



HAL
open science

Mise en oeuvre d'architectures de transcodage vidéo H.264/AVC et SVC

Christophe Deknudt

► **To cite this version:**

Christophe Deknudt. Mise en oeuvre d'architectures de transcodage vidéo H.264/AVC et SVC. Sciences de l'ingénieur [physics]. Université de Valenciennes et du Hainaut-Cambrésis, 2011. Français. NNT : 2011VALE0012 . hal-00799450

HAL Id: hal-00799450

<https://hal.science/hal-00799450v1>

Submitted on 18 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

présentée à

L'UNIVERSITÉ DE VALENCIENNES ET DU HAINAUT-CAMBRÉSIS

Pour l'obtention du grade de

DOCTEUR ÈS SCIENCES

Discipline : Micro et nano technologies, Acoustique et Télécommunications

Spécialité : Électronique

École Doctorale : Sciences pour l'Ingénieur (072)

par

CHRISTOPHE DEKNUDT

**Mise en œuvre d'architectures de transcodage vidéo
H.264/AVC et SVC : application à la transmission
optimisée de la vidéo haute définition**

Soutenance prévue le 18 juillet 2011 devant le jury composé de :

Rapporteurs :

- M. Y. POUSSET Maître de Conférences HDR, Institut de Recherche XLIM
UMR CNRS 6172, Département SIC, Poitiers
- M. P. LE CALLET Professeur des Universités, Institut de Recherche IRCCyN
UMR CNRS 6597, Nantes

Examineur :

- M. E. RENAN Ingénieur, THALES Communications, Colombes

Co-encadrant :

- M. P. CORLAY Maître de Conférences, IEMN OAE UMR CNRS 8520, Valenciennes

Directeurs de thèse :

- M. F.-X. COUDOUX Professeur des Universités, IEMN OAE
UMR CNRS 8520, Valenciennes
- M. M.-G. GAZALET Professeur des Universités, IEMN OAE
UMR CNRS 8520, Valenciennes

UNIVERSITÉ DE VALENCIENNES ET DU HAINAUT-CAMBRÉSIS
Institut d'Électronique, de Microélectronique et de Nanotechnologie (UMR CNRS 8520)
Département Opto-Acousto-Électronique

*Thèse présentée en vue de l'obtention du grade de
Docteur de l'Université de Valenciennes et du Hainaut-Cambrésis*

Discipline : Micro et nano technologies, Acoustique et Télécommunications
Spécialité : Électronique
École Doctorale : Sciences pour l'Ingénieur (072)

par

CHRISTOPHE DEKNUDT

**Mise en œuvre d'architectures de transcodage vidéo
H.264/AVC et SVC : application à la transmission
optimisée de la vidéo haute définition**

Soutenance prévue le 18 juillet 2011 devant le jury composé de :

Rapporteurs :

M. Y. POUSSET Maître de Conférences HDR, Institut de Recherche XLIM
UMR CNRS 6172, Département SIC, Poitiers
M. P. LE CALLET Professeur des Universités, Institut de Recherche IRCCyN
UMR CNRS 6597, Nantes

Examineur :

M. E. RENAN Ingénieur, THALES Communications, Colombes

Co-encadrant :

M. P. CORLAY Maître de Conférences, IEMN OAE UMR CNRS 8520, Valenciennes

Directeurs de thèse :

M. F.-X. COUDOUX Professeur des Universités, IEMN OAE
UMR CNRS 8520, Valenciennes
M. M.-G. GAZALET Professeur des Universités, IEMN OAE
UMR CNRS 8520, Valenciennes

Remerciements

Ce travail de thèse a été réalisé au sein du département Opto-Acousto-Électronique de l'Institut d'Électronique, de Microélectronique et de Nanotechnologies (IEMN-UMR8520) de l'Université de Valenciennes et du Hainaut Cambrésis. Je tiens à remercier les deux directeurs successifs du laboratoire, qui m'ont donné les moyens d'effectuer cette thèse, ainsi que le CNRS et l'Université de Valenciennes et du Hainaut Cambrésis de m'avoir permis de concilier travail et doctorat.

Je remercie également chaleureusement mes directeurs de thèse, Monsieur François-Xavier Coudoux et Monsieur Marc Gazalet, ainsi que mon co-encadrant, Monsieur Patrick Corlay, pour la qualité de leur encadrement et leur disponibilité, me permettant ainsi de mener à bien ces années de doctorat. En particulier merci à François-Xavier Coudoux de m'avoir proposé de rejoindre le projet TOSCANE, élément déclencheur de mon travail de thèse, mais aussi de m'avoir fait confiance dès le début.

Je tiens également à remercier Messieurs Patrick Le Callet et Yannis Pousset d'avoir accepté de juger la qualité de mon travail de thèse en tant que rapporteurs et leur exprime ma sincère reconnaissance.

Je tiens à exprimer ma gratitude envers Erwann Renan d'avoir accepté de faire partie du jury de thèse.

Ces années de thèse se sont déroulées très agréablement grâce à la compagnie de mes collègues de bureau : Anne-Sophie, Imade, Fabien, Thomas, Benoit, Sabrina, Julien, Kahena, Éloïse...

Un merci particulier à l'ensemble des membres du groupe Communications Numériques (COMNUM) de l'IEMN-DOAE, à mes voisins du bureau 112, ainsi que mes collègues IATOS, Fred, Marc et Liévin. Je remercie tous les autres membres du laboratoire IEMN-DOAE pour leur accueil chaleureux et leur bonne humeur et plus particulièrement nos trois secrétaires, Edith, Anne et Virginie, pour leur aide précieuse lors des formalités administratives.

Une mention spéciale à mes collègues de l'équipe DREAM, du Département Audiovisuel, avec qui je prends également beaucoup de plaisir à travailler.

D'un point de vue plus personnel, j'adresse un merci particulier à ma famille, mes parents et mes beaux-parents pour leur soutien.

Un remerciement particulier à ma jeune épouse Fanny pour son dévouement et sa compréhension.

Enfin, merci à tous ceux que je ne peux citer personnellement : membres du projet TOSCANE, enseignants et personnels de l'Université de Valenciennes...

Table des matières

Introduction générale.....	1
Chapitre 1: La compression vidéo H.264/AVC et son extension échelonnable	7
1.1 Introduction	7
1.2 La norme H.264/AVC	8
1.2.1 Organisation générale des données.....	8
1.2.1.1 Profils et niveaux.....	8
1.2.1.2 Organisation d'un flux H.264/AVC	9
1.2.1.3 Système colorimétrique et échantillonnage	10
1.2.1.4 Slices, macroblocs et blocs	10
1.2.1.5 Architecture d'un encodeur H.264/AVC	12
1.2.2 Prédiction des macroblocs	13
1.2.2.1 Codage des macroblocs de type intra	15
1.2.2.2 Codage des macroblocs de type P	18
1.2.2.3 Codage des macroblocs de type B	21
1.2.3 Traitement des coefficients résiduels	22
1.2.3.1 Transformées	22
1.2.3.2 Balayage en zig-zag	24
1.2.3.3 Quantification	25
1.2.3.4 Codage Run/Level.....	25
1.2.3.5 Le codage entropique CAVLC (Context-based Adaptive Variable Length Coding)	26
1.2.4 Évaluation des performances de H.264/AVC.....	29
1.3 L'extension échelonnable : H.264 SVC	31
1.3.1 Organisation d'un flux SVC	31
1.3.2 Les différents types d'échelonnabilité dans H.264 SVC.....	31
1.3.2.1 L'échelonnabilité temporelle	32
1.3.2.2 L'échelonnabilité spatiale	33
1.3.2.3 L'échelonnabilité en qualité	37
1.3.3 Évaluation des performances de H.264 SVC.....	40
1.4 Conclusion	43
Chapitre 2: Solution proposée de transcodage des flux vidéo haute définition en H.264/AVC	45
2.1 Introduction	45
2.2 État de l'art du transcodage H.264/AVC	46
2.2.1 Transcodage par requantification	49
2.2.2 Transcodage par sélection fréquentielle	50
2.3 Solution proposée de transcodage H.264/AVC à complexité réduite.....	51
2.3.1 Architecture	51
2.3.1.1 Traitement des macroblocs intra.....	52
2.3.1.2 Traitement des macroblocs inter.....	54
2.3.2 Performances de l'architecture de transcodage en sélection fréquentielle.....	55
2.4 Comparaison de la qualité visuelle du transcodage par sélection en position fréquentielle (FP) et requantification.....	56
2.4.1 Paramètres de transcodage utilisés	56
2.4.2 Critères de qualité visuelle utilisés.....	57
2.4.2.1 PSNR (RSBC)	57
2.4.2.2 Métrique VQM	58
2.4.2.3 Impact du transcodage des images intra et inter sur la qualité vidéo	58

Table des matières

2.4.3	Analyse PSNR	60
2.4.4	Analyse selon la métrique VQM	65
2.4.5	Approche subjective	70
2.5	Comparaison de la qualité visuelle du transcodage par sélection en position fréquentielle (FP) et sélection par niveau fréquentiel (FL)	71
2.5.1	Paramètres de transcodage utilisés	72
2.5.2	Analyse PSNR	72
2.5.3	Analyse VQM	75
2.6	Estimation du débit de flux H.264/AVC transcodés par sélection en position fréquentielle	77
2.6.1	Analyse du débit en fonction des paramètres de transcodage (FP _{intra} et FP _{inter})	78
2.6.2	Modélisation du débit	78
2.7	Conclusion	80
Chapitre 3:	Solution proposée de transcodage à complexité réduite des flux H.264 SVC en échelonnabilité spatiale	81
3.1	Introduction	81
3.2	Échelonnabilité spatiale des images intra et inter dans H.264 SVC	82
3.2.1	Prédiction des macroblocs	82
3.2.2	Configuration des encodages H.264 SVC en échelonnabilité spatiale	86
3.2.3	Statistiques sur l'utilisation du type de codage utilisé pour les macroblocs de type intra-image de la couche d'amélioration haute	86
3.2.4	Estimation du surcoût de débit engendré par un forçage en Intra_BL des macroblocs intra et impact sur le PSNR	88
3.3	Transcodage H.264 SVC en échelonnabilité spatiale par sélection en position fréquentielle des coefficients résiduels de la couche d'amélioration haute	92
3.3.1	Architecture proposée	92
3.4	Résultats expérimentaux du transcodage de flux en échelonnabilité spatiale	93
3.4.1	Réduction de débit	94
3.4.2	Analyse PSNR	94
3.5	Limitation de l'architecture de transcodage à la composante de luminance	105
3.6	Étude de l'influence du paramètre de transcodage FP _{inter} sur la réduction de débit et la qualité vidéo	107
3.7	Conclusion	113
Chapitre 4:	Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL	115
4.1	Introduction	115
4.2	Extension de la zone de couverture des services de télévision haute définition sur le canal ADSL	117
4.2.1	Principes de l'ADSL	117
4.2.1.1	Perturbations rencontrées sur les lignes ADSL	118
4.2.1.2	Principales caractéristiques de l'ADSL	119
4.2.1.3	Évolutions de l'ADSL	121
4.2.1.4	La télévision sur IP	122
4.2.2	Objectif du scénario	124
4.2.3	Paramètres d'encodage et de transcodage vidéo	125
4.2.4	Paramètres de simulation de la transmission ADSL	127
4.2.4.1	Caractéristiques des lignes ADSL utilisées lors de la simulation	128
4.2.4.2	Détermination de la zone initiale d'éligibilité	130
4.2.5	Résultats expérimentaux	132
4.2.5.1	Transcodage H.264 SVC par sélection en position fréquentielle des coefficients résiduels de la couche d'amélioration	132
4.2.5.2	Expérimentations sur la boucle 1	135
4.2.5.3	Expérimentations sur la boucle 3	140
4.2.5.4	Expérimentations sur la boucle 8	142

Table des matières

4.3	Amélioration de la qualité de service de la transmission des flux H.264/AVC sur le canal CPL.....	144
4.3.1	Présentation du canal CPL.....	146
4.3.2	Description du système adaptatif de transmission vidéo proposé.....	148
4.3.3	Paramètres d'encodage et de transcodage vidéo.....	148
4.3.4	Paramètres de simulation de la modulation DMT sur canal CPL.....	149
4.3.5	Résultats expérimentaux.....	150
4.4	Conclusion.....	154
	Conclusion générale et perspectives.....	157
	Annexes.....	161
	Annexe A.....	161
	Annexe A.1 : Identification des NALU dans un flux H.264/AVC.....	161
	Annexe A.2 : Équivalence entre le paramètre de quantification et le pas de quantification.....	162
	Annexe A.3 : Comparaison des performances des normes MPEG-2 et H.264/AVC (expérimentations).....	162
	Annexe A.4 : Influence des modes de prédiction intra sur l'encodage H264/AVC ...	164
	Annexe B.....	165
	Annexe B.1 : Comparaison de la qualité visuelle du transcodage par sélection en position fréquentielle et requantification.....	165
	Annexe B.2 : Comparaison de la qualité visuelle du transcodage par sélection en position fréquentielle et requantification.....	168
	Séquences 1080p.....	168
	Séquences 4CIF et 4SIF.....	172
	Annexe C : Le standard Homeplug AV.....	178
	Liste des figures.....	183
	Liste des tableaux.....	187
	Glossaire.....	189
	Bibliographie.....	193

Introduction générale

Le développement en nombre des systèmes de télécommunication permettant la diffusion de contenus audiovisuels tels que la télévision numérique (par câble, satellite, réseau hertzien ou ligne ADSL), l'Internet, la téléphonie mobile 3G ont contribué à la multiplication des contenus audiovisuels et obligé les opérateurs à tenir compte de l'hétérogénéité croissante des réseaux de transmission (de type filaire, sans-fil ou mixte) ainsi que des terminaux vidéo (affichage en définition standard, haute définition, à différentes fréquences image...) en adaptant la transmission aux multiples combinaisons service/canal/récepteur.

Une technique couramment employée pour la diffusion d'un contenu à destination de récepteurs divers est le *simulcast*. Cette solution consiste à décliner un même contenu en plusieurs versions adaptées aux différents formats requis (i.e. aux différents types de terminaux visés), et à partager les ressources (mémoire, bande passante) entre ces différentes versions. Le *simulcast* implique nécessairement complexité lors de la création et de la gestion des contenus (multiples encodages, gestion de multiples versions) et gaspillage des ressources. De plus il offre une flexibilité très limitée qui ne permet pas d'adapter finement les contenus à la diversité des besoins des clients. Deux alternatives au *simulcast* existent : le transcodage vidéo et l'utilisation d'une norme de compression permettant une forme d'échelonnabilité du flux vidéo.

Le transcodage inter-format (par exemple de la norme de compression MPEG-2 vers le H.264/AVC), ou le transcodage intra-format (par exemple du H.264/AVC haut débit vers du H.264/AVC bas débit), consistent à ré-encoder un contenu vidéo pour l'adapter aux besoins du client. Ce type de solution peut être mis en oeuvre dans des scénarios *multicast*, qui consistent à diffuser un contenu unique vers un groupe d'utilisateurs. En certains points du réseau, le flux vidéo est ré-encodé à débit réduit pour être diffusé vers une branche du réseau ayant une bande passante plus limitée. Alors, cette solution implique des traitements additionnels à l'intérieur du réseau soit au niveau du serveur, de la passerelle ou du client. Les variations dynamiques de la bande passante et de sa qualité dans des environnements mobiles nécessitent également des fonctions d'adaptation.

L'utilisation d'une norme de compression vidéo échelonnable permet d'utiliser un unique flux vidéo contenant plusieurs représentations d'un même contenu à des définitions spatiales, fréquences image et/ou niveaux de qualité différents. Une dépendance est créée entre les

couches inférieures, indispensables au décodage du flux vidéo, et les couches supérieures. Alors, il est possible de sélectionner la couche de base (première couche) et éventuellement des couches successives d'amélioration afin de disposer d'un flux vidéo à un débit paramétrable, limité par le nombre de couches.

Le projet TOSCANE (Transmission vidéo Optimisée Source/ Canal échelonnable) s'inscrit dans ce contexte et porte sur le développement d'un système de diffusion mettant en oeuvre conjointement des techniques de codage de source, s'appuyant sur des techniques de compression vidéo non-échelonnables et échelonnables, et de codage de canal, reposant sur l'optimisation des paramètres de transmission.

Les retombées des résultats du projet visent l'amélioration de la couverture globale de diffusion des contenus audiovisuels au sein des réseaux filaires de type ADSL, et sans fil (WiFi en particulier), et la simplification de la mise en oeuvre de réseaux multimédia domestiques. L'optimisation de la transmission vidéo est effectuée au moyen de la norme de compression vidéo H.264/AVC et de son extension échelonnable (H.264 SVC), mais également d'une adaptation des contenus vidéo via transcodage et d'une architecture matérielle temps réel reconfigurable.

Le transcodage ou l'échelonnabilité permettent d'accéder à différents niveaux de définition (spatiale, temporelle et/ou de qualité). Ceci permet d'envisager chez un même abonné la possibilité d'utiliser différents types de terminaux ou de réseaux d'accès, afin de visualiser un même programme ou des programmes différents. Alors, les opérations d'adaptation et d'optimisation conjointe peuvent être réalisées soit au niveau du central ADSL, au niveau du modem routeur de l'abonné, ou bien au niveau du terminal. Pour y parvenir, il est nécessaire de prendre également en considération les caractéristiques et paramètres de transmission (puissance d'émission, perturbations réseau) afin d'optimiser la qualité de service et la qualité d'expérience utilisateur. Le choix d'une métrique temps réel de la qualité de service du réseau a permis, d'une part, l'optimisation et l'adaptation de contenus en boucle fermée. D'autre part, un second niveau d'adaptation a été rendu possible grâce à la conception d'architectures matérielles temps réel.

Enfin, la méthode de transport RTP/UDP a été choisie pour la diffusion des flux vidéo sur le réseau en raison de ses avantages techniques : flexibilité offerte en matière d'adaptation des contenus, filtrage de couches simplifié, multi-diffusion et interopérabilité avec les autres services.

Dans le domaine de l'optimisation du codage source, le projet a permis l'élaboration de techniques d'adaptation de contenus par transcodage, filtrage, échelonnabilité (SNR, temporelle ou spatiale) selon le format de codage H.264/AVC ou SVC.

Dans le domaine de l'optimisation du codage canal, des techniques de protection inégale des flux vidéo compressés vis-à-vis des erreurs de transmission ont été élaborées (optimisation du transcodage conjoint à partir de métriques sans référence, allocation multi-résolution du débit et de la puissance, mise en oeuvre des résultats obtenus dans le scénario diffusion live dans la plateforme de démonstration du projet).

Inclus dans les schémas d'optimisation cités précédemment, les travaux de thèse décrits dans ce manuscrit s'inscrivent dans le projet TOSCANE et sont essentiellement consacrés à l'étude et la mise en oeuvre d'architectures de transcodage vidéo pour H.264/AVC et son extension échelonnable. Le manuscrit se décompose en quatre chapitres : le premier est bibliographique et les trois autres présentent les solutions originales. Le premier chapitre de ce manuscrit est consacré à la norme H.264/AVC et à son extension échelonnable. Nous présentons les divers outils de codage employés et en particulier les solutions innovantes apportées par cette norme (comme la prédiction intra-image), nécessaires à la compréhension des architectures de transcodage présentées dans les chapitres suivants.

Dans le deuxième chapitre, nous présentons une première architecture de transcodage des flux H.264/AVC par sélection des coefficients résiduels fréquentiels. Cette solution de faible complexité applique le traitement sur l'ensemble des macroblocs. Le chapitre débute par un état de l'art du transcodage vidéo dans le cas de la norme H.264/AVC. Nous poursuivrons par une comparaison des performances en termes de réduction de débit et de qualité vidéo entre notre solution de transcodage vidéo H.264/AVC et la requantification, solution alternative couramment utilisée. Nous montrons que notre solution donne dans la majorité des cas de meilleurs résultats. Ensuite, nous présentons une comparaison des performances des deux techniques implémentées de transcodage vidéo par sélection fréquentielle : la sélection en position fréquentielle et la sélection en niveau fréquentiel. Nous verrons que le transcodage par sélection en position fréquentielle a été retenu pour sa faible complexité, sa meilleure flexibilité en termes de répartition des débits intermédiaires et sa qualité vidéo justifiée par les mesures PSNR et VQM. Puis, nous proposons une modélisation simple et efficace du débit dans le cadre du transcodage vidéo par sélection en position fréquentielle afin d'estimer le débit d'un flux vidéo transcodé.

Nous proposons au chapitre 3 une extension de l'architecture de transcodage des flux H.264/AVC au cas des flux H.264 SVC en échelonnabilité spatiale. À l'encodage, il est

possible de concevoir des flux H.264 SVC combinant l'échelonnabilité spatiale et l'échelonnabilité en qualité. Cependant cette technique est rarement employée car elle est au détriment de l'efficacité d'encodage des flux et nécessite de prévoir dès l'encodage les débits vidéo visés. L'objectif de ce chapitre est de fournir des débits vidéo intermédiaires des flux tout en conservant la définition spatiale la plus haute. Le principe de cette solution de transcodage consiste à sélectionner les coefficients résiduels de la couche d'amélioration la plus haute en fonction de leur position en fréquence. Afin d'assurer un traitement temps réel des flux vidéo, l'architecture de transcodage doit être de faible complexité. Alors, nous proposons de transposer la partie en boucle ouverte de l'architecture de transcodage H.264/AVC au cas H.264 SVC. Ainsi, nous pourrions disposer de débits intermédiaires situés entre le débit de l'ensemble des couches inférieures et le débit total du flux vidéo, assurant une qualité progressive de l'image avec la définition la plus élevée. Dans un premier temps, nous rappelons les modes de prédiction intervenant dans l'échelonnabilité spatiale. Puis nous proposons et évaluons les performances d'une première architecture de transcodage appliquant le traitement sur l'ensemble des macrobloccs de la couche d'amélioration. Nous montrerons qu'en fonction du paramètre de quantification utilisé lors de l'encodage, la réduction de débit peut atteindre 66%. Enfin, dans l'objectif de limiter les distorsions visuelles dues au traitement et de minimiser la complexité de notre architecture de transcodage SVC, nous proposons et évaluons deux variantes de cette architecture. La première consiste à restreindre le traitement aux macrobloccs de luminance. Quant à la seconde, elle consiste à restreindre le traitement aux macrobloccs utilisant la prédiction inter-image.

Finalement, le chapitre 4 est consacré à la mise en oeuvre des deux architectures de transcodage proposées dans les chapitres précédents à travers deux scénarios de transmission vidéo haute définition. Nous verrons que ces deux solutions de transcodage permettent de s'adapter aux variations des voies de diffusion. Dans la première partie de ce chapitre, nous proposons un scénario de transmission sur le canal ADSL permettant d'étendre la zone de couverture des services de télévision haute définition, mettant en oeuvre l'architecture de transcodage des flux H.264 SVC présentée au chapitre 3. Nous montrons que la zone d'éligibilité des services de télévision haute définition peut être significativement étendue d'un minimum de 650 mètres, en fonction de la distance du domicile au DSLAM et des caractéristiques de la ligne, tout en fournissant une qualité vidéo progressive aux abonnés ADSL. Dans la deuxième partie du chapitre, nous proposons une solution améliorant la qualité de service des flux vidéo haute définition transmis sur le réseau électrique d'un domicile par courant porteur en ligne, soumis à de fréquents changements d'états. Ce dernier scénario de transmission utilise l'architecture de transcodage des flux

H.264/AVC proposée dans le chapitre 2. Moyennant une légère baisse de PSNR, les résultats des simulations prouvent que notre solution assure une qualité vidéo constante contrairement à une transmission avec erreurs caractérisée par des variations arbitraires de la qualité vidéo.

Le manuscrit se termine par une conclusion générale ainsi que les perspectives de poursuite de ces travaux de thèse.

Chapitre 1: La compression vidéo H.264/AVC et son extension échelonnable

1.1 Introduction

Ces dernières décennies, deux besoins majeurs ont poussé le développement de la compression vidéo : la transmission d'une vidéo sous contrainte de débit et l'optimisation du stockage vidéo. Depuis la fin des années 1990, la norme de compression vidéo *MPEG-2* est devenue incontournable. Elle s'est imposée comme format de compression vidéo des *DVD-vidéo* et a été utilisée également par le *SVCD* ou *super video compact disc*. Nous retrouvons également la norme *MPEG-2* avec l'apparition de la télévision numérique par satellite (*DVB-S*) en définition standard, tout comme la télévision numérique sur *ADSL*. Plus récemment, depuis février 2005 en France, la télévision numérique terrestre (*TNT*) en définition standard est conforme à la norme *DVB-T* et utilise également la compression *MPEG-2* pour les chaînes sans abonnement.

Afin de faire évoluer et de remplacer les normes existantes (*H.263*, *MPEG-2* et *MPEG-4 part 2*) les groupes *VCEG (ITU-T)* et *MPEG* ont travaillé conjointement en créant le groupe de travail *JVT*. Son but est de concevoir une nouvelle norme plus efficace en terme de compression et plus flexible afin de s'adapter à de nombreux terminaux. L'avancement technologique des microprocesseurs a permis d'élaborer des algorithmes plus coûteux en temps de calcul mais plus performants. C'est ainsi qu'est apparue en 2003 la première version de la norme *H.264/AVC (Advanced Video Coding)*, appelée également *MPEG-4 part 10*. Elle est utilisée pour la compression des flux vidéo transmis en *TNT HD*, par les chaînes payantes de la *TNT*, mais également de plus en plus par les opérateurs *ADSL*, sans oublier les disques vidéo *Blu-ray*.

Nous commencerons par présenter la norme *H.264/AVC* dans la section 1.2, puis nous évoquerons ses principes de base, ainsi que les nouveaux outils de compression qu'elle met à disposition. Certains outils sont également utilisés dans les architectures de transcodage vidéo proposées dans le cadre de nos travaux (chapitres 2 et 3) et sont ici davantage détaillés. Depuis quelques années, de nombreux terminaux vidéo sont apparus dans les foyers en plus de la set-top box traditionnelle couplée à une télévision. Les téléphones portables, consoles de jeux ou ordinateurs sont également des terminaux à prendre en compte avec des caractéristiques spécifiques. Aussi, la définition des écrans des téléphones portables ne permet pas pour le moment d'afficher des vidéos en haute définition. A cela

s'ajoute également des problématiques de puissance de calcul nécessaires au décodage des flux vidéo ou de bande passante avec les connexions sans fil. Il est alors nécessaire d'avoir à disposition le même contenu vidéo avec des définitions différentes. L'extension échelonnable de H.264/AVC (H.264 SVC pour *Scalable Video Coding*), introduite en 2005, répond à ces besoins en évitant de transmettre des informations redondantes communes aux divers flux afin d'économiser de la bande passante. Le H.264 SVC est présenté dans la section 1.3.

1.2 La norme H.264/AVC

Dans cette section, nous allons présenter la norme H.264/AVC (*Advanced Video Coding*), appelée également MPEG-4 Part 10 [ITU09]. Cette norme a pour objectif de s'adapter à de multiples domaines, allant de la vidéo sur téléphone mobile à la télévision haute définition, en passant par la vidéo surveillance. Cette section est dédiée à la présentation des différents outils de codage vidéo utilisés par la norme H.264/AVC mais également par nos différentes architectures de transcodage vidéo.

1.2.1 Organisation générale des données

1.2.1.1 Profils et niveaux

Les *profils et niveaux* ont été créés dès la mise en place de la norme MPEG-2. Ils facilitent la compatibilité entre des applications aux caractéristiques similaires. Nous pouvons prendre pour exemple le cas d'un opérateur ADSL qui se doit d'encoder les flux vidéo des chaînes de télévision conformément aux caractéristiques de décodage du boîtier vidéo présent chez l'abonné.

Ainsi, le *profil* précise l'ensemble des outils de codage et algorithmes utilisés pour créer le flux vidéo. Un décodeur conforme à un profil donné doit supporter toutes les fonctionnalités de celui-ci. À l'origine, H.264/AVC définissait trois profils : *Baseline*, *Main* et *Extended*. D'autres profils ont été ajoutés afin d'étendre le standard à d'autres types d'applications (profils *High*, *High 10*, *High 4:2:2* et *High 4:4:4*) et notamment professionnelles (profils *High 10 Intra*, *High 4:2:2 Intra*, *High 4:4:4 Intra*, *CAVLC 4:4:4 Intra* et *High 4:4:4*). Dans le cadre des travaux du projet TOSCANE, nous avons opté pour le profil *Main* qui est le plus adapté à la diffusion vidéo grand public. Nous noterons que contrairement au profil *Baseline*, il permet de coder des images bidirectionnelles.

Le *niveau* quant à lui impose des limitations sur certains paramètres du flux. Quinze niveaux sont définis afin de préciser la taille maximale de l'image en nombre de macroblocs

(cf 1.2.1.4) allant du QCIF à la très haute définition (4096x2304), la capacité de la mémoire du décodeur, les performances du décodeur ou le débit vidéo.

1.2.1.2 Organisation d'un flux H.264/AVC

La norme H.264/AVC s'ouvrant à de multiples applications et types de réseaux, il a fallu créer une organisation flexible et évolutive des données. Dans ce but, les données relatives à la compression de l'image sont regroupées dans la couche de codage vidéo ou Video Coding Layer (VCL). Elle est englobée par une autre couche orientée réseau ou Network Abstraction Layer (NAL). Cette couche NAL formate la couche VCL et y ajoute des informations afin de permettre le stockage de la vidéo ou son acheminement par des couches de transport réseau (Figure 1.1).

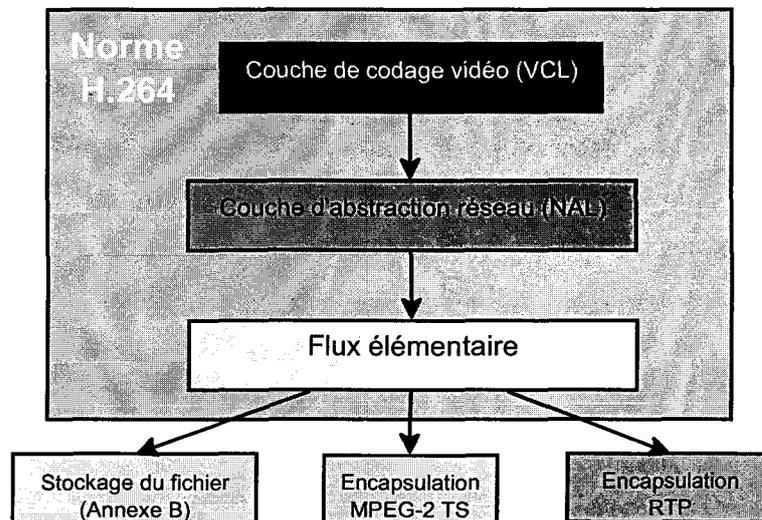


Figure 1.1 Représentation en couches d'un flux H.264

La couche NAL crée des sous-ensembles de données représentant des données vidéo et/ou des informations nécessaires au bon fonctionnement du décodage, appelés *NALU* ou unités de NAL. L'ensemble des unités de NAL nécessaires au décodage d'une image est appelé unité d'accès. Le flux élémentaire alors créé peut être transmis après paquetsation RTP (*Real-time Transport Protocol*) ou MPEG-2 TS (*Transport Stream*) sur un réseau, ou bien encore stocké sur un support de données. Afin de marquer le début d'une nouvelle NALU, une succession de trois octets appelée *start code prefix* est insérée. Celui-ci se compose de deux octets égaux à zéro et d'un octet égal à un. Il peut être précédé dans certains cas d'un octet égal à zéro (*zero_byte*). Pour prévenir un éventuel problème lié à l'apparition d'un start code en plein milieu d'une NALU, une règle appelée *emulation_prevention_three_byte* est appliquée. Un octet égal à 0x03 est inséré avant l'octet égal à 0x01 de la séquence d'octets qui pose problème. La Figure 1.2 illustre le flux binaire

d'une NALU avec son start code et son entête, suivis des données de la NALU ([ITU09] section B.1).

(Zero_byte) (1 octet)	Start code prefix (3 octets)	Entête (1 octet)			Données de la NALU (longueur variable)
0x00	0x000001	f	nr	nauty	
<i>f: forbidden_zero_bit (1 bit)</i> <i>nr: nal_ref_idc (2 bits)</i> <i>nauty: nal_unit_type (5 bits)</i>					

Figure 1.2 Représentation du flux binaire d'une NALU

Dans l'entête de la NALU, un mot de cinq bits représente le type de NALU. Il est appelé *nal_unit_type* dans la norme. À titre indicatif, l'annexe A.1 fournit une table de correspondance entre la désignation de la NALU et sa numérotation.

1.2.1.3 Système colorimétrique et échantillonnage

La vision humaine étant plus sensible aux variations de luminosité qu'aux variations des couleurs, la norme propose de base un codage des vidéos au format 4:2:0 [RIC10]. Ainsi, le nombre d'échantillons de chrominance est divisé par deux horizontalement et verticalement. Les échantillons de luminance ou de chrominance sont codés de base sur 8 bits (0 à 255). Afin de satisfaire les exigences du cinéma numérique et des autres professionnels de l'image, H.264/AVC peut gérer des échantillons codés sur 10 bits (0 à 1023, comprenant des infra-noirs et super-blancs). Il faut noter que le profil High permet de coder des vidéos au format 4:2:2 voire 4:4:4. Dans le cadre de nos travaux, nous avons utilisé des images codées sur 8 bits par composante, sous-échantillonnées en 4:2:0 afin de rester dans un scénario de transmission vidéo utilisant des débits réalistes.

1.2.1.4 Slices, macrobloccs et blocs

La norme H.264/AVC introduit la notion de *slice* ou découpage de l'image en plusieurs régions [SUL05]. La Figure 1.3 illustre un exemple d'une image divisée en trois slices. Nous noterons qu'une slice peut être au minimum égale à un macrobloc et au maximum égale à une image complète.

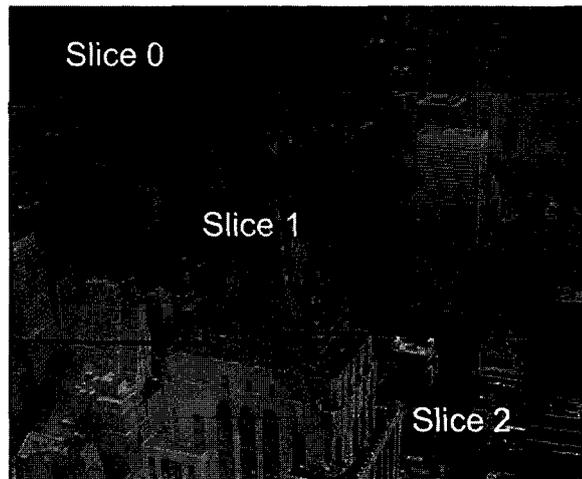


Figure 1.3 Découpage d'une image en slices

Comme les normes précédentes, H.264/AVC repose sur un découpage de chaque slice en sous-ensembles de taille 16x16 pixels, appelés *macrobloks*. La Figure 1.4 présente un exemple de partitionnement d'une image composée de trois slices en macrobloks et leur numérotation. Cette numérotation s'effectue par un balayage de gauche à droite, du haut vers le bas, celui-ci est appelé *raster scan*. Ces macrobloks peuvent être également découpés en plus petits sous-ensembles de 4 par 4 pixels, appelés *blocs* ou *sous-macrobloks*. H.264/AVC permet également de travailler avec des blocs de 8 par 8 pixels mais nous n'avons pas utilisé ce type de blocs dans nos travaux car ils ne sont pas disponibles dans le profil *Main*.

0	1	2	3	4	5	6	7	8	9	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24	25
26	27	28	29	30	31	32	33	34	35	36	37	38
39	40	41	42	43	44	45	46	47	48	49	50	51
52	53	54	55	56	57	58	59	60	61	62	63	64
65	66	67	68	69	70	71	72	73	74	75	76	77
78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100	101	102	103
104	105	106	107	108	109	110	111	112	113	114	115	116
117	118	119	120	121	122	123	124	125	126	127	128	129
130	131	132	133	134	135	136	137	138	139	140	141	142
143	144	145	146	147	148	149	150	151	152	153	154	155

Figure 1.4 Exemple de numérotation des macrobloks dans une image composée de trois slices

A l'encodage et au décodage, le traitement des blocs s'effectue selon un balayage spécifique ([ITU09] section 6.4.3). La Figure 1.5 illustre l'ordre de balayage des blocs de 8 par 8 pixels au sein d'un macrobloc de luminance. La Figure 1.6 fait de même avec un macrobloc de luminance partitionné en 16 blocs de 4 par 4 pixels. Nous remarquerons que le balayage suit la même logique qu'avec les blocs de 8 par 8 pixels.

0	1
2	3

Figure 1.5 Ordre de balayage des blocs luminance 8x8

0	1	4	5
2	3	6	7
8	9	12	13
10	11	14	15

Figure 1.6 Ordre de balayage des blocs luminance 4x4

Nous ajouterons que la création d'une slice se fait par le regroupement de macroblocs successifs dans l'ordre du balayage de type raster scan. Il n'est pas donc possible que les slices suivent un motif spécifique.

Un outil de H.264/AVC nommé flexible macrobloc ordering (*FMO*) permet de séparer des macroblocs [LAM06] afin de créer des groupes. Il peut être utilisé comme outil de robustesse aux erreurs et de codage par régions d'intérêt ou Region of Interest (*ROI*).

1.2.1.5 Architecture d'un encodeur H.264/AVC

La Figure 1.7 fournit un schéma de principe d'un encodeur H.264/AVC [WIE03a]. Chaque macrobloc ou bloc de la source vidéo est codé en calculant l'erreur entre celui-ci et la prédiction intra ou inter-image. Le résultat de cette opération aboutit à un bloc de coefficients résiduels notés $R(x, y)$. Ce bloc de coefficients subit ensuite une transformée entière (T), une quantification (Q) et un codage entropique afin d'intégrer ces données dans le flux vidéo. L'architecture de l'encodeur est en boucle fermée afin de réaliser les prédictions sur les blocs précédemment encodés puis décodés et non sur l'image originale.

Ainsi, le résidu transformé et quantifié, subit une quantification inverse et une transformée inverse. Nous obtenons un bloc de coefficients résiduels notés $R'(x, y)$. La prédiction intra ou inter-image est ajoutée à ce bloc de coefficients afin de reconstituer le bloc dans le domaine spatial. Ce bloc est ensuite ajouté au *buffer image* afin d'être utilisé par la prédiction. Il faut noter que ce bloc constitue une version approchée de l'original car les fréquences éliminées par la quantification sont définitivement perdues.

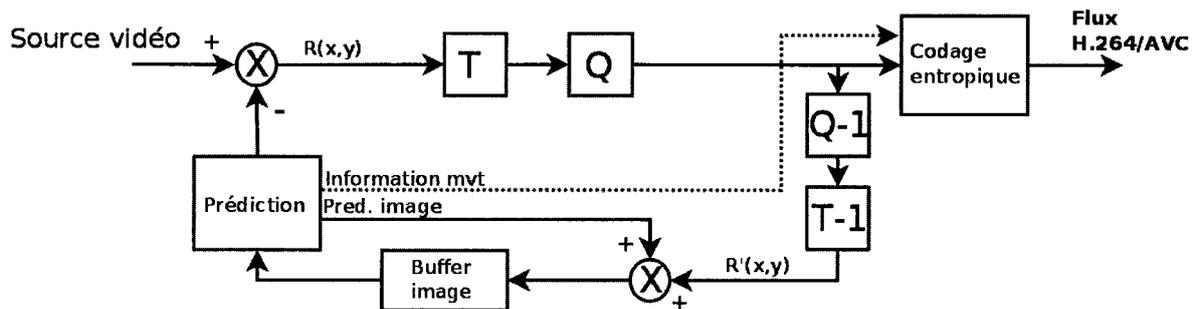


Figure 1.7 Schéma bloc d'un encodeur H.264/AVC

Le H.264/AVC permet également l'utilisation ou non d'un filtre anti-bloc (également appelé *loop filter* ou *deblocking filter*) à l'encodage et au décodage. Ce filtre est appliqué à chaque bloc décodé en vue de réduire la distorsion due à l'effet de bloc. Au niveau du décodeur, ce filtre est appliqué juste après la transformée inverse du bloc de coefficients résiduels. Il faut noter que contrairement à la prédiction inter-image, la prédiction intra-image utilise les blocs intra avant filtrage. Les avantages visés sont les suivants : lissage des bords des blocs afin d'améliorer l'apparence des images surtout pour les taux de compression élevés, limitation des résidus après compensation de mouvement. Le filtre de reconstruction est appliqué sur les bords verticaux puis horizontaux des blocs 4×4 de luminance, puis de chrominance. Enfin, nous noterons que les performances de ce filtre ne sont pas toujours au rendez-vous ou sont négligeables par rapport à sa complexité dans le cas d'un traitement temps réel [ZHO05].

1.2.2 Prédiction des macroblocs

Dans une slice de H.264/AVC, un macrobloc peut être codé par deux types de prédiction [WIE03a] [RIC10] :

- La prédiction de type inter-image existait déjà en MPEG-2, elle repose sur le principe que des images successives se ressemblent et qu'il est possible de trouver de fortes corrélations entre l'image courante et l'image précédente et/ou suivante, dans des régions plus ou moins proches du macrobloc à coder. La prédiction inter-image débute par la recherche dans une autre image (image $i-1$, voire image $i+1$) du macrobloc le plus ressemblant possible au macrobloc à

coder. Quand le meilleur candidat est identifié, le vecteur de mouvement est calculé afin de caractériser le déplacement du macrobloc candidat au macrobloc à coder. Le macrobloc prédit est calculé à partir de ce vecteur mouvement. Puis le macrobloc résiduel est obtenu en calculant l'erreur entre le macrobloc à coder et le macrobloc prédit. Notre macrobloc codé peut être de type P (Prédit - utilise une prédiction basée sur l'image précédente) ou B (Bidirectionnel - utilise une prédiction basée sur l'image précédente et suivante).

- La prédiction intra-image utilise les redondances spatiales de l'image pour prédire le macrobloc courant. Elle se base sur le fait que des pixels voisins sont fortement corrélés. Elle a été étudiée et introduite dans le cadre de cette norme. Un macrobloc peut devenir candidat à la prédiction intra-image si son numéro (ou adresse) est supérieur ou égal à zéro et inférieur ou égal à celui du macrobloc courant, tout en se situant dans la slice en cours de codage.

Chaque macrobloc peut être traité par un ou plusieurs modes de codage, dépendant du type de la slice à laquelle il appartient. Dans H.264/AVC, nous distinguerons trois types de slices : les slices de type intra (contenant exclusivement des macroblochs intra), les slices de type P (contenant des macroblochs intra et prédits) et les slices de type B (contenant des macroblochs intra, prédits et bidirectionnels).

Ces slices sont organisées par groupes d'images ou *Group of Pictures* (GOP) afin de définir la structure cyclique de la prédiction inter-image de la séquence vidéo. Un GOP débute par une image Intra. À l'encodage, nous spécifions le nombre d'images à ignorer avant l'insertion d'une image P et le nombre d'images P présentes dans le GOP. Les images B sont insérées entre les images I et P. La Figure 1.8 représente un exemple de GOP composé d'une image intra, de trois images prédites et de huit images bidirectionnelles. La longueur du GOP ou la période d'apparition d'une image de type intra correspond à la somme des images intra, prédites et bidirectionnelles.

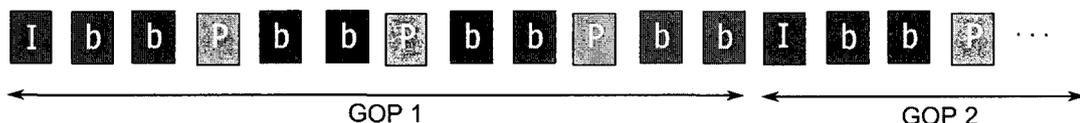


Figure 1.8 Exemple de représentation d'un GOP de 12 images

1.2.2.1 Codage des macroblocs de type intra

Les macroblocs de type intra peuvent être présents dans tous les types de slices. Le codage d'un macrobloc intra peut se faire avec prédiction en utilisant les échantillons des blocs voisins déjà codés et situés en haut et à gauche du bloc courant, ou sans prédiction, dans ce cas un codage appelé Intra PCM est utilisé.

Contrairement aux normes précédentes (H.263, MPEG-2 ou MPEG-4 Part 2), où la prédiction intra-image se faisait dans le domaine transformé, la prédiction des macroblocs de type intra se réalise dans le domaine spatial et peut se faire selon deux codages pour la luminance :

- Soit le macrobloc courant fait partie d'une zone de l'image peu détaillée, comme un aplat de couleur, dans ce cas un codage appelé INTRA_16x16 sera utilisé.
- Soit le macrobloc fait partie d'une zone très détaillée de l'image, celui-ci contient déjà des pixels bien différents, dans ce cas un codage plus adapté sera utilisé. Il est appelé INTRA_4x4 et divise le macrobloc en seize blocs de quatre par quatre pixels.

À l'encodage, le choix du type de codage d'un bloc (ou macrobloc) intra pour un mode de prédiction donné se fait en calculant la corrélation entre le bloc à coder et le bloc prédit pour chaque combinaison (type de codage – mode de prédiction) selon un critère donné. Le meilleur candidat correspond au choix qui minimise ce critère. Un critère d'évaluation couramment utilisé est la somme des valeurs absolues des différences ou *Sum of Absolute Differences (SAD)* entre le bloc à coder b et le bloc prédit p :

$$C_{SAD} = \sum_{x,y} |b(x,y) - p(x,y)| \quad (1.1)$$

Le codage d'un macrobloc INTRA_4x4 suit donc cette logique. Pour chacun des blocs 4x4, le module de prédiction va calculer neuf blocs prédits en utilisant les neuf modes de prédiction. Le meilleur candidat est le bloc, pour un mode donné, qui minimise l'erreur résultant de la différence entre le bloc courant et le bloc prédit.

Le bloc prédit est représenté en Figure 1.9. Il comporte seize échantillons de luminance, désignés par les lettres minuscules a-p. Les échantillons des blocs voisins sont quant à eux notés A-M.

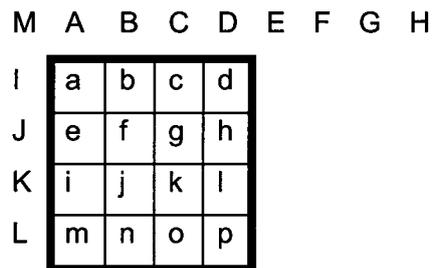


Figure 1.9 Désignation des échantillons d'un bloc prédit 4x4 et des échantillons voisins

La Figure 1.10 illustre les neuf modes de prédiction intra. Nous remarquons le mode DC qui représente le niveau lumineux moyen des échantillons voisins, ainsi que huit modes de direction afin de s'adapter aux différentes transitions éventuelles dans le bloc. Ceux-ci sont appelés : vertical, horizontal, diagonal bas/gauche, diagonal bas-droit, vertical-droit, horizontal-bas, vertical-gauche et horizontal-haut.

Les huit modes de direction ne sont utilisables que lorsque tous les échantillons nécessaires des pixels voisins sont disponibles. Il existe cependant une exception à cette règle lorsque les échantillons E-H ne sont pas disponibles. Dans ce cas l'échantillon D est dupliqué afin de rendre les modes considérés utilisables. Ceci peut arriver quand le bloc voisin n'est pas encore codé, n'est pas de type intra, ou se situe à l'extérieur de la slice. Par contre, le mode de prédiction DC est toujours utilisable, même si d'autres échantillons ne sont pas disponibles. Ainsi, une valeur médiane de 128 est attribuée aux échantillons manquants (codés sur 8 bits) lors de cette prédiction.

Notons par exemple que le mode de prédiction vertical utilise les quatre échantillons de la dernière ligne du bloc supérieur, notés A-D, afin de construire le bloc de prédiction. Ainsi, la première colonne du bloc prédit (échantillons a, e, i et m) ne comporte que des échantillons identiques à A. De même pour la deuxième colonne (échantillons b, f, j et n) qui ne contient que des échantillons identiques à B.

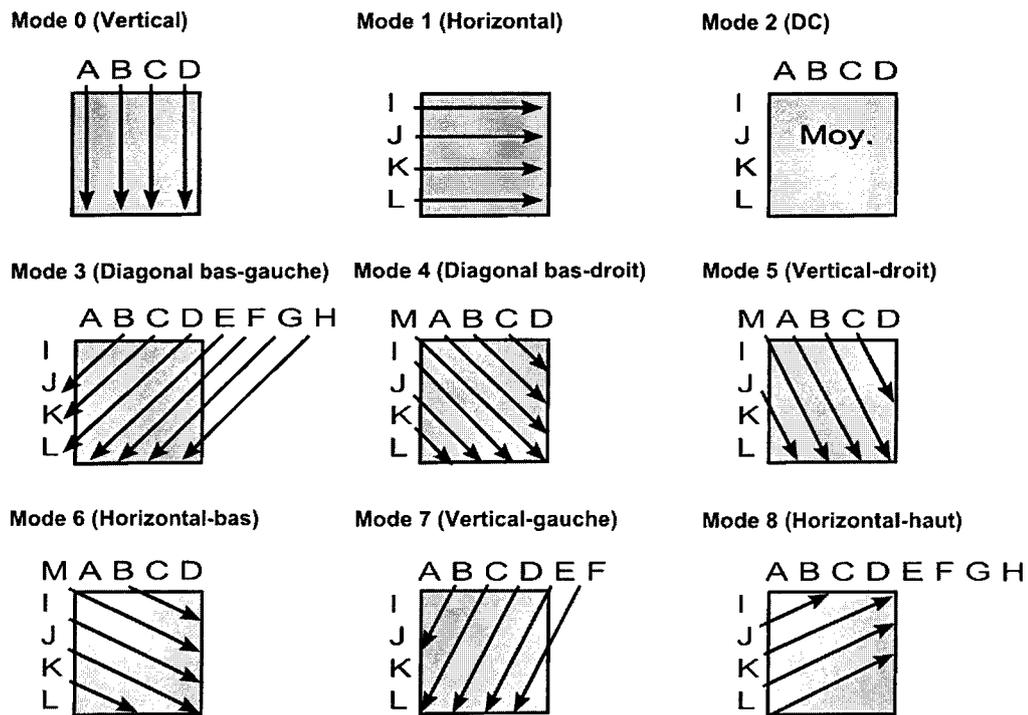


Figure 1.10 Les 9 modes de prédiction du codage INTRA_4x4

Le codage d'un macrobloc de type INTRA_16x16 se réalise de la façon suivante. Quatre modes de prédiction sont utilisés : vertical, horizontal, DC et Plan. Le fonctionnement est similaire à la prédiction INTRA_4X4 à l'exception que la prédiction se fait sur l'ensemble du macrobloc et que les échantillons voisins sont au nombre de seize en haut du macrobloc et à sa gauche.

Le mode Intra PCM permet de transmettre un macrobloc à l'identique. Ainsi, les étapes de prédiction, transformée entière et quantification ne sont pas utilisées. Chaque échantillon de luminance ou de chrominance est transmis sans perte d'information (par défaut un échantillon est codé sur 8 bits).

Ainsi, afin de déterminer le type de codage d'un macrobloc en luminance, il va être nécessaire de calculer le coût pour le codage I_PCM, le codage INTRA_16X16 (4 modes de prédiction) et le codage INTRA_4X4 (9 modes de prédiction pour chacun des 16 blocs de 4x4 pixels du macrobloc).

Le codage de la chrominance d'un macrobloc de type intra (hors Intra PCM) se réalise similairement au codage de la luminance d'un macrobloc de type INTRA_16x16. Chaque composante de chrominance (Cr - différence au rouge ou Cb - différence au bleu) d'un macrobloc est prédite à partir des échantillons voisins déjà codés et situés en haut et/ou à gauche. À l'exception de leur numérotation, les quatre modes de prédiction sont identiques à ceux du codage de la luminance en INTRA_6x16. Ces modes sont : DC (mode 0), horizontal

(mode 1), vertical (mode 2) et plan (mode 3). Notons que le même mode de prédiction est utilisé pour les deux composantes.

Dans un flux H.264/AVC, un drapeau appelé *constrained_intra_pred_flag* est présent dans les NALU de type PPS afin de préciser si la prédiction intra-image autorise (drapeau égal à 0) ou non (drapeau égal à 1) l'utilisation de résidus ou d'échantillons des blocs voisins issus de la prédiction inter-image. Cette restriction peut s'avérer être très utile dans le cas des slices de type P ou B, comprenant des macroblocs de type intra et inter.

1.2.2.2 Codage des macroblocs de type P

Il existe plusieurs types de macroblocs P. Un type correspond à un partitionnement spécifique du macrobloc qui est utilisé lors de la prédiction et de la compensation de mouvement (Figure 1.11). La norme autorise la gestion des blocs de luminance de taille 16x16, 16x8, 8x16 et 8x8 pixels. Dans le cas des blocs 8x8, un élément syntaxique est ajouté afin de gérer un partitionnement additionnel et les échantillons de chrominance.

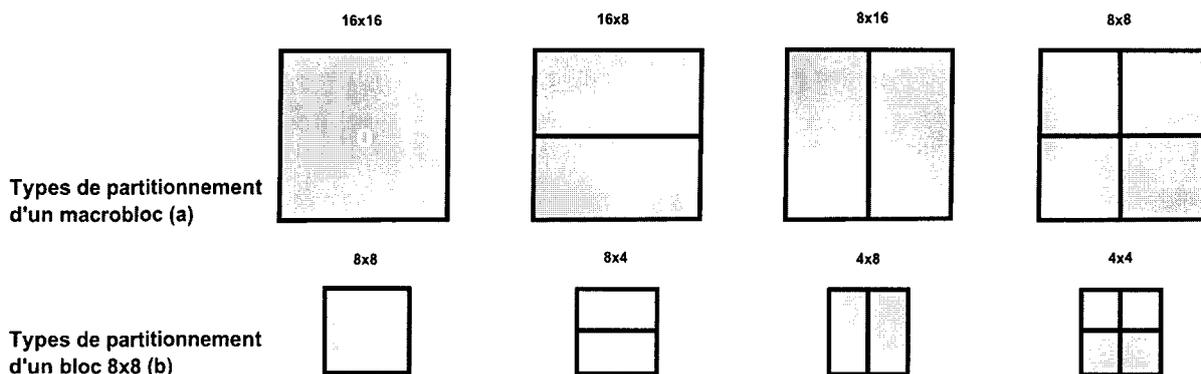


Figure 1.11 Partitionnement d'un macrobloc pour la prédiction/compensation de mouvement (a) - Partitionnement d'un bloc 8x8 (b)

Pour chaque bloc de luminance codé, le bloc de prédiction est obtenu en translatant un bloc de l'image de référence grâce au vecteur de mouvement calculé [WIE03a]. Un maximum de seize vecteurs de mouvement peut être utilisé pour un unique macrobloc P si celui-ci est partitionné en quatre blocs de 8x8 échantillons, eux-mêmes partitionnés en quatre blocs de 4x4 échantillons.

Dans un macrobloc codé par prédiction inter-image, chaque partition est prédite à partir d'une zone de pixels de taille identique dans l'image de référence. Le vecteur de mouvement caractérisant le déplacement nécessaire entre ces deux zones, a une précision de l'ordre du quart de pixel pour la luminance et du huitième de pixel pour la chrominance. Comme les échantillons de luminance et de chrominance n'existent pas à cette position intermédiaire dans l'image de référence, il est nécessaire de les créer par interpolation des

pixels avoisinants. La Figure 1.12 (a) illustre une zone de 6x6 pixels, où les échantillons de luminance A-T sont mis en évidence.

Les valeurs de prédiction à la moitié des positions entières des échantillons sont obtenues en appliquant un filtre RIF à six coefficients et à une dimension, horizontalement et verticalement. Par exemple dans la Figure 1.12 (b), l'échantillon *c* est obtenu à partir du filtrage des échantillons *M*, *N*, *C*, *I*, *O* et *P* (horizontalement). Aussi, l'échantillon *i* est obtenu à partir du filtrage des échantillons *A-F* (verticalement). De manière similaire, dans la Figure 1.12 (c), l'échantillon *m* est obtenu à partir du filtrage des échantillons déjà interpolés *g-l*.

Les valeurs de prédiction au quart de la position entière d'un échantillon sont générées en moyennant les échantillons voisins à position entière et à position intermédiaire (demi-pixel). Ainsi, dans la Figure 1.12 (d), l'échantillon *aa* est obtenu par moyennage des échantillons *C* et *c* (horizontalement), l'échantillon *bb* est obtenu par moyennage des échantillons *C* et *i* (verticalement) et l'échantillon *cc* est obtenu par moyennage des échantillons *C* et *m* (diagonalement).

Il faut noter que les vecteurs de mouvement ont la possibilité de dépasser les frontières de l'image. Dans ce cas, l'image de référence est agrandie en y répétant les échantillons frontaliers.

La norme prévoit la possibilité d'utiliser plusieurs images lors de la prédiction et de la compensation de mouvement. Il est alors indispensable d'implémenter à l'encodage et au décodage un *buffer* permettant de stocker ces images. Le décodeur recrée à l'identique le *buffer* de l'encodeur suivant les spécifications du flux binaire. À moins que la taille de ce *buffer* soit fixée à une image, l'index de l'image de référence doit être signalé. Cet index est transmis pour chaque mouvement compensé d'un bloc de luminance 16x16, 16x8, 8x16 ou 8x8. Une compensation de mouvement pour des régions plus petites que 8x8 utilise le même index de référence pour tous les blocs internes à la région 8x8.

En plus des précédents types de macroblocs, un macrobloc *P* peut être également codé par le type *P_SKIP*. Dans le flux vidéo, un macrobloc de type *P_SKIP* est codé en quelques bits. Il ne contient pas de résidu fréquentiel quantifié, ni de vecteur de mouvement ou d'un index d'image de référence. Le signal de prédiction reconstruit est obtenu similairement au signal de prédiction d'un macrobloc de type *P_16x16* qui fait référence à une image qui est stockée à l'index 0 du *buffer* image. Le vecteur mouvement utilisé pour reconstruire le macrobloc de type *P_SKIP* est construit à partir des vecteurs mouvements (pouvant être nuls) des macroblocs voisins. L'intérêt de ce type de codage est de représenter avec quelques bits de larges régions sans changement ou avec un mouvement constant.

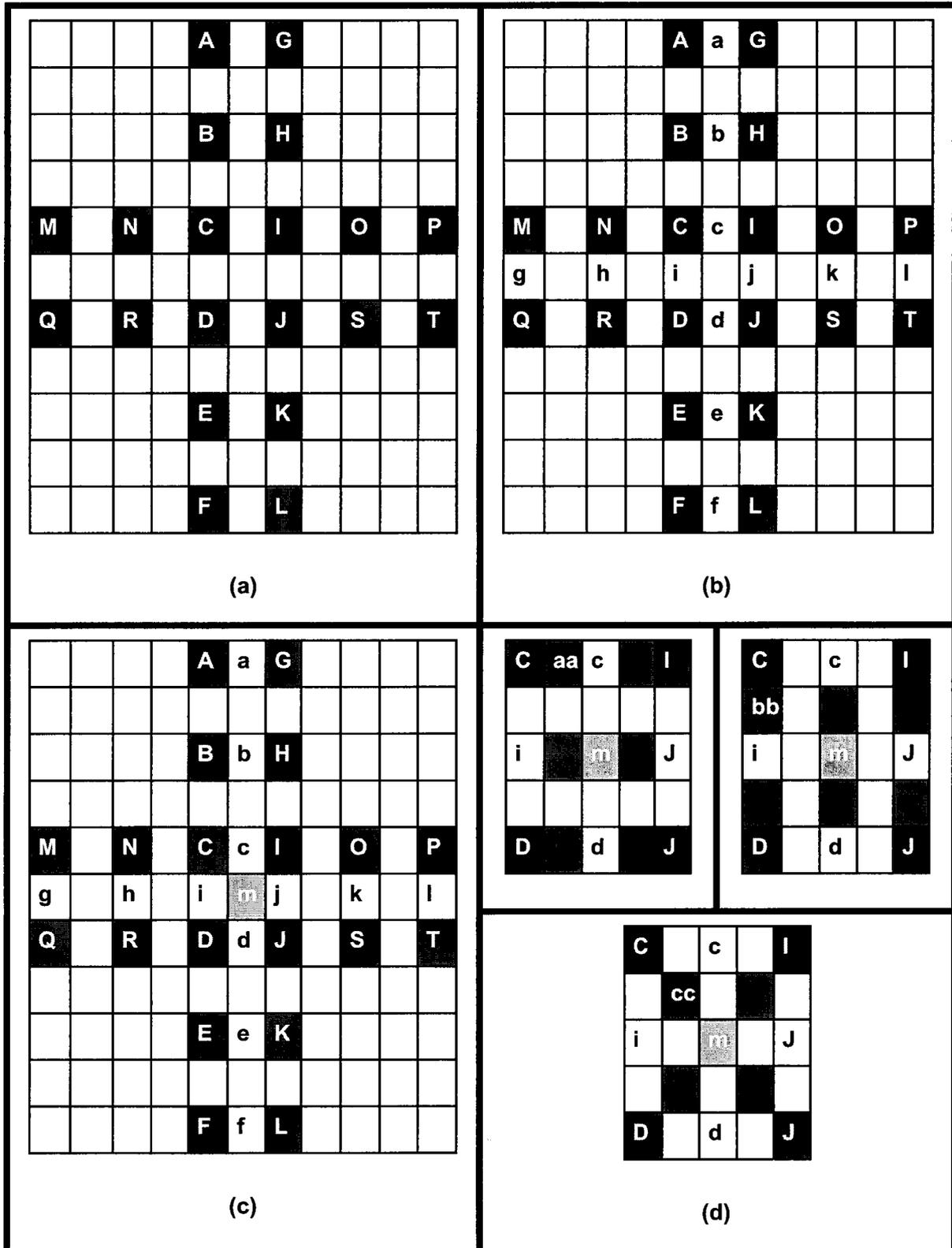


Figure 1.12 Interpolation des échantillons de luminance jusqu'au quart de pixel

Dans un flux H.264/AVC, l'élément syntaxique *mb_skip_run* précise le nombre de macroblocs consécutifs de type *P_SKIP* ou de type *B_SKIP*, si la slice courante est respectivement de type P ou de type B.

1.2.2.3 Codage des macroblocs de type B

La différence majeure entre les macroblocs de type B et P est que les macroblocs de type B peuvent utiliser deux images de référence, pouvant précéder et suivre l'image en cours de codage. Ainsi, ces macroblocs peuvent utiliser une moyenne pondérée de deux valeurs distinctes de *prédiction de mouvement afin de construire le signal de prédiction*.

Les macroblocs de type B utilisent deux listes distinctes d'images de référence. Elles sont respectivement notées *list 0* et *list 1*. Le partitionnement des macroblocs des slices B est similaire à celui des slices de type P. Pour chaque partitionnement de type 16x16, 16x8, 8x16 et 8x8, un type de codage est choisi parmi :

- *L0*, où uniquement les images de la liste 0 peuvent être utilisées comme référence.
- *L1*, où uniquement les images de la liste 1 peuvent être utilisées comme référence.
- *Bi* ou *bi-predictive*, où les images de la liste 0 et de la liste 1 peuvent être utilisées comme référence.
- *Direct*, réservé aux partitionnements de type 16x16 ou 8x8 ; il est déduit des éléments syntaxiques précédemment transmis et peut être soit *list 0*, *list 1* ou *bi-predictive*. Une partition 8x8 d'un macrobloc B peut être codée également en mode direct.
- *B_Skip*, réservé au partitionnement de type 16x16, similaire au mode *P_Skip* dans les slices de type P. Le codage du vecteur de mouvement est similaire à celui des macroblocs de type *P_Skip*, avec quelques modifications car les blocs voisins peuvent être codés avec des modes de prédiction différents.

La Figure 1.13 illustre un exemple de dépendance entre les slices d'un groupe de 16 images. Les slices annotées *b3* sont des images bidirectionnelles, elles sont codées en utilisant comme référence des images intra, des images prédites ou d'autres images bidirectionnelles. Les slices annotées *B2* sont des images bidirectionnelles utilisées comme référence par les slices *b3*. Elles utilisent comme référence des images intra, des images prédites ou d'autres images bidirectionnelles. Les slices annotées *B1* sont des images bidirectionnelles utilisées comme référence par les slices *B2* et *b3*. Elles utilisent comme référence des images intra et des images prédites. Les slices *P0* sont des images prédites à partir des images intra.

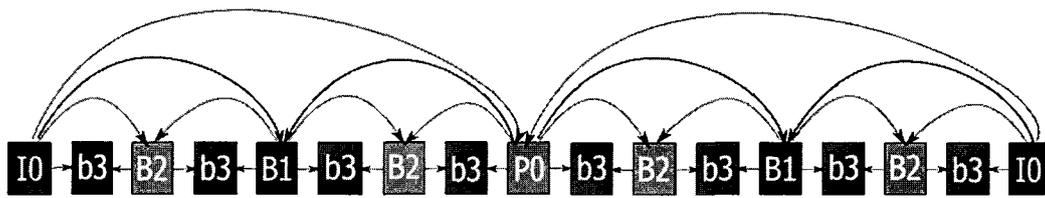


Figure 1.13 Exemple de dépendance entre tranches sur un GOP de 16 images

1.2.3 Traitement des coefficients résiduels

1.2.3.1 Transformées

La norme H.264/AVC utilise une transformée des coefficients résiduels afin de faciliter le codage des macroblocs ou blocs d'échantillons [MAL03]. Il ne s'agit pas de la classique transformée en cosinus discrète ou *DCT* mais d'une transformée entière aux propriétés similaires. Cette transformée s'applique sur des blocs de quatre par quatre échantillons. Ainsi, chaque macrobloc ou sous-macrobloc est décomposé en blocs de quatre par quatre échantillons.

Le choix de la taille de cette transformée peut se justifier en trois points [WIE03a] :

- Une conséquence de l'amélioration de la prédiction intra et inter-image est que le bloc résiduel est désormais moins corrélé du bloc du domaine spatial. Ainsi, la décorrélation issue de la transformée est moins importante. La suppression de la corrélation peut être effectuée par une transformée de petite taille ou de grande taille (transformée 8x8 dans les normes précédentes).
- Quatre transformées de petite taille (4x4) sont moins coûteuses en temps de calcul qu'une grande transformée (8x8) et nécessitent une précision de calcul moins importante.
- Visuellement, le bruit autour des contours est moins flagrant et minimise les artefacts de types *mosquito noise* ou scintillement des contours.

