

Rapport sur le manuscrit de thèse de

Mathieu Léonardon

intitulé

**« Décodage de codes polaires sur des architectures
programmables »**

par Amer BAGHDADI

Professeur à l'IMT Atlantique, Institut Mines-Télécom

La thèse de Monsieur Mathieu Léonardon s'est déroulée en cotutelle entre le laboratoire IMS à l'Université de Bordeaux et le département de Génie Electrique à l'Ecole Polytechnique de Montréal, sous la direction des Professeurs Christophe Jégo et Yvon Savaria et le co-encadrement de Dr. Camille Leroux.

Le sujet abordé s'inscrit dans un thème de recherche lié à la proposition de nouvelles solutions pour le décodage de codes polaires sur des architectures programmables avec une perspective d'offrir un compromis original entre flexibilité, performance et efficacité énergétique.

Il s'agit d'un sujet d'actualité, en particulier avec la récente adoption de cette nouvelle famille de codes correcteurs d'erreurs pour le codage des canaux de contrôle dans la 5^{ème} génération de réseaux mobiles (5G). Inventés il y a juste dix ans par Erdal Arikan, les codes polaires suscitent l'intérêt de nombreuses équipes de recherche académiques et industrielles. Dans ce contexte, en plus des travaux sur la définition de nouvelles constructions de codes polaires et sur la proposition de nouveaux algorithmes et techniques de décodage, il devient crucial d'investiguer et de proposer des implémentations logicielles et matérielles efficaces. Par ailleurs, les applications émergentes, adoptant des codes correcteurs d'erreurs, imposent des contraintes de plus en plus fortes en termes de qualité de transmission, de débit, de latence et d'efficacité énergétique, et souvent spécifient différents modes de communication avec des paramètres variés (e.g. construction du code, taille du code, rendement). Ainsi, la généricité/flexibilité des solutions d'implémentation proposées devient un critère de haute priorité. Ceci est d'autant plus vrai dans une tendance de virtualisation et de coopération des traitements numériques du signal et de l'infrastructure de réseau, soulignée dans ce manuscrit de thèse.

Les travaux présentés dans ce manuscrit s'inscrivent dans ce contexte et visent ainsi des implémentations logicielles sur des architectures programmables en vue d'améliorer la flexibilité du réseau. Plusieurs contributions originales sont proposées aussi bien au niveau de la description logicielle des algorithmes de décodage qu'au niveau de la conception matérielle d'architectures programmables spécialisées. Ces contributions significatives et cohérentes sont regroupées en trois parties : (1) proposition d'un décodeur logiciel pour les algorithmes de décodage de codes polaires à liste avec un grand degré de flexibilité dynamique et des optimisations permettant de dépasser les débits des implémentations logicielles existantes sur cible processeurs x86, (2) conception d'un processeur à jeu d'instructions dédié au décodage par annulation successive (XTensa Polaire) basé sur un cœur de processeur XTensa de Tensilica/Cadence permettant de réduire d'un ordre de grandeur la consommation énergétique, (3) conception d'une nouvelle architecture d'ASIP (TT-SC) basée sur l'approche TTA (Transport Triggered Architecture) permettant de réduire de deux ordres de grandeur la consommation énergétique tout en améliorant le débit.

Analyse du document :

Le manuscrit de thèse soumis est constitué d'un document de 125 pages, rédigé en français et composé d'une introduction générale précédé d'un glossaire d'acronymes et d'une liste de définitions des notations utilisées, d'un chapitre dédié à la présentation des concepts fondamentaux des codes polaires, de trois chapitres détaillant les contributions scientifiques de la thèse et d'un chapitre succinct de conclusions et perspectives.

Dans l'introduction générale, Mathieu Léonardon présente brièvement le contexte et les motivations de son sujet de thèse en s'appuyant sur les enjeux et les défis techniques des codes correcteurs d'erreurs et de leurs architectures. La structure du manuscrit est ensuite détaillée en récapitulant les contributions principales de la thèse.

