of channel estimation errors) derived in [157]. For MIMO-OFDM systems, the expected outage rates of the theoretical decoder is expressed as [146]

$$\overline{C}_{\mathcal{T}}^{\text{out}}(\gamma) = \mathbb{E}_{\widehat{\mathbf{H}}}\big[C_{\mathcal{T}}^{\text{out}}(\gamma, \widehat{\mathbf{H}})\big], \tag{4.52}$$

where the outage rates $C_{\mathcal{T}}^{\text{out}}(\gamma, \widehat{\mathbf{H}})$ are computed by using the achievable rates

$$C_{\mathcal{T}}(\mathbf{H}) = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \det \left\{ \mathbb{I}_{M_R} + \frac{\bar{P}\, \mathbf{H}_k \mathbf{H}_k^{\dagger}}{\sigma_z^2} \right\} \tag{4.53}$$

in expressions similar to (4.49) and (4.50).

4.7 Simulation Results and Discussions

In this section, we provide numerical results to evaluate the performance provided by the proposed detector in the presence of channel estimation errors, in comparison with more classical

approaches. We focus on the impact of imperfect channel estimation on receiver BER and the achievable outage rates associated to improved and mismatched detectors. The situation of interest is a BICM scheme, involving the rate 1/2 NRNSC code of constraint length 3 defined in octal form by the polynomials $(5,7)_8$. Data symbols belong to a 16-QAM constellation with Gray or set-partition (SP) labeling, and the impact of this labeling is discussed below, since it may enlarge or reduce the improvement brought by the improved metric.

We employ the multipath channel model CM1 specified in the IEEE802.15.3a channel modeling subcommittee report [65]. The choice of the CM1 channel is motivated by the fact that for this channel, the CP is longer than the maximum channel delay spread, and consequently the ISI can be neglected. The CM1 channel impulse response corresponding to a single subband has $L = 32$ multipath components and is power-normalized. The path gain coefficients $\alpha_{ij}(l)$ (see (4.3)) for different transmit/receive antennas i, j are generated independently. Channel coefficients are kept constant during each frame and changed to new independent realizations from one frame to the next frame. Following the IEEE802.15.3a standard proposal [4], the IFFT is calculated on 128 points and the bandwidth of each subband is 528 MHz. Thus, each OFDM tone has a bandwidth of 528 MHz/128=4.13 MHz. The number of subcarriers is respectively equal to 100 and 50 for single- and multi-antenna MB-OFDM configurations. Throughout the simulations, each frame is assumed to consist of three OFDM symbols. At each time-slot, multibanding is performed by sending OFDM symbols over different subbands according to the time-frequency code {1,2,3,1,2,3,...}.

The channel is assumed to be estimated at the receiver with a number of N_P training symbols so that the channel estimate follows the linear model (4.10). We use mutually-orthogonal pilot sequences for channel estimation and the average pilot-symbol power is set equal to the average data-symbol power.

The interleaver is pseudo-random, operating over the entire frame of size $N_I = MM_TB$ bits (excluding pilots, obviously). Moreover, unless otherwise mentioned, the number of decoding iterations is set to 4. The average signal-to-noise ratio (SNR) is considered in the form of E_b/N_0 in dB and includes the antenna array gain at receiver, M_R. We assume that the noise variance is known at the receiver.

4.7.1 Bit Error Rate Analysis

Case of Single-antenna Multiband OFDM: Let us first address the case of the single-antenna MB-OFDM iterative MAP detection with a 16-QAM modulation and Gray labeling. It can be observed from Fig. 4.3 that for $N_P = 1$ (the shortest possible training sequence), the improvement in terms of required E_b/N_0 in order to attain a given BER is about 2 dB, compared

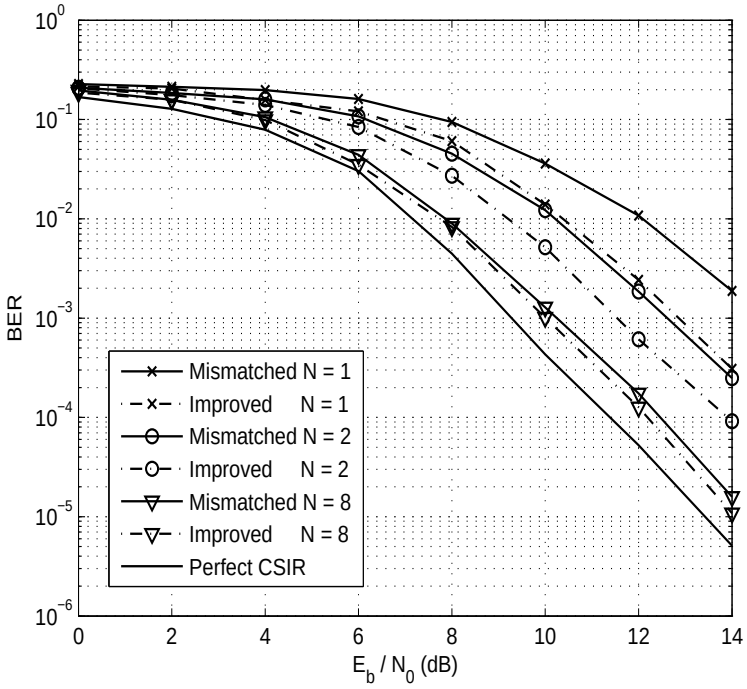

Figure 4.3: BER performance of improved and mismatched MAP detectors over the CM1 channel, training sequence lengths $N_P \in \{1, 2, 8\}$, 16-QAM modulation with Gray labeling.

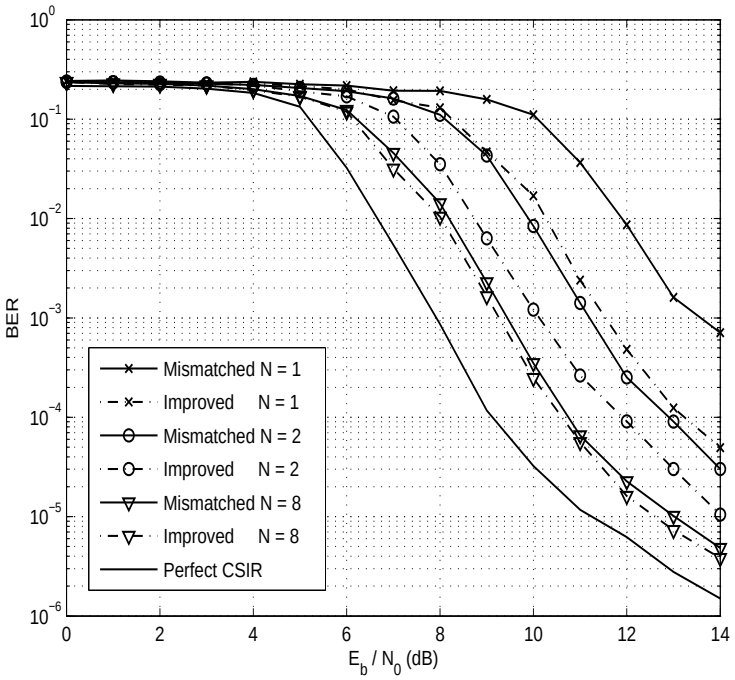

Figure 4.4: BER performance of the improved and mismatched MAP detectors over the CM1 channel, training sequence lengths $N_P \in \{1, 2, 8\}$, 16-QAM modulation with SP labeling.

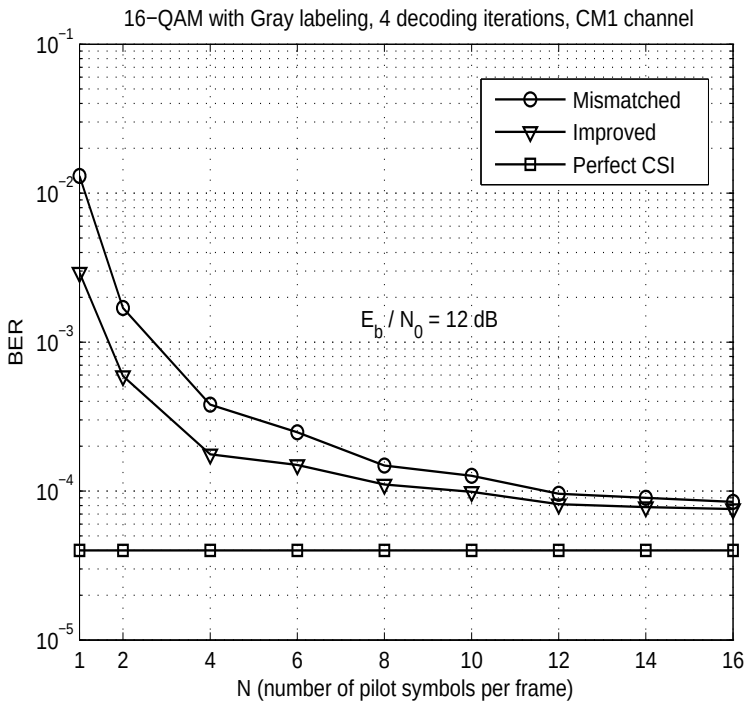

Figure 4.5: Reduction of the number of training symbols at $E_b/N_0 = 12$ dB over the CM1 channel, 16-QAM modulation with Gray labeling .

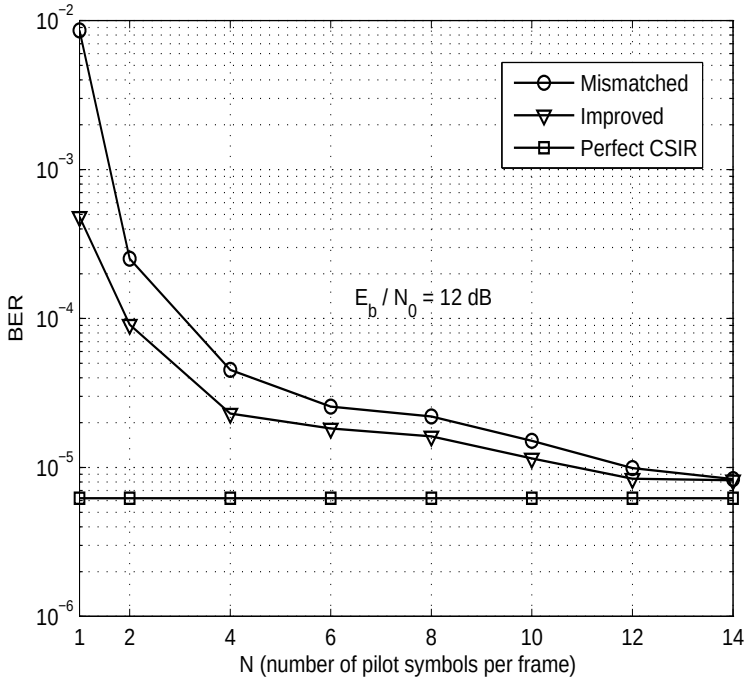

Figure 4.6: Reduction of the number of training symbols at $E_b/N_0 = 12$ dB over the CM1 channel, 16-QAM modulation with SP labeling.

to the mismatched situation, while the perfect channel knowledge would even be about 3 dB better. Obviously, it is also observed that these quantities are reduced (the performance of both the mismatched decoder and the improved one get closer to the perfect channel one) when increasing the length of the training sequence. Note that the performance of the improved receiver with 1 pilot is very close to that of the mismatched receiver with 2 pilots.

Let us now analyze the impact of labeling on the performance. It is well known that when iterative decoding is used, Gray labeling may not be the best choice. This was recognized by Ritcey *et al.* [158, 159] in which the authors consider the use of a SP labeling. Actually, this labeling can substantionally improve the performance of a BICM compared to Gray labeling when iterative decoding is used. This is illustrated in Fig. 4.4 for the case of 16-QAM and SP labeling on the CM1 channel. Before analyzing the performance of the improved detector, note that at high SNR (here at $E_b/N_0 = 13$ dB for the mismatched detector and $N_P = 1$ and at $E_b/N_0 = 9$ dB with perfect CSIR), the slope of the curve changes according to the free distance of the code. This happens when the iterative decoder attains the performance of the ML decoder [160]. A first observation is that the distance between the mismatched detector and the perfect channel knowledge is even larger (about 6 dB for $N_P = 1$ at a BER of 7×10^{-4}). Another observation is that, even if the global performance is largely improved by using the SP labeling when perfect channel knowledge is available, the difference between SP and Gray labeling is not very large with the mismatched decoder, i.e., the sensitivity of the iterative decoder to the channel knowledge seems to be larger for SP labeling. However, the use of the improved metric allows to recover most of the improvement, since, even with a single training sample ($N_P = 1$), the BER is improved with SP compared to Gray labeling by about 0.9 dB at a BER of 5×10^{-4}. In other words, iterative decoding with SP labeling benefits more from the improved metric than the one with Gray labeling. Otherwise, similar conclusions hold between the SP-labeling curves.

Figures 4.5 and 4.6 show the BER performance versus the number of pilot symbols N_P at a fixed E_b/N_0 of 12 dB for 16-QAM with Gray and SP labeling, respectively. This allows to evaluate the number of training sequences necessary to achieve a fixed BER. For instance, from Fig. 4.5, we observe that the improved detector requires 10 pilots per frame to achieve a BER of 10^{-4} at $E_b/N_0 = 12$ dB while the mismatched detector attains this BER for 12 training symbols. Our results show that the improved detector outperforms the mismatched detector especially when only a few pilots are dedicated for channel estimation. Actually, for the displayed transmission configuration, the performance loss due to the mismatched receiver with respect to the improved receiver becomes insignificant for $N_P \geq 12$ (about 11 % of the overall frame of pilot and data symbols).

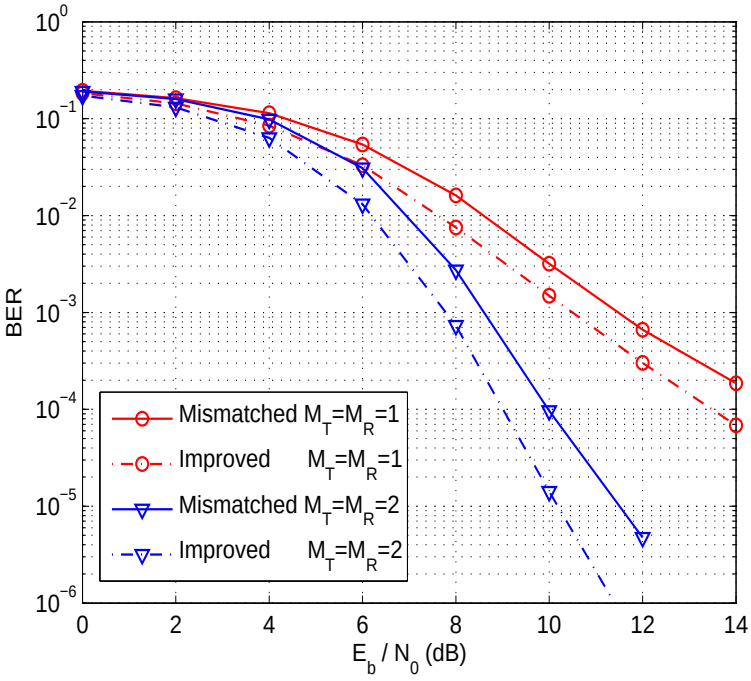

Figure 4.7: BER performance for single-antenna and MIMO MB-OFDM transmission; improved and mismatched MAP detectors; CM1 channel; training sequence lengths $N_P = 2$; 16-QAM modulation with Gray labeling.

Case of Multiband MIMO-OFDM: First, we compare in Fig. 4.7, the BER performance of single- and multi-antenna MB-OFDM transmission with 16-QAM constellation, and Gray labeling where the channel is estimated by sending 2 pilot symbols. We observe an important performance improvement achieved by employing multiple antennas for MB-OFDM transmission. This can be seen by comparing the slope of the BER curves of improved and mismatched detectors for one and two transmit/receive antennas.

We now compare the BER performance of the improved and mismatched MAP detectors under imperfect channel estimation. Let us first address the case of BICM iterative decoding with 16-QAM and Gray labeling. Figure 4.8 depicts the BER curves of mismatched and improved receivers for $M_T = M_R = 2$ (refered as the 2×2 MB-MIMO-OFDM system), for different training sequence lengths $N_P \in \{2, 4, 8\}$. As a reference, we have also presented the BER curve in the case of perfect CSIR. We notice that the required SNR to attain the BER of 10^{-5} with $N_P = 2$ is reduced by about 1.3 dB for the improved detector, as compared to the mismatched detector. By increasing N_P, the channel estimation errors become less important and the performance difference between the two detectors decreases: the achieved gain in SNR at BER $= 10^{-5}$ is about 0.5 dB and 0.2 dB, for $N_P = 4$ and $N_P = 8$, respectively. Actually, the performance loss of the mismatched receiver with respect to the improved receiver becomes insignificant for $N_P \geq 8$.

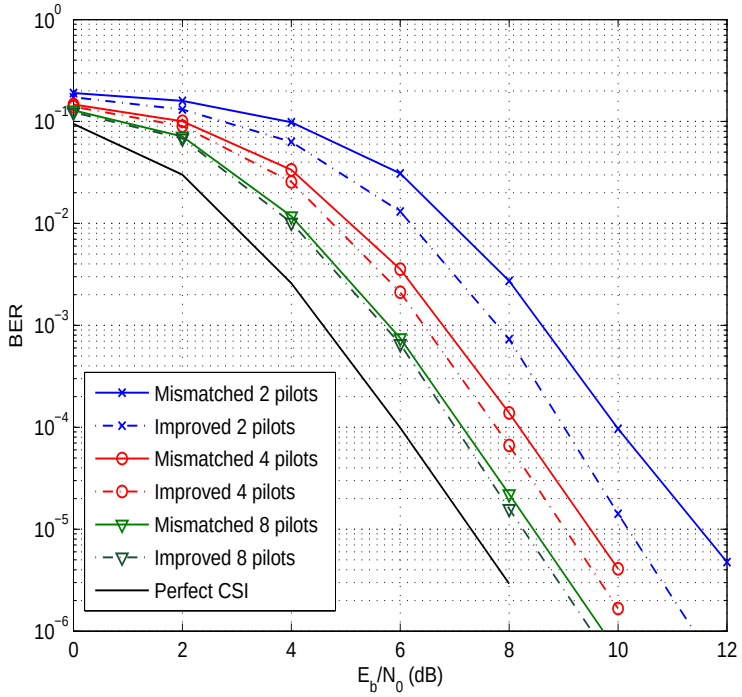

Figure 4.8: BER performance of improved and mismatched MAP detectors over the CM1 channel; 2×2 MB-MIMO-OFDM; training sequence lengths $N_P \in \{2, 4, 8\}$; 16-QAM modulation with Gray labeling.

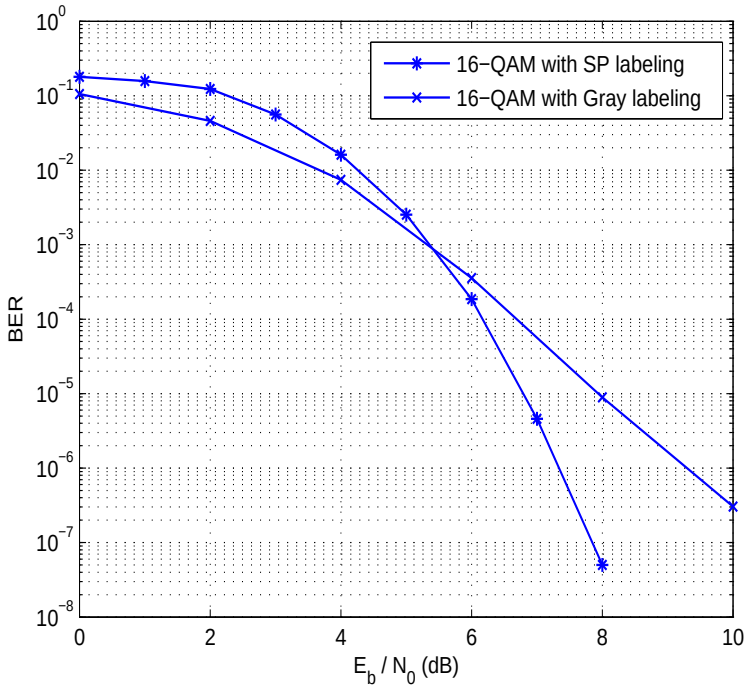

Figure 4.9: BER performance of iterative MAP detection with Gray and SP labeling under perfect CSIR; 16-QAM modulation; 2×2 MB-MIMO-OFDM; CM1 channel.

Let us now consider a 2×2 MB-MIMO-OFDM system with 16-QAM modulation and SP labeling. We compare in Fig. 4.9, the BER curve obtained with Gray and SP labeling under perfect CSIR. This figure clearly shows the advantage that SP labeling has above an SNR threshold (here for $E_b/N_0 = 5.4$ dB). In Fig. 4.10, we show the BER curves when the channel is estimated by different numbers of pilot symbols $N_P \in \{2, 4, 8\}$. The slope of the curves shows that above a given SNR, the global performance is largely improved thanks to SP labeling. In particular, we observe a considerable amount of SNR reduction for the improved receiver when one or two training symbols are devoted to channel estimation. However, we note that the mismatched detector with SP labeling seems to be more robust to channel estimation errors as compared to the single-antenna MB-OFDM system. In fact, we see that the performance gain provided by the improved detector with Gray and SP labelings are very close.

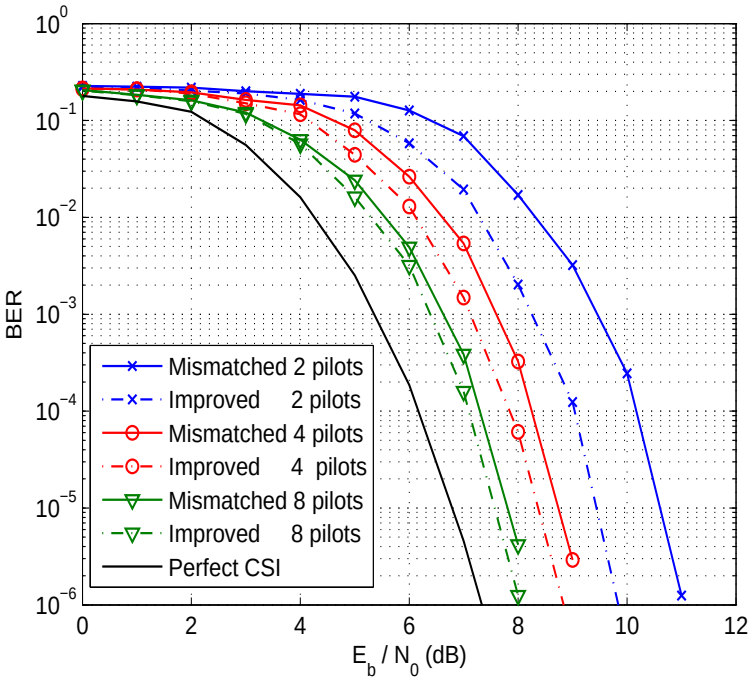

Figure 4.10: BER performance of improved and mismatched MAP detectors over the CM1 channel; 2×2 MB-MIMO-OFDM; training sequence lengths $N_P \in \{2, 4, 8\}$; 16-QAM modulation with set-partition labeling.

Finally, we have contrasted in Fig. 4.11, the BER curves of the improved and mismatched detectors for different decoding iterations. This reveals another interesting feature of the improved detector, namely, the reduction of receiver iterations needed to attain a target BER. As observed, the BER curve of the improved receiver after 2 iterations is very close to that of the mismatched receiver after 4 iterations. This can be exploited for reducing the receiver complexity and latency.

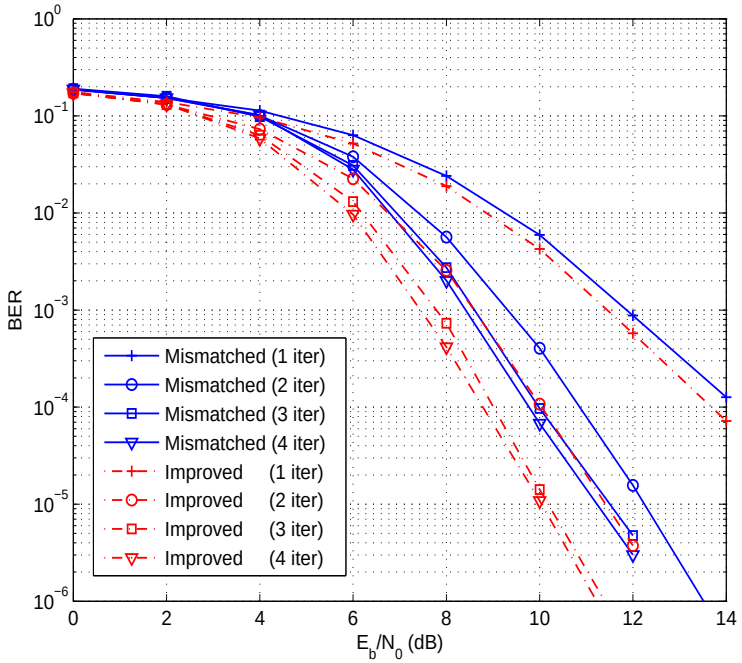

Figure 4.11: Reduction of the number of receiver iterations by using the improved MAP detector; 2×2 MB-MIMO-OFDM; CM1 channel; training sequence lengths $N_P = 2$; 16-QAM modulation with Gray labeling.