Le résultat de la transformée entière d'un bloc de 4x4 pixels est présenté en Figure 1.14, nous remarquerons la présence d'un coefficient basse fréquence (*DC*) ou niveau lumineux moyen et de quinze coefficients haute fréquence (*AC*).

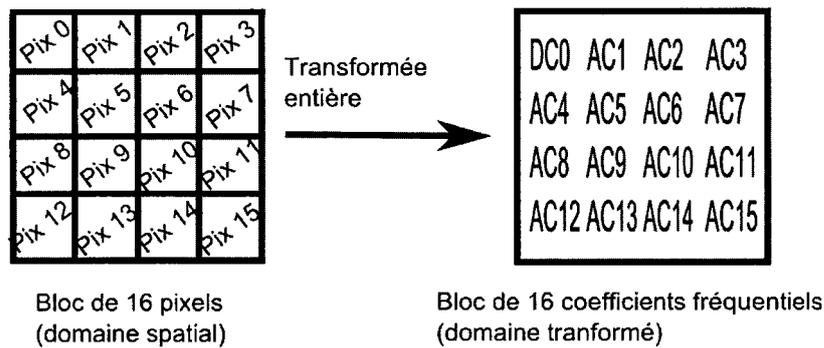


Figure 1.14 Transformée entière d'un bloc de 4x4 échantillons

La matrice de cette transformée entière est présentée ci-dessous :

$$T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

Comme l'ensemble des calculs se réalise sur des entiers, la transformée et son inverse sont réversibles et permettent de préserver la précision des résultats sans se soucier des arrondis.

Dans le cas des macroblocs utilisant le mode de prédiction Intra_16x16 pour la luminance, une transformée d'Hadamard est utilisée sur l'ensemble des seize coefficients DC des blocs 4x4. La Figure 1.15 illustre les étapes d'encodage d'un macrobloc de type Intra_16x16. Un macrobloc peut se décomposer en seize blocs de 4 par 4 pixels. Chaque bloc subit une transformée entière (a). À titre indicatif, le résultat de la transformée entière est représenté en (b) pour le bloc 0. Les coefficients vont alors être triés afin de créer un bloc regroupant les 16 coefficients DC de l'ensemble des blocs (c) et seize blocs de quinze coefficients AC (d). Les coefficients DC vont quant à eux subir une transformée d'Hadamard 4x4 afin d'exploiter la corrélation entre les coefficients DC.

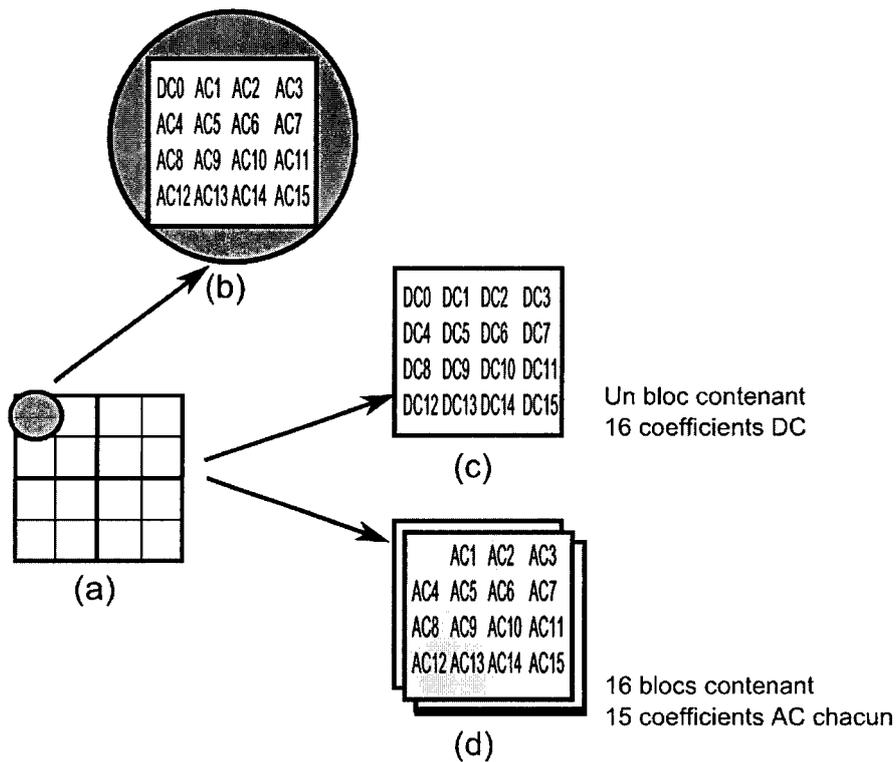


Figure 1.15 Étapes d'encodage d'un macrobloc de type Intra_16x16

Le codage de la chrominance se réalise de façon similaire. Mais il faut noter que les échantillons de chrominance sont moins nombreux après le sous-échantillonnage 4:2:0 (voire 4:2:2), ainsi les coefficients DC de chrominance subiront une transformée d'Hadamard de taille 2x2.

1.2.3.2 Balayage en zig-zag

Le balayage en zig-zag est utilisé afin d'établir un ordre de traitement des coefficients résiduels par rang fréquentiel approximativement croissant. Ainsi, les basses fréquences horizontales, diagonales et verticales seront traitées avant les hautes fréquences. Il permet de coder en dernier les hautes fréquences, aux amplitudes les moins élevées, qui ont plus de chances d'être annulées par la quantification. La Figure 1.16 illustre le balayage en zig-zag sur un bloc (de type INTRA_4x4) de 4x4 coefficients résiduels fréquentiels.

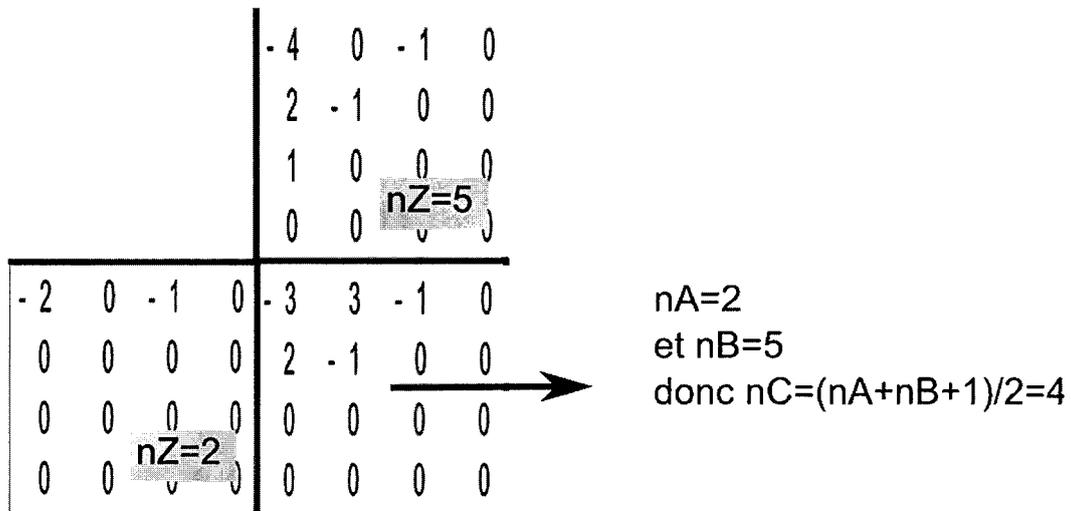


Figure 1.18 Prédiction du nombre de coefficients non nuls

Le Tableau 1.1 précise la table utilisée pour coder le `coeff_token` en fonction de la valeur du paramètre nC . Ainsi, la *Table 1* associe un code de petite longueur aux valeurs faibles de `TotalCoeffs` (un code long pour les valeurs élevées de `TotalCoeffs`). À l'inverse, la *Table 3* est plus adaptée aux valeurs de `TotalCoeffs` élevées. La table à longueur fixe quant à elle associe un code de 6 bits pour chaque valeur de `TotalCoeffs`. Ces tables sont disponibles en annexe A.3.

nC	Table utilisée
0, 1	Table 1 (codes à longueur variable)
2, 3	Table 2 (codes à longueur variable)
4, 5, 6, 7	Table 3 (codes à longueur variable)
8 et plus	Table 4 (codes à longueur fixe)

Tableau 1.1 Choix de la table utilisée pour coder le `coeff_token`

Le codage spécifique des trailing ones vient du constat que les coefficients résiduels non nuls les plus hauts en fréquence ont généralement de très petites amplitudes après l'étape de quantification. Ainsi, il est courant que ces coefficients aient une amplitude égale à « 1 » avec un signe positif ou négatif. Pour chaque trailing one, uniquement un bit est utilisé pour coder le signe. Le zéro représente un nombre positif et le un représente un nombre négatif.

Le signe et l'amplitude des autres coefficients sont codés dans l'ordre du zig-zag inverse en partant du coefficient le plus élevé en fréquence jusqu'au coefficient DC. Le choix du code à longueur variable (VLC) pour coder chaque niveau successif se fait par adaptation contextuelle en fonction de l'amplitude du niveau précédemment codé.

Ensuite, l'étape suivante consiste à coder le nombre total de zéros (successifs ou non) précédents le dernier coefficient. Nous ne parlons pas ici des zéros suivants le dernier coefficient non nul.

Enfin, la dernière étape sert à coder le nombre de zéros précédant chaque coefficient non nul (`run_before`) dans le sens inverse du balayage en zig-zag.

Notons qu'un autre type de codage entropique est disponible dans H.264/AVC. Il s'agit du CABAC ou codage arithmétique binaire adaptatif fondé sur le contexte (*Context-based Adaptive Binary Arithmetic Coding*). Nous ne l'avons pas utilisé dans les travaux du projet TOSCANE car il n'est pas disponible pour l'ensemble des profils de la norme, contrairement au codage CAVLC. Il donne des performances de codage intéressantes et permet notamment de coder l'ensemble des éléments syntaxiques (drapeaux, modes de prédiction, vecteurs mouvement...) avec les coefficients résiduels [MAR03].

Le CABAC fournit un schéma de codage amélioré avec des rapports bit/symbole optima. Le rôle principal du CABAC est de dépasser les performances du CAVLC en se rapprochant plus rapidement de l'entropie. Cela est permis notamment par sa dimension arithmétique qui permet de passer en dessous de la limite d'un bit par symbole du CAVLC. Le codage arithmétique signifie que le jeu de probabilités est transposé sur un intervalle allant de 0 à 1. Les probabilités sont représentées par les valeurs de cet intervalle. La dimension contexte signifie que le jeu de probabilité est mis à jour à chaque codage d'élément binaire. Le CABAC se décompose en quatre étapes :

- La première étape est la binarisation. Elle permet de simplifier l'étape de mise à jour du contexte (étape 4) en passant de sources de taille N à des sources de taille 2. Lors de cette étape, on binarise tout symbole dont la valeur est différente de 0 ou de 1. Par exemple, ces symboles peuvent être des coefficients résiduels transformés et quantifiés, des vecteurs de mouvement ou encore des partitions de macrobloc. Pour chaque symbole, une table de binarisation adaptée est choisie.
- La deuxième étape s'applique à chaque élément binaire. Un modèle de contexte est choisi pour le codage. Il dépend du type d'informations à coder et d'autres paramètres comme les valeurs des blocs voisins dans le cas des vecteurs de mouvement. On recense 459 contextes disponibles qui sont initialisés au début de chaque slice en fonction du paramètre de quantification. Chaque contexte contient alors la probabilité que le bit courant est égal à 0 ou 1.
- La troisième étape correspond à l'encodage. Le contexte initialisé est utilisé pour coder l'élément binaire de manière arithmétique.

- La dernière étape correspond à la mise à jour du contexte. Si le symbole 1 est transmis, sa probabilité d'apparition est augmentée dans le contexte sélectionné.

1.2.4 Évaluation des performances de H.264/AVC

Dans cette section, nous proposons d'évaluer les performances de H.264/AVC en termes de réduction de débit et de qualité d'image par une comparaison avec la norme MPEG-2.

Dans la littérature, de nombreux articles sont consacrés à la comparaison des performances de H.264/AVC aux normes de compression vidéo précédentes. Dans [WIE03b], le H.264/AVC est comparé au MPEG-2 à travers trois types d'applications : le streaming vidéo, la visio-conférence et la diffusion vidéo de qualité supérieure (sur supports de type DVD ou Blu-Ray, télévision numérique...). Nous reprendrons les résultats de cette dernière catégorie car ils sont plus proches de nos travaux de recherche avec des débits supérieurs à 3 Mb/s. Deux groupes de séquences ont été utilisés :

- Quatre séquences en définition standard (720x576 pixels, 25 Hz) en balayage entrelacé.
- Quatre séquences en haute définition 720p60 (1280x576 pixels, 60 Hz) en balayage progressif.

Les paramètres principaux d'encodage des flux en MPEG-2 et H.264/AVC sont les suivants :

- Utilisation de deux images bidirectionnelles entre chaque image clef.
- Utilisation d'un contrôle de débit de type Lagrangien.
- Estimation de mouvement par recherche complète sur ± 32 pixels.

Les flux H.264/AVC utilisent le codage entropique CABAC.

Dans ces expérimentations, la métrique PSNR (cf 3.4.2.1) est utilisée afin de déterminer la distorsion entre les échantillons de l'image d'origine et ceux de l'image encodée puis décodée. La Figure 1.19 (gauche) illustre la distorsion (dB) due à ces deux types d'encodages en fonction du débit vidéo pour trois séquences vidéo (deux séquences au format SD et une au format HD). La Figure 1.19 (droite) illustre la réduction de débit supplémentaire apportée par le H.264/AVC par rapport au MPEG-2 en fonction du PSNR (dB).

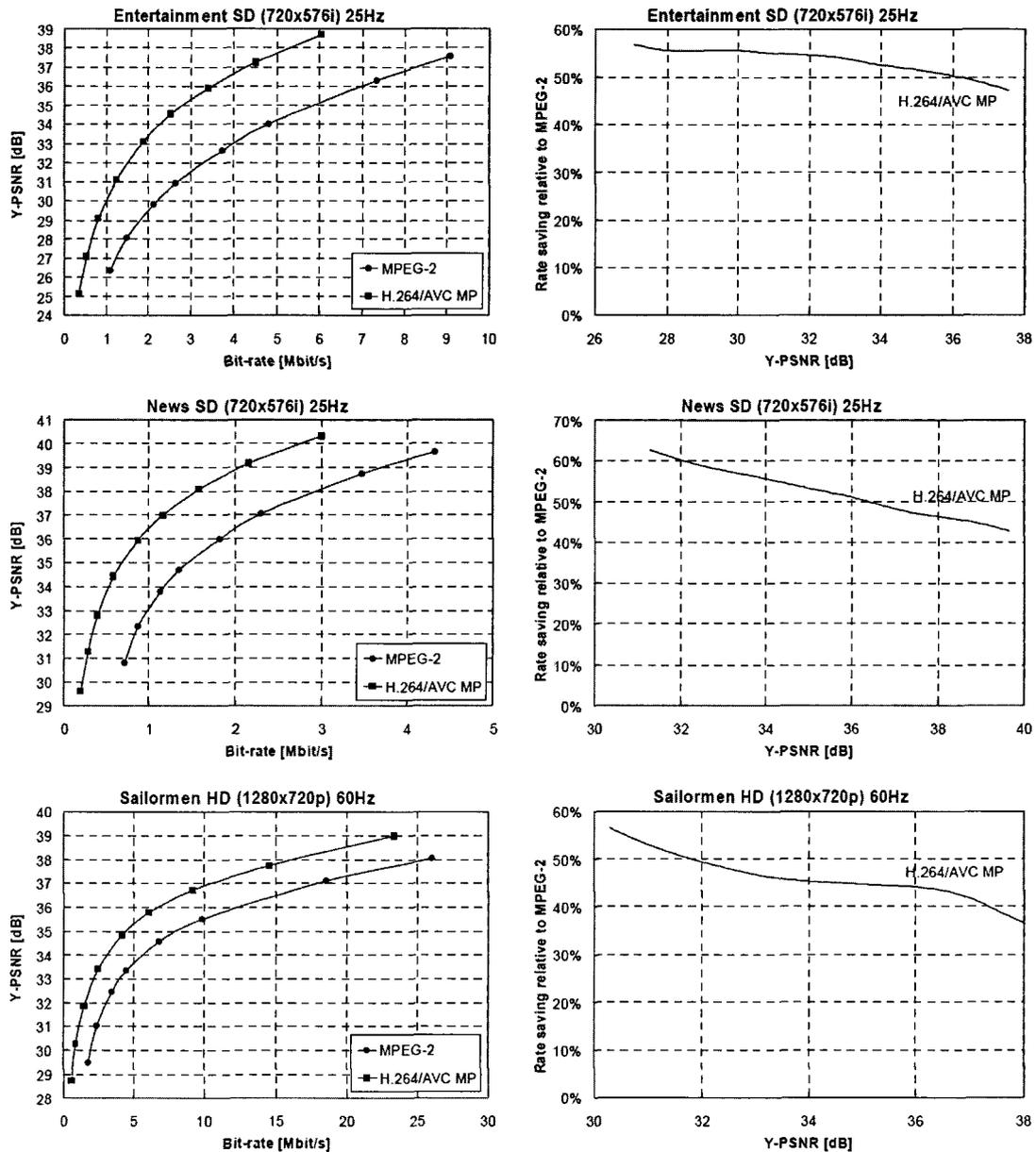


Figure 1.19 PSNR et réduction de débit apportée par le MPEG-2 et le H.264/AVC [WIE03b]

Pour les séquences vidéo en définition standard, les auteurs constatent que le H.264/AVC permet une réduction de débit de l'ordre de 45-65% par rapport à un encodage MPEG-2, à PSNR équivalent. L'encodage H.264/AVC des séquences au format 720p60 permet une réduction de débit de l'ordre de 25-45% par rapport à un encodage MPEG-2, à PSNR équivalent. Sur l'ensemble des séquences, les auteurs constatent une réduction de débit moyenne de l'ordre de 45% (à PSNR équivalent). À titre indicatif, quelques expérimentations supplémentaires sont disponibles dans [RIC10] ou en annexe A.3 en utilisant le codage entropique CAVLC pour l'encodage H.264/AVC.

L'efficacité de codage de H.264/AVC est réellement supérieure à celle de la norme MPEG-2. Cette efficacité s'explique notamment par l'amélioration de la prédiction inter-image, ainsi qu'à l'apparition de la prédiction intra-image dans le domaine spatial. En annexe A.4, quelques expérimentations ont été réalisées afin d'évaluer l'influence des modes de prédiction (en particulier le mode DC) sur l'encodage des images intra.

1.3 L'extension échelonnable : H.264 SVC

Le principe de l'*échelonnabilité* repose sur l'idée de créer un unique flux binaire contenant plusieurs versions de la source vidéo, tout en évitant de coder les informations redondantes. Nous trouvons dans ce flux vidéo une couche de base et une ou plusieurs couches d'amélioration. La couche de base représente la version à la plus basse qualité ou définition de la source vidéo. Le décodage des couches d'amélioration requiert des informations présentes dans la couche de base, voire d'une couche inférieure à la couche d'amélioration courante. Une couche d'amélioration a vocation d'améliorer le rendu de l'image décodée en augmentant sa définition par une échelonnabilité spatiale, le nombre d'images par une échelonnabilité temporelle ou la qualité par une échelonnabilité *SNR*.

L'échelonnabilité n'est pas une nouveauté que nous devons à la norme H.264/AVC. Depuis des années, d'autres normes de compression vidéo ont proposé des solutions d'échelonnabilité mais le succès industriel et commercial n'a pas été au rendez-vous (MPEG-2 par exemple permet l'échelonnabilité avec les profils *SNR*, *Spatial* et *High*). Désormais, la multiplication des terminaux en tout genre capables de décoder de la vidéo crée un marché avec de réels besoins d'échelonnabilité. L'annexe G de H.264/AVC [ITU09], communément appelée H.264 SVC, arrive donc dans une conjoncture économique favorable qui peut faciliter son essor industriel. D'un point de vue plus pessimiste, dans le cas où son implantation n'est pas au rendez-vous, nous pouvons penser que les algorithmes développés dans le cadre de H.264 SVC serviront aux autres normes à venir, orientées également sur un découpage de l'image en macroblocs.

1.3.1 Organisation d'un flux SVC

1.3.2 Les différents types d'échelonnabilité dans H.264 SVC

Cette section décrit les trois types d'échelonnabilité présents dans l'annexe G de H.264 [SCH07]. En particulier, l'échelonnabilité spatiale sera utilisée dans le chapitre 3 dans une solution de transcodage de la couche d'amélioration haute.

1.3.2.1 L'échelonnabilité temporelle

Dans ce type d'échelonnabilité, la couche de base est codée avec un nombre d'images par seconde relativement faible par rapport à la séquence source. Le décodage supplémentaire des couches d'amélioration va permettre d'augmenter ce nombre d'images. L'annexe G de H.264/AVC n'apporte pas d'innovation algorithmique en ce qui concerne l'échelonnabilité temporelle. Cette échelonnabilité est présente nativement dans H.264/AVC dans les profils Main et High par les algorithmes de codage des images prédites et/ou bidirectionnelles.

La norme MPEG-2 est limitée à trois niveaux d'échelonnabilité pour la simple raison que les images sont retirables par type d'image, conformément à leur dépendance. Ainsi, un GOP de 24 images se présentant sous la forme **lbbbPbbbPbbbPbbbPbbbPbbb**, peut se représenter sous trois niveaux d'échelonnabilité temporelle :

- Nous conservons uniquement les images intra : affichage à 1fps
- Nous conservons les images intra et les images prédites : affichage à 6fps (GOP **IPPPPP**)
- GOP complet : affichage à 24ips

H.264/AVC innove dans le sens où chaque image peut être utilisée comme référence lors de la compensation de mouvement. Ceci est désormais possible avec l'apparition d'images bidirectionnelles utilisées comme référence par d'autres images bidirectionnelles. Nous parlons alors de GOP en structure hiérarchique ou pyramidale.

La Figure 1.20 illustre un exemple d'une séquence vidéo codée avec un GOP de 16 images. Au niveau le plus élevé d'échelonnabilité temporelle (niveau 3), toutes les images sont présentes. Au deuxième niveau d'échelonnabilité temporelle, toutes les images bidirectionnelles qui ne sont pas utilisées comme référence sont supprimées (il y a deux fois moins d'images à afficher). Au premier niveau d'échelonnabilité temporelle, les images bidirectionnelles utilisées comme référence par les images précédemment supprimées au niveau 2, sont également supprimées (il ne reste plus qu'un quart des images à afficher). La procédure peut être renouvelée par la suppression des images bidirectionnelles restantes (étape 0).

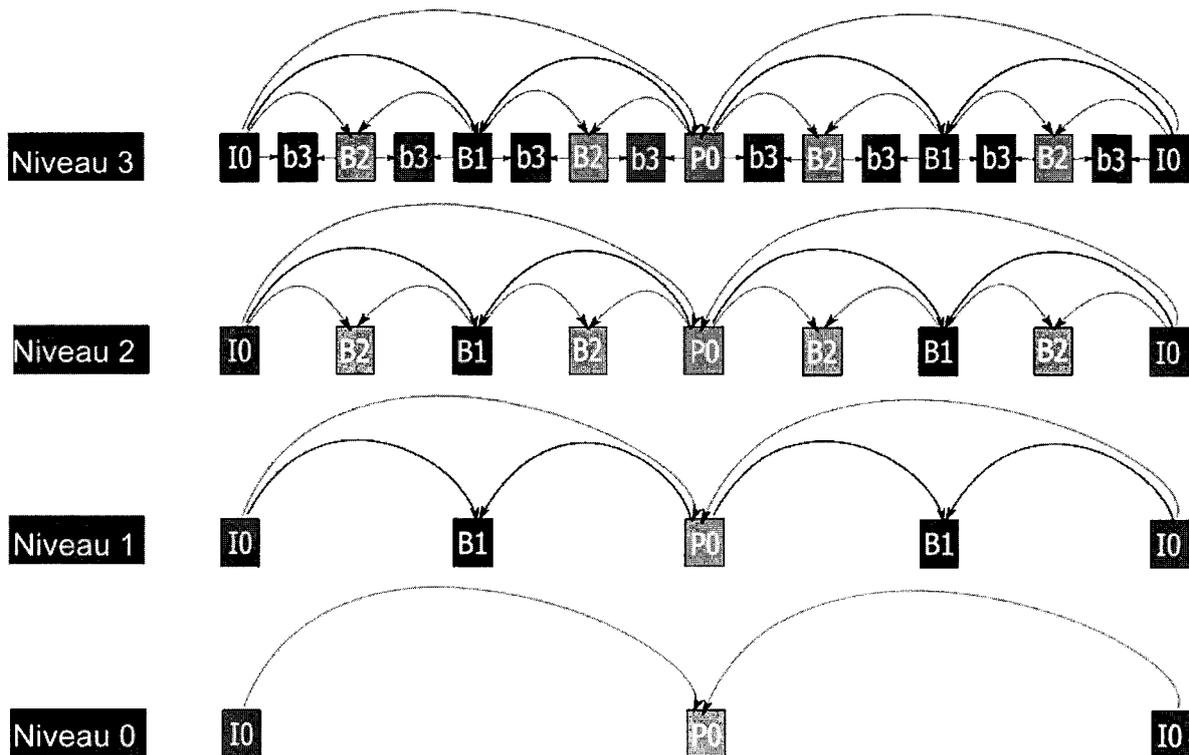


Figure 1.20 Exemple d'un GOP de 16 images avec 4 niveaux d'échelonnabilité temporelle

L'amélioration de H.264 SVC repose en réalité sur la possibilité d'utiliser cette échelonnabilité temporelle sans avoir besoin d'utiliser un profil spécifique, contrairement à H.264/AVC.

1.3.2.2 L'échelonnabilité spatiale

L'échelonnabilité spatiale est certainement le type d'échelonnabilité qui a le plus d'applications dans l'industrie. Les terminaux susceptibles de l'utiliser vont du téléphone portable à la télévision haute définition, en passant par les ordinateurs et les consoles de jeux. Uniquement dans le cas de la téléphonie mobile, l'opérateur téléphonique se retrouve avec une multitude de terminaux ou téléphones portables aux caractéristiques différentes. Ainsi, les définitions supportées par les écrans des téléphones et la puissance de calcul disponible pour le décodage de l'image diffèrent d'un modèle à un autre. Il est alors inévitable de disposer de la même source vidéo sous différents formats d'image.

Dans ce type d'échelonnabilité, la couche de base correspondra inévitablement au sous-échantillonnage spatial le plus bas de la source vidéo. L'ajout d'une couche d'amélioration a pour incidence d'augmenter la définition de la vidéo. Le schéma ci-dessous (Figure 1.21) illustre l'encodage d'une source vidéo en trois niveaux d'échelonnabilité spatiale. La couche la plus haute, ou couche d'amélioration 2, possède la même définition

spatiale que la source vidéo ; notons que ceci n'est pas imposé par la norme et dépend uniquement du scénario utilisé. La couche d'amélioration 1 possède une définition spatiale inférieure à celle de la source vidéo mais supérieure à celle de la couche de base.

L'encodeur commence par encoder la couche de base comme spécifié dans H.264/AVC. Pour chaque bloc, l'erreur est calculée entre le bloc sous-échantillonné et la prédiction intra ou inter-image qui est calculée à partir des échantillons déjà codés. Cette opération aboutit à un résidu dans le domaine spatial qui est ensuite transformé, quantifié et encodé entropiquement. Afin d'éviter toute propagation d'erreur lors de la prédiction intra ou inter-image, l'architecture de l'encodeur est en boucle fermée. Ainsi le résidu transformé puis quantifié, subit les opérations inverses afin d'obtenir le nouveau résidu dans le domaine spatial. Ceci permet de mettre à jour le buffer image qui est utilisé lors de la prédiction intra ou inter-image. De plus, les blocs de notre buffer image de la couche de base, les résidus ou éventuellement les vecteurs mouvement peuvent également être utilisés par la couche d'amélioration 1.

L'encodage de la couche d'amélioration 1 débute par la récupération de l'image courante dans le buffer image de la couche de base. Nous noterons ici qu'il est nécessaire que la couche de base soit déjà encodée, ou du moins que l'encodage de la couche de base ait terminé le codage du bloc courant. Alors, l'image de la couche de base va être sur-échantillonnée à la même définition spatiale que la source vidéo de référence de la couche d'amélioration 1. Ainsi, cette image pourra servir de référence à la prédiction inter-couche. Il s'agit d'une image de référence intéressante malgré les distorsions subies au sous-échantillonnage, au sur-échantillonnage et au codage. Pour chaque bloc, l'erreur est calculée entre le bloc de la source vidéo de référence de la couche d'amélioration 1, la prédiction inter-couche et/ou la prédiction intra ou inter-image.

L'encodage des couches d'amélioration supérieures se réalise similairement, en utilisant comme référence les blocs de la couche inférieure sur-échantillonnés spatialement.

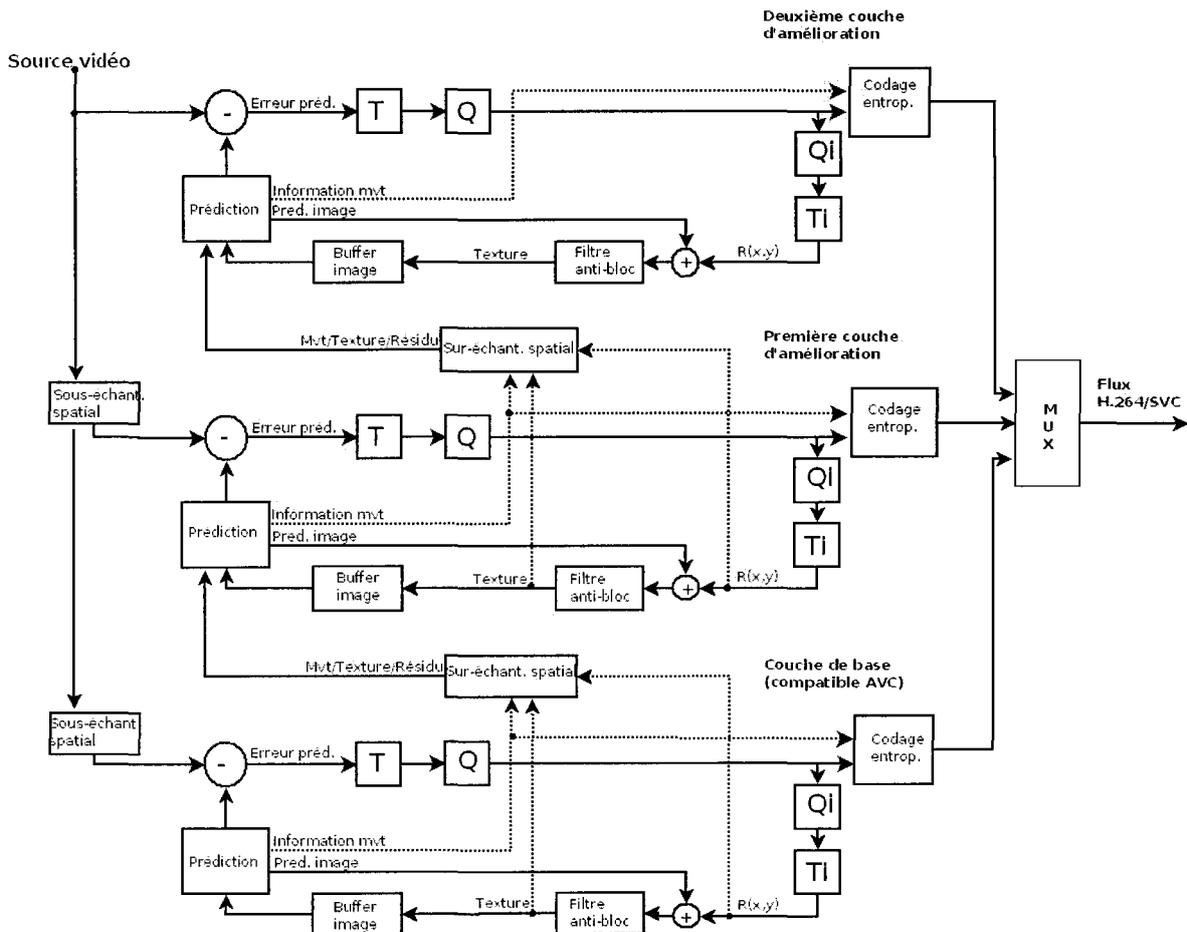


Figure 1.21 Schéma de l'encodeur H.264 SVC en échelonnabilité spatiale

Contrairement aux normes précédentes supportant l'échelonnabilité spatiale, H.264 SVC ne se contente pas uniquement de réaliser un sur-échantillonnage de la couche d'amélioration inférieure. Des modes de prédiction [DEW07] ont été créés afin d'améliorer l'efficacité de la compression de l'échelonnabilité spatiale.

Ainsi, dans une couche d'amélioration tous les modes habituels de prédiction sont utilisables : prédiction intra à partir des blocs voisins de l'image courante dans cette couche d'amélioration, prédiction inter à partir des images déjà codées et décodées dans cette couche d'amélioration.

Comme décrit précédemment, il est également possible d'utiliser la couche inférieure comme couche de référence en la sur-échantillonnant. Dans le cas d'un sous-échantillonnage dyadique, un macrobloc (16 par 16 pixels) de la couche d'amélioration courante a pour équivalent un bloc de 8 par 8 pixels dans la couche inférieure. Les options de prédiction disponibles pour coder ce macrobloc sont décrites dans les paragraphes suivants.

H.264 SVC introduit un nouveau type de macrobloc pour les couches d'amélioration en échelonnabilité spatiale [SCH07]. Ce type de macrobloc est signalé dans le flux vidéo par un élément syntaxique de type drapeau mis à 1, appelé *base mode flag*. Il est composé uniquement d'un résidu, sans modes de prédiction ou autres informations.

Le macrobloc peut être codé par prédiction de mouvement inter-couche par l'utilisation des choix de prédiction du bloc de la couche de référence (*Base Mode*). Quand le drapeau *Base Mode* est mis à zéro, le bloc de la couche d'amélioration va être codé à partir des modes de prédiction du bloc de la couche de référence. Dans le cas d'un bloc de la couche inférieure codé en inter, le bloc correspondant de la couche d'amélioration sera codé également par une prédiction inter-image, avec les mêmes indices d'images de référence, le même type de partitionnement mais sur-échantillonné et avec également les vecteurs de mouvement sur-échantillonnés de la couche inférieure. Quand le drapeau *Base Mode* est mis à 1, seul le résidu est transmis à la couche d'amélioration.

Si le drapeau *Motion Prediction* est mis à 1, la partition courante du macrobloc de la couche d'amélioration est prédite à partir de la prédiction inter-image avec les mêmes indices d'images de référence de la couche de référence mais avec un codage différentiel entre le vecteur mouvement calculé et celui sur-échantillonné de la couche inférieure (MVD).

Dans le cas où le drapeau *Base Mode* est mis à 1 et que le bloc correspondant de la couche inférieure est codé en intra, celui-ci va être sur-échantillonné par un filtre RIF à quatre coefficients afin de servir de macrobloc de prédiction. Le résidu de la couche d'amélioration est obtenu en calculant l'erreur entre le macrobloc courant et le bloc prédit.

Si le drapeau *Residual Prediction* est mis à 1, le résidu du macrobloc de la couche d'amélioration est prédit à partir du résidu du bloc (8x8) de la couche de référence. Tout d'abord, le résidu du bloc (8x8) de la couche de référence est sur-échantillonné en utilisant une interpolation bi-linéaire. Puis, l'erreur est calculée entre le résidu du macrobloc de la couche d'amélioration et le résidu sur-échantillonné de la couche de référence. Dans le flux vidéo, c'est cette erreur qui est codée à la place du résidu du macrobloc.

Des limitations sont imposées sur la prédiction inter-couche. La prédiction inter-couche intra-image est uniquement applicable à un macrobloc de la couche d'amélioration dont le bloc correspondant de la couche de référence est de type intra. On parle de *constrained inter-layer prediction*. Pour des raisons de complexité de décodage, cette limitation de la prédiction intra est indispensable aux couches servant de référence à la prédiction inter-couche. Ainsi, les macroblocs codés en intra dans la couche de référence peuvent être reconstruits directement, sans avoir à décoder des macroblocs codés par la prédiction inter-couche [SCH05].

1.3.2.3 L'échelonnabilité en qualité

Dans H.264 SVC, il existe deux types d'échelonnabilité en qualité : *Coarse Grain Scalability (CGS)* et *Medium Grain Scalability (MGS)*. L'échelonnabilité en qualité ou échelonnabilité SNR a pour objectif de fournir une même séquence vidéo sous plusieurs débits à qualité différente mais à définition spatiale identique.

L'échelonnabilité CGS peut se définir comme un cas particulier de l'échelonnabilité spatiale, où chaque couche a une définition spatiale identique. Son fonctionnement est donc très proche de celui de l'échelonnabilité spatiale avec la particularité que le paramètre de quantification va être modifié d'une couche à l'autre. Ainsi, la couche de base aura le paramètre de quantification le plus élevé, la couche d'amélioration la plus haute (*dependency identifier* le plus élevé) aura le paramètre de quantification le plus petit. La Figure 1.22 nous propose un exemple d'encodeur H.264 SVC en échelonnabilité CGS. Trois couches d'échelonnabilité en qualité y sont proposées, elles utilisent toutes un paramètre de quantification différent ($Q1 > Q2 > Q3$).

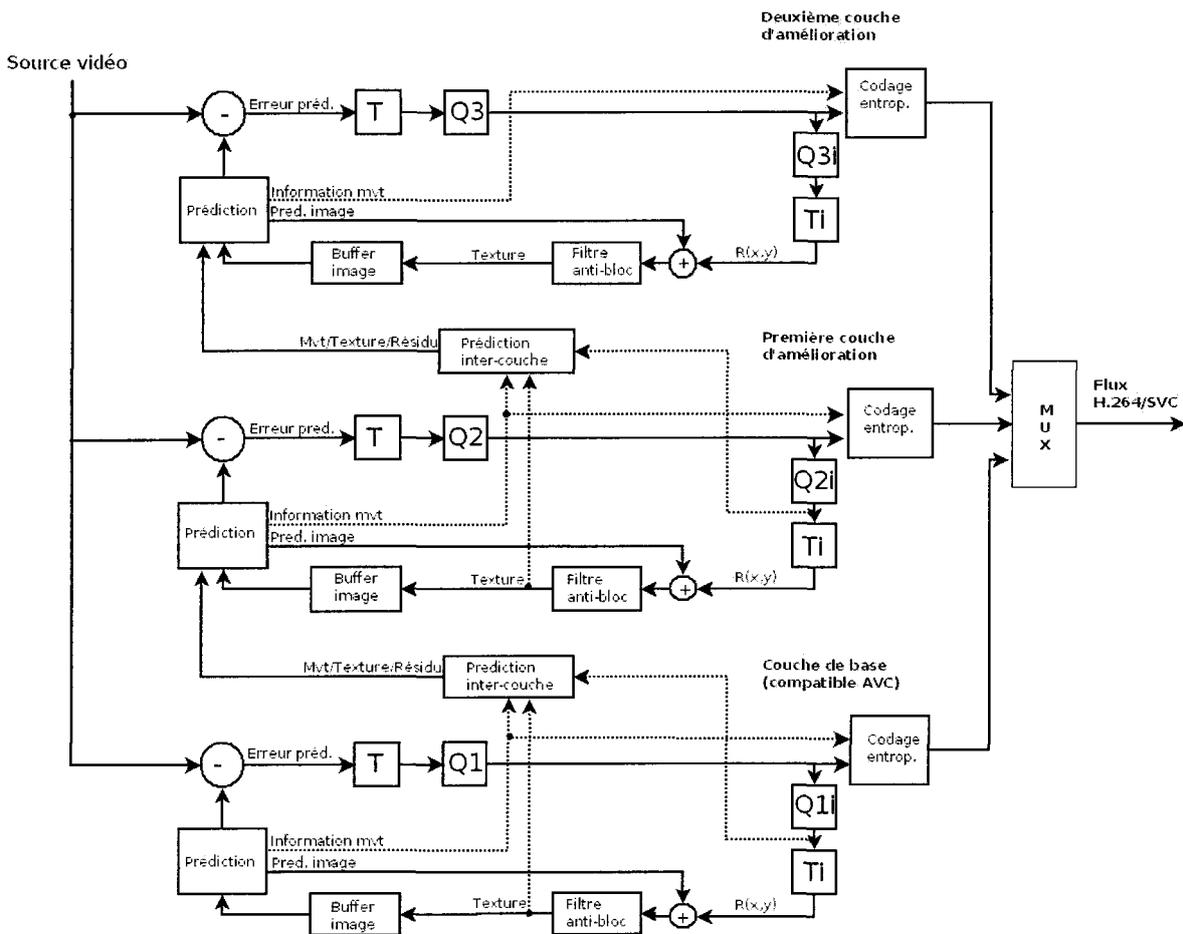


Figure 1.22 Schéma d'un encodeur H.264 SVC en échelonnabilité CGS

Nous précisons que la prédiction inter-couche intra-image ou résiduelle est réalisée dans le domaine transformé [SCH07]. L'inconvénient majeur de ce type d'échelonnabilité est de ne pas pouvoir fournir plus de débits différents que le nombre de couches présentes dans le flux. L'augmentation du nombre de couches n'est pas vraiment une solution à ce problème car cela rend plus complexe l'encodage vidéo et demeure généralement moins efficace.

L'échelonnabilité *MGS* ou *Medium Grain scalability* propose de résoudre ce problème en se basant sur l'échelonnabilité *CGS* mais en y ajoutant des informations de signalisation d'extraction de couche. Contrairement à l'échelonnabilité *CGS*, l'adaptation de débit ne se réalise pas par l'extraction de l'ensemble des NALU d'une couche d'amélioration mais par l'extraction individuelle d'une couche d'amélioration pour une unité d'accès (ou image). Le surcoût de débit (informations de signalisation) engendré par cette échelonnabilité n'est pas négligeable et se chiffre à deux ou trois pourcents.

La Figure 1.23 illustre un exemple de structure de codage par défaut d'une séquence vidéo encodée en *MGS*. Dans cet exemple, l'extraction des couches d'amélioration se fait par niveau temporel hiérarchique. L'information d'extraction de couche est représentée par des numéros sur la figure. Les premières couches d'amélioration supprimées vont être celles des images bidirectionnelles qui ne sont pas utilisées comme référence par d'autres images bidirectionnelles (images b). Il faut noter que ces couches d'amélioration ne sont pas forcément toutes supprimées et qu'il peut être nécessaire d'en supprimer une pour aboutir au débit visé. La suppression des couches d'amélioration se poursuit avec celles des images bidirectionnelles servant de référence (images B). Puis les couches d'amélioration des images prédites et intra sont supprimées.

L'inconvénient de cette technique est l'apparition d'effet de drift, voire de changement de qualité, quand une couche d'amélioration est extraite alors qu'elle était utilisée comme référence par une autre image. Ce problème peut arriver dans notre exemple si nous supprimons la couche d'amélioration de l'image intra alors que la couche de base de l'image P utilise comme référence l'image intra dans sa qualité maximale (couche de base + couche d'amélioration).

Ce problème peut être solutionné en utilisant des images clefs. Cette technique permet de créer une hiérarchisation entre les images et de coordonner les liens de prédiction.

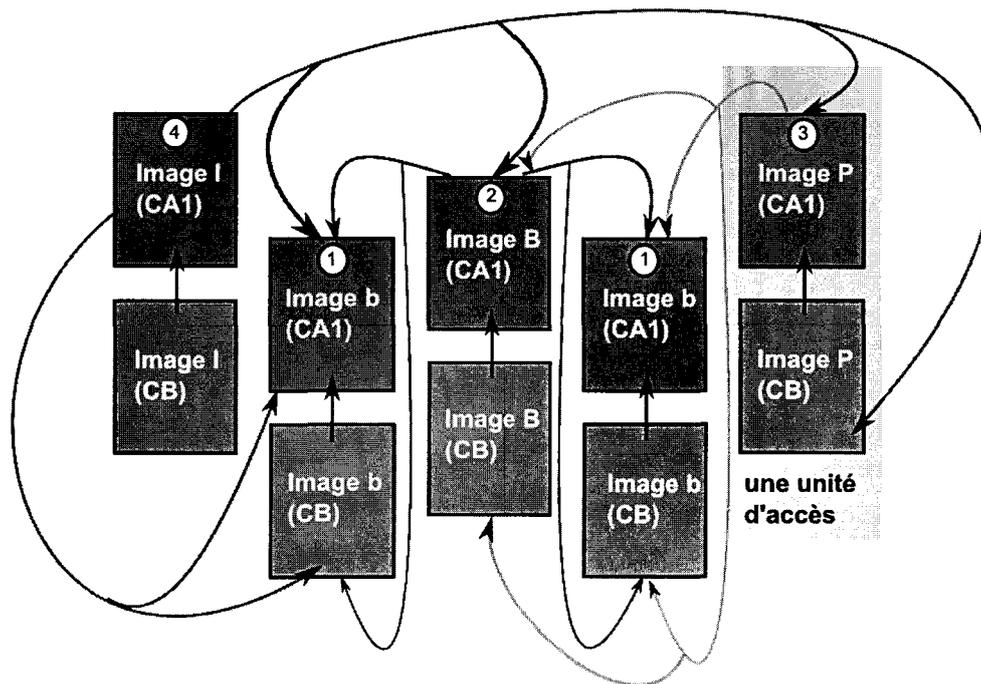


Figure 1.23 Exemple de structure de codage par défaut d'une séquence codée en MGS

La Figure 1.24 illustre un exemple de structure de codage par défaut d'une séquence vidéo encodée en MGS. Dans cet exemple, la couche de base d'une image clef utilise une prédiction vers la couche de base d'une autre image. Le codage de la couche d'amélioration d'une image clef utilise uniquement la prédiction inter-couche avec la couche de base de l'image courante. Les images qui ne sont pas clef prennent pour référence les couches de base et/ou d'amélioration.

Nous ajouterons que l'échelonnabilité MGS permet également de créer des NALU intermédiaires avec un certain nombre de coefficients fréquentiels résiduels. L'un des inconvénients majeurs de ce type d'échelonnabilité est également le surcoût de débit dû à la création des multiples NALU.

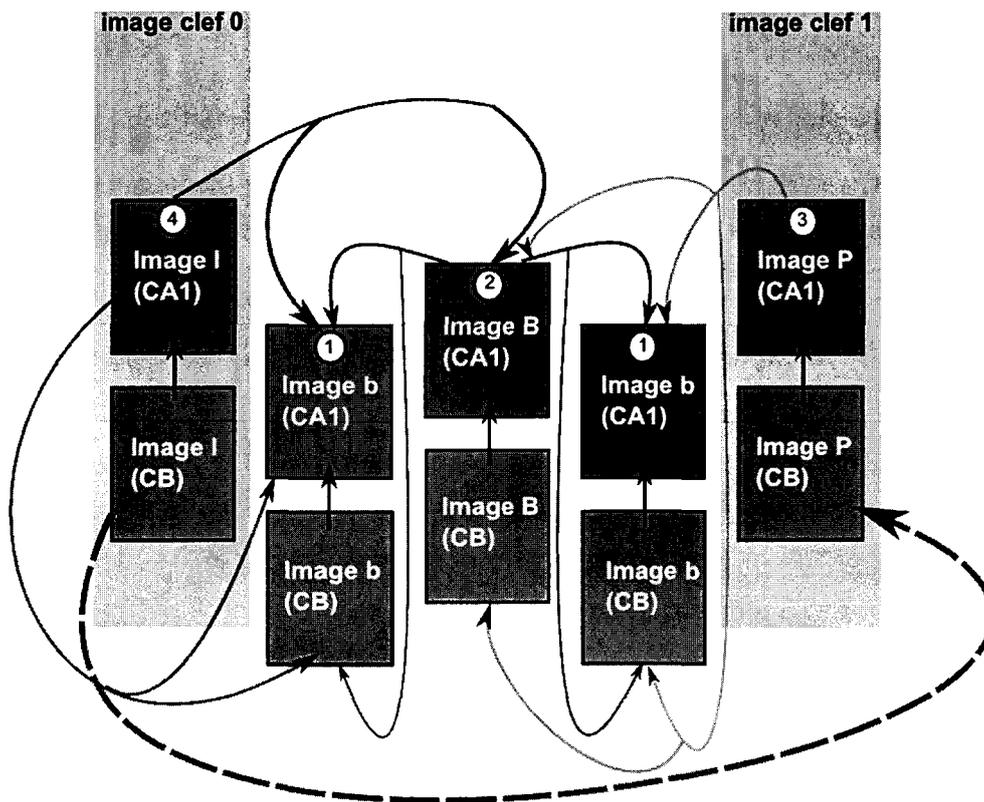


Figure 1.24 Exemple de structure de codage d'une séquence codée en MGS utilisant des images clefs

1.3.3 Évaluation des performances de H.264 SVC

L'objectif de cette section est de présenter les performances de l'extension échelonné de H.264/AVC à travers différents types d'échelonnabilité.

Dans [SCH07], des expérimentations ont été réalisées sur les différents types d'échelonnabilité de H.264 SVC. La Figure 1.25 est issue de cet article et reprend en partie les résultats sur l'échelonnabilité en qualité. Les échelonnabilités CGS et MGS (utilisant les images clefs) sont utilisées avec les séquences *City* et *Crew* au format CIF 15 Hz. Ces flux sont comparés à un flux H.264/AVC utilisant un GOP de 16 images et des images bidirectionnelles hiérarchiques. Pour ces flux échelonnables, les auteurs précisent qu'il y a une différence de 12 entre le paramètre de quantification du débit le plus faible (QP le plus élevé) et celui du débit le plus haut (QP le plus petit). Le paramètre DQP utilisé dans les encodages CGS correspond à la différence entre les valeurs du paramètre de quantification de deux couches successives. Il est mentionné que l'efficacité de codage du CGS diminue lorsque le nombre de débits intermédiaires est augmenté. Les auteurs ont ajouté à titre indicatif des résultats avec l'échelonnabilité CGS combinée à un décodage à boucles

multiples bien qu'il ne soit pas supporté par le H.264 SVC. Ceci permet de vérifier que le gain en efficacité de codage est très faible et ne justifie pas une augmentation de la complexité au décodage. La figure contient également la courbe PSNR en fonction du débit pour l'encodage MGS. Il est configuré pour utiliser les images clefs et un DQP égal à 6.

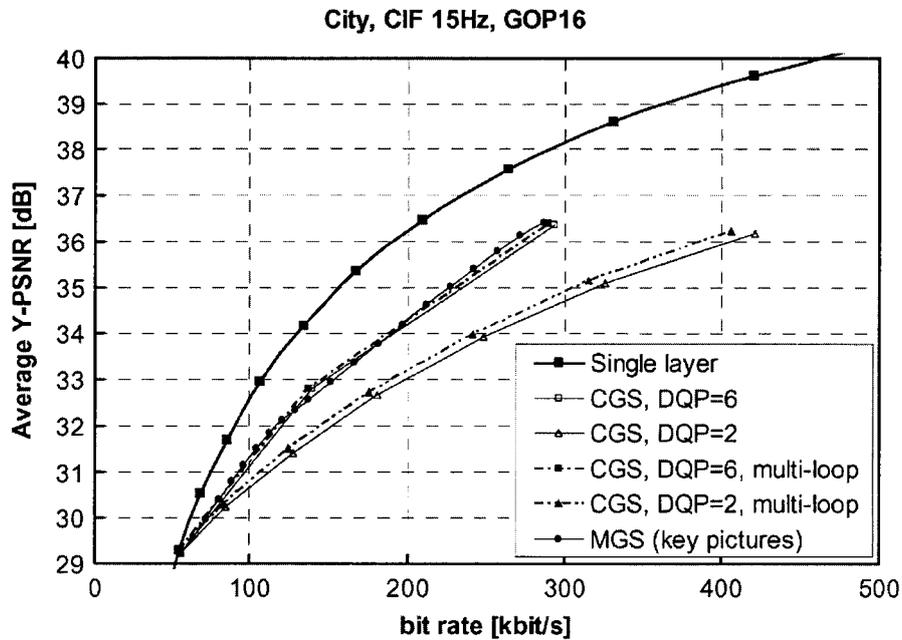


Figure 1.25 Comparaison des échelonnabilités CGS et MGS pour la séquence *City* (CIF 15 Hz) [SCH07]

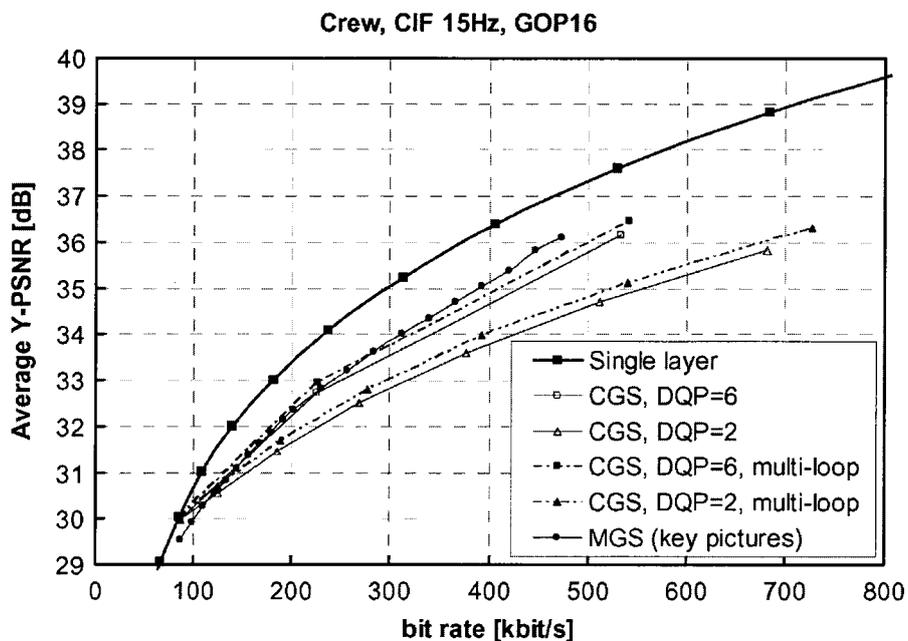


Figure 1.26 Comparaison des échelonnabilités CGS et MGS pour la séquence *Crew* (CIF 15 Hz) [SCH07]

Nous remarquons une amélioration de l'efficacité de codage pour le MGS par rapport au CGS. Les auteurs mentionnent également que le codage MGS permet d'augmenter significativement le nombre de débits intermédiaires en comparaison au codage CGS.

Dans [HUA07], les auteurs proposent une comparaison des performances de l'échelonnabilité spatiale par rapport à un encodage H.264/AVC à la même définition spatiale. Les flux H.264 SVC en échelonnabilité spatiale sont composés de trois couches pour une représentation de la séquence Crew aux formats QCIF, CIF et 4CIF. Le débit nécessaire au décodage d'une couche d'amélioration correspond au débit de cette couche additionné au débit des éventuelles couches de référence. Sur la Figure 1.27, nous pouvons constater que la couche de base (format QCIF) a une évolution du PSNR de la luminance en fonction du débit similaire à celle de l'encodage H.264/AVC. À débit identique, le décodage de la couche d'amélioration au format CIF a un PSNR inférieur d'environ 0,5 dB par rapport à l'encodage H.264/AVC. Similairement, le décodage de la couche d'amélioration au format 4CIF a un PSNR inférieur d'environ 1 dB (à faible débit) et 0,3 dB (à haut débit) par rapport à l'encodage H.264/AVC. Les auteurs concluent que l'échelonnabilité spatiale est obtenue au détriment d'une légère perte de l'efficacité de codage. Des résultats similaires ont été obtenus dans [SCH08].

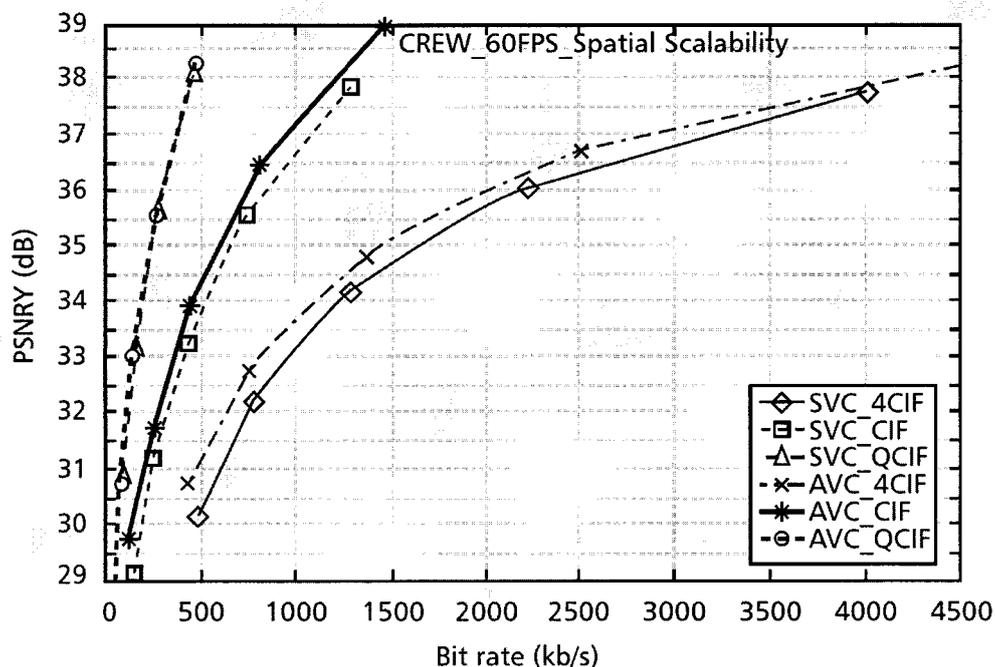


Figure 1.27 Efficacité de l'échelonnabilité spatiale - séquence Crew [HUA07]

1.4 Conclusion

Dans ce chapitre, nous avons présenté la norme H.264/AVC et ses différents outils de codage. La prédiction intra-image et les avancées dans la prédiction inter-image mais également le codage entropique, permettent une nette amélioration du taux de compression des séquences vidéo par rapport aux normes précédentes. L'extension échelonnable de H.264/AVC permet trois types d'échelonnabilité : temporelle, spatiale et en qualité. Le H.264 SVC permet de s'adapter aux divers terminaux vidéo et offre de multiples solutions aux scénarii de transmission vidéo du marché professionnel et grand public.

Dans la suite de ce manuscrit, l'ensemble de nos travaux est fondé sur cette norme et son extension échelonnable.

Chapitre 2: Solution proposée de transcodage des flux vidéo haute définition en H.264/AVC

2.1 Introduction

Le transcodage vidéo peut se définir comme une étape de modification d'une ou plusieurs caractéristiques d'un flux vidéo. Ces caractéristiques sont par exemple la norme (ou dans certains cas le standard) de compression vidéo utilisée et les paramètres de configuration d'encodage (définition spatiale, débit, nombre d'images par seconde...). Le principe du transcodage vidéo est illustré en Figure 2.1. Il consiste à changer un ou plusieurs de ces paramètres dans le but de faciliter la transmission vidéo ou l'interopérabilité entre l'émetteur et le récepteur vidéo. Nous pouvons illustrer ceci par un exemple de chaîne de transmission vidéo complexe, composée d'un encodeur-émetteur vidéo et de plusieurs récepteurs vidéo aux caractéristiques de décodage vidéo différentes. Le flux vidéo est émis dans une norme de compression vidéo A, compatible avec le terminal 1 mais non-compatible avec le terminal 2 (norme de compression vidéo B). Ainsi, le terminal 2 nécessitera un transcodage vidéo en amont afin de passer de la norme de compression vidéo A à la norme de compression vidéo B.

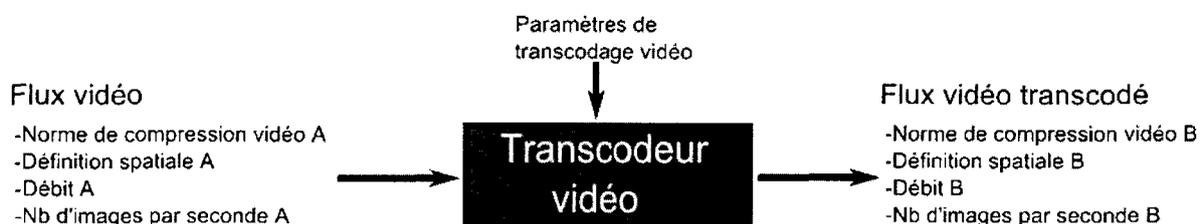


Figure 2.1 Schéma général d'un transcodeur vidéo

Un cas particulier du transcodage vidéo consiste à garder la même norme de compression vidéo que le flux vidéo compressé entrant mais à diminuer par contre le débit vidéo. Ce chapitre est consacré à ce cas particulier du transcodage vidéo, appelé également *transrating*.

L'objectif principal du projet TOSCANE était d'améliorer la couverture globale de diffusion des contenus audiovisuels en optimisant la transmission sur les réseaux filaires de type ADSL, mais également la diffusion (filaire ou sans fil) au sein du réseau local de l'abonné. Ceci a été rendu possible par l'optimisation conjointe du codage source et du codage canal. En amont de la chaîne de transmission, les flux vidéo étaient encodés en H.264/AVC ou en H.264/AVC SVC. Dans le cas des flux en codage échelonné,

l'adaptation de débit se réalisait par suppression des NALU de la couche d'amélioration la plus haute (échelonnabilité CGS). Dans le cas des flux H.264/AVC, non-échelonnables, l'adaptation de débit pouvait se réaliser par transcodage spatial, temporel ou en qualité. La solution de transcodage vidéo proposée dans ce chapitre repose sur ce dernier type de transcodage. Elle a été retenue lors du projet pour sa flexibilité et sa faible complexité afin d'adapter le débit vidéo des flux H.264/AVC au débit disponible sur le canal de transmission. Ce même transcodeur H.264/AVC est utilisé pour les expérimentations de ce chapitre. Ce logiciel n'est pas un outil de simulation de transcodage mais un programme écrit en *langage C*, composé de milliers de lignes de code. Des mois de lecture de la norme, de compréhension du logiciel de référence *JM* (encodeur et décodeur) et de développement informatique ont été nécessaires à ce travail. Comme le programme n'utilise pas de bibliothèques spécifiques mais uniquement des fonctions *C standard*, il peut être compilé sur n'importe quel système d'exploitation.

Nous débuterons par un état de l'art du transcodage vidéo dans le cas de la norme H.264/AVC dans la section 2.2. Notre architecture de transcodage basée sur la sélection des coefficients résiduels fréquentiels est présentée dans la section 2.3. Nous poursuivrons en section 2.4 par une comparaison des performances en termes de réduction de débit et de qualité vidéo entre notre solution de transcodage vidéo H.264/AVC et celle couramment utilisée mais basée sur la requantification. La section 2.5 est consacrée à la comparaison des performances des deux techniques implémentées de transcodage vidéo par sélection fréquentielle : la sélection en position fréquentielle et la sélection en niveau fréquentiel. La section 2.6 propose une modélisation simple et efficace du débit dans le cadre du transcodage vidéo par sélection fréquentielle dans le but d'estimer le débit d'un flux vidéo transcodé. Nos conclusions sont proposées en section 2.7.

2.2 État de l'art du transcodage H.264/AVC

Xin et al. [XIN05] décrivent trois types d'architecture de transcodage : le transcodeur en boucle ouverte, le transcodeur en cascade dans le domaine spatial appelé Cascaded Pixel-Domain Transrater (CPDT) et le transcodeur dans le domaine de la transformée en cosinus discrète appelé Discrete cosine transform-Domain Transrater (DDT).

Un transcodeur en boucle ouverte est basé sur une architecture peu complexe. Un traitement est réalisé sur le flux entrant (modification ou suppression de certains coefficients résiduels fréquentiels) sans qu'il y ait de retour sur ces informations. Le principe d'un transcodeur H.264/AVC en boucle ouverte est illustré en Figure 2.2. Un décodage entropique est nécessaire afin de récupérer les coefficients résiduels, suivi d'un traitement

sur ces coefficients (requantification ou sélection fréquentielle), puis un codage entropique afin de reconstituer le flux vidéo. Nous noterons que les éléments syntaxiques du flux vidéo (codés en Exp-Golomb) sont copiés à l'identique. Le transcodeur en boucle ouverte nécessite peu de mémoire et de temps de calcul.

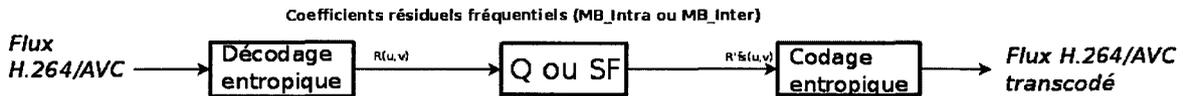


Figure 2.2 Principe d'un transcodeur en boucle ouverte

La requantification, tout comme la sélection fréquentielle, modifie la valeur des pixels d'un bloc par un traitement dans le domaine fréquentiel. A l'encodage, une prédiction intra-image est utilisée sur les macroblocs voisins de l'image encodée. Cette prédiction réalisée dans le domaine spatial est notée $P(x,y)$. Le bloc résiduel qui sera codé correspond à la différence entre le bloc de pixels $B(x,y)$ et le bloc de prédiction $P(x,y)$:

$$R(x, y) = B(x, y) - P(x, y)$$

Le transcodeur en boucle ouverte applique directement la requantification ou la sélection fréquentielle au résidu $R(u,v)$ dans le domaine transformé pour obtenir le nouveau résidu transcodé $R'(u,v)$.

Au décodage du flux transcodé, la prédiction intra-image va utiliser les blocs voisins transcodés puis décodés afin de créer un nouveau signal de prédiction $P'(x,y)$ contenant une erreur due au transcodage :

$$P'(x, y) = P(x, y) - e(x, y)$$

Cette erreur va augmenter lors du décodage et aura pour conséquence de dénaturer complètement le bloc reconstruit :

$$B_{reconstruit}(x, y) = R'(x, y) + P'(x, y)$$

Nous pouvons visualiser en Figure 2.3 ce phénomène d'effet de glissement ou d'*effet de drift* [LEO08]. Il apparaît en cascade dans l'ordre de décodage des macroblocs.