Le **chapitre 1** donne un aperçu général sur les codes polaires et les différents algorithmes de décodage proposés dans la littérature. Il introduit judicieusement les éléments nécessaires à la lecture des chapitres suivants. Le chapitre présente tout d'abord le modèle de base d'une chaîne de communications numériques et précise le modèle de canal considéré pour caractériser les performances en termes de taux d'erreurs. Après cette brève introduction, les codes polaires et le principe de polarisation sont présentés. Ensuite, le chapitre détaille les différentes variantes algorithmiques concernant le décodage de ces codes. Ainsi, le principe de l'algorithme de décodage par Annulation Successive (SC) est illustré et les niveaux de parallélisme associés sont explicités. La version améliorée de type SC Liste est détaillée par la suite avec plusieurs variantes : concaténation avec un CRC, décodage adaptatif, SC Flip, SC Stack, ainsi que deux algorithmes itératifs à sortie souple (Belief Propagation et SCAN). Plusieurs courbes de performance de décodage sont aussi fournies pour comparer ces différents algorithmes et les impacts des divers paramètres associés (taille de trame, de liste, de CRC, rendement). La présentation et les discussions menées sont structurées de manière particulièrement claire, cohérente et didactique. Le chapitre se termine par la présentation de la technique d'élagage de l'arbre de décodage en vue de réduire sa complexité.

Le **chapitre 2** présente la première contribution concernant la proposition d'un décodeur logiciel qui implémente les algorithmes de décodage à liste de codes polaires. Il commence par motiver l'intérêt de telles solutions dans un contexte d'évolution des réseaux mobiles vers une infrastructure virtualisée. Ensuite, une analyse des solutions logicielles existantes est présentée. Ces solutions exploitent principalement les techniques de vectorisation (instructions SIMD) et de déroulage du code source avant compilation pour réduire le temps d'exécution. Pour mettre en valeur l'originalité de la solution logicielle proposée, une partie importante du chapitre est consacrée à la description et à la justification des différents niveaux de flexibilité supportés (ou de généricité et de flexibilité comme définies pour M. Léonardon). Dans ce cadre, plusieurs analyses et discussions pertinentes, appuyées par des simulations et des résultats originaux, sont menées sur la flexibilité en termes de : (1) taille de code polaire, (2) profondeur liste de l'algorithme CASCL, (3) paramétrage fin de l'élagage de l'arbre de décodage, (4) format de représentation des données et (5) version adaptative du décodage à liste (PASCL et FASCL). Ce travail est ensuite complété par la présentation de trois nouvelles techniques d'optimisation pour accélérer l'exécution de l'implémentation logicielle proposée : (1) utilisation nouvelle de la méthode de Schreier pour le tri des métriques et des LLR, (2) stockage de séquences calculées à l'avance du CRC et extraction parallèle des bits d'information avant vérification du CRC, (3) gestion des sommes partielles des arbres de décodage par copy ou à l'aide de pointeurs. Le chapitre se termine par la présentation de résultats d'expérimentation sur des CPU Intel i5 et i7 et des comparaisons avec l'état de l'art. Malgré le grand degré de flexibilité dynamique de l'implémentation proposée, des gains significatifs en débit sont illustrés en particulier avec l'algorithme complètement adaptatif FASCL. Cet algorithme implique cependant une latence plus élevée dans le pire cas. Les performances sont évaluées également sur d'autres cibles matérielles (Intel, AMD, quatre versions différentes de processeurs ARM). Une discussion et des comparaisons pertinentes sont menées dans ce contexte entre les débits atteints et la consommation énergétique sur ces différentes cibles. M. Léonardon indique pour ces résultats qu'un seul cœur de processeur a été utilisé. Il indique également que la puissance mesurée est celle du cœur de processeur seulement, excluant la consommation d'énergie de la mémoire principale et de la mémoire cache de niveau 3. Ces éléments et la précision de mesure de la consommation énergétique pour les différentes

ciblés considérées auront mérités, selon moi, une discussion et une justification plus détaillées. Globalement, les nombreuses analyses et résultats originaux présentés dans ce chapitre illustrent la flexibilité, l'efficacité et la portabilité de l'implémentation proposée qui a été d'ailleurs mise à la disposition de la communauté scientifique et qui constituera certainement une référence dans le domaine.