4.7.2 Achievable Outage Rates Analysis

Case of Single-antenna Multiband OFDM: We now analyze the achievable outage rates provided by a receiver based on the improved and mismatched detection techniques for MB-OFDM under imperfect CSIR. Channel statistics are assumed according to the uncorrelated i.i.d. Rayleigh fading. Note that due to the very complex (if not impossible) characterization of the integration boundaries in (4.49), we have employed empirical methods to calculate the expected outage rates of both single- and multi-antenna MB-OFDM systems. Figure 4.12 shows the expected outage rates (in bits per channel-use) versus the SNR, obtained by adopting mismatched and improved detection approaches. The outage probability has been fixed to $\gamma = 0.01$ and the channel is estimated by sending $N_P = 1$ pilot per frame. For comparison, we also display the ergodic capacity as well as the upper bound on the expected outage rates provided by the theoretical decoder, given by equations (4.34) and (4.52), respectively. The figure clearly shows the sub-optimality of mismatched detection in terms of expected outage rates, compared to the rates provided by the theoretical decoder. It can be seen that the mismatched outage rate is about 4.8 dB (at a mean outage rate of 5 bits per channel-use) of SNR far from the rates achieved by the theoretical decoder. We note that by adopting the improved receiver, this SNR gap is reduced by about 1.8 dB.

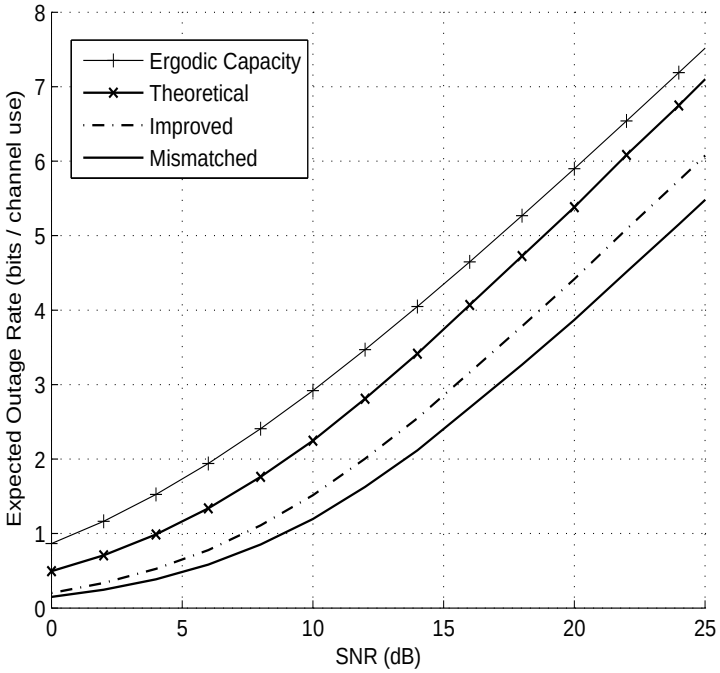

Figure 4.12: Expected outage rates of MB-OFDM transmission versus SNR for $M = 16$ subcarriers; $N_P = 1$ pilot per frame and different detection approaches; outage probability $\gamma = 0.01$; i.i.d. Rayleigh fading channel.

In Fig. 4.13, we compare the achieved throughputs of the improved and mismatched receivers with respect to the number of training symbols at fixed SNR values of 10 and 15 dB. For each case, we also display the corresponding upper bound on the achievable outage rates provided by the theoretical decoder and the ergodic capacity. As it can be seen, the improved detector requires fewer pilot symbols in order to provide a prescribed mean outage rate. For instance, the gain in the number of channel-uses for pilot transmission is 2 at a mean outage rate of 3.5 bits per channel-use.

Case of Multiband MIMO-OFDM: Figure 4.14 shows the expected outage rates (in bits per channel-use) of both the improved (expression (4.51)) and the mismatched ML detector corresponding to the transmission of one OFDM symbol with $M = 16$ subcarriers and a 2×2 MB-MIMO-OFDM system. For comparison, we have also shown the upper bound on these achievable outage rates (expression (4.52)) and the ergodic capacity with perfect channel knowledge (expression (4.34)). The MIMO-OFDM channel is estimated by sending 2 pilot symbols per frame and the outage probability has been fixed to $\gamma = 0.01$. It can be observed that at a mean outage rate of 8 bits per channel-use, the achievable rates of the mismatched ML detector is about 5 dB of SNR far from those provided by the theoretical decoder. We note that by

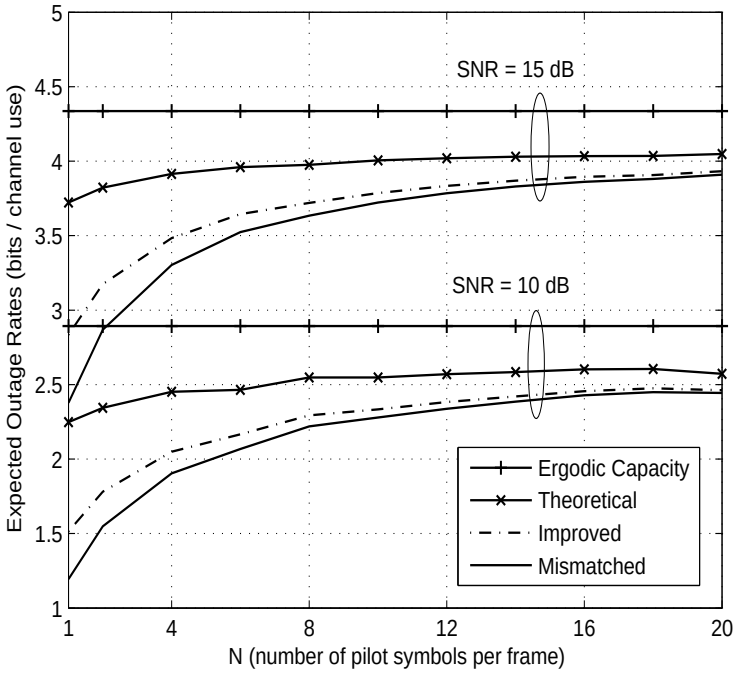

Figure 4.13: Expected outage rates of MB-OFDM transmission versus N_P for $M = 16$ subcarriers and different detection approaches at SNRs of 10 and 15 dB; outage probability $\gamma = 0.01$; i.i.d. Rayleigh fading channel.

using the improved ML detector, higher rates are obtained for any considered SNR value and the aforementioned SNR gap is reduced by about 1.8 dB.

Similar plots are shown in Fig. 4.15 in the case of a 4×4 MIMO-OFDM channel estimated by sending 4 training symbols. Again, it can be observed that the improved detector achieves higher rates than the mismatched detector. However, we note that the SNR increase (at a given mean outage rate) induced by using the mismatched detector rather than the improved detector is less than that obtained for a 2×2 MIMO-OFDM channel, as compared with Fig. 4.14. For example, the performance degradation due to the mismatched detector has decreased to less than 1 dB at a mean outage rate of 14 bits per channel use. This can be explained by noting that for estimating the $M_R M_T$ entries of the channel matrix $\mathbf{H}_k$, we must send $N_P \geq M_T$ (here $N_P = M_T = 4$) pilots per frame[2]; this yields a more accurate estimate of the channel and consequently brings closer the performance of the two detection metrics according to (4.27). This observation is in consistence with that presented in [37] where it is reported that the performance degradation due to imperfect channel estimation can be reduced by increasing the number of antennas. Note that, although the improved detector provides a significant gain compared to the mismatched approach, it is still sub-optimal since its achievable rates are about

[2]Satisfying the condition $N_P \geq M_T$ is necessary to ensure the channel matrix identifiability.

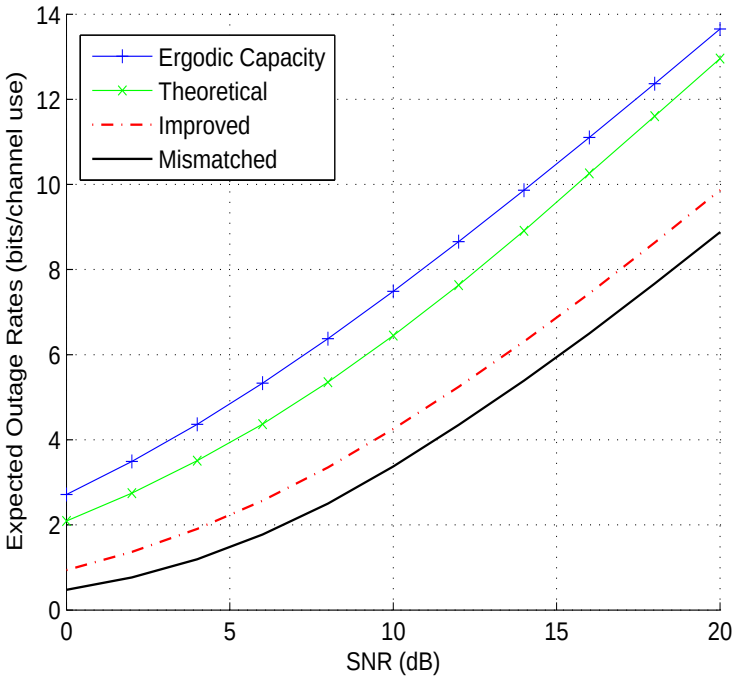

Figure 4.14: Expected outage rates of a 2×2 MB-MIMO-OFDM transmission versus SNR for $M = 16$ subcarriers and $N_P = 2$ pilots per frame and different detection approaches; outage probability $\gamma = 0.01$; CM1 channel.

4 dB of SNR far from those provided by the theoretical decoder. Thus, a still open question is how the gap to the estimation induced outage capacities could be filled.

Figures 4.16 and 4.17 show the expected outage rates with respect to the number of pilots at a fixed SNR of 15 dB for a 2×2 and 4×4 configuration, respectively. We observe that the improved ML detector reduces the number of required pilot symbols at a given mean outage rate. For instance, when $M_T = M_R = 2$, the gain in the number of channel-uses for pilot transmission is about 4 pilots at a mean outage rate of 9 bits per channel-use. These results confirm our expectation that under near perfect channel estimation (obtained when the number of pilot symbols increases), the performance of the improved and mismatched receivers become almost the same.

4.8 Conclusion

The problem of signal detection in a practical MB-OFDM communication system where the receiver has only access to a noisy estimate of the channel provided by pilot symbols was investigated. Both the case of single- and of multi-antenna MB-OFDM systems was addressed. Based on a statistical characterization of the channel estimation process, we proposed a general

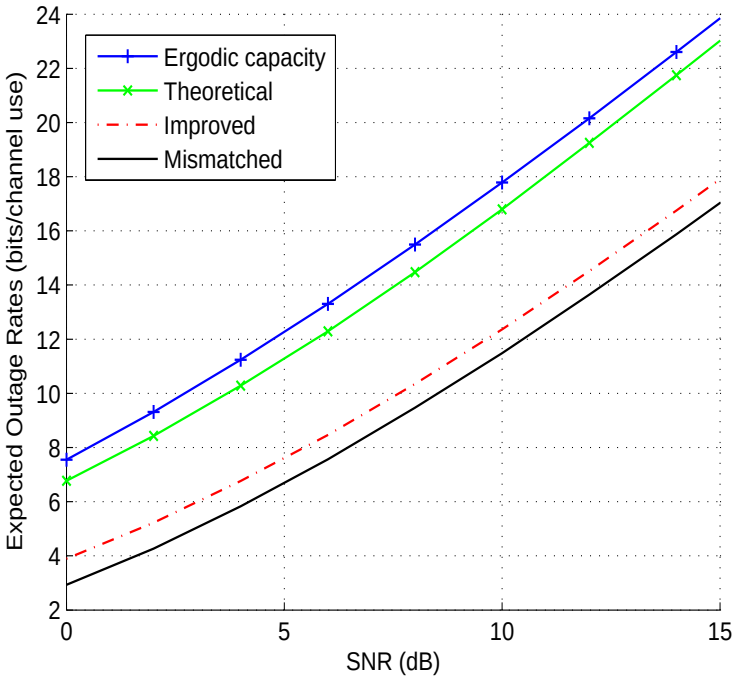

Figure 4.15: Expected outage rates of a 4×4 MB-MIMO-OFDM transmission versus SNR for $M = 16$ subcarriers and $N_P = 4$ pilots per frame and different detection approaches; outage probability $\gamma = 0.01$; CM1 channel.

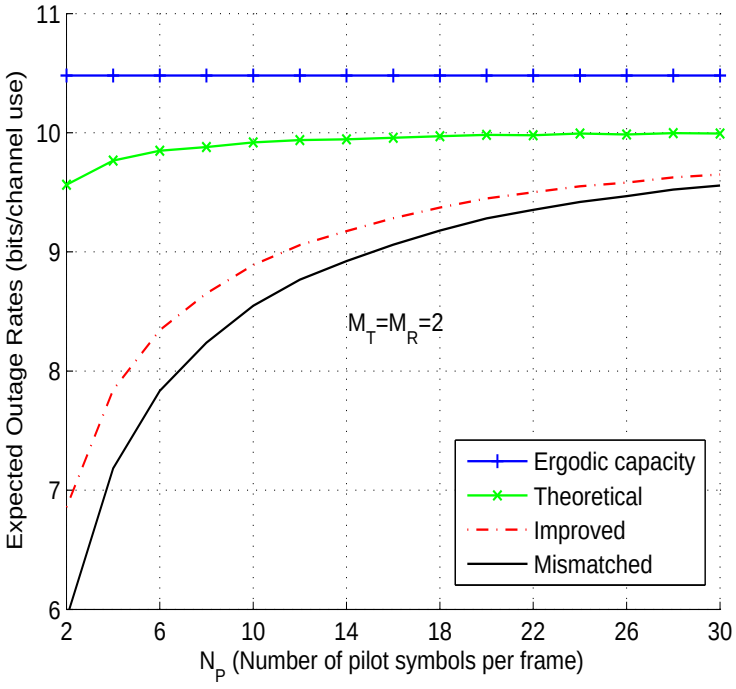

Figure 4.16: Expected outage rates of a 2×2 MB-MIMO-OFDM transmission versus N_P for $M = 16$ subcarriers and different detection approaches at SNR=15 dB; outage probability $\gamma = 0.01$; CM1 channel.

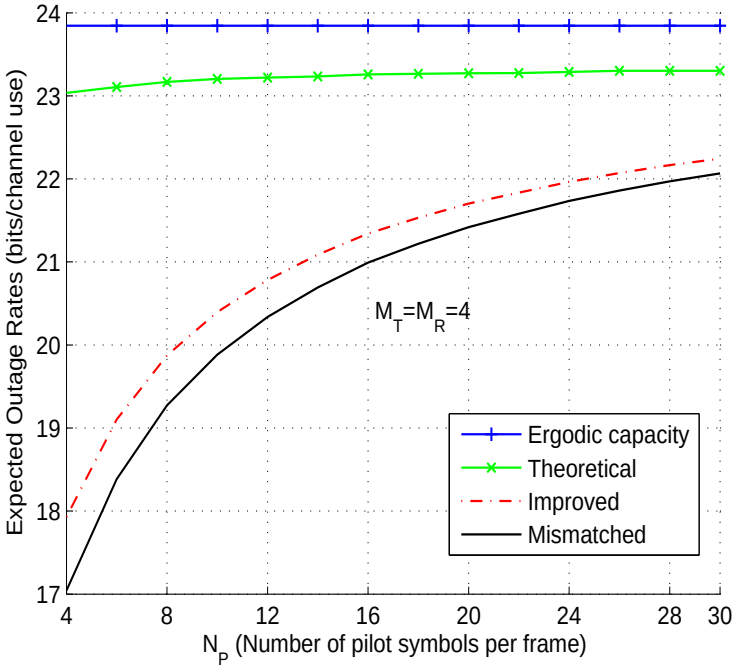

Figure 4.17: Expected outage rates of a 4×4 MB-MIMO-OFDM transmission versus N_P for $M = 16$ subcarriers and different detection approaches at SNR=15 dB; outage probability $\gamma = 0.01$; CM1 channel.

detector design that takes into account the imperfect channel available at the receiver. In the case of ML detection, this approach led to an improved ML metric that we used in the derivation of a modified iterative MAP detector. We also derived the expressions of the achievable outage rates associated to the improved and mismatched ML metrics. Our numerical results indicated that the mismatched detector is sub-optimal in terms of BER and achievable outage rates, especially for short training sequence. They also confirmed the adequacy of the improved detector under imperfect channel estimation. It was shown that the proposed receiver can save the improvements brought by set-partitioning labeling in the case of poor channel estimates (obtained with very short training sequences). **This performance improvement was obtained without requiring additional complexity in the receiver**. A practical application of our results is in the evaluation of the trade-off between the required QoS (in terms of BER and achieved throughputs) and system parameters (e.g., training power, transmission power, period of training, decoding latency, outage probability).

Chapter 5

Low-complexity Iterative MIMO Signal Detection Accounting for Channel Estimation Errors

5.1 Introduction and Motivations

In the previous chapter, we presented an improved turbo receiver based on maximum *a posteriori* (MAP) detection for MIMO-OFDM systems working in the presence of channel estimation errors. In this chapter, we consider again an iterative detection composed of a MAP detector and a soft-input soft-output (SISO) channel decoder exchanging soft information with each other through several iterations [43,149]. Although MAP demodulation is the optimal solution in the sense of the bit error rate, its complexity grows exponentially with both the number of transmit antennas and the signal constellation size. Hence, considerable interest has been devoted to the design of less complex MIMO detectors conserving performance as close as possible to the optimal MAP detector.

Reduced-complexity receivers usually fall into two main classes, whether they rely on reduced state trellis-based algorithms or on linear filtering. Among the trellis-based solutions, several contributions involve the use of Max-Log-MAP detector in combination with "hard" demodulation algorithms such as sphere decoding and semidefinite relaxation [12], see also [13,14]. Other trellis-based reduced-complexity detectors are based on the idea of list decoding, in which the likelihood of each bit is approximated by partial marginalization over a list of dominant bit-vectors, rather than complete marginalization over the list of all possible vectors. Most of the existing schemes in this area are based on "hard" sphere decoding [161,162] and tree search algorithms [15,16]. However, the complexity of trellis-based detectors is still higher than filter-based

solutions. For this reason, in this chapter we will focus our attention on filter-based solutions based on the minimum mean square error (MMSE) criterion.

A low-complexity filter-based detector is the one based on soft parallel interference cancellation (soft-PIC) and linear MMSE filtering. This scheme was first proposed by Wang and Poor [17] in the context of multiuser detection for CDMA. Based on Wang's work, Tuchler *et al.* [18] proposed an MMSE-based equalizer, where the filter parameters are updated by taking into account the data *a priori* probabilities available from the decoder. This equalizer replaced the MAP equalizer in a turbo detection scheme. Then, Dejonghe and Vandendorpe extended the results of [18] to fractionally-spaced equalization and to higher-level modulations in [19] (see also [20] for the case of MIMO fading channels). Sellathurai and Haykin also proposed a turbo MIMO detector based on soft-PIC and MMSE filtering by considering the simple spatial multiplexing scheme (also known as the V-BLAST scheme) [21].

To our knowledge, all of the filter-based detector designs proposed in the literature are based on a *perfect* knowledge of the channel at the receiver. Obviously, in practical communication systems, the receiver has only access to an *imperfect* (and possibly very poor) estimate of the unknown channel. In this situation, one may use a *mismatched* detector by simply using the channel estimate instead of the perfect channel. We showed in the previous chapters that mismatched detection greatly degrades the performance of the optimal MAP detector. Our aim in this chapter is to propose an improved MMSE-based detector, which takes into account the imperfect channel estimation obtained via some training sequences. To this end, we propose a filter design based on the general Bayesian framework, introduced in the previous chapter. This enables us to derive an improved soft-PIC detector by taking into account the channel estimation error in the formulation of the instantaneous linear MMSE filter, as well as in the interference cancellation part. For simplicity, we first derive our detector for the V-BLAST scheme and then generalize it to the case of an arbitrary space-time (ST) coded MIMO systems.

The outline of this chapter is as follows. In Section 5.2, we describe our system model and recall some results concerning pilot-based channel estimation from the previous chapter. In Section 5.3, we provide the formulation of the considered MMSE-based soft-PIC detector. Using this material in Section 5.4, we propose an improved turbo-PIC detector under imperfect CSIR. Section 5.5 is devoted to a simplified method for computing the MMSE filter coefficients. The simulations of Section 5.6, illustrate a comparative performance study of the proposed detector with the classical mismatched detector, and Section 5.7 concludes the chapter.

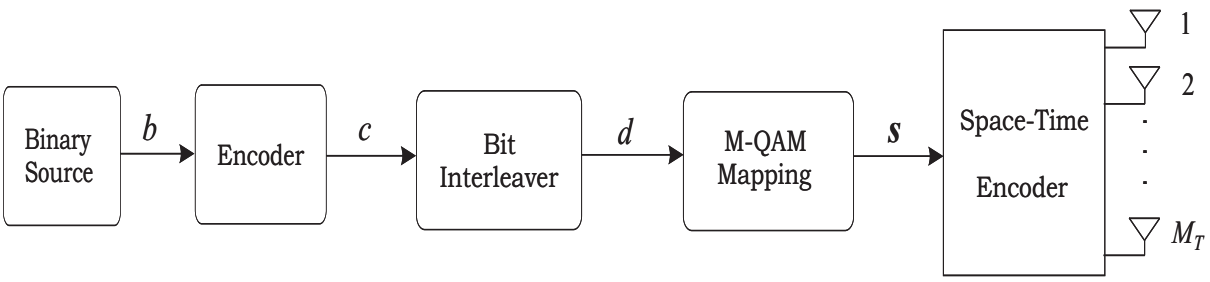

Figure 5.1: Block diagram of MIMO-BICM transmission scheme.

5.2 System Model

5.2.1 MIMO Fading Channel

We consider a single-user MIMO system with M_T transmit and M_R receive antennas ($M_R \geq M_T$), transmitting over a frequency non-selective channel and refer to it as an ($M_T \times M_R$) MIMO channel. Figure 5.1 shows the block diagram of the transmitter that employs the bit-interleaved coded modulation (BICM) scheme which is known to be a simple and efficient method for exploiting channel time-selectivity [44]. The binary data sequence $\boldsymbol{b}$ are encoded by a non-recursive non-systematic convolutional (NRNSC) code before being interleaved by a quasi-random interleaver. The output bits $\boldsymbol{d}$ are gathered in subsequences of B bits and mapped to complex M-QAM ($\mathrm{M} = 2^B$) symbols s, before being passed to the ST encoder.

In what follows, unless otherwise mentioned, we consider the ST coding in its simplest form, that is, just spatial multiplexing of data symbols. We present our model for this simple case, and we shall generalize it later to the case of an arbitrary ST code.

Let us denote by $\boldsymbol{x}_k$ the $(M_T \times 1)$ vector of transmitted data symbols (after the ST encoder) at a sample time k. As we consider the simple spatial multiplexing, we use $\boldsymbol{s}_k$ instead of $\boldsymbol{x}_k$. Assuming a frame of symbols corresponding to L_f channel-uses transmitted over the channel matrix $\mathbf{H}$, the received signal vector $\boldsymbol{y}_k$ of dimension $(M_R \times 1)$ is given by

$$\boldsymbol{y}_k = \mathbf{H}\,\boldsymbol{s}_k + \boldsymbol{z}_k \qquad k = 1, ..., L_f, \tag{5.1}$$

where $\boldsymbol{s}_k$ is the $(M_T \times 1)$ vector of transmitted symbol with average power $E_s \triangleq \frac{1}{M_T}\mathbb{E}[\,\mathrm{tr}(\boldsymbol{s}_k \boldsymbol{s}_k^\dagger)\,]$. We assume that the entries $H_{i,j}$ of the random matrix $\mathbf{H}$ are i.i.d. zero-mean circularly symmetric complex Gaussian (ZMCSCG) random variables. Thus, the channel matrix $\mathbf{H}$ is distributed as $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbb{I}_{M_T} \otimes \mathbf{\Sigma}_H)$, where

$$\mathcal{CN}(\mathbf{0}, \mathbb{I}_{M_T} \otimes \mathbf{\Sigma}_H) = \frac{1}{\pi^{M_R M_T} \det\{\mathbf{\Sigma}_H\}^{M_T}} \exp\left\{ -\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_H^{-1}\mathbf{H}^\dagger\right)\right\}. \tag{5.2}$$

Here, $\mathbf{\Sigma}_H$ is the $(M_R \times M_R)$ covariance matrix of the rows of $\mathbf{H}$. Obviously, with our assumptions of i.i.d. channel, $\mathbf{\Sigma}_H$ is a diagonal matrix with equal diagonal entries σ_h^2. The noise vector $\boldsymbol{z}_k$

is assumed to be a ZMCSCG random vector with covariance matrix $\mathbf{\Sigma}_z \triangleq \mathbb{E}(\boldsymbol{z}_k \boldsymbol{z}_k^{\mathcal{H}}) = \sigma_z^2 \, \mathbb{I}_{M_R}$. Both $\mathbf{H}$ and $\boldsymbol{z}_k$ are assumed to be ergodic and stationary random processes.