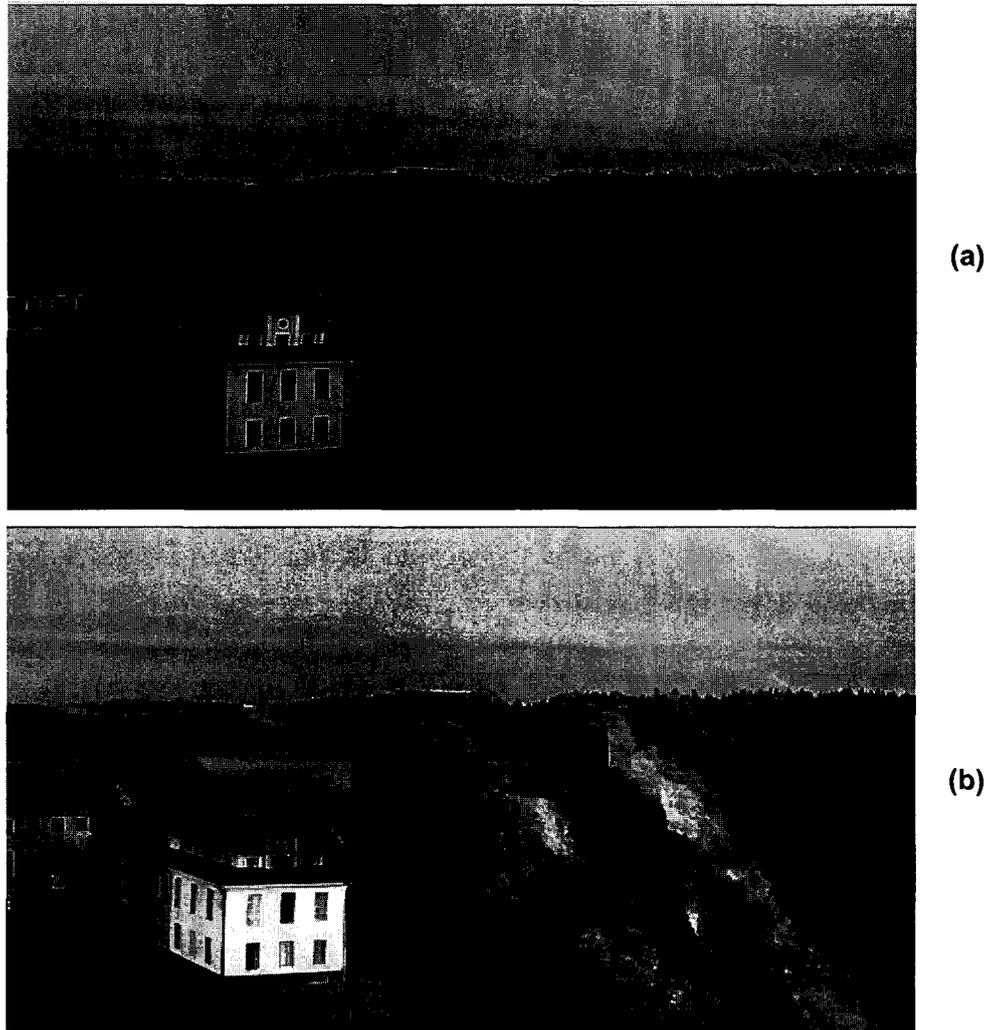


Figure 2.3 Image de référence du flux à l'entrée du transcodeur (a) - Illustration de l'effet de drift dû au transcodage en boucle ouverte d'une image intra en H.264/AVC (b)

Le transcodeur CPDT est composé d'un décodeur et d'un encodeur à complexité réduite réutilisant des données issues de l'encodage initial comme les modes de prédiction ou les vecteurs de mouvement (« mode reuse »). Les calculs de compensation de l'effet de drift se réalisent dans le domaine spatial, également appelé domaine pixel, à la sortie du décodeur. Cette méthode permet notamment d'éviter tout problème de drift mais augmente la complexité du transcodeur.

Le transcodage dans le domaine transformé (*DDT*) prend pour hypothèse que certaines opérations inverses sont linéaires et peuvent être calculées aussi bien dans le domaine spatial que transformé. Ceci permet notamment de s'affranchir du calcul d'une

transformée inverse et de rester dans le domaine transformé tout en réalisant l'opération. La complexité du transcodeur est alors minimisée.

2.2.1 Transcodage par requantification

Le transcodage H.264/AVC par requantification des coefficients résiduels fréquentiels est couramment proposé dans la littérature. Dans [DEC06a], les auteurs proposent d'adapter les architectures CPDT et DDT au standard H.264/AVC en requantifiant uniquement les coefficients résiduels fréquentiels des macroblochs codés en INTRA_16x16 (cf 1.2.2.1). Cependant, ils mentionnent que les hypothèses de linéarité des opérations ne sont pas toujours vérifiées. Ainsi, le PSNR d'une séquence vidéo transcodée par les deux architectures peut être légèrement différent. Ils expliquent que cette différence vient d'erreurs d'arrondi après certaines divisions comme lors du calcul des modes de prédiction *DC* et *plan* ou après la transformée inverse. Ces erreurs créent alors un effet de glissement plus ou moins prononcé.

Dans [LEF06], les auteurs proposent une architecture hybride de transcodage H.264/AVC par requantification, opérant dans le domaine spatial et le domaine fréquentiel, réutilisant les modes de prédiction et vecteurs de mouvement présents dans le flux vidéo. Similairement à [DEC06a], des expérimentations ont été réalisées en utilisant une architecture de type DDT, opérant dans le domaine transformé, sur les images intra et inter. Les auteurs mentionnent que l'effet de glissement provoqué par ces erreurs d'arrondi n'est pas acceptable pour les images intra. Cet effet de glissement est par contre négligeable dans le cas des images inter, similairement au transcodage des flux MPEG-2. Ainsi, l'architecture proposée opère dans le domaine spatial pour les images intra et dans le domaine fréquentiel pour les images inter. Par la minimisation de cet effet de glissement, ce transcodeur est un bon compromis entre qualité d'image et complexité.

La Figure 2.4 propose un exemple de coefficients résiduels fréquentiels avant et après requantification. Les coefficients sont présentés dans l'ordre du zig-zag après le décodage entropique et le décodage run-level (a). La quantification inverse utilise le paramètre de quantification initial (QP1) afin d'obtenir les coefficients résiduels (b). La requantification consiste à appliquer une nouvelle quantification sur les coefficients résiduels (b) en utilisant un paramètre de quantification plus élevé (QP2>QP1). Nous remarquons que certains coefficients (c) sont supprimés (disparition d'une fréquence) ou que leur amplitude est modifiée (diminution de la précision).

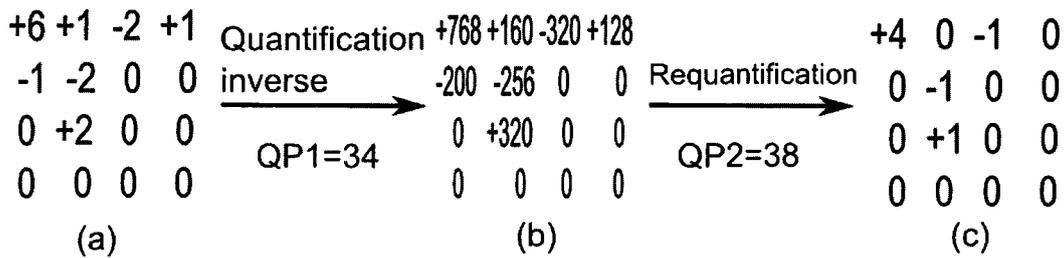


Figure 2.4 Exemple de coefficients résiduels fréquentiels d'un bloc 4x4 dans l'ordre du zig-zag après quantification inverse et requantification

Dans la suite de chapitre, nous utiliserons le paramètre de requantification dQP , suivant la relation : $dQP = QP_requantification - QP_initial$

2.2.2 Transcodage par sélection fréquentielle

Une autre technique de transcodage consiste à supprimer certains coefficients résiduels fréquentiels en fonction de leur position ou de leur amplitude dans le codage en zig-zag. Ceci permet notamment de garder la précision initiale des coefficients restants. Le filtrage par sélection fréquentielle génère un flou global plus ou moins prononcé sur les images transcodées, qui peut être moins gênant visuellement que les effets de blocs ou de scintillement des contours dus à la requantification.

Les architectures existantes de transcodage des flux MPEG-2 ont l'avantage de s'affranchir d'un décodage puis d'un encodage entropique [COU08]. Malheureusement, ces algorithmes ne sont pas facilement transposables au standard H.264/AVC à cause des nouveaux outils de codage comme la prédiction intra et le codage entropique CAVLC ou CABAC.

De Cock et al. [DEC06b] ont étudié une architecture de transcodage fondée sur la sélection des coefficients résiduels transformés suivant deux approches : une sélection par position suivant l'ordre d'apparition dans le codage en zig-zag et une sélection par niveau. Cependant, cette architecture traite uniquement les macroblocs de type inter. Les auteurs précisent qu'étendre leur architecture au cas des macroblocs intra réduirait inacceptablement la qualité visuelle du GOP entier par la présence de l'effet de drift. L'architecture que nous proposons étend la précédente solution au cas des macroblocs intra par un transcodage en boucle fermée, elle est présentée dans la section 3.3.

La Figure 2.5 illustre un exemple de coefficients résiduels fréquentiels traités par les deux types de sélection fréquentielle. Les coefficients résiduels fréquentiels (a) sont présentés dans l'ordre du balayage en zig-zag. La sélection fréquentielle par position utilise un paramètre FP (*Frequency Position*) égal à trois. Ainsi, le coefficient DC et les deux premiers coefficients dans l'ordre du balayage en zig-zag sont gardés (b). La sélection

fréquentielle par niveau utilise un paramètre FL (*Frequency Level*) égal à trois. Ainsi, le coefficient DC et les deux coefficients les plus élevés (le signe n'est pas pris en compte) dans l'ordre du balayage en zig-zag sont gardés (c).

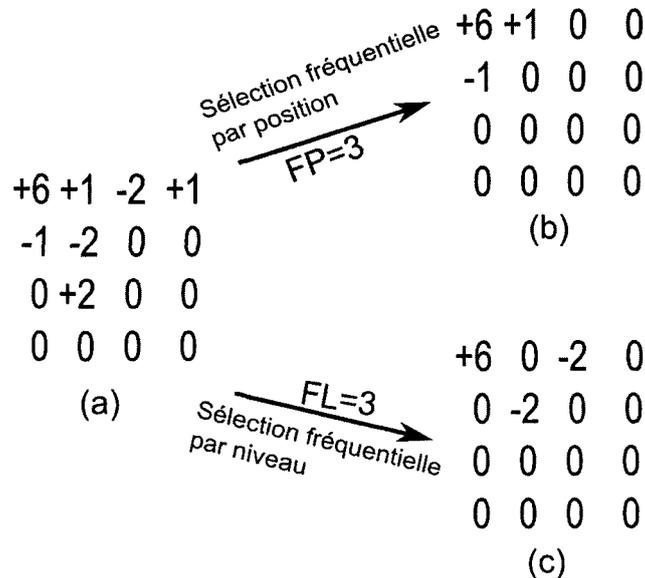


Figure 2.5 Exemple de traitement des coefficients résiduels par sélection en position fréquentielle et en niveau fréquentiel

2.3 Solution proposée de transcodage H.264/AVC à complexité réduite

Dans cette section, nous présentons notre architecture de transcodage des flux vidéo H.264/AVC [DEK09] [DEK10]. Notre objectif était de concevoir une solution de transcodage à faible complexité afin de permettre un traitement temps réel des flux vidéo. La sélection des coefficients résiduels transformés a été choisie pour la minimisation de la complexité mais également pour le type de dégradation engendré : uniquement du flou. Comme précisé ci-dessus, nous nous concentrerons sur deux approches : une sélection en position fréquentielle dans l'ordre de balayage en zig-zag (FP) et une sélection en niveau fréquentiel (FL).

2.3.1 Architecture

L'architecture de transcodage proposée est illustrée en Figure 2.6. Elle est fondée sur le CPDT pour les macroblocs codés par la prédiction intra-image et sur le transcodeur en boucle ouverte pour les macroblocs codés par la prédiction inter-image. Le choix du traitement de ces macroblocs est expliqué dans la sous-section 2.4.2.3. Cette architecture

permet de concilier une qualité vidéo diffusable avec un traitement temps réel. Dans un flux vidéo H.264/AVC, un élément syntaxique permet de différencier les macroblochs de type intra des macroblochs de type inter. Il s'agit du type de macrobloc (*mb_type*). Celui-ci est utilisé dans notre solution afin de basculer sur le traitement adéquat. Il faut noter que le type d'image n'est pas utilisé car des macroblochs intra peuvent être présents dans des images inter.

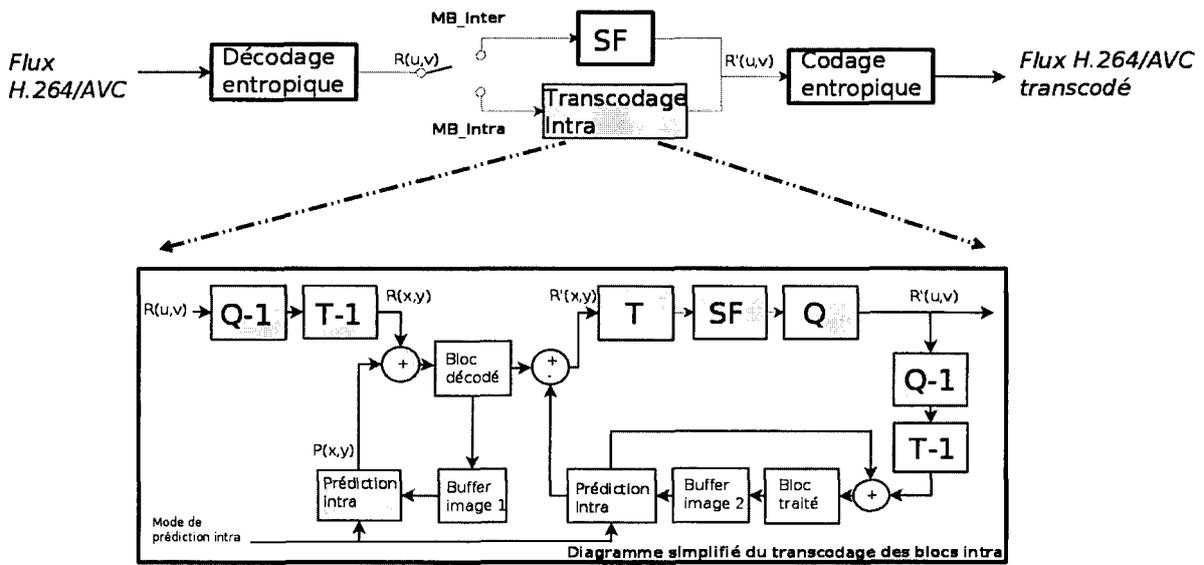


Figure 2.6 Diagramme simplifié du transcodeur H.264/AVC en sélection fréquentielle (SF) pour les macroblochs intra et inter

Le descriptif du traitement des macroblochs intra et inter est présenté dans les sections suivantes. Nous noterons que le bloc nommé *SF* correspond au traitement de sélection fréquentielle des coefficients résiduels fréquentiels. Ce traitement s'effectue soit par position fréquentielle ou par niveau fréquentiel.

Comme nous l'avons vu dans le chapitre 1, le codage CAVLC a été retenu dans le projet TOSCANE pour sa présence dans chaque profil de la norme. L'architecture de transcodage utilise le CAVLC mais une évolution vers le CABAC est facilement envisageable.

2.3.1.1 Traitement des macroblochs intra

Le traitement des macroblochs intra est effectué dans le domaine spatial afin d'empêcher toute propagation d'erreur. Nous réutilisons le type de codage des macroblochs (INTRA_4x4 ou INTRA_16x16) ainsi que leur mode de prédiction intra-image afin de limiter la complexité de calcul du transcodeur. L'architecture est composée de deux étapes de traitement consécutives : un décodeur intra-image et un encodeur à complexité réduite.

La première partie de l'architecture de transcodage correspond au décodage d'un macrobloc intra. Elle se compose d'un décodage entropique (CAVLD), d'une quantification

inverse, d'une transformée inverse et d'une prédiction intra-image. Dans la Figure 2.6, Q^{-1} représente la quantification inverse et T^{-1} désigne la transformée inverse. Le résultat de la transformée inverse est appelé $R(x,y)$ et correspond aux coefficients résiduels dans le domaine spatial. Ces coefficients sont ajoutés à la prédiction intra-image $P(x,y)$ afin de construire le bloc décodé. Ce bloc est enregistré dans l'image décodée à la position de la partition adéquate du macrobloc courant, dans un tampon mémoire nommé *Buffer Image 1*.

La seconde partie de l'architecture de transcodage est un encodeur à complexité réduite. Un autre tampon mémoire est ajouté, nommé *Buffer Image 2*. Il permet d'enregistrer les blocs qui ont subi la sélection fréquentielle. La prédiction intra-image du *Buffer Image 2* est soustraite au bloc résultant de la première partie de l'architecture de transcodage, afin de calculer et coder le nouveau résidu. Puis le nouveau résidu est transformé, sélectionné en fréquence (algorithme de sélection par position fréquentielle ou par niveau fréquentiel) puis quantifié. Le traitement est poursuivi par le décodage du nouveau résidu. Cette opération nécessite une quantification inverse, une transformée inverse et une mise à jour du *Buffer Image 2* afin d'assurer le bon fonctionnement de la nouvelle prédiction intra-image. Le nouveau résidu quantifié est codé au format *Run-Level* puis entropiquement par le CAVLC. Les paramètres de transcodage pour les macroblocs intra sont appelés position fréquentielle (FP_Intra) ou niveau fréquentiel (FL_Intra).

Une image intra issue de notre architecture de transcodage (SF) est illustrée en Figure 2.7. À l'entrée du transcodeur, nous avons utilisé le même flux H.264/AVC qu'en Figure 2.3 avec un paramètre FP_Intra égal à six. Le flux H.264/AVC a été encodé avec uniquement des images intra, composées de macroblocs codés en *INTRA_4x4* ou *INTRA_16x16*. Dans chaque bloc 4x4, seulement le coefficient DC et les cinq coefficients AC suivants dans l'ordre de balayage du zig-zag sont gardés, incluant d'éventuels coefficients nuls. L'image montre clairement que l'effet de drift a été évité, fournissant une qualité vidéo très satisfaisante.



Figure 2.7 Image intra résultant de notre architecture de transcodage en boucle fermée (FS)

Le transcodage des macroblocs *INTRA_16x16* (le transcodage des blocs de chrominance n'est pas utilisé dans ces travaux mais peut s'effectuer de la même manière) se réalise similairement à celui des macroblocs *INTRA_4x4*. Dans ce cas, l'ensemble des seize coefficients *DC* (*LUMA_INTRA16x16DC*) des seize blocs *4x4* qui composent le macrobloc (cf 1.2.3.1) ne subissent pas de transcodage. Par contre les blocs *LUMA_INTRA16x16AC*, composés au maximum de quinze coefficients résiduels fréquentiels sont sélectionnés en fréquence.

2.3.1.2 Traitement des macroblocs inter

L'architecture de transcodage inter-image est également illustrée en Figure 2.6. Cette architecture en boucle ouverte a été proposée par De Cock et al. dans [DEC06b]. De faible complexité, elle ne nécessite pas le décodage des macroblocs inter avant et/ou après transcodage. Notons que si la prédiction intra-image d'un macrobloc utilise comme voisin un macrobloc inter, un effet de drift est prévisible si ce macrobloc inter a été transcodé. Dans le cas d'un flux vidéo utilisant les prédictions intra-image et inter-image, il est nécessaire de spécifier à 1 l'option d'encodage *constrained_intra_pred_flag* afin d'empêcher l'utilisation des macroblocs inter lors de la prédiction intra-image.

Le transcodage inter-image par sélection fréquentielle (*FP* ou *FL*) est un peu moins complexe que le transcodage par requantification. Uniquement un décodage entropique, une sélection fréquentielle et un codage entropique sont nécessaires. D'autre part, le transcodage inter-image par requantification nécessite un décodage entropique, une

quantification inverse avec le paramètre de quantification initial $QP1$, une quantification avec un nouveau paramètre de quantification $QP2 > QP1$ et enfin un codage entropique.

2.3.2 Performances de l'architecture de transcodage en sélection fréquentielle

Notre architecture de transcodage (*FP* et *FL*) a été implémentée sur ordinateur PC en prenant pour base le logiciel de référence du standard H.264/AVC appelé JM (version 12.4). Notre programme est écrit en langage C et sans optimisation en langage assembleur (ou instructions SSE).

Le Tableau 2.1 synthétise le nombre d'images transcodées par seconde par notre transcodeur en *FP* pour des flux H.264/AVC au format 720p.

Nous commençons par utiliser un flux vidéo encodé en H.264/AVC avec uniquement des images intra. Le transcodage de ce flux se réalise à environ six images par seconde.

Nous poursuivons les expérimentations avec un flux vidéo encodé sous forme de GOP de douze images, respectivement une image intra et onze images prédites. Le transcodage de ce flux, composé majoritairement d'images prédites, se réalise à environ quinze images par seconde. Nous en déduisons que les images prédites sont transcodées plus de deux fois plus rapidement que les images intra.

Nous finissons les expérimentations avec un flux vidéo encodé sous forme de GOP de douze images. Ce GOP est composé d'une image intra, de deux images prédites et de neuf images bidirectionnelles avec codage hiérarchique. Ainsi, les images nommées « b » utilisent comme référence les images notées « B ». Notre transcodeur permet un traitement quasi temps réel sur ce type de flux (21 images par seconde), composé majoritairement d'images bidirectionnelles. Nous en déduisons que les images bidirectionnelles sont transcodées plus de trois fois plus rapidement que les images intra.

Séquences yuv 720p utilisées : Parkrun, Mobcal, Shields, Stockholm, Intotree et Oldtown		
CPU: Intel Core 2 Duo E8500 @ 3.16Ghz (un coeur utilisé)		
Encodage H.264/AVC	Paramètres de transcodage	Nombre d'images par seconde (ips)
Encodage Intra	Paramètre FP Intra = 3	6.0
Encodage Intra/Inter-image avec un GOP de 12 (IPPPPPPPPPP)	Paramètre FP Intra/Inter = 3	14.7
Encodage Intra/Inter-image avec un GOP de 12 (IbBbPbBbPbBb)	Paramètre FP Intra/Inter = 3	21.0

Tableau 2.1 Performances du transcodeur par sélection en position fréquentielle

Dans le cadre du projet TOSCANE, notre architecture de transcodage pour les images intra a été portée sur *FPGA* par le laboratoire LIEN de Nancy. Il s'agit d'une architecture Virtex 5 du constructeur Xilinx. Le transcodage y est réalisé en temps réel sur des vidéos au format 1080p.

Nous avons remarqué que le transcodeur par sélection en niveau fréquentiel était 8% plus lent que le transcodeur en sélection par position. Ceci peut s'expliquer par les multiples boucles utilisées par l'algorithme de tri de la sélection par niveau. Il faut d'ailleurs noter que cet algorithme de tri n'est pas optimisé. Ceci peut constituer un léger avantage pour le transcodage par sélection en position fréquentielle.

2.4 Comparaison de la qualité visuelle du transcodage par sélection en position fréquentielle (FP) et requantification

Dans cette partie, nous comparons la qualité visuelle de flux H.264/AVC en haute définition transcodés par sélection en position fréquentielle à la qualité visuelle obtenue par le transcodage en requantification [DEK10]. Nous avons choisi d'utiliser ici la sélection en position fréquentielle plutôt que la sélection en niveau fréquentiel car cette première solution a une complexité légèrement inférieure. Une comparaison de ces deux solutions en termes de qualité visuelle et de réduction de débit est présentée en section 2.5.

2.4.1 Paramètres de transcodage utilisés

Nous avons utilisé l'encodeur de JM 12.4 pour réaliser l'encodage H.264/AVC sur 400 images des séquences 720p *Parkun*, *Mobcal*, *Shields* et *Stockholm* (illustrées en Figure B.12), disponibles en YUV sous-échantillonné en 4:2:0. L'encodeur a été configuré avec le *main profile* du standard H.264/AVC et l'encodage entropique CAVLC. Nous avons préféré désactiver le filtre de deblocking lors de l'encodage des flux vidéo car son efficacité est contestable (cf 1.2.1.5) et son utilisation pourrait fausser l'interprétation de nos résultats.

Dans les paragraphes suivants, le paramètre de transcodage représentant la position fréquentielle (*FP*) peut théoriquement varier de 1 (uniquement le coefficient DC) à 16 (aucun changement). Afin de garder une qualité vidéo acceptable, nous avons jugé nécessaire de garder au minimum le coefficient DC, le premier coefficient fréquentiel horizontal et le premier coefficient fréquentiel vertical. Ainsi, le paramètre *FP* sera supérieur ou égal à trois.

Nous utilisons également un paramètre de requantification constant (*dQP*), représentant la valeur ajoutée au paramètre de quantification initial afin d'obtenir le nouveau paramètre de quantification.

Afin de réaliser cette comparaison, un transcodeur en requantification a été développé sur la même architecture que le transcodeur en sélection fréquentielle. Dans les sections suivantes, nous comparons les performances en réduction de débit et qualité vidéo de notre architecture de transcodage (*FP*) à celles obtenues par requantification. Les deux transcodeurs utilisent une architecture très proche, basée sur le mode « reuse ». Le module de sélection par position fréquentielle, tout comme celui de requantification, est appliqué uniquement aux coefficients de la luminance. Notons que notre architecture de transcodage peut facilement être modifiée afin de réaliser le traitement des blocs de chrominance.

2.4.2 Critères de qualité visuelle utilisés

Dans cette sous-section, nous comparons l'impact du transcodage des images intra ou des images inter sur la qualité vidéo. Pour mener à bien cette expérimentation, nous avons encodé les séquences en H.264/AVC en utilisant un GOP de 12, comprenant des images intra, prédites et bidirectionnelles. Afin de rester dans un scénario de diffusion vidéo haute définition réaliste, nous avons utilisé des débits d'encodage de respectivement 12, 8 et 6 Mb/s. L'évaluation de la qualité vidéo est effectuée à la fois par le PSNR et par la métrique VQM.

2.4.2.1 PSNR (RSBC)

Pour commencer, nous allons utiliser le *rapport signal sur bruit crête (RSBC)*, ou *Peak Signal to Noise Ratio (PSNR)*. Dans notre cas, il demeure le moyen le plus facile pour mesurer la distorsion entre la composante d'une image transcodée et celle de l'image de référence (séquence d'origine, jamais compressée). Dans les expérimentations suivantes, nous procéderons au calcul du PSNR uniquement sur la composante de luminance de l'image car aucun traitement n'est appliqué sur les composantes de chrominance. La formule de calcul du PSNR sur la luminance est présentée ci-dessous :

$$PSNR_Y = 10 \log_{10} \left(\frac{d^2}{EQM} \right) (dB)$$
$$EQM = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (Y_{trans}(i, j) - Y_{ref}(i, j))^2$$

La composante de luminance a une largeur de m échantillons et une hauteur de n échantillons. La variable d représente la dynamique du signal. Dans notre cas, les signaux sont numérisés généralement sur 8 bits par composante afin d'obtenir des niveaux allant de 0 à 255. Ainsi d est égal à 255. L'erreur quadratique moyenne (*EQM*) est calculée entre les échantillons de luminance de l'image transcodée Y_{trans} et ceux de leur référence Y_{ref} . Le

PSNR calcule uniquement la distorsion entre deux images et ne peut donc pas être considéré comme une métrique de qualité subjective.

2.4.2.2 Métrique VQM

VQM est une méthode standardisée par l'institut américain NTIA/ITS pour mesurer la qualité d'une vidéo en quantifiant sa détérioration. Cette métrique a été conçue afin de prédire précisément les décisions de votes subjectifs qui sont issues d'un panel d'utilisateurs [ANS03] [VQE03] [PIN04]. Cette méthode utilise divers algorithmes pour identifier d'éventuels changements visuels entre une vidéo après traitement ou transmission et sa référence. Ces changements peuvent être localisés sur les échantillons de luminance ou de chrominance, dans le domaine spatial ou temporel. Le flux vidéo va être divisé en régions spatio-temporelles, composées de zones rectangulaires de pixels sur une ou plusieurs images, afin d'extraire des informations précises sur le contenu de la vidéo. Les techniques employées utilisent entre autres des filtres de rehaussement des contours (pour détecter par exemple l'apparition de flou), des gradients (pour mesurer la variation du bruit dans l'image, l'amplitude d'un mouvement sur plusieurs images) et calculs de luminosité/contraste.

Pour calculer la note VQM, nous nous sommes servi du logiciel VQM pc 2.2, en utilisant le modèle « HRC Television ». Les notes VQM vont de zéro à un, 0.0 signifiant que la qualité vidéo est égale à celle de l'original et 1.0 signifiant que la vidéo est de mauvaise qualité. Comme pour le calcul du PSNR, nous prenons pour référence la séquence vidéo originale et non pas le flux à l'entrée du transcodeur (déjà encodé). Ainsi, la mesure VQM du flux à l'entrée du transcodeur, équivalent à un transcodage avec FP=16 ou dQP=0, n'est pas égale à 0.0 dans les résultats suivants.

2.4.2.3 Impact du transcodage des images intra et inter sur la qualité vidéo

Le Tableau 2.2 nous donne les mesures *PSNR* de la luminance et VQM à titre d'exemple sur la séquence Parkrun en utilisant le transcodage par sélection en position fréquentielle uniquement sur les images intra. Des résultats similaires ont été obtenus avec d'autres séquences. Pour les trois débits, la note VQM augmente pour de faibles valeurs de FP, ceci signifiant que la qualité visuelle diminue. De plus, les résultats des mesures de PSNR montrent une différence de plus de 3 dB entre FP_Intra égal à seize et FP_Intra égal à trois. Cette baisse de la qualité vidéo montre l'importance des images intra sur la qualité vidéo de l'ensemble de la séquence.

Séquence <i>Parkrun</i> – fp_inter=16			
Débit d'encodage H.264/AVC (GOP lbBbPbBbPbBb)	fp_intra	VQM	Y-PSNR (dB)
12 Mb/s	16	0.24	29.40
	5	0.27	25.79
	3	0.29	24.14
8 Mb/s	16	0.30	27.93
	5	0.33	25.19
	3	0.34	23.77
6 Mb/s	16	0.37	26.63
	5	0.39	24.57
	3	0.41	23.38

Tableau 2.2 Impact du transcodage des images intra sur la qualité vidéo

La Figure 2.8 illustre la dépendance entre les images du GOP utilisé. Les images intra sont utilisées comme référence par les images prédites et bidirectionnelles. Le transcodage des images intra a pour conséquence de diminuer la qualité des images intra mais également celle des autres images les utilisant comme référence. La qualité vidéo de l'ensemble de la séquence est alors affectée.

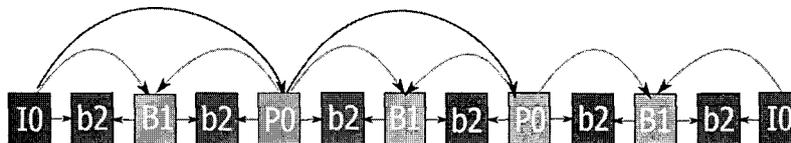


Figure 2.8 Dépendance entre les images du GOP utilisé

Le Tableau 2.3 montre les résultats du transcodage par sélection en position fréquentielle (*FP*) pour les images inter. Chaque flux détient une note VQM constante et une mesure PSNR peu variable. De plus, visuellement nous n'avons pas trouvé de changements importants dans la qualité visuelle.

Séquence <i>Parkrun</i> – fp_intra=16			
Débit d'encodage H.264/AVC (GOP IbBbPbBbPbBb)	FP Inter	VQM	PSNR (dB)
12 Mb/s	16	0.24	29.40
	5	0.24	27.76
	3	0.24	27.22
8 Mb/s	16	0.30	27.93
	5	0.30	26.89
	3	0.30	26.78
6 Mb/s	16	0.37	26.63
	5	0.36	25.99
	3	0.36	25.69

Tableau 2.3 Impact du transcodage des images inter sur la qualité vidéo

Comme le transcodage par sélection en position fréquentielle des images inter a un impact moins significatif que celui des images intra sur la qualité visuelle des séquences utilisées, nous avons décidé d'encoder les flux uniquement en intra pour les manipulations suivantes.

Dans les expérimentations suivantes, nous n'avons pas utilisé de contrôle de débit lors de l'encodage des flux vidéo afin d'éviter tout saut de qualité dû à un changement de paramètre de quantification. Ainsi, nous avons utilisé un paramètre de quantification constant. Sa valeur a été déterminée à partir de la moyenne des paramètres de quantification utilisés lors de l'encodage des flux avec des images intra et inter à des débits de 12, 8 ou 6 Mb/s. Ce paramètre de quantification a été configuré avec une valeur constante entre 25 et 38, dépendant de la séquence utilisée. Ainsi, les débits présentés sont plus élevés que ceux résultant de l'encodage intra/inter-image et se situent entre 30 et 60 Mb/s. Ceci s'explique par une utilisation de l'encodeur restreinte aux images intra. Les débits sont donnés à titre indicatif.

2.4.3 Analyse PSNR

Une partie de nos résultats expérimentaux est reportée en annexe B, du Tableau B.1 au Tableau B.4, respectivement pour les séquences *Parkrun*, *Mobcal*, *Stockholm* et *Shields*. A l'entrée du transcodage, le flux original a été encodé avec un paramètre de quantification constant égal à 34, 36 ou 38. Nous avons choisi des valeurs de FP et dQP qui aboutissent à un débit transcodé similaire.

Les résultats du Tableau B.2 pour la séquence *Mobcal* sont reportés en Figure 2.9. Les paramètres de quantification utilisés sont égaux à 25, 27 et 28. La figure représente l'évolution du PSNR de la luminance en fonction du débit en sortie du transcodeur par sélection en position fréquentielle et par requantification. Le débit indiqué est en pourcentage du débit du flux initial.

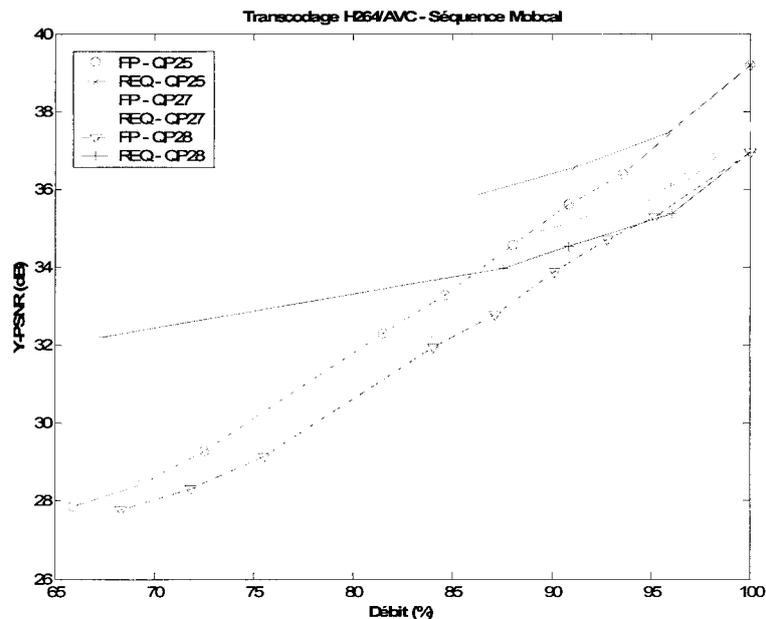


Figure 2.9 Évolution du PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et requantification –Séquence *Mobcal*.

Nous constatons que pour la séquence *Mobcal*, ainsi que les séquences *Stockholm* (Figure 2.10) et *Shields* (Figure 2.11), les résultats des mesures PSNR en luminance sont en faveur de la requantification pour des débits inférieurs à environ 95% du débit initial. Ainsi, pour la séquence *Mobcal* encodée avec un paramètre de quantification égal à 25, cette différence peut atteindre 6 dB à débit équivalent, avec les paramètres $dQP=4$ et $FP=4$. Nous noterons que sur l'ensemble des séquences, le transcoding par requantification fournit uniquement deux points de transcoding ($dQP=4$ et $dQP=3$) dans la plage de débit de 63% à 87% du débit initial. Par contre, le transcoding par sélection en position fréquentielle offre six points. Dans le but d'adapter finement le débit, ceci peut être un avantage non-négligeable pour le transcoding par sélection en position fréquentielle.

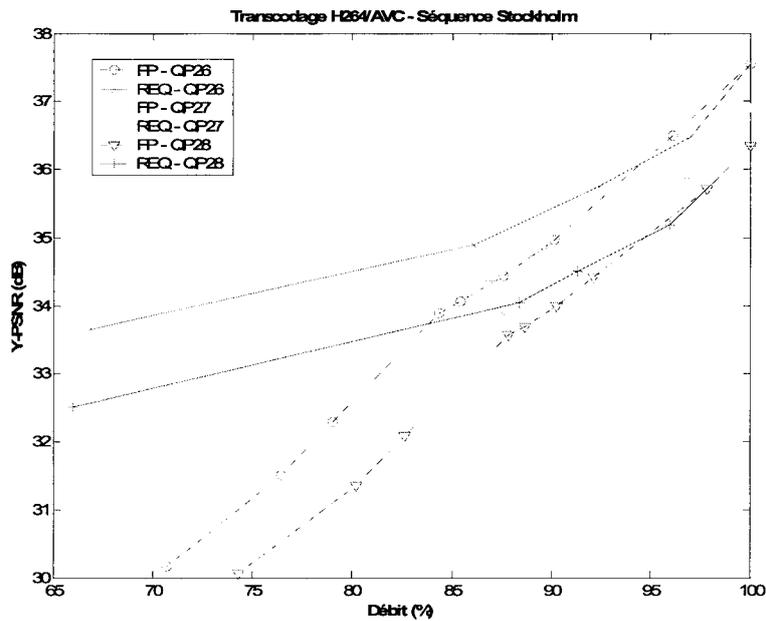


Figure 2.10 Évolution du PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et requantification –Séquence Stockholm

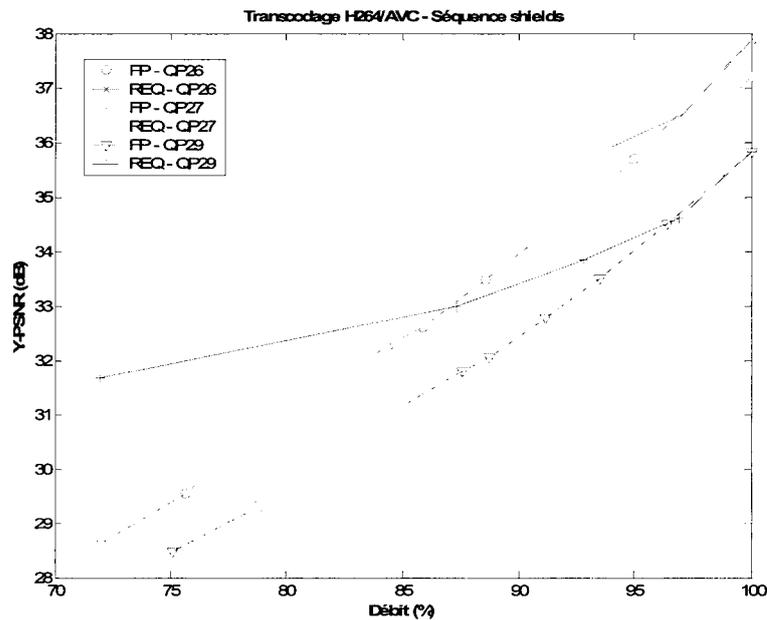


Figure 2.11 Évolution du PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et requantification –Séquence Shields

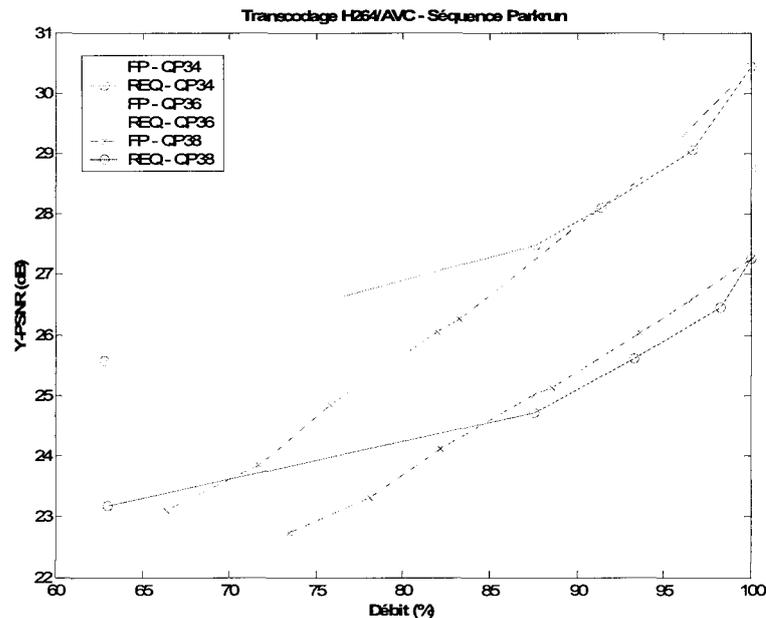


Figure 2.12 Évolution du PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et requantification –Séquence Parkrun.

D'une manière moins significative, les résultats des mesures PSNR en luminance pour la séquence *Parkrun* (Figure 2.12) sont en faveur de la requantification pour des débits inférieurs à environ 85% du débit initial.

À première vue, les résultats des mesures PSNR ne sont pas en faveur du transcodage par sélection en position fréquentielle. Et pourtant, cette différence n'est pas observée visuellement. La Figure 2.13 illustre deux images issues du transcodage par sélection en position fréquentielle (a) et par requantification (b), à débit similaire, avec respectivement les paramètres $FP=3$ et $dQP=4$. Nous pouvons constater que de nombreux effets de blocs sont présents dans le cas de la requantification, très prononcés au niveau des branchages. Le résultat du transcodage par sélection en position fréquentielle est beaucoup plus agréable malgré un léger flou et un PSNR inférieur de plus de 2 dB à celui du transcodage par requantification. Nous ajouterons que dans le cas d'un encodage intra-image, les erreurs liées à la requantification évoluent à la fois en amplitude et en position entre des images successives ce qui crée une sensation de scintillement plutôt désagréable.



(a)
FP=3
Y-
PSNR :
23.11 dB
31.6
Mb/s



(b)
dQP=4
Y-
PSNR :
25.58 dB
29.9
Mb/s

Figure 2.13 Exemple d'images issues du transcodage par sélection en position fréquentielle (a) et par requantification (b), avec respectivement les paramètres FP=3 et dQP=4

Comme le PSNR est bien connu pour être mal corrélé à l'évaluation subjective de la qualité vidéo, nous avons également décidé d'évaluer la qualité vidéo en utilisant la métrique VQM, qui est connue pour être plus proche de l'évaluation subjective.

2.4.4 Analyse selon la métrique VQM

Les résultats de la comparaison de la qualité visuelle avec la métrique VQM entre les deux transcodeurs sont présentés en Figure 2.14 pour la séquence Mobcal. Pour un flux encodé avec un paramètre de quantification égal à 25, 27 ou 28 et une réduction de débit de l'ordre de 35% maximum, la requantification fournit quatre débits intermédiaires en augmentant le paramètre de quantification de 1 à 4 (dQP=1 à dQP=4). Le transcodage par sélection en position fréquentielle offre quant à lui treize points (FP=15 à FP=3).

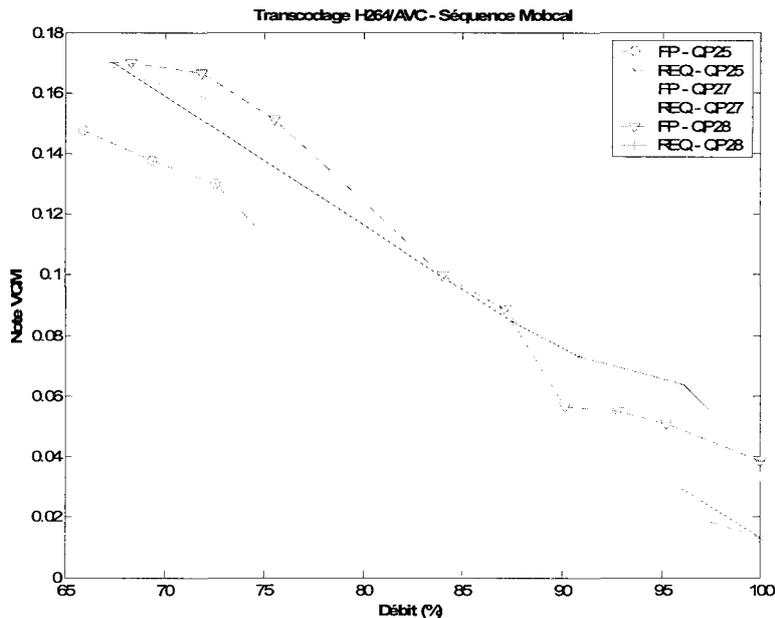


Figure 2.14 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence Mobcal.

Pour chaque couple de courbes comparées (Transcodage par FP ou requantification), nous distinguons deux zones délimitées par l'intersection des deux courbes :

- Pour une réduction de débit jusqu'à 15%, les notes VQM du transcodage par sélection en position fréquentielle (FP=15 à FP=7) sont meilleures (note VQM plus petite) qu'en requantification (dQP=1 à dQP3).
- Pour une réduction de débit entre 15% et 35%, la requantification ne fournit pas de débits intermédiaires mais uniquement un point (dQP4). Par contre, le transcodage par sélection en position fréquentielle fournit quatre points (FP=6 à FP=3). Sur l'ensemble des paramètres de quantification utilisés, le débit intermédiaire donné par le paramètre FP=6 est intéressant puisqu'il offre une qualité vidéo meilleure qu'avec le paramètre dQP=4, à un débit inférieur à celui fournit par dQP=3. Les paramètres

FP=5 à FP=3 fournissent des débits intermédiaires avec une qualité vidéo inférieure à celle de dQP=4 pour un paramètre de quantification égal à 25 ou 27. Par contre pour un paramètre de quantification égal à 28, les paramètres FP=5 à FP=3 donnent un débit inférieur à celui de dQP=4 avec une meilleure note VQM.

Les résultats de la séquence *Shields* sont illustrés en Figure 2.15. Ils sont très proches de ceux de la séquence *Mobcal*. Nous constatons également que les débits intermédiaires sont plus nombreux avec la solution de transcodage par sélection en position fréquentielle et nous retrouvons les deux zones précédemment évoquées.

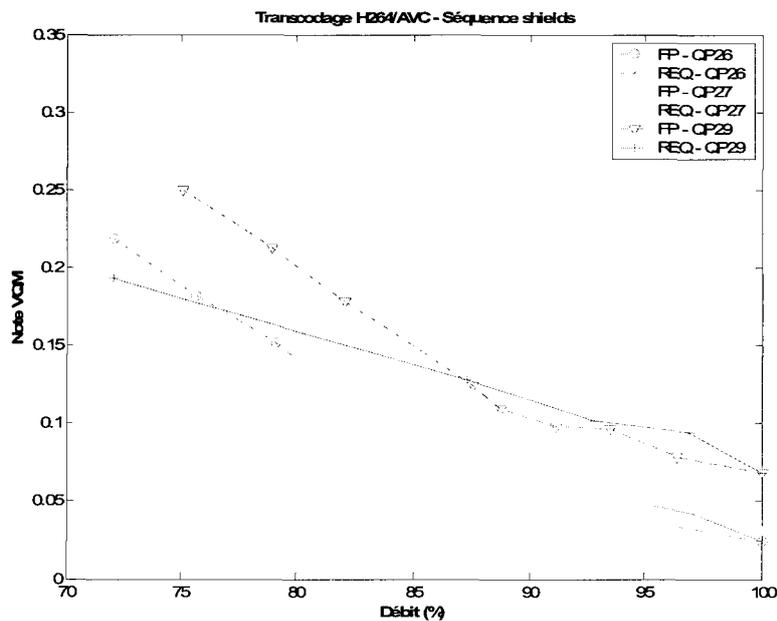


Figure 2.15 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence Shields.

La séquence *Stockholm* (Figure 2.16) donne des résultats très satisfaisants pour une réduction de débit jusqu'à 25%, toutes les notes VQM du transcodage par sélection en position fréquentielle sont meilleures (plus petites) que celles obtenues par requantification, excepté pour FP=3 avec un paramètre de quantification égal à 25.

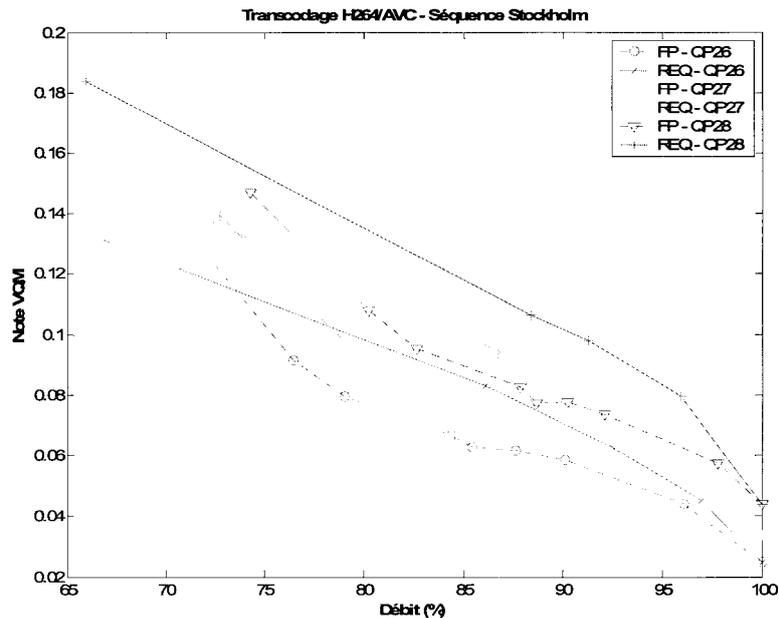


Figure 2.16 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence Stockholm.

Les résultats VQM pour la séquence *Parkrun*, encodée avec un paramètre de quantification égal à 34, 36 ou 38 sont reportés en Figure 2.17. Généralement, toutes les notes VQM de transcodage par FP sont meilleures que celles de la requantification à débit similaire. Le paramètre FP=3 donne un débit équivalent à 66% du débit initial et une note VQM de 0.19. Par la requantification, le paramètre dQP=4 donne un débit équivalent à 63% du débit initial et une note VQM critique de 0.32.

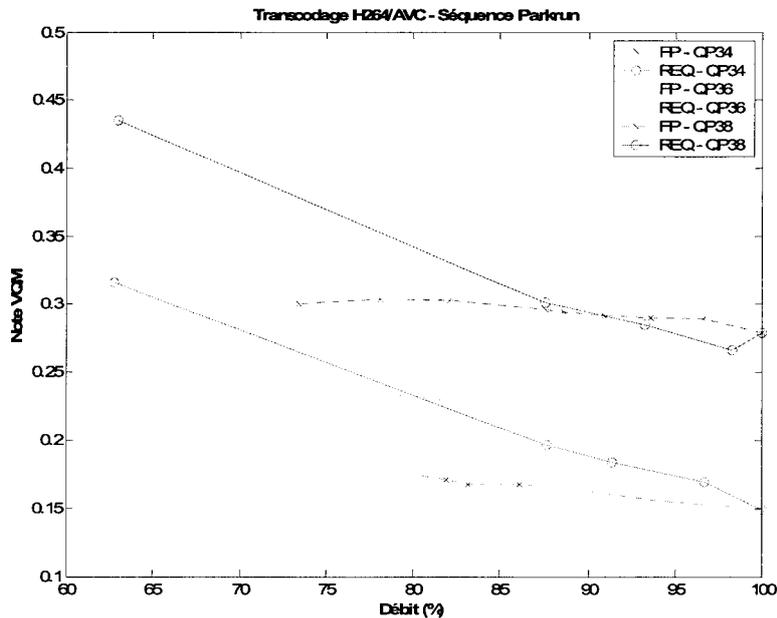


Figure 2.17 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence Parkrun.

Nous poursuivons notre analyse de la qualité vidéo des flux transcodés en essayant d'identifier et de quantifier les défauts visuels dus au transcodage. Comme mentionné dans [BER08], le logiciel *VQM pc* fournit certaines informations additionnelles concernant les éventuels défauts visuels à travers l'analyse *RCA (Root Cause Analysis)*. Cinq défauts sont pris en compte : le flou, les mouvements saccadés, le bruit global de l'image, la distorsion des blocs et les blocs en erreur. Une notation de 100% signifie que le défaut serait perçu majoritairement par les utilisateurs, 50% indique que le défaut serait perçu secondairement et 0% qu'il ne serait pas perçu.

Sur les courbes de transcodage par requantification de la séquence *Parkrun* (Figure 2.17), avec un flux initial encodé avec un paramètre de quantification égal à 38 (ou 36), nous constatons que la note VQM obtenue pour $dQP=1$ est légèrement meilleure (inférieure) que celle du flux encodé initialement. Le Tableau 2.4 dévoile les mesures RCA pour le flux initial et le flux transcodé avec $dQP=1$. Nous remarquons que le flux encodé contient déjà du flou et de la distorsion des blocs. Les mesures RCA détectent un flou moins prononcé et une distorsion des blocs plus élevée pour le flux transcodé avec $dQP=1$. Bien que cette mesure soit indépendante de VQM, nous pouvons penser que la métrique VQM accorde plus d'importance aux défauts de netteté. Ceci expliquerait les bons résultats des notes VQM obtenues avec $dQP=1$.

Chapitre 2: Solution proposée de transcodage des flux vidéo haute définition en H.264/AVC

Aussi, si le paramètre de transcodage de position en fréquence (*FP*) diminue, le logiciel détecte de plus en plus de flou sur les flux transcodés. Par contre la distorsion des blocs est constante. En augmentant le paramètre de requantification, nous constatons une évolution des défauts de distorsion des blocs et du flou.

Le Tableau 2.5 compare les résultats de l'analyse RCA pour les séquences *Stockholm* et *Ducks* en utilisant le transcodage par sélection en position fréquentielle et le transcodage par requantification à des débits similaires. Nous remarquons que le flux encodé contient uniquement du flou. L'analyse des flux transcodés par la requantification met en évidence l'apparition de flou et de distorsion des blocs. Il est intéressant de noter que le transcodage par FS produit uniquement du flou et aucun autre défaut indésirable.

Séquence : <i>Parkrun</i> (720p50) HRC: Modèle Television									
QP	dQP ou PBP	Y-PSNR (dB)	Note VQM	Flou (%)	Mouv. saccadé (%)	Bruit Global (%)	Distorsion des blocs (%)	Blocs en erreur (%)	Débit Mb/s
38		27.25	0.2794	45	0	0	18	0	29.7
38	dQP=1	26.46	0.2664	40	1	0	27	0	29.2
38	FP=10	26.58	0.2891	44	0	0	19	0	28.7
38	FP=9	26.04	0.2902	45	1	0	19	0	27.8
38	dQP=2	25.60	0.2850	36	0	0	32	0	27.7
38	FP=8	25.56	0.2916	45	0	0	19	0	27.0
38	FP=7	25.14	0.2942	45	0	0	19	0	26.3
38	dQP=3	24.71	0.3014	35	0	0	35	0	26.0
38	FP=6	25.01	0.2959	45	0	0	19	0	26.0
38	FP=5	24.11	0.3029	46	0	0	19	0	24.4
38	FP=4	23.32	0.3042	48	0	0	17	0	23.2
38	FP=3	22.74	0.3002	50	0	0	14	0	21.8
38	FP=2	21.48	0.3728	61	6	0	8	0	19.3
38	dQP=4	23.17	0.4345	68	13	0	19	0	18.7
38	dQP=5	22.87	0.4404	69	14	0	20	0	17.4

Tableau 2.4 Mesures RCA pour la séquence *Parkrun* encodée avec QP=38

Séquence	dQP ou FP	Flou	Distorsion des blocs	Note VQM	Débit (Mb/s)
<i>Stockholm</i>	-	19	0	0.09	27.0
	dQP=1	15	1	0.16	23.8
	dQP=4	34	15	0.24	18.8
	FP=6	24	0	0.12	24.6
	FP=3	38	0	0.18	21.2
<i>Ducks</i>	-	11	0	0.10	52.2
	dQP=3	13	1	0.18	45.6
	dQP=4	27	9	0.25	37.4
	FP=6	20	0	0.13	44.8
	FP=3	39	0	0.19	38.4

Tableau 2.5 Mesure des défauts perçus par l'analyse RCA pour les séquences Stockholm et Ducks

2.4.5 Approche subjective

Afin de confirmer nos précédents résultats, nous avons décidé d'organiser des tests subjectifs avec vingt six utilisateurs, experts ou non dans le domaine de la vidéo. Chaque utilisateur a effectué individuellement les tests, analysant vingt quatre séquences de test. Il était positionné à environ cinq fois la hauteur de l'écran (2,5 mètres) de l'écran 28 pouces (LCD 16/9) pour une visualisation correcte des séquences 720p.

Chaque séquence de test utilise une méthode de comparaison avec un unique stimuli et un choix forcé [WU06]. Chaque séquence de test est composée de deux séquences provenant du même flux encodé en H.264/AVC intra, transcodées par la sélection en position fréquentielle ou la requantification puis décodées, regroupées à débit similaire et affichées dans un ordre aléatoire. Le paramètre de quantification initial varie de 25 à 40 afin de couvrir une large gamme de débits. Les paramètres de transcodage varient de 1 à 6 pour la requantification (dQP) et de 3 à 9 pour la sélection en position en fréquence (FP). Le contenu des vidéos est au format 720p avec les séquences *Parkrun*, *Mobcal*, *Stockholm*, *Shields*, *Ducks* et *Crowdrun*.

La Figure 2.18 illustre les préférences des utilisateurs entre le transcodage par sélection en position fréquentielle et la requantification. Ce « barre-graphe » représente le numéro de séquence de test en abscisse. Pour chaque numéro de test, nous retrouvons en ordonnée la préférence des utilisateurs pour chaque solution de transcodage. Les utilisateurs n'ont préféré que trois séquences de tests sur vingt quatre (numéros 4, 13 et 18) utilisant la requantification. Ces séquences de tests étaient précisément *Ducks*, *Mobcal* et

Shields, utilisant respectivement un FP égal à 3 et un dQP égal à 3 ou 4. Le paramètre de quantification initial était configuré respectivement à 25, 40 et 30.

Sur l'ensemble des tests, les utilisateurs ont choisi la sélection en position fréquentielle (FP) à 74.7%. Six séquences transcodées par la FP ont été choisies par plus de 90% des utilisateurs. En définitive, nous remarquons que cette ébauche d'analyse subjective confirme la tendance des résultats de l'analyse VQM.

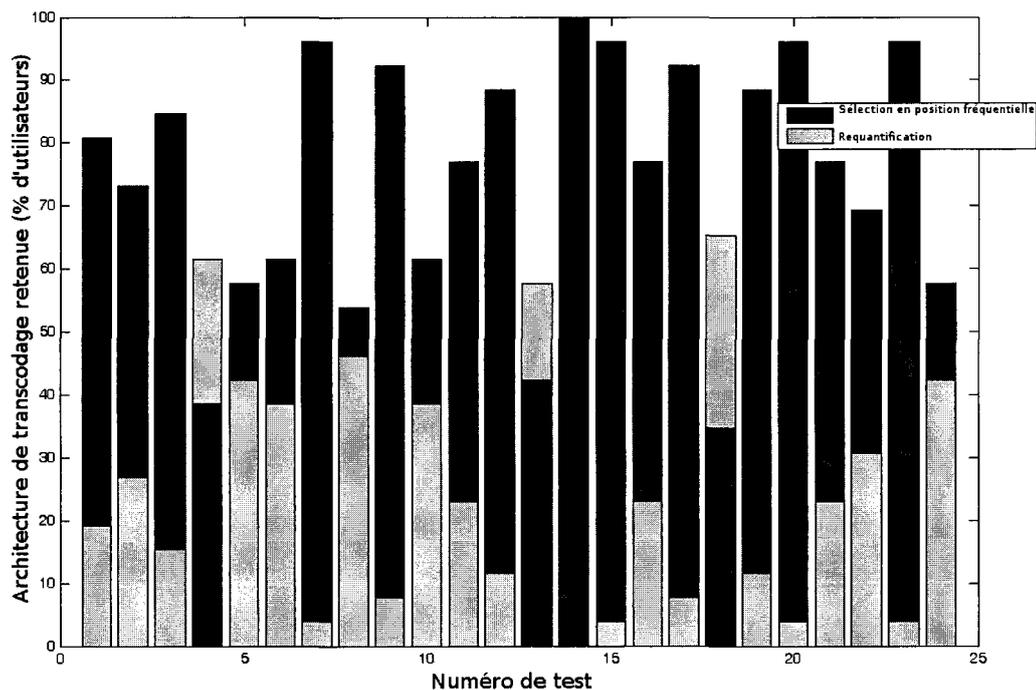


Figure 2.18 Résultats des préférences des observateurs entre le transcodage par sélection en position fréquentielle et la requantification

2.5 Comparaison de la qualité visuelle du transcodage par sélection en position fréquentielle (FP) et sélection par niveau fréquentiel (FL)

Dans cette section, nous comparons la qualité visuelle issue du transcodage par sélection par position fréquentielle (FP) à celle obtenue par sélection par niveau fréquentiel (FL), à débit équivalent [DEK11b]. La sélection par position fréquentielle supprime les coefficients résiduels suivant l'ordre inverse du codage en zig-zag. La sélection par niveau supprime les coefficients résiduels par ordre croissant de la valeur absolue de leur amplitude (cf Figure 2.5). Pour les raisons indiquées en 2.4.2.3 concernant l'impact du transcodage des

images intra sur la qualité visuelle des séquences vidéo, les expérimentations sont réalisées sur des flux H.264/AVC encodés uniquement avec des images intra.

2.5.1 Paramètres de transcodage utilisés

Comme précédemment, nous avons utilisé l'encodeur de JM 12.4 pour réaliser l'encodage H.264/AVC intra-image des flux vidéo. Les séquences vidéo utilisées sont au format 720p. Elles sont au nombre de neuf : *Crowdrun*, *Ducks*, *Intotree*, *Mobcal*, *Oldtown*, *Parkjoy*, *Parkrun*, *Shields* et *Stockholm* (illustrées en Figure B.12). D'autres tests sont disponibles en annexe B.2 pour les formats 1080p et 576p (définition standard). L'ensemble des tests a été réalisé sur 200 images à cause du nombre limité d'images de certaines séquences. Ces séquences sont disponibles en YUV sous-échantillonné en 4:2:0. L'encodeur a été configuré avec le *main profile* du standard H.264/AVC et l'encodage entropique CAVLC.

Dans les paragraphes suivants, les paramètres de transcodage représentant la position fréquentielle (FP) ou le niveau fréquentiel (FL) peuvent varier théoriquement de 1 (uniquement le coefficient DC) à 16 (aucun changement).

Dans les sous-sections suivantes, nous comparons les performances de notre architecture de transcodage par sélection en position fréquentielle à celles obtenues par sélection en niveau fréquentiel en termes de réduction de débit et de qualité vidéo. Nous utiliserons également les métriques PSNR et VQM. Les deux transcodeurs utilisent une architecture très proche, basée sur le mode « reuse ». Le module de sélection par position fréquentielle, tout comme celui de sélection par niveau fréquentiel, est appliqué uniquement aux coefficients de la luminance.

2.5.2 Analyse PSNR

Sur l'ensemble des séquences vidéo 720p, nous constatons une réduction de débit maximale de 58% avec la séquence *Stockholm* pour un paramètre de quantification QP=18 et un paramètre de transcodage FP=1 ou FL=1. Pour l'ensemble des séquences, la réduction de débit décroît en augmentant le paramètre de quantification. Pour une réduction de débit supérieure à 10%, nous remarquons la présence d'au moins 30% de plus de débits intermédiaires pour le transcodage par sélection en position fréquentielle que pour le transcodage en sélection en niveau fréquentiel. Cet écart a tendance à augmenter avec des paramètres de quantification plus élevés (QP=30 et QP=36). Ainsi, dans cette plage de réduction de débit supérieure à 10%, pour la séquence *Ducks* encodée avec un QP=30, huit

débites intermédiaires sont disponibles par la sélection en position fréquentielle contre trois pour la sélection en niveau fréquentiel.

Nous constatons également pour certaines séquences (*Mobcal*, *Oldtown*, *Shields*, *Stockholm*) un effet de bruit de type *mosquito* pour un paramètre FL=2, voire FL=3. La Figure 2.19 illustre le flux *Oldtown* encodé avec un paramètre de quantification égal à 18 (a), puis transcodé par sélection en position fréquentielle (b) ou par niveau fréquentiel (c). L'effet *mosquito* est présent en (c), il se traduit par une alternance de disparition et d'apparition d'une ou plusieurs fréquences. Bien qu'il apparaisse localement dans l'image, cet effet de scintillement, localisé le long des contours des objets, attire l'attention de l'observateur et devient fortement gênant. A débit équivalent, le transcodage par sélection en position fréquentielle n'introduit pas d'effet de ce type.