En vue d'améliorer les performances du décodage logiciel des codes polaires, en particulier en terme d'efficacité énergétique, le **chapitre 3** présente une architecture de processeur à jeu d'instructions dédié au décodage des codes polaires (XTensa Polaire). Le chapitre commence par une introduction sur les architectures des processeurs RISC et les méthodologies de conception d'ASIP. Le flot de conception retenu est celui des processeurs extensibles proposé par Tensilica qui permet de configurer et d'étendre un cœur RSIC avec des instructions spécialisées. Considérant la version SC de l'algorithme de décodage, plusieurs choix architecturaux et techniques de spécialisation ont été proposés : activation de la fonctionnalité FLIX permettant le support d'un parallélisme d'instructions de type VLIW, configuration des mémoires caches et augmentation de la taille des registres et des interfaces mémoires, ajout de nouvelles instructions spécialisées et les unités matérielles dédiées associées. Pour ce dernier, des instructions *multi-registres* et *simple-registre* ont été définies pour accélérer les fonctions polaires élémentaires avec un niveau de parallélisme de 64 et une file de registres vectoriels. De plus, pour réduire la durée des opérations liées au parcours de l'arbre de décodage et aux tests associés, cette fois la technique de déroulage du code source avant compilation est adopté. Pour conserver la généricité du code logiciel, ce déroulage est automatisé astucieusement à partir des paramètres du code polaire, donnés en entrée du programme. L'impact de ces extensions est évalué dans la section d'expérimentation. Un gain impressionnant en débit, avec un facteur d'accélération d'environ x50 par rapport à l'architecture XTensa de base, est illustré. Concernant la comparaison avec un processeur ARM Cortex A57, les résultats montrent également une réduction significative en nombre de cycles d'horloges nécessaires au décodage d'un mot de code (rapport d'environ 2 à 4). Cependant, cette réduction semble être contrebalancée par le rapport entre les fréquences d'horloge des deux processeurs. Le résultat original de la solution proposée reste principalement dans la consommation énergétique qui a été réduite d'un ordre de grandeur. Dans ce contexte, bien que M. Léonardon a souligné explicitement la limitation de la version académique des outils de conception de Tensilica, il aurait été utile d'étendre et de justifier davantage l'analyse de certaines estimations considérées dans la présentation des résultats. En particulier, si la puissance consommée estimée prend bien en compte les extensions spécialisées de l'ASIP, les mémoires caches et RAM, et quel modèle d'accès aux mémoires est considéré pour l'estimation du nombre de cycles d'horloge.

Le **chapitre 4** présente une nouvelle architecture de processeur à jeu d'instructions dédié au décodage des codes polaires, basée cette fois sur l'approche TTA (Transport Triggered Architecture). Le chapitre commence par justifier le choix de cette approche qui permet un degré de spécialisation plus important que l'approche Tensilica avec en plus la génération automatique du modèle matériel et un ensemble complet d'outils de conception et de développement. Le modèle de base et les concepts principaux des processeurs TTA sont illustrés à travers un exemple. Le flot de conception et le rôle des différents outils de la suite logicielle libre utilisée (TCE) sont également présentés. Une architecture de processeur dédié au décodage SC est détaillée (TT-SC). L'architecture intègre une partie avec des unités fonctionnelles de base, une unité de chargement et sauvegarde (LSU) vectorielle et des unités de calcul polaire pour réaliser les fonctions élémentaires de décodage SC. Certains choix architecturaux sont similaires aux instructions spécialisées de l'architecture de processeur XTensa. La latence de l'unité LSU vectorielle a été réduite à un cycle d'horloge et des unités matérielles ajoutées pour permettre un chargement et sauvegarde de données non-alignées. L'ASIP intègre également une unité matérielle de décodage d'un sous-arbre avec un traitement déroulé et multi-cycles. Concernant la description logicielle du décodage, il est proposé de dérouler le code source pour réduire le nombre de calculs d'adresse et le nombre d'indirections. Dans ce contexte, une analyse intéressante est présentée concernant le taux d'occupation des bus et des unités fonctionnelles. L'architecture de cet ASIP a été également étendue pour supporter l'algorithme SCAN (décodage itératif à sorties souples) en plus de l'algorithme SC. Ainsi, une seconde architecture de processeur dédié a été conçue, nommée TT-SCAN. L'organisation de cette nouvelle architecture, avec partage et adaptation de certaines ressources matérielles, illustre bien la modularité de l'approche et la pertinence des choix effectués. Le

