Conception pour l'efficacité énergétique des circuits intégrés

Sylvain Engels

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : NANO ELECTRONIQUE ET NANO TECHNOLOGIES

Arrêté ministériel : 25 mai 2016

Présentée par

Sylvain ENGELS

Thèse dirigée par **Laurent FESQUET**,
Maître de Conférence Grenoble INP / Phelma, Université Grenoble Alpes

préparée au sein du **Laboratoire des Techniques de l'Informatique et de la Microélectronique pour l'Architecture des systèmes intégrés**

dans **l'École Doctorale Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)**

Conception pour l'efficacité énergétique des circuits intégrés

Energy efficiency technics for advanced CMOS technologies

Thèse soutenue publiquement le **1 juillet 2022**,
devant le jury composé de :

Monsieur Ian O'CONNORS
PROFESSEUR, Ecole Centrale de Lyon, Président
Monsieur Luc HEBRARD
PROFESSEUR, Université de Strasbourg, Rapporteur
Monsieur Lionel TORRES
PROFESSEUR, Université de Montpellier, Rapporteur
Madame Lorena ANGHEL
PROFESSEUR, Université Grenoble Alpes - Grenoble INP, Examinatrice
Madame Andreia CATHELIN
INGENIEUR HDR, STMicroelectronics, Examinatrice
Monsieur Laurent FESQUET
MAITRE DE CONFERENCE, Université Grenoble Alpes, Directeur de thèse

"You Can Hide the Latency,

But,

You Cannot Hide the ENERGY !!"

Peter M Kogge

Table des Matières

Introduction

A ses débuts, la filière microélectronique s'est concentrée uniquement sur l'augmentation des performances des puces et la densité d'intégration des composants par unité de surface. Cette course à l'intégration a fait émerger progressivement la problématique de la densité de puissance au point d'en faire dès les années 90 un élément majeur à prendre en compte lors de la conception. En effet, même si la gestion de l'énergie a toujours été un élément important de la conception des circuits (montre à quartz, calculatrice), l'avènement des solutions nomades et l'augmentation de la densité des transistors l'ont propulsée sur le devant de la scène, au point d'en devenir le facteur prédominant à prendre en compte lors de la conception des circuits intégrés.

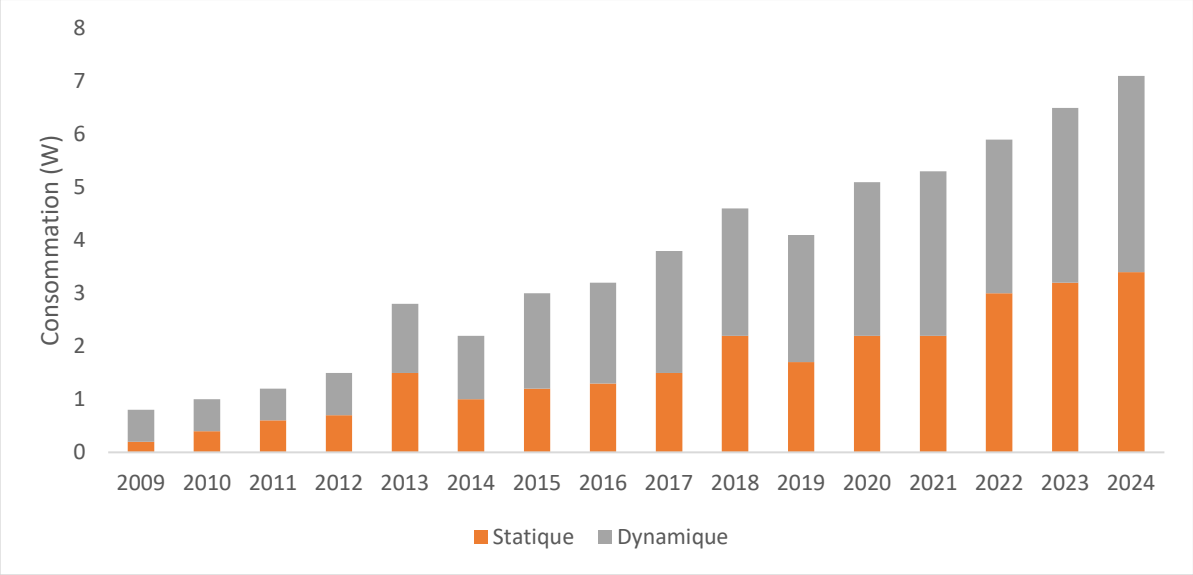


Figure 1 : Consommation dans les Système Portable [ITRS, 2010 Update]