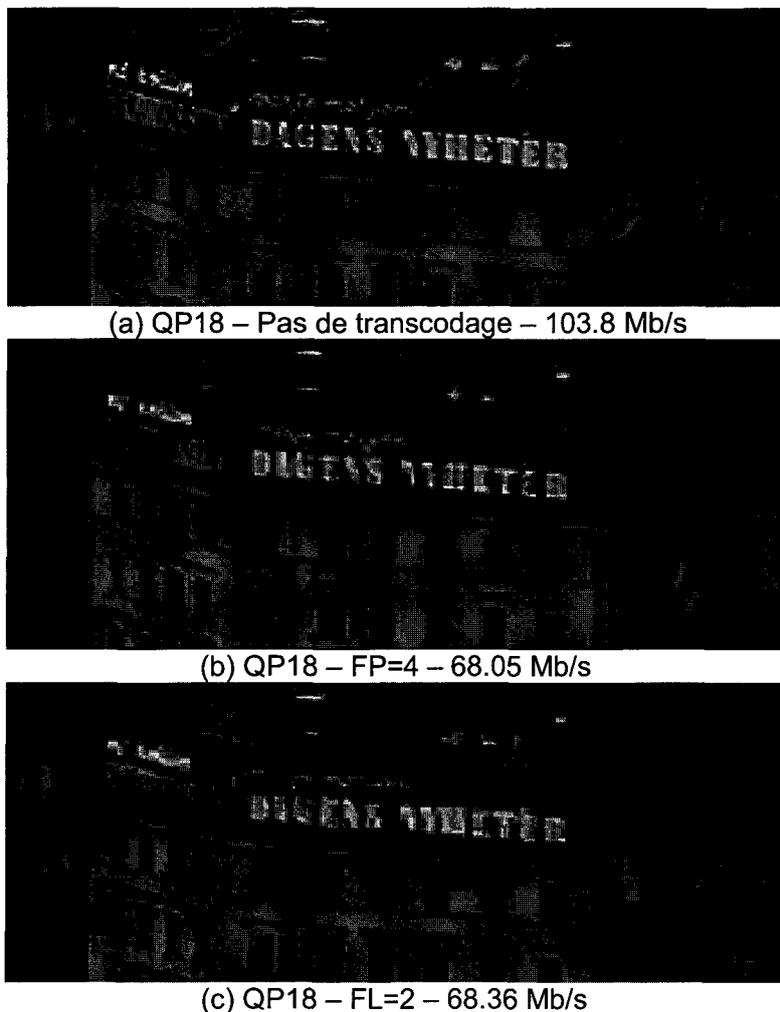


Figure 2.19 Illustration du résultat de transcodage d'une partie de la séquence *Oldtown*

Les mesures PSNR sont en faveur de la sélection par position fréquentielle dans environ 70% des cas. Pour les quatre paramètres de quantification utilisés (QP=18, 24, 30 ou 36), le PSNR après transcodage des séquences *Intotree*, *Oldtown*, *Shields* et *Stockholm* est toujours en faveur de la sélection en position fréquentielle. Cependant les résultats PSNR des séquences *Ducks*, *Mobcal*, *Parkjoy*, *Parkrun* et *Crowdrun* sont meilleurs pour :

- le transcodage par sélection en niveau fréquentiel pour les paramètres de quantification QP 18 et QP=24
- le transcodage par sélection en position fréquentielle pour les paramètres de quantification QP=30 et QP=36

A débit similaire, nous constatons un écart atteignant jusque 2 dB entre les deux types de transcodage. Par exemple, la séquence *Ducks* encodée avec un paramètre de quantification égal à 18, sur la Figure 2.20, avec FP=8 (128,7Mb/s – PSNR-Y= 32,72dB) et FL=4 (125,9Mb/s – PSNR-Y=33,17dB). Ou encore, la séquence *Oldtown* encodée avec un paramètre de quantification égal à 18, sur la Figure 2.21, avec FP=6 (76,4Mb/s – PSNR-Y= 35,20dB) et FL=3 (77,0Mb/s – PSNR-Y=33,98dB).

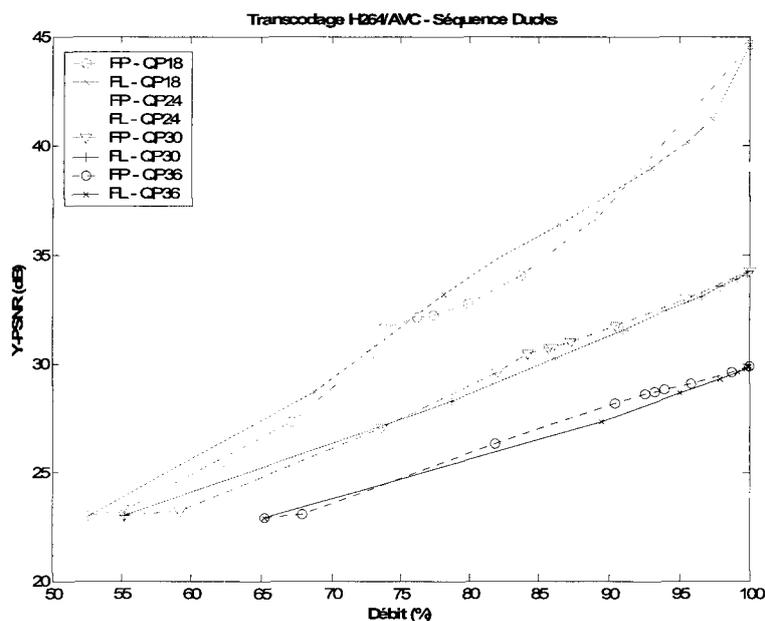


Figure 2.20 Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Ducks*.

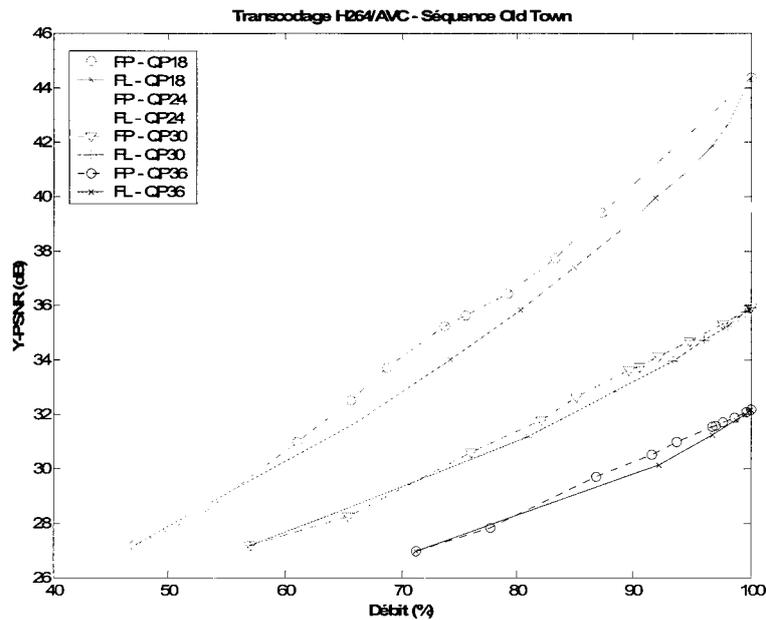


Figure 2.21 Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Oldtown*.

2.5.3 Analyse VQM

Les résultats VQM sont illustrés par les figures 3.22 et 3.23 pour les séquences *Oldtown* et *Ducks* respectivement. Nous constatons que similairement à l'analyse PSNR, les notes VQM pour la séquence *Oldtown* sont en faveur de la sélection en position fréquentielle avec les paramètres de quantification QP=18 et QP=36.

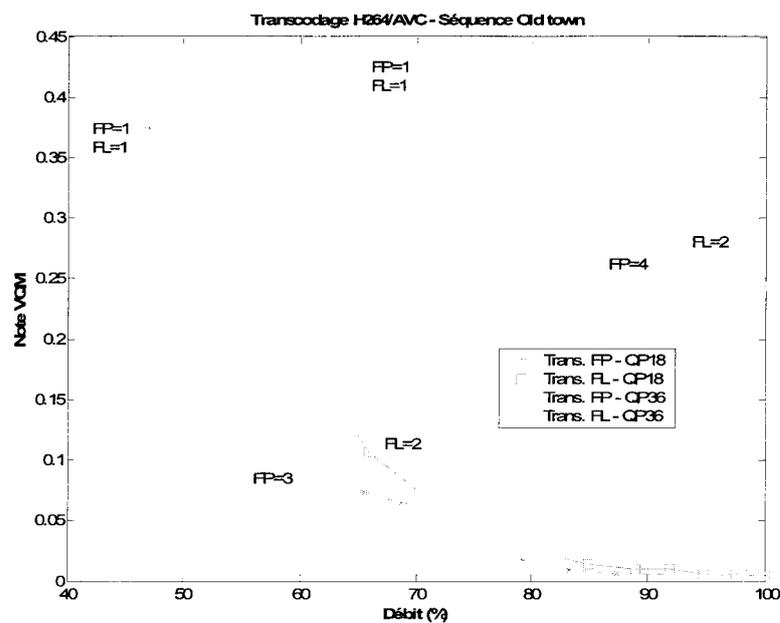


Figure 2.22 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *OldTown*.

Dans la figure 3.23, dédiée à la séquence *Ducks*, les résultats VQM pour le paramètre de quantification QP=18 ne suivent pas la tendance du PSNR. Ainsi, pour une réduction de débit supérieure à 30% les résultats VQM sont favorables à la sélection en position fréquentielle, utilisant deux débits intermédiaires (FP=2 et FP=3). Dans le cas d'une réduction de débit inférieure, les résultats VQM sont très proches pour les deux types de transcodage.

Similairement, pour un paramètre de quantification QP=36, le transcodage par sélection en position fréquentielle est retenu pour une réduction de débit supérieure à 12%. Dans le cas d'une réduction de débit inférieure, les résultats VQM sont également très proches pour les deux types de transcodage.

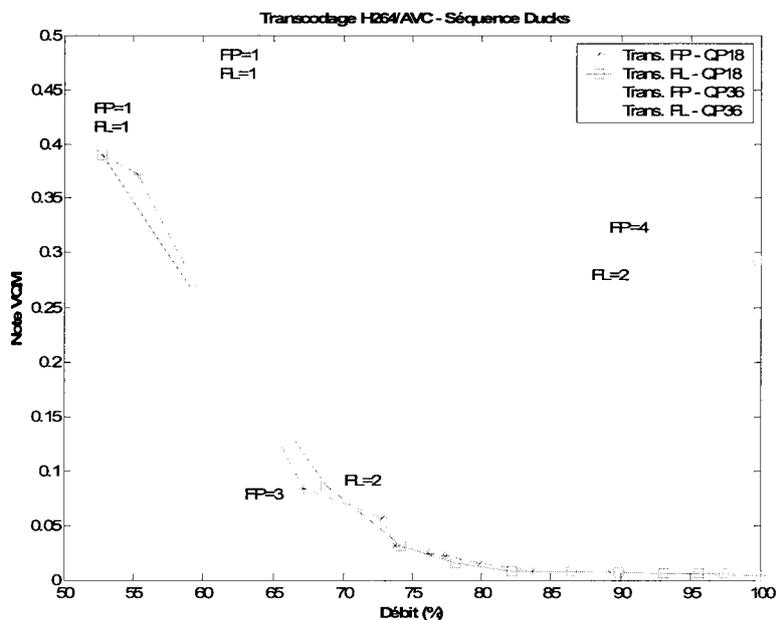


Figure 2.23 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Ducks*.

L'ensemble des résultats de transcodage des séquences 720p est disponible en annexe B.2. Nous trouverons également en annexe les résultats de nos expérimentations pour des séquences vidéo au format 1080p et définition standard. Les résultats sont partagés pour les séquences en définition standard mais ils sont nettement en faveur du transcodage par sélection en position fréquentielle pour les séquences au format 1080p.

Dans cette section, nous avons constaté par les analyses PSNR et VQM que, selon la séquence vidéo et le paramètre de quantification, le transcodage par sélection en position

fréquentielle donnait des résultats similaires ou meilleurs que ceux obtenus lors du transcodage par sélection en niveau fréquentiel.

Les résultats du transcodage par sélection en position fréquentielle étaient meilleurs lorsque :

- les séquences sont en 1080p
- les séquences 720p et SD sont encodées avec un taux de compression élevé

Dans les autres cas, les résultats des analyses PSNR et VQM sont très proches. Lors de l'analyse PSNR, nous avons remarqué quelques séquences favorables au transcodage par sélection en niveau fréquentiel pour des paramètres de quantification QP=18 et QP=24. Cependant, ces résultats ne sont pas confirmés par l'analyse VQM pour l'ensemble des points de transcodage mais uniquement sur des faibles réductions de débit. Aussi, le transcodage par sélection en position fréquentielle apporte des débits intermédiaires ponctuels et non pas massivement regroupés dans les faibles réductions de débit. Ceci facilite l'adaptation du débit du codage source au débit du canal. Enfin, le transcodage par sélection en niveau fréquentiel peut provoquer un effet gênant, de type mosquito noise, pour des réductions de débit relativement élevées.

2.6 Estimation du débit de flux H.264/AVC transcodés par sélection en position fréquentielle

L'un des objectifs du projet TOSCANE est d'étendre la zone d'éligibilité des abonnés au service de télévision sur IP. Nous verrons au chapitre 4 qu'il est possible d'estimer par simulation le débit de transmission en réception d'une ligne ADSL pour un taux erreur binaire (TEB) très faible (de l'ordre de 10^{-11}) après décodage des codes correcteurs d'erreurs. Après avoir eu la connaissance de ce débit, l'étape suivante consiste à utiliser notre architecture de transcodage afin d'adapter le débit du flux vidéo à celui du canal de transmission.

Dans cette section, nous allons estimer le débit du flux transcodé dans le but de l'adapter à la bande passante du canal [DEK10]. Afin de réaliser ceci, nous avons besoin de modéliser l'évolution du débit en fonction des paramètres de transcodage pour les macroblocs intra- et inter-image, respectivement, FP_{intra} et FP_{inter}. De nombreux modèles précis de débits ont été présentés dans la littérature [LEI03] mais ils s'appuient souvent sur des algorithmes de contrôle de débit avancé, utilisant le transcodage par requantification. Ci-dessous, dans le cas du transcodage par sélection en position fréquentielle, nous montrons qu'un modèle de débit simple peut être utilisé. Il s'agit d'un avantage considérable pour une éventuelle implémentation temps réelle de l'architecture de transcodage.

2.6.1 Analyse du débit en fonction des paramètres de transcodage (FPintra et FPinter)

Nous avons cherché à estimer les paramètres de transcodage (FPintra et FPinter) pour un débit spécifique en utilisant dix séquences haute définition en 720p (*Parkrun, Mobcal, Stockholm, Crowdrun, Intotree, Parkjoy, Oldtown...*), encodées à 8Mb/s, utilisant un groupe d'images ou GOP égal à 12, reposant sur une structure de type IBBBP.

La courbe moyenne de l'évolution du débit en fonction des paramètres de transcodage FPintra et FPinter pour l'ensemble des séquences vidéo est illustrée en Figure 2.24. Comme le montre cette figure, nous constatons une évolution stable du débit en fonction des paramètres de transcodage. Cette courbe moyenne nous donne un écart type relatif moyen de 4,13%, avec un minimum de 1,44% et un maximum de 12,22%. Ceci nous permet de prédire fiablement l'évolution du débit.

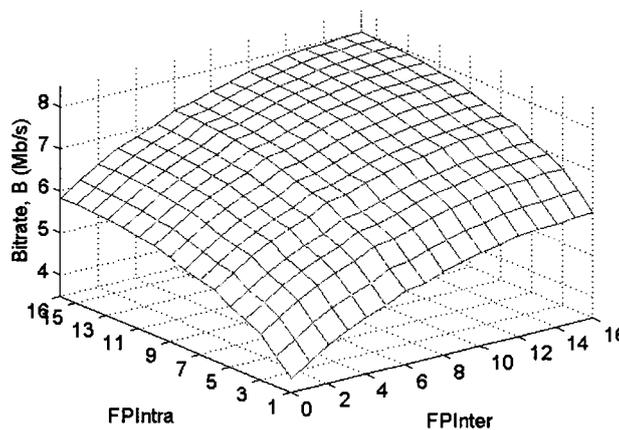


Figure 2.24 Courbe moyenne illustrant l'évolution du débit (B) en fonction des paramètres de transcodage FPintra et FPinter.

2.6.2 Modélisation du débit

Les paramètres de transcodage FPintra et FPinter ont été utilisés pour transcoder les macroblocs intra- ou inter-image, en fonction du type du macrobloc. Ces paramètres sont indépendants et peuvent être utilisés pour exprimer la fonction mathématique du débit d'un flux vidéo. Le débit modélisé, appelé \hat{B} , peut être exprimé de la manière suivante :

$$\hat{B} (FPintra, FPinter) = f(FPintra) + g(FPinter) \quad (1)$$

où: $\hat{B} (16, 16) = 8$ Mb/s (débit initial); $f(x)$ et $g(x)$ sont les fonctions croissantes des paramètres FPintra et FPinter, respectivement.

La Figure 2.24 nous dévoile une symétrie au niveau du débit qui dépend des paramètres FP_{intra} et FP_{inter} . Ces paramètres ont une influence similaire sur le débit vidéo en sortie du transcodeur. Alors, nous pouvons considérer que les fonctions $f(x)$ et $g(x)$ ont la même expression mathématique :

$$f(x) = g(x) \quad (2)$$

La Figure 2.25 illustre l'évolution du débit du modèle proposé. Les fonctions $B(FP_{intra}, 16)$ et $B(16, FP_{inter})$ dépendent respectivement des paramètres de transcodage FP_{intra} et FP_{inter} .

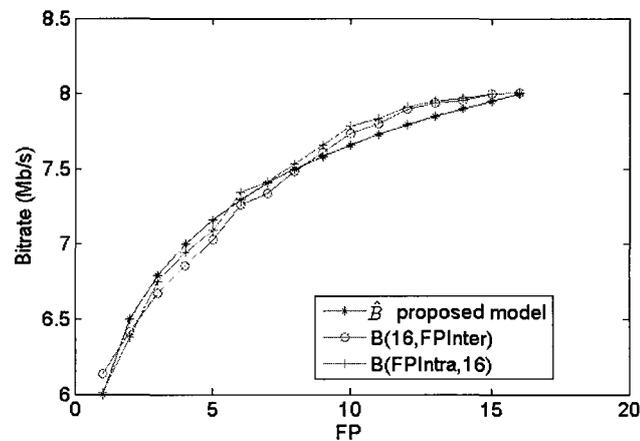


Figure 2.25 Évolution du débit du modèle proposé, $B(FP_{intra}, 16)$ et $B(16, FP_{inter})$, dépendant du paramètre FP_{intra} ou FP_{inter} .

En utilisant les relations (1) et (2), les fonctions $B(FP_{intra}, 16)$ et $B(16, FP_{inter})$ peuvent être modélisées par la même fonction $f(x)$, telle que :

$$\hat{B}(x, 16) = \hat{B}(16, x) = f(x) + f(16) \quad (3)$$

Dans [FER08], un modèle empirique simple a été développé afin de favoriser une implémentation matérielle peu complexe, sans une multitude de tests d'encodage ou de contrôle de débit avancé. Le modèle repose sur une relation logarithmique entre le paramètre de quantification et le débit. Dans notre cas, le débit évolue également dans une relation linéaire du logarithme en base 2 du paramètre de transcodage FP. Nous avons constaté que la division par 2 du paramètre de transcodage FP aboutissait à une réduction de débit de l'ordre de 0,5 Mb/s. Ainsi, la fonction $f(x)$ peut être exprimée de la façon suivante :

$$f(x) = \alpha \log_2(x) + \beta \quad (4)$$

avec: $\alpha = 0.5$

Nous pouvons en déduire β avec :

$$\hat{B}(16, 16) = 4 + 2\beta = 8 \text{ Mb/s, donc : } \beta = 2 \quad (5)$$

Par conséquent, le débit peut être modélisé de la manière suivante :

$$\hat{B}(FP_{Intra}, FP_{Inter}) = 0.5 \log_2 (FP_{Intra} FP_{Inter}) + 4$$

où : $1 \leq FP_{Intra} \leq 16$ et $1 \leq FP_{Inter} \leq 16$ (6)

Nous constatons un écart type relatif moyen de 4,12% entre le modèle et les courbes de débit, avec un minimum de 1,47% et un maximum de 11,86%. Le modèle proposé est simple et relativement fiable. Il nous permet de minimiser la complexité de notre architecture de transcodage en évitant d'utiliser des algorithmes de contrôle de débit avancés lors du transcodage.

2.7 Conclusion

Dans ce chapitre, nous avons présenté une architecture de transcodage H.264/AVC innovante. Permettant à la fois le transcodage des images intra et inter, cette architecture fondée sur la sélection fréquentielle des coefficients résiduels transformés a une complexité relativement réduite favorisant une implémentation temps réelle. L'architecture est de type boucle fermée pour les macroblocs intra et de type boucle ouverte pour les macroblocs inter. La qualité vidéo du transcodage par sélection en position fréquentielle a été ensuite comparée à celle d'une solution utilisant la requantification des coefficients résiduels transformés. Nous avons pu constater que notre architecture donnait bien souvent de meilleurs résultats en termes de qualité vidéo par la métrique VQM et l'analyse subjective.

Ensuite, nous avons implémenté deux types de transcodage par sélectivité fréquentielle : la sélection en position fréquentielle et la sélection en niveau fréquentiel. Nous avons comparé les performances de ces deux types de sélectivité fréquentielle en termes de complexité, de débit et de qualité vidéo. Nous avons décidé de retenir le transcodage par sélection en position fréquentielle pour sa faible complexité, sa répartition des débits intermédiaires et la qualité vidéo justifiées par les mesures PSNR et VQM.

Pour finir, nous avons proposé un modèle de débit de transcodage par sélection en position fréquentielle en fonction des paramètres de transcodage (FP_{Intra} et FP_{Inter}). Il permet d'adapter facilement, sans coût de calcul supplémentaire, le débit du flux vidéo transcodé à la bande passante du canal.

Chapitre 3: Solution proposée de transcodage à complexité réduite des flux H.264 SVC en échelonnabilité spatiale

3.1 Introduction

L'échelonnabilité spatiale permet de disposer dans un même flux vidéo de plusieurs versions d'une source vidéo en termes de définition spatiale. Dans ce type d'échelonnabilité, la couche de base possède la définition spatiale la plus basse de la source vidéo. L'ajout d'une couche d'amélioration a pour incidence d'augmenter la définition spatiale du flux vidéo. Ainsi, le débit nécessaire au décodage de la couche de base est le plus faible. Pour décoder une couche d'amélioration, il est nécessaire de disposer des données de cette couche tout comme celles des couches inférieures. Par contre, un inconvénient majeur de l'échelonnabilité spatiale dans H.264 SVC est de ne pas pouvoir fournir davantage de choix de débits vidéo que le nombre de couches présentes dans le flux.

Dans ce chapitre, nous proposons une extension de l'architecture de transcodage des flux H.264/AVC, présentée au chapitre précédent, au cas des flux H.264 SVC en échelonnabilité spatiale. L'objectif est de fournir des débits vidéo intermédiaires des flux tout en conservant la définition spatiale la plus haute. Le principe de cette solution de transcodage consiste à sélectionner par position fréquentielle les coefficients résiduels de la couche d'amélioration haute par leur position en fréquence. Afin d'assurer un traitement temps réel des flux vidéo, l'architecture de transcodage doit être de faible complexité. Alors, nous proposons de transposer la partie en boucle ouverte de l'architecture de transcodage H.264/AVC, issue de **[DEC06b]**, au cas H.264 SVC. Ainsi, nous pourrions disposer de débits intermédiaires situés entre le débit de l'ensemble des couches inférieures et le débit total du flux vidéo, assurant une qualité progressive de l'image avec la définition la plus élevée.

À notre connaissance, il n'existe pas de solution similaire dans la littérature scientifique. Il n'est donc pas possible de se comparer à une solution existante. Cependant, nous proposons de comparer les flux vidéo transcodés à ceux de la couche d'amélioration haute en pleine qualité et de la couche immédiatement inférieure (sur-échantillonnée spatialement) en termes de réduction de débit et de PSNR. Un type d'application de ce transcodage est la transmission vidéo. Selon les caractéristiques du canal de transmission, le principe consiste à adapter le débit vidéo à celui du canal afin de transmettre la vidéo dans la meilleure définition mais avec une qualité légèrement inférieure, tout en restant supérieure à celle de

la couche inférieure sur-échantillonnée. Les applications visées sont variées, allant notamment de la télévision sur IP à la vidéo-surveillance. Un exemple d'application sera détaillé au chapitre 4.

Le chapitre est organisé de la manière suivante. Dans un premier temps nous rappelons les modes de prédiction intervenant dans l'échelonnabilité spatiale. Puis nous proposons et évaluons les performances d'une première architecture de transcodage appliquant le traitement sur l'ensemble des macroblocs de la couche d'amélioration. Enfin, dans l'objectif de minimiser la complexité de notre architecture de transcodage SVC, nous proposons et évaluons deux variantes de cette architecture. La première consiste à restreindre le traitement aux macroblocs de luminance. Quant à la seconde, elle consiste à restreindre le traitement aux macroblocs utilisant la prédiction inter-image. Nous finissons par une conclusion générale et les perspectives offertes par cette architecture de transcodage.

3.2 Échelonnabilité spatiale des images intra et inter dans H.264 SVC

Nous avons déjà présenté l'échelonnabilité spatiale de H.264 SVC au chapitre 1 (section 1.3.2.2). Nous revenons ici un peu plus en détail sur certaines notions de H.264 SVC nécessaires à la compréhension de l'architecture de transcodage proposée. Tout d'abord, une couche de base et une ou plusieurs couches d'amélioration sont présentes dans ce type de flux vidéo échelonnable. La couche de base représente la version à la plus basse définition de la source vidéo. Le décodage des couches d'amélioration requiert des informations présentes dans la couche de base, voire d'une couche inférieure à la couche d'amélioration courante. Une couche d'amélioration a vocation d'améliorer le rendu de l'image décodée en augmentant sa définition via une échelonnabilité spatiale.

Une couche de base comme une couche d'amélioration peut être composée d'une ou plusieurs NALU. L'appartenance d'une NALU à une couche est mentionnée par l'intermédiaire d'un *dependency identifier* (nommé *D*). Notons que la définition spatiale ne doit pas diminuer d'une couche à une couche supérieure. Aussi, le terme *access unit* désigne l'ensemble des couches nécessaires au décodage d'une image dans sa définition spatiale la plus haute.

3.2.1 Prédiction des macroblocs

Dans un flux H.264 SVC en échelonnabilité spatiale, un macrobloc de la couche d'amélioration peut être codé par une prédiction de type intra-couche ou inter-couche.

La prédiction intra-couche se réalise conformément à H.264/AVC en utilisant les images de cette même couche d'amélioration (cf 1.2.2). Nous distinguons deux cas :

- Soit le macrobloc est codé par une prédiction intra-image, nous parlerons de prédiction intra-couche intra-image. Ce type de prédiction est illustré dans la partie supérieure de la Figure 3.1 (couche d'amélioration) par une flèche pleine notée (a). La prédiction intra-image se réalise conformément à H.264/AVC en utilisant les échantillons de luminance ou de chrominance des blocs voisins. Il est possible d'utiliser 9 modes de prédiction en mode INTRA_4X4 et 4 modes de prédiction en mode INTRA_16x16.
- Soit le macrobloc est codé par une prédiction inter-image, nous parlerons de prédiction intra-couche inter-image. La prédiction inter-image se réalise conformément à H.264/AVC en utilisant les échantillons des blocs situés dans les images précédentes ou suivantes (flèches pleines notées (b) de la Figure 3.1).

La prédiction inter-couche utilise un bloc de la couche inférieure comme prédiction du macrobloc courant. Ce type de prédiction est signalé dans le flux vidéo par le mode *base layer*. Nous distinguons également deux cas :

- Soit le macrobloc est codé par une prédiction intra-image, nous parlerons de prédiction inter-couche intra-image (flèches en pointillés notées (c) dans la Figure 3.1). Il s'agit d'un nouveau mode de prédiction intra appelé INTRA_BL. Ainsi, le bloc correspondant dans la couche de référence est sur-échantillonné et va servir de signal de prédiction.
- Soit le macrobloc est codé par une prédiction inter-image, nous parlerons de prédiction inter-couche inter-image (flèche notée (d) dans la Figure 3.1). La prédiction inter-image est réalisée en utilisant les échantillons des blocs situés dans les images précédentes ou suivantes de la couche d'amélioration inférieure.

Aussi, notons l'existence de la prédiction résiduelle inter-couche dont le principe consiste à calculer l'erreur entre le résidu de la couche d'amélioration et celui-ci sur-échantillonné de la couche de référence. Le résidu de la couche d'amélioration peut être codé auparavant par prédiction intra-couche ou inter-couche.

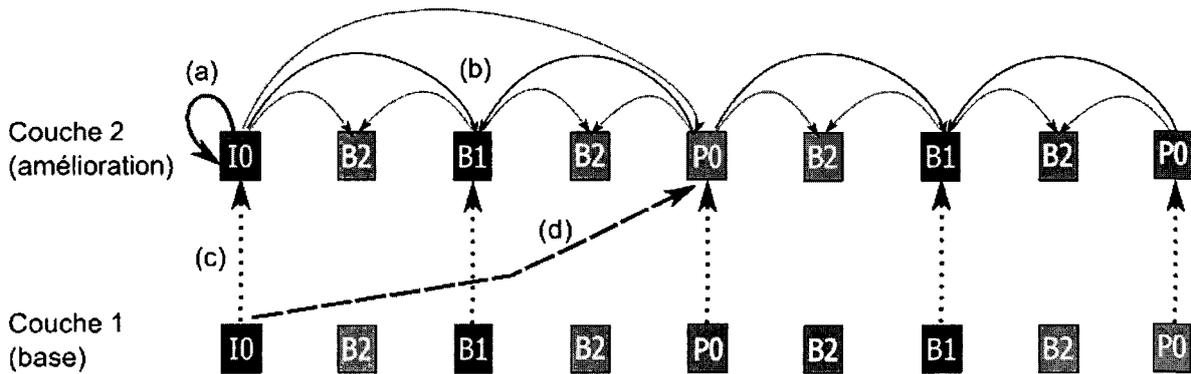


Figure 3.1 Illustration de l'échelonnabilité spatiale utilisant la prédiction intra-couche (flèches pleines) et la prédiction inter-couche (flèches en pointillés)

Dans le cas où la couche d'amélioration à transcoder contient des macroblocs utilisant la prédiction intra-couche intra-image, un effet de drift comparable à celui constaté dans le chapitre 3 est prévisible. La Figure 2.3 illustre ce phénomène en montrant une image I (Intra) du flux à l'entrée du transcodeur (a) et une image résultant d'un traitement sur l'ensemble des macroblocs de la couche d'amélioration haute (b). Un effet de drift apparaît horizontalement et verticalement. Il est un peu moins prononcé qu'en H.264/AVC car nous verrons en section 3.2.3 que les macroblocs INTRA_BL sont également utilisés et sont bien souvent majoritaires dans cette couche d'amélioration. Ainsi, l'effet de drift est stoppé par l'apparition de ce type de macrobloc dans l'ordre de décodage (identique à l'ordre de transcodage) des macroblocs.

Notons que le transcodage des macroblocs INTRA_BL ne crée pas d'effet de drift de type intra-image puisque la référence est située dans la couche inférieure et demeure inchangée. Par contre, nous verrons par la suite qu'un effet de drift que nous pouvons supposer négligeable existe lorsqu'un macrobloc INTRA_BL est modifié et utilisé comme référence lors d'une éventuelle prédiction inter-image. Plus généralement, les macroblocs utilisant la prédiction intra-couche inter-image sont sujets à cet effet de drift négligeable, similairement à H.264/AVC.

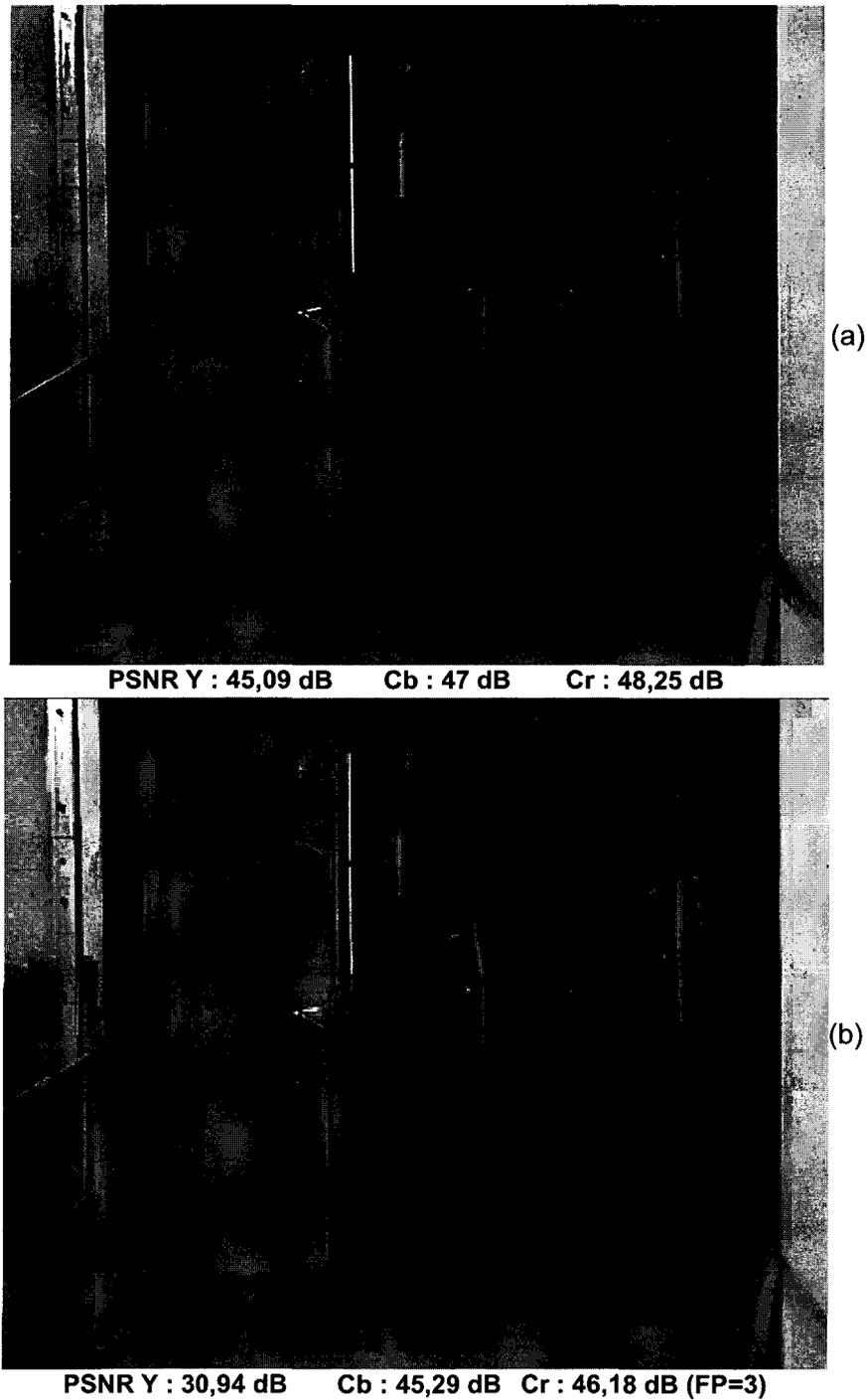


Figure 3.2 Illustration d'une image du flux à l'entrée du transcodeur (a) et de l'effet de drift dû au transcodage en boucle ouverte de macroblobs utilisant la prédiction intra-couche intra-image sur la couche d'amélioration haute (b)

3.2.2 Configuration des encodages H.264 SVC en échelonnabilité spatiale

L'ensemble des encodages a été réalisé avec l'encodeur JSVM (version 9.15) configuré en échelonnabilité spatiale avec une couche de base et une seule couche d'amélioration (bien qu'il est soit possible d'utiliser plusieurs couches d'amélioration). La couche de base a une définition spatiale au format CIF (352x288 pixels). La couche d'amélioration a une définition spatiale au format 4CIF (704x576 pixels). Nous avons utilisé les séquences *City*, *Crew*, *Harbour* et *Soccer*. Chaque séquence a été encodée avec un GOP de 12 images (IbBbPbBbPbBbI) et un paramètre de quantification fixé, identique pour la couche de base et la couche d'amélioration. Afin d'évaluer les techniques utilisées avec l'évolution du débit et sur une large plage de qualité, chaque séquence est encodée avec un paramètre de quantification égal à 18, 24, 30 puis 36. Nous précisons que le codage entropique utilisé est le CAVLC.

3.2.3 Statistiques sur l'utilisation du type de codage utilisé pour les macrobloccs de type intra-image de la couche d'amélioration haute

Comme dans le cas du transcodeur H.264/AVC, le traitement sera réalisé en fonction du type de macrobloc. Cette étude va nous permettre d'estimer la proportion de macrobloccs utilisant la prédiction intra-image pour chaque type de codage. Nous considérons uniquement les macrobloccs situés dans les NALU de la couche d'amélioration. Le Tableau 3.1 regroupe l'ensemble de ces résultats expérimentaux.

La Figure 3.3 synthétise la proportion moyenne du type de codage des macrobloccs utilisés sur l'ensemble des séquences encodées.

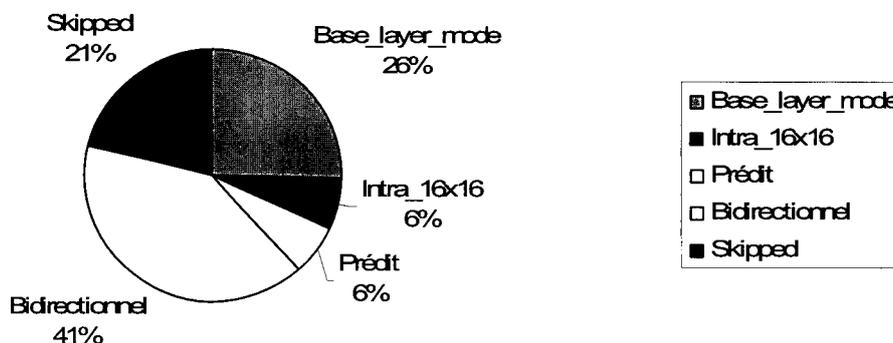


Figure 3.3 Types de macrobloccs sur l'ensemble des slices (couche d'amélioration)

Nous constatons que les macrobloccs bidirectionnels sont les plus utilisés ce qui est logique puisque les images B (composées majoritairement de macrobloccs bidirectionnels)

représentent les 3/4 des images du GOP (IbBbPbBbPbBbI). Les macroblochs en mode *base layer* utilisent la prédiction inter-couche et sont soit de type intra-image ou soit de type inter-image. Sur l'ensemble des images, les macroblochs codés par la prédiction intra-image sont soit de type INTRA_16x16 (6% - prédiction intra-couche) ou soit de type INTRA_BL (inclus dans les 26% du mode *base layer* – prédiction inter-couche). Nous remarquons que les macroblochs utilisant la prédiction intra-couche intra-image (prédiction intra de H.264/AVC) sont uniquement de type INTRA_16x16. Ceci s'explique par notre choix d'échelonnabilité spatiale. Ces macroblochs ont un bloc correspondant (ici de taille 8x8 pixels) dans la couche de base et ne peuvent être partitionnés autrement qu'en 16x16 pixels.

Nous poursuivons notre analyse dans les slices de type intra avec la Figure 3.4. Les macroblochs utilisant la prédiction inter-couche intra-image (Intra – mode *base layer*) sont majoritaires (avec 71%) par rapport aux macroblochs utilisant la prédiction intra-couche intra-image (INTRA_16x16) représentés à 29%.

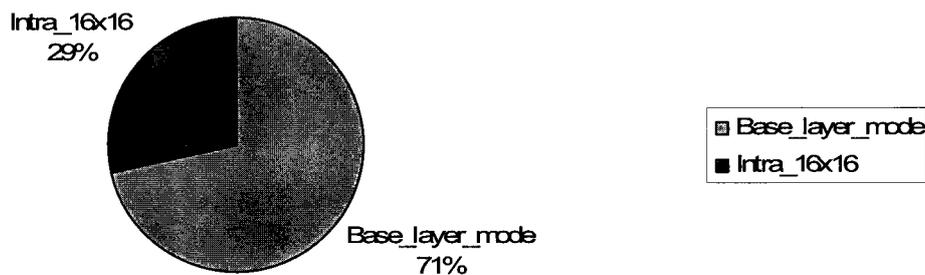


Figure 3.4 Types de macroblochs dans les slices de type Intra (couche d'amélioration)

Dans le cas des slices de type P (Figure 3.5), nous constatons que les deux modes de codage INTRA_16x16 et INTRA_BL sont quasiment autant utilisés (respectivement 22% et 25%).

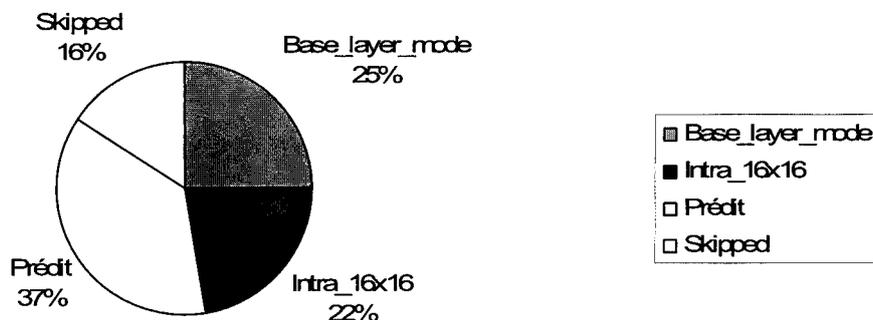


Figure 3.5 Types de macroblochs dans les slices de type P (couche d'amélioration)

Dans le cas des slices de type B (Figure 3.6), nous constatons que le mode de codage INTRA_16x16 n'est pas utilisé. Le mode de codage INTRA_BL représente 21 % des macroblochs présents dans les slices de type B.

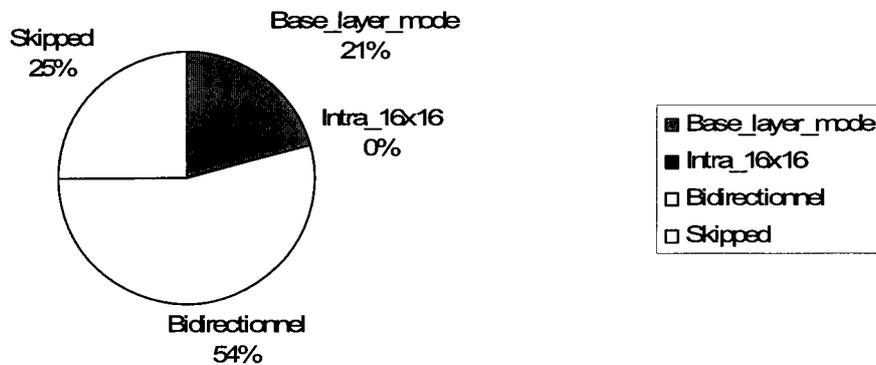


Figure 3.6 Types de macroblochs dans les slices de type B (couche d'amélioration)

Dans la section suivante nous allons étudier le surcoût de débit et l'impact sur le PSNR d'un encodage utilisant uniquement les macroblochs INTRA_BL lors de la prédiction intra-image. Ainsi, afin d'éviter le problème de glissement d'erreur de la prédiction intra-image, nous pourrions éventuellement envisager une première version de l'architecture de transcodage n'utilisant pas les macroblochs intra-couche intra-image.

3.2.4 Estimation du surcoût de débit engendré par un forçage en Intra_BL des macroblochs intra et impact sur le PSNR

Nous avons remarqué en section 3.2.1 qu'un traitement sur les macroblochs de la couche d'amélioration haute utilisant la prédiction intra-couche intra-image avait pour conséquence un effet de glissement d'erreur sur l'ensemble de l'image à cause d'une mauvaise prédiction intra-image. Nous avons également constaté que les macroblochs de type intra-couche intra-image étaient présents dans les slices de type I et P. Bien que moins présents que les macroblochs INTRA_BL, ces macroblochs ne sont pas négligeables en nombre. Nous proposons dans cette sous-section de réaliser un encodage des séquences vidéo dépourvu de macroblochs utilisant la prédiction intra-couche intra-image dans la couche d'amélioration haute. Nous comparerons ensuite le PSNR des séquences encodées en fonction du débit par rapport à celui des séquences encodées sans modification (comprenant les macroblochs INTRA_16x16 et INTRA_BL).

La Figure 3.7 illustre les résultats pour la séquence City. La courbe est composée de quatre points. Ils représentent le PSNR moyen (en luminance ou chrominance) de la séquence en fonction du débit (couche de base comprise) pour un encodage utilisant un paramètre de quantification égal à 18, 24, 30 ou 36. Nous constatons que pour l'ensemble des points, les

mesures sont très proches. Des résultats similaires ont été obtenus avec les trois autres séquences vidéo. Les mesures correspondantes sont regroupées à partir du Tableau 3.2 jusqu'au Tableau 3.5. Sur l'ensemble des expérimentations, nous constatons un surcoût de débit moyen de 1,56 % (min : 0,55 % et max : 2,71 %) et une diminution moyenne du PSNR de 0,04 dB (min : 0,02 dB et max : 0,12 dB) dus au forçage en mode INTRA_BL.

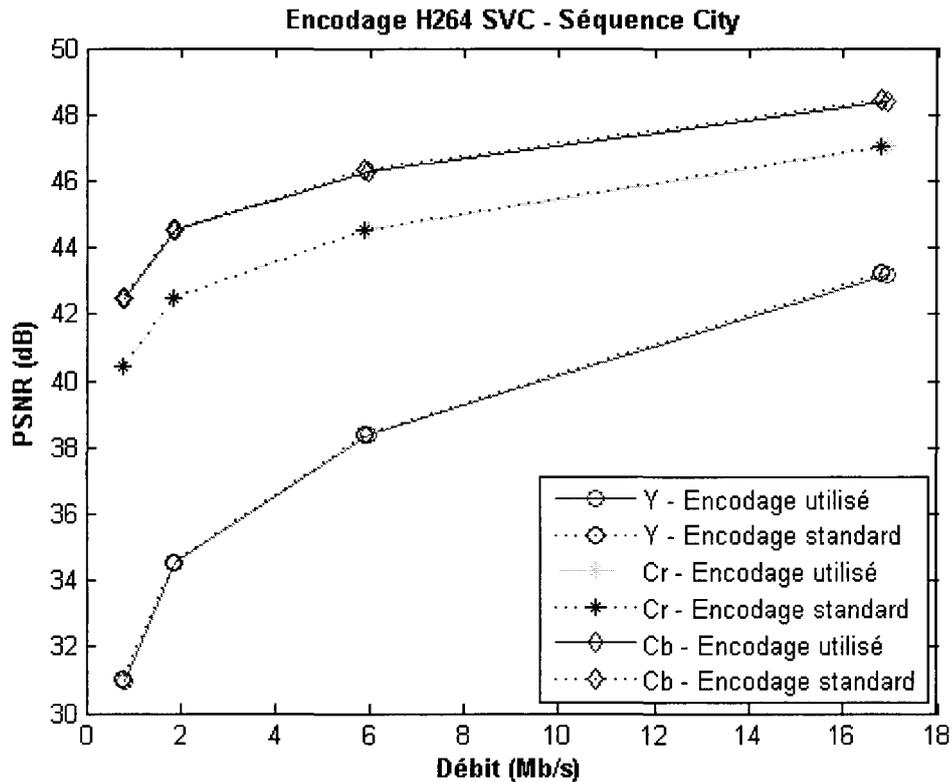


Figure 3.7 Évolution du PSNR en fonction du débit pour la séquence *City* encodée avec ou sans macroblochs utilisant la prédiction intra-couche intra-image

Dans cette section, nous avons remarqué que la limitation à l'encodage du codage des macroblochs utilisant la prédiction intra-image à celui des macroblochs utilisant la prédiction inter-couche intra-image (INTRA_BL) avait un impact négligeable sur la qualité d'image et le débit. Ainsi, dans la suite du chapitre nous emploierons des flux encodés avec cette limitation.

	Séquence	City				Crew				Harbour				Soccer				Moyenne
Statistiques (%)	QP	18	24	30	36	18	24	30	36	18	24	30	36	18	24	30	36	
Ensemble du flux	IBL	17,24	19,95	18,93	18,14	31,24	36,10	35,90	33,41	23,70	26,57	27,86	22,95	24,91	26,02	24,45	23,19	25,66
	Intra_16x16	8,17	6,16	4,46	2,79	7,79	5,50	4,94	4,55	10,40	8,32	6,74	4,68	9,09	6,94	5,08	4,14	6,23
	Prédit	10,53	9,88	7,18	5,24	2,86	4,82	4,79	3,82	7,71	8,98	9,23	8,35	5,62	5,57	4,85	3,51	6,43
	Bidirectionnel	63,86	51,11	26,49	15,17	57,99	45,78	32,36	21,12	58,14	54,49	42,47	23,43	59,46	44,14	28,22	18,85	40,19
	Skipped	0,19	12,90	42,94	58,66	0,12	7,81	22,02	37,11	0,04	1,64	13,70	40,60	0,93	17,34	37,40	50,30	21,48
Slice I	IBL	57,83	61,70	66,83	74,38	81,71	81,78	80,78	81,09	59,08	61,47	65,04	69,90	70,94	74,60	76,84	78,66	71,42
	Intra_16x16	42,17	38,30	33,17	25,62	18,29	18,22	19,22	18,91	40,92	38,53	34,96	30,10	29,06	25,40	23,16	21,34	28,58
Slice P	IBL	11,93	12,70	12,33	13,60	47,30	45,95	42,05	36,14	15,15	17,31	18,13	19,49	28,39	28,15	25,76	24,25	24,91
	Intra_16x16	26,88	17,11	9,79	3,78	36,17	22,98	19,23	17,13	40,30	29,46	22,05	12,50	38,48	27,81	18,19	13,58	22,22
	Prédit	60,76	57,00	41,41	30,22	16,50	27,78	27,63	22,00	44,50	51,81	53,26	48,13	32,40	32,12	27,98	20,27	37,11
	Skipped	0,43	13,19	36,47	52,39	0,04	3,28	11,10	24,73	0,05	1,41	6,56	19,88	0,73	11,92	28,06	41,90	15,76
Slice B	IBL	13,92	16,96	15,09	12,89	21,84	28,68	29,43	27,41	21,73	24,82	25,96	18,49	18,93	20,08	18,26	16,71	20,70
	Intra_16x16	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Bidirectionnel	85,92	68,76	35,64	20,40	78,01	61,58	43,54	28,42	78,22	73,30	57,14	31,52	79,98	59,38	37,96	25,37	54,07
	Skipped	0,16	14,28	49,26	66,70	0,15	9,74	27,04	44,17	0,05	1,88	16,90	49,98	1,08	20,54	43,77	57,92	25,23

Tableau 3.1 Estimation de l'utilisation des modes de prédiction

Séquence CITY - 704x576 - 300 images - 30ips (scenario2:CIF-4CIF)										
QP	Encodage modifié de la couche d'amélioration				Encodage standard				Augmentation du débit (%)	Différence de PSNR (dB)
	Débit (Mbps)	Y-PSNR (dB)	Cb-PSNR (dB)	Cr-PSNR (dB)	Débit (Mbps)	Y-PSNR (dB)	Cb-PSNR (dB)	Cr-PSNR (dB)		
18	16,95	43,2	47,03	48,41	16,84	43,22	47,06	48,44	0,66	0,03
24	5,99	38,39	44,52	46,3	5,9	38,41	44,53	46,33	1,46	0,02
30	1,9	34,53	42,53	44,56	1,86	34,56	42,49	44,54	2,4	0,03
36	0,8	30,97	40,51	42,5	0,78	31,02	40,45	42,48	2,71	0,06

Tableau 3.2 Encodage de la séquence City sans macrobloc utilisant la prédiction intra-couche intra-image (dans la C.A.)

Chapitre 3 : Solution proposée de transcodage à complexité réduite des flux H.264 SVC en échelonnabilité spatiale

Séquence <i>CREW</i> - 704x576 - 300 images - 30ips (scenario2: CIF-4CIF)										
QP	Encodage modifié de la couche d'amélioration				Encodage standard				Augmentation du débit (%)	Différence de PSNR (dB)
	Débit (Mbps)	Y-PSNR (dB)	Cb-PSNR (dB)	Cr-PSNR (dB)	Débit (Mbps)	Y-PSNR (dB)	Cb-PSNR (dB)	Cr-PSNR (dB)		
18	15,63	43,9	45,41	46,93	15,54	43,95	45,45	46,97	0,55	0,05
24	5,74	39,85	42,84	44,25	5,68	39,87	42,85	44,27	1,02	0,02
30	2,29	36,86	41,23	41,76	2,25	36,91	41,25	41,78	1,46	0,05
36	1,03	33,73	39,27	39	1,01	33,85	39,32	39,05	1,94	0,12

Tableau 3.3 Encodage de la séquence Crew sans macrobloc utilisant la prédiction intra-couche intra-image (dans la C.A.)

Séquence <i>HARBOUR</i> - 704x576 - 300 images - 30ips (scenario2: CIF-4CIF)										
QP	Encodage modifié de la couche d'amélioration				Encodage standard				Augmentation du débit (%)	Différence de PSNR (dB)
	Débit (Mbps)	Y-PSNR (dB)	Cb-PSNR (dB)	Cr-PSNR (dB)	Débit (Mbps)	Y-PSNR (dB)	Cb-PSNR (dB)	Cr-PSNR (dB)		
18	21,62	43,13	46,66	48,2	21,43	43,17	46,69	48,24	0,87	0,03
24	10,24	38,54	43,99	45,93	10,11	38,56	44	45,93	1,2	0,02
30	4,15	34,44	42,03	43,98	4,08	34,47	42	43,95	1,63	0,03
36	1,58	30,54	40,05	41,57	1,55	30,58	39,98	41,52	2,17	0,04

Tableau 3.4 Encodage de la séquence Harbour sans macrobloc utilisant la prédiction intra-couche intra-image (dans la C.A.)

Séquence <i>SOCCKER</i> - 704x576 - 300 images - 30ips (scenario2: CIF-4CIF)										
QP	Encodage modifié de la couche d'amélioration				Encodage standard				Augmentation du débit (%)	Différence de PSNR (dB)
	Débit (Mbps)	Y-PSNR (dB)	Cb-PSNR (dB)	Cr-PSNR (dB)	Débit (Mbps)	Y-PSNR (dB)	Cb-PSNR (dB)	Cr-PSNR (dB)		
18	13,94	43,9	47,36	48,85	13,77	43,94	47,39	48,89	1,25	0,04
24	5,5	39,53	45,04	46,89	5,41	39,56	45,07	46,92	1,63	0,03
30	2,27	35,84	43,1	44,91	2,23	35,87	43,09	44,92	1,93	0,03
36	1,01	32,35	40,87	42,75	0,99	32,44	40,84	42,75	2,14	0,09

Tableau 3.5 Encodage de la séquence Soccer sans macrobloc utilisant la prédiction intra-couche intra-image (dans la C.A.)

3.3 Transcodage H.264 SVC en échelonnabilité spatiale par sélection en position fréquentielle des coefficients résiduels de la couche d'amélioration haute

3.3.1 Architecture proposée

Dans cette section, nous proposons une première architecture de transcodage H.264 SVC en échelonnabilité spatiale. Ce traitement de faible complexité est réalisé par sélection en position fréquentielle des coefficients résiduels de la couche d'amélioration haute.

Tout d'abord, le transcodage est uniquement réalisé sur la couche d'amélioration la plus haute afin d'éviter d'éventuelles propagations d'erreurs entre une couche de référence et une couche supérieure. Le traitement opère sur les coefficients résiduels fréquentiels des composantes de luminance et de chrominance. Nous avons choisi d'utiliser des flux H.264 SVC encodés sans macroblocs utilisant la prédiction intra-couche intra-image (INTRA_16x16) pour la couche d'amélioration la plus haute. Ainsi, le traitement est appliqué à tous les macroblocs de la couche d'amélioration haute du flux.

Comme illustré en Figure 2.6, l'architecture de transcodage H.264 SVC en échelonnabilité spatiale est peu complexe. L'étape la plus délicate consiste à localiser les coefficients résiduels fréquentiels en effectuant une lecture rigoureuse des différents éléments syntaxiques consécutifs dans le flux H.264 SVC. Ensuite, il est nécessaire d'effectuer un décodage entropique des coefficients afin de reconstituer les blocs résiduels fréquentiels. Ensuite, le module de sélection par position fréquentielle est appliqué au bloc afin de réduire le nombre de coefficients non-nuls. Le traitement s'achève par le codage entropique du nouveau bloc.

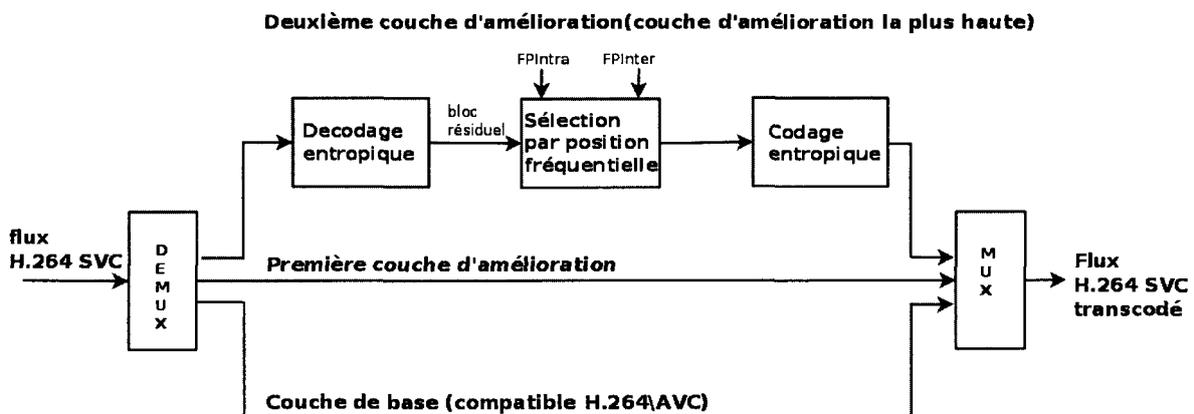


Figure 3.8 Diagramme simplifié du transcodeur H.264 SVC en sélection par position fréquentielle pour les macroblocs intra et inter

3.4 Résultats expérimentaux du transcodage de flux en échelonnabilité spatiale

Afin de réaliser ces expérimentations, nous avons utilisé les paramètres d'encodage présentés dans la section 3.2.2 avec la limitation du codage des macroblocs utilisant la prédiction intra-image au mode INTRA_BL (cf 3.2.4). Dans notre scénario, les flux sont composés uniquement d'une couche de base (CIF) et d'une couche d'amélioration (4CIF). Dans les paragraphes suivants, le paramètre de transcodage représentant la position fréquentielle (*FP*) peut théoriquement varier de 1 (uniquement le coefficient DC) à 16 (aucun changement). Le paramètre *FP* est volontairement configuré à l'identique pour les macroblocs codés en prédiction intra-image et inter-image ($FP_{Intra} = FP_{Inter}$). Notons qu'il s'agit d'une première approche et qu'il est possible de faire varier ces paramètres indépendamment les uns des autres et ainsi obtenir jusqu'à 256 combinaisons de débits intermédiaires.

Afin de situer les points ou couples (débit, PSNR) issus du transcodage, nous utiliserons comme référence le PSNR de la couche de base (CIF) sur-échantillonnée au format 4CIF avec son débit respectif, ainsi que le PSNR de la couche d'amélioration reconstruite avec son débit (débit de la couche de base compris). Chaque point (débit, PSNR) issu du transcodage prend évidemment en compte le débit de la couche de base, nécessaire au bon fonctionnement de l'échelonnabilité. Nous utilisons l'utilitaire *DownConvert* fourni avec le logiciel JSVM afin d'effectuer le sur-échantillonnage de la couche de base. L'algorithme de sur-échantillonnage choisi est celui par défaut, correspondant à celui de la norme H.264 SVC. Il est utilisé dans l'extension échelonnable de H.264/AVC lors de l'encodage de la luminance des macroblocs de type INTRA_BL dans la couche d'amélioration. Le filtre effectue le sur-échantillonnage du bloc correspondant dans la couche de base et constitue le bloc de prédiction. Il s'agit pour la composante de luminance d'un filtre RIF à une dimension composé de 4 coefficients et appliqué horizontalement et verticalement, et d'un filtre bilinéaire pour les composantes de chrominance [SCH07].

Pour chaque séquence vidéo et pour un paramètre de quantification donné, les résultats comprennent le débit équivalent ainsi que le PSNR de la luminance et de la chrominance pour :

- La couche de base (format CIF) sur-échantillonnée au format 4CIF. Ces points sont situés dans la partie gauche des courbes et ne sont pas reliés aux autres mesures.
- La couche d'amélioration (format 4CIF) transcodée est représentée par 11 mesures, allant de $FP=1$ à $FP=10$, puis $FP=16$ (flux initial, sans transcodage). Les mesures de $FP=11$ à $FP=15$ ne sont pas représentées par souci de visibilité des résultats.

3.4.1 Réduction de débit

Les mesures pour les séquences *City*, *Crew*, *Harbour* et *Soccer* sont illustrées en Figure 3.10, Figure 3.11, Figure 3.12 et Figure 3.13 respectivement. Les quatre points non reliés en bas à gauche des courbes sont les mesures de la couche de base sur-échantillonnée pour chaque paramètre de quantification. Les points situés à droite des courbes correspondent au débit et au PSNR des flux encodés (Débit = 100%). L'ensemble des manipulations montre que la réduction de débit maximale diminue lorsque le paramètre de quantification augmente. Ceci n'est pas surprenant puisque l'augmentation du paramètre de quantification conduit à diminuer l'amplitude des coefficients résiduels jusqu'à les rendre nuls. Il y a donc moins de coefficients résiduels potentiellement supprimables par le transcodage par sélection en position fréquentielle.

Nous constatons que la réduction maximale de débit (66%) est atteinte pour la séquence *City* (Figure 3.10) encodée avec un paramètre de quantification égal à 18 et transcodée avec le paramètre de transcodage FP égal à 1. Dans ce cas, les données de la couche d'amélioration sont réduites de 76%.

D'autre part, nous avons une réduction minimale de débit pour la séquence *Crew* (Figure 3.11) encodée avec un paramètre de quantification égal à 38. Même avec un paramètre de transcodage FP égal à 1, la réduction de débit est de l'ordre de 11%. Ces résultats peuvent s'expliquer par la nature de la séquence vidéo et la valeur relativement élevée du paramètre de quantification. La séquence est composée de larges plages uniformes (mur, porte...) et de peu de zones très détaillées (visages, écussons des combinaisons) riches en hautes fréquences. L'utilisation d'un paramètre de quantification élevé a pour incidence de réduire l'amplitude des coefficients résiduels fréquentiels jusqu'à les rendre nuls. Alors, le flux encodé est relativement pauvre en hautes fréquences et le transcodage ne peut supprimer qu'un nombre limité de coefficients résiduels.

Il est intéressant de noter que chaque paramètre de transcodage FP procure un débit intermédiaire. D'autres débits intermédiaires peuvent être obtenus en choisissant de ne pas mettre les paramètres de transcodage FP_{Intra} et FP_{Inter} égaux. Des résultats sont proposés à ce sujet dans la section 3.6.

3.4.2 Analyse PSNR

Dans cette section, nous évaluons tout d'abord le PSNR des échantillons de luminance pour l'ensemble des débits intermédiaires issus du transcodage. Puis, nous considérons le PSNR des échantillons de chrominance.

Tout d'abord, nous remarquons que la différence la plus importante de PSNR en luminance entre les deux paramètres de transcodage extrêmes FP=16 (pas de transcodage) et FP=1 (uniquement coefficient DC) est atteinte pour la séquence *City* encodée avec un paramètre de quantification QP=18. Elle s'élève à 15 dB. À l'inverse, la différence de PSNR la moins significative entre ces deux extrema est de l'ordre de 1,7 dB. Elle est atteinte pour la séquence *Crew* encodée avec un paramètre de quantification QP=36.

Le PSNR de la luminance de la couche de base (CIF) sur-échantillonnée au format 4CIF est logiquement toujours inférieur à celui de la couche d'amélioration pour l'ensemble des paramètres de transcodage. Nous constatons que l'écart de PSNR entre la couche de base sur-échantillonnée et la couche d'amélioration transcodée avec le paramètre FP=1 a tendance à s'agrandir pour un taux de compression élevé (hausse du paramètre de quantification). La différence la plus importante s'élève à 1,59 dB. Elle est remarquée pour la séquence vidéo *Harbour* encodée avec un paramètre de quantification QP=36. À l'inverse, l'écart le plus petit est de 0,55 dB. Il est constaté pour la séquence vidéo *City* encodée avec un paramètre de quantification QP=18. Cette différence de PSNR s'explique par une réduction de débit plus ou moins importante après transcodage.