We consider the block fading model for the channel time variations. So, channel coefficients are assumed to be constant during a block of symbols, and change to new independent values from one block to another. We assume that each frame of data symbols corresponds to N_c independent fading blocks. Notice that $N_c = 1$ returns to the quasi-static channel model.

5.2.2 Pilot-based Channel Estimation

In order to estimate the MIMO transmission channel matrix $\mathbf{H}$ at the receiver, corresponding to each fading block, we send a number of N_P pilot symbols in addition to the ST-coded data symbols. Recall from the previous chapter that with orthogonal training sequences, the maximum likelihood (ML) estimate of $\mathbf{H}$ is given by

$$\begin{aligned} \widehat{\mathbf{H}}_{\mathrm{ML}} &= \mathbf{Y}_P \, \mathbf{S}_P^{\dagger} \, (\mathbf{S}_P \mathbf{S}_P^{\dagger})^{-1} \\ &= \mathbf{H} + \boldsymbol{\mathcal{E}} \end{aligned} \tag{5.3}$$

where the j-th column of $\boldsymbol{\mathcal{E}}$ has the covariance matrix $\mathbf{\Sigma}_{\mathcal{E}} = \sigma_{\mathcal{E}}^2 \mathbb{I}_{M_R}$.
This pilot-based channel estimation can be characterized by the *posterior* distribution of the perfect channel conditioned on its ML estimate, as follows (see Appendix B)

$$p\,(\mathbf{H}|\widehat{\mathbf{H}}_{\mathrm{ML}}) = \mathcal{CN}\big(\delta\,\widehat{\mathbf{H}}_{\mathrm{ML}},\, \delta\sigma_{\mathcal{E}}^2\,\mathbb{I}_{M_T} \otimes \mathbb{I}_{M_R}\big). \tag{5.4}$$

The latter distribution constitutes a Bayesian framework which is exploited in the following for the design of appropriate turbo-PIC detectors under imperfect channel estimation.
For the sake of simplicity, we will not specify in the sequel the subscript ML for $\widehat{\mathbf{H}}$.

5.3 General Formulation of MMSE-based Turbo-PIC Detection

At the receiver, we perform iterative symbol detection and channel decoding. As shown in Fig. 5.2, the receiver principally consists of a MIMO detector (also shortly called demapper) and a SISO channel decoder that exchange *extrinsic* soft information with each other. Here, we consider this soft information in the form of LLRs.
As we are looking for the best trade-off between simple implementation and higher performance, the SISO decoder is based on the Max-Log-MAP algorithm, described in [11, 152]. In the following, we present the general formulation of the MIMO detector part based on soft-PIC and MMSE filtering assuming perfect CSIR. This constitutes the basis for the next section where we present an improved soft-PIC detector for the case of imperfect CSIR.

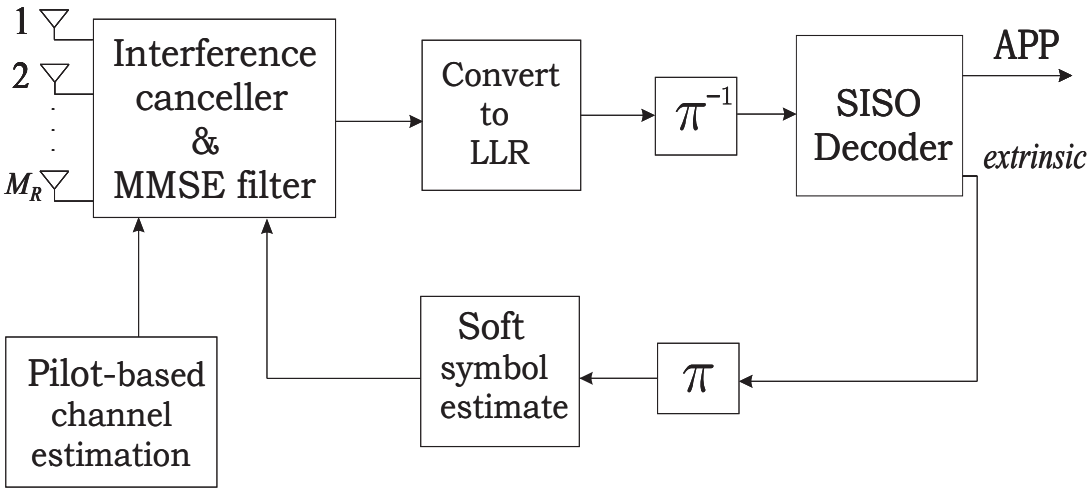

Figure 5.2: Structure of the turbo-PIC receiver with pilot based channel estimation.

5.3.1 A Glance at MAP Detection

Let $d_k^{i,m}$ be the m-th ($m = 1, 2, ..., BM_T$) bit corresponding to the symbol vector $\boldsymbol{s}_k$, transmitted at the k-th time-slot from the i-th antenna. We denote by $L(d_k^{i,m})$ the LLR of the bit $d_k^{i,m}$ at the output of the MIMO detector. Recall from the previous chapter that

$$L(d_k^{i,m}) = \log \frac{\displaystyle\sum_{\boldsymbol{s}_k \in \mathcal{S}_1^m} e^{-\mathcal{D}_{\mathrm{ML}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \mathbf{H})} \prod_{\substack{n=1 \\ n \neq m}}^{BM_T} P_{\mathrm{dec}}^1(d_k^{i,n})}{\displaystyle\sum_{\boldsymbol{s}_k \in \mathcal{S}_0^m} e^{-\mathcal{D}_{\mathrm{ML}}(\boldsymbol{s}_k, \boldsymbol{y}_k, \mathbf{H})} \prod_{\substack{n=1 \\ n \neq m}}^{BM_T} P_{\mathrm{dec}}^0(d_k^{i,n})}. \tag{5.5}$$

It is clear that the computational complexity of the MAP detector becomes prohibitively large for large size signal constellations and/or for large number of transmit antennas, as each of the sets $\mathcal{S}_1^m$ and $\mathcal{S}_0^m$ in (5.5) contains $2^{(BM_T-1)}$ vectors $\boldsymbol{s}_k$. For such cases, the suboptimal soft-PIC detector would make a good compromise between complexity and performance [163].

5.3.2 Soft-PIC Detection for the Case of Spatial Multiplexing

Here, to detect a symbol transmitted from a given antenna, we first make use of the soft information available from the SISO channel decoder to reduce and hopefully to cancel the interfering signals arising from other transmit antennas. Then, an MMSE filter is applied to further reduce the residual interference. Note that this differs from a conventional MMSE filtering in the sense that the MMSE criterion is evaluated over both the distribution of the noise and the distribution over the symbols. For turbo detection, the symbol distribution is no longer i.i.d. as typically assumed for classical MMSE-based detection.

Let us consider the transmitted vector $\boldsymbol{s}_k = [s_k^1, ..., s_k^{M_T}]^T$ at time k and assume that we

are interested in the detection of its i-th symbol s_k^i. We start by evaluating the parameters $\hat{s}_k^j$ and $\sigma^2_{s_k^j}$ for the interfering symbols s_k^j for $j \neq i$ from the SISO decoder as follows:

$$\hat{s}_k^j = \mathbb{E}\big[s_k^j\big] = \sum_{\substack{j=1 \\ j\neq i}}^{2^B} s_k^j \, P[s_k^j] \tag{5.6}$$

$$\sigma^2_{s_k^j} = \mathbb{E}\big[\,|s_k^j|^2\,\big] = \sum_{\substack{j=1 \\ j\neq i}}^{2^B} |s_k^j|^2 \, P[s_k^j] \tag{5.7}$$

where $P[s_k^j]$ is the a posteriori probability of the transmission of s_k^j and is evaluated using the probabilities $P_{\text{dec}}(d_k^{j,n})$ at the decoder output:

$$P[s_k^j] = K \prod_{n=1}^{B} P_{\text{dec}}(d_k^{j,n}),$$

where K is a normalization factor. We further introduce the following definitions.
$\underline{\mathbf{H}}_i$ is the $(M_R \times (M_T - 1))$ matrix constructed from $\mathbf{H}$ by discarding its i-th column, namely $\mathbf{h}_i$. We also define the $((M_T - 1) \times 1)$ vectors

$$\underline{\boldsymbol{s}}_k^i \triangleq \big[s_k^1, s_k^2, ..., s_k^{i-1}, s_k^{i+1}, ..., s_k^{M_T}\big]^T$$

and

$$\underline{\hat{\boldsymbol{s}}}_k^i \triangleq \big[\hat{s}_k^1, \hat{s}_k^2, ..., \hat{s}_k^{i-1}, \hat{s}_k^{i+1}, ..., \hat{s}_k^{M_T}\big]^T.$$

Now, given the received signal vector $\boldsymbol{y}_k$, a soft interference cancellation is performed for detecting the symbol s_k^i by subtracting to $\boldsymbol{y}_k$ the estimated signals of the other transmit antennas as:

$$\begin{aligned} \underline{\boldsymbol{y}}_k^i &= \boldsymbol{y}_k - \underline{\mathbf{H}}_i \, \underline{\hat{\boldsymbol{s}}}_k^i \\ &= \mathbf{h}_i s_k^i + \underline{\mathbf{H}}_i \, \underline{\boldsymbol{s}}_k^i - \underline{\mathbf{H}}_i \, \underline{\hat{\boldsymbol{s}}}_k^i + \boldsymbol{z}_k, \quad \text{for } \; i = 1, ..., M_T. \end{aligned} \tag{5.8}$$

Except under perfect prior information on the symbols which leads to $\hat{s}_k^j = s_k^j$ for all $j \neq i$, there remains a residual interference in $\underline{\boldsymbol{y}}_k^i$. In order to reduce further this interference, an instantaneous linear MMSE filter $\mathbf{w}_k^i$ is applied to $\underline{\boldsymbol{y}}_k^i$ to minimize the mean square error value of the error e_k^i defined as

$$e_k^i = s_k^i - r_k^i \tag{5.9}$$

where the filter output r_k^i is equal to

$$r_k^i = \mathbf{w}_k^i \, \underline{\boldsymbol{y}}_k^i. \tag{5.10}$$

Here, the $(1 \times M_R)$ vector $\mathbf{w}_k^i$ is obtained as

$$\mathbf{w}_k^i = \underset{\mathbf{w}_k^i \in \mathbb{C}^{1\times M_R}}{\arg\min} \; \mathbb{E}_{\boldsymbol{s}_k, \boldsymbol{z}_k} \left[\left| s_k^i - \mathbf{w}_k^i \, \underline{\boldsymbol{y}}_k^i \right|^2 \right]. \tag{5.11}$$

By invoking the orthogonality principle [84], the coefficients of the MMSE filter $\mathbf{w}_k^i$ are given by

$$\mathbf{w}_k^i = \mathbf{h}_i^\dagger \left[\mathbf{h}_i \mathbf{h}_i^\dagger + \frac{\underline{\mathbf{H}}_i \big(\mathbf{\Lambda}_{k,i} - \widetilde{\mathbf{\Lambda}}_{k,i} \big) \underline{\mathbf{H}}_i^\dagger}{\sigma^2_{s_k^i}} + \frac{\sigma_z^2}{\sigma^2_{s_k^i}} \mathbb{I}_{M_R} \right]^{-1} \tag{5.12}$$

where

$$\mathbf{\Lambda}_{k,i} = \mathbb{E}\big[\underline{\boldsymbol{s}}_k^i \, \underline{\boldsymbol{s}}_k^{i\,\dagger} \big] \approx \mathrm{diag}\Big(\mathbb{E}\big[|s_k^1|^2\big], ..., \mathbb{E}\big[|s_k^{i-1}|^2\big], \mathbb{E}\big[|s_k^{i+1}|^2\big], ..., \mathbb{E}\big[|s_k^{M_T}|^2\big] \Big), \quad \text{and}$$

$$\widetilde{\mathbf{\Lambda}}_{k,i} = \underline{\hat{\boldsymbol{s}}}_k^i \, \underline{\hat{\boldsymbol{s}}}_k^{i\,\dagger} \approx \mathrm{diag}\Big(|\hat{s}_k^1|^2, ..., |\hat{s}_k^{i-1}|^2, |\hat{s}_k^{i+1}|^2, ..., |\hat{s}_k^{M_T}|^2 \Big).$$

Note that the off-diagonal entries in $\mathbf{\Lambda}_{k,i}$ and $\widetilde{\mathbf{\Lambda}}_{k,i}$ have been neglected to reduce the complexity without causing significant performance loss [164].

At the first decoding iteration, we have no prior information available on the transmitted data, i.e., $\mathbf{\Lambda}_{k,i} = \sigma^2_{s_k^i} \mathbb{I}_{M_T-1}$ and $\widetilde{\mathbf{\Lambda}}_{k,i} = \mathbf{0}_{M_T-1}$. Consequently, (5.12) reduces to

$$\mathbf{w}_k^i = \mathbf{h}_i^\dagger \left[\mathbf{H}\mathbf{H}^\dagger + \frac{\sigma_z^2}{\sigma^2_{s_k^i}} \mathbb{I}_{M_R} \right]^{-1} \tag{5.13}$$

which is no more than the linear MMSE detector for s_k^i.

Before being passed to the SISO decoder, the detected symbols r_k at the output of the MMSE filter in (5.10), must be converted to LLR. This is done assuming a Gaussian distribution for the residual interference after soft-PIC detection. Details on the LLR conversion can be found in [17]. For the sake of brevity, this will be presented only for the case of the improved detector in Section 5.4.

5.3.3 Generalization to the Case of Arbitrary Space-time Coding

Up to now, we have considered the simple spatial multiplexing (V-BLAST) scheme at the transmitter. We show here that the detector formulation presented for the case of spatial multiplexing, can also be applied to the general case of ST coding with a slight modification. More details are provided in Appendix D.

Let us consider the general formulation of linear dispersion (LD) codes [165]. Let $\boldsymbol{s}$ of dimension $(Q \times 1)$ be the vector of data symbols prior to ST coding

$$\boldsymbol{s} = \big[s^1 \; s^2 \; \cdots \; s^Q \big]^T. \tag{5.14}$$

We removed the time index k for simplicity. By ST coding, these symbols are mapped into a matrix $\mathbf{X}$ of dimension $(M_T \times T_u)$ with T_u the number of channel-uses. Corresponding to an encoded matrix $\mathbf{X}$, we receive the matrix $\mathbf{Y}$ of dimension $(M_R \times T_u)$.

In order to obtain a general formulation for the receiver, we separate the real and imaginary

parts of the entries of $\boldsymbol{s}$, $\mathbf{Y}$, and $\mathbf{X}$, and stack them row-wise in vectors $\breve{\mathbf{s}}$, $\breve{\mathbf{X}}$, and $\breve{\mathbf{Y}}$, of dimension $(2Q \times 1)$, $(2M_T T_u \times 1)$, and $(2M_R T_u \times 1)$, respectively [165]. Vectors $\breve{\mathbf{X}}$ and $\breve{\mathbf{Y}}$ are related through a $(2M_R T_u \times 2M_T T_u)$ matrix $\breve{\mathbf{H}}$ as (see Appendix D for more details):

$$\breve{\mathbf{Y}} = \breve{\mathbf{H}}\, \breve{\mathbf{X}} + \breve{\mathbf{Z}} \tag{5.15}$$

where $\breve{\mathbf{Z}}$ is the vector of real AWGN of zero mean and variance $\sigma_z^2/2$. Now, we can write the "ST code + channel" input/output relationship by considering an equivalent channel matrix $\breve{\mathbf{H}}_{eq}$ of dimension $(2M_R T_u \times 2Q)$:

$$\breve{\mathbf{Y}} = \breve{\mathbf{H}}\, \breve{\mathbf{F}}\, \breve{\mathbf{s}} + \breve{\mathbf{Z}} = \breve{\mathbf{H}}_{eq}\, \breve{\mathbf{s}} + \breve{\mathbf{Z}}. \tag{5.16}$$

We see that, in the expressions of the detector in subsection 5.3.2, we have just to consider $\breve{\mathbf{H}}_{eq}$, $\breve{\mathbf{s}}$, and $\breve{\mathbf{Y}}$, instead of $\mathbf{H}$, $\boldsymbol{s}$, and $\boldsymbol{y}$, respectively.

5.4 Improved MMSE-based Turbo-PIC Detection Under Imperfect CSIR

We propose here modifications to the turbo-PIC detector to mitigate the impact of imperfect channel estimation on the receiver performance.

As we see from (5.8) and (5.12), we need the channel $\mathbf{H}$ for both interference canceling and MMSE filtering. As the receiver has only an imperfect channel estimate $\widehat{\mathbf{H}}$, the sub-optimal *mismatched* solution consists in replacing $\underline{\mathbf{H}}_i$ and $\mathbf{h}_i$ in (5.8) and (5.12) by their estimates $\underline{\widehat{\mathbf{H}}}_i$ and $\widehat{\mathbf{h}}_i$, respectively.

As a first step toward a realistic design, we make use of the available channel estimate $\widehat{\mathbf{H}}$ for interference cancellation. That is, equation (5.8) is rewritten as

$$\begin{aligned} \underline{\boldsymbol{y}}_k^i &= \boldsymbol{y}_k - \underline{\widehat{\mathbf{H}}}_i\, \underline{\hat{\boldsymbol{s}}}_k^i \\ &= \mathbf{h}_i s_k^i + \underline{\mathbf{H}}_i\, \underline{\boldsymbol{s}}_k^i - \underline{\widehat{\mathbf{H}}}_i\, \underline{\hat{\boldsymbol{s}}}_k^i + \boldsymbol{z}_k, \quad \text{for } \; i = 1, ..., M_T \end{aligned} \tag{5.17}$$

where $\underline{\widehat{\mathbf{H}}}_i$ is the $(M_R \times (M_T - 1))$ matrix constructed from $\widehat{\mathbf{H}}$ by discarding its i-th column, namely $\widehat{\mathbf{h}}_i$. We now propose a novel *improved* PIC detector under imperfect CSIR. We note that (5.17) still depends on the unknown channel matrix $\mathbf{H}$. To overcome this problem, we use the posterior distribution (5.4) and make two modifications to the detector described in Subsection 5.3.2, as follows.

The first modification we propose, concerns the design of the filter $\mathbf{w}_k^i$ in (5.11). Since the cost function $f(\boldsymbol{y}_k, s_k^i, \mathbf{H}) = \mathbb{E}_{\boldsymbol{s}_k, \boldsymbol{z}_k}\left[\left|s_k^i - \mathbf{w}_k^i\, \underline{\boldsymbol{y}}_k^i\right|^2\right]$ is a function of the perfect channel $\mathbf{H}$ (via $\underline{\boldsymbol{y}}_k^i$), according to the method developed in the previous chapter (Section 4.3), we propose a

modified filter $\widetilde{\mathbf{w}}_k^i$, chosen to minimize the average of the mean square error over all realizations of channel estimation errors. Using (5.4), we propose the following filter design

$$\begin{aligned}\widetilde{\mathbf{w}}_k^i &= \underset{\widetilde{\mathbf{w}}_k^i \in \mathbb{C}^{1\times M_R}}{\arg\min}\ \mathbb{E}_{\mathbf{H},\boldsymbol{s}_k,\boldsymbol{z}_k}\left[\left|s_k^i - \widetilde{\mathbf{w}}_k^i\,\underline{\boldsymbol{y}}_k^i\right|^2\middle|\widehat{\mathbf{H}}\right] \\ &= \underset{\widetilde{\mathbf{w}}\in \mathbb{C}^{1\times M_R}}{\arg\min}\ \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left[\mathbb{E}_{\boldsymbol{s}_k,\boldsymbol{z}_k}\left[\left|s_k^i - \widetilde{\mathbf{w}}_k^i\,\underline{\boldsymbol{y}}_k^i\right|^2\right]\right] \end{aligned} \tag{5.18}$$

where in the latter expression, we have assumed the independence between $\mathbf{H}$, $\boldsymbol{s}_k$, and $\boldsymbol{z}_k$. From (5.18) and after invoking the orthogonality principle [84], we obtain

$$\widetilde{\mathbf{w}}_k^i = \left(\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left[\mathbb{E}_{\boldsymbol{s}_k,\boldsymbol{z}_k}\left[s_k^i\,\underline{\boldsymbol{y}}_k^{i\,\dagger}\right]\right]\right)\left(\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left[\mathbb{E}_{\boldsymbol{s}_k,\boldsymbol{z}_k}\left[\underline{\boldsymbol{y}}_k^i\,\underline{\boldsymbol{y}}_k^{i\,\dagger}\right]\right]\right)^{-1}. \tag{5.19}$$

After some algebraic manipulations provided in the Appendix D, we get the modified filter $\widetilde{\mathbf{w}}_k^i$, directly as a function of $\underline{\widehat{\mathbf{H}}}_i$ and $\widehat{\mathbf{h}}_i$, as follows.

$$\boxed{\widetilde{\mathbf{w}}_k^i = \overline{\mathbf{R}}_{s_k^i\underline{\boldsymbol{y}}_k^i}\,\overline{\mathbf{R}}_{\underline{\boldsymbol{y}}_k^i}^{-1}} \tag{5.20}$$

where

$$\overline{\mathbf{R}}_{s_k^i\underline{\boldsymbol{y}}_k^i} = \delta\,\sigma^2_{s_k^i}\,\widehat{\mathbf{h}}_i^{\dagger} + (\delta-1)\,\mathbf{m}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger} \tag{5.21}$$

with $\mathbf{m}_{k,i} = \hat{s}_k^i\,\underline{\hat{\boldsymbol{s}}}_k^{i\,\dagger}$ and δ is given by equation (4.18) in the previous chapter, and

$$\begin{aligned}\overline{\mathbf{R}}_{\underline{\boldsymbol{y}}_k^i} &= \delta^2\sigma^2_{s_k^i}\,\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\dagger} + \delta^2\,\underline{\widehat{\mathbf{H}}}_i\,\boldsymbol{\Lambda}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger} + (\delta^2-\delta)\,\widehat{\mathbf{h}}_i\,\mathbf{m}_{k,i}\,\underline{\mathbf{H}}_i^{\dagger} + (\delta^2-\delta)\,\underline{\widehat{\mathbf{H}}}_i\,\mathbf{m}_{k,i}^{\dagger}\,\widehat{\mathbf{h}}_i^{\dagger} \\ &\quad + (1-2\delta)\,\underline{\widehat{\mathbf{H}}}_i\,\widetilde{\boldsymbol{\Lambda}}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger} + \big(\sigma_z^2 + (1-\delta)\,\sigma^2_{s_k^i} + (1-\delta)\,\mathrm{tr}(\boldsymbol{\Lambda}_{k,i})\big)\mathbb{I}_{M_R}. \end{aligned} \tag{5.22}$$

To get more insight on the proposed detector, let us consider the ideal case where perfect channel knowledge is available at the receiver, i.e., $\widehat{\mathbf{H}} = \mathbf{H}$ and $\sigma^2_{\mathcal{E}} = 0$. We note that in this case $\delta = 1$ and the posterior pdf (5.4) reduces to a Dirac delta function; consequently, the two filters $\mathbf{w}_k^i$ (5.12) and $\widetilde{\mathbf{w}}_k^i$ (5.20) coincide. However, in the presence of estimation errors, the proposed improved and mismatched detectors become different due to the inherent averaging in (5.18), which provides a robust design that adapts itself to the channel estimate available at the receiver.