Cette gestion de l'énergie peut s'analyser sous différents aspects en fonction du besoin et du type d'applications visées. Nous pouvons ainsi classer les produits en trois classes : HP (*High Performance*), LOP (*Low Operating Power*) et LPST (*Low STandby Power*). Nous retrouvons tout d'abord les circuits nécessitant une performance de calcul très élevée (fermes de calcul, processeurs graphiques, routeurs, …). La puissance dissipée par ces circuits classifiés HP se retrouve sous forme de chaleur. La problématique, dans ce cas, est de trouver une solution de refroidissement efficace ou de dissiper de la meilleure façon cette chaleur générée lors du fonctionnement. Les deuxièmes catégories sont les produits nomades classifiés LOP pour qui le téléphone portable est sûrement le représentant le plus connu. Dans

ce cas, la performance reste un élément important mais l'autonomie de la batterie fait aussi partie de l'équation. Enfin, la dernière catégorie LSTP est définie par les produits majoritairement en veille. C'est ici le domaine de l'internet des objets où nous retrouvons des produits quasi-exclusivement sur batterie et qui passent la majorité de leur temps en veille.

Bien sur ces catégories ne sont pas exclusives et, en fonction des applications, certains circuits pourront avoir besoin de basculer d'une catégorie à l'autre. Ceci est particulièrement vrai pour les applications mobiles où il est apparu de nouveaux besoins (jeux, vidéos, connectivité radio...) qui sont demandeurs de performance.

Le fil conducteur du travail de recherche mené tout au long de cette thèse a été de proposer des solutions visant à optimiser l'efficacité énergétique des circuits. Du fait de sa durée particulière (~15ans), les travaux ont suivi l'évolution des techniques liées à la réduction de consommation. Les premières solutions proposées nous sembleront évidentes à l'aube des années 2020 mais elles resteront les premières briques qui ont permis l'essor des téléphones portables et plus récemment de l'internet des objets.

Ce travail de recherche a donné lieu à de nombreuses publications, communications et à plusieurs brevets dont les références sont regroupées à la fin de ce manuscrit de thèse. Il se focalise essentiellement sur les aspects de conception des circuits intégrés. Il existe aussi de nombreux leviers du côté de la technologie et des systèmes pour améliorer l'efficacité énergétique et nous positionnerons donc ces travaux comme un pont entre ces deux mondes : « Comment tirer le meilleur parti de la technologie offerte pour offrir les meilleures possibilités aux concepteurs de systèmes ? ».

Dans le premier chapitre, nous allons introduire les éléments constituants de l'efficacité énergétique – à savoir les consommations dynamique et statique – et nous introduirons les concepts fondamentaux qui seront travaillés dans les chapitres suivants. Ainsi, le vol d'horloge, les îlots d'alimentation mais aussi la technologie FDSOI et la logique asynchrone seront brièvement introduits.

Le deuxième chapitre se focalisera sur deux techniques qui s'attaqueront respectivement à réduire les consommations statique et dynamique. Le registre double-front proposé nous offrira la possibilité d'utiliser le front descendant de l'horloge et nous introduirons un flot de conception associé à l'utilisation de ce registre particulier. Pour la consommation statique, une première implémentation d'îlots et de bascules à rétention sera présenté. Ces

travaux ont servi de base pour la mise en place de solution industrielle dans la gestion des îlots d'alimentation.

Dans le troisième chapitre, nous irons chercher le « pont » avec les procédés de fabrication en s'attachant à proposer des solutions qui s'appuient sur les particularités des procédés submicroniques et/ou utilisant un substrat isolé SOI. L'augmentation des variations dans les procédés de fabrication sera mitigée par des solutions de capteur et actuateur performant nous permettant de compenser ces effets sur l'efficacité énergétique. L'essor de la technologie FD-SOI et, plus particulièrement, sa capacité intrinsèque au contrôle de la tension de substrat nous incitera à proposer des solutions de conception utilisant ce nouveau levier.

Le dernier chapitre ouvrira la voie vers l'exploration de nouvelles techniques. Nous nous focaliserons particulièrement sur la technologie asynchrone. La propriété de calcul sur évènements liée à cette technologie lui prodigue un avantage intrinsèque que seule la facilité de conception synchrone arrive à compenser. Nous proposerons ici à travers un exemple de réseau de capteurs, un flot de conception ainsi qu'un circuit réalisé avec cette technologie quasi-insensible au délai. Une variante de cette solution s'appuyant sur des données groupées nous permettra de contrôler non seulement le délai mais aussi l'émission électromagnétique de notre circuit. Cet avantage semble primordial dans les circuits réalisés pour l'internet des objets qui sont connectés par nature et donc sensibles à ces perturbations.

Enfin une conclusion globale nous permettra de rassembler ces travaux à « l'histoire de l'efficacité énergétique » et de tenter de proposer différentes perspectives à ces mêmes travaux.