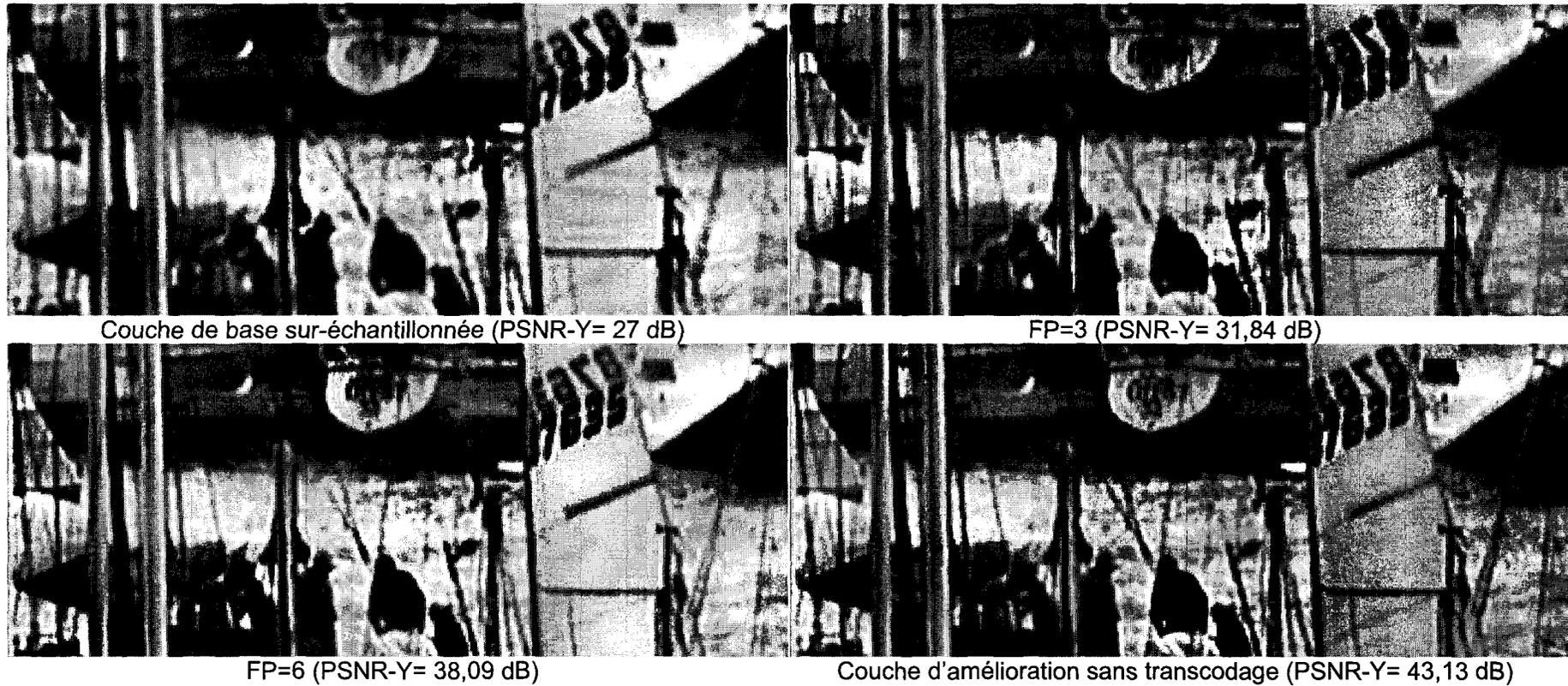


Figure 3.9 Illustration de l'image numéro 0 (image intra) de la séquence *Harbour* (QP=18)

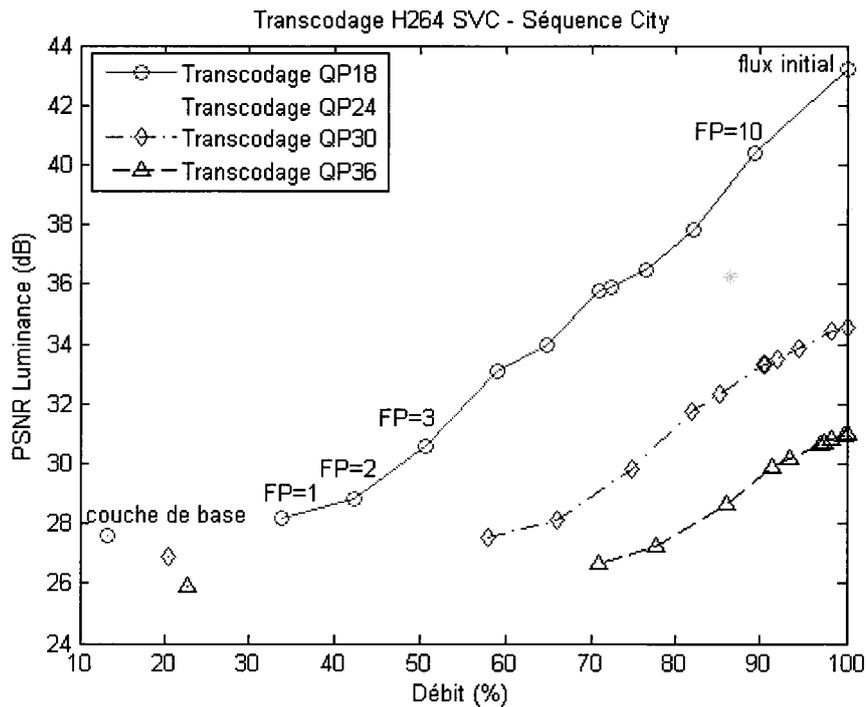


Figure 3.10 Évolution du PSNR (Y) en fonction du débit - Séquence City

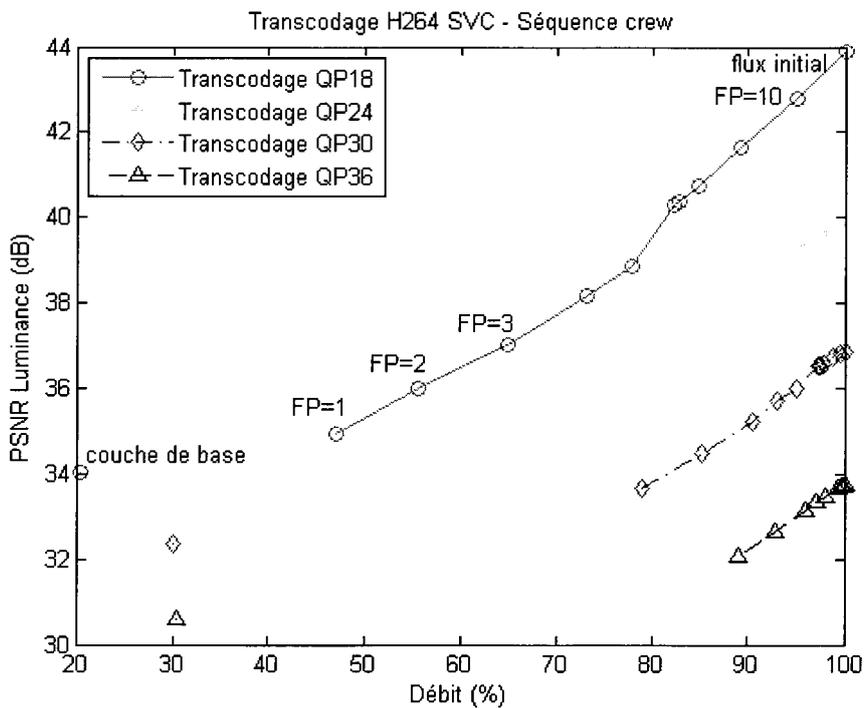


Figure 3.11 Évolution du PSNR (Y) en fonction du débit – Séquence Crew

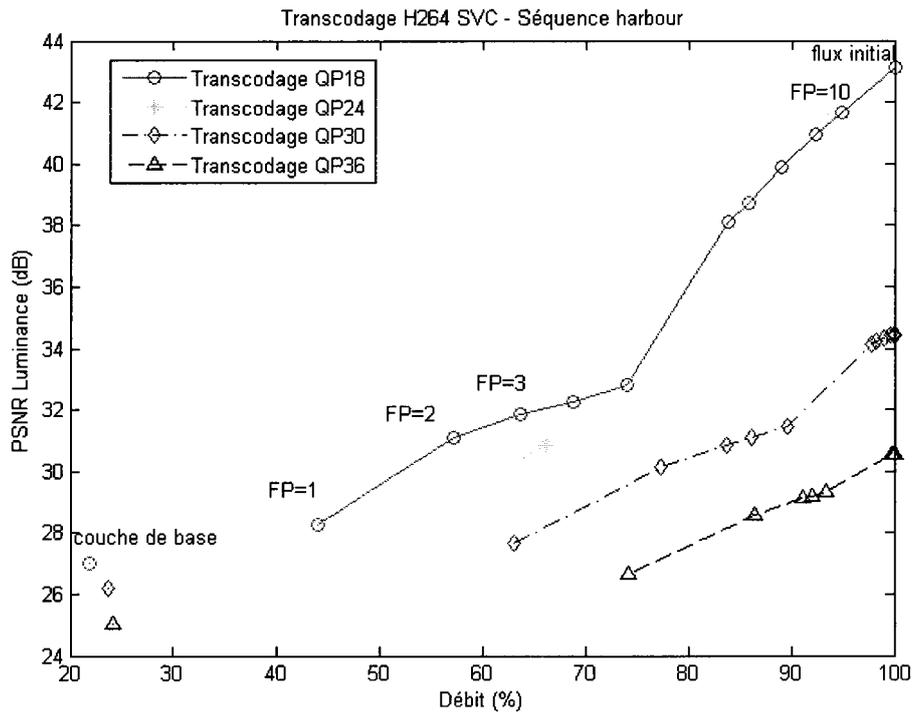


Figure 3.12 Évolution du PSNR (Y) en fonction du débit – Séquence *Harbour*

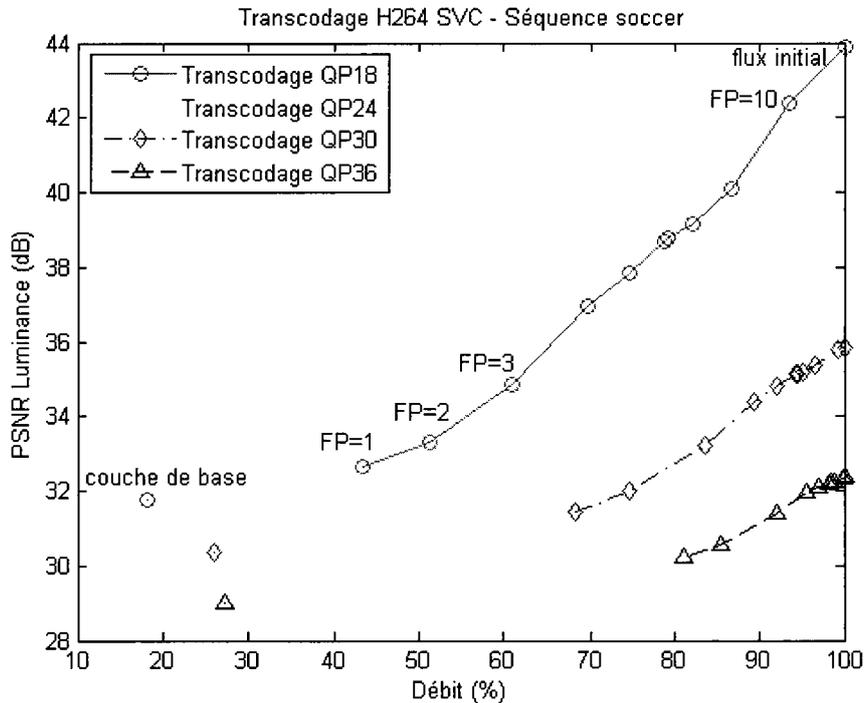


Figure 3.13 Évolution du PSNR (Y) en fonction du débit – Séquence *Soccer*

En considérant la composante de chrominance Cb (U) et l'ensemble des séquences vidéo, nous constatons une différence maximale de PSNR de 3,7 dB entre les deux

paramètres de transcodage extrêmes FP=16 (pas de transcodage) et FP=1 (uniquement coefficient DC) pour la séquence *City* encodée avec un paramètre de quantification QP=18 (Figure 2.9). Dans le cas de la composante Cr (V), cette différence maximale de PSNR est de 3,9 dB pour la séquence *Crew* encodée avec un paramètre de quantification également fixé à 18 (Figure 3.15).

À l'inverse, la différence de PSNR la moins significative entre ces deux extrema est de l'ordre de 0,1 dB et 0,05 dB pour les composantes Cb et Cr respectivement. Elle est atteinte pour les séquences *Harbour* et *City* encodées avec un paramètre de quantification QP=36.

Le PSNR des composantes de chrominance de la couche de base (CIF) sur-échantillonnée au format 4CIF est toujours inférieur à celui de la couche d'amélioration pour l'ensemble des paramètres de transcodage. Nous constatons que l'écart de PSNR entre la couche de base sur-échantillonnée et la couche d'amélioration transcodée avec le paramètre FP=1 a tendance à s'agrandir pour un taux de compression élevé (hausse du paramètre de quantification). La différence la plus importante s'élève à 2,04 dB pour Cb et 2,61 dB pour Cr. Cette différence est constatée pour la séquence vidéo *Crew* encodée avec un paramètre de quantification QP=36. À l'inverse, l'écart le plus petit est de 0,57 dB pour Cb et 0,86 dB pour Cr. Il est constaté pour les séquences vidéo *City* et *Harbour* encodées avec un paramètre de quantification QP=18.

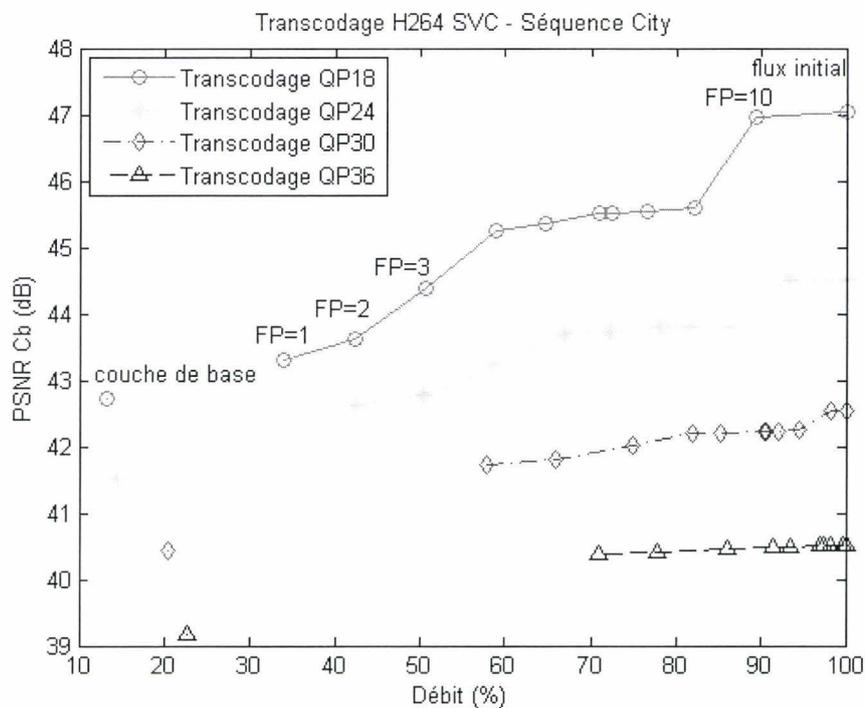


Figure 3.14 Évolution du PSNR de la composante de chrominance Cb (U) en fonction du débit — Séquence *City*



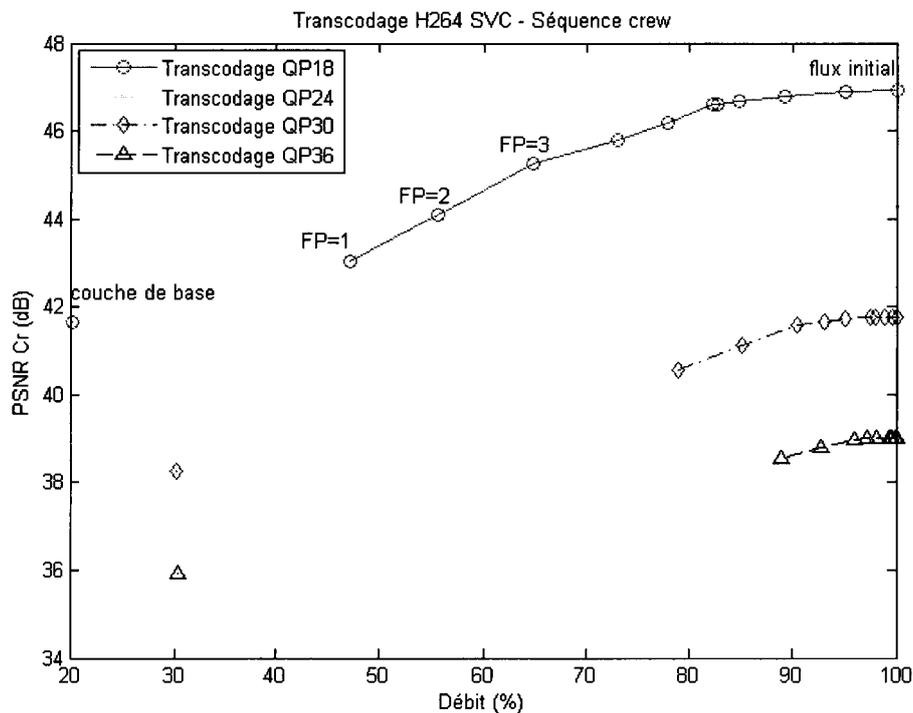


Figure 3.15 Évolution du PSNR de la composante de chrominance Cr (V) en fonction du débit - Transcodage sur flux SVC –Séquence Crew

L'ensemble des mesures est regroupé sous forme de tableaux du Tableau 3.6 au Tableau 3.9, pour les séquences *City*, *Crew*, *Harbour* et *Soccer* respectivement.

Dans cette section, nous avons constaté que :

- Nous avons une faible réduction de débit pour des taux de compression élevés (QP=36). Ce qui correspond à utiliser un paramètre de quantification élevé. Le débit de la couche d'amélioration transcodée (FP=1) est proche de celui de la couche d'amélioration sans transcodage (FP=16). Alors, la différence de débit entre la couche de base et le flux transcodé (couche de base + couche d'amélioration) est plus importante. Par contre, la différence entre le PSNR de la couche de base et celui de la couche d'amélioration (FP=1) est plus significative.
- Nous avons une importante réduction de débit pour des débits relativement élevés (QP=18). Ce qui correspond à utiliser un paramètre de quantification plus petit. Le débit de la couche d'amélioration transcodée (FP=1) est largement inférieur à celui de la couche d'amélioration sans transcodage (FP=16). Alors, la différence de débit entre la couche de base et le flux

Chapitre 3 : Solution proposée de transcodage à complexité réduite des flux H.264 SVC en échelonnabilité spatiale

transcodé (couche de base + couche d'amélioration) est moins importante. La différence entre le PSNR de la couche de base et celui de la couche d'amélioration (FP=1) est moins flagrante.

Séquence CITY - 704x576 - 300 images - 30ips (H.264/AVC SVS – spatial -CIF-4CIF)							
QP	Couche(s)	FP	Débit (Mb/s)	PSNR Y(dB)	PSNR Cb(dB)	PSNR Cr(dB)	
18	CB + CA	-	16,95	43,2	47,03	48,41	
		10	15,14	40,38	46,97	48,4	
		9	13,93	37,84	45,6	47,94	
		8	12,95	36,48	45,53	47,91	
		7	12,27	35,91	45,52	47,91	
		6	12,03	35,77	45,52	47,9	
		5	10,97	33,95	45,37	47,87	
		4	9,99	33,09	45,26	47,8	
		3	8,57	30,6	44,39	47,35	
		2	7,19	28,81	43,63	46,77	
		1	5,73	28,17	43,3	46,55	
		CB up	-	2,24	27,62	42,73	45,67
		24	CB + CA	-	5,99	38,39	44,52
	10			5,58	37,68	44,51	46,3
9	5,18			36,27	43,81	46,22	
8	4,71			35,42	43,8	46,22	
7	4,89			35,06	43,8	46,22	
6	4,66			34,98	43,8	46,22	
5	4,32			33,48	43,72	46,21	
4	4,01			32,71	43,69	46,19	
3	3,52			30,38	43,22	46,06	
2	3,02			28,61	42,78	45,79	
1	2,55			27,96	42,61	45,69	
CB up	-			0,86	27,38	41,51	44,21
30	CB + CA			-	1,9	34,53	42,53
			10	1,87	34,42	42,53	44,56
		9	1,8	33,86	42,24	44,56	
		8	1,75	33,48	42,24	44,56	
		7	1,72	33,33	42,24	44,56	
		6	1,72	33,3	42,24	44,56	
		5	1,62	32,33	42,2	44,56	
		4	1,56	31,73	42,19	44,56	
		3	1,42	29,83	42,02	44,53	
		2	1,26	28,15	41,8	44,44	
		1	1,1	27,53	41,73	44,39	
		CB up	-	0,39	26,87	40,42	42,84
		36	CB + CA	-	0,8	30,97	40,51
	10			0,8	30,96	40,51	42,5
9	0,79			30,79	40,5	42,5	
8	0,78			30,68	40,5	42,5	
7	0,77			30,65	40,5	42,5	
6	0,77			30,64	40,5	42,5	
5	0,75			30,2	40,5	42,5	
4	0,73			29,89	40,49	42,5	
3	0,69			28,63	40,46	42,49	
2	0,62			27,22	40,4	42,47	
1	0,57			26,65	40,38	42,45	
CB up	-			0,18	25,87	39,16	41,17

Tableau 3.6 Évolution du PSNR en fonction du débit pour la séquence City

**Chapitre 3 : Solution proposée de transcodage à complexité réduite des flux H.264
SVC en échelonnabilité spatiale**

Séquence <i>CREW</i> - 704x576 - 300 images - 30ips (H.264/AVC SVS – spatial -CIF-4CIF)						
QP	Couche(s)	FP	Débit (Mbps)	PSNR Y(dB)	PSNR Cb(dB)	PSNR Cr(dB)
18	CB + CA	-	15,63	43,9	45,41	46,93
		10	14,86	42,8	45,35	46,9
		9	13,94	41,63	45,08	46,79
		8	13,26	40,75	44,93	46,66
		7	12,93	40,36	44,91	46,61
		6	12,86	40,28	44,91	46,6
		5	12,17	38,88	44,79	46,18
		4	11,41	38,16	44,52	45,78
		3	10,13	37,03	43,93	45,25
		2	8,69	35,99	43,17	44,09
		1	7,37	34,94	42,67	43,02
		CB up	-	3,16	34,02	41,89
	24	CB + CA	-	5,74	39,85	42,84
10			5,63	39,63	42,83	44,25
9			5,46	39,26	42,82	44,23
8			5,32	38,89	42,81	44,21
7			5,26	38,71	42,81	44,2
6			5,24	38,68	42,81	44,2
5			5,04	37,74	42,79	44,04
4			4,83	37,24	42,73	43,87
3			4,52	36,43	42,59	43,62
2			4,08	35,5	42,21	42,83
1			3,63	34,52	41,9	42
CB up			-	1,47	33,45	40,55
30		CB + CA	-	2,29	36,86	41,23
	10		2,28	36,83	41,23	41,76
	9		2,26	36,72	41,23	41,76
	8		2,24	36,59	41,23	41,76
	7		2,23	36,53	41,23	41,76
	6		2,23	36,52	41,23	41,76
	5		2,17	36,01	41,22	41,72
	4		2,13	35,72	41,21	41,67
	3		2,07	35,23	41,17	41,58
	2		1,95	34,47	40,97	41,11
	1		1,81	33,66	40,8	40,56
	CB up		-	0,69	32,35	38,97
	36	CB + CA	-	1,03	33,73	39,27
10			1,03	33,72	39,27	39
9			1,02	33,7	39,27	39
8			1,02	33,66	39,27	39
7			1,02	33,65	39,27	39
6			1,02	33,65	39,27	39
5			1,01	33,45	39,27	38,99
4			1	33,33	39,27	38,99
3			0,98	33,11	39,26	38,98
2			0,95	32,63	39,18	38,8
1			0,91	32,08	39,14	38,54
CB up			-	0,31	30,57	37,1

Tableau 3.7 Évolution du PSNR en fonction du débit pour la séquence Crew

**Chapitre 3 : Solution proposée de transcodage à complexité réduite des flux H.264
SVC en échelonnabilité spatiale**

Séquence HARBOUR - 704x576 - 300 images - 30ips (H.264/AVC SVS – spatial -CIF-4CIF)						
QP	Couche(s)	FP	Débit (Mbps)	PSNR Y(dB)	PSNR Cb(dB)	PSNR Cr(dB)
18	CB + CA	-	21,62	43,13	46,66	48,2
		10	20,51	41,64	46,63	48,2
		9	19,96	40,96	46,53	48,18
		8	19,23	39,86	46,46	48,16
		7	18,56	38,74	46,41	48,16
		6	18,11	38,09	46,37	48,15
		5	16	32,79	45,89	47,79
		4	14,86	32,28	45,71	47,74
		3	13,79	31,84	45,14	47,49
		2	12,35	31,08	44,39	47,07
		1	9,49	28,27	43,64	46,15
		CB up	-	4,73	27	42,94
24	CB + CA	-	10,23	38,54	43,99	45,93
		10	10,08	38,33	43,99	45,93
		9	9,96	38,14	43,97	45,93
		8	9,74	37,7	43,96	45,93
		7	9,52	37,15	43,96	45,93
		6	9,4	36,8	43,95	45,93
		5	8,41	32,46	43,81	45,85
		4	7,88	31,98	43,77	45,84
		3	7,47	31,62	43,53	45,77
		2	6,77	30,85	43,11	45,56
		1	5,3	28,12	42,63	44,92
		CB up	-	2,36	26,77	41,29
30	CB + CA	-	4,15	34,44	42,03	43,98
		10	4,14	34,43	42,03	43,98
		9	4,13	34,41	42,03	43,98
		8	4,11	34,34	42,03	43,98
		7	4,07	34,21	42,03	43,98
		6	4,05	34,13	42,03	43,98
		5	3,71	31,43	42,01	43,98
		4	3,57	31,09	42	43,97
		3	3,47	30,87	41,93	43,96
		2	3,2	30,14	41,73	43,89
		1	2,62	27,66	41,49	43,54
		CB up	-	0,98	26,2	39,75
36	CB + CA	-	1,58	30,54	40,05	41,57
		10	1,58	30,54	40,05	41,57
		9	1,58	30,54	40,05	41,57
		8	1,58	30,53	40,05	41,57
		7	1,58	30,52	40,05	41,57
		6	1,57	30,52	40,05	41,57
		5	1,47	29,33	40,04	41,57
		4	1,45	29,2	40,04	41,57
		3	1,44	29,13	40,04	41,57
		2	1,36	28,59	40	41,55
		1	1,17	26,63	39,95	41,49
		CB up	-	0,38	25,04	38,16

Tableau 3.8 Évolution du PSNR en fonction du débit pour la séquence Harbour

**Chapitre 3 : Solution proposée de transcodage à complexité réduite des flux H.264
SVC en échelonnabilité spatiale**

Séquence SOCCER - 704x576 - 300 images - 30ips (H.264/AVC SVS – spatial -CIF-4CIF)						
QP	Couche(s)	FP	Débit (Mbps)	PSNR Y(dB)	PSNR Cb(dB)	PSNR Cr(dB)
18	CB + CA	-	13,94	43,9	47,36	48,85
		10	13,04	42,36	47,3	48,78
		9	12,08	40,09	47,01	48,67
		8	11,44	39,16	46,94	48,59
		7	11,05	38,76	46,9	48,55
		6	10,99	38,71	46,87	48,5
		5	10,42	37,83	46,74	48,38
		4	9,72	36,95	46,55	48,21
		3	8,51	34,85	45,71	47,8
		2	7,15	33,32	44,65	47,11
		1	6,06	32,67	44,28	46,74
		CB up	-	2,53	31,77	43,24
	24	CB + CA	-	5,5	39,53	45,04
10			5,32	39,17	45,03	46,88
9			5,03	38,04	44,98	46,86
8			4,86	37,55	44,96	46,85
7			4,76	37,34	44,95	46,84
6			4,75	37,32	44,95	46,83
5			4,57	36,71	44,91	46,8
4			4,35	36,01	44,84	46,73
3			3,93	34,26	44,45	46,58
2			3,4	32,84	43,78	46,16
1			3	32,23	43,51	45,91
CB up			-	1,21	31,29	42,05
30		CB + CA	-	2,27	35,84	43,1
	10		2,25	35,79	43,1	44,91
	9		2,19	35,38	43,09	44,91
	8		2,16	35,2	43,09	44,91
	7		2,14	35,12	43,08	44,9
	6		2,14	35,12	43,08	44,9
	5		2,09	34,81	43,08	44,9
	4		2,03	34,39	43,05	44,87
	3		1,9	33,22	42,91	44,84
	2		1,69	31,98	42,51	44,63
	1		1,55	31,46	42,36	44,49
	CB up		-	0,59	30,35	40,58
	36	CB + CA	-	1,01	32,35	40,87
10			1,01	32,35	40,86	42,75
9			1	32,25	40,86	42,75
8			0,99	32,21	40,86	42,75
7			0,99	32,2	40,86	42,75
6			0,99	32,2	40,86	42,75
5			0,98	32,09	40,86	42,75
4			0,96	31,94	40,86	42,75
3			0,93	31,39	40,84	42,74
2			0,86	30,56	40,61	42,68
1			0,82	30,22	40,54	42,63
CB up			-	0,27	29,01	38,59

Tableau 3.9 Évolution du PSNR en fonction du débit pour la séquence Soccer

3.5 Limitation de l'architecture de transcodage à la composante de luminance

Dans la section précédente, nous avons transcodé les échantillons de luminance tout comme ceux de chrominance. Nous proposons ici d'effectuer un transcodage des échantillons de luminance uniquement afin de comparer l'impact sur la réduction de débit. Cette étude nous permettra d'estimer la réduction de débit relative au transcodage de la luminance mais également de proposer une version moins complexe de notre transcodeur. Afin de réaliser ces expérimentations, nous avons également utilisé les paramètres d'encodage présentés dans la section 3.2.2. Le mode opératoire est identique à celui de la section 3.4 hormis que notre transcodeur H.264 SVC opère une sélection en position fréquentielle des coefficients résiduels de luminance uniquement de la couche d'amélioration haute.

Pour l'ensemble des manipulations, la courbe légèrement supérieure (en pointillés) représente le PSNR en fonction du pourcentage du débit initial dans le cas où la composante de chrominance est également transcodée (section 3.4). Pour chaque point correspondant d'une courbe à une autre, nous avons un PSNR identique mais un débit relatif différent. Le débit en sortie du transcodage en luminance uniquement est donc supérieur ou égal à celui du transcodage pour les composantes de luminance et de chrominance.

Dans la section 3.4, nous avons remarqué que la réduction de débit était plus importante pour des débits relativement élevés (i.e. en utilisant un paramètre de quantification petit, ici $QP=18$). La Figure 3.16 regroupe les résultats des quatre séquences utilisées et encodées avec un paramètre de quantification égal à 18. Pour chaque séquence, deux courbes sont représentées. Dans le cas d'un paramètre de quantification égal à 18, le manque à gagner de réduction de débit dû à l'absence de transcodage des échantillons de chrominance est le plus significatif. Il s'élève au maximum à 0,85 Mb/s pour la séquence *Crew* transcodée avec le paramètre $FP=1$. À l'inverse, la différence est respectivement de 0,25 Mb/s, 0,24 Mb/s et 0,29 Mb/s, pour les séquences *City*, *Harbour* et *Soccer*.

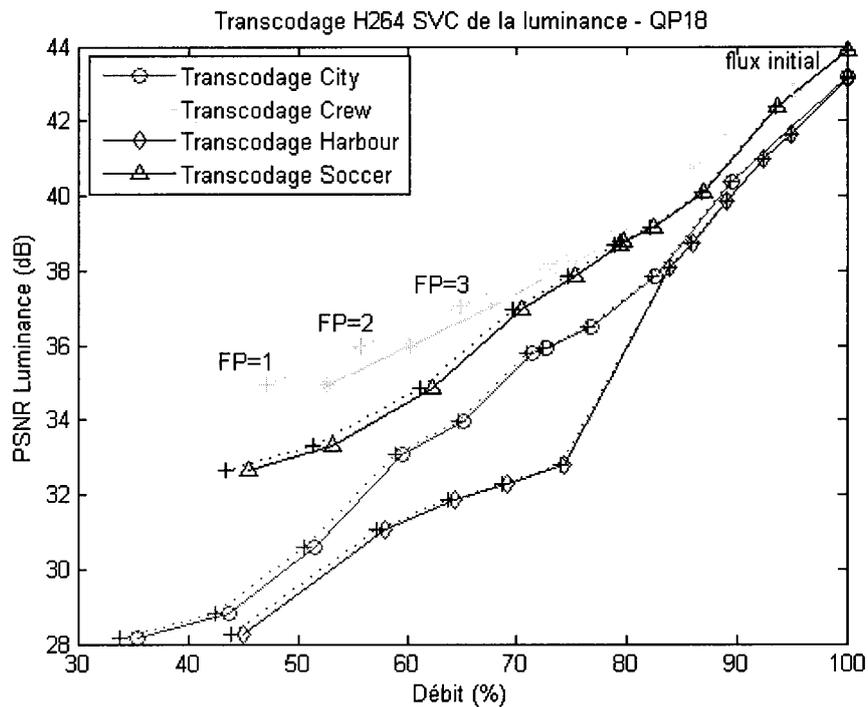


Figure 3.16 Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des échantillons de luminance – QP=18

Pour un paramètre de quantification plus élevé la différence de réduction de débit est de moins en moins significative. La Figure 3.17 illustre les résultats pour le paramètre de quantification QP=24. Les courbes sont quasiment confondues. Ceci s'explique par le peu de fréquences présentes dans les blocs résiduels de chrominance après une quantification relativement élevée.

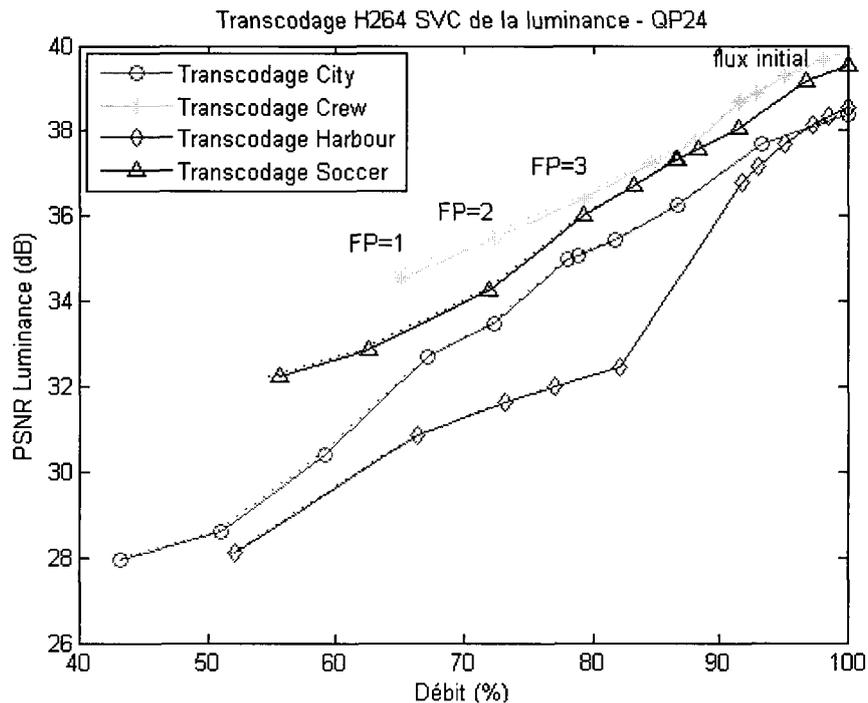


Figure 3.17 Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des échantillons de luminance – QP=24

Cette étude nous a permis d'estimer la réduction de débit relative au transcodage de la luminance sur une plage de débits relativement large. Moyennant une baisse d'efficacité, il est envisageable de concevoir une version moins complexe de notre transcodeur traitant uniquement les blocs de luminance.

3.6 Étude de l'influence du paramètre de transcodage FPinter sur la réduction de débit et la qualité vidéo

Dans cette section nous étudions la réduction de débit et la qualité vidéo lorsque la sélectivité fréquentielle du transcodage est uniquement appliquée aux macroblocs utilisant la prédiction inter-image. Par convention, nous appellerons « mode inter » ce mode particulier du transcodeur, en opposition au « mode standard » lorsque les deux types de macroblocs sont transcodés identiquement ($FP_{intra}=FP_{inter}$). L'objectif principal de cette étude est d'évaluer l'intérêt de ne transcoder que ces macroblocs afin d'envisager une réduction de la complexité de l'architecture de transcodage.

Afin de réaliser ces expérimentations, nous avons également utilisé les paramètres d'encodage présentés dans la section 3.2.2. Le mode opératoire est identique à celui de la section 3.4 hormis la limitation du traitement aux macroblocs inter-image. Sur les figures suivantes, nous représentons à la fois pour un paramètre de quantification donné:

- Les points issus du transcodage des macroblochs utilisant la prédiction intra- et inter-image (FPIntra=FPInter), reliés par des pointillés (mode standard).
- Les points issus du transcodage des macroblochs utilisant la prédiction inter-image (FPInter variable et FPIntra=16), reliés par un trait plein (mode inter).

La Figure 3.18 illustre l'évolution du PSNR de la luminance en fonction du débit pour la séquence *City*. Cette séquence détient la différence la plus importante entre les deux modes du transcodeur. Ainsi, pour un paramètre de quantification égal à 18, nous constatons que le PSNR résultant du transcodage en mode inter (paramètre FPInter=1, 0 (b)) est supérieur de plus de 7 dB à celui issu du transcodage en mode standard (FPIntra=FPInter=2, 0 (a)). Notons qu'entre les deux modes du transcodeur, le mode standard permet une réduction maximale de débit (67%) plus importante que celle du mode inter (45%).

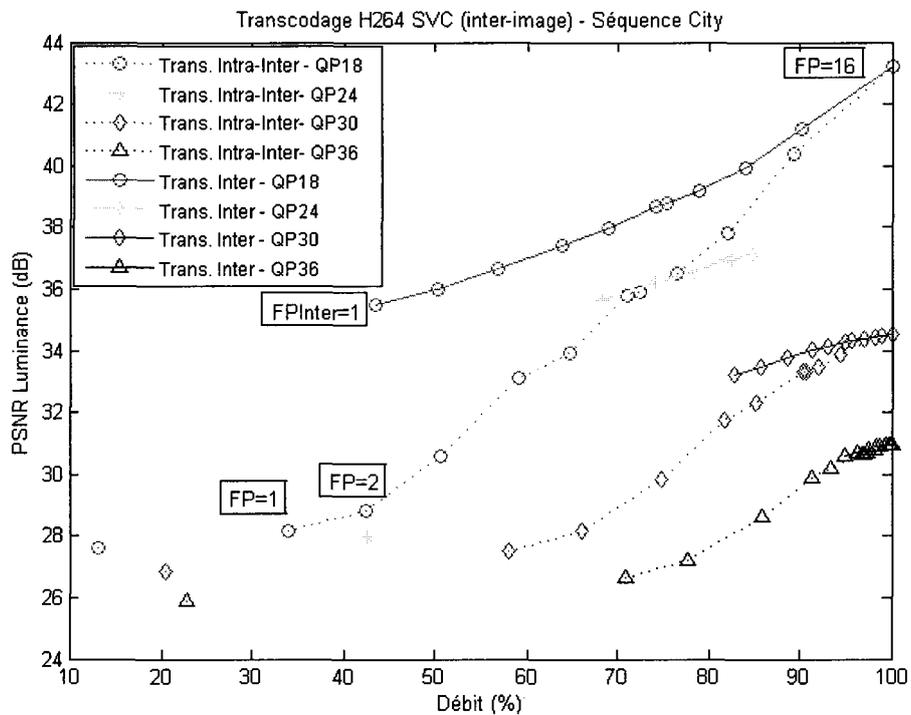
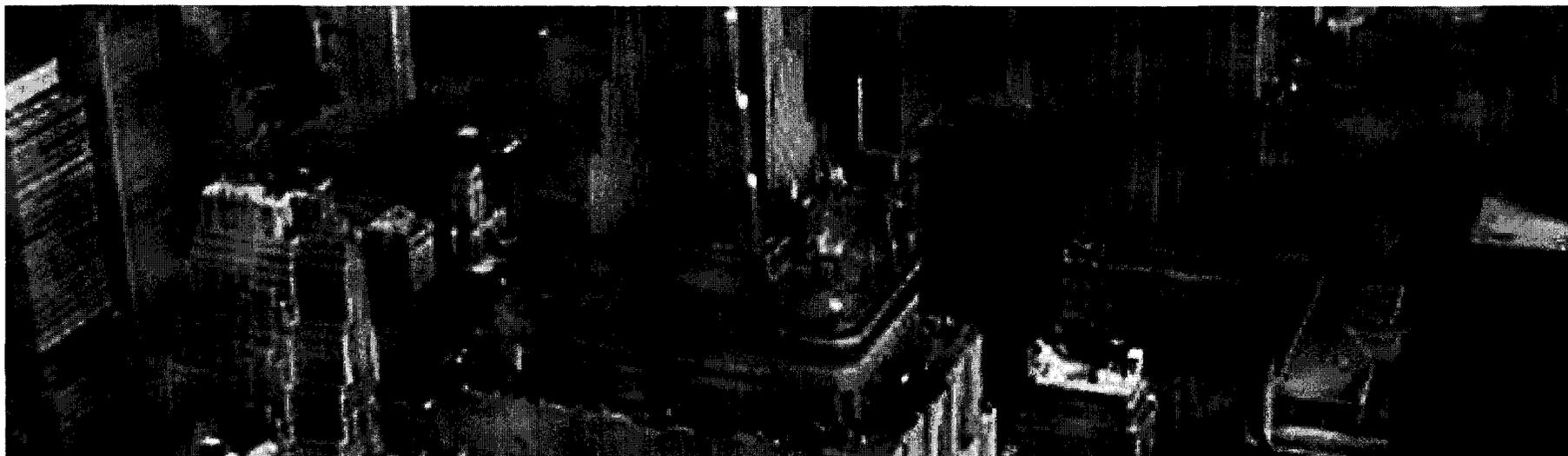


Figure 3.18 Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des macroblochs inter-image – Séquence City



(a) Transcodage en mode standard - $FP_{Intra}=FP_{Inter}=2$ (7,19 Mb/s, PSNR-Y=28,81 dB)



(b) Transcodage en mode inter - $FP_{Inter}=1$ et $FP_{Intra}=16$ (8,52 Mb/s, PSNR-Y=35,47 dB)

Figure 3.19 Illustration de l'image numéro 1 (image bidirectionnelle) de la séquence City (QP=18)

Nous remarquons que l'augmentation du paramètre de quantification accentue cette différence. Ainsi, pour un paramètre de quantification égal à 36, la réduction de débit maximale du mode inter est quasiment inexistante et se chiffre à environ 6%. Ceci peut s'expliquer par l'efficacité de la prédiction inter-image à bas débit vidéo. Dans ce cas les effets de bloc sont plus prononcés et l'estimation de mouvement trouve facilement un bloc similaire dans l'image de référence sans la nécessité de coder des résidus fréquentiels. Il y a donc peu de coefficients résiduels fréquentiels dans les macroblochs de type P ou B. Par contre, les macroblochs de type I contiennent encore quelques coefficients résiduels fréquentiels.

L'évolution du PSNR de la luminance en fonction du débit pour les séquences *Crew*, *Harbour* et *Soccer* est représentée dans la Figure 3.20, la Figure 3.21 et la Figure 3.22 respectivement. Nous constatons que le PSNR résultant du transcodage en mode inter est toujours supérieur à celui issu du transcodage en mode standard mais également que :

- Pour la séquence *Crew*, le mode standard permet une réduction maximale de débit en moyenne supérieure à 8% de celle du mode inter. Contrairement à la séquence *City*, nous remarquons que l'augmentation du paramètre de quantification accentue peu cette différence.
- Pour la séquence *Harbour*, nous notons qu'entre les deux modes du transcodeur, le mode standard permet une réduction maximale de débit (56%) plus importante que celle du mode inter (49%). Nous remarquons que l'augmentation du paramètre de quantification accentue cette différence. Ainsi, pour un paramètre de quantification égal à 36, la réduction maximale de débit du mode standard (26%) est plus significative et correspond au double de celle du mode inter (13%).
- Pour la séquence *Soccer*, nous notons qu'entre les deux modes du transcodeur, le mode standard permet une réduction maximale de débit (57%) plus importante que celle du mode inter (48%). Aussi, nous remarquons que l'augmentation du paramètre de quantification accentue quelque peu cette différence.

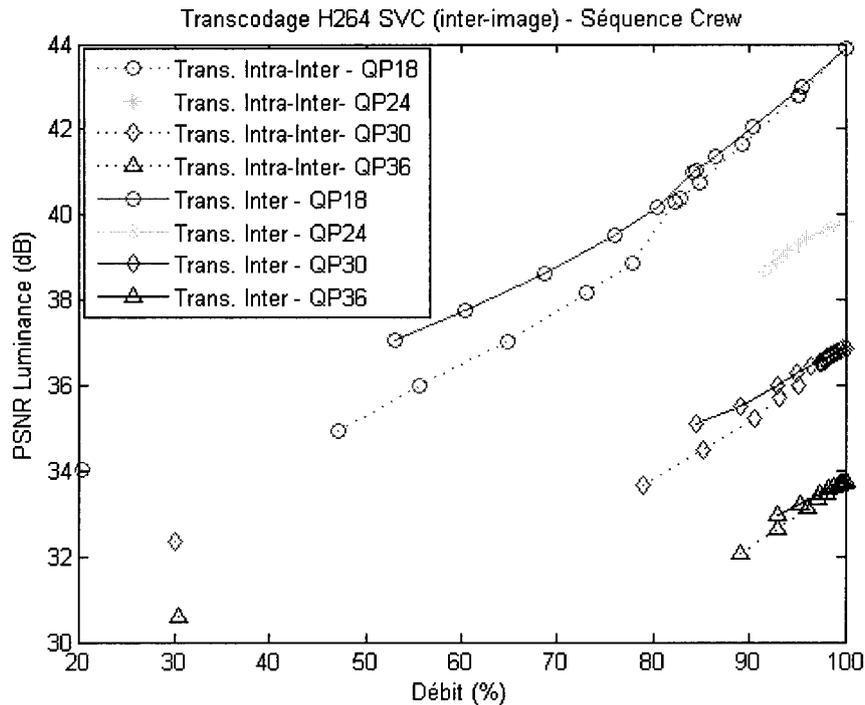


Figure 3.20 Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des macroblochs inter-image – Séquence *Crew*

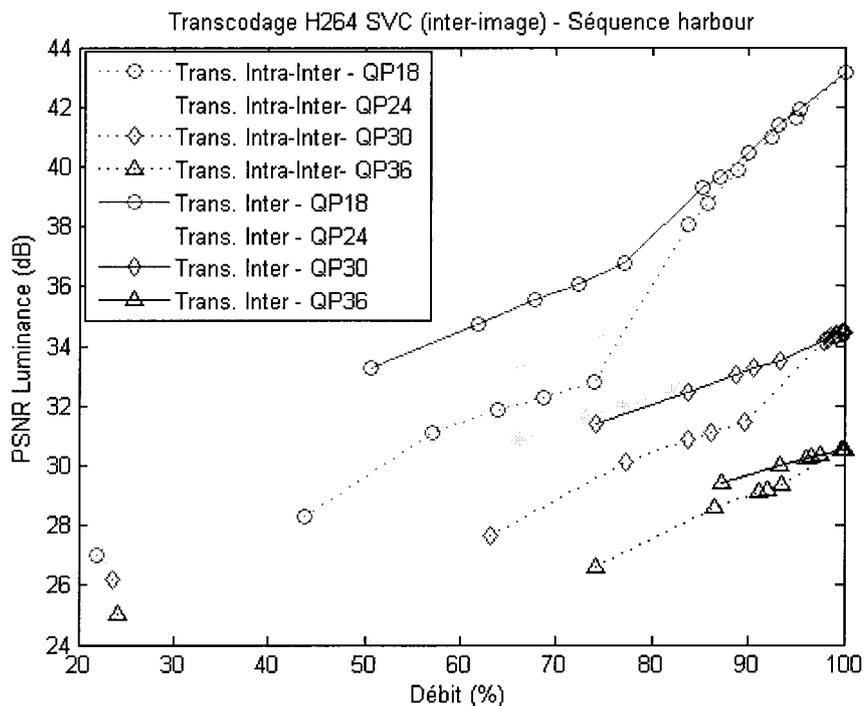


Figure 3.21 Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des macroblochs inter-image – Séquence *Harbour*

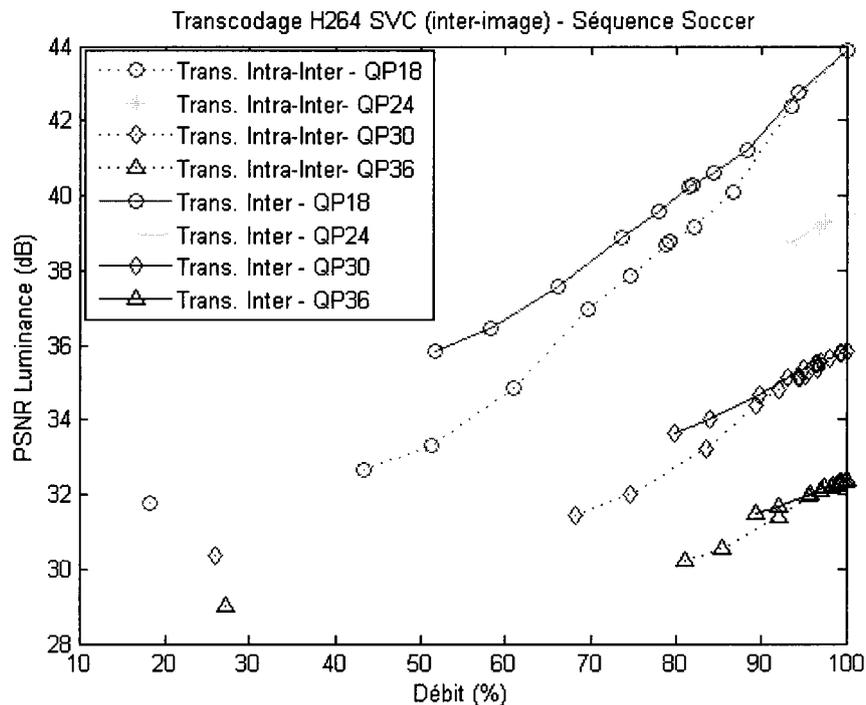


Figure 3.22 Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des macroblochs inter-image – Séquence Soccer

Sur l'ensemble des séquences et des paramètres de quantification utilisés, nous constatons à débit équivalent un meilleur PSNR pour le transcodeur en mode inter. La différence de qualité avec la couche de base sur-échantillonnée est encore plus accentuée avec le mode inter. Par contre, la réduction de débit maximale du mode inter est toujours inférieure à celle du mode standard. Aussi, cette différence a tendance à augmenter lorsque nous utilisons des paramètres de quantification plus élevés.

Ces résultats sont encourageants et il est envisageable de proposer une architecture de transcodage des flux H.264 SVC selon ce mode inter afin de limiter la complexité du transcodeur. Ainsi, cette architecture ne nécessite pas de forçage des macroblochs intra-couche intra-image en inter-couche intra-image (INTRA_BL). L'avantage de cette solution n'est pas négligeable puisqu'il est possible de transcoder n'importe quel flux H.264 SVC en échelonnabilité spatiale. Rappelons que dans ce cas, il est cependant nécessaire d'implémenter les deux types de codage entropique : le CAVLC (choisi dans le cadre de ces travaux) et le CABAC.

3.7 Conclusion

Nous avons présenté dans ce chapitre, à notre connaissance, la première solution de transcodage des flux H.264 SVC en échelonnabilité spatiale. L'architecture de transcodage, en boucle ouverte, est peu complexe afin de garantir un traitement temps réel des coefficients résiduels de la couche d'amélioration haute. Elle peut se résumer en trois étapes : décodage entropique, sélection par position fréquentielle puis encodage entropique des coefficients résiduels de la couche d'amélioration haute.

Tout d'abord, nous avons constaté que le transcodage des macroblochs utilisant la prédiction intra-couche intra-image (INTRA_16x16) crée un effet de glissement d'erreur due à la prédiction intra-image comparable à celui des macroblochs de type intra de H.264/AVC. Puis, nous avons remarqué que le forçage à l'encodage de ce type de macroblochs par des macroblochs utilisant la prédiction inter-couche intra-image (INTRA_BL) avait un impact négligeable sur la qualité d'image et le débit. L'ensemble des flux utilisés dans ce chapitre utilise ce forçage.

En premier lieu, nous avons commencé par appliquer ce traitement à l'ensemble des macroblochs de la couche d'amélioration haute. Nous avons constaté que cette architecture de transcodage permettait d'obtenir des flux vidéo avec une qualité progressive de la vidéo, située entre celle de la couche de base sur-échantillonnée et celle du flux dans sa qualité maximale. En fonction du paramètre de quantification utilisé à l'encodage, le débit vidéo peut être réduit jusqu'à 66%.

Puis, nous avons essayé de minimiser davantage la complexité du transcodeur en ne considérant qu'un traitement de la luminance. Dans ce cas, un léger manque à gagner au niveau de la réduction de débit est à prévoir pour des débits relativement élevés (i.e., QP=18 et QP=24).

En dernier lieu, nous avons également essayé de minimiser davantage la complexité du transcodeur en ne considérant qu'un traitement sur les macroblochs utilisant la prédiction inter-image (mode inter, appliqué aux échantillons de luminance et de chrominance). En considérant l'ensemble des séquences vidéo et des paramètres de quantification utilisés, nous constatons à débit équivalent un meilleur PSNR pour le transcodeur en mode inter. Il est intéressant de noter que ce mode de transcodage est compatible avec n'importe quel flux H.264 SVC en échelonnabilité spatiale.

Les perspectives de cette étude sont multiples. Nous envisageons d'utiliser cette architecture avec des flux H.264 SVC en échelonnabilité en qualité (cf 1.3.2.3) de type CGS. Les résultats pourraient être comparés à ceux obtenus par l'échelonnabilité en qualité de type MGS. Notons qu'en remplaçant le module de sélection fréquentielle par un module de

Chapitre 3 : Solution proposée de transcodage à complexité réduite des flux H.264 SVC en échelonnabilité spatiale

requantification des coefficients résiduels, il est envisageable de réduire davantage le débit mais au détriment de la qualité vidéo et d'une légère hausse de la complexité du transcodeur. Enfin, similairement à notre transcodeur H.264/AVC, il pourrait être intéressant d'effectuer un traitement en boucle fermée des macroblochs utilisant la prédiction intra-couche intra-image. Cette architecture serait de complexité plus élevée mais permettrait de réduire davantage le débit sans avoir à effectuer un encodage spécifique.

Chapitre 4: Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

4.1 Introduction

Afin de s'adapter aux variations dans le temps des propriétés des réseaux de vidéocommunication et/ou aux caractéristiques hétérogènes des terminaux d'affichage, il est souvent nécessaire de pouvoir adapter de façon dynamique les contenus vidéo au sein de la chaîne de diffusion. Pour ce faire, nous avons présenté dans ce manuscrit deux architectures de transcodage des flux H.264/AVC ou H.264 SVC. Nous proposons dans ce chapitre d'utiliser chaque architecture de transcodage dans le cadre d'une transmission vidéo haute définition afin d'optimiser conjointement les paramètres de codage de source et de canal. La Figure 4.1 illustre le principe d'optimisation conjointe d'une transmission vidéo sur un réseau afin de maximiser la qualité vidéo reçue. Plusieurs paramètres entrent en jeu lors de l'optimisation de la transmission :

- La connaissance des propriétés du canal (réponse en fréquence, bruit) est nécessaire afin d'ajuster les paramètres de transmission (débit, puissance).
- L'estimation de la distorsion du signal vidéo reconstruit permet de déterminer les paramètres du système qui maximisent la qualité perçue par l'utilisateur.

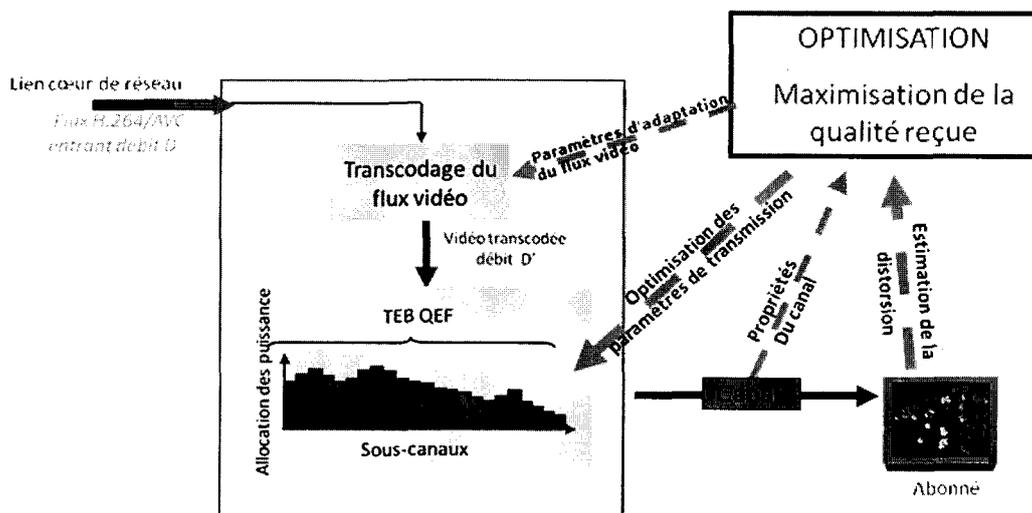


Figure 4.1 Illustration du principe d'optimisation conjointe d'une transmission sur un réseau filaire

Chapitre 4 : Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

En sortie, le module d'optimisation fournit les paramètres de transcodage vidéo et de transmission qui permettent d'obtenir la meilleure qualité visuelle en réception sous les contraintes de transmission données par l'application en terme de puissance d'émission et de taux erreur binaire (TEB, probabilité qu'un bit soit reçu en erreur).

Dans ce qui suit, nous considérons deux types de voies de diffusion : le canal ADSL (Asymmetric Digital Subscriber Line) et le canal CPL (Courant Porteur en Ligne). Il s'agit dans les deux cas de réseaux filaires qui s'appuient sur une technologie multi-porteuses. Le principe des modulations multi-porteuses est bien connu et consiste à diviser la bande passante du canal en sous-canaux et d'y utiliser une sous-porteuse par sous-canal.

Ensuite, les propriétés du canal sont utilisées afin d'optimiser les paramètres de transmission lors de l'étape d'allocation des bits et des puissances. Dans notre cas, cette étape consiste à répartir les bits et la puissance sous les contraintes de puissance d'émission et de TEB fixés de façon à obtenir le débit visé. Puis, le débit en sortie du transcodeur vidéo est adapté au moyen des paramètres de transcodage au débit disponible sur le canal de transmission après l'étape d'allocation des bits et des puissances.

Dans ce qui suit, nous proposons de mettre en oeuvre les deux architectures de transcodage proposées dans les chapitres précédents à travers deux scénarios de transmission vidéo haute définition. Nous verrons que ces deux solutions de transcodage permettent de s'adapter aux variations des voies de diffusion. Actuellement, les opérateurs ADSL n'assurent les services de télévision haute définition qu'à un nombre d'abonnés restreint, éligibles à ce type de service. Il s'agit typiquement des abonnés situés à proximité du central téléphonique, et qui possèdent une ligne téléphonique ayant un gain suffisant pour transmettre le débit requis pour le critère de qualité retenu par l'opérateur. Les clients inéligibles à cette offre peuvent prétendre aux programmes vidéo en définition standard ou ne sont pas du tout éligibles aux services de télévision. Ainsi, dans la première partie de ce chapitre, nous proposons un scénario de transmission sur le canal ADSL permettant d'étendre la zone de couverture des services de télévision haute définition, mettant en oeuvre l'architecture de transcodage des flux H.264 SVC présentée au chapitre 3.

Dans la deuxième partie du chapitre, nous proposons une solution améliorant la qualité de service des flux vidéo haute définition transmis sur le réseau électrique d'un domicile par courant porteur en ligne, soumis à de fréquents changements d'états. Ce dernier scénario de transmission utilise l'architecture de transcodage des flux H.264/AVC proposée dans le chapitre 2.

4.2 Extension de la zone de couverture des services de télévision haute définition sur le canal ADSL

Dans cette section, nous commençons par rappeler les principales caractéristiques de la technologie ADSL, puis nous décrivons notre solution d'adaptation de la transmission vidéo. La transmission ADSL ou *Asymmetric Digital Subscriber Line* utilise la ligne téléphonique traditionnelle de l'abonné afin de recevoir et transmettre des données numériques. L'ADSL est proposée pour la première fois en France en 1999 par France Télécom. Elle fait partie de la famille des technologies DSL qui regroupe un ensemble de solutions permettant la transmission haut débit de données sur la ligne téléphonique traditionnelle. Comme son nom l'indique, le débit d'une ligne ADSL est asymétrique. Les fournisseurs d'accès à Internet (FAI) proposent aux abonnés un débit descendant (du central téléphonique au domicile de l'abonné - réception) beaucoup plus important que le débit ascendant (du domicile de l'abonné au central téléphonique - émission). Il est alors possible d'accéder à Internet et à de multiples services comme la téléphonie et la télévision sur IP (*Internet Protocol*).

4.2.1 Principes de l'ADSL

La Figure 4.2 illustre une ligne ADSL reliant le réseau domestique de l'abonné au central téléphonique. La ligne ADSL, également appelée boucle locale, est composée d'une paire torsadée de cuivre allant de la prise téléphonique de l'abonné jusqu'au répartiteur téléphonique, situé dans un central téléphonique.

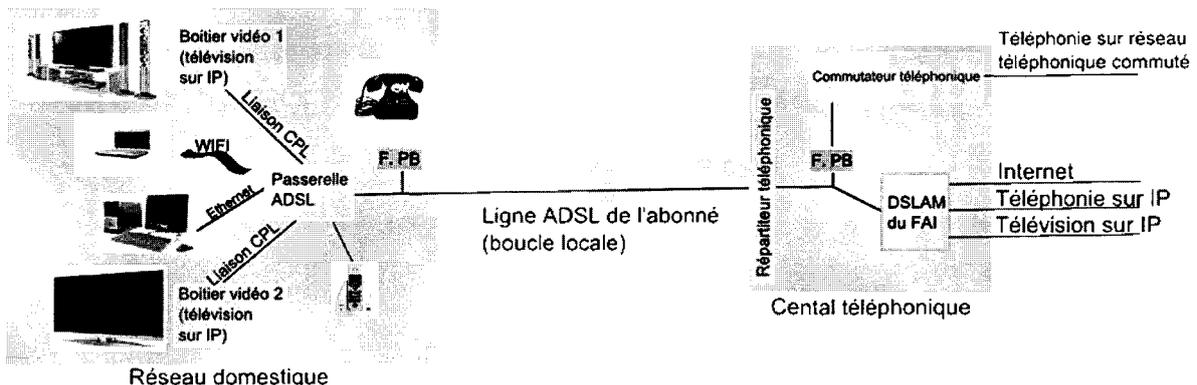


Figure 4.2 Illustration d'une ligne ADSL reliant le réseau domestique au central téléphonique

Chez l'abonné, un filtre passe-bas est placé entre une prise téléphonique et un téléphone afin d'assurer le bon fonctionnement des appels par le réseau téléphonique commuté traditionnel. Ce filtre permet la suppression des hautes fréquences utilisées par l'ADSL puisque le téléphone utilise la bande de fréquences de 300 à 3400 Hz (Figure 4.3). La

passerelle ADSL ou boîtier ADSL de l'abonné, spécifique au fournisseur d'accès à Internet, est généralement reliée directement à la prise téléphonique car elle intègre un filtre passe-haut. Nous remarquerons que le téléphone sur IP est connecté directement à la passerelle ADSL.

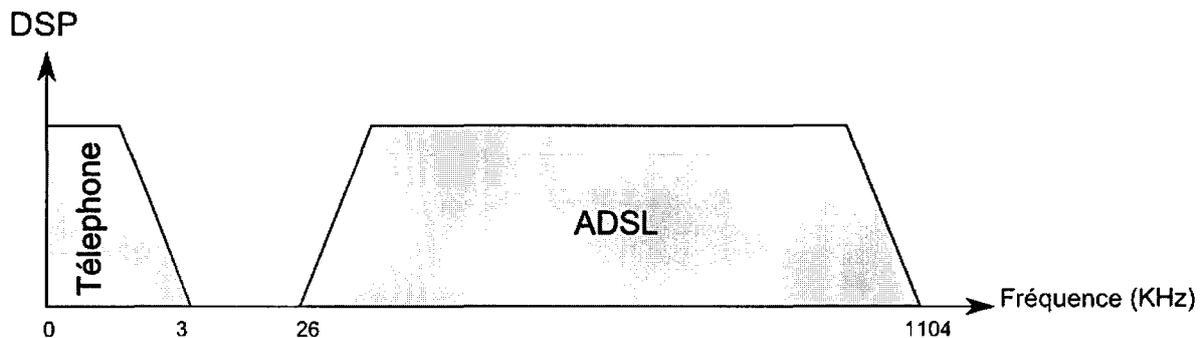


Figure 4.3 Utilisation de la bande passante téléphonique entre le téléphone traditionnel commuté et l'ADSL

À l'intérieur du central téléphonique, la paire torsadée de l'abonné arrive à l'entrée du répartiteur téléphonique. Le signal va être soit filtré par un filtre passe-bas pour une utilisation du service de téléphonie commuté, ou soit envoyé vers le DSLAM (*Digital Subscriber Line Access Multiplexer*) du fournisseur d'accès à Internet. Le DSLAM ou multiplexeur d'accès DSL récupère le flux de données ADSL transitant sur les lignes téléphoniques qui lui sont raccordées. Le flux de données ADSL comprend la « navigation » Internet mais également la téléphonie et la télévision sur IP. Ensuite le DSLAM transmet le flux de données des différentes lignes d'abonnés raccordées vers le réseau du fournisseur d'accès à Internet par fibre optique selon un multiplexage temporel. Ces données sont transmises vers le réseau Internet, un serveur de diffusion de télévision sur IP ou un serveur de téléphonie sur IP du FAI.

4.2.1.1 Perturbations rencontrées sur les lignes ADSL

Sur les lignes ADSL, nous pouvons rencontrer comme sources de perturbation : le bruit électronique, l'écho, la diaphonie et les bruits impulsifs.

Le bruit électronique regroupe le bruit thermique, le bruit des amplificateurs et le bruit de quantification du convertisseur analogique-numérique que nous retrouvons à la réception du signal. Ce bruit peut se modéliser par un bruit blanc additif Gaussien de densité spectrale de puissance mono latérale comprise entre -140 et 170 dBm/Hz [ASL92].

Dans le cas où la technique de multiplexage fréquentiel est employée, il y aura présence d'un écho résiduel correspondant aux réflexions dues à la présence d'éventuels

ponts ou bridge taps. Les ponts entraînent des perturbations dans la transmission du signal du central téléphonique vers l'abonné. Le signal émis par le central téléphonique est réfléchi au niveau du pont générant un signal atténué et retardé. Alors, le récepteur reçoit le signal émis auquel s'ajoute une partie du signal réfléchi qui est vu par le récepteur comme du bruit. L'émetteur reçoit l'autre partie du signal réfléchi constituant un écho. Si par contre les signaux descendants et montants de l'ADSL partagent la même bande de fréquences, il faudra traiter en supplément la technique d'annulation d'écho [HO96].

La diaphonie provient du couplage des paires torsadées entre elles au sein d'un même câble. Deux cas sont à distinguer : la télédiaphonie et la para-diaphonie. La télédiaphonie se produit quand les signaux sont transmis dans le même sens, appelée également Far End CrossTalk ou FEXT. La para-diaphonie se produit lorsque les signaux transitent dans des sens opposés, appelée également Near EXT ou NEXT.

Les bruits impulsifs sont provoqués par les activités de commutation dans le réseau ou par des perturbations électromagnétiques diverses. Il a été constaté que leurs caractéristiques sont très variables d'une ligne à l'autre. Ces bruits peuvent détériorer sensiblement la qualité d'une transmission ADSL si leur effet n'est pas rendu « aléatoire » en utilisant un entrelacement des données avant émission suivi d'un désentrelacement à la réception [ASL92].

4.2.1.2 Principales caractéristiques de l'ADSL

C'est en 1991 que le professeur J.M. Cioffi a introduit l'idée d'employer une modulation multi-porteuses pour l'ADSL. La transmission sur ligne ADSL utilise la modulation DMT (*Discrete Multi Tone*). Le principe de la modulation DMT est bien connu dans la littérature scientifique [CIO91]. Il consiste à diviser la bande passante du canal en N sous-canaux puis à transmettre des bits sur N sous-porteuses orthogonales utilisant chacune une modulation QAM indépendante. Un symbole DMT correspond à la superposition de N symboles QAM modulés sur N sous-porteuses indépendantes.