Our second modification concerns the application of the derived filter $\widetilde{\mathbf{w}}_k^i$ to the received signal $\underline{\boldsymbol{y}}_k^i$. Since the latter is a function of the perfect channel[1], we propose to apply the MMSE filter of (5.20) to a modified received signal, evaluated with reference to (5.8) by:

$$\begin{aligned}\underline{\widetilde{\boldsymbol{y}}}_k^i &= \widetilde{\mathbf{h}}_i s_k^i + \underline{\widetilde{\mathbf{H}}}_i\,\underline{\boldsymbol{s}}_k^i - \underline{\widehat{\mathbf{H}}}_i\,\underline{\hat{\boldsymbol{s}}}_k^i + \boldsymbol{z}_k, \\ &= (\delta\,\widehat{\mathbf{h}}_i)s_k^i + (\delta\,\underline{\widehat{\mathbf{H}}}_i)\,\underline{\boldsymbol{s}}_k^i - \underline{\widehat{\mathbf{H}}}_i\,\underline{\hat{\boldsymbol{s}}}_k^i + \boldsymbol{z}_k, \end{aligned} \tag{5.23}$$

[1] Actually, the unknown channel is required for LLR evaluation, see (5.24) and (5.26).

where

$$\widetilde{\underline{\mathbf{H}}}_i = \mathbb{E}_{\underline{\mathbf{H}}_i|\widehat{\underline{\mathbf{H}}}_i}[\underline{\mathbf{H}}_i] = \delta\,\widehat{\underline{\mathbf{H}}}_i \quad \text{and} \quad \widetilde{\mathbf{h}}_i = \mathbb{E}_{\mathbf{h}_i|\widehat{\mathbf{h}}_i}[\mathbf{h}_i] = \delta\,\widehat{\mathbf{h}}_i.$$

Now, by applying the modified filter $\widetilde{\mathbf{w}}_k^i$ to $\widetilde{\underline{\boldsymbol{y}}}_k^i$ of (5.23), the output of the improved MMSE detector is obtained as

$$\widetilde{r}_k^i = \widetilde{\mathbf{w}}_k^i\,\widetilde{\underline{\boldsymbol{y}}}_k^i = \underbrace{\delta\,\widetilde{\mathbf{w}}_k^i\,\widehat{\mathbf{h}}_i}_{\mu_{k,i}}\,s_k^i + \underbrace{\delta\,\widetilde{\mathbf{w}}_k^i\,\widehat{\underline{\mathbf{H}}}_i\,\underline{\boldsymbol{s}}_k^i - \widetilde{\mathbf{w}}_k^i\,\widehat{\underline{\mathbf{H}}}_i\,\underline{\hat{\boldsymbol{s}}}_k^i + \widetilde{\mathbf{w}}_k^i\,\boldsymbol{z}_k}_{\eta_{k,i}} \tag{5.24}$$

where $\eta_{k,i}$ is the interference-plus-noise affecting the output of the instantaneous MMSE filter $\widetilde{r}_k^i$. From (5.24) it is clear that the output of the improved MMSE filter can be viewed as an equivalent AWGN channel having s_k^i at its input:

$$\widetilde{r}_k^i = \mu_{k,i}\,s_k^i + \eta_{k,i}. \tag{5.25}$$

It is shown in [17,19] that under perfect CSIR, a similar expression of $\eta_{k,i}$ is well approximated by a zero-mean Gaussian random variable with variance $\sigma^2_{\eta_{k,i}}$. The parameters $\mu_{k,i}$ and $\sigma^2_{\eta_{k,i}}$ are calculated at each sample-time by using the symbols statistics. The exact derivation of the variance $\sigma^2_{\eta_{k,i}}$ is provided in Appendix D for the more general case of the improved detector.

From (5.24), we can calculate the LLRs on the corresponding bits of the detected symbols at the output of the MMSE filter, that will be used by the SISO channel decoder:

$$L(d_k^{i,m}) = \log \frac{P_{\text{dem}}\big(\,d_k^{i,m}=1|\,\widetilde{r}_k^i, \mu_{k,i}\,\big)}{P_{\text{dem}}\big(\,d_k^{i,m}=0|\,\widetilde{r}_k^i, \mu_{k,i}\,\big)}$$

$$= \log \frac{\displaystyle\sum_{s_k^i \in \underline{\mathcal{S}}_1^m} \exp\left\{-\frac{|\widetilde{r}_k^i - \mu_{k,i}\,s_k^i|^2}{\sigma^2_{\eta_{k,i}}}\right\} \prod_{\substack{n=1 \\ n\neq m}}^{B} P^1_{\text{dec}}(d_k^{i,n})}{\displaystyle\sum_{s_k^i \in \underline{\mathcal{S}}_0^m} \exp\left\{-\frac{|\widetilde{r}_k^i - \mu_{k,i}\,s_k^i|^2}{\sigma^2_{\eta_{k,i}}}\right\} \prod_{\substack{n=1 \\ n\neq m}}^{B} P^0_{\text{dec}}(d_k^{i,n})}. \tag{5.26}$$

It is very important to note that in contrast to the case of MAP detection where in (5.5) $\mathcal{S}_1^m$ and $\mathcal{S}_0^m$ are of size $2^{(M_T B-1)}$, here the cardinality of the sets $\underline{\mathcal{S}}_1^m$ and $\underline{\mathcal{S}}_0^m$ is only equal to $2^{(B-1)}$.

5.5 Simplified Turbo-PIC Detection

As clear from equations (5.12) and (5.20), the MMSE filter coefficients depend on the index k and hence must be recomputed at each time-sample k. This motivates us to propose here a simplified version of our proposed detector in order to reduce the receiver complexity further.

5.5.1 Case of Perfect CSIR

Consider first the filter expression (5.12). We may assume that after the second iteration, the soft estimates of transmitted symbols are reliable enough. In other words, we assume that $\underline{\hat{\boldsymbol{s}}}_k^i = \underline{\boldsymbol{s}}_k^i$ and as a result $\widetilde{\boldsymbol{\Lambda}}_{k,i} = \boldsymbol{\Lambda}_{k,i}$. With this approximation, the expression of $\mathbf{w}_k^i$ in (5.12) simplifies to the following expression [21]

$$\begin{aligned}\mathbf{w}_k^i &= \mathbf{h}_i^\dagger \left[\mathbf{h}_i\mathbf{h}_i^\dagger + \frac{\sigma_z^2}{\sigma_{s_k^i}^2}\mathbb{I}_{M_R}\right]^{-1} \\ &= \frac{1}{\mathbf{h}_i^\dagger\mathbf{h}_i + \frac{\sigma_z^2}{\sigma_{s_k^i}^2}}\ \mathbf{h}_i^\dagger. \end{aligned} \tag{5.27}$$

where the filter coefficients do not depend on k anymore, since $\sigma_{s_k^i}^2$ will be constant for all k. In other words, with this approximation we do not need to update the MMSE filter for each sample-time since we effectively need to calculate $\mathbf{w}_k^i$ once for a given realization of $\mathbf{H}$, i.e., for the underlying fading block. Furthermore, the matrix inversion in (5.12) is replaced by a scalar inversion. Obviously, this simplification causes a degradation to the receiver performance. Yet, this performance loss would be justified, regarding the complexity reduction, and hopefully, thanks to iterative detection, it would be acceptable. We will compare the performance of this simplified detector with the exact PIC detector in Section 5.6.

Notice that unless otherwise mentioned, in this chapter, by turbo-PIC under perfect CSIR, we mean the exact formulation of the soft-PIC detector, i.e., expression (5.12).

5.5.2 Case of Improved Detector and Imperfect CSIR

It can be easily verified that, replacing $\widetilde{\boldsymbol{\Lambda}}_{k,i}$ by $\boldsymbol{\Lambda}_{k,i}$ in (5.20) does unfortunately not lead to a compact expression involving a simple scalar inversion. However, with a slight modification in the interference cancellation part in (5.8), we can also derive a simple improved detector similarly to the expression (5.27). This is achieved by cancelling the residual interference by using (5.8) instead of (5.17) as

$$\begin{aligned}\underline{\boldsymbol{y}}_k^i &= \boldsymbol{y}_k - \underline{\mathbf{H}}_i\,\underline{\hat{\boldsymbol{s}}}_k^i \\ &= \mathbf{h}_i s_k^i + \underline{\mathbf{H}}_i\,\underline{\boldsymbol{s}}_k^i - \underline{\mathbf{H}}_i\,\underline{\hat{\boldsymbol{s}}}_k^i + \boldsymbol{z}_k, \end{aligned} \tag{5.28}$$

By following a similar procedure to the one of Section 5.4, it can then be shown that the improved MMSE filter of (5.20) reduces to

$$\widetilde{\mathbf{w}}_k^i = \delta\,\widehat{\mathbf{h}}_i^\dagger \left[\delta^2\,\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^\dagger + (1-\delta)\left(1 + \frac{\mathrm{tr}\big(\boldsymbol{\Lambda}_{k,i} - \widetilde{\boldsymbol{\Lambda}}_{k,i}\big)}{\sigma_{s_k^i}^2} + \frac{\sigma_z^2}{(1-\delta)\,\sigma_{s_k^i}^2}\right)\mathbb{I}_{M_R} + \frac{\delta^2\,\underline{\widehat{\mathbf{H}}}_i\big(\boldsymbol{\Lambda}_{k,i} - \widetilde{\boldsymbol{\Lambda}}_{k,i}\big)\,\underline{\widehat{\mathbf{H}}}_i^\dagger}{\sigma_{s_k^i}^2}\right]^{-1}. \tag{5.29}$$

Now, by setting $\widetilde{\mathbf{\Lambda}}_{k,i}$ equal to $\mathbf{\Lambda}_{k,i}$ in (5.29), we obtain the following simplified expression for the MMSE filter of the improved detector:

$$\begin{aligned}\widetilde{\mathbf{w}}_k^i &= \delta\,\widehat{\mathbf{h}}_i^\dagger\left[\delta^2\,\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^\dagger + (1-\delta)\mathbb{I}_{M_R} + \frac{\sigma_z^2}{\sigma^2_{s_k^i}}\mathbb{I}_{M_R}\right]^{-1} \\ &= \frac{\delta}{\delta^2\widehat{\mathbf{h}}_i^\dagger\widehat{\mathbf{h}}_i + (1-\delta) + \frac{\sigma_z^2}{\sigma^2_{s_k^i}}}\;\widehat{\mathbf{h}}_i^\dagger. \end{aligned} \tag{5.30}$$

The modification of the interference cancellation and the calculation of the LLRs are as before, provided by equations (5.24) and (5.26), respectively.

5.6 Simulation Results and Discussions

We now provide some numerical results to show the performance improvement by using the proposed modified detector in the presence of channel estimation errors. The performance is evaluated in terms of receiver bit error rates (BER).

For channel coding in our BICM scheme, we consider the 1/2 rate NRNSC code of constraint length $K = 7$, defined in octal form by $(133, 171)_8$. Uncorrelated Rayleigh fading channel is considered and each frame of symbols corresponds to N_c fading blocks. Channel coefficients are kept constant during each fading block and changed to new independent realizations from one block to the next. Corresponding to each fading block, we devote N_P channel-uses to the transmission of pilot sequences. Throughout the simulations, each frame is composed of $L_f = 128$ channel-uses for data symbols plus a total number of N_cN_P channel-uses for pilot transmission. Data symbols belong to QPSK or 16-QAM constellations with Gray labeling. We use mutually orthogonal QPSK pilot sequences for channel estimation and the average pilot-symbol power is set equal to the average data-symbol power.

The interleaver is pseudo-random, operating over the entire frame of size $N_I = L_fBM_T$ bits (excluding pilots, obviously). Moreover, the number of receiver iterations is set to 5. The SNR is considered in the form of E_b/N_0 and includes the antenna array gain M_R at the receiver. We assume that the noise variance is known at the receiver.

5.6.1 Case of Turbo-PIC Detector with Spatial Multiplexing

For the case of spatial multiplexing, Fig. 5.3 shows BER curves of the mismatched and improved receivers for the case of QPSK modulation and $M_T = 2$ and $M_R = 2$ that we denote by (2×2) MIMO system. The number of channel-uses for pilot transmission is $N_P \in \{2, 4, 8\}$. As a reference, we have also presented the BER curve in the case of perfect CSIR. We see that the

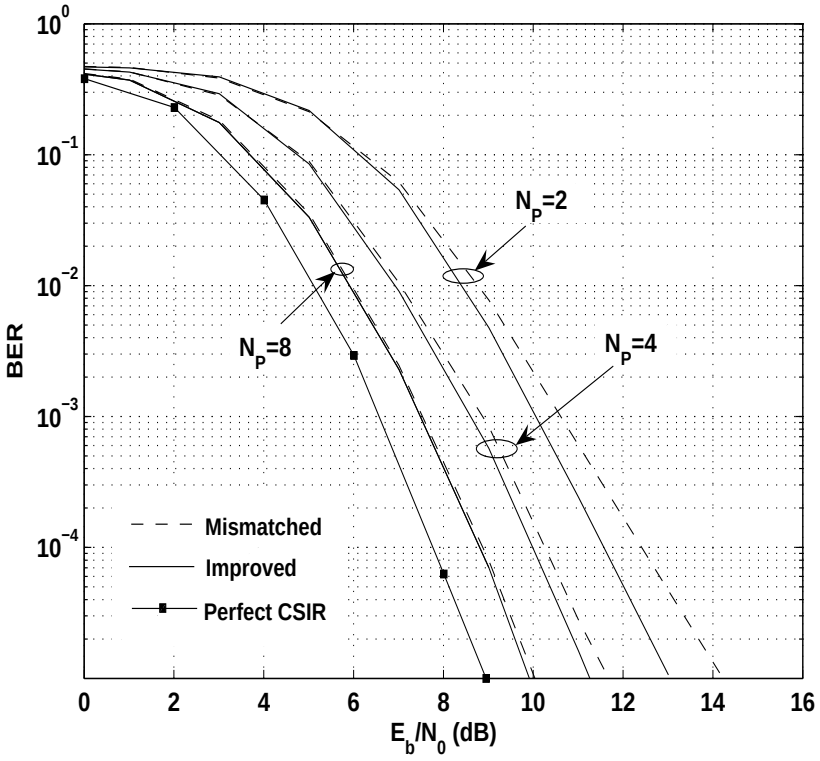

Figure 5.3: BER performance of improved and mismatched turbo-PIC; (2×2) MIMO with V-BLAST ST scheme, i.i.d. Rayleigh fading with $N_c = 4$, QPSK modulation, training sequence length $N_P \in \{2, 4, 8\}$.

gain in SNR of the improved detector to attain the BER of 10^{-5} is about 1.4 dB, 0.5 dB, and 0.2 dB, respectively for $N_P = 2$, 4, and 8.

Let us consider the case of the (2×2) system with now a 16-QAM modulation. Results are shown in Fig. 5.4. We notice that the gain in SNR by using the improved detector at a BER of 10^{-5} is now about 0.9 dB, 0.25 dB, and 0.05 dB, for $N_P = 2$, 4, and 8, respectively. The obtained gains by using the improved detector are thus less important for the 16-QAM than for the QPSK modulation. As a matter of fact, even with the perfect CSIR knowledge, for larger constellation sizes, turbo-PIC becomes more suboptimal, except in the presence of high (time or space) diversity. This can be seen in Fig. 5.4 by comparing the slopes of the BER curves for turbo-PIC and for the turbo-MAP detector in the case of perfect CSIR.[2] Consequently, in this case, the improved turbo-PIC does not offer a considerable gain, as compared to the mismatched detector.

[2]Notice that the turbo-MAP detector here becomes computationally complex and the simulations for obtaining the corresponding curve have taken a considerable long time.

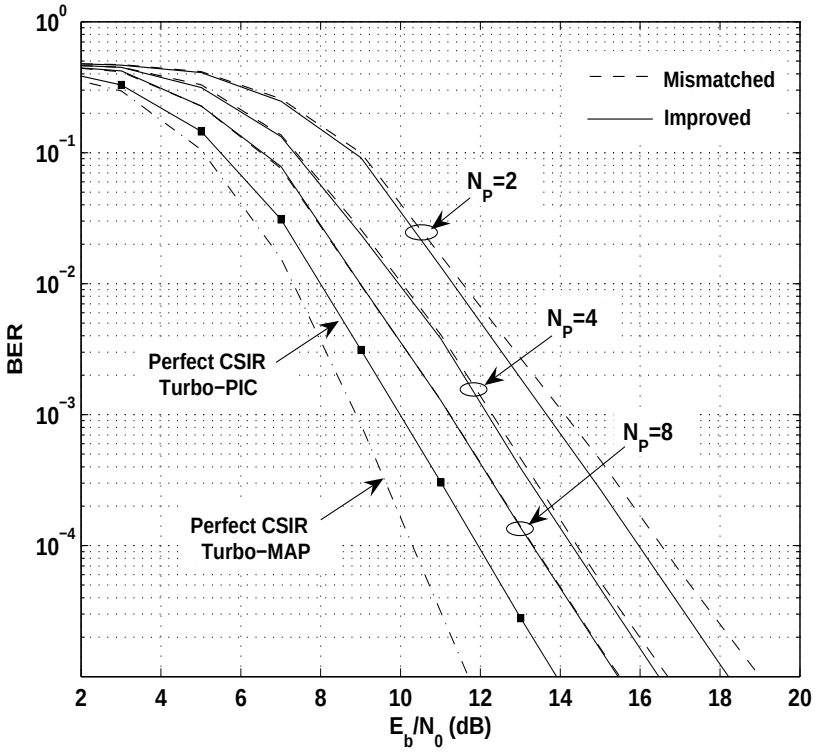

Figure 5.4: BER performance of improved and mismatched turbo-PIC; (2×2) MIMO with V-BLAST ST scheme, i.i.d. Rayleigh fading with $N_c = 4$, 16-QAM modulation, training sequence length $N_P \in \{2, 4, 8\}$.

5.6.2 Case of Turbo-PIC Detector with Space-time Coding

Up to now, we have considered the simple spatial multiplexing at the transmitter. It is interesting to study the performance of the improved detector for *more powerful* space-time codes. We consider here, as the ST scheme, the optimized *Golden code* (denoted here by GLD), presented in [166] for the case of two transmit antennas and $M_R \geq M_T$, which offers full-rate full-diversity. With this ST scheme, each vector of four symbols $\boldsymbol{s} = \left[s^1\ s^2\ s^3\ s^4\right]^T$ is mapped into a (2×2) matrix $\mathbf{X}$ as described below:

$$\mathbf{X} = \frac{1}{\sqrt{5}} \begin{bmatrix} \alpha\ [s^1 + \theta\ s^2] & \alpha\ [s^3 + \theta\ s^4] \\ \gamma\ \overline{\alpha}\ [s^3 + \overline{\theta}\ s^4] & \overline{\alpha}\ [s^1 + \overline{\theta}\ s^2] \end{bmatrix}, \tag{5.31}$$

where

$$\theta = \frac{1+\sqrt{5}}{2}\ \ ,\ \ \alpha = 1 + j(1-\theta)\ \ ,\ \ \overline{\theta} = 1-\theta\ \ ,\ \ \overline{\alpha} = 1 + j(1-\overline{\theta})\ \ ,\ \ \gamma = j\ \ ,\ \ j = \sqrt{-1}.$$

Figure 5.5 contrasts the BER curves of the receiver for the cases of the improved and of the mismatched detectors. It is seen that the performance gain by using the improved detector can be quite considerable in this case. The SNR gain at a BER of 10^{-5} is about 2.3 dB, 0.9 dB, and 0.4 dB, for $N_P = 2$, 4, and 8, respectively.

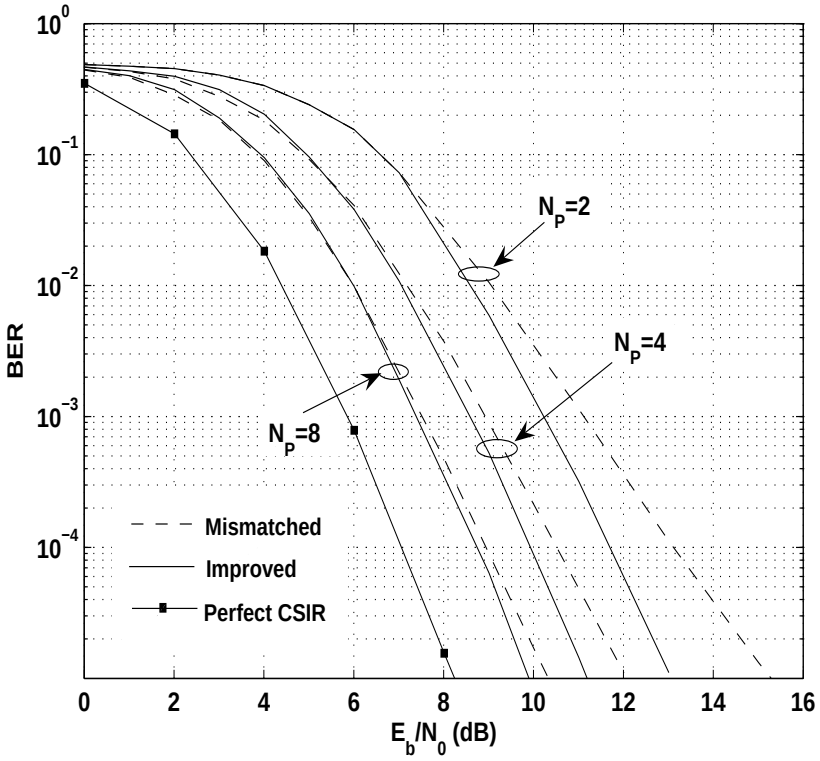

Figure 5.5: BER performance of improved and mismatched turbo-PIC; (2×2) MIMO with GLD ST scheme, i.i.d. Rayleigh fading with $N_c = 4$, QPSK modulation, training sequence length $N_P \in \{2, 4, 8\}$.

5.6.3 Case of Simplified Turbo-PIC Detector

Let us now consider the case of the simplified soft-PIC and study the performance gain by using the improved detector. Remember that here both the mismatched and the improved simplified detectors need to calculate the MMSE filter once per channel realization, and not for each channel-use, as it is the case for the exact implementation of the soft-PIC.

First, we have compared in Fig. 5.6 the performances of the exact and of the simplified turbo-PIC for the case of perfect CSIR, where a (2×2) system is considered with the simple V-BLAST (for QPSK and 16-QAM modulations) and with the GLD scheme (for a QPSK modulation). We notice that the performance degradation by using the simplified soft-PIC detector at a BER of 10^{-5} is about 0.6 dB and 0.7 dB, respectively for V-BLAST and GLD schemes with a QPSK modulation, and of more than 3 dB for the V-BLAST scheme with a 16-QAM modulation. In effect, for large constellation sizes, the performance degradation by the simplification made in soft-PIC becomes considerable.

Now, consider Fig. 5.7 which compares the performances of the mismatched and of the improved simplified detectors under the same conditions than that of Fig. 5.3, that is, for a V-BLAST scheme with a QPSK modulation. We notice that the SNR gain by using the improved

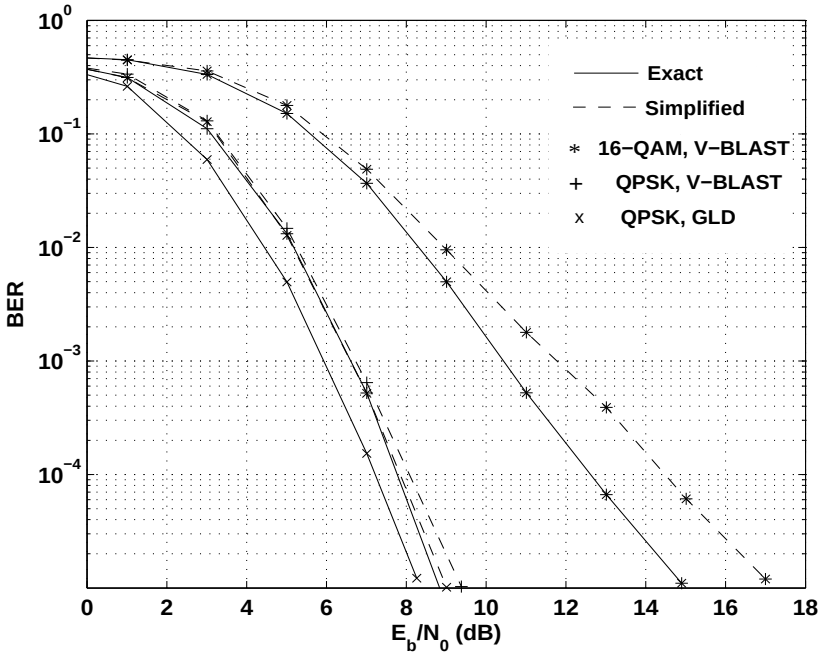

Figure 5.6: BER performance of exact and simplified turbo-PIC implementations; (2×2) MIMO, i.i.d. Rayleigh fading with $N_c = 4$, perfect CSIR.