Chapitre I. Les bases de l'efficacité énergétique

I. 1. La consommation dans les circuits

I. 1. 1. La problématique d'une surconsommation

I. 1. 1. 1 Le silicium noir

Une des raisons principales de la diffusion des circuits intégrés dans notre vie de tous les jours est la réduction du coût unitaire du transistor lié à la capacité d'intégration toujours plus importante offerte par les nouvelles technologies. La conjecture de Moore que toute l'industrie s'efforce de suivre exprime parfaitement cette idée et s'est révélée particulièrement juste jusqu'à la technologie 130nm.

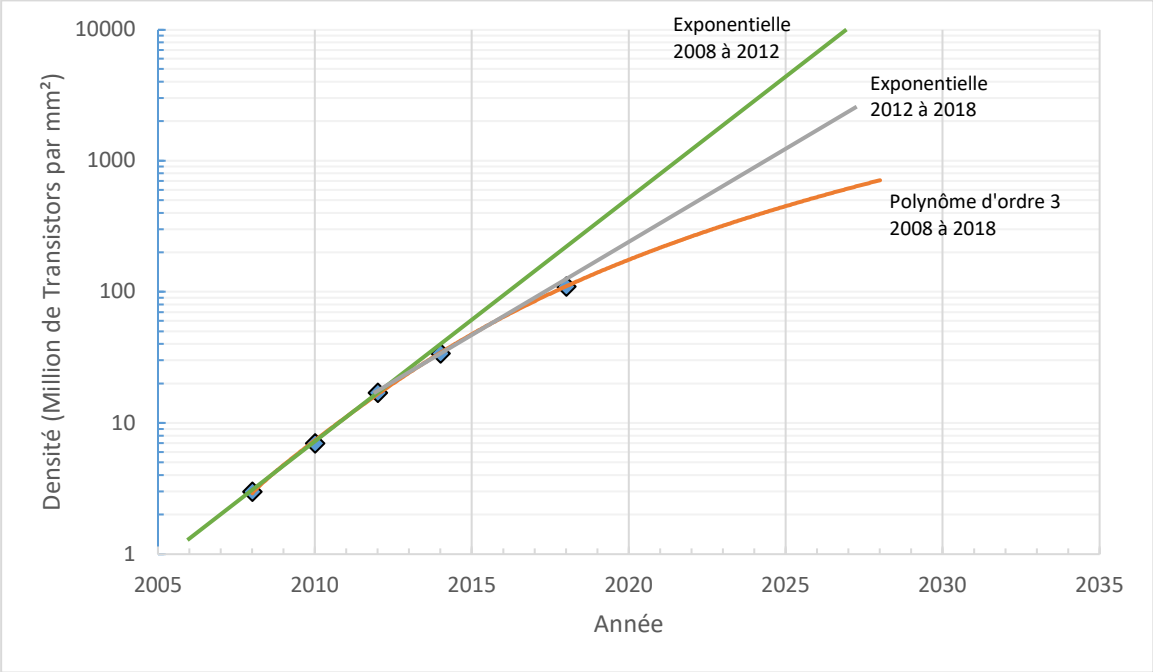


Figure 2 : Vers la fin de la loi de Moore (Source : Quora.com)

Cette conjecture proposée par Moore en 1965[1] , qui décrit le doublement de la densité des transistors tous les deux ans, s'accompagne généralement de la loi de Dennard qui exprime la relation entre les paramètres du transistor. L'hypothèse de base étant que la densité énergétique du transistor reste constante. En d'autres termes, si on réduit la taille du transistor, nous devons également réduire sa tension d'alimentation.

Tout irait pour le mieux si l'augmentation de l'intégration s'alignait parfaitement avec la réduction de la tension d'alimentation. Alors que la beauté du CMOS venait de l'absence de consommation statique, l'arrivée de technologies toujours plus fines crée une augmentation de cette consommation qui dès 2004 avec l'apparition de nœud technologique en 90nm ne peut plus être négligée [2]. Cette consommation parasite signe la fin de la loi de Dennard et de l'époque du *happy scaling*.

La rupture de la loi de Dennard s'accompagne donc d'une augmentation de la densité de puissance. Afin d'éviter que notre circuit atteigne des densités proches d'une « centrale nucléaire », nous sommes obligés de limiter l'usage des transistors. C'est ce que nous appelons le silicium noir ou *dark silicon* en anglais. A contrario du silicium appelé blanc, le silicium noir est la partie du circuit que nous ne pouvons plus utiliser du fait de l'échauffement du circuit lié à sa consommation dynamique et statique. Ramené au niveau transistor, cela veut dire que certains transistors ne peuvent plus être utilisés pour le calcul sous peine de voir exploser l'enveloppe de puissance de notre circuit. Un des nouveaux défis est d'utiliser ces transistors pour contrôler la puissance des circuits. C'est ce que nous voyons apparaitre avec l'apparition sur nos systèmes d'ilots d'alimentation, d'empilement de transistors, d'interrupteurs de puissance et autres solutions pour réduire la consommation ou du moins essayer de la stabiliser.