La Figure 4.4 illustre le schéma général d'une transmission ADSL utilisant la modulation DMT. Le signal est transmis de l'émetteur (partie supérieure du schéma) au récepteur (partie inférieure du schéma) via la ligne téléphonique. Notons que la modulation DMT est facilement mise en oeuvre grâce à une transformée de Fourier discrète inverse (TFDI) sur 2N points.

À l'émetteur, une estimation du gain et du bruit de chaque sous-canal est réalisée. L'étape d'allocation des bits et des puissances consiste à optimiser la transmission en ajustant le nombre de bits et la puissance sur chacun des sous-canaux. Un canal dont le bruit est faible se voit attribuer un nombre de bits et une puissance plus élevés. De nombreux algorithmes

d'allocation des bits et des puissances ont été proposés dans la littérature. Le premier algorithme développé pour les modulations multi-porteuses est celui de Hughes-Hartogs [HUG87] [COL99]. Par la suite, d'autres algorithmes ont été proposés avec des performances comparables mais convergeant beaucoup plus rapidement. Nous citerons l'algorithme de Peter Chow, conçu dans le cadre du système ADSL [CHO95] mais également l'algorithme de Campello [CAM98]. Afin de réaliser nos simulations, nous utiliserons l'algorithme d'allocation des bits et des puissances proposé dans la thèse de Charlene Goudemand au sein du laboratoire IEMN-DOAE [GOU05], algorithme utilisé dans le cadre des travaux du projet TOSCANE. Cet algorithme permet de minimiser la puissance totale utilisée lors de la transmission pour un TEB et un débit fixés. Contrairement à l'algorithme de Hughes-Hartogs, l'algorithme a l'avantage de converger beaucoup plus rapidement.

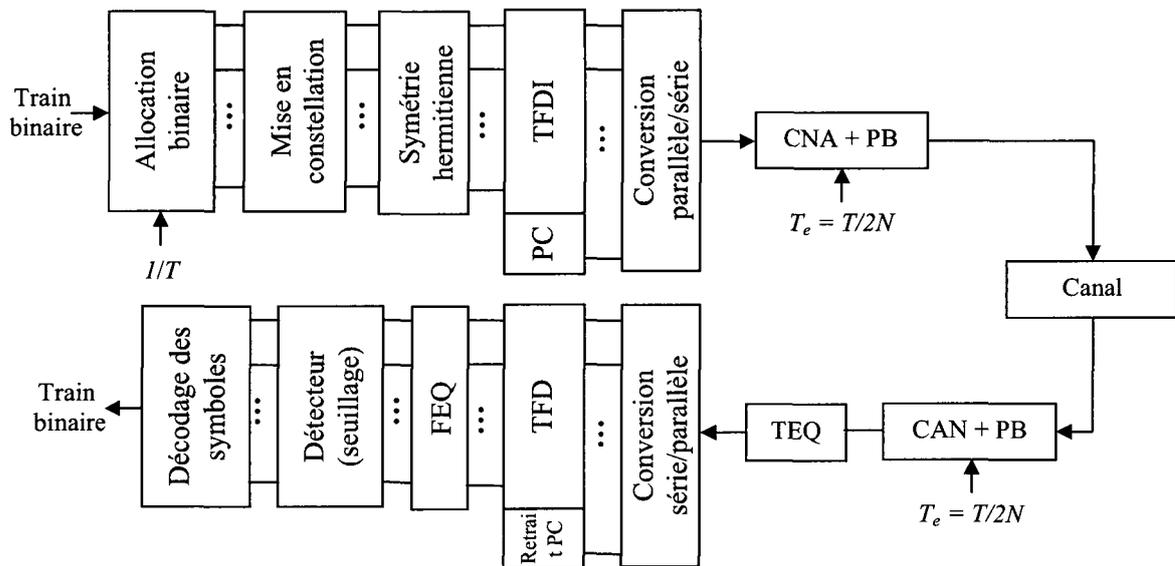


Figure 4.4 Illustration de la chaîne DMT complète [GOU05]

Notons également la présence d'un égaliseur fréquentiel (FEQ – Frequency domain Equalizer) et d'un égaliseur temporel ou TEQ (*Time domain Equalizer*). Le TEQ permet de réduire artificiellement l'allongement des symboles en diminuant la durée de la réponse impulsionnelle du canal. Il est constitué d'un filtre numérique à réponse impulsionnelle finie, inséré dans le récepteur du signal DMT entre le convertisseur analogique-numérique et le convertisseur série-parallèle [BIN00].

Le Tableau 4.1 reprend les principaux paramètres de la modulation DMT sur ADSL [ANS97]. La TFDI est réalisée sur $2N=512$ points. Ainsi, il est possible d'utiliser $N=256$ sous-canaux pour la transmission. L'ajout d'un préfixe cyclique d'une longueur de 32 échantillons à chaque symbole DMT ainsi que d'une trame de synchronisation toutes les 63

Chapitre 4 : Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

trames ADSL fait passer la fréquence d'échantillonnage de $1/T_e=2,048$ MHz à $1/T_e'=2,208$ MHz. L'intervalle fréquentiel entre les sous-canaux noté $\Delta f = \frac{1}{2NT_e'}$ est égal à 4,3125 kHz. Enfin, en ADSL, le nombre de bits alloués à chaque sous-canal DMT est compris entre 2 et 14.

Paramètre	Symbole	Valeur
Nombre de sous-canaux	N	256
Nombre de points en entrée de la IFFT	2N	512 points
Longueur du préfixe cyclique	τ	32 points
Longueur d'un symbole	$2N+\tau$	544 points
Fréquence d'échantillonnage avant ajout du préfixe cyclique et de la trame de synchronisation	$F_e = \frac{1}{T_e}$	2,048 MHz
Fréquence d'échantillonnage après ajout du préfixe cyclique et de la trame de synchronisation	$F_e' = \frac{1}{T_e'}$	2,208 MHz
Espacement de fréquence	$\Delta f = \frac{1}{2NT_e'}$	4,3125 KHz
Nombre minimum de bits par sous-canal	b_{\min}	2
Nombre maximum de bits par sous-canal	b_{\max}	14

Tableau 4.1 Paramètres utilisés par la modulation DMT de l'ADSL

4.2.1.3 Évolutions de l'ADSL

L'ADSL2 ou G992.3 (juillet 2002) est une évolution de la technologie ADSL qui permet de passer d'une vitesse maximum de 8 Mb/s à 10 Mb/s. Cette norme permet notamment une meilleure fiabilité du signal par une modulation plus efficace et une meilleure immunité contre les interférences. Nous noterons également la possibilité de reconfiguration en ligne et la diminution du débit des signaux de signalisation au profit du débit utile. Par ailleurs, la consommation de puissance ainsi que le temps d'initialisation de la liaison sont diminués (3 secondes au lieu de 10). Ces avancées technologiques permettent l'augmentation de la distance d'éligibilité des lignes ADSL (distance maximale entre le DSLAM et le domicile de

l'abonné pour laquelle l'abonné est éligible aux services de télévision sur IP). Du point de vue du fournisseur d'accès à Internet, les procédures de diagnostics et de maintenance ont été perfectionnées. L'ADSL2 procure pour les fournisseurs d'accès une surveillance en temps réel des performances aux deux extrémités de la ligne torsadée [STA02].

L'ADSL2+ ou G.992.5 (mai 2003) procure un débit pour la liaison descendante significativement plus important que celui de l'ADSL/ADSL2 en doublant la bande passante utilisée. La fréquence maximale utilisée en ADSL et ADSL2 est de 1,1 MHz ; elle passe ici à 2,2 MHz. Cette norme garde la compatibilité et les avantages de l'ADSL2. Le nombre de sous-porteuses est doublé en passant de 256 à 512. Le même espacement fréquentiel entre les sous-porteuses est gardé afin d'assurer la compatibilité. Un débit de plus de 10 Mb/s peut être obtenu pour une ligne dont la longueur est inférieure à 1800 m. Comme l'atténuation augmente avec la longueur de la ligne et avec la fréquence, les performances en termes de débit de l'ADSL2+ rejoignent celles de l'ADSL2 pour des lignes longues (plus de 4 km).

Le gain étant faible pour l'abonné lors du passage de l'ADSL à l'ADSL2, l'ADSL2+ se développe davantage et peut justifier le coût du remplacement des modems ADSL par le FAI.

4.2.1.4 La télévision sur IP

Contrairement à la diffusion par satellite (DVB-S) ou par télévision numérique terrestre (TNT, DVB-T), l'ADSL ne permet pas de recevoir l'ensemble des chaînes d'un bouquet de télévision à cause de la limitation de la bande passante de la ligne de l'abonné. Par contre, le canal de transmission n'est pas commun à l'ensemble des abonnés puisque chaque abonné possède une ligne distincte. Il est alors possible de transmettre des programmes différents à chaque abonné. La sélection du programme ou service de télévision par l'utilisateur se réalise par la voie ascendante (du domicile de l'abonné au central téléphonique) de la ligne ADSL. Ce procédé permet également l'accès à des contenus audiovisuels différents des chaînes de télévision en direct, nous parlerons de vidéo à la demande (VoD).

Dans le cas de la télévision sur IP en ADSL, le flux vidéo est généralement encapsulé dans un flux de transport MPEG2 (MPEG2-TS), lui-même encapsulé dans une paquetsation de type RTP/UDP/IP ou UDP/IP. Le protocole RTP (*Real-time Transport Protocol*) est utilisé afin de garantir une transmission des données en temps réel. Dans le cadre du projet TOSCANE, la solution d'encapsuler directement les paquets vidéo dans une paquetsation RTP/UDP/IP a été retenue afin de s'adapter à la fois aux scénarii de diffusion vidéo en direct et de vidéo à la demande.

Chapitre 4 : Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

A titre d'illustration, le Tableau 4.2 regroupe les principales informations sur les flux vidéo de différentes chaînes vidéo issues du bouquet de télévision du FAI Free. La plupart des chaînes de télévision ont une définition à l'affichage de 720x576 pixels ou 1980x1080 pixels. Elles ont cependant subies un sous-échantillonnage horizontal à l'encodage (respectivement 480x576 pixels ou 1440x1080 pixels) indiqué par le rapport SAR ou Storage Aspect Ratio choisi (ici égal à 2/3 ou 3/4) de la norme de compression vidéo. Depuis de nombreuses années, l'opérateur propose à ces abonnés un certain nombre de chaînes en définition standard et encodées en MPEG-2 à débit moyen de 2,7 Mb/s.

Désignation de la chaîne de télévision	FAI Free Source : www.digitalbitrate.com
TF1	N.C.
TF1 HD	N.C.
France 2 (Bas-débit)	480x576 (16/9) 2,7 Mb/s (MPEG-2) 1,7 Mb/s (H.264/AVC)
France 2 HD	1440x1080 (16/9) 4,4 Mb/s (H.264/AVC)
France 3 (Bas-débit)	480x576 (16/9) 2,8 Mb/s (MPEG-2) 720x576 1,6 Mb/s (H.264/AVC)
France 3 HD	1440x1080 (16/9) 4,4 Mb/s (H.264/AVC)
Arte (Bas-débit)	480x576 (16/9) 2,5 Mb/s (MPEG-2) 720x576 1,4 Mb/s (H.264/AVC)
Arte HD	1440x1080 (16/9) 4,3 Mb/s (H.264/AVC)
Direct8 (Bas-débit)	544x576 (16/9) 2,7 Mb/s (MPEG-2) 544x576 1,7 Mb/s (H.264/AVC)
Direct8 HD	1440x1080 (16/9) 4,4 Mb/s (H.264/AVC)
NRJ12 (Bas-débit)	544x576 (16/9) 2,7 Mb/s (MPEG-2) 720x576 1,7 Mb/s (H.264/AVC)
NRJ12 HD	1440x1080 (16/9) 4,3 Mb/s (H.264/AVC)

Tableau 4.2 Informations sur le flux vidéo de différentes chaînes issues du bouquet de télévision du FAI Free

Dès 2007, Free étend la zone d'éligibilité à son service de télévision sur IP sans changer son infrastructure ADSL mais grâce à sa nouvelle Freebox HD (version 5). Celle-ci est équipée d'une puce ATEME permettant de décoder du H.264/AVC en temps réel. Il est alors possible

de visualiser certaines chaînes de télévision avec une connexion à relativement bas débit grâce au H.264/AVC. Les débits indiqués sur le tableau sont de l'ordre de 1,7 Mb/s.

L'opérateur propose également certaines chaînes en haute définition encodées en H.264/AVC à des débits relativement faibles (de l'ordre de 4,4 Mb/s). La qualité n'est pas tout à fait en rendez-vous pour de la haute définition mais apporte un changement notable par rapport à la définition standard. D'autres opérateurs (comme Orange) proposent la télévision haute définition à un débit beaucoup plus élevé (environ 10 Mb/s) mais n'est proposée qu'à une minorité des abonnés ADSL dont la ligne le permet.

4.2.2 Objectif du scénario

L'objectif de ce premier scénario de transmission vidéo est d'étendre la zone d'éligibilité des services de télévision haute définition d'un opérateur ADSL.

Nous considérons que le flux vidéo a été encodé par l'extension échelonnée de H.264/AVC. Afin de concevoir un scénario réaliste de transmission vidéo, nous avons opté pour une couche de base en définition standard (720x576 pixels – 25 images par seconde) et une couche d'amélioration en haute définition (1280x720 pixels – 50 images par seconde) utilisant les échelonnabilités spatiale et temporelle.

Comme illustré en Figure 4.5 : en utilisant ce type de flux vidéo H.264 SVC, il est possible de transmettre le flux intégralement (couche de base et couche d'amélioration) jusqu'à une distance notée d_1 entre le domicile du dernier abonné éligible aux services de télévision haute définition IPTV sur ADSL et le DSLAM. Cette distance correspond à la distance initiale d'éligibilité aux services de télévision haute définition.

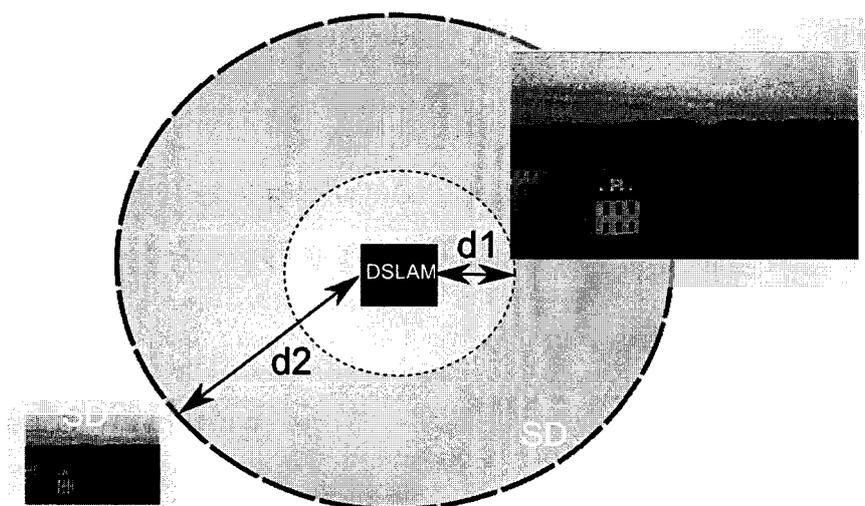


Figure 4.5 Représentation de la zone d'éligibilité initiale pour le service de télévision haute définition et le service de télévision en définition standard

Moyennant un traitement du flux par sélection de couche au niveau du DSLAM, il est également possible de transmettre uniquement la couche de base en définition standard pour des longueurs de lignes plus importantes. La distance $d2$ est la distance initiale d'éligibilité aux services de télévision en définition standard, correspondant à la longueur maximale de la ligne au delà de laquelle il n'est plus possible de bénéficier de programmes en définition standard.

Hormis le filtrage des images bidirectionnelles de la couche d'amélioration (rendu possible par l'échelonnabilité temporelle), ce type de transmission vidéo ne permet pas de fournir une qualité vidéo intermédiaire entre la haute définition et la définition standard. Comme illustré en Figure 4.6, nous allons utiliser le transcodeur H.264 SVC (proposé dans le chapitre 3) afin de diminuer le débit du flux vidéo et de pouvoir le transmettre dans la définition la plus haute (celle de la couche d'amélioration) mais avec une qualité inférieure (HD Q-). Ainsi, pour toute distance $d3$ telle que $d1 < d3 < d2$, nous proposons une qualité vidéo intermédiaire.

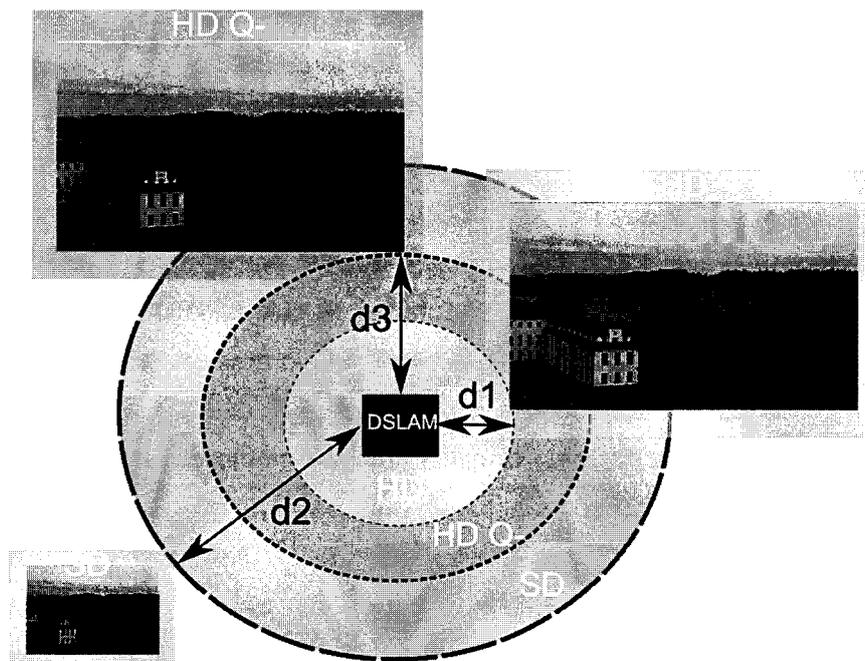


Figure 4.6 Extension de la zone d'éligibilité du service de télévision haute définition

4.2.3 Paramètres d'encodage et de transcoding vidéo

Les flux SVC utilisés dans cette section ont été encodés avec l'encodeur JSVM 9.15 configuré en échelonnabilité spatiale avec une couche de base et une seule couche d'amélioration. La couche de base a une définition spatiale standard (720x576 pixels – 25

images par seconde). La couche d'amélioration est au format 720p50 (1280x720 pixels – 50 images par seconde). Nous avons utilisé les séquences *Mobcal*, *Parkrun*, *Shields* et *Stockholm*. Les représentations des séquences vidéo en définition standard ont été obtenues en sous-échantillonnant les séquences vidéo haute définition par le logiciel *downconvert* fourni avec le JSVM et configuré par défaut selon la désignation JVT-R006.

Comme illustré dans la Figure 4.7, la couche d'amélioration des séquences vidéo a été encodée avec un groupe d'images (GOP) égal à 12 utilisant les images bidirectionnelles hiérarchiques (notées B). L'intégralité de ces images permet une restitution maximale à 50 images par seconde. Les images bidirectionnelles non hiérarchiques (notées b) ne sont pas présentes dans la couche de base. La présence de deux fois moins d'images dans la couche de base permet une restitution maximale à 25 images par seconde.

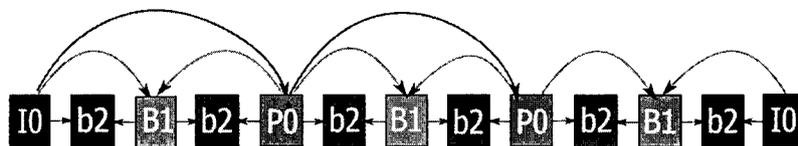


Figure 4.7 Illustration du type de GOP utilisé dans la couche d'amélioration

Notons également qu'un paramètre de quantification identique a été choisi pour l'encodage de la couche de base et la couche d'amélioration afin de réaliser une échelonnabilité spatiale indépendante d'une éventuelle échelonnabilité en qualité. Aussi, notons que l'utilisation d'un contrôle de débit fait évoluer la valeur du paramètre de quantification entre deux images ou bien entre des macroblocs successifs. Nous pouvons alors constater des sauts de qualité vidéo qui complexifieraient l'analyse de la qualité vidéo. Les débits indiqués dans la suite de cette section sont des débits moyens, calculés sur l'ensemble des 300 images de chaque séquence vidéo. En fonction des séquences, un paramètre de quantification spécifique a été utilisé afin d'obtenir des débits vidéo réalistes dans le cas d'une transmission vidéo sur ligne ADSL. Ces débits vidéo vont de 7 Mb/s à 10 Mb/s (pour des paramètres de quantification allant de 28 à 36), incluant la couche de base (débit situé entre 1,8 Mb/s et 2,5 Mb/s) et la couche d'amélioration. Nous précisons également que dans la suite de cette section, les débits présentés pour la couche d'amélioration intègrent également le débit de la couche de base car celle-ci est indispensable au décodage.

Bien que la plupart des décodeurs SVC supportent le CAVLC et le CABAC, il est intéressant que la couche de base soit décodable par la majorité des décodeurs. Le CAVLC apparaît comme le plus adapté puisqu'il est présent dans l'ensemble des profils de la norme H.264/AVC. Ainsi, nous employons le codage entropique CAVLC. Notons que notre architecture de transcodage est compatible avec une éventuelle implémentation du CABAC.

Enfin, dans une première approche du scénario, nous précisons que dans l'ensemble des expérimentations le paramètre de transcodage FP est identique pour les macroblocs utilisant la prédiction intra-image ou la prédiction inter-image ($FP_{Intra}=FP_{Inter}$).

4.2.4 Paramètres de simulation de la transmission ADSL

Dans ce manuscrit, nous avons considéré jusqu'ici que le débit d'une transmission vidéo était un débit vidéo utile, contenant uniquement le flux élémentaire H.264/AVC échelonnable ou non. Celui-ci est composé d'une succession d'unités NAL (Network Abstraction Layer) sans aucune paquetsation (cf 1.2.1.2). Afin d'établir un scénario de transmission vidéo réaliste, nous allons également prendre en compte dans cette section le surcoût de débit total de la vidéo. Dans le cadre du projet TOSCANÉ, la paquetsation RTP/UDP/IP/ATM a été retenue, nous avons chiffré ce surcoût de débit à 15%, il est principalement dû :

- Aux encapsulations successives RTP, UDP, IP nécessaires à la télévision sur IP. La taille utile des paquets RTP a été fixée à 800 octets (en-tête non compris). Comme nous avons également retenu le mode de transport de trame Asynchronous Transfer Mode (ATM) utilisé dans la norme ADSL2, il convient de prendre en compte la paquetsation sous forme de cellules ATM. Cette dernière encapsulation engendre à elle seule un surcoût de débit de l'ordre de 10% avec une taille de paquet de 53 octets et un en-tête de 5 octets. Le Tableau 4.3 synthétise la taille des en-têtes et des paquets pour ces encapsulations.
- Aux codes correcteurs d'erreurs (FEC - Forward Error Correction) utilisés afin de rendre le flux binaire moins sensible aux bruits et perturbations subies lors de la transmission. Ces codes introduisent une redondance dans le flux binaire. Ainsi, le flux est encapsulé puis protégé par codes Reed Solomon [PRO95] au niveau de la couche physique. Ces codes introduisent une redondance dans le flux se caractérisant par un surcoût de débit que nous considérons également dans les simulations suivantes. Les codes de Reed Solomon permettent d'effectuer une transmission à un taux d'erreur binaire (TEB de transmission ou TEB avant décodage FEC) beaucoup plus élevé et accroissent les performances du système. Notons que ce TEB de transmission est utilisé au sein de l'algorithme d'allocation binaire utilisé.

Encapsulation	Taille de l'en-tête (en octets)	Taille d'un paquet (en octets)
RTP	16	816
UDP	8	824
IP	20	844
ATM	5	53

Tableau 4.3 Taille des en-têtes et paquets de l'encapsulation vidéo utilisée

4.2.4.1 Caractéristiques des lignes ADSL utilisées lors de la simulation

Chaque abonné ADSL possède une ligne avec des caractéristiques bien spécifiques. En connaissant celles-ci, il est possible de déterminer par un algorithme d'allocation des bits et des puissances si la ligne est éligible ou pas aux services de télévision haute définition. Cette notion d'éligibilité sous entend la définition d'un critère ou seuil limite au delà duquel la vidéo n'est plus reçue avec un minimum de qualité suffisant qui peut varier selon les opérateurs. Dans notre cas, ce seuil a été fixé à environ une erreur non corrigée par heure. La modélisation de la fonction de transfert d'une ligne téléphonique peut se réaliser sous la forme de quadripôles dépendant de la longueur de la ligne, des différentes sections de câble. Les paramètres de résistance, d'induction, de capacité et de conductance de chaque section de la ligne sont pris en compte. Notons également la présence éventuelle d'un « pont » (ou bridge taps) qui correspond à une partie de la ligne dont l'extrémité est connectée à la ligne de l'abonné et l'autre est ouverte. L'European Telecommunications Standards Institute (ETSI) donne les valeurs de ces paramètres en fonction du diamètre de la section de ligne [ETS96].

Dans les simulations suivantes, nous allons utiliser trois modèles de lignes parmi les neuf définis par l'ETSI. Ces modèles de lignes sont également appelés boucles et sont représentés en Figure 4.8. Nous remarquons que la longueur d'une section est exprimée en kilomètres et que son diamètre est exprimé en millimètres. À part la boucle 0, les autres lignes ont une longueur ajustable (notée X) sur certaines sections. Afin de réaliser nos simulations, nous utiliserons des modèles de lignes aux caractéristiques bien distinctes : la boucle 1, la boucle 3 (avec une longueur minimale de 1500 mètres) et la boucle 8 (avec une longueur minimale de 1100 mètres et la présence de deux « ponts » de part et d'autre de cette première section). Nous précisons que l'extrémité gauche des lignes correspond au domicile de l'abonné et l'extrémité droite de la ligne correspond au DSLAM.

diminue. Aussi, nous constatons une forte atténuation du canal pour les fréquences les plus élevées.

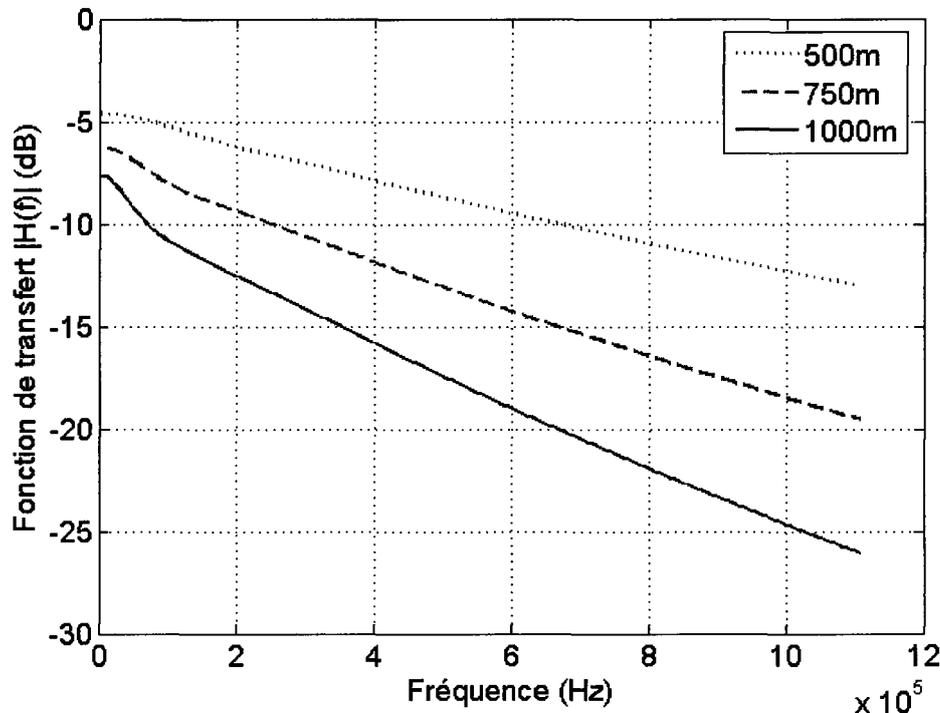


Figure 4.9 Représentation de la fonction de transfert de la ligne « boucle 1 » en fonction de la fréquence pour différentes longueurs

Comme indiqué dans la sous-section 4.2.1.1, le bruit d'une ligne ADSL peut être modélisé. Dans le cadre des simulations réalisées, nous avons utilisé un bruit de type B modélisé par l'ETSI. Il s'agit d'un bruit blanc additif Gaussien de densité spectrale de puissance mono latérale comprise entre -140 et 170 dBm/Hz.

4.2.4.2 Détermination de la zone initiale d'éligibilité

Nous avons évoqué précédemment que la distance initiale d'éligibilité à un service était la distance limite permettant la transmission du débit nécessaire au bon fonctionnement de ce service. Au delà de cette distance (un mètre supplémentaire suffit), il n'est plus possible de garantir ce débit de transmission vidéo en respectant les contraintes de TEB fixé par l'opérateur (moins d'une erreur non corrigée par heure de transmission dans le cadre du projet TOSCANE - TEB quasi error free) [DVB09] et de puissance maximale autorisée sur les lignes ADSL (19,9 dBm).

Lors des simulations, le calcul des distances initiales d'éligibilité est réalisé de manière itérative. Pour chaque ligne de longueur spécifique, la distribution optimale des bits et des puissances est déterminée grâce à l'algorithme d'allocation des bits et des puissances. Cet

Chapitre 4 : Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

l'algorithme utilise le TEB de transmission équivalent au TEB quasi error free retenu après décodage FEC (ici égal à 10^{-11}). Comme évoqué précédemment, le codage FEC ainsi que les diverses encapsulations sont prises en compte lors du calcul du débit en bit/symbole DMT, alloué par cet algorithme. La longueur de la ligne en entrée de l'algorithme d'allocation est déterminée par dichotomie :

- Elle est augmentée lorsque la puissance totale respecte la contrainte de puissance
- Elle est diminuée lorsque la puissance totale dépasse 19,9 dBm.

Enfin, le calcul est stoppé lorsque la méthode converge vers une distance stable. Alors, la distance d'éligibilité correspond à la dernière distance pour laquelle la puissance en sortie de l'allocation binaire respecte la contrainte de puissance autorisée en ADSL (19,9 dBm).

Le Tableau 4.4 synthétise la distance initiale d'éligibilité des modèles de lignes (les huit modèles de ligne sont représentés) pour trois exemples de débits, respectivement 12 Mb/s, 8 Mb/s et 6 Mb/s. Nous remarquerons que seules les lignes 1 et 2 permettent une transmission à 12 Mb/s. Aussi, les lignes 4, 7 et 8 ne permettent pas non plus une transmission de l'ordre de 8 Mb/s.

À titre indicatif, nous avons également réalisé quelques expérimentations en présence de perturbateurs ADSL sur la Boucle 1 bien que ceux-ci ne soient pas utilisés dans la suite des travaux. Comme évoqué dans la sous-section 4.2.1.1, les perturbateurs ADSL comprennent le bruit électronique, l'écho, la diaphonie et les bruits impulsifs. Nous constatons que leur présence entraîne une réduction de la distance d'éligibilité pour chaque débit proposé.

Désignation de la ligne	Nombre de perturbateurs	Distance minimale (en mètres)	Distance d'éligibilité pour un flux à 12 Mb/s (en mètres)	Distance d'éligibilité pour un flux à 8 Mb/s (en mètres)	Distance d'éligibilité pour un flux à 6 Mb/s (en mètres)
Boucle 1	0	0	44	1415	2221
Boucle 1	24	0	27	1170	2174
Boucle 1	49	0	21	1027	2135
Boucle 2	0	0	46	1726	2838
Boucle 3	0	1500	N.E.	1686	2546
Boucle 4	0	2200	N.E.	N.E.	2679
Boucle 5	0	1750	N.E.	1913	2828
Boucle 6	0	1750	N.E.	1790	2671
Boucle 6	24	1750	N.E.	N.E.	2619
Boucle 6	49	1750	N.E.	N.E.	2577
Boucle 7	0	4200	N.E.	N.E.	4312
Boucle 8	0	1100	N.E.	N.E.	1865

N.E. : non éligible

Tableau 4.4 Distance initiale d'éligibilité des modèles de lignes pour des débits de 12 Mb/s, 8 Mb/s et 6 Mb/s

Après avoir déterminé la distance initiale d'éligibilité à un service, dans les simulations suivantes, nous prolongeons la ligne ADSL par pas de 50 mètres afin de déterminer le débit de transmission correspondant, respectant les contraintes de TEB fixé et de puissance d'émission. Pour chaque tronçon de ligne avec un débit maximal, nous sélectionnons les paramètres de transcodage donnant un débit vidéo inférieur ou égal et maximisant le PSNR. Pour ces raisons, tous les points de transcodage ne sont pas forcément représentés dans les figures issues des simulations.

4.2.5 Résultats expérimentaux

4.2.5.1 Transcodage H.264 SVC par sélection en position fréquentielle des coefficients résiduels de la couche d'amélioration

Nous présentons dans cette sous-section les débits après transcodage H.264 SVC par sélection en position fréquentielle des coefficients résiduels de la couche d'amélioration pour les quatre séquences choisies. Les résultats sont regroupés du Tableau 4.5 au Tableau 4.8 pour respectivement les séquences *Mobcal*, *Parkrun*, *Shields* et *Stockholm*. La réduction de débit due au transcodage est significative et peut atteindre jusqu'à 50% du débit initial du flux total (couche de base comprise).

Nous présentons également les mesures PSNR pour la luminance et les composantes de chrominance. Nous commenterons uniquement les mesures de PSNR de la luminance dans les expérimentations suivantes dans un souci de pertinence des résultats.

La couche de base ayant une définition de 720x576 pixels à 25 images par seconde et la couche d'amélioration ayant une définition de 1280x720 pixels à 50 images par seconde, il était nécessaire de fixer une séquence de référence afin de calculer le PSNR des flux après transcodage. Nous avons choisi comme référence la séquence vidéo d'origine au format 720p50 (n'ayant pas subi d'encodage). Afin d'évaluer le PSNR des flux transcodés, nous avons choisi de le comparer à celui du flux non transcodé et à celui de la couche de base. La couche de base ayant une définition de 720x576 pixels à 25 images par seconde et la couche d'amélioration ayant une définition de 1280x720 pixels à 50 images par seconde, nous avons décidé de sur-échantillonner spatialement la couche de base (logiciel downconvert fourni avec le JSVM – algorithme de sur-échantillonnage spatial des blocs de luminance dans la norme H.264 SVC) et de la sur-échantillonner temporellement en doublant chaque image. Cette dernière technique quoique rudimentaire a l'avantage d'être facilement mise en œuvre dans les boîtiers vidéo ou les téléviseurs.

Nous constatons une différence moyenne de PSNR de plus de 8 dB entre le PSNR de la couche de base et le PSNR du flux sans transcodage. Nous pouvons observer également

Chapitre 4 : Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

une différence moyenne de PSNR de 3,8 dB entre le PSNR de la couche de base et le PSNR du flux transcodé avec un paramètre de transcodage FP égal à 1.

Enfin, nous considérons qu'un flux haute définition est de qualité vidéo acceptable si nous gardons au minimum le coefficient DC, le premier coefficient fréquentiel horizontal et le premier coefficient fréquentiel vertical de la couche d'amélioration. Ainsi l'extension de la zone d'éligibilité aux services de télévision haute définition prendra en compte les paramètres de transcodage FP allant de 3 à 16. Pour les paramètres FP=1 et FP=2, nous constatons une amélioration de qualité par rapport à la couche de base sur-échantillonnée mais nous ne pouvons pas qualifier la définition de ces flux comme étant de la haute définition.

Séquence <i>Mobcal</i> - QP=28 - 1280x720 - 300 images - 50ips (SVC – spatial -SD-720p)					
Couche(s)	FP	Débit (Mb/s)	PSNR Y(dB)	PSNR Cb(dB)	PSNR Cr(dB)
	-	9,11	35,22	37,17	40,66
	15	9,10	35,21	37,17	40,66
	14	9,01	35,12	37,17	40,66
	13	8,96	35,08	37,17	40,66
	12	8,83	34,89	37,17	40,66
	11	8,65	34,57	37,16	40,66
	10	8,48	34,39	37,16	40,66
	9	8,18	34,04	37,12	40,65
	8	7,95	33,66	37,10	40,65
	7	7,61	32,82	37,07	40,65
	6	7,32	32,31	37,05	40,64
	5	6,54	30,14	36,88	40,59
	4	6,16	29,73	36,80	40,57
	3	5,84	29,54	36,68	40,51
	2	5,53	29,40	36,48	40,39
CB + CA	1	4,74	28,26	36,19	40,16
CB up	-	2,30	25,64	35,55	39,15

Tableau 4.5 Transcodage H.264 SVC de la séquence *Mobcal*

Chapitre 4 : Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

Séquence <i>Parkrun</i> - QP=36 - 1280x720 - 300 images - 50ips (SVC – spatial -SD-720p)					
Couche(s)	FP	Débit (Mb/s)	PSNR Y(dB)	PSNR Cb(dB)	PSNR Cr(dB)
	-	9,95	27,93	36,74	39,07
	15	9,93	27,92	36,74	39,07
	14	9,82	27,86	36,74	39,07
	13	9,76	27,83	36,74	39,07
	12	9,65	27,76	36,74	39,07
	11	9,40	27,55	36,74	39,07
	10	9,18	27,40	36,74	39,07
	9	8,78	26,98	36,74	39,07
	8	8,45	26,64	36,74	39,07
	7	8,06	26,27	36,74	39,07
	6	7,89	26,17	36,74	39,07
	5	7,23	25,47	36,74	39,07
	4	6,71	25,04	36,74	39,07
	3	6,24	24,69	36,72	39,06
	2	5,79	24,38	36,65	39,04
CB + CA	1	4,96	23,72	36,63	39,03
CB up	-	2,39	20,78	35,71	38,65

Tableau 4.6 Transcodage H.264 SVC de la séquence *Parkrun*

Séquence <i>Shields</i> - QP=28 - 1280x720 - 300 images - 50ips (SVC – spatial -SD-720p)					
Couche(s)	FP	Débit (Mb/s)	PSNR Y(dB)	PSNR Cb(dB)	PSNR Cr(dB)
	-	8,50	35,38	38,17	39,71
	15	8,49	35,38	38,17	39,71
	14	8,43	35,32	38,17	39,71
	13	8,40	35,29	38,17	39,71
	12	8,35	35,20	38,16	39,71
	11	8,23	34,98	38,13	39,71
	10	8,08	34,81	38,11	39,71
	9	7,65	34,09	38,00	39,67
	8	7,43	33,79	37,88	39,62
	7	7,19	33,34	37,80	39,61
	6	7,09	33,21	37,79	39,61
	5	6,69	32,44	37,58	39,55
	4	6,36	32,04	37,35	39,42
	3	5,97	31,62	37,08	39,22
	2	5,61	31,38	36,76	38,89
CB + CA	1	5,04	30,76	36,36	38,58
CB up	-	2,53	25,51	34,45	36,50

Tableau 4.7 Transcodage H.264 SVC de la séquence *Shields*

Séquence <i>Stockholm</i> - QP=28 -1280x720 - 300 images - 50ips (SVC – spatial -SD-720p)					
Couche(s)	FP	Débit (Mb/s)	PSNR Y(dB)	PSNR Cb(dB)	PSNR Cr(dB)
	-	7,01	34,98	40,56	40,53
	15	7,01	34,98	40,56	40,53
	14	6,97	34,97	40,56	40,53
	13	6,96	34,97	40,56	40,53
	12	6,95	34,96	40,56	40,53
	11	6,87	34,93	40,56	40,53
	10	6,75	34,86	40,56	40,53
	9	6,32	34,22	40,53	40,49
	8	6,15	34,11	40,51	40,47
	7	5,98	34,02	40,51	40,47
	6	5,92	33,99	40,51	40,47
	5	5,52	33,44	40,41	40,40
	4	5,28	33,25	40,36	40,35
	3	4,88	32,79	40,21	40,19
	2	4,53	32,53	39,94	39,93
CB + CA	1	4,02	31,87	39,66	39,69
CB up	-	1,77	27,46	38,57	38,55

Tableau 4.8 Transcodage H.264 SVC de la séquence *Stockholm*

Il est intéressant de noter que le PSNR de la chrominance après transcodage est très peu variable. Similairement aux explications des chapitres précédents, ce phénomène peut tout d'abord s'expliquer par le sous-échantillonnage 4:2:0 de la séquence d'origine non compressée qui supprime définitivement certaines informations de chrominance. Ainsi, à l'encodage les blocs de chrominance contiennent moins de hautes fréquences (i.e. de coefficients résiduels fréquentiels). Aussi, l'utilisation d'un paramètre de quantification élevé (QP=36) à l'encodage accentue cette différence en annulant de nombreux coefficients résiduels fréquentiels dans les blocs de chrominance

4.2.5.2 Expérimentations sur la boucle 1

Nous avons vu précédemment que la boucle 1 n'a pas de longueur minimale et possède un diamètre de 0,4 mm. Le débit de nos quatre flux SVC est compris entre 7 Mb/s et 10 Mb/s. Par simulation, nous avons vérifié que ce type de ligne permet un débit maximal de 10 Mb/s jusqu'à une distance de 358 mètres.

Sur la Figure 4.10, nous constatons que la séquence *Mobcal* encodée avec un paramètre de quantification égal à 28 peut être transmise dans sa version haute définition jusqu'à une

distance de 608 mètres (point « a »). Cette distance est notée $d1$ et correspond à la distance initiale d'éligibilité au service de télévision haute définition. Le débit du flux vidéo est égal à 9,11 Mb/s et le PSNR de la luminance est de 35,22 dB. Sans notre solution de transcodage vidéo des flux SVC, uniquement la couche de base aurait pu être transmise au delà de cette distance. Dans ce cas une technique de sélection de couche est utilisée afin de supprimer la couche d'amélioration et de ne garder uniquement que la couche de base. Le PSNR de la couche de base sur-échantillonnée spatialement et temporellement est d'environ 9 dB inférieur à celui de la couche d'amélioration en pleine qualité.

Pour les abonnés dont la ligne téléphonique a une longueur comprise entre 608 mètres (point « a ») et 2508 mètres (point « c »), l'architecture de transcodage proposée permet de recevoir quand même les services de télévision haute définition moyennant une réduction progressive de la qualité d'image sur une distance d'environ 1900 mètres. Comme proposé en section 4.2.5.1, l'extension de la zone d'éligibilité à la vidéo haute définition est rendue possible avec un paramètre de transcodage FP supérieur ou égal à 3 (point « b »). Ainsi, notre solution permet une extension significative de la zone d'éligibilité aux services de télévision haute définition de 1650 mètres.

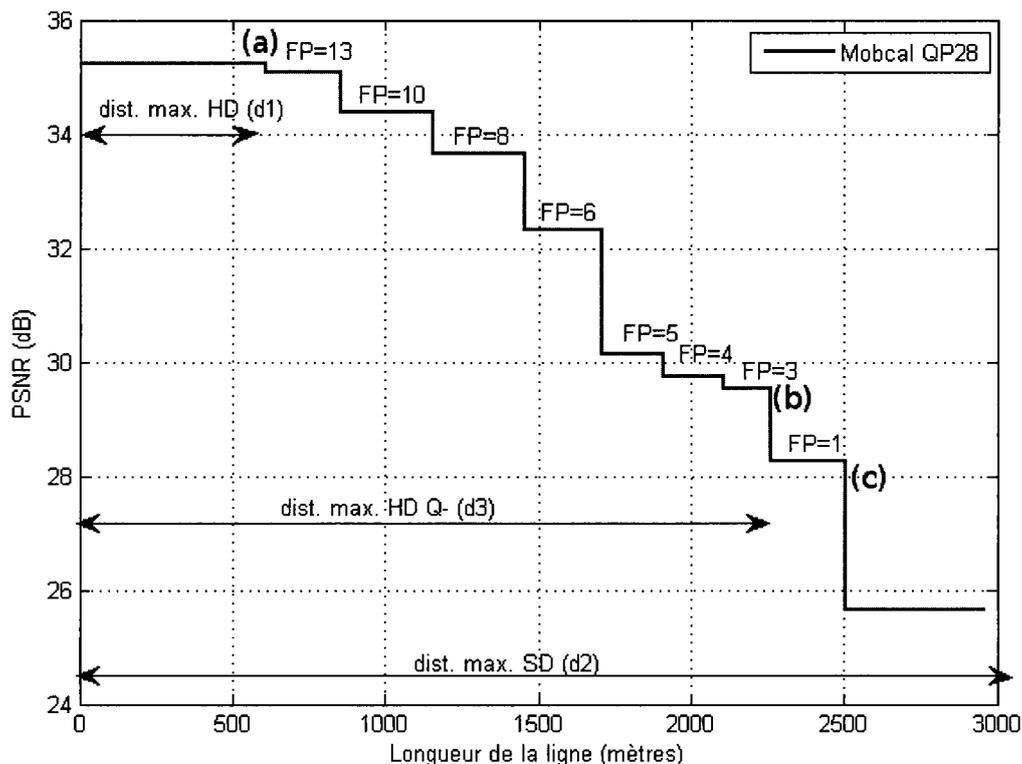


Figure 4.10 Évolution du PSNR de la luminance en fonction de la longueur de la boucle 1 pour la séquence *Mobcal*

Les résultats pour la séquence *Shields* encodée avec un paramètre de quantification égal à 28 sont représentés sur la Figure 4.11. Ce flux vidéo peut être transmis jusqu'à une distance de 1158 mètres. Le débit du flux est égal à 8,5 Mb/s et le PSNR de la luminance est de 35,38 dB. Nous constatons également avec enthousiasme que la zone d'éligibilité aux services de télévision haute définition est allongée significativement de 1100 mètres. La Figure 4.12 illustre la différence de qualité d'image entre la couche de base sur-échantillonnée et la couche d'amélioration transcodée avec un paramètre FP=8 sur une zone d'image de la séquence *Shields*.

Notons également, une différence d'environ 5 dB entre la couche de base sur-échantillonnée et le flux transcodé avec le paramètre de transcodage FP égal à 1. Cette différence de qualité vidéo s'explique par l'ajout d'un coefficient résiduel (coefficient DC) et de l'ensemble des modes de prédiction et informations de signalisation de la couche d'amélioration. Nous mentionnons que cette différence à un coût au niveau du débit vidéo puisque le débit nécessaire au décodage de la couche d'amélioration (débit de la couche de base compris) correspond au double de celui nécessaire au décodage de la couche de base.

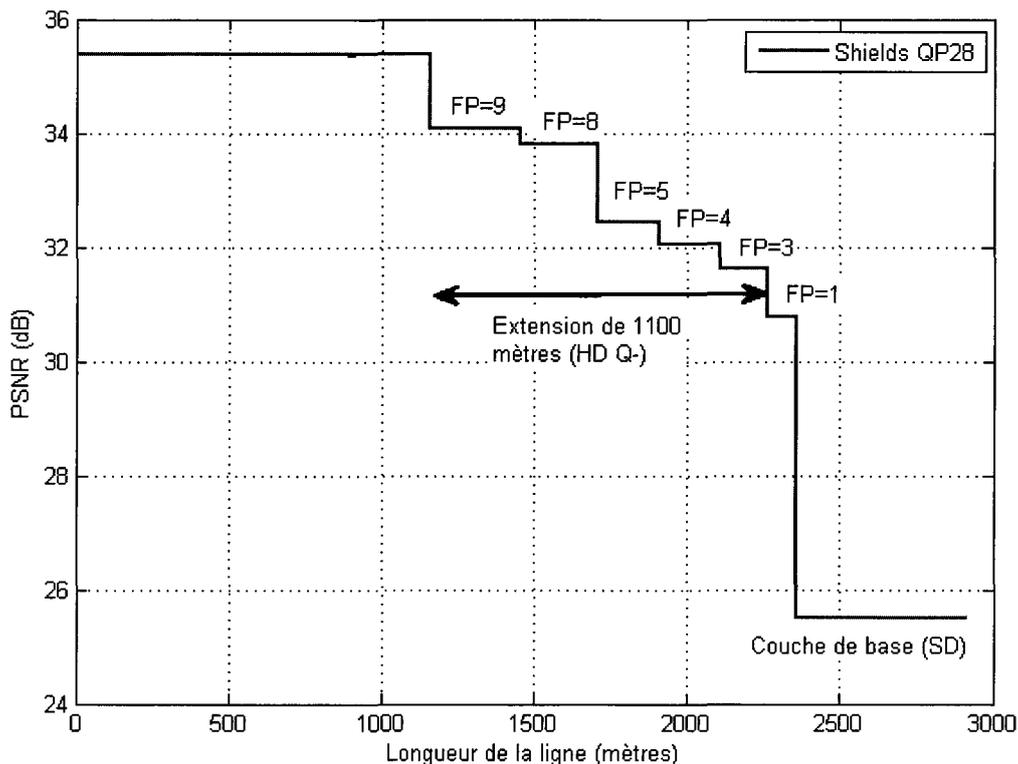


Figure 4.11 Évolution du PSNR de la luminance en fonction de la longueur de la boucle 1 pour la séquence *Shields*

Chapitre 4 : Application des solutions de transcoding vidéo aux transmissions de type ADSL et CPL

Les résultats des simulations sont similaires pour les autres séquences et sont résumés dans le Tableau 4.9. La distance initiale d'éligibilité aux services de télévision haute définition ($d1$) augmente pour des débits vidéo initiaux plus élevés.

La séquence *Stockholm* (7 Mb/s) peut être transmise jusqu'à une distance de 1708 mètres sous la contrainte du TEB et de la puissance d'émission. L'extension de la zone d'éligibilité aux services de télévision haute définition (différence entre $d3$ et $d1$) est comprise entre 800 mètres et 1750 mètres. Le PSNR de la couche de base sur-échantillonnée spatialement et temporellement est au minimum 7,0 dB inférieur à celui de la couche d'amélioration en pleine qualité.

Séquence vidéo	<i>Mobcal</i> QP=28	<i>Parkrun</i> QP=36	<i>Shields</i> QP=28	<i>Stockholm</i> QP=28
Débit initial	9,11 Mb/s	9,95 Mb/s	8,5 Mb/s	7,01 Mb/s
Distance initiale d'éligibilité HD ($d1$) et PSNR	608 m 35,22 dB	358 m 27,93 dB	1158 m 35,38 dB	1708 m 34,98 dB
Distance maximale d'éligibilité HD après extension ($d3$ max) et PSNR	2258 m 29,54 dB (FP=3)	2108 m 24,69 dB (FP=3)	2258 m 31,62 dB (FP=3)	2508 m 32,79 db (FP=3)
Distance initiale d'éligibilité SD ($d2$) et PSNR	>3000 m 25,64 dB	>3000 m 20,78 dB	>3000 m 25,51 dB	>3000 m 27,46 dB

Tableau 4.9 Distances d'éligibilité (HD, HD Q-, SD) pour les différentes séquences transmises sur la Boucle 1



Séquence *Shields* – Couche de base sur-échantillonnée



Séquence *Shields* – Couche d'amélioration transcodée avec FP=8

Figure 4.12 Illustration sur une zone de la séquence *Shields*

4.2.5.3 Expérimentations sur la boucle 3

Le modèle de ligne appelé boucle 3 est composé de deux sections. La première section a une longueur fixe de 1500 mètres et un diamètre de 0,5 mm. La deuxième section a une longueur variable et un diamètre de 0,4 mm.

Le Tableau 4.10 synthétise une partie des résultats pour les quatre séquences vidéo transmises sur la boucle numéro 3. Cette ligne autorise un débit de transmission vidéo maximal de 8,2 Mb/s pour sa longueur minimale de 1500 mètres.

Les séquences *Mobcal*, *Parkrun* (Figure 4.13) et *Shields* (Figure 4.14) ne peuvent pas être transmises au débit initial d'encodage (respectivement 9,11 Mb/s, 9,95 Mb/s et 8,5 Mb/s). Le paramètre de transcodage FP de notre transcodeur H.264 SVC permet d'ajuster le débit des flux considérés afin qu'il soit inférieur ou égal au débit maximal autorisé à cette distance (8,2 Mb/s). Après transcodage, les flux vidéo peuvent être transmis jusqu'à une distance de 1557 mètres à PSNR maximal. Le PSNR de ces flux est en moyenne inférieur de 1,1 dB à celui du flux original. Sans notre solution de transcodage, seule la couche de base aurait été transmise. Le PSNR de la couche de base sur-échantillonnée spatialement et temporellement est en moyenne inférieur de 8,9 dB à celui du flux original. Aussi, l'éligibilité à la télévision haute définition peut s'étendre jusque 2507 mètres pour les séquences *Mobcal* et *Shields*, et 2357 mètres pour la séquence *Parkrun*. Notons que pour un paramètre de transcodage FP égal à 1, il est possible de transmettre jusqu'à une distance de 2807 mètres. Le PSNR correspondant est en moyenne supérieur à 3,6 dB à celui de la couche de base.

Par contre, la séquence *Stockholm* (Figure 4.14) a été encodée avec un débit de l'ordre de 7 Mb/s et peut être transmise sur ce type de ligne sans transcodage jusqu'à une distance de 2157 mètres.

Séquence	<i>Mobcal</i> QP=28	<i>Parkrun</i> QP=36	<i>Shields</i> QP=28	<i>Stockholm</i> QP=28
Débit initial	9,11 Mb/s	9,95 Mb/s	8,5 Mb/s	7,01 Mb/s
Distance initiale d'éligibilité HD (d1) et PSNR	-	-	-	1707 m 34,98 dB
Distance maximale d'éligibilité HD après extension (d3 max) et PSNR	2507m 29,73 dB (FP=4)	2357 m 24,69 dB (FP=3)	2507 m 31,62 dB (FP=3)	2807 m 32,79 dB (FP=3)
Distance initiale d'éligibilité SD (d2) et PSNR	3307 m 25,64 dB	3307 m 20,78 dB	3307 m 25,51 dB	3457 m 27,46 dB

Tableau 4.10 Distances d'éligibilité (HD, HD Q-, SD) pour les différentes séquences transmises sur la Boucle 3

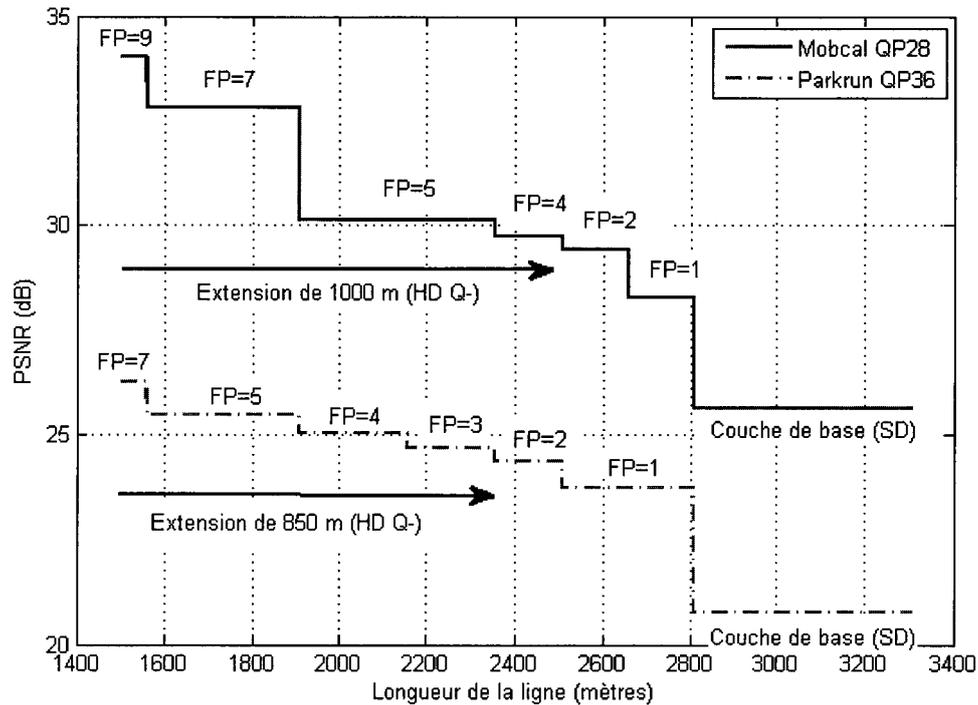


Figure 4.13 Évolution du PSNR de la luminance en fonction de la longueur de la boucle 3 pour les séquences *Mobcal* et *Parkrun*

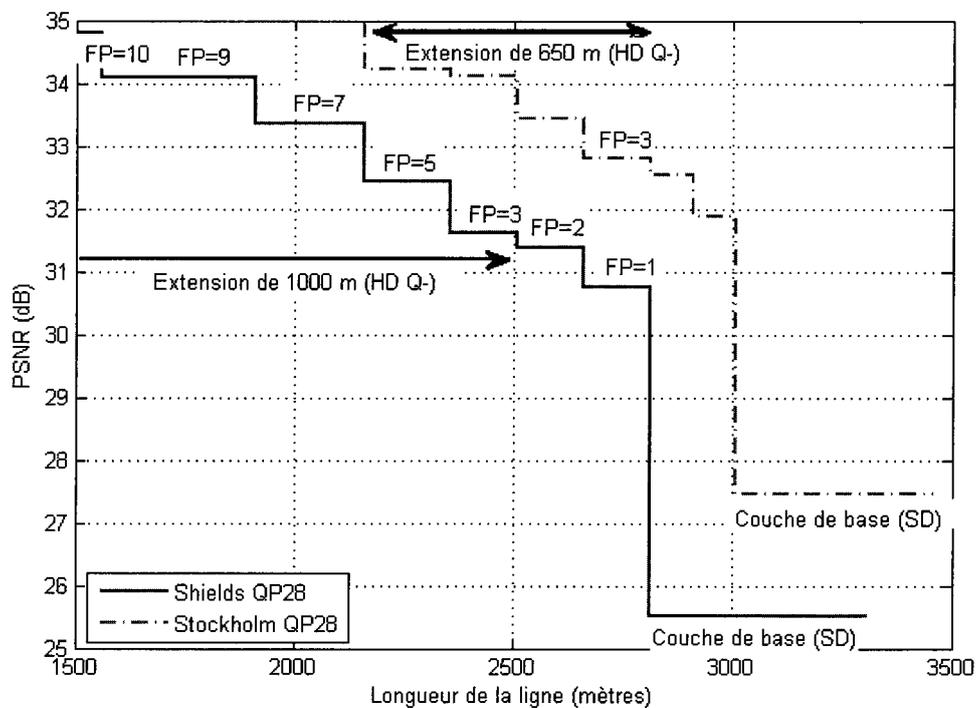


Figure 4.14 Évolution du PSNR de la luminance en fonction de la longueur de la boucle 3 pour les séquences *Shields* et *Stockholm*

4.2.5.4 Expérimentations sur la boucle 8

Le modèle de ligne appelé boucle 8 est composé de deux sections. La première section a une longueur de 1100 mètres et un diamètre de 0,4 mm. La deuxième section a une longueur variable et un diamètre également égal à 0,4 mm. La particularité de ce modèle de ligne réside en la présence de « ponts » sur la ligne de longueur égale à 500 mètres. Le premier est situé chez l'abonné et le second est localisé à 1100 mètres du domicile de l'abonné.

Le Tableau 4.11 synthétise une partie des résultats pour les quatre séquences vidéo transmises sur la boucle numéro 8. Cette ligne autorise un débit de transmission vidéo maximal de 7,9 Mb/s pour sa longueur minimale de 1100 mètres.

Tout comme pour le modèle de ligne numéro 3, les séquences *Mobcal*, *Parkrun* (Figure 4.15) et *Shields* (Figure 4.16) ne peuvent pas être transmises dans leur débit initial (respectivement 9,11 Mb/s, 9,95 Mb/s et 8,5 Mb/s). Pour chacune de ces séquences vidéo, nous avons ajusté le paramètre de transcodage FP afin que le débit soit inférieur ou égal au débit maximal autorisé à cette distance (7,9 Mb/s). Après transcodage, les flux vidéo peuvent être transmis jusqu'à une distance de 1134 mètres à PSNR maximal. Le PSNR de ces flux est en moyenne inférieur de 1,8 dB à celui du flux original. Sans notre solution de transcodage, seule la couche de base aurait été transmise. Comme indiqué précédemment, le PSNR de la couche de base sur-échantillonnée spatialement et temporellement est en moyenne inférieur de 8,9 dB à celui du flux original. Aussi, l'éligibilité à la télévision haute définition (paramètre FP supérieur ou égal à 3) peut s'étendre jusque 1784 mètres pour les séquences *Parkrun* et *Shields*, et 1934 mètres pour la séquence *Mobcal*. Notons que pour un paramètre de transcodage FP égal à 1, il est possible de transmettre jusqu'à une distance de 2034 mètres, voire 2184 mètres pour la séquence *Mobcal*. Le PSNR correspondant est en moyenne supérieur de 3,6 dB à celui de la couche de base.

La séquence *Stockholm* (Figure 4.16) a été encodée avec un débit de l'ordre de 7 Mb/s et peut être transmise sur ce type de ligne sans transcodage jusqu'à une distance de 1384 mètres. Notons que pour un paramètre de transcodage FP égal à 1, il est possible de transmettre jusqu'à une distance de 2284 mètres. Le PSNR correspondant est supérieur de 4,4 dB à celui de la couche de base.

Séquence	Mobcal QP=28	Parkrun QP=36	Shields QP=28	Stockholm QP=28
Débit initial	9,11 Mb/s	9,95 Mb/s	8,5 Mb/s	7,01 Mb/s
Distance initiale d'éligibilité HD (d1) et PSNR	-	-	-	1384 m 34,98 dB
Distance maximale d'éligibilité HD après extension (d3 max) et PSNR	1934 m 29,54 dB (FP=3)	1784 m 24,69 dB (FP=3)	1784 m 32,04 dB (FP=4)	2184 m 32,79 dB (FP=3)
Distance initiale d'éligibilité SD (d2) et PSNR	2734 m 25,64 dB	2734 m 20,78 dB	2584 m 25,51 dB	2834 m 27,46 dB

Tableau 4.11 Distances d'éligibilité (HD, HD Q-, SD) pour les différentes séquences transmises sur la Boucle 8

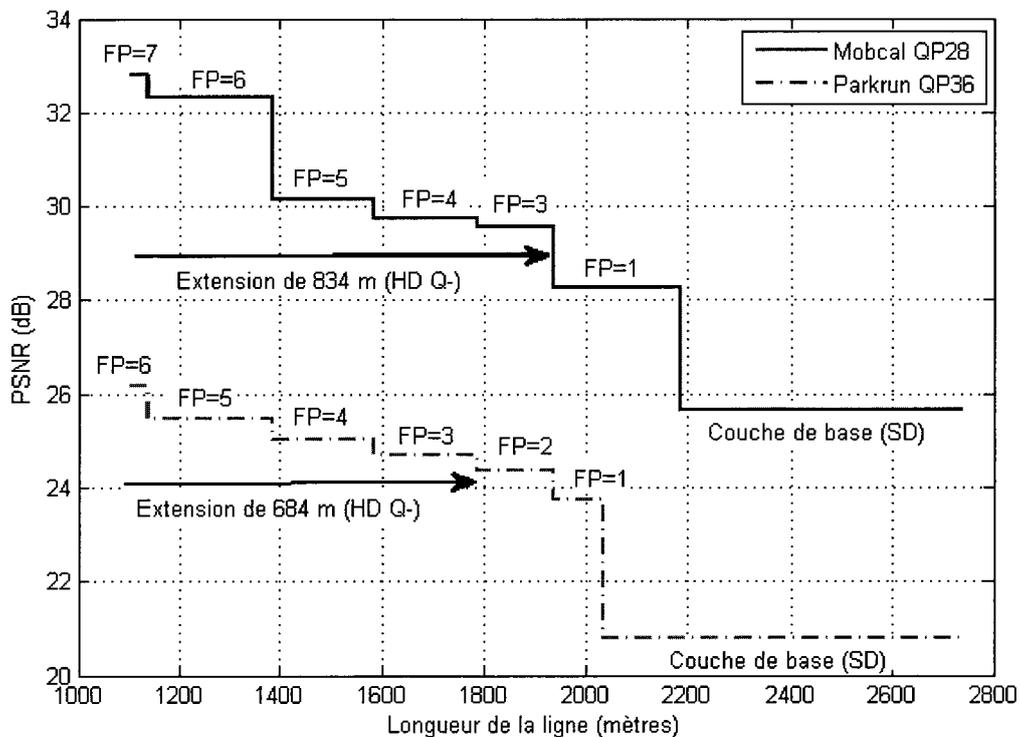


Figure 4.15 Évolution du PSNR de la luminance en fonction de la longueur de la boucle 8 pour les séquences *Mobcal* et *Parkrun*

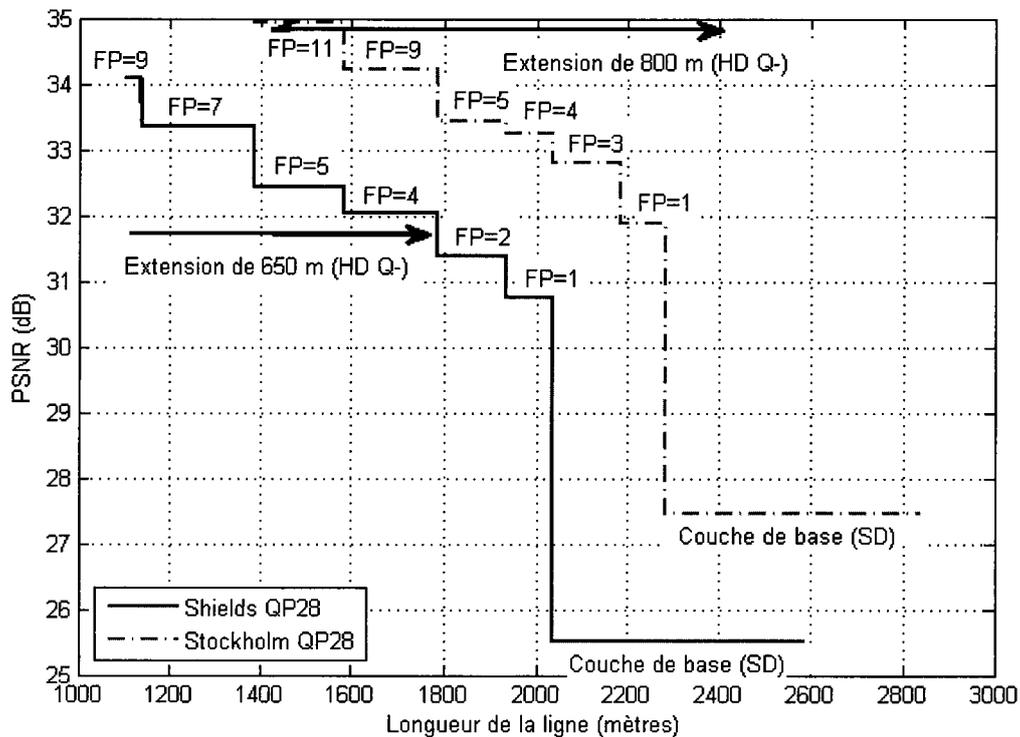


Figure 4.16 Évolution du PSNR de la luminance en fonction de la longueur de la boucle 8 pour les séquences *Shields* et *Stockholm*

Dans cette section, nous avons rappelé les principes fondamentaux de l'ADSL en citant notamment les différents types de perturbations pouvant être rencontrées et les caractéristiques principales de l'ADSL. Par diverses simulations sur des modèles de lignes différents, nous avons constaté que notre solution permettait d'élargir significativement la zone d'éligibilité des abonnés. Les flux vidéo initiaux utilisés ayant des débits légèrement différents (entre 7 et 10 Mb/s), il a été constaté que notre solution de transcodage permettait d'adapter finement le débit vidéo à celui du canal afin d'assurer une transmission vidéo haute définition moyennant une diminution progressive du PSNR. Initialement, certains flux vidéo ne pouvaient pas être transmis jusqu'à la distance minimale de la ligne (boucles 3 et 8), notre solution d'adaptation conjointe nous a permis de réaliser cette transmission en haute définition pour une légère baisse de PSNR.