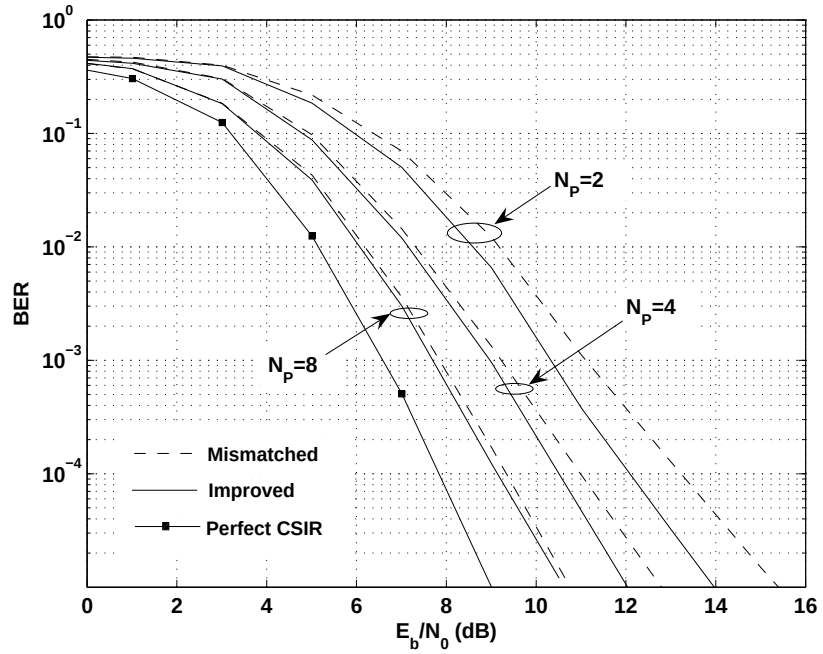

Figure 5.7: BER performance of improved and mismatched simplified turbo-PIC; (2×2) MIMO with V-BLAST ST scheme, i.i.d. Rayleigh fading with $N_c = 4$, QPSK modulation, training sequence length $N_P \in \{2, 4, 8\}$.

detector is as important as in the case of the exact formulation of the soft-PIC and is about 1.4 dB, 0.7 dB, and 0.2 dB, for $N_P = 2$, 4, and 8, respectively.

5.7 Conclusion

By introducing a Bayesian approach characterizing the channel estimation process, we proposed an improved low-complexity turbo-PIC detector that mitigates the impact of channel uncertainty on the detection performance. Both cases of exact and simplified implementations of the soft-PIC detector were treated. The formulation of the improved detector was provided for the simple case of the V-BLAST scheme. We also provided its generalization to the case of an arbitrary ST scheme.

We found that the classically-used mismatched detector becomes largely suboptimal compared to the proposed detector, especially when only a few number of pilots is dedicated for channel estimation. **The important point is that the performance improvement is obtained while imposing practically no additional complexity to the receiver**.

For ST schemes designed by imposing more constraints on the coding rate and/or diversity, the receiver is more sensitive to the channel estimation errors, and the improved detector provides even more important gains. For instance, for the case of the full-rate full-diversity GLD scheme, we noticed a considerable SNR gain of 2.3 dB offered by the improved detector.

Chapter 6

Conclusions and Perspectives

The objective of this thesis was to study the problem of iterative data detection in a realistic wireless communication system, where the receiver disposes only of an *imperfect* (and possibly very noisy) estimate of the unknown channel parameters. To obtain the CSIR, we considered a commonly-used approach which consists in sending some training sequences at the beginning of the information frame. Obviously, obtaining an accurate estimate of the channel through the use of training symbols, would require a large number of pilots per frame, which can result in a considerable reduction of the system throughput due to the pilot overhead. Thus, in practice, the widely-used assumption of perfect CSIR is not a valid assumption. This brings many theoretical and practical challenges which motivated our research. For instance, the well-established theoretical capacity limits and their associated so-called optimal decoding schemes may turn out to be somewhat limited in the described scenario. Furthermore, new transceiver structures designed by taking into account the presence of channel uncertainty should be investigated. The work presented in this thesis tried to establish a connection between the common pilot-only based channel estimation technique and the design of practical iterative (turbo) reception schemes which have been shown to perform close to the ultimate Shannon limit. This research led to some theoretical and practical results and opened the issue for further investigations.

In **Chapter 2** we provided an overview of the UWB technology and introduced the MB-OFDM transmission scheme which is the application scenario considered in this work. However, the principle and ideas presented in this thesis are also applicable to conventional OFDM systems.

Our first step toward the design of an improved receiver in the presence of imperfect channel estimation was presented in **Chapter 3**. In this contribution, we proposed a semi-blind joint channel estimation and data detection algorithm based on the EM algorithm that integrates the advantages of wavelet based parameter estimation. By expressing the unknown UWB channel

in terms of its discrete wavelet coefficients, we were able to choose a prior distribution that captures the sparsness property of UWB channels in the wavelet domain. This led to a MAP estimator equivalent to a hard thresholding procedure at each iteration of the EM algorithm, which was used to reduce the number of estimated coefficients. It was observed that when the channel has a sparse wavelet expansion, the proposed scheme provides significant reduction in the number of estimated parameters (in average by about 60 %) and outperforms pilot-only based and traditional semi-blind methods.

A direct extension of this work is to consider a MB-OFDM system working over larger bandwidths (i.e., more than the first three subbands considered in this thesis). This leads to a sparser channel representation in the wavelet domain; consequently the considered prior model becomes more informative and the algorithm will likely reduce furthermore the number of estimated parameters. Another unexplored area is the use of the time-scale wavelet representation for modeling the time-selectivity of OFDM channels rather than using the standard AR modeling. Using the wavelet decomposition to model the two-dimensional OFDM channel with only few parameters known a priori at the receiver (similar to the scheme proposed in [6]), constitutes another possible extension of our work for reducing the receiver complexity.

In **Chapter 4**, by using the statistics of the channel estimation errors, we introduced a Bayesian framework; this allowed us to formulate an improved ML detection metric as an alternative to the sub-optimal mismatched ML metric which classically replaces the unknown channel by its estimate. Using this, we were able to integrate the imperfect channel knowledge into the design of turbo receivers. Then, we proposed to characterize the achieved throughputs versus outage associated to both improved and mismatched ML metrics. We also provided a comparison with the rates achieved by a theoretical decoder, defined as the best decoder in the presence of channel estimation errors. Our numerical results conducted for both single- and multi-antenna MB-OFDM systems showed a considerable amount of performance degradation in terms of BER and achievable information rates for the case of the mismatched detection, and confirmed the adequacy of the improved detector under imperfect channel estimation. In addition to SNR reduction, we observed that the improved turbo receiver also reduces the number of pilots as well as the number of decoder iterations required to achieve a target BER. A practical application of this research is in the evaluation of the trade-off between the required QoS (in terms of BER and achieved throughputs) and the system parameters (e.g., training power, transmission power, period of training, number of decoding iterations, outage probability).

Although we showed that the improved detector outperforms the classical mismatched detector in terms of achieved throughputs, there still remains a considerable SNR gap between the improved and the theoretical detector. Thus, the derivation of other detection metrics taking

into account the channel uncertainty and providing closer rates to the theoretical capacity limit is still an open problem.

The improved ML metric used in our work can also be applied for iterative decoding with other codes than simple convolutional codes. For instance, the decoding algorithms of low-density parity-check codes (LDPC) are based on a *Belief Propagation* principle [167, 168] which is very similar to the evaluation of LLRs in turbo-codes. Hence, the same modifications done in our work for turbo detectors can also be applied to improve the performance of LDPC codes under channel estimation errors. Another possible application of our results concerns turbo decoding of trellis-coded modulations impaired by channel estimation errors.

By transmitting pilots at the beginning of each communication session, we have intrinsically assumed a time-invariant channel during the whole period of data transmission. For a time-varying channel, this technique wastes bandwidth, causing a reduction of the overall data rate. For such situations, it might be beneficial to arithmetically *superimpose* the pilots to the data. The idea of superimposing known sequences to data is closely related to digital watermarking techniques, and the connection between watermarking and communications with side information has been made in [169]. The use of superimposed training (SIT) for channel estimation was introduced in [170]. However, there are several obstacles in practice. The most important is the degradation of the channel estimate quality due to the presence of unknown data during channel estimation, which obviously affects the BER performance. Our work concerning conventional pilot transmission constitutes the basis for further research on developing improved receivers for wireless systems based on SIT.

As an attractive alternative for turbo-MAP detectors, we introduced in **Chapter 5** a low-complexity detector based on soft-PIC and MMSE filtering. By using the same Bayesian framework as in Chapter 4, we were able to reduce the impact of channel uncertainty on the detection performance. Both cases of the exact and of the simplified implementations were treated. We found that the classically-used mismatched detector becomes once again largely sub-optimal compared to the proposed detector, especially when few number of pilots are dedicated to channel estimation. The most considerable sensitivity to channel uncertainty was observed for space-time (ST) coded systems. For instance, for the case of full-rate full diversity Golden code [166], we noticed a gain of 2.3 dB offered by the improved soft-PIC detector. This large performance gain motivates us to investigate in the future, the combination of our improved soft-PIC detector with other orthogonal and non-orthogonal ST codes.

One has to remember that in this thesis we have exclusively paid attention to transmission scenarios with no channel information at the transmitter. For improving the performance, one can estimate the channel and feed back this information to the transmitter in order to adapt the

modulation and the channel code and to optimally allocate the power according to the channel response. A future research direction may include providing informations regarding the quality of the channel estimate via a limited feedback at the transmitter. For example, the pdf of the channel given its estimate at the receiver can be partially fed back to the transmitter. In that case, depending on the feedback quality (e.g., the number of bits used for quantization), the channel uncertainty at the transmitter is likely to be larger than that at the receiver. In this regard, establishing a general statistical (Bayesian) framework at the transmitter, including both the effect of the channel estimation errors and of the degradation due to the limited feedback is a chalenging topic to investigate.

Finally, we notice that the improved detection schemes derived in Chapters 4 and 5 of this thesis require complete knowledge of the statistics of the fading process as well as those of the channel estimation errors. However, in some scenarios these distribution may not be available and consequently the statistical information about the channel estimation process (in this thesis, the pdf of the perfect channel given its estimate) cannot be computed. Designing improved receivers for these scenarios leads to a different mathematical framework and brings new challenges that would be interesting to investigate in future research.

Appendix A

Additional Computations Related to Chapter 3

A.1 Proof of the Equivalence Between Expressions (3.32) and (3.33)

Let us define $\mathbf{A} = \langle \widetilde{\mathbf{H}}^{(t)} \rangle$ for simplicity. We can expand (3.32) as

$$\begin{aligned}
\mathbf{g}^{(t+1)} &= \arg\max_{\mathbf{g}} \left\{ -\frac{\| \mathbf{A} - \mathbf{T}\mathbf{g} \|^2}{\alpha^2} + \log \pi(\mathbf{g}) \right\} \\
&= \arg\max_{\mathbf{g}} \left\{ -\frac{\mathbf{A}^\dagger \mathbf{A} - 2\,\mathbb{R}e(\mathbf{A}^\dagger \mathbf{T}\mathbf{g}) + \mathbf{g}^\dagger \mathbf{T}^\dagger \mathbf{T}\mathbf{g}}{\alpha^2} + \log \pi(\mathbf{g}) \right\} \\
&= \arg\max_{\mathbf{g}} \left\{ -\frac{\mathbf{A}^\dagger \mathbf{A} - 2\,\mathbb{R}e(\mathbf{A}^\dagger \mathbf{T}\mathbf{g}) + \mathbf{g}^\dagger \mathbf{g}}{\alpha^2} + \log \pi(\mathbf{g}) \right\} \qquad \text{(A.1)}
\end{aligned}$$

where in (A.1) we have used the fact that $\mathbf{T}^\dagger \mathbf{T} = \mathbb{I}_L$.

Similarly expanding (3.33) for $\widetilde{\mathbf{g}}^{(t)} = \mathbf{T}^\dagger \mathbf{A}$ yields

$$\begin{aligned}
\mathbf{g}^{(t+1)} &= \arg\max_{\mathbf{g}} \left\{ -\frac{\| \mathbf{T}^\dagger \mathbf{A} - \mathbf{g} \|^2}{\alpha^2} + \log \pi(\mathbf{g}) \right\} \\
&= \arg\max_{\mathbf{g}} \left\{ -\frac{\mathbf{A}^\dagger \mathbf{T}\mathbf{T}^\dagger \mathbf{A} - 2\mathbb{R}e(\mathbf{A}^\dagger \mathbf{T}\mathbf{g}) + \mathbf{g}^\dagger \mathbf{g}}{\alpha^2} + \log \pi(\mathbf{g}) \right\} \qquad \text{(A.2)}
\end{aligned}$$

We notice that due the truncation of the FFT matrix $\mathbf{F}_{M,L}$, we have $\mathbf{T}\mathbf{T}^\dagger \neq \mathbb{I}_M$ and consequently (A.1) and (A.2) have different arguments. However, it is easy to see that (A.2) can be written as

$$\mathbf{g}^{(t+1)} = \arg\max_{\mathbf{g}} \left\{ -\frac{\| \mathbf{A} - \mathbf{T}\mathbf{g} \|^2}{\alpha^2} + \log \pi(\mathbf{g}) + K \right\} \qquad \text{(A.3)}$$

where

$$K = \frac{\mathbf{A}^\dagger \mathbf{A} - \mathbf{A}^\dagger \mathbf{T}\mathbf{T}^\dagger \mathbf{A}}{\alpha^2}$$

is a constant term not depending on $\mathbf{g}$.

Thus, from (A.3) and (A.1), the equivalence between (3.32) and (3.33) is straightforward.

A.2 Derivation of the MAP Estimate (3.42)

From (3.35), we derive the simplified model for a given wavelet coeficient

$$\widetilde{g} = g + Z_1' \tag{A.4}$$

where $Z_1' \sim \mathcal{CN}(0, \alpha^2)$; the time index t and the coefficient index j ($j = 1, ..., L$) have been omitted for simplicity. Our aim is to estimate (or to update) from (A.4), the coefficient g in the MAP sense as

$$\begin{aligned} \hat{g}_{\text{map}} &= \arg\max_g \left\{ p(g|\hat{g}) \right\} \\ &= \arg\max_g \left\{ p(\hat{g}|g)\ \pi_{\text{MOG}}(g) \right\} \end{aligned} \tag{A.5}$$

where

$$\pi_{\text{MOG}}(g) = \lambda \mathcal{CN}_g(0, \tau_1^2) + (1-\lambda)\mathcal{CN}_g(0, \tau_2^2)$$

is the prior distribution chosen for the unknown wavelet coefficient g. Our derivation is based on this more general prior called mixture of Gaussians (MOG). Then, we address the case of the Bernoulli-Gaussian prior model as a special case of MOG.

The a posteriori density $p(g|\hat{g})$ is proportional to

$$p(g|\widetilde{g}) \propto p(\widetilde{g}|g)\big(\lambda \mathcal{CN}_g(0, \tau_1^2) + (1-\lambda)\mathcal{CN}_g(0, \tau_2^2)\big). \tag{A.6}$$

Actually, from (A.6) it is clear that maximizing directly the logarithm of $p(\widetilde{g}|g)$ requires the maximization of the logarithm of a sum, which is not evident analytically. In order to avoid this problem, we propose to introduce a state variable (or indicator) $\beta \in \{0, 1\}$. This indicator variable enables us to express the MOG prior model as

$$\begin{cases} (g|\beta = 0) & \sim & \mathcal{CN}_g\left(0, \tau_1^2\right) & \text{with probability } \lambda, \\ \\ (g|\beta = 1) & \sim & \mathcal{CN}_g\left(0, \tau_2^2\right) & \text{with probability } 1-\lambda. \end{cases} \tag{A.7}$$

Furthermore we have

$$p(\beta = 0) = \lambda, \text{ and } p(\beta = 1) = 1 - \lambda.$$

By using the indicator variable β, we can express the MAP estimation in (A.5) as

$$\begin{aligned}\hat{g}_{\text{map}} &= \arg\max_{g} \left\{ p(g, \beta|\widetilde{g}) \right\} \\ &= \arg\max_{g} \left\{ p(g|\widetilde{g}, \beta)\, p(\beta|\widetilde{g}) \right\}. \end{aligned} \tag{A.8}$$

In what follows, we evaluate $p(\beta|\widetilde{g})$ and $p(g|\widetilde{g}, \beta)$.

Evaluation of $p(\beta|\widetilde{g})$: We have

$$p(\beta = 0|\widetilde{g}) \propto p(\widetilde{g}|\beta = 0)\, p(\beta = 0), \quad \text{and} \quad p(\beta = 1|\widetilde{g}) \propto p(\widetilde{g}|\beta = 1)\, p(\beta = 1).$$

It is clear from (A.4) that conditioned on $\beta = 0$ and $\beta = 1$, $\widetilde{g}$ is distributed as $\mathcal{CN}(0, \alpha^2 + \tau_1^2)$ and $\mathcal{CN}(0, \alpha^2 + \tau_2^2)$, respectively. Thus, it is easy to see that

$$\begin{aligned} p\left(\beta = 0|\widetilde{g}\right) &= \lambda\, \mathcal{N}\left(0, \alpha^2 + \tau_1^2\right)/c \\ p\left(\beta = 1|\widetilde{g}\right) &= (1 - \lambda)\, \mathcal{N}\left(0, \alpha^2 + \tau_2^2\right)/c \end{aligned} \tag{A.9}$$

where the constant $c = \lambda \mathcal{N}\left(0, \alpha^2 + \tau_1^2\right) + (1 - \lambda)\mathcal{N}\left(0, \alpha^2 + \tau_2^2\right)$.
The indicator variable can be estimated in the MAP sense as

$$\beta^* = \begin{cases} 0, & \text{if } \; p\left(\beta = 0|\widetilde{g}\right) \geq 0.5 \\ \\ 1, & \text{if } \; p\left(\beta = 1|\widetilde{g}\right) \geq 0.5. \end{cases} \tag{A.10}$$

Evaluation of $p(g|\widetilde{g}, \beta)$: We have

$$p(g|\widetilde{g}, \beta) \propto p(\widetilde{g}|g, \beta)\, p(g|\beta).$$

It is important to notice that at this step, all the pdfs are conditioned on the estimated indicator variable β. That is, the pdf $p(g|\widetilde{g}, \beta)$ is expressed as

$$p(g|\widetilde{g}, \beta = \beta^*) \propto p(\widetilde{g}|g, \beta = \beta^*)\, p(g|\beta = \beta^*). \tag{A.11}$$

Since conditioned on g, $\widetilde{g}$ is independent of β, we have $p(\widetilde{g}|g, \beta = \beta^*) = p(\widetilde{g}|g)$. Moreover, according to (A.7), $p(g|\beta = \beta^*) = \mathcal{CN}_g\left(0, \tau_*^2\right)$ where $\tau_*^2 = \tau_1^2$ if $\beta^* = 0$ and $\tau_*^2 = \tau_2^2$ if $\beta^* = 1$. Thus, (A.10) is given by

$$p(g|\widetilde{g}, \beta = \beta^*) \propto \mathcal{CN}_g(g, \alpha^2)\; \mathcal{CN}_g(0, \tau_*^2). \tag{A.12}$$

Now, we are able to write the maximization problem in (A.8) as

$$\begin{aligned}\hat{g}_{\text{map}} &= \arg\max_{g} \left\{ \mathcal{CN}_g(g, \alpha^2)\; \mathcal{CN}_g(0, \tau_*^2)\; p(\beta|\widetilde{g}) \right\} \\ &= \arg\max_{g} \left\{ \mathcal{CN}_g(g, \alpha^2)\; \mathcal{CN}_g(0, \tau_*^2) \right\} \end{aligned} \tag{A.13}$$

Note that according to (A.9), $p(\beta|\widetilde{g})$ does not depend on g. This pdf is only used for the MAP estimation of the indicator variable in (A.10).

It is easy to see from (A.12) that

$$\log p(g|\widetilde{g}, \beta = \beta^*) = K - \frac{|\widetilde{g} - g|^2}{\alpha^2} - \frac{|g|^2}{\tau_*^2}, \tag{A.14}$$

where K is a constant parameter.

Taking the derivative of (A.14) with respect to g, and after some algebra we get

$$\hat{g}_{\text{map}} = \begin{cases} \dfrac{\tau_1^2}{\alpha^2 + \tau_1^2}\, \widetilde{g}, & \text{if } \beta^* = 0 \\ \\ \dfrac{\tau_2^2}{\alpha^2 + \tau_2^2}\, \widetilde{g}, & \text{if } \beta^* = 1. \end{cases} \tag{A.15}$$

The solution $\hat{g}_{\text{map}}$ in (A.15) is associated with the MOG prior model. We note that when the variance τ_1^2 tends to zero, the MOG model tends to the Bernoulli-Gaussian prior model, defined by

$$\pi_{\text{BG}}(g) = \lambda\, \delta(g) + (1 - \lambda)\, \mathcal{CN}_g(0, \tau_2^2).$$

Hence, the MAP estimate associated to the Bernoulli-Gaussian prior model $\hat{g}_{\text{map}}^{\text{BG}}$, can be derived from (A.15), by evaluating the limit of this solution when τ_1^2 tends to zero.

We obtain

$$\hat{g}_{\text{map}}^{\text{BG}} = \begin{cases} 0, & \text{if } \beta^* = 0 \\ \\ \dfrac{\tau_2^2}{\alpha^2 + \tau_2^2}\, \widetilde{g}, & \text{if } \beta^* = 1 \end{cases} \tag{A.16}$$

which is the expression (3.42).

Appendix B

Additional Computations Related to Chapter 4

B.1 Derivation of the A Posteriori Probability (4.15)

The following theorem is derived in [171].

Theorem B.1.1. *Let* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *be circularly symmetric complex Gaussian random vectors with zero means and full-rank covariance matrices* $\mathbf{\Sigma}_{ij} = \mathbb{E}[\mathbf{x}_i \mathbf{x}_j^\dagger]$. *Then the conditional random vector* $\mathbf{x}_1 | \mathbf{x}_2 \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ *is circularly symmetric complex Gaussian with mean* $\boldsymbol{\mu} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{x}_2$ *and covariance matrix* $\mathbf{\Sigma} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$.

We denote by $\mathbf{h}_{k,i}$ and $\widehat{\mathbf{h}}_{k,i}$ the i-th column of matrices $\mathbf{H}_k$ and $\widehat{\mathbf{H}}_k$, respectively; we set $\mathbf{x}_1 = \mathbf{h}_{k,i}$ and $\mathbf{x}_2 = \widehat{\mathbf{h}}_{k,i}$. From equations (4.13) and (4.14), we have $\mathbf{\Sigma}_{11} = \mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{H_k}$ and $\mathbf{\Sigma}_{22} = \mathbf{\Sigma}_{H_k} + \mathbf{\Sigma}_{\mathcal{E}}$ in Theorem B.1.1. According to this theorem, the conditional pdf of $\mathbf{h}_{k,i} | \widehat{\mathbf{h}}_{k,i}$ is a circularly symmetric complex Gaussian distribution with

$$\text{mean} = \mathbf{\Sigma}_\Delta \widehat{\mathbf{h}}_{k,i} \quad \text{where} \quad \mathbf{\Sigma}_\Delta \triangleq \mathbf{\Sigma}_{H_k}(\mathbf{\Sigma}_{H_k} + \mathbf{\Sigma}_{\mathcal{E}})^{-1} \quad \text{and} \tag{B.1}$$

$$\text{covariance matrix} = \mathbf{\Sigma}_{H_k} - \mathbf{\Sigma}_{H_k}(\mathbf{\Sigma}_{H_k} + \mathbf{\Sigma}_{\mathcal{E}})^{-1}\mathbf{\Sigma}_{H_k}^\dagger = \mathbf{\Sigma}_\Delta \mathbf{\Sigma}_{\mathcal{E}}. \tag{B.2}$$

The equivalence in (B.2) can be seen by left multiplying both sides of $(\mathbf{\Sigma}_{\mathcal{E}} + \mathbf{\Sigma}_{H_k}) - \mathbf{\Sigma}_{H_k}^\dagger = \mathbf{\Sigma}_{\mathcal{E}}$ by $\mathbf{\Sigma}_{H_k}(\mathbf{\Sigma}_{\mathcal{E}} + \mathbf{\Sigma}_{H_k})^{-1}$. Assuming the same covariance matrix for all columns of $\mathbf{H}_k$ and $\widehat{\mathbf{H}}_k$, we obtain the a posteriori pdf (4.15).

B.2 Evaluation of the Likelihood Function (4.24)

To evaluate the conditional expectation in (4.23), we use the following theorem from [172].