chapitre se termine par la présentation des expérimentations et des résultats de mesures et de comparaison des performances avec des processeurs à usage général (Intel i7 et ARM Cortex A57), ainsi qu'avec l'ASIP XTensa proposé dans le chapitre 3. Cette comparaison est également étendue aux architectures matérielles dédiées grâce à des implémentations sur cible FPGA et ASIC. Plusieurs résultats originaux, accompagnés de discussions et d'analyses pertinentes, sont présentés dans ce contexte. Le débit de l'ASIP TT-SC proposé dépasse celui du CPU Intel i7 avec une consommation énergétique réduite de deux ordres de grandeur. Sur cible ASIC 28 nm, le débit du décodeur SC atteint 352 Mb/s pour une consommation énergétique de 0,14 nJ/bit. Dans ce contexte, la méthode et la précision de mesure de la consommation énergétique auront pu être détaillées. Pour compléter l'étude, il aurait été peut-être intéressant d'utiliser des métriques combinées (e.g. débit/surface) et d'illustrer les niveaux de flexibilité supportés avec cette nouvelle architecture d'ASIP dans la section d'expérimentation et mesures. Enfin, les résultats déjà fournis sont convaincants et montrent comment la solution proposée réalise un compromis très prometteur entre les processeurs à usage général et les architectures matérielles dédiées.

Ces premières contributions sur la conception d'ASIP spécialisés dans le décodage par annulation successive des codes polaires, ainsi que les analyses et comparaisons détaillées des performances présentées dans ce travail de thèse, constitueront certainement une référence pour la littérature sur ce sujet d'actualité.

Finalement, le document se termine par un chapitre succinct présentant une conclusion générale qui résume les travaux réalisés ainsi que des perspectives de recherche pertinentes pour compléter et étendre les contributions proposées.

Conclusion :

Globalement, le manuscrit est rédigé de manière particulièrement claire, précise et didactique. Sa structure est tout à fait cohérente et présente clairement les contributions scientifiques de Mathieu Léonardon. Le contexte des travaux, les principales idées et les choix effectués sont illustrés avec pertinence. Les expérimentations effectuées et les analyses menées permettent d'évaluer et de comparer les solutions proposées et montrent des résultats originaux très prometteurs.

M. Léonardon a effectué un travail conséquent sur un sujet d'actualité qui nécessite des compétences avancées en conception d'architectures matérielles de traitement, en programmation logicielle, ainsi qu'en communications numériques. Les contributions scientifiques proposées sont très pertinentes et significatives en quantité et en qualité. Au niveau de la valorisation scientifique, le manuscrit de thèse affiche une liste de quatre publications dans des conférences internationales (SiPS'2017, NEWCAS'2017, ISCAS'2018, ISTC'2018) et deux soumissions d'articles en revues internationales.

Par conséquent, après la lecture de son manuscrit de thèse, j'estime que le travail présenté par Mathieu Léonardon est original, comprend des contributions scientifiques pertinentes et présente toutes les qualités requises. J'émet donc un avis très favorable à la tenue de la soutenance de thèse de Monsieur Mathieu Léonardon en vue de l'obtention du titre de docteur de l'Université de Bordeaux et de l'Ecole Polytechnique de Montréal.

Brest, le 30 novembre 2018



Amer BAGHDADI, HDR
Professeur à l'IMT Atlantique
CNRS UMR 6285 Lab-STICC
Institut Mines-Télécom