4.3 Amélioration de la qualité de service de la transmission des flux H.264/AVC sur le canal CPL

Au domicile de l'abonné, la transmission des données peut se réaliser de plusieurs façons : par transmission sur câble de type Ethernet, par liaison sans fil de type IEEE 802.11 ou bien par le *Courant Porteur en Ligne* (CPL), également appelé *Power Line*

Chapitre 4 : Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

Communications (CPL). Dans le cadre du projet TOSCANE, des travaux ont été réalisés sur les réseaux domestiques sans fil, nous proposons ici de compléter ce travail en abordant la problématique des transmissions sur CPL.

La transmission de données sur le canal électrique par courant porteur en ligne est une solution intéressante d'un point de vue économique car elle permet d'utiliser une infrastructure réseau déjà existante, mais également d'un point de vue pratique car la bande passante du canal est relativement importante.

La différence majeure entre les canaux ADSL et CPL est due au caractère changeant du canal CPL, soumis à de multiples perturbations. Ainsi, il est nécessaire de réaliser fréquemment une estimation du canal et l'étape d'allocation des bits et des puissances.

Dans les expérimentations suivantes, nous allons considérer une bande de fréquences spécifique et utiliser une chaîne de transmission DMT (couramment utilisée pour les transmissions par CPL) similaire à celle de l'ADSL. Notons l'existence du standard HomePlug AV mettant en oeuvre également la modulation DMT [HPP05]. Ce standard est brièvement présenté en annexe C. Aussi, nous mentionnons la création de la norme IEEE P1901 publiée en février 2011 [IEE11]. La norme intègre deux types de modulations multi-porteuses. L'une est une modulation DMT afin d'assurer la compatibilité avec HomePlug AV et l'autre est une modulation à base d'ondelettes.

En vue d'améliorer les performances de transmission vidéo sur ce type canal, nous proposons une nouvelle solution visant à améliorer la qualité de service lors de la transmission de flux H.264/AVC haute définition sur le réseau électrique d'un domicile par courant porteur en ligne (CPL) [DEK11a]. La bande passante du canal étant relativement importante, nous proposons de la diviser en sous-bandes afin d'y transmettre des programmes indépendants dans un contexte multi-utilisateurs. Ainsi, il est par exemple possible de transmettre sur le canal :

- Un ou plusieurs programme(s) TV disponible(s) par la TV sur IP. Le débit est partagé avec la connexion Internet.
- L'ensemble des chaînes de la TNT. Environ 18 chaînes en définition standard (soit 4 Mb/s au maximum par chaîne) et 6 chaînes (pour le moment) en haute définition (8Mb/s).
- Les données du réseau local du domicile : 100 Mb/s.
- Les flux d'un système de vidéo surveillance.

4.3.1 Présentation du canal CPL

D'un point de vue physique, le canal CPL peut être considéré comme un canal multi-trajets dû aux multiples réflexions causées par les différentes impédances des appareils électriques connectés. Les effets de réflexion font devenir le canal sélectif en fréquence [ZIM00]. Tant que les appareils électriques ne changent pas d'état de fonctionnement, le canal peut être considéré comme statique.

La Figure 4.17 représente en rouge la fonction de transfert du canal CPL en fonction de la fréquence à un instant donné. Puis, la connexion ou le débranchement d'un appareil du réseau électrique va perturber le canal CPL. La nouvelle réponse en fréquence du canal est représentée en bleu. Nous constatons que ces deux réponses en fréquence sont bien différentes. Désormais, l'estimation du canal de transmission réalisée précédemment n'est plus correcte, il est alors nécessaire de recommencer cette étape afin d'assurer le bon fonctionnement de la transmission.

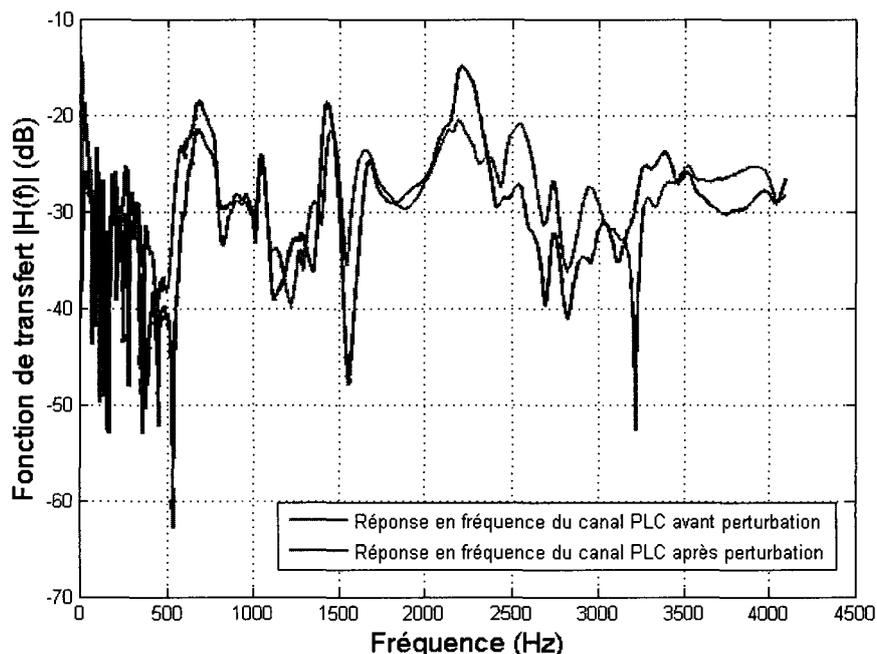


Figure 4.17 Représentation de la fonction de transfert du canal CPL en fonction de la fréquence avant (en rouge) et après (en bleu) perturbation.

Dans un premier temps, nous avons réalisé une expérimentation au sein du laboratoire afin d'évaluer la réponse temporelle du réseau électrique. Un signal sinusoïdal est transmis sur le canal puis un appareil électrique est mis sous tension. La réponse temporelle du canal est illustrée en Figure 4.18. Nous distinguons deux états stationnaires (état 1 et état 2) séparés par une phase transitoire.

L'état 1 correspond à l'état stable du canal avant la mise sous tension de l'appareil électrique. La phase transitoire se caractérise par de multiples rebonds dus à l'interrupteur d'alimentation de l'appareil. La réponse en fréquence du canal de transmission est perturbée par le branchement/débranchement d'un appareil électrique ou au changement de son mode de fonctionnement. L'état 2 correspond à l'état du canal après la mise sous tension de l'appareil électrique et stabilisation de son alimentation électrique. D'un point de vue général, le canal CPL intérieur peut être perçu comme une succession d'états stationnaires séparés par des phases transitoires [BUE06].

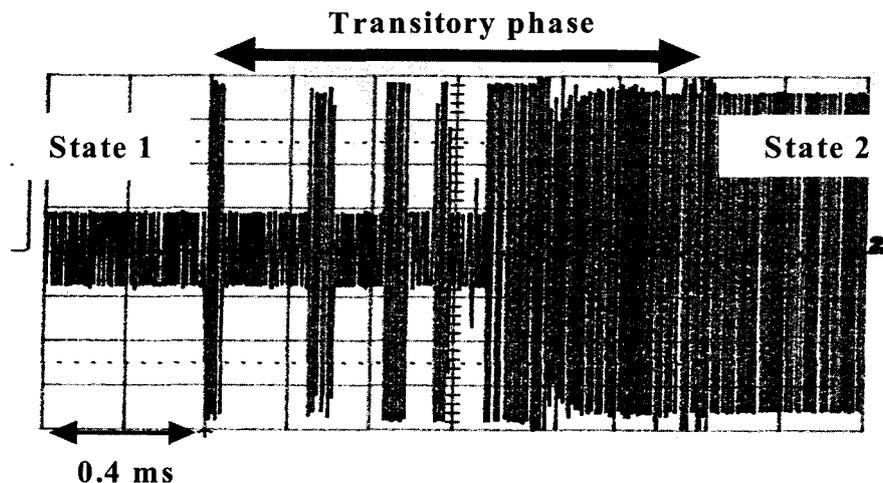


Figure 4.18 Réponse temporelle du canal à une sinusoïde transmise lors de la mise sous-tension d'un appareil électrique

Dans le cas d'un état stationnaire, les techniques présentées dans la section précédente pour la technologie ADSL sont utilisables directement en modifiant quelques caractéristiques comme le bruit du canal de transmission. Un autre paramètre à considérer est celui de la durée de l'état stationnaire. Celui-ci est beaucoup plus court qu'en ADSL. Il est donc nécessaire d'adapter la technique d'estimation du canal au CPL afin de réaliser cette opération en temps réel. Une solution envisageable consiste à introduire dans les symboles transmis des sous-porteuses pilotes aux caractéristiques connues (amplitude, phase, fréquence) préalablement par le récepteur [BUE11].

Comme la réponse en fréquence du canal varie d'un état à un autre (Figure 4.17), la capacité du canal [SHA48] peut brusquement diminuer sur une bande de fréquences. Ainsi, il est impossible d'assurer une transmission vidéo sans erreur. Afin d'assurer une qualité de service suffisante de la transmission vidéo, il est alors obligatoire d'adapter le débit vidéo à celui du canal de transmission.

4.3.2 Description du système adaptatif de transmission vidéo proposé

Le système de transmission vidéo proposé utilise la modulation multi-porteuses DMT, un algorithme d'allocation des bits et des puissances et notre solution de transcodage des flux H.264/AVC. Le fonctionnement du système de transmission vidéo est le suivant :

- Tant que le canal CPL reste dans son état initial (État 1 de la Figure 4.18), la réponse en fréquence du canal est stationnaire. L'algorithme d'allocation des bits et des puissances a convergé et permet d'assurer un débit de transmission R_1 sous une contrainte d'un TEB (QEF – Quasi Error Free) et de puissance d'émission maximale.
- Après une phase de transition, le canal de transmission se stabilise et forme un nouvel état stationnaire (État 2 de la Figure 4.18). Alors, nous pouvons effectuer une nouvelle estimation du canal et utiliser à nouveau l'algorithme d'allocation des bits et des puissances afin de déterminer le débit disponible, noté R_2 sous la même contrainte de TEB et de densité spectrale de puissance maximale autorisée (-50 dBm/Hz) [HPP05]. Nous considérons ici le cas critique où le nouveau débit du canal est diminué ($R_2 < R_1$). Il est alors nécessaire d'adapter le débit du flux vidéo au débit R_2 du canal de transmission.

4.3.3 Paramètres d'encodage et de transcodage vidéo

Les séquences vidéo utilisées dans cette section sont au format 720p50 (1280x720 pixels – 50 images par seconde). Elles ont été encodées en H.264/AVC par l'encodeur JM 12.4. Ces séquences sont respectivement *Mobcal*, *Parkrun*, *Shields* et *Stockholm*.

Afin de garantir une bonne qualité visuelle et rester cohérents avec les débits de diffusion utilisés actuellement, nous avons utilisé le contrôle de débit de l'encodeur en le paramétrant à 8 Mb/s. Similairement à la Figure 4.7, les séquences vidéo ont été encodées avec un groupe d'images (GOP) égal à 12 utilisant les images bidirectionnelles hiérarchiques (notées B). Aussi, nous employons le codage entropique CAVLC plutôt que le CABAC puisque celui-ci est présent dans l'ensemble des profils de la norme.

Après la nouvelle estimation du canal, une nouvelle allocation des bits et des puissances par sous-canal est réalisée et nous disposons d'un nouveau débit R_2 de transmission sous la contrainte d'un TEB et d'une puissance maximale d'émission. Alors, il est nécessaire de déterminer la combinaison des paramètres de transcodage (FPIIntra, FPIInter) qui fournit un débit inférieur ou égal à R_2 tout en assurant une qualité vidéo optimale au niveau du

récepteur. Il est désormais nécessaire de trouver la meilleure combinaison (FP_{Intra} , FP_{Inter}), parmi les 256 possibles, qui minimise la distorsion visuelle exprimée par une MSE normalisée, sous la contrainte du TEB fixé et de la puissance d'émission. En pratique, nous déterminons le débit en utilisant la solution de modélisation de débit présentée en section 2.6.2 :

$$\hat{B}(FP_{Intra}, FP_{Inter}) = 0.5 \log_2 (FP_{Intra} FP_{Inter}) + 4$$

$$\text{où : } 1 \leq FP_{Intra} \leq 16 \text{ et } 1 \leq FP_{Inter} \leq 16$$

$$\hat{B}(16, 16) = 8 \text{ Mb/s (débit initial)}$$

4.3.4 Paramètres de simulation de la modulation DMT sur canal CPL

Le scénario de transmission vidéo proposé est orienté multi-utilisateurs. La bande passante totale du canal CPL, allant de 0 à 25 MHz, est divisée en bandes de fréquences d'environ 1,5 MHz chacune. Des programmes ou services de télévision sont transmis simultanément aux différents utilisateurs par l'intermédiaire des différentes sous-bandes.

La réponse fréquentielle du canal illustrée en Figure 4.17 a été utilisée afin de modéliser le canal du réseau électrique et de prendre en compte des conditions de transmission réalistes. Nous considérons une bande de fréquences limitée, allant de 4,7 à 6,2 MHz. Lors des simulations, nous avons utilisé la densité spectrale de puissance du bruit mesuré avant et après la phase de transition afin de s'approcher du cas réel. La densité spectrale de puissance d'un bruit réel est représentée en Figure 4.19. Notons qu'il est possible d'utiliser des bruits modélisés lors d'une simulation de transmission sur CPL. Par exemple, un bruit de type Esmailian **[ESM00]** est une modélisation du bruit de fond basée sur de nombreuses mesures du canal CPL intérieur. La densité spectrale de puissance du bruit est modélisée par une relation décrivant la décroissance du niveau de bruit en fonction de la fréquence.

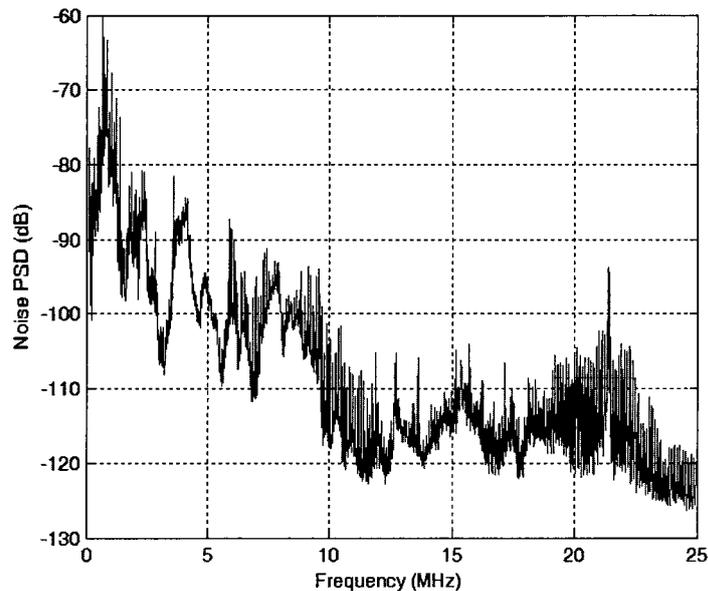


Figure 4.19 Densité spectrale de puissance d'un bruit réel sur le canal CPL

Comme dans le premier scénario de transmission vidéo, nous utilisons l'algorithme d'allocation des bits et des puissances de C. Goudemand [GOU05]. Il s'appuie sur l'algorithme de Hughes-Hartogs et prend en compte les caractéristiques de sélectivité en fréquence des sous-canaux mais également la puissance du bruit, obtenus après estimation du canal. L'algorithme permet d'atteindre un débit maximal sous la contrainte d'un TEB fixé et de densité spectrale de puissance maximale. Dans notre cas, nous gardons le même TEB afin que la transmission vidéo dans l'état 2 puisse être considérée comme quasiment sans erreur.

Le flux vidéo H.264/AVC est transmis par la modulation DMT combinée avec des codes correcteurs d'erreurs Reed-Solomon. Le flux après codage Reed-Solomon est transmis avec un TEB égal à $2,4 \cdot 10^{-4}$. Celui-ci correspond à un TEB de l'ordre de 10^{-11} après décodage Reed-Solomon.

4.3.5 Résultats expérimentaux

Nous avons représenté par simulation en Figure 4.20 l'évolution du PSNR de la luminance en fonction du numéro d'image de la séquence *Shields* (de l'image 0 à l'image 124) avant, pendant et après une perturbation. La courbe verte en pointillés représente l'évolution du PSNR de la luminance de la séquence source, encodée à 8Mb/s. Le PSNR moyen est d'environ 36 dB.

L'État 1 correspond à un premier état stationnaire. Le flux vidéo est transmis quasiment sans erreur à 8Mb/s sous la contrainte de TEB fixée. La Figure 4.21 (a) illustre une image de ce

flux vidéo. Le PSNR de la luminance de l'image considérée (numéro 93, image Bidirectionnelle) est égal à 36 dB.

Puis le canal CPL subit un changement d'état dû à la connexion ou la déconnexion d'un appareil sur le réseau électrique, nous obtenons une phase de transition. Il est très difficile de transmettre des informations pendant cet intervalle de temps. Pendant cette phase, nous simulons cette perturbation par un gel d'image sur la dernière image de l'état 1 (i.e. image 49). Nous pouvons remarquer la chute brusque du PSNR dans cet intervalle.

Ensuite, le canal se stabilise, nous obtenons un nouvel état stationnaire appelé État 2. Ici, deux cas de figures sont à prendre en compte :

- Soit aucune adaptation du débit vidéo n'est réalisée. Malheureusement, la capacité du canal de transmission a changé et il n'est plus possible de transmettre le flux vidéo au débit initial sous la contrainte d'un même TEB (QEF). Le nouveau taux d'erreur binaire (TEB_2) diminue considérablement de 10^{-11} à 10^{-7} . Alors, comme illustré en Figure 4.21 (b), la présence de bits en erreur dans le flux vidéo reçu réduit considérablement la qualité visuelle de l'image. Le PSNR de la luminance de l'image considérée est de 12,2 dB. Nous obtenons la courbe rouge, caractérisée par quelques erreurs à un TEB égal à 10^{-7} . Certaines erreurs sont présentes sur des images de référence et se répercutent sur d'autres images par la prédiction inter-image. Le décodage du flux a été réalisé grâce à un décodeur robuste H.264/AVC (fonctionnant sans dissimulation d'erreur) que nous avons développé. Notons que volontairement, nous n'avons pas injecté d'erreurs dans les entêtes des NALU afin de ne pas complexifier la gestion des erreurs par le décodeur.
- Soit nous utilisons la chaîne de transmission proposée. En utilisant cette solution, l'architecture de transcodage permet d'adapter dynamiquement le débit vidéo à celui du canal de transmission après une nouvelle allocation des bits et des puissances. Le débit vidéo retenu doit être inférieur ou égal au débit R_2 . Il est ici égal à 6,95 Mb/s au lieu de 8 Mb/s initialement. Nous pouvons constater sur la courbe bleue en trait plein, une légère baisse du PSNR par rapport à la séquence de référence. La transmission vidéo s'effectue sans problème sous la contrainte du TEB fixé préalablement. La Figure 4.21 (c) illustre une image du flux vidéo ayant subi le transcodage H.264/AVC par sélection en position fréquentielle après décodage au niveau du récepteur. La légère distorsion visuelle présente est uniquement due au transcodage. Le PSNR de la luminance de l'image est égal à 33,9 dB.

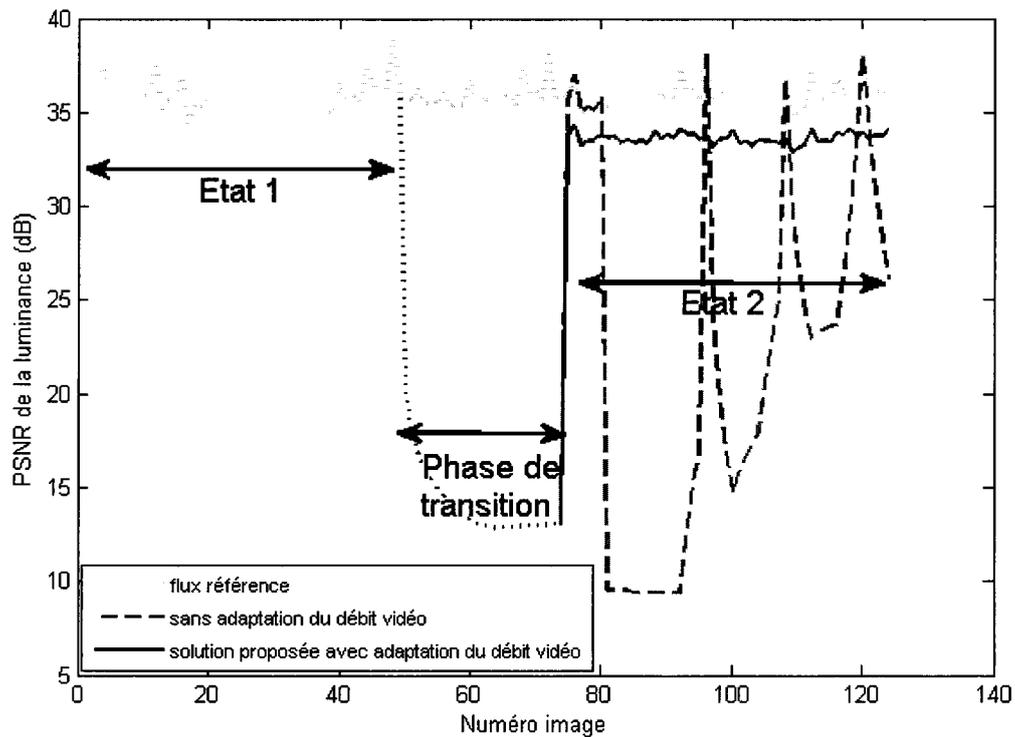


Figure 4.20 Évolution du PSNR de la luminance en fonction du numéro d'image avec ou sans adaptation du débit vidéo

Des résultats similaires ont été obtenus avec la séquence *Mobcal* (illustrée également en Figure 4.21). Le flux d'origine est transmis à 8 Mb/s et le PSNR de la luminance est égal à 38,28 dB. Sans notre solution d'adaptation du débit vidéo, le PSNR chute considérablement à cause des multiples erreurs de transmission, il est égal à 13,37 dB. Notre système permet alors la transmission de notre flux vidéo à 6,87 Mb/s. La qualité du flux vidéo résultant du transcodage est quant à elle tout à fait acceptable et s'élève à 34,04 dB.

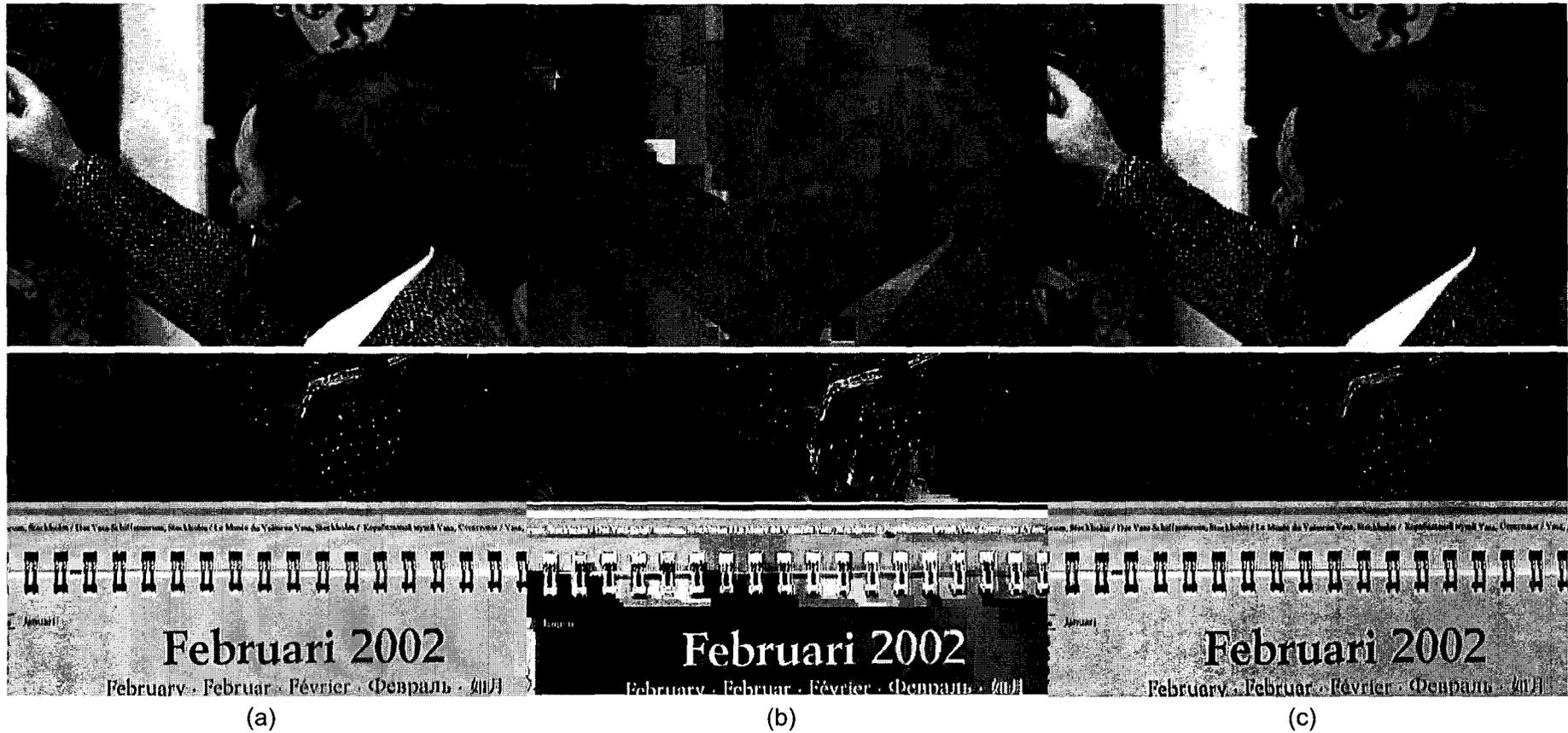


Figure 4.21 Illustration d'une même image décodée après transmission dans le canal (a) avant changement d'état (Débit = 8 Mb/s); (b) après changement d'état mais sans adaptation au canal; (c) après changement d'état et adaptation

Dans cette section, nous avons présenté un système de transmission vidéo sur réseau électrique combinant une chaîne de transmission DMT et une architecture de transcodage des flux H.264/AVC. Le but de ce système est d'assurer une transmission de haute qualité des flux vidéo sur le réseau électrique. L'algorithme d'allocation des bits et des puissances permet d'assurer un débit de transmission sous la contrainte d'un TEB. Ce débit est alors pris en considération afin d'adapter le débit du flux vidéo par l'architecture de transcodage H.264/AVC. Au niveau du récepteur, nous avons pu remarquer que la qualité vidéo était considérablement améliorée (par exemple de plus de 20 dB sur une image en erreur de la séquence *Shields*).

4.4 Conclusion

Dans ce chapitre, après avoir rappelé les principales caractéristiques du canal ADSL et les technologies mises en œuvre dans ce type de transmission, nous avons proposé un scénario de transmission vidéo visant à étendre la zone d'éligibilité aux services de télévision haute définition. En fonction de la distance du domicile de l'abonné au DSLAM, notre solution permet de recevoir quand même les services de télévision haute définition moyennant une réduction progressive de la qualité vidéo. Par diverses simulations sur des modèles de lignes différents, nous avons constaté qu'il était possible d'élargir considérablement la zone d'éligibilité des abonnés d'un minimum de 650 mètres. Les flux vidéo initiaux utilisés ayant des débits légèrement différents (entre 7 et 10 Mb/s), il a été constaté que notre solution de transcodage permettait d'adapter finement le débit vidéo à celui du canal de transmission pour une diminution progressive du PSNR. Initialement, certains flux vidéo ne pouvaient pas être transmis jusqu'à la distance minimale de la ligne (boucles 3 et 8), notre solution d'adaptation conjointe nous a permis de réaliser une transmission vidéo haute définition jusqu'à une distance maximale de 2500 mètres.

Ensuite, nous nous sommes concentrés sur la qualité de service de la transmission vidéo au sein du domicile dans le cas d'une transmission par courant porteur en ligne. Afin d'éviter une éventuelle interruption de la transmission vidéo, nous avons proposé une chaîne de transmission vidéo complète s'adaptant aux multiples variations du canal de transmission. Cette solution assure une transmission des flux vidéo haute définition H.264/AVC quasiment sans erreur par l'utilisation d'un algorithme d'allocation des bits et des puissance mais également par l'architecture de transcodage H.264/AVC. La comparaison de la qualité vidéo issue de notre solution à celle d'une transmission vidéo erronée sans adaptation conjointe est sans équivoque. La présence d'une erreur sur une image utilisée comme référence réduit considérablement la qualité vidéo sur une séquence d'images. Notre solution de

Chapitre 4 : Application des solutions de transcodage vidéo aux transmissions de type ADSL et CPL

transmission vidéo sur canal électrique résout ce problème en assurant la continuité de la transmission vidéo mais moyennant une légère baisse du PSNR du flux au niveau du récepteur.

Conclusion générale et perspectives

L'objectif de ce travail de thèse a été de proposer des architectures de transcodage vidéo de faible complexité afin de permettre l'optimisation de la transmission des flux vidéo H.264/AVC échelonnables et non-échelonnables sur canal filaire.

Une partie de ces travaux de thèse se sont inscrits au sein du projet TOSCANÉ (Transmission vidéo Optimisée Source/ Canal écheloNnable) qui s'est déroulé de 2007 à 2010, soutenu par l'Agence Nationale pour la Recherche. Ce projet avait pour objectif l'optimisation de la diffusion de flux vidéo haute définition échelonnables et non-échelonnables sur le canal ADSL puis sur le réseau domestique sans fil. Certains choix effectués au cours de cette thèse ont ainsi été guidés par le cahier des charges du projet.

Les principales contributions de ces travaux de thèse sont les suivantes :

- L'étude algorithmique et l'implémentation logicielle de plusieurs architectures de transcodage H.264/AVC à complexité réduite.
- L'étude algorithmique et l'implémentation logicielle d'une architecture de transcodage à complexité réduite des flux H.264 SVC.
- L'élaboration d'un démonstrateur temps réel en collaboration avec Thales Communications France (TCF) présentant les solutions de transcodage vidéo.

Dans ce manuscrit, nous avons commencé par présenter la norme H.264/AVC et son extension échelonnable à travers ses différents outils de codage. Afin d'adapter le débit des flux vidéo H.264/AVC à la bande passante disponible sur le canal de transmission, nous avons proposé deux principales architectures de transcodage.

La première architecture de transcodage proposée s'applique aux flux H.264/AVC non échelonnables. Cette architecture permet à la fois le transcodage des macroblocs codés par la prédiction intra-image et inter-image. Elle est fondée sur la sélection fréquentielle des coefficients résiduels transformés et se caractérise par une complexité relativement réduite, favorisant une implémentation temps réel. L'architecture de transcodage est de type boucle fermée pour les macroblocs intra et de type boucle ouverte pour les macroblocs inter. Nous avons vérifié, en termes de qualité vidéo par la métrique VQM (Video Quality Metric) et d'analyse subjective, que cette solution donnait dans la majorité des cas de meilleurs résultats qu'une approche par requantification des coefficients résiduels fréquentiels. Puis, un type de transcodage supplémentaire par sélectivité fréquentielle a été implémenté : la sélection par niveau fréquentiel. Nous avons comparé les performances de ces deux types de sélectivité fréquentielle (en termes de complexité, de réduction de débit et de qualité

vidéo) et avons décidé de retenir le transcodage par sélection en position fréquentielle pour sa faible complexité, sa meilleure flexibilité en termes de répartition des débits intermédiaires et la qualité vidéo justifiée par les mesures PSNR et VQM. Pour finir, nous avons proposé un modèle de débit de transcodage par sélection en position fréquentielle en fonction des paramètres de transcodage (*FPIntra* et *FPInter*). Il permet d'adapter facilement, sans coût de calcul supplémentaire, le débit du flux vidéo transcodé à la bande passante du canal.

La deuxième architecture de transcodage proposée s'applique aux flux H.264/AVC échelonnables. À notre connaissance, cette architecture constitue la première solution de transcodage des flux H.264 SVC en échelonnabilité spatiale. L'architecture de transcodage, en boucle ouverte, est peu complexe afin de garantir un traitement temps réel des coefficients résiduels de la couche d'amélioration haute. Afin d'éviter un effet de glissement d'erreur des macroblocs utilisant la prédiction intra-couche intra-image, nous avons d'abord opté pour un encodage des flux H.264 SVC sans ce type de macrobloc. Nous avons commencé par appliquer ce traitement à l'ensemble des macroblocs de la couche d'amélioration haute et avons constaté que cette architecture de transcodage permettait d'obtenir des flux vidéo avec une qualité progressive, située entre celle de la couche de base sur-échantillonnée et celle du flux dans sa qualité maximale. En fonction du paramètre de quantification utilisé à l'encodage, le débit vidéo peut être réduit jusqu'à 66%.

Puis, nous avons essayé de minimiser davantage la complexité du transcodeur en considérant deux modes de traitement. Le premier considère uniquement les macroblocs de luminance. Dans ce cas, un léger manque à gagner au niveau de la réduction de débit est à prévoir pour des débits relativement élevés. Le second considère uniquement les macroblocs utilisant la prédiction inter-image (mode inter). À débit équivalent, nous avons constaté un meilleur PSNR pour le transcodeur en mode inter. Dans ce mode de transcodage, un léger manque à gagner au niveau de la réduction de débit a été constaté dû à l'absence de traitement des macroblocs intra. Aussi, il est intéressant de noter que ce mode de transcodage autorise une compatibilité avec davantage de flux H.264 SVC en échelonnabilité spatiale.

Les deux architectures de transcodage sont ensuite mises en œuvre dans deux scénarios de transmission vidéo haute définition. Le premier scénario concerne une transmission de contenus vidéo haute définition au format H.264 SVC sur le canal ADSL. Jusqu'ici trois cas sont prévisibles : soit la ligne de l'abonné permet un débit compatible avec les services de télévision haute définition, soit le domicile de l'abonné est un peu trop éloigné du DSLAM et il est uniquement éligible aux services de télévision en définition standard, soit la ligne ne permet aucun service de télévision. Ainsi, nous avons vérifié que cette solution de

transcodage des flux H.264 SVC permettait d'étendre la zone d'éligibilité des services de télévision haute définition au minimum de 650 mètres en fournissant une qualité vidéo progressive aux abonnés ADSL en fonction de la distance du domicile au DSLAM.

Ensuite, nous avons proposé une solution permettant d'augmenter la qualité de service lors de la transmission par courant porteur en ligne de flux vidéo H.264/AVC en haute définition au sein du domicile. Ce type de canal de transmission étant caractérisé par de fréquents changements d'états, nous utilisons une chaîne de transmission vidéo complète intégrant notre architecture de transcodage H.264/AVC afin d'éviter une éventuelle interruption de la transmission vidéo. Les résultats montrent que moyennant une légère baisse de PSNR, notre solution assure une qualité vidéo constante contrairement à une transmission avec erreurs caractérisée par des sauts de qualité vidéo.

Nos travaux présentent plusieurs possibilités de poursuite. Tout d'abord, nous envisageons d'intégrer un contrôle de débit à l'architecture de transcodage H.264/AVC. Ce module pourrait être réalisé similairement à l'approche proposée lors du projet TOSCANE pour un transcodage par requantification, issue d'une étude de l'entropie d'un signal distribué selon une distribution de Cauchy ayant subi une quantification en fonction du nombre de zéros généré par la quantification [TOS08]. Cette solution pourrait être étendue au cas du transcodage H.264 SVC.

Aussi, l'architecture de transcodage H.264 SVC peut être utilisée avec des flux H.264 SVC en échelonnabilité en qualité de type CGS. Ainsi, les résultats pourraient être comparés à ceux obtenus par l'échelonnabilité en qualité de type MGS. Enfin, notons qu'en remplaçant le module de sélection fréquentielle des coefficients résiduels par un module de requantification, il est envisageable de réduire davantage le débit mais au détriment de la qualité vidéo et d'une légère hausse de la complexité du transcodeur.

Depuis 2010, l'annexe H de H.264/AVC propose un nouveau format d'encodage de flux stéréoscopiques ou multi-vues connu sous le nom de MVC (Multiview Video Coding). Cette extension a été retenue tout récemment comme format d'encodage des disques Blu-ray 3D permettant de visionner des films en trois dimensions avec une résolution de 1080p. Le marché audiovisuel actuel étant favorable aux systèmes stéréoscopiques et multi-vues, nous prévoyons également une adaptation de l'architecture de transcodage SVC aux flux H.264 MVC afin de réaliser une adaptation de leur transmission sur des canaux de transmission aux caractéristiques différentes. Comme évoqué précédemment, l'architecture de transcodage H.264/AVC a été implémentée sur FPGA (Field Programmable Gate Array) lors du projet TOSCANE afin de permettre un traitement temps réel des flux haute définition 1080p. Ainsi, nous pensons à moyen terme utiliser les potentialités offertes par les circuits

Conclusion générale et perspectives

FPGA (mémoires, CPU et DSP enfouis) et GPU (Graphic Processor Unit) afin d'y implémenter l'architecture de transcodage SVC puis MVC en temps réel.

Annexes

Annexe A

Annexe A.1 : Identification des NALU dans un flux H.264/AVC

Le NAL_UNIT_TYPE est une séquence de cinq bits présente dans l'entête d'une NALU (ou NAL Unit header). Il permet d'identifier le type de la NALU courante par une numérotation allant de 0 à 31.

Désignation	Numérotation	Remarques
NAL_UNIT_UNSPECIFIED_0	0	(3)
NAL_UNIT_CODED_SLICE	1	
NAL_UNIT_CODED_SLICE_DATAPART_A	2	
NAL_UNIT_CODED_SLICE_DATAPART_B	3	
NAL_UNIT_CODED_SLICE_DATAPART_C	4	
NAL_UNIT_CODED_SLICE_IDR	5	
NAL_UNIT_SEI	6	
NAL_UNIT_SPS	7	
NAL_UNIT_PPS	8	
NAL_UNIT_ACCESS_UNIT_DELIMITER	9	
NAL_UNIT_END_OF_SEQUENCE	10	
NAL_UNIT_END_OF_STREAM	11	
NAL_UNIT_FILLER_DATA	12	
NAL_UNIT_SPS_EXTENSION	13	(2)
NAL_UNIT_PREFIX	14	(1)
NAL_UNIT_SUBSET_SPS	15	(1)
NAL_UNIT_RESERVED_16	16	
NAL_UNIT_RESERVED_17	17	
NAL_UNIT_RESERVED_18	18	
NAL_UNIT_AUX_CODED_SLICE	19	(2)
NAL_UNIT_CODED_SLICE_SCALABLE	20	(1)
NAL_UNIT_RESERVED_21	21	
NAL_UNIT_RESERVED_22	22	
NAL_UNIT_RESERVED_23	23	
NAL_UNIT_UNSPECIFIED_X	24-31	(3)

- (1) Ajoutée par l'extension échelonnable (annexe G). NALU pouvant être filtrée par un décodeur non échelonnable.
- (2) NALU pouvant être filtrée par le décodeur sans avoir de répercussions sur le décodage des autres NALU.
- (3) NALU non utilisée dans la norme. Elle peut être utilisée par une application spécifique avec un encodeur et un décodeur adaptés.

Annexe A.2 : Équivalence entre le paramètre de quantification et le pas de quantification

Paramètre de quantif.	Pas de quantif.	Paramètre de quantif.	Pas de quantif.	Paramètre de quantif.	Pas de quantif.
0	0.625	19	5.5	38	52
1	0.6875	20	6.5	39	56
2	0.8125	21	7	40	64
3	0.875	22	8	41	72
4	1.0	23	9	42	80
5	1.125	24	10	43	88
6	1.25	25	11	44	104
7	1.375	26	13	45	112
8	1.625	27	14	46	128
9	1.75	28	16	47	144
10	2.0	29	18	48	160
11	2.25	30	20	49	176
12	2.5	31	22	50	208
13	2.75	32	26	51	224
14	3.25	33	28		
15	3.5	34	32		
16	4.0	35	36		
17	4.5	36	40		
18	5.0	37	44		

Annexe A.3 : Comparaison des performances des normes MPEG-2 et H.264/AVC (expérimentations)

Dans cette analyse, nous avons opté pour un encodage à débit équivalent pour les deux normes de compression et en comparant le PSNR entre la séquence vidéo d'origine (jamais compressée) et celles compressées puis décodées. L'encodage des séquences en MPEG-2 se réalise avec un GOP classique de 12 images utilisant une image intra et des images prédites et bidirectionnelles (IbbPbbPbbPbb), une prédiction de mouvement jusqu'au demi pixel. L'encodage des séquences en H.264/AVC utilise le même type de GOP, le codage entropique CAVLC mais n'utilise pas les images bidirectionnelles hiérarchiques.

Dans la Figure A.1, les résultats sont présentés pour la séquence *Crew* au format 4CIF. Les débits des flux encodés vont d'environ 700 kb/s à 9 Mb/s. À débit équivalent, nous remarquons un PSNR meilleur d'environ 3 dB pour le H.264/AVC. À PSNR équivalent, le débit du flux vidéo encodé en H.264/AVC est environ égal à la moitié de celui encodé en MPEG-2.

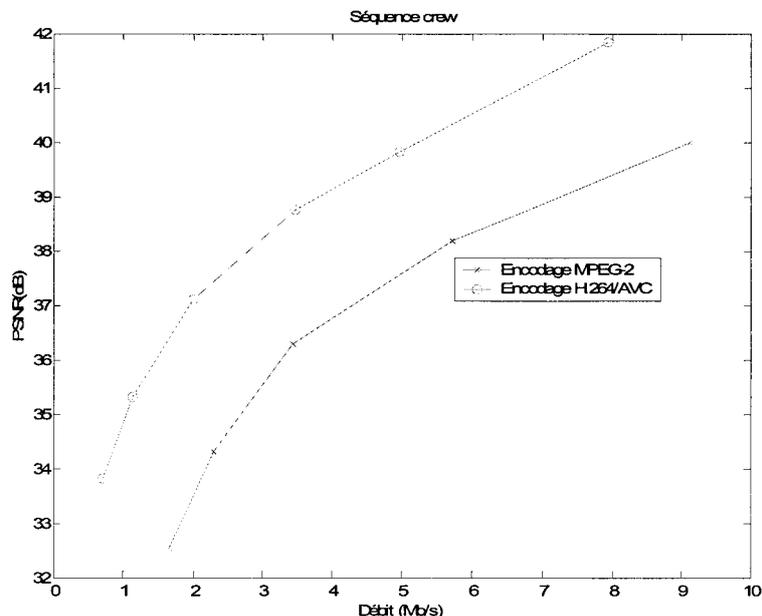


Figure A.1 Comparaison des performances de MPEG-2 et H.264/AVC pour la séquence *Crew* en 4CIF

La tendance des résultats est confirmée avec la séquence *City* (Figure A.2) au format CIF, encodée avec des paramètres identiques. À débit équivalent, nous remarquerons un écart de PSNR pouvant atteindre jusque 6 dB. Bien entendu, des débits de plusieurs méga bits ne sont pas justifiés pour les séquences vidéo au format CIF mais sont donnés à titre indicatif.

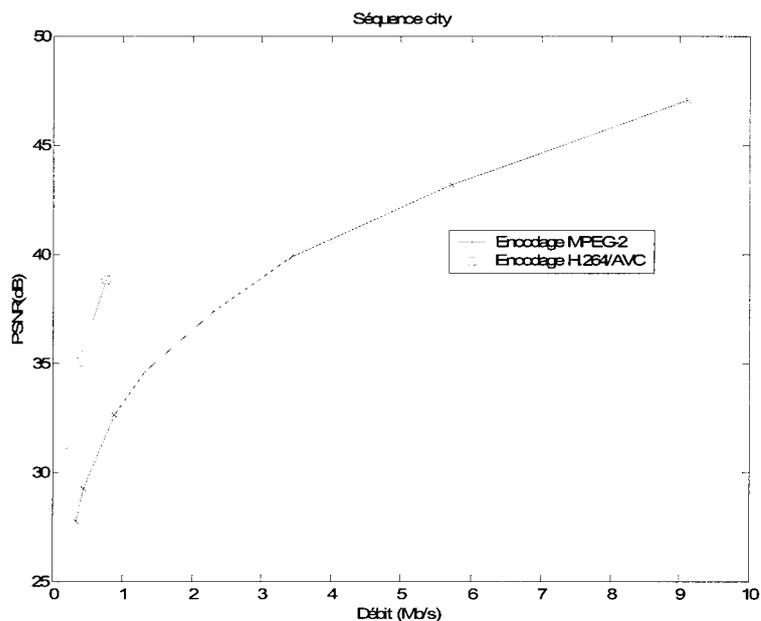


Figure A.2 Comparaison des performances de MPEG-2 et H.264/AVC pour la séquence *City* en CIF

L'efficacité de codage de H.264/AVC s'explique par l'amélioration de la prédiction inter-image par rapport à MPEG-2, ainsi qu'à l'apparition de la prédiction intra-image dans le domaine spatial.

Annexe A.4 : Influence des modes de prédiction intra sur l'encodage H264/AVC

L'objectif de cette sous-section est d'évaluer l'influence des modes de prédiction intra lors de l'encodage des images intra. Nous avons vu en sous-section 1.2.2.1 qu'il existe 9 modes de prédiction pour le mode Intra_4x4 (blocs de 4x4 pixels) et 4 modes de prédiction pour le mode intra_16x16 (macrobloc). L'idée est de confirmer que ce nombre de modes de prédiction intra est bien justifié malgré la complexité engendrée lors de l'encodage. Nous avons opté pour comparer le débit de séquences vidéo (CIF avec un paramètre de quantification constant égal à 30 ou 40) encodées exclusivement en intra-image avec l'ensemble des modes de prédiction ou uniquement le mode DC. Le mode DC a été retenu car il s'adapte à tous les cas de figure en prédisant un bloc égal à la moyenne des échantillons voisins (dans le cas de la luminance, il s'agit d'un signal lumineux moyen).

Les résultats sont présentés dans le Tableau A.1. Nous remarquons que le surcoût de débit dû au forçage en mode DC est de l'ordre de 13-23% à PSNR similaire. A part pour la séquence *Soccer* encodée avec un QP égal à 40 où le surcoût de débit atteint 36%.

Séquence	QP	Mode Intra	PSNR-Y (dB)	Débit (Mb/s)	Surcoût de débit dû au forçage en mode DC (%)
City	30	Libre	34.67	2,48	15.49
	30	DC	34.41	2,86	
	40	Libre	27.94	0,7	13.44
	40	DC	27.72	0,8	
Crew	30	Libre	36.92	1,58	23.19
	30	DC	36.58	1,94	
	40	Libre	30.81	0,5	22,05
	40	DC	30.58	0,61	
Soccer	30	Libre	35.70	1,66	22.91
	30	DC	35.57	2,04	
	40	Libre	30.28	0,45	35.96
	40	DC	30.13	0,61	

Tableau A.1 Estimation du surcoût de débit dû au forçage en mode DC en H.264/AVC Intra-image

Ces résultats justifient l'utilisation de l'ensemble des modes de prédiction intra lors de l'encodage. La différence de débit s'explique par une meilleure corrélation entre le bloc de prédiction et le bloc à coder, rendue possible par le choix du meilleur candidat parmi l'ensemble des modes de prédiction intra. Ainsi, les coefficients du bloc résiduel, issu de la différence entre le bloc à coder et le bloc de prédiction, peuvent être minimisés et codés entropiquement sur moins de bits.

Annexe B

Annexe B.1 : Comparaison de la qualité visuelle du transcodage par sélection en position fréquentielle et requantification

Séquence : Parkrun (720p50) Encodage Intra (Intra4x4-16x16) VQM-HRC : Television Model				
QP	dQP ou FP	Y-PSNR (dB)	Note VQM	Débit (Mbps)
34	-	30.43	0.1486	47.6
	dQP=1	29.05	0.1694	46.0
	FP=10	28.65	0.1554	44.7
	dQP=2	28.09	0.1840	43.5
	FP=9	27.70	0.1631	42.6
	dQP=3	27.49	0.1969	41.7
	FP=8	26.92	0.1678	41.0
	FP=7	26.26	0.1678	39.6
	FP=6	26.05	0.1708	39.0
	FP=5	24.84	0.1828	36.1
	FP=4	23.84	0.1857	34.1
	FP=3	23.11	0.1921	31.6
	dQP=4	25.58	0.3156	29.9
36	-	28.75	0.2086	37.6
	dQP=1	28.0	0.2029	37.0
	FP=10	27.64	0.2129	35.8
	dQP=2	26.78	0.2403	34.6
	FP=9	26.92	0.2160	34.5
	FP=8	26.30	0.2210	33.4
	dQP=3	26.09	0.2525	33.1
	FP=7	25.75	0.2195	32.4
	FP=6	25.59	0.2240	32.0
	FP=5	24.53	0.2332	29.8
	FP=4	23.62	0.2374	28.2
	FP=3	22.96	0.2444	26.2
	dQP=4	24.38	0.3745	24.1
38	-	27.25	0.2794	29.7
	dQP=1	26.46	0.2664	29.2
	FP=10	26.58	0.2891	28.7
	FP=9	26.04	0.2902	27.8
	dQP=2	25.60	0.2850	27.7
	FP=8	25.56	0.2916	27.0
	FP=7	25.14	0.2942	26.3
	dQP=3	24.71	0.3014	26.0
	FP=6	25.01	0.2959	26.0
	FP=5	24.11	0.3029	24.4
	FP=4	23.32	0.3042	23.2
	FP=3	22.74	0.3002	21.8
	dQP=4	23.17	0.4345	18.7

Tableau B.1 Mesures PSNR et VQM de la séquence *Parkrun*

Séquence : Mobcal (720p50) Encodage Intra (Intra4x4-16x16) VQM-HRC : Television Model				
QP	dQP ou FP	Y-PSNR (dB)	Note VQM	Débit (Mbps)
25	-	39.19	0.0134	63.4
	dQP=1	37.44	0.0297	60.8
	FP=10	36.38	0.0260	59.3
	dQP=2	36.56	0.0368	57.8
	FP=9	35.60	0.0270	57.6
	FP=8	34.56	0.0300	55.8
	dQP=3	35.88	0.0423	54.8
	FP=7	33.28	0.0402	53.7
	FP=6	32.30	0.0682	51.7
	FP=5	29.30	0.1302	46.0
	FP=4	28.43	0.1375	44.0
	dQP=4	34.25	0.0889	43.4
	FP=3	27.85	0.1478	41.8
27	-	37.60	0.0335	50.3
	dQP=1	36.26	0.0475	48.8
	FP=10	35.62	0.0423	47.6
	FP=9	34.98	0.0482	46.4
	dQP=2	35.12	0.0680	45.7
	FP=8	34.08	0.0440	45.1
	FP=7	32.93	0.0575	43.5
	dQP=3	34.44	0.0692	43.4
	FP=6	32.05	0.0852	42.0
	FP=5	29.20	0.1512	37.5
	FP=4	28.36	0.1605	35.8
	dQP=4	32.98	0.1369	35.4
	FP=3	27.80	0.1664	34.0
28	-	36.95	0.0386	45.8
	dQP=1	35.38	0.0639	44.0
	FP=10	35.29	0.0507	43.6
	FP=9	34.71	0.0550	42.5
	dQP=2	34.52	0.0730	41.6
	FP=8	33.87	0.0564	41.3
	dQP=3	33.98	0.0843	40.1
	FP=7	32.78	0.0886	39.9
	FP=6	31.95	0.0998	38.5
	FP=5	29.15	0.1511	34.6
	FP=4	28.32	0.1663	32.9
	FP=3	27.77	0.1697	31.3
	dQP=4	32.21	0.1697	30.9

Tableau B.2 Mesures PSNR et VQM de la séquence *Mobcal*

Séquence : Stockholm (720p50)				
Encodage Intra (Intra4x4-16x16)				
VQM-HRC : Television Model				
QP	dQP ou FP	Y-PSNR (dB)	Note VQM	Débit (Mbps)
26	-	37.55	0.0248	46.8
	dQP=1	36.48	0.0451	45.4
	FP=10	36.49	0.0440	45.0
	dQP=2	35.76	0.0628	43.3
	FP=9	34.97	0.0584	42.2
	FP=8	34.44	0.0616	41.0
	dQP=3	34.90	0.0831	40.3
	FP=7	34.06	0.0626	40.0
	FP=6	33.89	0.0665	39.5
	FP=5	32.29	0.0794	37.0
	FP=4	31.51	0.0914	35.8
	FP=3	30.16	0.1370	33.1
	dQP=4	33.65	0.1313	31.3
	27	-	36.91	0.0337
FP=10		36.09	0.0454	40.4
dQP=1		35.93	0.0577	40.2
FP=9		34.70	0.0625	38.0
dQP=2		35.04	0.0822	37.8
FP=8		34.22	0.0665	37.1
FP=7		33.88	0.0678	36.3
dQP=3		34.43	0.0938	36.1
FP=6		33.73	0.0756	36.0
FP=5		32.19	0.0839	33.7
FP=4		31.44	0.0993	32.7
FP=3		30.12	0.1393	30.2
dQP=4	33.18	0.1497	28.3	
28	-	36.34	0.0441	37.0
	FP=10	35.71	0.0573	36.2
	dQP=1	35.18	0.0797	35.5
	FP=9	34.42	0.0733	34.1
	dQP=2	34.51	0.0979	33.8
	FP=8	33.99	0.0777	33.4
	FP=7	33.69	0.0772	32.8
	dQP=3	34.04	0.1062	32.7
	FP=6	33.56	0.0825	32.5
	FP=5	32.08	0.0953	30.6
	FP=4	31.35	0.1078	29.7
	FP=3	30.06	0.1471	27.5
dQP=4	32.51	0.1836	24.4	

Tableau B.3 Mesures PSNR et VQM de la séquence *Stockholm*

Séquence : Shields (720p50)				
Encodage Intra (Intra4x4-16x16)				
VQM-HRC : Television Model				
QP	dQP ou FP	Y-PSNR (dB)	Note VQM	Débit (Mbps)
26	-	37.89	0.0240	57.9
	dQP=1	36.49	0.0422	56.2
	FP=10	35.69	0.0379	55.0
	FP=9	34.40	0.0531	52.9
	dQP=2	35.66	0.0546	53.7
	FP=8	33.48	0.0623	51.3
	dQP=3	34.76	0.0741	50.3
	FP=7	32.60	0.0741	49.7
	FP=6	32.29	0.0938	48.9
	FP=5	30.75	0.1513	45.8
	FP=4	29.55	0.1821	43.8
	FP=3	28.62	0.2184	41.7
	dQP=4	33.52	0.1219	41.5
	27	-	37.21	0.0310
dQP=1		35.96	0.0560	50.7
FP=10		35.36	0.0421	50.0
FP=9		34.16	0.0587	48.3
dQP=2		34.89	0.0743	47.8
FP=8		33.30	0.0665	46.9
dQP=3		34.23	0.0827	45.6
FP=7		32.46	0.0784	45.6
FP=6		32.17	0.1009	44.9
FP=5		30.67	0.1579	42.0
FP=4		29.49	0.1898	40.3
FP=3		28.59	0.2329	38.3
dQP=4	32.98	0.1425	37.9	
29	-	35.82	0.0685	41.8
	dQP=1	34.60	0.0945	40.5
	FP=10	34.50	0.0783	40.3
	FP=9	33.50	0.0967	39.1
	dQP=2	33.84	0.1019	38.8
	FP=8	32.77	0.0977	38.1
	FP=7	32.05	0.1093	37.1
	FP=6	31.80	0.1243	36.6
	dQP=3	33.0	0.1274	36.5
	FP=5	30.43	0.1780	34.3
	FP=4	29.33	0.2128	33.0
	FP=3	28.48	0.2502	31.4
dQP=4	31.67	0.1929	30.1	

Tableau B.4 Mesures PSNR et VQM de la séquence *Shields*

Annexe B.2 : Comparaison de la qualité visuelle du transcodage par sélection en position fréquentielle et requantification

Séquences 1080p

Sept séquences ont été utilisées en 1080p : *Blue Sky*, *Pedestrian Area*, *Riverbed*, *Rush Hour*, *Station2*, *Sunflower* et *tractor*.

Sur l'ensemble des séquences vidéo 1080p, nous constatons une réduction de débit maximale de 40% avec la séquence *Pedestrian area* pour un paramètre de quantification QP=18 et un paramètre de transcodage FP=1 ou FL=1. L'ensemble des résultats est présenté sous forme de courbes ci-dessous. La séquence dont la réduction de débit est la moins significative est la séquence *Sunflower*. La réduction de débit est de l'ordre de 15%. Pour l'ensemble des séquences, la réduction de débit décroît en augmentant le paramètre de quantification. Pour une réduction de débit supérieure à 10%, nous remarquons la présence de deux fois plus de débits intermédiaires pour le transcodage par sélection en position fréquentielle que pour le transcodage en sélection en niveau fréquentiel. Ce premier point est un avantage certain pour le transcodage par sélection en position fréquentielle, favorisant une adaptation fine du débit du codage source à celui du canal de transmission.

Comme abordé dans la section précédente, un paramètre de transcodage par position fréquentielle FP=1 ou FP=2 sera peu utilisé à cause de la médiocrité de la qualité vidéo du flux résultant. Ainsi, nous préférons utiliser un paramètre FP supérieur ou égal à 3. Dans le cas du transcodage par niveau fréquentiel, le débit associé au paramètre FL=2 est généralement supérieur ou égal à celui de FP=3. Nous constatons pour certaines séquences (*Pedestrian area*, *Tractor*) un effet de bruit de type *mosquito* pour un paramètre FL=2. A débit équivalent, le transcodage par sélection en position fréquentielle n'introduit pas d'effet de ce type. La Figure B.1 illustre une partie de la séquence *Pedestrian area* encodée avec un paramètre de quantification QP=18 (a), puis transcodée par sélection en position fréquentielle (b) ou transcodée par sélection en niveau fréquentiel (c).

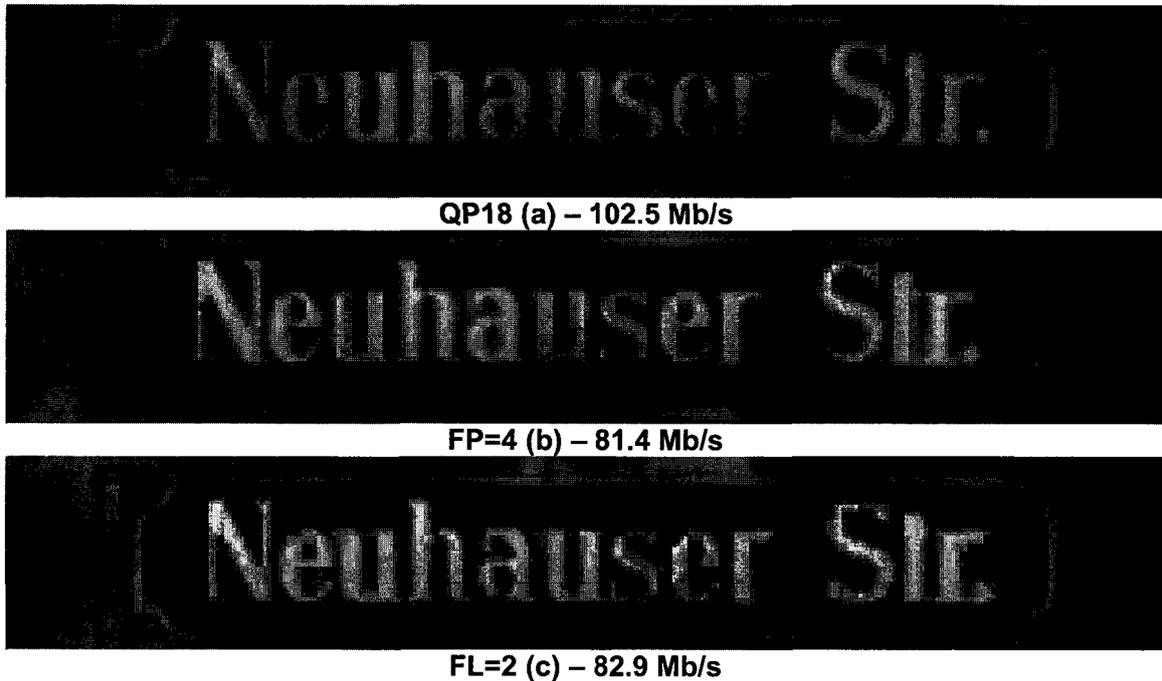


Figure B.1 Illustration du résultat de transcodage d'une partie de la séquence *Pedestrian area*

L'effet mosquito est présent en (c). Il se traduit par une alternance de disparition et d'apparition d'une ou plusieurs fréquences. Bien qu'il apparaisse localement dans l'image, cet effet de scintillement, localisé le long des contours des objets, attire l'attention de l'observateur et devient fortement gênant.

Sur les sept séquences 1080p, les mesures PSNR sont largement en faveur de la sélection par position fréquentielle. À débit similaire, nous constatons un écart allant jusqu'à 2 dB entre les deux types de transcodage. Par exemple, la séquence *Blue sky* encodée avec un paramètre de quantification égal à 18 (Figure B.2) avec FP=5 (142.9Mb/s – PSNR-Y= 39.13dB) et FL=3 (143.9Mb/s – PSNR-Y=37dB).

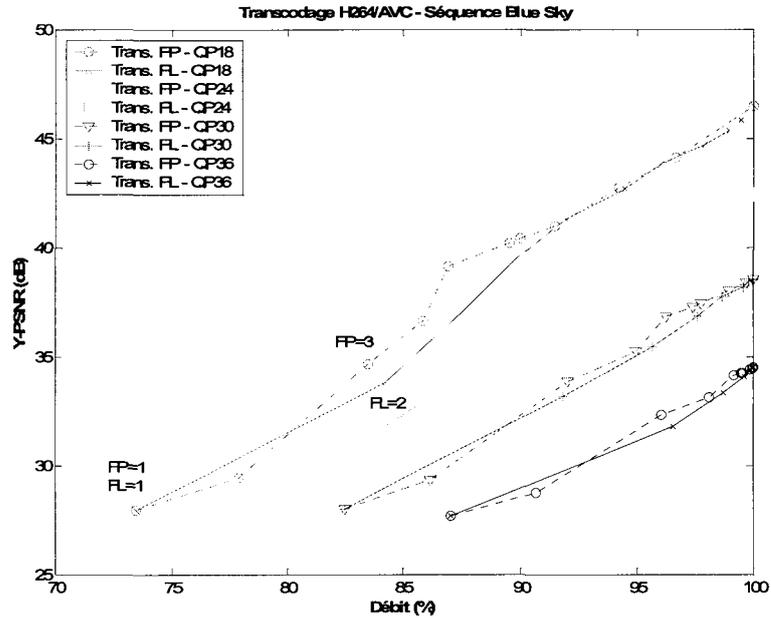


Figure B.2 Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Blue Sky*.

La Figure B.3 consacrée à la séquence *Pedestrian area* montre également un meilleur PSNR pour le transcodage par sélection en position fréquentielle. Rappelons également l'insuffisance des points du transcodage par sélection en niveau fréquentiel pour une réduction de débit supérieure à 10%, en particulier de FL=1 à FL=4.

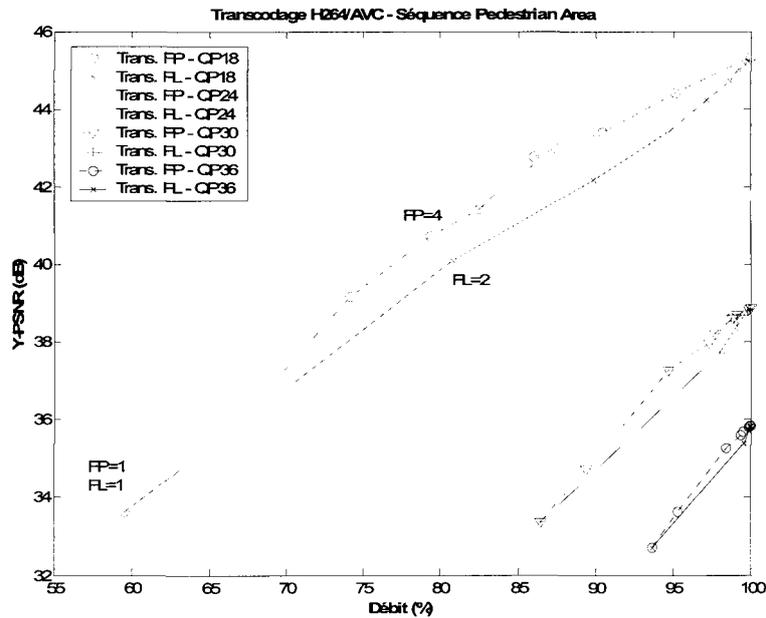


Figure B.3 Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Pedestrian Area*.