Theorem B.2.1. *For a circularly symmetric complex random vector* $\boldsymbol{u} \sim \mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$ *with mean* $\mathbf{m} = \mathbb{E}[\boldsymbol{u}]$ *and covariance matrix* $\mathbf{\Sigma} = \mathbb{E}[\boldsymbol{u}\boldsymbol{u}^\dagger] - \mathbf{m}\mathbf{m}^\dagger$, *and a Hermitian matrix* $\mathbf{A}$ *such that*

$\mathbb{I} + \mathbf{\Sigma A} > 0$, *we have*

$$\mathbb{E}_{\boldsymbol{u}}\Big[\exp\Big\{-\boldsymbol{u}^{\dagger}\mathbf{A}\boldsymbol{u}\Big\}\Big] = \frac{\exp\big\{-\mathbf{m}^{\dagger}\mathbf{A}(\mathbb{I}+\mathbf{\Sigma A})^{-1}\mathbf{m}\big\}}{\det\{\mathbb{I}+\mathbf{\Sigma A}\}}. \tag{B.3}$$

Let us define $\boldsymbol{u} = \boldsymbol{y}_k - \mathbf{H}_k\boldsymbol{s}_k$. Using the a posteriori distribution of (4.15) and after some algebra, we can derive the conditional pdf of $\boldsymbol{u}$ given $\boldsymbol{s}_k$ and $\widehat{\mathbf{H}}_k$ as $\boldsymbol{u}|(\boldsymbol{s}_k, \widehat{\mathbf{H}}_k) \sim \mathcal{CN}(\mathbf{m}_u, \mathbf{\Sigma}_u)$, where $\mathbf{m}_u = \boldsymbol{y}_k - \mathbf{\Sigma}_{\Delta}\widehat{\mathbf{H}}_k\boldsymbol{s}_k$ and $\mathbf{\Sigma}_u = \mathbf{\Sigma}_{\Delta}\mathbf{\Sigma}_{\mathcal{E}}\|\boldsymbol{s}_k\|^2$. We further define $\mathbf{A} = \mathbf{\Sigma}_z^{-1}$. By applying Theorem B.2.1, (4.23) is written as

$$\begin{aligned}\widetilde{W}(\boldsymbol{y}_k|\widehat{\mathbf{H}}_k, \boldsymbol{s}_k) &= \mathbb{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left[\frac{\exp\big\{-(\boldsymbol{y}_k - \mathbf{H}_k\boldsymbol{s}_k)^{\dagger}\ \mathbf{\Sigma}_z{}^{-1}(\boldsymbol{y}_k - \mathbf{H}_k\boldsymbol{s}_k)\big\}}{\det\big\{\pi\mathbf{\Sigma}_z\big\}}\right] \\ &= \frac{\exp\Big\{-(\boldsymbol{y}_k - \mathbf{\Sigma}_{\Delta}\widehat{\mathbf{H}}_k\boldsymbol{s}_k)^{\dagger}\ \mathbf{\Sigma}_z^{-1}(\mathbb{I} + \mathbf{\Sigma}_{\Delta}\mathbf{\Sigma}_{\mathcal{E}}\|\boldsymbol{s}_k\|^2\mathbf{\Sigma}_z^{-1})^{-1}(\boldsymbol{y}_k - \mathbf{\Sigma}_{\Delta}\widehat{\mathbf{H}}_k\boldsymbol{s}_k)\Big\}}{\det\big\{\pi\mathbf{\Sigma}_z(\mathbb{I} + \mathbf{\Sigma}_{\Delta}\mathbf{\Sigma}_{\mathcal{E}}\|\boldsymbol{s}_k\|^2\mathbf{\Sigma}_z^{-1})\big\}}\end{aligned} \tag{B.4}$$

Since $\mathbf{\Sigma}_z$, $\mathbf{\Sigma}_{\Delta}$ and $\mathbf{\Sigma}_{\mathcal{E}}$ are diagonal matrices, the latter equation is rewritten as

$$\begin{aligned}\widetilde{W}(\boldsymbol{y}_k|\widehat{\mathbf{H}}_k, \boldsymbol{s}_k) &= \frac{\exp\Big\{-\big(\boldsymbol{y}_k - \delta\widehat{\mathbf{H}}_k\boldsymbol{s}_k\big)^{\dagger}\big(\mathbf{\Sigma}_z + \delta\,\mathbf{\Sigma}_{\mathcal{E}}\|\boldsymbol{s}_k\|^2\big)^{-1}\big(\boldsymbol{y}_k - \delta\widehat{\mathbf{H}}_k\boldsymbol{s}_k\big)\Big\}}{\det\big\{\pi(\mathbf{\Sigma}_z + \delta\,\mathbf{\Sigma}_{\mathcal{E}}\|\boldsymbol{s}_k\|)\big\}} \\ &= \mathcal{CN}\Big(\delta\widehat{\mathbf{H}}_k\boldsymbol{s}_k\,,\ \mathbf{\Sigma}_z + \delta\,\mathbf{\Sigma}_{\mathcal{E}}\|\boldsymbol{s}_k\|^2\Big).\end{aligned} \tag{B.5}$$

B.3 Proof of the Equality (4.41)

Proof. Consider the quadratic expressions $Q_1(\mathbf{x}) = \|\mathbf{Ax}\|^2 + K_1$ and $Q_2(\mathbf{x}) = \|\mathbf{x}\|^2 + K_2$, where $\mathbf{A}$ is an arbitrary $(M_R \times M_T)$ complex matrix, $\mathbf{x}$ is a $(M_T \times 1)$ vector such that $Q_1, Q_2 > 0$ *almost surely* and K_1 and K_2 are real positive constants. The joint generating function of Q_1 and Q_2, is defined as $M_{Q_1,Q_2}(t_1, t_2) = \mathbb{E}_{\mathbf{x}}\big[\exp\big\{t_1Q_1(\mathbf{x}) + t_2Q_2(\mathbf{x})\big\}\big]$. After simple algebra, we obtain

$$M_{Q_1,Q_2}(t_1, t_2) = \exp\big\{t_1K_1 + t_2K_2\big\}\det\big\{\mathbb{I}_{M_T} - \big(t_1\bar{P}\mathbf{A}^{\dagger}\mathbf{A} + t_2\bar{P}\mathbb{I}_{M_T}\big)\big\}^{-1}. \tag{B.6}$$

Then, from the gamma integral and setting $t_2 = -z$ in (B.6) we have

$$\mathbb{E}_{\mathbf{x}}\Big[Q_1(\mathbf{x})Q_2^{-1}(\mathbf{x})\Big] = \int_0^{\infty}\mathbb{E}_{\mathbf{x}}\Big[Q_1(\mathbf{x})\exp\big\{-zQ_2(\mathbf{x})\big\}\Big]\,\mathrm{d}z, \tag{B.7}$$

where it is not difficult to show that

$$\begin{aligned}\mathbb{E}_{\mathbf{x}}\Big[Q_1(\mathbf{x})\exp\big\{-zQ_2(\mathbf{x})\big\}\Big] &= \frac{\partial M_{Q_1,Q_2}(t_1, -z)}{\partial t_1}\Big|_{t_1=0}, \\ &= \big[K_1 + \mathrm{tr}(\bar{P}\mathbf{A}\mathbf{A}^{\dagger})(1 + z\bar{P})^{-1}\big] \\ &\quad \times(1 + z\bar{P})^{-(M_T/2)}\exp\big\{-K_2z\big\}.\end{aligned} \tag{B.8}$$

Finally, Lemma 4.6.2 follows by solving the integral in (B.7), which leads to the equality (4.41).

□

Appendix C

Derivation of the Achievable Information Rates for the Case of Single-antenna MB-OFDM Systems

C.1 Preliminaries

Channel Model: The single-antenna MB-OFDM system model with M subcarriers per subband is written as:

$$\boldsymbol{y} = \mathbf{H}_d\, \boldsymbol{s} + \boldsymbol{z}, \tag{C.1}$$

where the $(M \times 1)$ vectors $\boldsymbol{y}$ and $\boldsymbol{s}$ respectively denote the received and transmitted symbols; the noise vector $\boldsymbol{z}$ is assumed to be zero-mean circularly symmetric complex Gaussian (ZMCSCG) with distribution $\boldsymbol{z} \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2\, \mathbf{I}_M)$; and $\mathbf{H}_d = \mathrm{diag}(\mathbf{H})$ is the $(M \times M)$ diagonal channel matrix with diagonal elements given by the vector $\mathbf{H} = [H_0, \ldots, H_{M-1}]^T$, where $H_k = \sum_{l=0}^{L-1} h_l\, e^{-j2\pi kl/M}$.
In the following, when there is no confusion, we use $\mathbf{H}$ instead of $\mathbf{H}_d$.

Improved and Mismatched ML Detection Metric: Following a similar approach as in Chapter 4, the mismatched $\mathcal{D}_{\text{MM}}$ and improved $\mathcal{D}_{\mathcal{M}}$ ML metrics can be obtained as follows [147]:

$$\mathcal{D}_{\text{MM}}(\boldsymbol{s}, \boldsymbol{y}, \widehat{\mathbf{H}}) = \|\boldsymbol{y} - \widehat{\mathbf{H}}_d\, \boldsymbol{s}\|^2, \tag{C.2}$$

$$\mathcal{D}_{\mathcal{M}}(\boldsymbol{s}, \boldsymbol{y}, \widehat{\mathbf{H}}) = \sum_{k=0}^{M-1} \log\, \pi\big(\sigma_z^2 + \delta\, \sigma_{\mathcal{E}}^2\, |s_k|^2\big) + \frac{\big|y_k - \delta\, \widehat{H}_k\, s_k\big|^2}{\sigma_z^2 + \delta\, \sigma_{\mathcal{E}}^2\, |s_k|^2}. \tag{C.3}$$

C.2 Instantaneous Achievable Information Rates of MB-OFDM

In this section, we derive the instantaneous achievable information rates $C_{\mathcal{M}}$ and C_{MM} associated to a receiver based on the improved and mismatched detection rules given by (C.3) and (C.2), respectively. To this end, we apply the following theorem provided in [155], to a transmission characterized by the likelihood $W(\boldsymbol{y}|\boldsymbol{s},\mathbf{H}) = \prod_{k=0}^{M-1} \mathcal{CN}(H_k s_k, \sigma_z^2)$ which is a mapping from the input symbols $\boldsymbol{s} \in \mathcal{S}$ with distribution $P_{\mathbf{S}}$, to the set of probability measures on the output alphabet $\boldsymbol{y} \in \mathcal{Y}$. Furthermore, for any input distribution $P_{\mathbf{S}}$, we define the output distribution as $P_{\mathbf{Y}} = \int W(\boldsymbol{y}|\boldsymbol{s},\mathbf{H})\,\mathrm{d}P_{\mathbf{S}}(\boldsymbol{s})$.

Theorem C.2.1. *Given a general detection metric $\mathcal{D}_{\mathcal{G}}(\boldsymbol{s},\boldsymbol{y},\widehat{\mathbf{H}})$ a channel and its estimate $(\mathbf{H},\widehat{\mathbf{H}})$, a detector using the metric $\mathcal{D}_{\mathcal{G}}$ achieves the following maximal information rate*

$$C_{\mathcal{D}_{\mathcal{G}}} = \max_{P_{\mathbf{S}}} \min_{V\in\mathcal{F}} I(S;Y), \tag{C.4}$$

where the maximization is performed over all probability distributions $P_{\mathbf{S}}$ on $\mathcal{S}$, and

$$I(S;Y) = D(V||P_{\mathbf{S}}P_{\mathbf{Y}}), \tag{C.5}$$

where $D(.||.)$ denotes the relative entropy functional [153]. In (C.4)*, the set $\mathcal{F}$ denotes the set of all likelihood functions $V(\boldsymbol{y}|\boldsymbol{s},\Upsilon) = \prod_{k=0}^{M-1}\mathcal{CN}(\upsilon_k s_k,\sigma_k^2)$ on $\mathcal{S}\times\mathcal{Y}$ that satisfy the following constraints*

$$(\mathrm{c}_1):\ \mathbb{E}_{P_{\mathbf{S}}}\big[V(\boldsymbol{y}|\boldsymbol{s},\Upsilon)\big] = \mathbb{E}_{P_{\mathbf{S}}}\big[W(\boldsymbol{y}|\boldsymbol{s},\mathbf{H})\big], \tag{C.6}$$

$$(\mathrm{c}_{2,\mathrm{k}}):\ \mathbb{E}_{P_{\mathbf{S}}}\left[\mathbb{E}_V\big[\mathcal{D}_{\mathcal{G}}(s_k,y_k,\widehat{H}_k)\big]\right] \le \mathbb{E}_{P_{\mathbf{S}}}\left[\mathbb{E}_W\big[\mathcal{D}_{\mathcal{G}}(s_k,y_k,\widehat{H}_k)\big]\right], \tag{C.7}$$

for every $k = 0,\ldots,M-1$, where $\Upsilon = [\upsilon_0,...,\upsilon_{M-1}]^T$.

Case of Improved ML Metric

In the following, we aim at solving the above constrained minimization problem for our specific MB-OFDM channel and metric $\mathcal{D}_{\mathcal{G}} = \mathcal{D}_{\mathcal{M}}$ of (C.3). To this end, we assume that the transmitter has no information about the channel estimate and consequently uniform power allocation is done across subcarriers. Furthermore, we assume a Gaussian i.i.d. input distribution $P_{\mathbf{S}} = \mathcal{CN}(\mathbf{0},\bar{P}\mathbb{I}_M)$. Under these conditions, the mutual information (C.5), written for one subband of the MB-OFDM transmission and averaged over all subcarriers is given by [173]

$$I(S;Y|\Upsilon) = \frac{1}{M}\log_2\det\left\{\mathbb{I}_M + \bar{P}\Upsilon_d\Upsilon_d^{\dagger}\mathbf{\Sigma}^{-1}\right\} = \frac{1}{M}\sum_{k=0}^{M-1}\log_2\left(1+\frac{\bar{P}|\upsilon_k|^2}{\sigma_k^2}\right), \tag{C.8}$$

where $\mathbf{\Sigma} = \mathrm{diag}([\sigma_0^2,...,\sigma_{M-1}^2])$ and $\Upsilon_d = \mathrm{diag}(\Upsilon)$. According to Theorem C.2.1, we have to find the covariance matrix $\mathbf{\Sigma}$ and the optimal channel vector Υ_{opt} (these two characterize the

pdf $V(\boldsymbol{y}|\boldsymbol{s},\Upsilon)$), so as to minimize the mutual information (C.8) and to satisfy constrains (C.6) and (C.7).

The unknown diagonal covariance matrix $\boldsymbol{\Sigma}$ is obtained from the constraint (c_1) of (C.6). It is easily seen that this constraint leads to the equality $V(\boldsymbol{y}|\Upsilon) = W(\boldsymbol{y}|\mathbf{H})$. Moreover, from the likelihoods $V(\boldsymbol{y}|\boldsymbol{s},\Upsilon)$ and $W(\boldsymbol{y}|\boldsymbol{s},\mathbf{H})$, it is clear that the conditional random vectors $(\boldsymbol{y}|\Upsilon)$ and $(\boldsymbol{y}|\mathbf{H})$ are distributed as $\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma}+\bar{P}\Upsilon_d\Upsilon_d^\dagger)$ and $\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{z}}+\bar{P}\mathbf{H}_d\mathbf{H}_d^\dagger)$, respectively. Thus the diagonal covariance matrix $\boldsymbol{\Sigma}$ is obtained as

$$\boldsymbol{\Sigma} = \bar{P}(\mathbf{H}_d\mathbf{H}_d^\dagger - \Upsilon_d\Upsilon_d^\dagger) + \boldsymbol{\Sigma}_{\boldsymbol{z}}, \tag{C.9}$$

with its k-th ($k = 0, ..., M-1$) diagonal element equal to

$$\sigma_k^2 = \bar{P}(|H_k|^2 - |\upsilon_k|^2) + \sigma_z^2. \tag{C.10}$$

Now, in order to find the optimal vector Υ_{opt}, we have to specify the set of inequality constraints ($c_{2,k}$) associated to the metric $\mathcal{D}_{\mathcal{M}}(s_k, y_k, \widehat{H}_k)$ of (C.3). As shown in Section C.3, the k-th inequality constraint set in (C.7) is equivalent to the set

$$\mathcal{V}_{\mathcal{M},k} = \left\{\upsilon_k :\ |\upsilon_k - a\,\widehat{H}_k|^2 \geq |H_k - a\,\widehat{H}_k|^2\right\}, \tag{C.11}$$

where $a = \delta(\,\lambda\sigma_z^2 - \bar{P}\delta\sigma_{\mathcal{E}}^2\,)/(\,\lambda\sigma_z^2 - \bar{P}\delta\sigma_{\mathcal{E}}^2(1-\lambda)\,)$; $\lambda = \exp\left\{\frac{\sigma_z^2}{\bar{P}\delta\sigma_{\mathcal{E}}^2}\right\}E_1\left(\frac{\sigma_z^2}{\bar{P}\delta\sigma_{\mathcal{E}}^2}\right)$ and $E_1(x) \triangleq \int_x^{+\infty}\frac{\exp\{-u\}}{u}\,\mathrm{d}u$ is defined as the exponential integral.

By using (C.8), (C.10) and (C.11) in Theorem C.2.1, the achievable rates $C_{\mathcal{M}}(\mathbf{H},\widehat{\mathbf{H}})$ associated to the improved metric $\mathcal{D}_{\mathcal{M}}$, can be obtained by solving the following minimization, which is equivalent to the initial formulation (C.4):

$$C_{\mathcal{M}}(\mathbf{H},\widehat{\mathbf{H}}) = \begin{cases} \min\limits_{\Upsilon}\ \dfrac{1}{M}\displaystyle\sum_{k=0}^{M-1}\log_2\left(1+\frac{\bar{P}|\upsilon_k|^2}{\sigma_k^2(|\upsilon_k|)}\right) \\ \\ \text{subject to}\quad |\upsilon_k - a\,\widehat{H}_k|^2 \geq |H_k - a\,\widehat{H}_k|^2 \quad \text{for } k = 0, ..., M-1. \end{cases} \tag{C.12}$$

Note that the k-th term in (C.12) is an increasing function of $|\upsilon_k|^2$ and the constraints are convex. Consequently, the minimum in (C.12) is obtained at the extremal of the sets [156]. Thus, the problem simplifies to the search of the optimal channel coefficients $\upsilon_k = \upsilon_{\text{opt},k}$ from the minimization of $\|\Upsilon\|_F^2$ under the constraint sets resulting from the equality in (C.11). This becomes a classical convex minimization problem that can be solved by using Lagrange multipliers. After simple algebra we get $\Upsilon_{\text{opt}} = [\upsilon_{\text{opt},0}, ..., \upsilon_{\text{opt},M-1}]^T$, where

$$\upsilon_{\text{opt},k} = \eta_{\mathcal{M},k}\widehat{H}_k, \quad k = 0, \ldots, M-1, \tag{C.13}$$

and

$$\eta_{\mathcal{M},k} = \begin{cases} a - \dfrac{|H_k - a\,\widehat{H}_k|}{|\widehat{H}_k|}, & \text{if } a \geq 0 \\ \\ a + \dfrac{|H_k - a\,\widehat{H}_k|}{|\widehat{H}_k|}, & \text{if } a < 0. \end{cases} \tag{C.14}$$

Finally, the achievable rates associated to metric $\mathcal{D}_{\mathcal{M}}$ are given as follows:

$$\boxed{C_{\mathcal{M}}(\mathbf{H}, \widehat{\mathbf{H}}) = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P}\, \eta_{\mathcal{M},k}^2 |\widehat{H}_k|^2}{\sigma_z^2 + \bar{P}(|H_k|^2 - \eta_{\mathcal{M},k}^2 |\widehat{H}_k|^2)} \right).} \tag{C.15}$$

Case of Mismatched ML Metric

Under the same assumptions as above, we can compute the achievable rates C_{MM} associated to the mismatched metric $\mathcal{D}_{\text{MM}}$ of (C.2) in the same way. The equality constraint (C.6) leads to the same result as equation (C.10). By following similar steps as in Section C.3, we can easily derive the k-th inequality constraint set associated to the metric D_{MM} as

$$\mathcal{V}_{\text{MM},k} = \left\{ v_k : \ \mathbb{R}e\big(v_k^* \, \widehat{H}_k\big) \geq \mathbb{R}e\big(H_k^* \, \widehat{H}_k\big) \right\} \quad \text{for } \ k = 0, ..., M-1. \tag{C.16}$$

Now, by considering the minimization problem of equation (C.12) under the constraints of (C.16), and using the Lagrange method, we obtain

$$\boxed{C_{\text{MM}}(\mathbf{H}, \widehat{\mathbf{H}}) = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P}\, \eta_{\text{MM},k}^2 |\widehat{H}_k|^2}{\sigma_z^2 + \bar{P}(|H_k|^2 - \eta_{\text{MM},k}^2 |\widehat{H}_k|^2)} \right)} \tag{C.17}$$

where $\eta_{\text{MM},k}$ is given by

$$\eta_{\text{MM},k} = \frac{\mathbb{R}e(H_k^* \, \widehat{H}_k)}{|\widehat{H}_k|^2}. \tag{C.18}$$

In order to get insight about the expressions (C.15) and (C.17), assume that perfect CSIR is available at the receiver, i.e., $H_k = \widehat{H}_k$ for $k = 0, ..., M-1$. In this situation, $\sigma_{\mathcal{E}}^2 = 0$, $\delta = 1$, and we get $\eta_{\text{MM},k} = \eta_{\mathcal{M},k} = 1$ for all k. Consequently, both detectors provide the instantaneous rate

$$C(\mathbf{H}) = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P}\, |H_k|^2}{\sigma_z^2} \right). \tag{C.19}$$

In this case, assuming that there is no delay constraint and the transmission time is longer than the channel coherence time, both of the detectors can achieve the ergodic capacity [174] which is defined as

$$C_{\text{erg}} = \mathbb{E}_{\mathbf{H}}\big[C(\mathbf{H})\big]. \tag{C.20}$$

Finally, note that the derivation of the achievable outage rates is similar to the case of multi-antenna MB-OFDM systems presented in Chapter 4.

C.3 Details on the Derivation of the Inequality Constraint (C.11)

Using the expression of the modified metric $\mathcal{D}_{\mathcal{M}}$ from (C.3), the left-hand side of the k-th constraint in (C.7) can be expanded as follows (the index k is omitted for notational brevity).

$$\begin{aligned}
&\mathbb{E}_{P_S}\left[\mathbb{E}_{V_{Y|S}}\big[\mathcal{D}_{\mathcal{M}}(s,y,\widehat{H})\big]\right] \\
&= \underbrace{\mathbb{E}_{P_S}\big[\log\pi(\sigma_z^2+\delta\sigma_{\mathcal{E}}^2|s|^2)\big]}_{K} + \mathbb{E}_{P_S}\,\mathbb{E}_{V_{Y|S}}\left[\frac{|y|^2-2\mathbb{R}e\big(y^*\delta\widehat{H}s\big)+\delta^2|\widehat{H}|^2|s|^2}{\sigma_z^2+\delta\sigma_{\mathcal{E}}^2|s|^2}\right] \\
&= K + \mathbb{E}_{P_S}\left[\frac{\mathbb{E}_{V_{Y|S}}[|y|^2]-2\mathbb{R}e(\mathbb{E}_{V_{Y|S}}[y^*]\delta\widehat{H}s)+\delta^2|\widehat{H}|^2|s|^2}{\sigma_z^2+\delta\sigma_{\mathcal{E}}^2|s|^2}\right] \\
&= K + \mathbb{E}_{P_S}\left[\frac{\sigma^2+|\upsilon|^2|s|^2-2\delta\mathbb{R}e(\upsilon^*\widehat{H})|s|^2+\delta^2|\widehat{H}|^2|s|^2}{\sigma_z^2+\delta\sigma_{\mathcal{E}}^2|s|^2}\right] \\
&= K + \mathbb{E}_{P_S}\left[\frac{\sigma^2+|\upsilon-\delta\widehat{H}|^2|s|^2}{\sigma_z^2+\delta\sigma_{\mathcal{E}}^2|s|^2}\right]. \qquad \text{(C.21)}
\end{aligned}$$

Similarly, the right-hand side of (C.7) can be obtained as

$$\mathbb{E}_{P_S}\left[\mathbb{E}_{W_{Y|S}}\big[\mathcal{D}_{\mathcal{M}}(s,y,\widehat{H})\big]\right] = K + \mathbb{E}_{P_S}\left[\frac{\sigma_z^2+|H-\delta\widehat{H}|^2|s|^2}{\sigma_z^2+\delta\sigma_{\mathcal{E}}^2|s|^2}\right]. \qquad \text{(C.22)}$$

Using (C.21) and (C.22) in (C.7), the inequality constraint is written as

$$\mathbb{E}_{P_S}\left[\frac{\sigma^2+|\upsilon-\delta\widehat{H}|^2|s|^2}{\sigma_z^2+\delta\sigma_{\mathcal{E}}^2|s|^2}\right] \le \mathbb{E}_{P_S}\left[\frac{\sigma_z^2+|H-\delta\widehat{H}|^2|s|^2}{\sigma_z^2+\delta\sigma_{\mathcal{E}}^2|s|^2}\right]. \qquad \text{(C.23)}$$

Now, in order to evaluate the expectation in (C.23), we introduce the following lemma.

Lemma C.3.1. *Assume $s\sim\mathcal{CN}(0,\beta)$ and $x=|s|^2$ be a centered Chi-squared random variable with two degrees of freedom with pdf $p(x)=\frac{1}{\beta}e^{-x/\beta}$, and let a, b, c and d be real and positive scalars. We have*

$$\mathbb{E}_X\left[\frac{a+b\,x}{c+d\,x}\right] = \frac{b}{d} + \left(\frac{a}{d\beta}-\frac{bc}{d^2\beta}\right)\exp\left\{\frac{c}{d\beta}\right\}E_1\left(\frac{c}{d\beta}\right), \qquad \text{(C.24)}$$

where $E_1(\alpha)\triangleq\int_\alpha^\infty\frac{\exp\{-u\}}{u}\,\mathrm{d}u$ is the exponential integral.

Proof. The proof is easy and thus omitted for brevity. □

By using Lemma C.3.1 and replacing a,b,c and d by their respective values from (C.23), we get

$$\frac{|\upsilon-\delta\widehat{H}|^2}{\delta\sigma_{\mathcal{E}}^2}+\left[\frac{\sigma^2}{(\delta\sigma_{\mathcal{E}}^2)\bar{P}}-\frac{|\upsilon-\delta\hat{h}|^2\sigma_z^2}{(1-\delta^2)\bar{P}}\right]\lambda \le \frac{|H-\delta\widehat{H}|^2}{\delta\sigma_{\mathcal{E}}^2}+\left[\frac{\sigma_z^2}{\delta\sigma_{\mathcal{E}}^2\bar{P}}-\frac{|H-\delta\hat{h}|^2\sigma_z^2}{(1-\delta^2)\bar{P}}\right]\lambda \qquad \text{(C.25)}$$

where $\lambda\triangleq\exp\big\{\frac{\sigma_z^2}{\delta\sigma_{\mathcal{E}}^2\bar{P}}\big\}E_1\big(\frac{\sigma_z^2}{\delta\sigma_{\mathcal{E}}^2\bar{P}}\big)$.
After some algebraic manipulations, (C.11) can be derived from (C.25).

Appendix D

Additional Computations Related to Chapter 5

D.1 Generalization of the Detector Expression to the Case of Arbitrary Space-time Coding

Here we provide more details on the modification of the expressions of the detectors for the general case of arbitrary space-time coding. For this purpose, we briefly recall the general formulation of linear dispersion (LD) codes from [165]. Actually, almost every ST code can be considered as a special case of LD codes.

Remember from subsection 5.3.3 that we denote by $\boldsymbol{s}$ and by $\mathbf{X}$, the $(Q \times 1)$ vector and the $(M_T \times T_u)$ matrix of symbols prior to and after ST coding, respectively, where T_u denotes the number of channel-uses. The ST coding rate is defined as $R_{\text{STC}} = Q/T_u$. For instance, for the V-BLAST scheme we have $Q = 2$, $T_u = 1$, and $R_{\text{STC}} = 2$, whereas for the GLD scheme, considered in subsection 5.6.2, $Q = 4$, $T_u = 2$, and still $R_{\text{STC}} = 2$.

Let us denote by α^q the real and by β^q the imaginary part of the q-th entry of $\boldsymbol{s}$, s^q, i.e., $s^q = \alpha^q + j\beta^q$. An ST scheme is described by its constructing matrices $\mathbf{A}^q$ and $\mathbf{B}^q$, $q = 1, \cdots, Q$, all of dimension $(M_T \times T_u)$ and assumed of real-value entries, such that:

$$\mathbf{X} = \sum_{q=1}^{Q} \left(\alpha^q \mathbf{A}^q + j\beta^q \mathbf{B}^q \right) . \tag{D.1}$$

We separate the real and imaginary parts of the entries of $\boldsymbol{s}$ and $\mathbf{X}$ and stack them row-wise in vectors $\check{\mathbf{s}}$ of dimension $(2Q \times 1)$ and $\check{\mathbf{X}}$ of dimension $(2M_T T_u \times 1)$, respectively. For instance,

$$\check{\mathbf{s}} = \left[\; \alpha^1 \; \beta^1 \; \cdots \; \alpha^Q \; \beta^Q \; \right]^T. \tag{D.2}$$

We obtain hence, $\check{\mathbf{X}} = \check{\mathbf{F}} \; \check{\mathbf{s}}$, where the matrix $\check{\mathbf{F}}$ has the dimension $(2M_T T_u \times 2Q)$ and is

obtained from the matrices $\mathbf{A}^q$ and $\mathbf{B}^q$ (see [165]). Similarly, from $\mathbf{Y}$ which is the received matrix corresponding to $\mathbf{X}$, we construct the vector $\breve{\mathbf{Y}}$ of dimension $(2M_R T_u \times 1)$. Vectors $\breve{\mathbf{X}}$ and $\breve{\mathbf{Y}}$ are related through a matrix $\breve{\mathbf{H}}$ of dimension $(2M_R T_u \times 2M_T T_u)$:

$$\breve{\mathbf{Y}} = \breve{\mathbf{H}}\ \breve{\mathbf{X}} + \breve{\mathbf{Z}} \tag{D.3}$$

where $\breve{\mathbf{Z}}$ is the vector of real AWGN of zero mean and variance $\sigma_z^2/2$. Matrix $\breve{\mathbf{H}}$ is composed of segments $\breve{\mathbf{H}}_{ij}$, $i = 1, \cdots, M_R$, $j = 1, \cdots, M_T$, that are block diagonal matrices of dimension $(2T_u \times 2T_u)$ with equal diagonal blocks $\mathbf{H}_{ij}$. Submatrices $\mathbf{H}_{ij}$ are obtained from each entry H_{ij} of the initial matrix $\mathbf{H}$ as shown below (see [175]).

$$\mathbf{H}_{ij} = \begin{bmatrix} \mathbb{R}e(H_{ij}) & -\mathbb{I}m(H_{ij}) \\ \mathbb{I}m(H_{ij}) & \mathbb{R}e(H_{ij}) \end{bmatrix}. \tag{D.4}$$

Now, we can write the "ST code + channel" input/output relationship by considering an equivalent channel matrix $\breve{\mathbf{H}}_{eq}$ of dimension $(2M_R T_u \times 2Q)$:

$$\breve{\mathbf{Y}} = \breve{\mathbf{H}}\ \breve{\mathbf{F}}\ \breve{\mathbf{s}} + \breve{\mathbf{Z}} = \breve{\mathbf{H}}_{eq}\ \breve{\mathbf{s}} + \breve{\mathbf{Z}}. \tag{D.5}$$

Now in the expression of the detector, we have just to consider $\breve{\mathbf{H}}_{eq}$, $\breve{\mathbf{s}}$, and $\breve{\mathbf{Y}}$, instead of $\mathbf{H}$, $\boldsymbol{s}$, and $\boldsymbol{y}$, respectively.

D.2 Derivation of the Improved MMSE Filter (5.20)

The inner expectations involved in (5.19) can be easily evaluated from (5.17) as

$$\mathbf{R}_{s_k^i \underline{\boldsymbol{y}}_k^i} = \mathbb{E}_{\boldsymbol{s}_k, \boldsymbol{z}_k}\left[s_k^i\, \underline{\boldsymbol{y}}_k^{i\,\dagger}\right] = \sigma_{s_k^i}^2\, \mathbf{h}_i^\dagger + \mathbf{m}_{k,i}\big(\underline{\mathbf{H}}_i - \underline{\widehat{\mathbf{H}}}_i\big)^\dagger \tag{D.6}$$

and

$$\begin{aligned} \mathbf{R}_{\underline{\boldsymbol{y}}_k^i} = \mathbb{E}_{\boldsymbol{s}_k, \boldsymbol{z}_k}\left[\underline{\boldsymbol{y}}_k^i\, \underline{\boldsymbol{y}}_k^{i\,\dagger}\right] &= \sigma_{s_k^i}^2\, \mathbf{h}_i \mathbf{h}_i^\dagger + \underline{\mathbf{H}}_i\, \boldsymbol{\Lambda}_{k,i}\, \underline{\mathbf{H}}_i^\dagger + \mathbf{h}_i\, \mathbf{m}_{k,i}\, \underline{\mathbf{H}}_i^\dagger + \underline{\mathbf{H}}_i\, \mathbf{m}_{k,i}^\dagger\, \mathbf{h}_i^\dagger \\ &\quad - \mathbf{h}_i\, \mathbf{m}_{k,i}\, \underline{\widehat{\mathbf{H}}}_i^\dagger - \underline{\widehat{\mathbf{H}}}_i\, \mathbf{m}_{k,i}^\dagger\, \mathbf{h}_i^\dagger - \underline{\mathbf{H}}_i\, \widetilde{\boldsymbol{\Lambda}}_{k,i}\, \underline{\widehat{\mathbf{H}}}_i^\dagger - \underline{\widehat{\mathbf{H}}}_i\, \widetilde{\boldsymbol{\Lambda}}_{k,i}\, \underline{\mathbf{H}}_i^\dagger \\ &\quad + \underline{\widehat{\mathbf{H}}}_i\, \widetilde{\boldsymbol{\Lambda}}_{k,i}\, \underline{\widehat{\mathbf{H}}}_i^\dagger + \sigma_z^2 \mathbb{I}_{M_R}. \end{aligned} \tag{D.7}$$

We have thus to evaluate the outer expectations

$$\overline{\mathbf{R}}_{s_k^i \underline{\boldsymbol{y}}_k^i} = \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left[\mathbf{R}_{s_k^i \underline{\boldsymbol{y}}_k^i}\right]$$

and

$$\overline{\mathbf{R}}_{\underline{\boldsymbol{y}}_k^i} = \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left[\mathbf{R}_{\underline{\boldsymbol{y}}_k^i}\right].$$

In order to compute the above expectations, we use the following lemma [176].

Lemma D.2.1. *For a circularly symmetric complex Gaussian random row-wise vector* $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and a Hermitian matrix* $\mathbf{A}$*, we have*

$$\mathbb{E}_{\mathbf{x}}\left[\mathbf{x}\,\mathbf{A}\,\mathbf{x}^{\dagger}\right] = \mathrm{tr}\big(\mathbf{A}\boldsymbol{\Sigma}\big) + \boldsymbol{\mu}\,\mathbf{A}\,\boldsymbol{\mu}^{\dagger}. \tag{D.8}$$

By applying Lemma D.2.1 and using the a posteriori channel pdf (5.4), it is straightforward to find $\overline{\mathbf{R}}_{s_k^i \underline{\boldsymbol{y}}_k^i}$ in (5.21) as

$$\overline{\mathbf{R}}_{s_k^i \underline{\boldsymbol{y}}_k^i} = \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left[\mathbf{R}_{s_k^i \underline{\boldsymbol{y}}_k^i}\right] = \delta\,\sigma^2_{s_k^i}\,\widehat{\mathbf{h}}_i^{\dagger} + (\delta-1)\,\mathbf{m}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger}. \tag{D.9}$$

The evaluation of $\overline{\mathbf{R}}_{\underline{\boldsymbol{y}}_k^i}$ in (5.22) involves the following equalities.

$$\begin{aligned}
&\mathbb{E}_{\mathbf{h}_i|\widehat{\mathbf{h}}_i}\left[\mathbf{h}_i\,\mathbf{h}_i^{\dagger}\right] = \delta^2\widehat{\mathbf{h}}_i\,\widehat{\mathbf{h}}_i^{\dagger} + (1-\delta)\mathbb{I}_{M_R},\\
&\mathbb{E}_{\mathbf{h}_i|\widehat{\mathbf{h}}_i}\left[\mathbf{h}_i\,\mathbf{m}_{k,i}\,\underline{\mathbf{H}}_i^{\dagger}\right] = \delta^2\,\widehat{\mathbf{h}}_i\,\mathbf{m}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger},\\
&\mathbb{E}_{\underline{\mathbf{H}}_i|\underline{\widehat{\mathbf{H}}}_i}\left[\underline{\mathbf{H}}_i\,\widetilde{\boldsymbol{\Lambda}}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger}\right] = \delta\,\underline{\widehat{\mathbf{H}}}_i\,\widetilde{\boldsymbol{\Lambda}}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger}\\
&\mathbb{E}_{\underline{\mathbf{H}}_i|\underline{\widehat{\mathbf{H}}}_i}\left[\underline{\mathbf{H}}_i\,\boldsymbol{\Lambda}_{k,i}\,\underline{\mathbf{H}}_i^{\dagger}\right] = \delta^2\underline{\widehat{\mathbf{H}}}_i\,\boldsymbol{\Lambda}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger} + (1-\delta)\,\mathrm{tr}(\boldsymbol{\Lambda}_{k,i})\mathbb{I}_{M_R}.
\end{aligned}$$

Now, by using the above equalities we obtain

$$\begin{aligned}
\overline{\mathbf{R}}_{\underline{\boldsymbol{y}}_k^i} = {}& \delta^2\,\sigma^2_{s_k^i}\,\widehat{\mathbf{h}}_i\,\widehat{\mathbf{h}}_i^{\dagger} + (1-\delta)\,\sigma^2_{s_k^i}\,\mathbb{I}_{M_R} + \sigma_z^2\mathbb{I}_{M_R} + \delta^2\underline{\widehat{\mathbf{H}}}_i\,\boldsymbol{\Lambda}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger}\\
&+ (1-\delta)\,\mathrm{tr}(\boldsymbol{\Lambda}_{k,i})\mathbb{I}_{M_R} + \delta^2\,\widehat{\mathbf{h}}_i\,\mathbf{m}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger} + \delta^2\,\underline{\widehat{\mathbf{H}}}_i\,\mathbf{m}_{k,i}^{\dagger}\,\widehat{\mathbf{h}}_i^{\dagger}\\
&- \delta\,\widehat{\mathbf{h}}_i\,\mathbf{m}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger} - \delta\,\underline{\widehat{\mathbf{H}}}_i\,\mathbf{m}_{k,i}^{\dagger}\,\widehat{\mathbf{h}}_i^{\dagger} - \delta\,\underline{\widehat{\mathbf{H}}}_i\,\widetilde{\boldsymbol{\Lambda}}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger}\\
&- \delta\,\underline{\widehat{\mathbf{H}}}_i\,\widetilde{\boldsymbol{\Lambda}}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger} + \underline{\widehat{\mathbf{H}}}_i\,\widetilde{\boldsymbol{\Lambda}}_{k,i}\,\underline{\widehat{\mathbf{H}}}_i^{\dagger}.
\end{aligned} \tag{D.10}$$

Equation (5.22) follows directly after rearranging the terms in (D.10). As a result, from (D.9) and (D.10), we obtain the improved MMSE filter (5.20).

D.3 Derivation of the Variance $\sigma^2_{\eta_{k,i}}$ in (5.26)

From (5.9), we can evaluate the mean-squared error (MSE) at the output of the turbo-PIC detector as

$$\begin{aligned}
\sigma^2_{\mathrm{MSE}} &= \mathbb{E}_{\boldsymbol{s}_k,\boldsymbol{z}_k}\left[\left|s_k^i - \mathbf{w}_k^i\,\underline{\boldsymbol{y}}_k^i\right|^2\right]\\
&= \sigma^2_{s_k^i} - \mathbf{R}_{s_k^i\,\underline{\boldsymbol{y}}_k^i}\,{\mathbf{w}_k^i}^{\dagger} - \mathbf{w}_k^i\,\mathbf{R}_{\underline{\boldsymbol{y}}_k^i s_k^i} + \mathbf{w}_k^i\,\mathbf{R}_{\underline{\boldsymbol{y}}_k^i}\,{\mathbf{w}_k^i}^{\dagger}.
\end{aligned} \tag{D.11}$$

Since $\mathbf{w}_k^i = \mathbf{R}_{s_k^i\,\underline{\boldsymbol{y}}_k^i}\mathbf{R}^{-1}_{\underline{\boldsymbol{y}}_k^i}$, we have $\mathbf{R}_{s_k^i\,\underline{\boldsymbol{y}}_k^i}\,{\mathbf{w}_k^i}^{\dagger} = \mathbf{w}_k^i\,\mathbf{R}_{\underline{\boldsymbol{y}}_k^i}\,{\mathbf{w}_k^i}^{\dagger}$. Consequently, (D.11) reduces to

$$\sigma^2_{\mathrm{MSE}} = \sigma^2_{s_k^i} - \mathbf{w}_k^i\,\mathbf{R}_{\underline{\boldsymbol{y}}_k^i s_k^i}. \tag{D.12}$$

For the case of the improved detector, after using $\widetilde{\mathbf{w}}_k^i$ and $\overline{\mathbf{R}}_{s_k^i \underline{\boldsymbol{y}}_k^i}$ from (5.20) and (5.21) instead of $\mathbf{w}_k^i$ and $\mathbf{R}_{s_k^i \underline{\boldsymbol{y}}_k^i}$ in (D.12), we obtain

$$\sigma^2_{\text{MSE-IM}} = \sigma^2_{s_k^i}\,(1-\mu_{k,i}) - (\delta-1)\widetilde{\mathbf{w}}_k^i\,\widehat{\underline{\mathbf{H}}}_i\,\mathbf{m}_{k,i}^{\dagger} \tag{D.13}$$

where $\mu_{k,i} = \delta\,\widetilde{\mathbf{w}}_k^i\,\widehat{\mathbf{h}}_i$.

Alternatively, from (5.25), we have

$$\begin{aligned}\sigma^2_{\text{MSE-IM}} &= \mathbb{E}_{\boldsymbol{s}_k,\boldsymbol{z}_k}\left[\left|s_k^i - \left(\mu_{k,i}\,s_k^i + \eta_{k,i}\right)\right|^2\right] \\ &= (1-\mu_{k,i})^2\,\sigma^2_{s_k^i} + \sigma^2_{\eta_k^i} - 2\,\mathbb{R}e\left((\delta-1)(1-\mu^*_{k,i})\widetilde{\mathbf{w}}_k^i\,\widehat{\underline{\mathbf{H}}}_i\,\mathbf{m}_{k,i}^{\dagger}\right).\end{aligned} \tag{D.14}$$

Comparing (D.13) and (D.14), leads to

$$\boxed{\sigma^2_{\eta_k^i} = (\mu_{k,i} - \mu^2_{k,i})\,\sigma^2_{s_k^i} + (\delta-1)\left[2\,\mathbb{R}e\left((1-\mu^*_{k,i})\widetilde{\mathbf{w}}_k^i\,\widehat{\underline{\mathbf{H}}}_i\,\mathbf{m}_{k,i}^{\dagger}\right) - \widetilde{\mathbf{w}}_k^i\,\widehat{\underline{\mathbf{H}}}_i\,\mathbf{m}_{k,i}^{\dagger}\right].} \tag{D.15}$$

Notice that by setting δ equal to one (corresponding to perfect CSIR) in (D.15), we retrieve the classical expression derived in the literature, i.e., $\sigma^2_{\eta_k^i} = (\mu_{k,i} - \mu^2_{k,i})\,\sigma^2_{s_k^i}$ (see [17, 19], for instance).

References

[1] M. Terré. Communication en ultra large bande (UWB). *Séminaire CNFRS/URSI*, 2003.

[2] M. G. Di Benedetto, T. Kaiser, A. F. Molish, I. Opperman, C. Politano, and D. Porcino. *UWB Communication Systems A Comprehensive Overview*. Series on Signal Processing and Communications. EURASIP, 2006.

[3] A. Batra, J. Balakrishnan, G. R. Aiello, J. R. Foerster, and A. Dabak. Design of multiband OFDM system for realistic UWB channel environments. *IEEE Trans. Microwave Theory and Techniques*, 52:2123–2138, Sep. 2004.

[4] A. Batra, J. Balakrishnan, and A. Dabak. Multiband OFDM physical layer proposal for IEEE 802.15 task group 3a. July 2003.

[5] X. Zhuang and F. W. Vook. Iterative channel estimation and decoding for a turbo-coded OFDM system via the EM algorithm. *in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2002.

[6] E. Jaffrot and M. Siala. Turbo channel estimation for OFDM systems on highly time and frequency selective channels. *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Jun. 2000.

[7] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley Series in probability and Statistics, Wiley, 1997.

[8] X. Ma, H. Kobayashi, and S. C. Schwartz. EM-based channel estimation algorithms for OFDM. *EURASIP J. Applied Signal Process.*, 10:1460–1477, 2004.

[9] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj. Semi-blind channel identification and equalization in OFDM: An expectation-maximization approach. *in Proc. IEEE Veh. Technol. Conf.*, pages 13–17, 2002.

[10] B. Lu, X. Wang, and Y. G. Li. Iterative receivers for space-time block coded OFDM systems in dispersive fading channels. *IEEE Trans. Wireless Commun.*, 1:213–225, Apr. 2002.

[11] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. Inform. Theory*, pages 284–287, Mar. 1974.

[12] W. K. Ma, T. N. Davidson, K. M. Wong, Z. Q. Luo, and P. Ching. Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA. *IEEE Trans. Signal. Process.*, 50(4):912–922, Apr. 2002.

[13] J. Jaldén and B. Ottersten. Parallel implementation of a soft output sphere decoder. *in Proc. Asilomar conf. on signals, systems and computers*, Oct. 2005.

[14] R. Wang and G. B. Giannakis. Approaching MIMO channel capacity with soft detection based on hard sphere decoding. *IEEE Trans. Wireless Commun.*, 54(4):587–590, Apr. 2006.

[15] S. Baro, J. Hagenauer, and M. Witzke. Iterative detection of MIMO transmission using a list-sequential LISS detector. *in Proc. Int. Commun. Conf. (ICC)*, 4:2653–2657, May 2003.

[16] Y. L. C. de Jong and T. J. Willink. Iterative tree search detection for MIMO wireless systems. *IEEE Trans. Commun.*, 53(6):930–935, June 2005.

[17] X. Wang and H. V. Poor. Iterative (turbo) soft interference cancellation and deoding for coded CDMA. *IEEE Trans. Commun.*, 47(7):1046–1061, July 1999.

[18] M. Tuchler, A. C. Singer, and R. Koetter. Minimum mean squared error equalization using a priori information. *IEEE Trans. Signal Process.*, 50(3):673–683, Mar. 2002.

[19] A. Dejonghe and L. Vandendorpe. Turbo-equalization for multilevel modulation:an efficient low-complexity scheme. *in Proc. Int. Commun. Conf. (ICC)*, pages 1863–1867, 2002.

[20] X. Wautelet, A. Dejonghe, and L. Vandendorpe. MMSE-based fractional turbo receiver for space-time BICM over frequency selective MIMO fading channels. *IEEE Trans. Signal Process.*, 52(6):1804–1809, June 2004.

[21] M. Sellathurai and S. Haykin. Turbo-BLAST for wireless communications: Theory and experiments. *IEEE Trans. Signal Process.*, 50(10):2538–2545, Oct. 2002.

[22] G.R. Aiello and G.D. Rogerson. Ultra-wideband wireless systems. *IEEE Microwave Mag.*, June 2003.

[23] M. Z. Win and R. A. Scholtz. Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications. *IEEE Trans. Commun.*, pages 679–691, April 2000.

[24] C. Limbodal. A spatial RAKE receiver for real-time UWB-IR applications. Master's thesis, University of Oslo, July 2005.

[25] L. Yang and G. B. Giannakis. Ultra-Wideband communications an idea whose time has come. *IEEE Signal Process. Mag.*, pages 26–54, Nov. 2004.

[26] WiMedia Alliance. *http://www.wimedia.org*.

[27] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, Oct. 1948.

[28] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error-correcting coding and decoding: Turbo-codes. In *Int. Conf. Commun. (ICC)*, pages 1064–1070, May 1993.

[29] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space-time block codes from orthogonal designs. *IEEE Trans. Inform. Theory*, 45(5):1456–1467, Jul. 1999.

[30] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. Technical report, Bell Labs Tech. J., Autumn 1996.

[31] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky. Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture. *Electronics Lett.*, 35(1):14–16, Jan. 1999.

[32] R. Prasad. *OFDM for Wireless Communications Systems*. Artech House Publishers, Sep. 2004.

[33] H. Bolcskei, D. Gesbert, and A. J. Paulraj. On the capacity of OFDM-based spatial multiplexing systems. *IEEE Trans. Commun.*, pages 225–234, Feb. 2002.

[34] V. Tarokh and H. Jafarkhani. A differential detection scheme for transmit diversity. *IEEE Trans. Select. Areas Commun.*, 18:1169–1174, Jul. 2000.

[35] B. M. Hochwald and W. Sweldens. Differential unitary space-time modulation. *IEEE Trans. Commun.*, 48:2041–2052, Dec. 2000.