Les résultats VQM (cf 2.4.4) sont majoritairement en faveur du transcodage par sélection en position fréquentielle. Ces résultats sont illustrés dans la Figure B.4 et la Figure B.5 pour les séquences *Blue sky* et *Pedestrian area* respectivement. Nous remarquerons une note VQM légèrement plus basse, donc meilleure pour certains débits intermédiaires du transcodage par sélection en niveau fréquentiel. Ceci s'observe pour une faible réduction de débit, inférieure à 10%, pour moins du tiers des débits intermédiaires du transcodage par sélection en niveau fréquentiel.

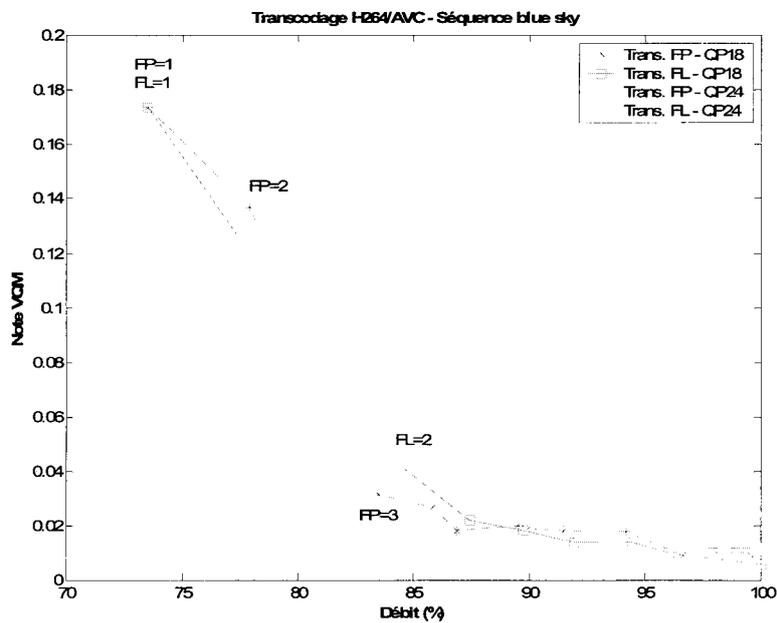


Figure B.4 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Blue sky*.

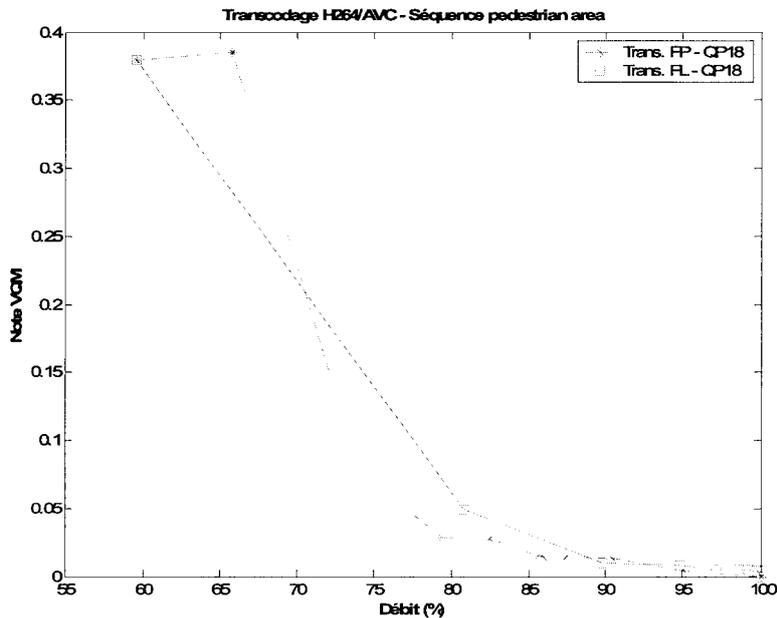


Figure B.5 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Pedestrian area*.

Séquences 4CIF et 4SIF

Dix séquences ont été utilisées en définition 4CIF (704x576) et 4SIF (640x480) afin d'être proche de la définition standard (SD) : *City, Crew, Harbour, Soccer, Ice, Flowergarden, Football, Galleon, Intros et Mobile Calendar*.

Sur l'ensemble des séquences vidéo 4CIF et 4SIF, nous constatons une réduction de débit maximale de 55% avec la séquence *City* pour un paramètre de quantification QP=18 et un paramètre de transcodage FP=1 ou FL=1. Pour l'ensemble des séquences, la réduction de débit décroît en augmentant le paramètre de quantification. Pour une réduction de débit supérieure à 10%, nous remarquons la présence d'au moins 30% de plus de débits intermédiaires pour le transcodage par sélection en position fréquentielle que pour le transcodage en sélection en niveau fréquentiel. Cet écart a tendance à augmenter avec les séquences *City, Crew, Soccer, Ice* et *Intros*; mais également avec des paramètres de quantification plus élevés (QP=30 et QP=36). Ainsi, dans cette plage de réduction de débit supérieure à 10%, pour la séquence *Soccer* encodée avec un QP=24, huit débits intermédiaires sont disponibles par la sélection en position fréquentielle contre trois pour la sélection en niveau fréquentiel.

Nous constatons également un effet de bruit de type *mosquito* pour certaines séquences (*Flower garden, Galleon* et *Mobile calendar*) pour un paramètre FL=2 ou FL=3

voire FL=4. A débit équivalent, le transcodage par sélection en position fréquentielle n'introduit pas d'effet de ce type.

Les mesures PSNR sont en faveur de la sélection par position fréquentielle dans environ 55% des cas. Pour les quatre paramètres de quantification utilisés (QP=18, 24, 30 ou 36), le PSNR après transcodage des séquences *Crew*, *Ice* (Figure B.6) et *Intros* est toujours en faveur de la sélection en position fréquentielle. Pour les paramètres de quantification QP=18, QP=24 et QP=30, le PSNR après transcodage des séquences *Mobile calendar* et *Harbour* (Figure B.7) est en faveur de la sélection en niveau fréquentiel. Les résultats PSNR des séquences *City*, *Flower garden*, *Football* (Figure B.8), *Galleon* et *Soccer* sont meilleurs pour :

- le transcodage par sélection en niveau fréquentiel pour les paramètres de quantification QP=18 et QP=24
- le transcodage par sélection en position fréquentielle pour les paramètres de quantification QP=30 et QP=36

À débit similaire, nous constatons un écart jusqu'à 2 dB entre les deux types de transcodage. Par exemple, nous remarquons la séquence *Ice* encodée avec un paramètre de quantification égal à 18, sur la Figure B.6, avec FP=8 (8.23Mb/s – PSNR-Y= 41.34dB) et FL=3 (8.25Mb/s – PSNR-Y=39.15dB).

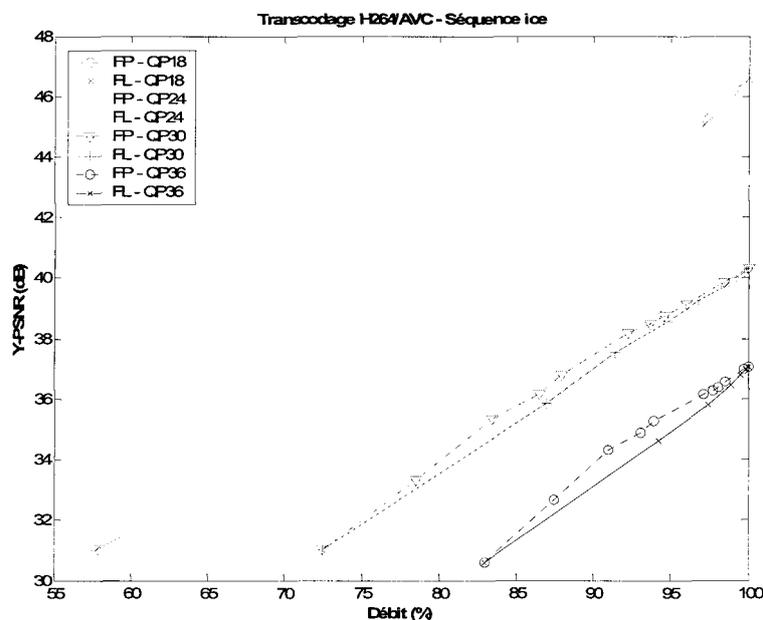


Figure B.6 Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence Ice.

Ou encore, la séquence *Harbour* encodée avec un paramètre de quantification égal à 18, sur la Figure B.7, avec FP=4 (19.3Mb/s – PSNR-Y= 30.0dB) et FL=2 (19.12Mb/s – PSNR-Y=31.51dB).

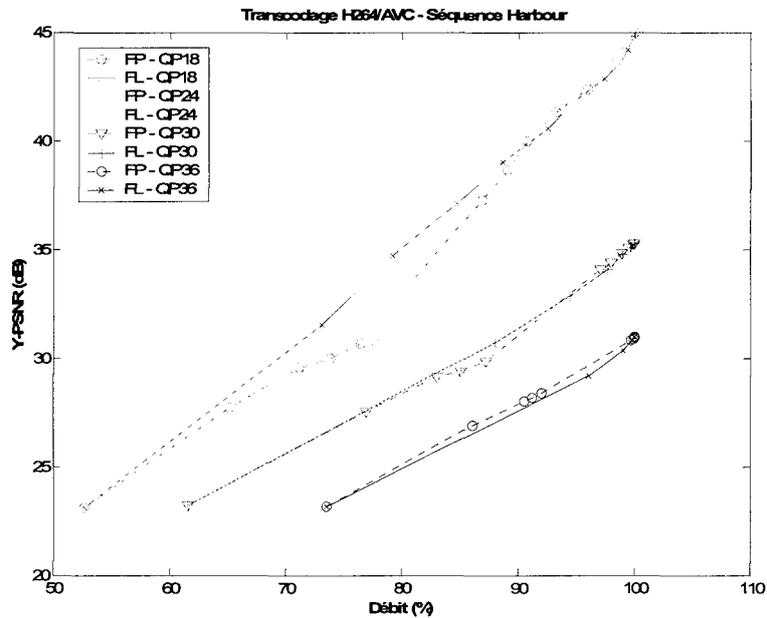


Figure B.7 Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Harbour*.

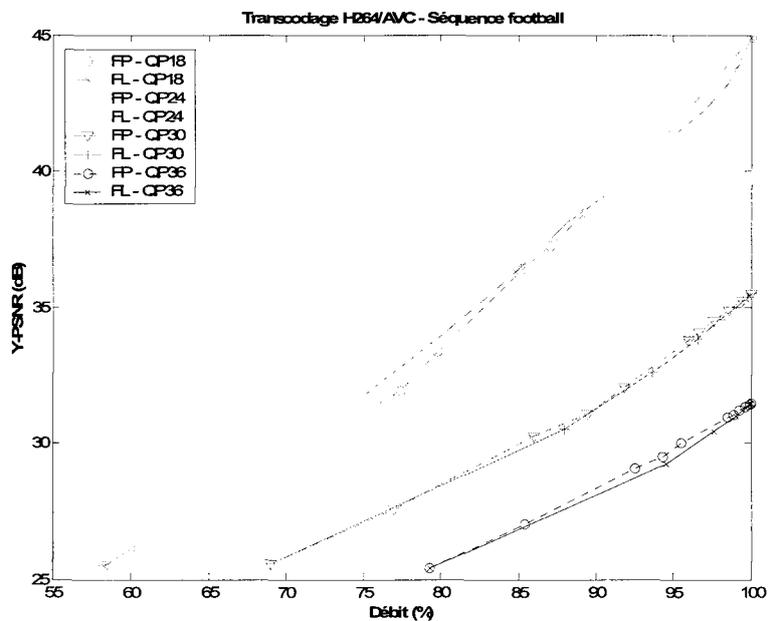


Figure B.8 Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Football*.

Les résultats VQM sont illustrés dans les Figures B.9, B.10 et B.11, pour les séquences *Ice*, *Harbour* et *Football* respectivement.

Nous constatons que similairement à l'analyse PSNR, les notes VQM pour la séquence *Ice* sont en faveur de la sélection en position fréquentielle avec les paramètres de quantification QP=18 et QP=36.

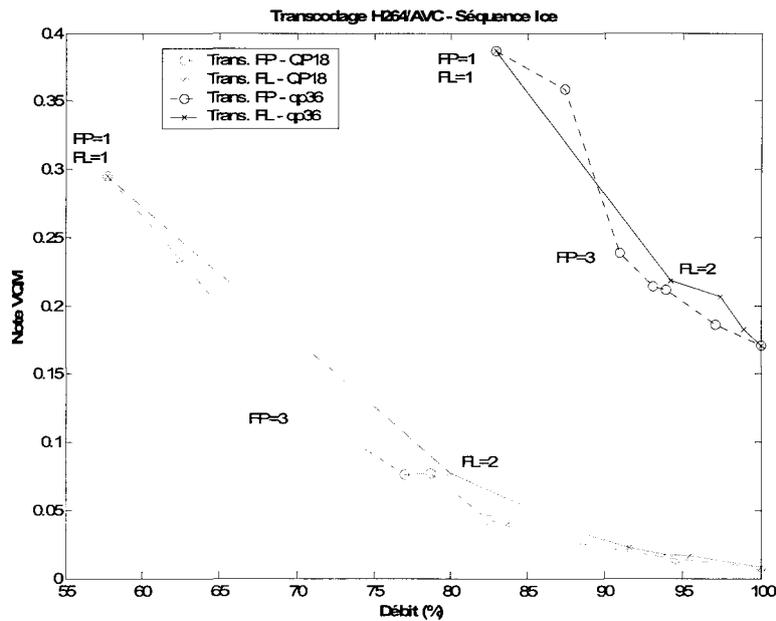


Figure B.9 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Ice*.

Dans la Figure B.10, dédiée à la séquence *Harbour*, les résultats VQM pour le paramètre de quantification QP=24 ne suivent pas la tendance du PSNR. Alors que les mesures PSNR étaient clairement en faveur de la sélection en niveau fréquentiel, les mesures VQM sont partagées en fonction du débit. Ainsi, pour une réduction de débit supérieure à 17% les résultats VQM sont favorables à la sélection en position fréquentielle, autorisant quatre débits intermédiaires (FP=2, FP=3, FP=4 et FP=5). Dans le cas d'une réduction de débit inférieure, les résultats VQM sont légèrement meilleurs pour la sélection en niveau fréquentiel.

Dans la Figure B.11, dédiée à la séquence *Football*, les résultats VQM pour le paramètre de quantification QP=18 ne suivent pas la tendance du PSNR. Alors que les mesures PSNR étaient clairement en faveur de la sélection en niveau fréquentiel, les mesures VQM sont partagées en fonction du débit. Ainsi, pour une réduction de débit supérieure à 20% les résultats VQM sont favorables à la sélection en position fréquentielle, autorisant trois débits

intermédiaires (FP=2, FP=3 et FP=4). Dans le cas d'une réduction de débit inférieure, les résultats VQM sont légèrement meilleurs pour la sélection en niveau fréquentiel.

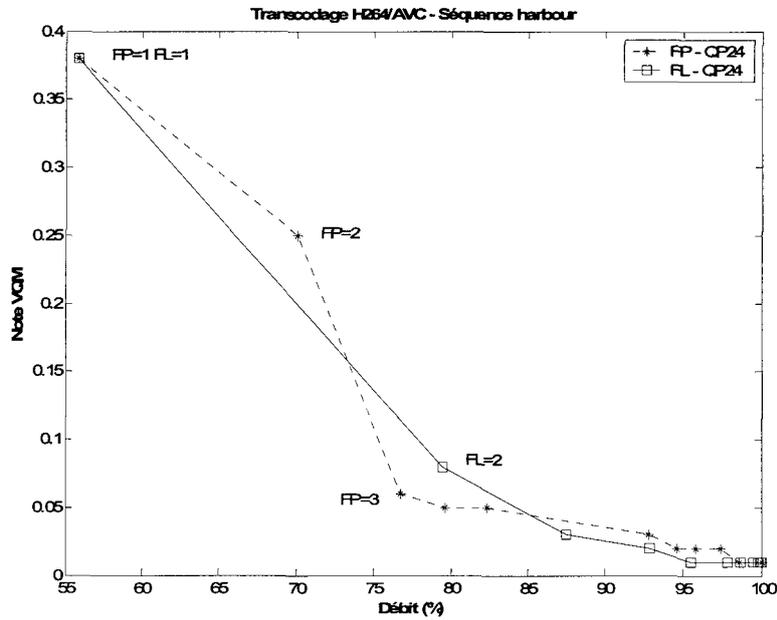


Figure B.10 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Harbour*.

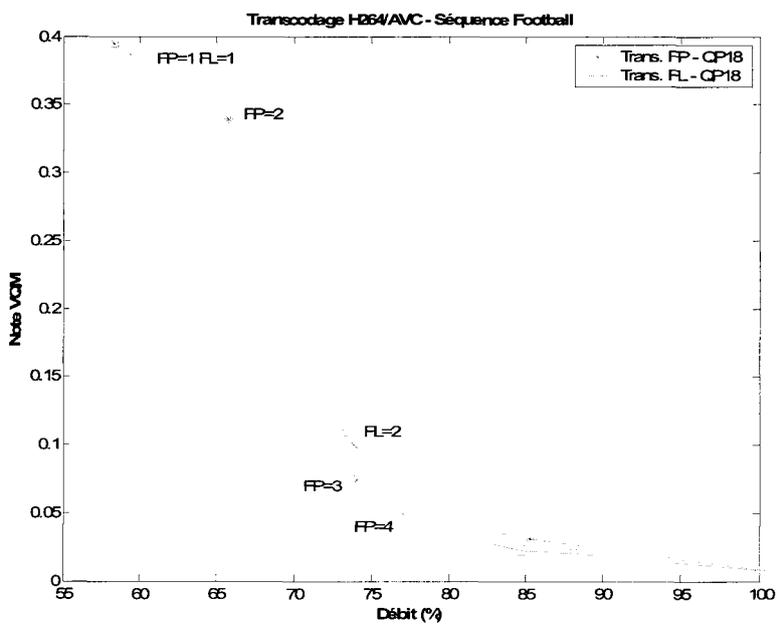


Figure B.11 Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence *Football*.



Figure B.12 Illustration des séquences 720p utilisées lors des expérimentations

Annexe C : Le standard Homeplug AV

La HomePlug Powerline Alliance (HPPA) est un consortium industriel créé en avril 2000. Sa motivation est d'établir une standardisation afin d'assurer une compatibilité entre les différentes technologies de communication par courant porteur en ligne et de promouvoir ce type de communication sur le marché industriel et grand public. Parmi les membres fondateurs de la HPPA, nous retrouvons : Intel, Intellon, Motorola, Panasonic et Texas Instruments [FER10]. Les différents standards issus de ce consortium font partie de la famille HomePlug.

Le premier standard issu de la HomePlug Powerline Alliance s'appelle Home Plug 1.0 [HPP01]. Il permet un débit de la couche physique (PHY) de 14 Mb/s (couche 1 du modèle OSI ou interconnexion des systèmes ouverts - Tableau C.1) utilisant une modulation DMT/OFDM (Orthogonal Frequency Division Multiplexing) de 84 sous-porteuses dans la bande de fréquences de 4,5 à 21 MHz. Afin d'éviter les interférences dans cette bande de fréquences avec les applications existantes (par exemple les transmissions radio-amateur) et de respecter les réglementations en vigueur, un masque de fréquences est utilisé afin de pouvoir spécifier un niveau d'amplitude à chaque sous-porteuse. Les sous-porteuses sont modulées soit par une modulation BPSK ou une modulation QPSK en fonction de la qualité du sous-canal. Le rapport signal sur bruit de chaque sous-canal est estimé afin d'établir une cartographie des sous-porteuses utilisables ou non (tone map) pour transmettre les données. Cette cartographie servira à définir le type de modulation et de code correcteur d'erreurs (FEC - Forward Error Correction) à utiliser (ici Viterbi et Reed Solomon) [PRO95]. Notons que l'industriel Intellon a fait évoluer ce standard en Home Plug 1.0.1 en utilisant d'autres types de modulations. Le débit de transmission des données peut atteindre jusque 80 Mb/s.

Numéro de couche	Désignation
7	Application
6	Présentation
5	Session
4	Transport
3	Réseau
2	Liaison
1	Physique

Tableau C.1 Les 7 couches du modèle OSI (Open Systems Interconnection)

La couche Liaison de données (couche 2 du modèle OSI) est composée de deux sous-couches : la couche de contrôle de la liaison logique (notée LLC) et la couche de contrôle

d'accès au support appelée Media Access Control (MAC). Sur cette dernière sous-couche, la méthode d'accès utilisée est la Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA).

Le standard HomePlug AV [**HPP05**] a été finalisé en 2005 et permet un débit de la couche physique de l'ordre de 200 Mb/s. Il utilise la bande de fréquences de 2 à 30 MHz et une couche de contrôle d'accès fondée sur du CSMA/CA et du TDMA.

La couche physique utilise un algorithme de bit loading allouant jusqu'à 10 bits par sous-porteuse. Le temps d'un symbole DMT/OFDM est relativement long afin de limiter la durée de l'intervalle de garde.

Le standard spécifie l'utilisation de codes correcteurs d'erreurs afin de rendre le flux binaire moins sensible aux bruits et perturbations subis lors de la transmission. Ces codes introduisent une redondance dans le flux binaire. Dans le standard HomePlug AV, les turbo codes [**BER93**] sont utilisés comme code correcteur d'erreurs. Ils permettent de s'approcher davantage de la limite théorique de Shannon.

HomePlug AV utilise 1155 sous-porteuses allant de 1,80 MHz à 30,0 MHz, espacées d'environ 24,4 kHz. Elles peuvent être modulées par des modulations BPSK, QPSK, 8-QAM, 16-QAM, 64-QAM, 256-QAM ou 1024-QAM en fonction du rapport signal sur bruit mesuré sur chaque sous-canal.

La Figure C.1 présente le diagramme de blocs des modules émetteur et récepteur de HomePlug AV.

La partie supérieure du diagramme représente l'émetteur (TX). La couche physique récupère ses entrées de la sous-couche MAC (couche 2 – liaison des données). Trois chaînes de traitement distinctes y sont représentées afin de gérer les codes correcteurs d'erreurs (FEC) différents utilisés par les informations de contrôle de HomePlug 1.0.1, les informations de contrôle de HomePlug AV, et la charge des paquets HomePlug AV. Les informations de contrôle contiennent par exemple la longueur de la trame et les adresses source et destination. La charge utile contient les données réelles à transmettre.

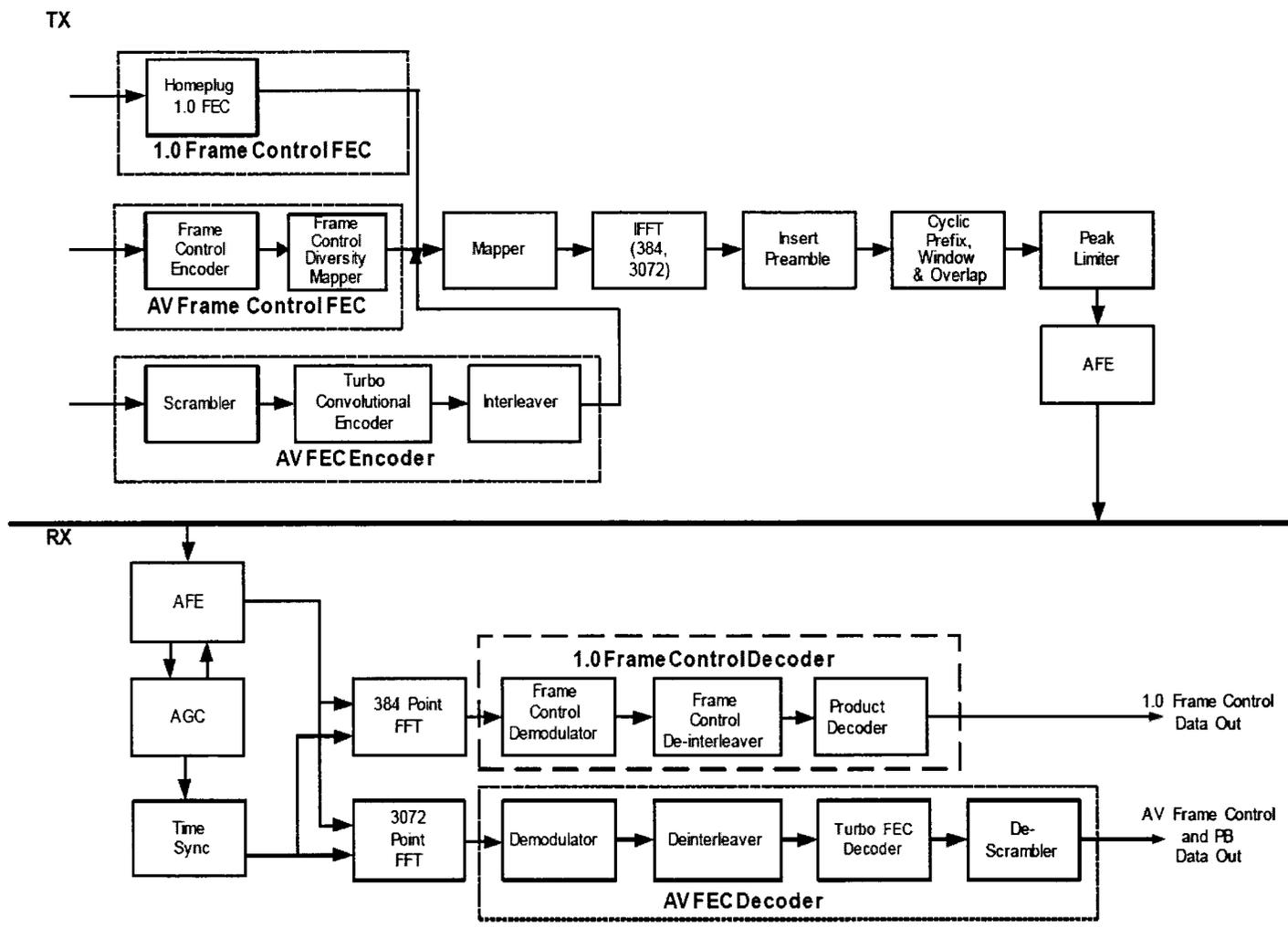


Figure C.1 Diagramme de blocs de l'émetteur et du récepteur de HomePlug AV [HPP05]

Les informations de contrôle HomePlug AV sont traitées par le bloc d'encodage. Le flux de données HomePlug AV subit un embrouillage, un encodage par turbo codes et un entrelacement des données. Les informations de contrôle HomePlug 1.0.1 subissent un encodage FEC de type Reed Solomon. Les sorties des trois encodeurs FEC sont raccordées à un module commun de modulation OFDM/DMT. Il est composé d'un mapper (servant au bit loading), du calcul de la Transformée de Fourier Rapide Inverse (IFFT) sur 384 ou 3072 points (respectivement pour HomePlug 1.01 et HomePlug AV) et de l'insertion du préfixe cyclique. Un convertisseur numérique-analogique (Analog Front End - AFE) est ensuite utilisé pour transmettre le signal sur le réseau électrique indoor.

La partie inférieure du diagramme représente le récepteur (RX). Un convertisseur analogique-numérique est utilisé avant d'appliquer la transformée de Fourier sur le signal et le décodage correcteur d'erreurs correspondant.

Liste des figures

Figure 1.1	Représentation en couches d'un flux H.264	9
Figure 1.2	Représentation du flux binaire d'une NALU	10
Figure 1.3	Découpage d'une image en slices	11
Figure 1.4	Exemple de numérotation des macroblocs dans une image composée de trois slices	11
Figure 1.5	Ordre de balayage des blocs luminance 8x8	12
Figure 1.6	Ordre de balayage des blocs luminance 4x4	12
Figure 1.7	Schéma bloc d'un encodeur H.264/AVC	13
Figure 1.8	Exemple de représentation d'un GOP de 12 images	14
Figure 1.9	Désignation des échantillons d'un bloc prédit 4x4 et des échantillons voisins	16
Figure 1.10	Les 9 modes de prédiction du codage INTRA_4x4	17
Figure 1.11	Partitionnement d'un macrobloc pour la prédiction/compensation de mouvement (a) - Partitionnement d'un bloc 8x8 (b)	18
Figure 1.12	Interpolation des échantillons de luminance jusqu'au quart de pixel	20
Figure 1.13	Exemple de dépendance entre slices sur un GOP de 16 images	22
Figure 1.14	Transformée entière d'un bloc de 4x4 échantillons	23
Figure 1.15	Étapes d'encodage d'un macrobloc de type Intra_16x16	24
Figure 1.16	Balayage en zig-zag d'un bloc de 4x4 coefficients résiduels fréquentiels	25
Figure 1.17	Exemple de codage d'un bloc 4x4 au format Run/Level	26
Figure 1.18	Prédiction du nombre de coefficients non nuls	27
Figure 1.19	PSNR et réduction de débit apportée par le MPEG-2 et le H.264/AVC [WIE03b]	30
Figure 1.20	Exemple d'un GOP de 16 images avec 4 niveaux d'échelonnabilité temporelle	33
Figure 1.21	Schéma de l'encodeur H.264 SVC en échelonnabilité spatiale	35
Figure 1.22	Schéma d'un encodeur H.264 SVC en échelonnabilité CGS	37
Figure 1.23	Exemple de structure de codage par défaut d'une séquence codée en MGS39	
Figure 1.24	Exemple de structure de codage d'une séquence codée en MGS utilisant des images clefs	40
Figure 1.25	Comparaison des échelonnabilités CGS et MGS pour la séquence <i>City</i> (CIF 15 Hz) [SCH07]	41
Figure 1.26	Comparaison des échelonnabilités CGS et MGS pour la séquence <i>Crew</i> (CIF 15 Hz) [SCH07]	41
Figure 1.27	Efficacité de l'échelonnabilité spatiale - séquence <i>Crew</i> [HUA07]	42
Figure 2.1	Schéma général d'un transcodeur vidéo	45
Figure 2.2	Principe d'un transcodeur en boucle ouverte	47
Figure 2.3	Image de référence du flux à l'entrée du transcodeur (a) - Illustration de l'effet de drift dû au transcodage en boucle ouverte d'une image intra en H.264/AVC (b)	48
Figure 2.4	Exemple de coefficients résiduels fréquentiels d'un bloc 4x4 dans l'ordre du zig-zag après quantification inverse et requantification	50
Figure 2.5	Exemple de traitement des coefficients résiduels par sélection en position fréquentielle et en niveau fréquentiel	51
Figure 2.6	Diagramme simplifié du transcodeur H.264/AVC en sélection fréquentielle (SF) pour les macroblocs intra et inter	52
Figure 2.7	Image intra résultant de notre architecture de transcodage en boucle fermée (FS)	54

Liste des figures

Figure 2.8	Dépendance entre les images du GOP utilisé	59
Figure 2.9	Évolution du PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence <i>Mobcal</i>	61
Figure 2.10	Évolution du PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence <i>Stockholm</i>	62
Figure 2.11	Évolution du PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence <i>Shields</i>	62
Figure 2.12	Évolution du PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence <i>Parkrun</i>	63
Figure 2.13	Exemple d'images issues du transcodage par sélection en position fréquentielle (a) et par requantification (b), avec respectivement les paramètres FP=3 et dQP=4	64
Figure 2.14	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence <i>Mobcal</i>	65
Figure 2.15	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence <i>Shields</i>	66
Figure 2.16	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence <i>Stockholm</i>	67
Figure 2.17	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et requantification – Séquence <i>Parkrun</i>	68
Figure 2.18	Résultats des préférences des observateurs entre le transcodage par sélection en position fréquentielle et la requantification	71
Figure 2.19	Illustration du résultat de transcodage d'une partie de la séquence <i>Oldtown</i> .	73
Figure 2.20	Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Ducks</i>	74
Figure 2.21	Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Oldtown</i>	75
Figure 2.22	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>OldTown</i>	75
Figure 2.23	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Ducks</i>	76
Figure 2.24	Courbe moyenne illustrant l'évolution du débit (B) en fonction des paramètres de transcodage FP _{intra} et FP _{inter}	78
Figure 2.25	Évolution du débit du modèle proposé, B(FP _{intra} , 16) et B(16, FP _{inter}), dépendant du paramètre FP _{intra} ou FP _{inter}	79
Figure 3.1	Illustration de l'échelonnabilité spatiale utilisant la prédiction intra-couche (flèches pleines) et la prédiction inter-couche (flèches en pointillés).....	84
Figure 3.2	Illustration d'une image du flux à l'entrée du transcodeur (a) et de l'effet de drift dû au transcodage en boucle ouverte de macroblochs utilisant la prédiction intra-couche intra-image sur la couche d'amélioration haute (b)	85
Figure 3.3	Types de macroblochs sur l'ensemble des slices (couche d'amélioration).....	86
Figure 3.4	Types de macroblochs dans les slices de type Intra (couche d'amélioration) ..	87
Figure 3.5	Types de macroblochs dans les slices de type P (couche d'amélioration).....	87
Figure 3.6	Types de macroblochs dans les slices de type B (couche d'amélioration).....	88
Figure 3.7	Évolution du PSNR en fonction du débit pour la séquence <i>City</i> encodée avec ou sans macroblochs utilisant la prédiction intra-couche intra-image.....	89
Figure 3.8	Diagramme simplifié du transcodeur H.264 SVC en sélection par position fréquentielle pour les macroblochs intra et inter	92
Figure 3.9	Illustration de l'image numéro 0 (image intra) de la séquence <i>Harbour</i> (QP=18)	96

Liste des figures

Figure 3.10	Évolution du PSNR (Y) en fonction du débit - Séquence <i>City</i>	97
Figure 3.11	Évolution du PSNR (Y) en fonction du débit – Séquence <i>Crew</i>	97
Figure 3.12	Évolution du PSNR (Y) en fonction du débit – Séquence <i>Harbour</i>	98
Figure 3.13	Évolution du PSNR (Y) en fonction du débit – Séquence <i>Soccer</i>	98
Figure 3.14	Évolution du PSNR de la composante de chrominance Cb (U) en fonction du débit — Séquence <i>City</i>	99
Figure 3.15	Évolution du PSNR de la composante de chrominance Cr (V) en fonction du débit - Transcodage sur flux SVC –Séquence <i>Crew</i>	100
Figure 3.16	Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des échantillons de luminance – QP=18.....	106
Figure 3.17	Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des échantillons de luminance – QP=24.....	107
Figure 3.18	Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des macroblocs inter-image – Séquence <i>City</i>	108
Figure 3.19	Illustration de l'image numéro 1 (image bidirectionnelle) de la séquence <i>City</i> (QP=18)	109
Figure 3.20	Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des macroblocs inter-image – Séquence <i>Crew</i>	111
Figure 3.21	Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des macroblocs inter-image – Séquence <i>Harbour</i>	111
Figure 3.22	Évolution du PSNR (Y) en fonction du débit - Transcodage sur flux SVC des macroblocs inter-image – Séquence <i>Soccer</i>	112
Figure 4.1	Illustration du principe d'optimisation conjointe d'une transmission sur un réseau filaire	115
Figure 4.2	Illustration d'une ligne ADSL reliant le réseau domestique au central téléphonique	117
Figure 4.3	Utilisation de la bande passante téléphonique entre le téléphone traditionnel commuté et l'ADSL.....	118
Figure 4.4	Illustration de la chaîne DMT complète [GOU05].....	120
Figure 4.5	Représentation de la zone d'éligibilité initiale pour le service de télévision haute définition et le service de télévision en définition standard.....	124
Figure 4.6	Extension de la zone d'éligibilité du service de télévision haute définition ...	125
Figure 4.7	Illustration du type de GOP utilisé dans la couche d'amélioration.....	126
Figure 4.8	Lignes test proposées par l'ETSI [ITU01].....	129
Figure 4.9	Représentation de la fonction de transfert de la ligne « boucle 1 » en fonction de la fréquence pour différentes longueurs.....	130
Figure 4.10	Évolution du PSNR de la luminance en fonction de la longueur de la boucle 1 pour la séquence <i>Mobcal</i>	136
Figure 4.11	Évolution du PSNR de la luminance en fonction de la longueur de la boucle 1 pour la séquence <i>Shields</i>	137
Figure 4.12	Illustration sur une zone de la séquence <i>Shields</i>	139
Figure 4.13	Évolution du PSNR de la luminance en fonction de la longueur de la boucle 3 pour les séquences <i>Mobcal</i> et <i>Parkrun</i>	141
Figure 4.14	Évolution du PSNR de la luminance en fonction de la longueur de la boucle 3 pour les séquences <i>Shields</i> et <i>Stockholm</i>	141
Figure 4.15	Évolution du PSNR de la luminance en fonction de la longueur de la boucle 8 pour les séquences <i>Mobcal</i> et <i>Parkrun</i>	143
Figure 4.16	Évolution du PSNR de la luminance en fonction de la longueur de la boucle 8 pour les séquences <i>Shields</i> et <i>Stockholm</i>	144

Liste des figures

Figure 4.17	Représentation de la fonction de transfert du canal CPL en fonction de la fréquence avant (en rouge) et après (en bleu) perturbation.....	146
Figure 4.18	Réponse temporelle du canal à une sinusoïde transmise lors de la mise sous-tension d'un appareil électrique.....	147
Figure 4.19	Densité spectrale de puissance d'un bruit réel sur le canal CPL.....	150
Figure 4.20	Évolution du PSNR de la luminance en fonction du numéro d'image avec ou sans adaptation du débit vidéo.....	152
Figure 4.21	Illustration d'une même image décodée après transmission dans le canal (a) avant changement d'état (Débit = 8 Mb/s); (b) après changement d'état mais sans adaptation au canal; (c) après changement d'état et adaptation.....	153
Figure A.1	Comparaison des performances de MPEG-2 et H.264/AVC pour la séquence <i>Crew</i> en 4CIF	163
Figure A.2	Comparaison des performances de MPEG-2 et H.264/AVC pour la séquence <i>City</i> en CIF	163
Figure B.1	Illustration du résultat de transcodage d'une partie de la séquence <i>Pedestrian area</i>	169
Figure B.2	Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Blue Sky</i>	170
Figure B.3	Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Pedestrian Area</i> ...	170
Figure B.4	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Blue sky</i>	171
Figure B.5	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Pedestrian area</i>	172
Figure B.6	Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Ice</i>	173
Figure B.7	Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Harbour</i>	174
Figure B.8	Évolution des mesures PSNR en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Football</i>	174
Figure B.9	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Ice</i>	175
Figure B.10	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Harbour</i>	176
Figure B.11	Évolution des mesures VQM en fonction du débit - Transcodage par sélection en position fréquentielle et en niveau fréquentiel – Séquence <i>Football</i>	176
Figure B.12	Illustration des séquences 720p utilisées lors des expérimentations.....	177
Figure C.1	Diagramme de blocs de l'émetteur et du récepteur de HomePlug AV [HPP05]	180

Liste des tableaux

Tableau 1.1	Choix de la table utilisée pour coder le coeff_token.....	27
Tableau 2.1	Performances du transcodeur par sélection en position fréquentielle	55
Tableau 2.2	Impact du transcodage des images intra sur la qualité vidéo.....	59
Tableau 2.3	Impact du transcodage des images inter sur la qualité vidéo.....	60
Tableau 2.4	Mesures RCA pour la séquence <i>Parkrun</i> encodée avec QP=38	69
Tableau 2.5	Mesure des défauts perçus par l'analyse RCA pour les séquences Stockholm et Ducks	70
Tableau 3.1	Estimation de l'utilisation des modes de prédiction.....	90
Tableau 3.2	Encodage de la séquence City sans macrobloc utilisant la prédiction intra-couche intra-image (dans la C.A.)	90
Tableau 3.3	Encodage de la séquence Crew sans macrobloc utilisant la prédiction intra-couche intra-image (dans la C.A.)	91
Tableau 3.4	Encodage de la séquence Harbour sans macrobloc utilisant la prédiction intra-couche intra-image (dans la C.A.)	91
Tableau 3.5	Encodage de la séquence Soccer sans macrobloc utilisant la prédiction intra-couche intra-image (dans la C.A.)	91
Tableau 3.6	Évolution du PSNR en fonction du débit pour la séquence City.....	101
Tableau 3.7	Évolution du PSNR en fonction du débit pour la séquence Crew.....	102
Tableau 3.8	Évolution du PSNR en fonction du débit pour la séquence Harbour	103
Tableau 3.9	Évolution du PSNR en fonction du débit pour la séquence Soccer.....	104
Tableau 4.1	Paramètres utilisés par la modulation DMT de l'ADSL	121
Tableau 4.2	Informations sur le flux vidéo de différentes chaînes issues du bouquet de télévision du FAI Free	123
Tableau 4.3	Taille des en-têtes et paquets de l'encapsulation vidéo utilisée	128
Tableau 4.4	Distance initiale d'éligibilité des modèles de lignes pour des débits de 12 Mb/s, 8 Mb/s et 6 Mb/s	131
Tableau 4.5	Transcodage H.264 SVC de la séquence <i>Mobcal</i>	133
Tableau 4.6	Transcodage H.264 SVC de la séquence <i>Parkrun</i>	134
Tableau 4.7	Transcodage H.264 SVC de la séquence <i>Shields</i>	134
Tableau 4.8	Transcodage H.264 SVC de la séquence <i>Stockholm</i>	135
Tableau 4.9	Distances d'éligibilité (HD, HD Q-, SD) pour les différentes séquences transmises sur la Boucle 1	138
Tableau 4.10	Distances d'éligibilité (HD, HD Q-, SD) pour les différentes séquences transmises sur la Boucle 3	140
Tableau 4.11	Distances d'éligibilité (HD, HD Q-, SD) pour les différentes séquences transmises sur la Boucle 8	143
Tableau A.1	Estimation du surcoût de débit dû au forçage en mode DC en H.264/AVC Intra-image	164
Tableau B.1	Mesures PSNR et VQM de la séquence <i>Parkrun</i>	166
Tableau B.2	Mesures PSNR et VQM de la séquence <i>Mobcal</i>	166
Tableau B.3	Mesures PSNR et VQM de la séquence <i>Stockholm</i>	167

Liste des tableaux

Tableau B.4	Mesures PSNR et VQM de la séquence <i>Shields</i>	167
Tableau C.1	Les 7 couches du modèle OSI (Open Systems Interconnection)	178

Glossaire

1080p : format vidéo 16/9 avec une largeur d'image de 1920 pixels et une hauteur de 1080 pixels.

4CIF : format vidéo avec une largeur d'image de 704 pixels et une hauteur de 576 pixels.

720p : format vidéo 16/9 avec une largeur d'image de 1280 pixels et une hauteur de 720 pixels.

AC transform coefficient : Coefficient résiduel dont un indice fréquentiel est différent de zéro (dans au moins une des deux dimensions).

ADSL : Asymmetric Digital Subscriber Line.

ANR: Agence Nationale pour la Recherche.

ATM : Asynchronous Transfer Mode.

AVC : Advanced Video Coding.

Base_mode_flag : drapeau indiquant le nouveau type de macrobloc introduit par H.264 SVC pour les couches d'amélioration en échelonnabilité spatiale.

Buffer : Zone mémoire réservée pour stocker des données.

CAVLC : Context-based Adaptive Variable Length Coding.

CAVLD : Context-based Adaptive Variable Length Decoding. Opération inverse ou de décodage du CAVLC.

CABAC : Context-based Adaptive Binary Arithmetic Coding.

CGS : Coarse Grain Scalability.

CIF : Format vidéo avec une largeur d'image de 352 pixels et une hauteur de 288 pixels.

Coded_block_pattern : ensemble de bits désignant les blocs 8x8 de luminance et de chrominance d'un macrobloc qui contiennent des coefficients résiduels fréquents non nuls.

Coeff_token : mot code à longueur variable (utilisé dans le CAVLC) contenant le nombre de coefficients non nuls (*TotalCoeffs*, variant de 0 à 16) et de coefficients dont la valeur absolue est égale à 1 (*trailing ones*).

Constrained_intra_pred_flag : drapeau présent dans les NALU de type PPS afin de préciser si la prédiction intra-image autorise (drapeau égal à 0) ou non (drapeau égal à 1) l'utilisation des échantillons des blocs voisins issus de la prédiction inter-image.

Couche macrobloc: ensemble des données relatives au codage d'un macrobloc, présent dans une NALU de type slice.

CPDT : Cascaded Pixel-Domain Transrater.

CSMA/CA : Carrier Sense Multiple Access with Collision Avoidance.

DC transform coefficient : Coefficient résiduel dont l'indice fréquentiel est égal à zéro (dans au moins une des deux dimensions).

DCT : Discrete Cosine Transform, Transformée en cosinus discrète.

DDT : Discrete cosine transform-Domain transrater.

Deblocking filter : voir filtre anti-bloc.

DMT : Discrete Multi Tone.

dQP : delta quantization parameter.

Drift : propagation d'erreur.

DVB : Digital Video Broadcasting.

DVD : Digital Versatile Disc.

Emulation_prevention_three_byte : Pour prévenir un éventuel problème lié à l'apparition d'un start code en plein milieu d'une NALU, un octet égal à 0x03 est inséré avant l'octet égal à 0x01 de la séquence d'octets qui pose problème.

EQM : Erreur Quadratique Moyenne.

ETSI : European Telecommunications Standards Institute.

FEC : Forward Error Correction.

FGS : Fine Grain Scalability.

Filtre anti-bloc : Filtre (également appelé *deblocking filter* ou *loop filter*) appliqué aux blocs décodés et conçu pour réduire la distorsion de type effet de bloc.

FL : frequency level.

FMO : flexible macroblock ordering.

FP : frequency position.

FPGA : Field Programmable Gate Array.

Frame : Image complète.

GOP : Group of Pictures.

H.264/AVC : Norme de compression vidéo également appelée MPEG-4 Part 10 ou Advanced Video Coding. Les normes UIT-T H.264 et ISO/CEI MPEG-4 Part 10 (ISO/CEI 14496-10) sont techniquement identiques.

HPPA : HomePlug Powerline Alliance.

Intra PCM : Intra Pulse Code Modulation ou I_PCM. Ce mode permet de transmettre un macrobloc sans aucune compression. Ainsi, les étapes de prédiction, transformée entière et quantification ne sont pas utilisées. Chaque échantillon de chrominance ou de luminance est transmis bit à bit (par défaut un échantillon est codé sur 8 bits).

ITU-T : International Telecommunication Union - Telecommunication Standardization Sector.

JM : logiciel de référence de H.264/AVC. La version 12.4 a été utilisée comme point de départ de l'implémentation de notre transcodeur H.264/AVC en sélection fréquentielle.

JVT : Joint Video Team.

Loop filter : voir filtre anti-bloc.

MAC : Media Access Control.

Macrobloc : Ensemble de 16x16 échantillons.

MGS : Medium Grain Scalability.

Mosquito noise : Cet effet se traduit par une alternance de disparition et d'apparition d'une ou plusieurs fréquences.

MPEG : Moving Picture Experts Group.

MPEG-2: standard de compression vidéo.

MPEG-2 TS: MPEG-2 Transport Stream.

MVD: Motion Vector Differences.

NAL: Network Abstraction Layer.

NALU: Unité de NAL (NAL Unit). Sous-ensemble de données représentant des données vidéo et/ou des informations nécessaires au bon fonctionnement du décodage.

OFDM : Orthogonal Frequency Division Multiplexing.

OSI : Open Systems Interconnection.

PSNR : Peak Signal to Noise Ratio.

QAM : Modulation d'amplitude en quadrature.

QCIF : Format vidéo avec une largeur d'image de 176 pixels et une hauteur de 144 pixels.

QEF : quasi-error free. Transmission considérée comme quasiment sans erreur.

QP : quantization parameter - paramètre de quantification.

Raster scan : balayage utilisé pour la numérotation des macroblocs. Il se réalise de la gauche vers la droite et de haut en bas.

RCA : Root Cause Analysis.

Reuse mode: Mode de fonctionnement d'un transcodeur réutilisant des informations issues de l'encodage afin de minimiser sa complexité.

ROI : Region of Interest.

RTP : Real Time Protocol, protocole de transport de données temps réel.

SAD: Sum of Absolute Differences ou somme des valeurs absolues des différences.

SAR : Storage Aspect Ratio.

Slice : Ensemble de macroblocs successifs dans l'ordre du raster scan, peut regrouper la totalité des macroblocs d'une image ou bien un seul macrobloc.

SNR : Signal to Noise Ratio.

SSE: Streaming SIMD extensions. Instructions SSE2 et SSE3 pour ordinateurs PC.

Start Code: Ensemble d'octets permettant d'identifier le début d'une NALU.

SVC: Scalable Video Coding.

SVCD: Moins connu en Europe, le *super video compact disc* a été encouragé par le gouvernement chinois afin de stocker des vidéos compressées au format MPEG-2 sur un disque compact.

TEB : Taux d'erreur binaire.

TES : Taux d'erreur symbole.

TNT: Télévision Numérique Terrestre.

TOSCANE : projet TOSCANE (Transmission vidéo Optimisée Source/ Canal échelonnable).

TotalCoeffs: nombre de coefficients non nuls dans un bloc de coefficients résiduels fréquentiels.

Trailing ones : Nombre de coefficients dont la valeur absolue est égale à 1.

Transcodeur : Nom général donné à une application changeant ou non le format de compression du flux vidéo à l'entrée, modifiant éventuellement son débit, la taille de l'image ou d'autres paramètres. Notre solution de transcodage H.264/AVC en qualité est appelée ici transcodeur mais aurait pu être appelée *transrater*, l'équivalent français n'étant pas forcément évident.

Transrater : Transcodeur vidéo diminuant le débit du flux vidéo en entrée tout en restant dans le même standard de compression vidéo.

VCD : Video Compact Disc.

VCEG : Video Coding Experts Group (UIT-T).

VCL : Video Coding Layer.

VDSL : Very high bit rate Digital Subscriber Line.

VLC : Variable-length coding.

VoD : Video on Demand.

VQM : Video Quality Metric.

Water-filling : Théorie permettant une allocation optimale des puissances [GAL68].

Zig-zag : balayage utilisé afin d'établir un ordre de traitement des coefficients résiduels par rang fréquentiel approximativement croissant.

Bibliographie

- [ANS97]** ANSI T1E1.4 Standard, "Asymmetric Digital Subscriber Line", 1997.
- [ANS03]** ANSI T1.801.03 – 2003, "American National Standard for Telecommunications – Digital transport of one-way video signals -Parameters for objective performance assessment," American National Standards Institute.
- [ASL92]** J.T. ASLANIS et J.M. CIOFFI, "Achievable information rates on digital subscriber loops : limiting information rate with Xtalk noise", IEEE Transactions on communications, Vol. 40, No. 2, February 1992.
- [BER93]** C. BERROU, A. GLAVIEUX ET P. THITIMAJSHIMA, "Near Shannon Limit Error – Correcting Coding and Decoding : Turbo Codes", IEEE ICC, Geneva, 1993.
- [BER08]** M. BERTOCCO, M. FARIAS, D. FORTIN, A. SONA, "Cross-Layer Measurement for the Analysis of DVB-T System Performance", IEEE Transactions on Instrumentation and Measurement, VOL. 57, NO.7, July 2008
- [BIN00]** J. BINGHAM, "ADSL, VDSL and Multicarrier Modulation", Wiley Series in Telecommunications and Signal Processing, A wiley Interscience publication, 2000.
- [BUE06]** D. BUECH, « Estimation de canaux sélectifs en fréquence par interpolation pour les systèmes DMT, application aux PLC », Thèse soutenue à l'Université de Valenciennes, Laboratoire IEMN-DOAE, 12 juillet 2006.
- [BUE11]** D. BUECHE, P. CORLAY, F-X COUDOUX, M. GAZALET, C. DEKNUDT, "Analysis of Optimal Power Distribution over Pilot tones for Multi-carrier Communications over PLC", Proceedings of the 2011 IEEE International Symposium on Power Line Communications and its Applications, ISPLC 2011, Udine, Italy, April 3-6, 2011.
- [CAM98]** J. CAMPELLO, "Optimal Discrete Bit Loading for Multicarrier Modulation Systems", 1998, IEEE Intern. Symp. On Inform. Th., Août 1998, Cambridge, MA.

- [CHO95]** P.S. CHOW, J.M. CIOFFI, J.A.C. BINGHAM, "A Practical Discrete Multitone Transceiver Loading Algorithm for Data Transmission over Spectrally Shaped Channels", IEEE Transactions on Communications, vol. 43, n^{os} 2/3/4, February/March/April 1995.
- [CIO91]** J.M. CIOFFI, "A multicarrier primer", ANSI T1E1.4 comittee contribution n^o91-157, November 1991.
- [COL99]** M. COLIN, « Étude de l'Optimisation d'un Système DMT-ADSL, Application à la transmission vidéo MPEG-2 en mode hiérarchique », Thèse soutenue à l'Université de Valenciennes, Laboratoire IEMN-DOAE, 28 Janvier 1999.
- [COU08]** F.-X. COUDOUX, M. GAZALET, C. MOUTON-GOUDEMAND, P. CORLAY, M.GHARBI, "Extended coverage for DSL video distribution using a quality-oriented JSCC architecture", IEEE Trans. On Broadcasting, Special Issue: Quality Issues in Multimedia Broadcasting, VOL.54, NO.3, pp.525-531, Sept. 2008
- [DEC06a]** J. DE COCK, S. NOTEBAERT, P. LAMBERT, D. DE SCHRIJVER, R. VAN DE WALLE, "Requantization Transcoding in Pixel and Frequency Domain for Intra 16x16 in H.264/AVC. ", Proceedings of Advanced Concepts for Intelligent Vision Systems Conference, pp.533-544, Antwerp 2006.
- [DEC06b]** J. DE COCK, S. NOTEBAERT, P. LAMBERT, K. DE WOLF, AND R. VAN DE WALLE, "Low-Complexity SNR Transcoding for H.264/AVC", Proceedings of the Fourth IASTED International Conference on Communications, Internet and Information Technology, December 2006, St. Thomas, US Virgin Islands.
- [DEK09]** C. DEKNUDT, P. CORLAY, A.S. BACQUET, F.X. COUDOUX, M. GUARISCO, H. RABAH, Y. BERVILLER, S. WEBER, "Transrating by frequencies selectivity for H.264/AVC Intra pictures", IEEE BMSB, May 13-15, Bilbao, 2009.
- [DEK10]** C. DEKNUDT, P. CORLAY, A.S. BACQUET, M. ZWINGELSTEIN-COLIN et F.X. COUDOUX, "Reduced Complexity H.264/AVC Transrating based on Frequency Selectivity for High-Definition Streams", IEEE Transactions on Consumer Electronics, Vol. 56, No. 4, pp.2430-2437, Novembre 2010.

[DEK11a] C. DEKNUDT, A-S BACQUET, P. CORLAY, F-X. COUDOUX, M. SLACHCIAK, "Efficient Adaptation to Channel State Changes Based on H.264/AVC Transrating for Power Line Transmission of Video Streams", Proceedings of the 2011 IEEE International Symposium on Power Line Communications and its Applications, ISPLC 2011, Udine, Italy, April 3-6, 2011, 141-145.

[DEK11b] C. DEKNUDT, P. CORLAY, F-X COUDOUX, A-S. BACQUET, "Comparison between H.264/AVC Intra Transrating Schemes Based on Frequency Position and Frequency Level Selectivity", Proceedings of the 2011 IEEE International Symposium on Broadband Multimedia Systems and Braodcasting, BMSB 2011, Erlangen, Germany, June 8-10, 2011.

[DEW07] K. DE WOLF, D. DE SCHRIJVER, W. DE NEVE, S. DE ZUTTER, P. LAMBERT, R. VAN DE WALLE, "Analysis of Prediction Mode Decision in Spatial Enhancement Layers in H.264/AVC SVC", Springer-Verlag Berlin Heidelberg 2007

[DVB09] Digital Video Broadcasting, "Framing Structure, channel coding and modulation for Satellite Services to Handheld devices (DVB-SH) below 3 GHz", Décembre 2009.

[ESM00] T. ESMAILIAN, P. G. GULAK, KSCHISCHANG F. R., "A discrete multitone power line communication system", proc. ICASSP Istanbul, p. 2953-2956, 2000.

[ETS96] ETSI Technical Report ETR328, "Transmission and Multiplexing (TM), Asymmetric Digital Subscriber Line (ADSL)", Requirements and Performance, Novembre 1996.

[FER08] P. FERRE, J. CHUNG-HOW, D. BULL and A. NIX, "Distortion-Based Link Adaptation for Wireless Video Transmission", EURASIP Journal on Advances in Signal Processing, Volume 2008, Article ID 253706, March 2008.

[FER10] H. C. FERREIRA, L. LAMPE, J. NEWBURY, T. G. SWART, "Power Line Communications – Theory and Applications for Narrowband and Broadband Communications over Power Lines", Wiley Press, 2010.

Bibliographie

- [GAL68]** R. GALLAGER, Information Theory and Reliable Communication, JohnWiley & Sons, New York, 1968.
- [GOU05]** C. GOUEMAND, «Amélioration des transmissions vidéo MPEG-2 sur les lignes ADSL basée sur une optimisation conjointe source/canal de la qualité vidéo », Thèse soutenue à l'Université de Valenciennes, Laboratoire IEMN-DOAE, 15 Décembre 2005.
- [HO96]** M. HO, J.M. CIOFFI, J. BINGHAM, "Discrete multitone echo cancellation", IEEE Transactions on communications, Vol. 44, No. 7, July 1996.
- [HPP01]** HOMEPLUG POWERLINE ALLIANCE, "HomePlug 1.0 Technology White Paper", 2005
- [HPP05]** HOMEPLUG POWERLINE ALLIANCE, "HomePlug AV White Paper", 2005
- [HUA07]** H-C. HUANG, W-H. PENG, T. CHIANG, H-M. HANG, "Advances in the Scalable Amendment of H.264/AVC", IEEE Communications Magazine, Janvier 2007.
- [HUG87]** D. HUGHES-HARTOGS, "Ensemble modem structure for imperfect transmission media", US Patents, 1987.
- [IEE11]** IEEE p1901 draft standard for broadband over power line networks: Medium access control and physical layer specifications.
- [ITU01]** ITU-T, Recommendation G.996.1, "Test procedures for digital subscriber line (DSL) transceivers", Février 2001.
- [ITU09]** ITU-T, Recommendation H.264, "Advanced video coding for generic audiovisual services Corrigendum 1", Janvier 2009.
- [LAM06]** P. LAMBERT, W. DE NEVE, Y. DHONDT, R. VAN DE WALLE, "Flexible macroblock ordering in H.264/AVC", Journal of Visual Communication and Image Representation, 17 (2006) 358–375

- [LEF06]** D. LEFOL, D. BULL, N. CANAGARAJAH, " Performance evaluation of transcoding algorithms for H.264", IEEE Transactions on Consumer Electronics, VOL.52, Issue.1, pp.215-222, Feb 2006.
- [LEI03]** Z. LEI and N.D. GEORGANAS, "An accurate bit-rate control algorithm for video transcoding", Journal of Visual Communication and Image Representation, vol. 14, no. 3, pp. 321-339, Sept. 2003.
- [LEO08]** A. LEONTARIS and A. M. TOURAPIS, "Drift Characterization of Intra Prediction and Quantization in H.264", in Proc. IEEE Data Compression Conference, Snowbird, Utah, pp. 212-221, March 25-27, 2008.
- [MAL03]** H. S. MALVAR A. HALLAPURO, M. KARCZEWICZ, L. KEROFISKY, "Low-Complexity Transform and Quantization in H.264/AVC", IEEE Transactions on Circuits and Systems for Video Technology, VOL. 13, NO. 7, pp. 598-603, Juillet 2003.
- [MAR03]** D. MARPE, H. SCHWARZ et T.WIEGAND "Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard", IEEE Transactions on Circuits and Systems for Video Technology, VOL. 13 , NO. 7, pp. 620-636, Juillet 2003.
- [PIN04]** M. H. PINSON, S. WOLF, "A new standardized method for objectively measuring video quality", IEEE Trans. On Broadcasting, p312-322, Sept. 2004.
- [PRO95]** J.G. PROAKIS, Digital Communications, Mc Graw Hills International, 3ème édition, 1995.
- [RIC10]** I. RICHARDSON, "The H.264 Advanced Video Compression Standard 2d Edition", Wiley Press, 2010.
- [SCH05]** H. SCHWARZ, T. HINZ, D. MARPE, T. WIEGAND, "Constrained Inter-Layer Prediction for Single-Loop Decoding in Spatial Scalability", International Conference on Image Processing – ICIP, pp. 870-873, 2005.

- [SCH07]** H. SCHWARZ, D. MARPE et T. WIEGAND, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE transactions on circuits and systems for video technology, vol. 17, no. 9, september 2007.
- [SCH08]** H. SCHARZ et M. WIEN "The Scalable Video Coding Extension of the H.264/AVC Standard" IEEE signal processing magazine, standards in a nutshell, pp. 135-141, Mars 2008.
- [SHA48]** C. SHANNON, "A Mathematical Theory of Communications: Part 1", The Bell Systems Technical Journal, Juillet 1948
- [STA02]** T. STARR, M. SORBARA, J.M. CIOFFI et P. SILVERMAN, "DSL Advances", Upper Saddle River: Prentice Hall, 2002.
- [SUL05]** G.SULLIVAN et T.WIEGAND "Video Compression – From Concepts to the H.264/AVC Standard", proceedings of the IEEE vol 93,pp. 18-31, Janvier 2005
- [TOS08]** Projet ANR TOSCANE, Délivrable D3.4, « Détermination conjointe des paramètres de codage de source et de canal (cas non scalable) », juillet 2008.
- [VQE03]** VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II," August 25, 2003.
- [WIE03a]** T. WIEGAND, G. J. SULLIVAN, G. BJONTEGAARD et A. LUTHRA, "Overview of the H.264 / AVC Video Coding Standard", IEEE transactions on circuits and systems for video technology, july 2003
- [WIE03b]** T. WIEGAND, H. SCHWARZ, A. JOCH, F. KOSENTINI et G. SULLIVAN, "Rate-Constrained Coder Control and Comparison of Video Coding Standards", IEEE transactions on circuits and systems for video technology, july 2003
- [WU06]** H.R. WU and K.R. RAO, "Digital Video Image Quality and Perceptual Coding", CRC Press, Taylor & Francis Group, 2006, pp. 123-154

Bibliographie

[XIN05] J. XIN, C. LIN, M. SUN, "Digital Video Transcoding", Proceedings of the IEEE, VOL.93, NO.1, pp.84-97, January 2005.

[ZHO05] Y. ZHONG, I. RICHARDSON, A. MILLER and Y. ZHAO, "Perceptual Quality of H.264/AVC Deblocking Filter", IEE VIE 2005, pp379-384, Glasgow, 4-6 April, 2005.

[ZIM00] M. ZIMMERMANN, K. DOSTERT, "The low voltage power distribution network as last mile access network", AEU International journal of electronics and commun., vol 54, n°1, p. 13-22, 2000.

Titre : Mise en oeuvre d'architectures de transcodage vidéo H.264/AVC et SVC : application à la transmission optimisée de la vidéo haute définition

Résumé : Les travaux présentés dans ce manuscrit de thèse se sont déroulés majoritairement dans le cadre du projet ANR TOSCANE dont l'objectif était d'optimiser les transmissions de flux vidéo haute définition. Pour cela, deux types d'architectures de transcodage à complexité réduite des flux H.264/AVC sont proposés afin d'adapter le débit vidéo à la bande passante du canal de transmission.

La première architecture s'applique aux flux H.264/AVC et réalise une sélection fréquentielle des coefficients résiduels. Après comparaison en termes de réduction de débit et de qualité vidéo entre cette solution et la technique de requantification, nous avons constaté que l'architecture proposée donnait majoritairement de meilleurs résultats. Puis, cette solution est intégrée dans un scénario de transmission vidéo haute définition par courant porteur en ligne. Ce type de canal est soumis à des changements d'états fréquents nécessitant une nouvelle estimation du canal et allocation des bits et des puissances afin de garantir un débit quasiment sans erreur. Au prix d'une légère baisse de PSNR, les flux sont dynamiquement transcodés afin d'adapter leur débit à celui du canal et rendre ainsi la transmission possible.

La seconde architecture s'applique aux flux H.264 SVC utilisant une échelonnabilité spatiale. Elle consiste à sélectionner par position fréquentielle les coefficients résiduels de la couche d'amélioration la plus haute afin d'obtenir des débits intermédiaires tout en restant dans la définition spatiale la plus élevée. Cette solution de transcodage est mise en œuvre dans le cas d'une transmission ADSL, permettant d'augmenter la zone d'éligibilité des services vidéo en haute définition et fournissant aux abonnés une qualité vidéo progressivement réduite.

Mots clés : Transmission vidéo numérique, H.264/AVC, H.264 SVC, Transcodage vidéo, Réduction de débit, compression vidéo, adaptation de contenu, Échelonnabilité.

Title: Implementation of H.264/AVC and SVC transrating architectures: application to optimized transmission of high definition video streams

Abstract: The works presented in this thesis are mainly part of the French ANR TOSCANE project which aims to optimize high definition video transmission. We propose two types of low complexity transrating architectures for H.264/AVC streams to adapt video bitrate to channel bandwidth.

The first transrating architecture is based on frequency selectivity of residual coefficients from H.264/AVC streams. A comparison between this solution and the well known requantization processing in terms of bitrate reduction and video quality shows that our solution often gives better results. Then, our solution is included in a high definition video transmission scheme using power line communications. This type of channel can exhibit sudden changes states; as a consequence transmission needs a new channel estimation, bit and power allocation to guarantee quasi error free transmission. By means of a slight video distortion, transrating architecture is used to dynamically adapt video bitrate to channel one.

The second transrating architecture is applied to H.264 SVC streams using spatial scalability. It consists in selectively removing residual coefficients of upper enhancement layer to obtain intermediate bitrates with upper spatial definition. This transrating solution is used in an ADSL transmission, allowing the extension of the area of eligibility for high definition video services while providing a progressive reduced video quality to subscribers.

Key words: Digital video transmission, H.264/AVC, H.264 SVC, Transrating, Bitrate reduction, Digital video compression, Content adaptation, Scalability.