[36] B. L. Hughes. Differential space-time modulations. *IEEE Trans. Inform. Theory*, 46:2567–2578, Nov. 2000.

[37] P. Garg, R. K. Mallik, and H. M. Gupta. Performance analysis of space-time coding with imperfect channel estimation. *IEEE Trans. Wireless Commun.*, 4:257–265, Jan. 2005.

[38] G. Taricco and E. Biglieri. Space-time decoding with imperfect channel estimation. *IEEE Trans. Wireless Commun.*, 4(4):1874–1888, July 2005.

[39] J. K. Cavers and P. Ho. Analysis of the error performance of trellis-Coded modulations in Rayleigh-Fading channels. *IEEE Trans. Commun.*, 40(1):74–83, Jan. 1992.

[40] M. Medard. The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel. *IEEE Trans. Inform. Theory*, 46:933–946, May 2000.

[41] T. Yoo and A. Goldsmith. Capacity of fading MIMO channels with channel estimation error. *in Proc. Int. Commun. Conf. (ICC)*, Jun. 2004.

[42] B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Trans. Inform. Theory*, 49(4):951–963, Apr. 2003.

[43] A. M. Tonello. Space-time bit-interleaved coded modulation with an iterative decoding strategy. *in Proc. Veh. Technol. Conf.*, pages 473–478, Sep. 2000.

[44] G. Caire, G. Taricco, and E. Biglieri. Bit-interleaved coded modulation. *IEEE Trans. Inform. Theory*, 44:927–945, May 1998.

[45] S. Sadough and E. Jaffrot. A wavelet packet based model for an ultra-wideband propagation channel. *in Proc. European Conf. Propagation and Systems ECPS*, March 15-18. 2005.

[46] T. W. Barrett. History of ultraWideband UWB radar & communications: Pioneers and innovators. *in Proc. Progress in Electromagnetics Symposium*, July 2000.

[47] FCC. First report and order, revision of part 15 of the commission's rules regarding ultra-wideband transmission systems. Technical report, Feb. 2004.

[48] J. D. Taylor. *Introduction to Ultra-wideband Radar Systems*. CRC Press, 1995.

[49] M. Ghavami, L. B. Michael, and R. Kohno. *Ultra Wideband Signals and Systems in Communication Engineering*. John Wiley & Sons, 2004.

[50] D. Porcino and W. Hirt. Ultra-wideband radio technology: potential and challenges ahead. *IEEE Commun. Mag.*, 41(7):66–74, July 2003.

[51] I. Opperman, M. Hamamainen, and J. Iinatti. *UWB Theory and Applications*. John Wiley, 2004.

[52] M. Z. Win and R. A. Scholtz. Impulse radio: How it works. *IEEE Commun. Letters*, 2(2):36–38, Feb. 1998.

[53] R. A. Scholtz. Multiple access with time-hopping impulse modulation. In *Military Communications Conf.*, pages 447–450, Oct. 1993.

[54] S. Roy, J. R. Foerster, V. S. Somayazulu, and D. G. Leeper. Ultrawideband radio design: The promise of high-speed, short range wireless conectivity. *Proceedings of the IEEE*, 92, Feb. 2004.

[55] B. Allen, M. Ghavami, A. Armogida, and H. Aghvami. UWB–A ubiquitous wire replacement technology ?

[56] J. T. Conroy, J. L. LoCicero, and D.R. Ucci. Communication techniques using monopulse waveforms. *in Proc. Military Commun. Conf.*, Oct. 1999.

[57] W. Zhuang, X. Shen, and Q. Bi. Ultra-wideband wireless communications. *Wireless Commun. and Mobile Computing*, 3(6):663–685, 2003.

[58] P. Withington, R. Reinhardt, and R. Stanley. Preliminary results of an ultra-wideband (impulse) scanning receiver. *in. Proc. Military Commun. Conf.*, 2, Oct. 1999.

[59] B. R. Vojcic and R. L. Picholtz. Direct-sequence code division multiple access for ultra-wide bandwidth impulse radio. *in Proc. Military Commun. Conf.*, Oct. 2003.

[60] J. G. Proakis. *Digital Communications*. McGraw-Hill, third edition, 1995.

[61] J. M. Wozencraft and I. M. Jacobs. *Principles of Communication Engineering*. Willey, 1965.

[62] G. R. Aiello and G. D Rogerson. Ultra-wideband wireless systems. *IEEE Microwave Mag.*, 4(2):36–47, 2003.

[63] G. D. Forney. The Viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, March 1973.

[64] B. Vucetic and J. Yuan. *Turbo Codes Principle and Applications*. Kluwer Academic Publishers, 2000.

[65] J. Foerster. Channel modeling sub-committee report final. Technical report, IEEE802.15-02/490, 2003.

[66] A. Saleh and A. Valenzuela. A statistical model for indoor multipath propagation. *IEEE J. Selected Areas in Commun.*, pages 128–137, February 1987.

[67] I. M. Johnstone and B. W. Silverman. Neddles and straw in haystacks: Empirical bayes estimates of possibly sparse sequences. *Annals of Statistics*, 32:1594–1649, 2004.

[68] European Telecommunications Standards Institute ETSI. Radio broadcasting systems, digital audio broadcating (DAB) to mobile, portable and fixed receivers. *prETS 300 744*, Feb. 1995.

[69] V. Engels and H. Rohling. Multilevel differential modulation techniques 64-(dapsk) for multicarrier transmission systems. *European Trans. Telecommun. Related Technol.*, pages 633–640, Nov. 1995.

[70] European Telecommunications Standards Institute ETSI. Digital broadcating for television, sound and data services. *prETS 300 744*, Apr. 1996.

[71] M. L. Moher and J. H. Lodge. TCMP a modulation and coding strategy for Rician fading channels. *IEEE Trans. Select. Areas Commun.*, 7(9):1347–1355, Dec. 1989.

[72] J. K. Cavers. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *IEEE Trans. Veh. Technol.*, 40:686–693, Nov. 1991.

[73] J. Rinne and M. Renfors. Pilot spacing in orthogonal frequency division multiplexing systems on practical channels. *IEEE Trans. Consum. Electr.*, 42(4):959–962, Nov. 1996.

[74] M. X. Chang and Y. T. Su. 2D regression channel estimation for equalizing OFDM signals. *in Proc. Vehicular Technol. Conf*, pages 240–244, 2000.

[75] A. Huang and Y. Zhao. Estimating channel response from pilot subcarrier pairs for OFDM systems. *in Proc. Midwest Symp. Circuits and Systems*, Aug. 1997.

[76] J. K. Moon and S. I. Choi. Performance of channel estimation methods for OFDM systems in a multipath fading channels. *IEEE Trans. Consum. Elect.*, 46(1):161–170, Feb. 2000.

[77] S. Coleri, M. Ergen, A. Puri, and A. Bahai. Channel estimation techniques based on pilot arrangment in OFDM systems. *IEEE Trans. Broadcasting*, 48(3):223–229, Sep. 2002.

[78] T. Onizawa, M. Mizoguchi, T. Sakata, and M. Morikura. A simple adaptive channel estimation scheme for OFDM systems. *in Proc. Vehicular Technol. Conf.*, pages 279–283, 1999.

[79] P. Hoeher, S. Kaiser, and P. Robertson. Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering. *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pages 1845–1848, Apr. 1997.

[80] P. Hoeher. TCM on frequency-selective land-mobile fading channels. *in Proc. Tirrenia Int. Workshop Digit. Commun;*, Sep. 1991.

[81] G. L. Stuber and M. Russel. Terrestrial digital video broadcasting for mobile reception using OFDM. *in Proc. IEEE Global Telecommun. Conf. (Globecom)*, pages 2049–2053, Nov. 1995.

[82] J. J. van de Beek, O. Edfors, and M. Sandell. On channel estimation in OFDM systems. *in Proc. Vehicular Technol. Conf.*, pages 815–819, Sep. 1995.

[83] O. Edfors, M. Sandell, J. J. Van de Beek, S. K. Wilson, and P. O. Borjesson. OFDM channel estimation by singular value decomposition. *IEEE Trans. Commun.*, 46(7):931–939, 1998.

[84] L. Scharf. *Statistical Signal Processing.* Addison-Wesley, 1991.

[85] M. H. Hsieh and C. H. Wei. Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels. *IEEE Trans. Consumer Electronics*, 44(1):217–225, Feb. 1998.

[86] Y. Cimini Li and N. R. Sollenberger. Robust channel estimation for ofdm systems with rapid dispersive fading channels. *IEEE Trans. Commun.*, 46(7):902–915, Jul. 1998.

[87] F. Sanzi and J. Speidel. An adaptive two-dimensional channel estimator for wireless OFDM with application to mobile DVB-T. *IEEE Trans. Broadcast.*, 46(2):128–133, Jun. 2000.

[88] N. Seshadri, Y. Li, and S. Ariyavisitakul. Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels. *IEEE J. Select. Areas Commun.*, 17(3):461–471, Mar. 1999.

[89] F. Tufvesson and T. Maseng. Pilot assisted channel estimation for OFDM in mobile cellular systems. *in Proc. Vehicular Technol. Conf.*, pages 1639 – 1643, May 1997.

[90] P. K. Frenger and N. A. B. Svensson. A decision directed coherent detector for OFDM. *in Proc. Vehicular Technol. Conf.*, pages 1584–1588, Apr. 1996.

[91] P. K. Frenger and N. A. B. Svensson. Decision-directed coherent detection in multicarrier systems on Rayleigh fading channels. *IEEE Trans. Veh. Technol.*, 48(2):490–498, Mar. 1999.

[92] V. Mignone and A. Morello. CD3-OFDM : A novel demodulation scheme for fixed and mobile receivers. *IEEE Trans. Commun.*, 44:1144–1151, Sep. 1996.

[93] S. B. Bulumulla, S. A. Kassam, and S. S. Venkatesh. A systematic approach to detecting OFDM signals in fading channels. *IEEE Trans. Commun.*, 48(5):725–728, May 2000.

[94] J. Akhtman and L. Hanzo. Decision directed channel estimation aided OFDM employing. *IEEE Trans. Wireless Commun.*, 6(4):1171–1175, Apr. 2007.

[95] H. Liu, G. Xu, L. Tong, and T. Kailath. Recent developments in blind channel equalization: From cyclostationnarity to subspaces. *IEEE Trans. Signal Process.*, 50(1-2):83–99, Apr. 1996.

[96] R. W. Heath and G. B. Giannakis. Exploiting input cyclostationarity for blind channel identification in OFDM systems. *IEEE Trans. Signal Process.*, 47:848–856, Mar. 1999.

[97] B. Muquet and M. de Courville. Blind and semi-blind channel identification methods using second order statistics for OFDM systems. *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 5:2745–2748, 1999.

[98] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue. Subspace method for the blind identification of multichannel FIR filters. *IEEE Trans. Signal Process.*, 43:516–525, Feb. 1995.

[99] B. Muquet, M. de Courville, and P. Duhamel. Subspace-based blind and semi-blind channel estimation for OFDM systems. *IEEE Trans. Signal Process.*, 50(7):1699–1711, July 2002.

[100] X. Cai and A. N. Akansu. A subspace method for blind channel identification in OFDM systems. *in Proc. Int. Commun. Conf. (ICC)*, pages 929–933, 2000.

[101] H. Bolcskei, P. Duhamel, and R. Hleiss. A subspace-based approach to blind channel identification in pulse shaping OFDM/OQAM systems. *IEEE Trans. Signal Process.*, 49(7):1594–1598, July 2001.

[102] S. Zhou, B. Muquet, and G. B. Giannakis. Subspace-based (semi-) blind channel estimation for block precoded space-time OFDM. *IEEE Trans. Signal. Process.*, 50(5):1215, May 2002.

[103] Z. Wang and G. B. Giannakis. Linear precoded or coded OFDM against wireless channel fades ? *in Proc. Signal Process. Advances Wireless Commun. (SPAWC)*, pages 267–270, Mar. 2001.

[104] R. Lin and a. P. Petropulu. Linear precoding assisted blind channel estimation for OFDM systems. *IEEE Trans. Veh. Technol.*, 54(3):983–995, May 2005.

[105] C. Li and S. Roy. Subspace-based blind channel estimation for OFDM by exploiting virtual carriers. *IEEE Trans. Wireless Commun.*, 2(1):141–149, Jan. 2003.

[106] N. Chotikakamthorn and H. B. Suzuki. On identifiability of OFDM blind channel estimation. *in Proc. Vehicular Technol. Conf.*, pages 2358–2361, Sep. 1999.

[107] S. Zhou and G. B. Giannakis. Finite-alphabet based channel estimation for OFDM and related multicarrier systems. *IEEE Trans. Commun.*, 49:1402–1414, Aug. 2001.

[108] M. C. Necker and G. L. Stuber. Totally blind channel estimation for OFDM on fast varying mobile radio channels. *IEEE Trans. Wireless. Commun.*, 3(5):1514–1524, Sep. 2004.

[109] E. de Carvalho and D. Slock. Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation. *in Proc. Signal Process. Advances Wireless Commun. (SPAWC)*, 1997.

[110] M. Muck, M. de Courville, and P. Duhamel. A pseudorandom postfix OFDM modulator semi-blind channel estimation and equalization. *IEEE Trans. Signal. Process.*, 54(3):1005–1017, Mar. 2006.

[111] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Royal Statistical Soc.*, 39:1–38, 1977.

[112] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux, and L. Vandendorpe. Comparison of EM-Based algorithms for MIMO channel estimation. *IEEE Trans. Commun.*, 55(1):216–226, Jan. 2007.

[113] X. Wautelet, C. Herzet, A. Dejonghe, and L. Vandendorpe. MMSE-based and EM iterative channel estimation methods. *in Proc IEEE 10th Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, Nov. 2003.

[114] W. Ma, H. Kobayashi, and S. C. Schwartz. EM-based channel estimation algorithms for OFDM. *EURASIP J. Applied Signal Process.*, 10:1460–1477, 2004.

[115] Y. Xie and C. N. Georghiades. Two EM-type channel estimation algorithms for OFDM with transmitter diversity. *IEEE Trans. Commun.*, 51(1):106–115, Jan. 2003.

[116] W. Zou and Y. Wu. COFDM: an overview. *IEEE Trans. Broadcast.*, 41:1–8, Mar. 1995.

[117] L. Mazet, V. Buzenac, M. Decourville, and P. Duhamel. An EM based semi-blind channel estimation algorithm designed for OFDM systems. *in Proc. Asilomar Conf.*, Nov. 2002.

[118] S. Touati, J. M. Mamfoumbi Ocloo, P. Duhamel, and F. Alberge. Semi-blind channel estimation for OFDM systems via an EM-block algorithm. *in Proc. EUSIPCO*, September 2004.

[119] S. ten Brink, F. Sandi, and J. Speidel. Two-dimensional iterative APP channel estimation and decoding for OFDM systems. *in Proc. IEEE Global Telecommun. Conf. (Globecom)*, pages 741–745, 2000.

[120] S. Y. Park, Y. G. Kim, C. G. Kang, and D. E. Kang. Iterative receiver with joint detection and channel estimation for OFDM system with multiple receiver antennas in mobile radio environment. *in Proc. IEEE Global Telecommun. Conf. (Globecom)*, Nov. 2001.

[121] G. A. Al-Rawi, T. Y. Al-Naffouri, A. Bahai, and J. Cioffi. Exploiting error-control coding and cyclic-prefix in channel estimation for coded OFDM systems. *IEEE Commun. Lett.*, 7:388–390, Jul. 2003.

[122] X. Zhuang and F. W. Vook. On reduced-complexity iterative channel tracking in turbo coded OFDM. *in Proc. IEEE Global Telecommun. Conf. Globecom*, pages 1064–1068, Dec. 2003.

[123] X. Ma, H. Kobayashi, and S. C. Schwartz. An enhanced channel estimation algorithm for OFDM: combined EM algorithm and polynomial fitting. *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 4:680–683, Apr. 2003.

[124] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj. Semi-blind channel identification and equalization in OFDM: An expectation-maximization approach. *in Proc. Vehicular Technol. Conf.*, pages 13–17, 2002.

[125] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj. An EM-based OFDM receiver for time-variant channels. *in Proc. IEEE Global Telecommun. Conf. (Globecom)*, pages 589–593, 2002.

[126] L. Mazet, V. Buzenac-Settineri, M. de Courville, and P. Duhamel. EM-based semi-blind estimation of time-varying channels. *in Proc. Signal Process. Advances Wireless Commun. (SPAWC)*, 2003.

[127] H. A. Cirpan, E. Panayirci, and H. Dogan. Non-data-aided channel estimation for OFDM systems with space-frequency transmit diversity. *IEEE Trans. Veh. Technol.*, 55(2):449–457, Mar. 2006.

[128] T. K. Moon. The expectation-maximization algorithm. *IEEE Signal Process. Mag.*, pages 47–59, Nov. 1996.

[129] C. Wu. On the convergence properties of the EM algorithm. *Ann. Statist*, 11:95–103, 1983.

[130] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1999.

[131] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Amer. Statist. Assoc*, 90:1200–1224, 1995.

[132] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook.* Technical University of Denmark, Feb. 2006.

[133] J. N. Bernardo and A. F. M. Smith. *Bayesian Theory.* Wiley and Sons, Chichester, England, 1994.

[134] E. T. Jaynes. *Probability Theory: the Logic of Science.* Cambridge University Press, 2003.

[135] P. Magniez, B. Muquet, P. Duhamel, V. Buzenac, and M. Decourville. Optimal decoding of bit-interleaved modulations: Theoretical aspects and practical algorithms. *Proc Int. Symp. on Turbo Codes and related topics*, pages 169–172, September 2004.

[136] S. Bories, A. Sibille, and C. Roblin. UWB indoor channel measurement. *in Proc. IWAT*, pages 466–469, March 2005.

[137] W. P. Siriwongpairat, M. Olfat, and K. J. R. Liu. Performance analysis and comparison of time hopping and direct sequence UWB-MIMO systems. *EURASIP J. Appl. Signal. Process. Special Issue on "UWB-state of the art"*, 2005:328–345, Mar. 2005.

[138] L. Yang and G. B. Giannakis. Analog space-time coding for multiantenna ultra-wideband transmissions. *IEEE Trans. Commun.*, 52(3):507–517, Mar. 2004.

[139] W. P. Siriwongpairat, W. Su, M. Olfat, and K. J. R. Liu. Multiband-OFDM MIMO coding framework for UWB communication systems. *IEEE Trans. Signal Process.*, 54(1):214–224, Jan. 2006.

[140] X. Zhang and B. Ottersten. Performance analysis of V-BLAST structure with channel estimation errors. *in Proc. Signal Process. Advances Wireless Commun. (SPAWC)*, pages 487–491, 2003.

[141] G. Jongren, M. Skoglund, and B. Ottersten. Design of channel-estimate-dependent space-time block codes. *IEEE Trans. Commun.*, 52(7):1191–1203, July 2004.

[142] K. Ahmed, C. Tepedelenhoglu, and A. Spanias. Effect of channel estimation on pair-wise error probability in OFDM. *in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pages 745–748, May 2004.

[143] A. Leke and J. M. Cioffi. Impact of imperfect channel knowledge on the performance of multicarrier systems. *in Proc. IEEE Global Telecommun. Conf. (Globecom)*, pages 951–955, Nov. 1998.

[144] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths. *IEEE Trans. Commun.*, 47:199–207, Feb. 1999.

[145] A. Lapidoth and S. Shamai. Fading channels: How perfect need 'perfect side information' be ? *IEEE Trans. Inform. Theory*, 48:1118–1134, May 2002.

[146] S. Sadough, P. Piantanida, and P. Duhamel. MIMO-OFDM optimal decoding and achievable information rates under imperfect channel estimation. *in Proc. Signal Process. Advances Wireless Commun. (SPAWC)*, Jun. 2007.

[147] S. Sadough, P. Piantanida, and P. Duhamel. Achievable outage rates with improved decoding of BICM Multiband OFDM under channel estimation errors. *in Proc. Asilomar Conf. on Signals, Systems and Computers*, Oct.-Nov. 2006.

[148] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Inform. Theory*, 44(6):2619–2692, Oct. 1998.

[149] J. J. Boutros, F. Boixadera, and C. Lamy. Bit-interleaved coded modulations for multiple-input multiple-output channels. *in Proc. Int. Symp. on Spread Spectrum Techniques and Applications*, pages 123–126, Sept. 2000.

[150] A. J. Paulraj and T. Kailath. Increasing capacity in wireless broadcast systems using distributed transmission/directional reception. Technical report, U.S. Patent 5 345 899, 1994.

[151] H. Bolcskei, D. Gesbert, and A. J. paulraj. On the capacity of OFDM-based spatial multiplexing systems. *IEEE Trans. Commun.*, 50(2):225–234, Feb. 2002.

[152] P. Robertson, P. Hoeher, and E. Villebrun. Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding. *European Trans. Telecommun.*, 8(2):119–125, Mar.-Apr. 1997.

[153] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley Series in Telecomunications. Wiley & Sons New York, 1991.

[154] G. G. Raleigh and J. M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Trans. Commun.*, 46:357–366, Mar. 1998.

[155] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai. On information rates for mismatched decoders. *IEEE Trans. Inform. Theory*, 40:1953–1967, Nov. 1994.

[156] J. Hirriart-Urrty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I.* Springer-Verlag, 1993.

[157] P. Piantanida, G. Matz, and P. Duhamel. Estimation-induced outage capacity of Ricean channels. *in Proc. Signal Process. Advances Wireless Commun. (SPAWC)*, July 2006.

[158] X. Li and J. A. Ritcey. Bit-interleaved coded modulation with iterative decoding. *in Proc. Int. Commun. Conf. (ICC)*, 2:858–863, June 1999.

[159] A. Chindapol and J. A. Ritcey. Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels. *IEEE J. Select. Areas Commun.*, 19(5):944–957, May 2001.

[160] A. Gorokhov and M. van Dijk. Optimized labeling maps for bit-interleaved transmission with turbo-demodulation. *in Proc. Vehicular Technol. Conf.*, 2001.

[161] B. M. Hochwald and S. ten Brink. Achieving near capacity on a multiple-antenna channel. *IEEE Trans. Commun.*, 51(3):389–399, Mar. 2003.

[162] H. Vikalo and B. Hassibi. Iterative decoding for MIMO channels via modified sphere decoding. *IEEE Trans. Wireless Commun.*, 3(3):2299–2311, Nov. 2004.

[163] M. Sellathurai and S. Haykin. Turbo-BLAST for wireless communications: theory and experiments. *IEEE Trans. Signal Process.*, 50(10):2538– 2546, Oct. 2002.

[164] H. Lee, B. Lee, and I. Lee. Iterative detection and decoding with improved V-BLAST for MIMO-OFDM systems. *IEEE J. Selected Areas Commun.*, 24(3):504–513, March 2006.

[165] B. Hassibi and B.M. Hochwald. High-rate codes that are linear in space and time. *IEEE Trans. Inform. Theory*, 48(7):1804–1824, July 2002.

[166] J.-C. Belfiore, G. Rekaya, and E. Viterbo. The golden code: a 2×2 full-rate space-time code with nonvanishing determinants. *IEEE Trans. Inform. Theory*, 51(4):1432–1436, Apr. 2005.

[167] R. G. Gallager. *Low-Density Parity-Check Codes*. PhD thesis, MIT, Cambridge, MA, 1963.

[168] D. J. C. Mackay. Good error-correcting codes based on very sparse matrices. *IEEE Trans. Inform. Theory*, 45:399–431, Mar. 1999.

[169] I. J. Cox, M. L. Miller, and A. L. McKellips. Watermarking as communications with side information. *Proceedings of the IEEE*, 87(7):1127–1141, Jul. 1999.

[170] B. Farhang-Boroujeny. Pilot-based channel identification: Proposal for semi-blind identification of communication channels. *Electronics Letters*, 31(13):1044–1046, 1995.

[171] M. Bilodeau and D. Brenner. *Theory of Multivariate Statistics*. New York: Springer, 1999.

[172] M. Schwartz, W. R. Bennett, and S. Stein. *Communications Systems and Techniques*. McGraw-Hill, 1966.

[173] I. E. Telatar. Capacity of multi-antenna Gaussian channels. Technical report, AT&T Bell Labs Tech. Memo., 1995.

[174] L. Ozarow, S. Shamai, and A. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Trans. Inform. Theory*, 43:359–378, May 1994.

[175] M. A. Khalighi, J.-F. Hélard, and S. Bourennane. Contrasting orthogonal and non-orthogonal space-time schemes for perfectly-known and estimated MIMO channels. *in Proc. ICCS*, Oct.-Nov. 2006. Singapore.

[176] G. Seber and A. Lee. *Linear Regression Analysis*. John Wiley and Sons, 2002.