Nous allons voir dans le chapitre suivant quelles sont les caractéristiques de ces deux consommations, principale cause de la perte d'efficacité énergétique.

I. 1. 1. 1 L'autonomie

Contrairement aux processeurs, il n'existe pas de loi de Moore pour les batteries. On ne peut pas réduire les ions qui transfèrent la charge dans la batterie. La seule solution est d'utiliser une chimie différente mais les temps d'évolution sont nécessairement plus longs et ne peuvent donc suivre l'évolution de performance des transistors.

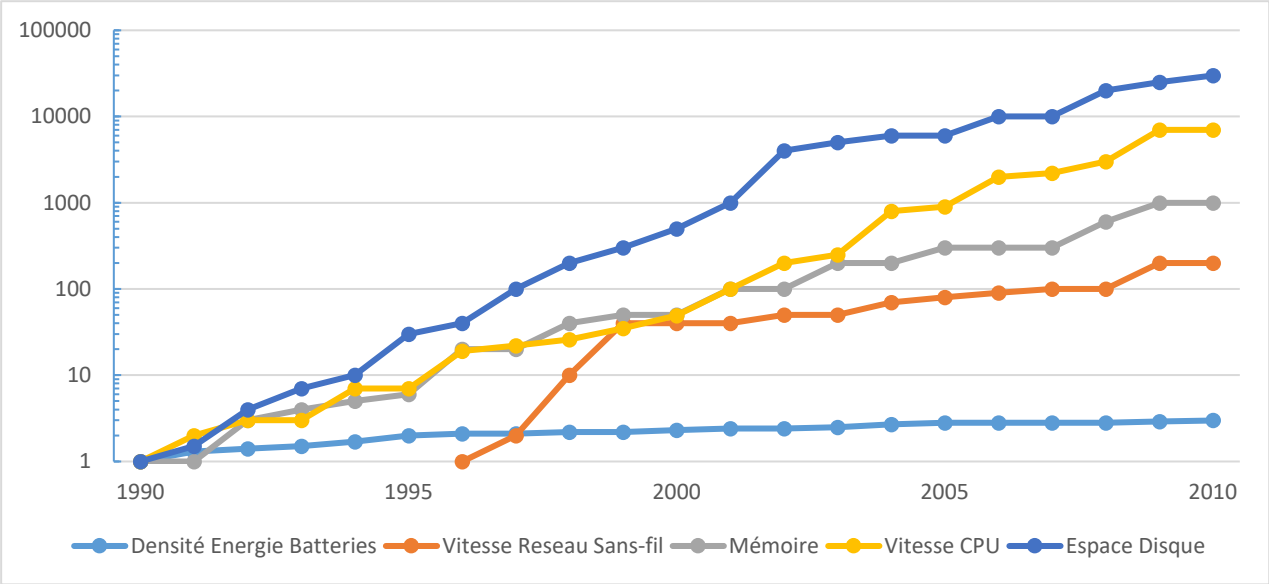


Figure 3 : Augmentation des performances des principaux blocks de circuit [3]

Malgré tout, l'évolution est bien là et les premières batteries nickel-cadmium apparues en 1899 qui permettaient de stocker entre 45 et 80 Wh/kg ont été remplacées par des batteries nickel-hydrure métallique (NiMH - *Nickel-Metal-Hydride* en anglais) en 1990 offrant entre 60

et 120Wh/kg de densité d'énergie. La technologie actuelle apparue en 1992 est la technologie lithium-ion qui a une densité comprise entre 110 et 160Wh/kg.

De nouvelles technologies sont en cours de développement : on peut citer la technologie lithium-polymère ou encore les nanotubes de silicium. Malgré tous ces efforts, il est impossible pour la batterie de suivre la courbe exponentielle de la consommation des circuits intégrés. Cela rend indispensable la recherche de nouvelles solutions pour augmenter l'efficacité énergétique des circuits et permettre le développement de systèmes basse consommation ou encore pouvant fonctionner en récupérant eux même l'énergie nécessaire à leur fonctionnement.

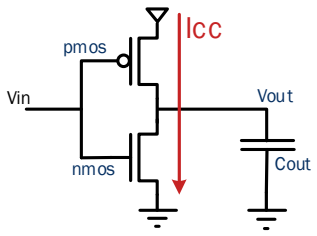
Si nous considérons que la majorité de la consommation de nos circuits nomades vient de la performance des unités de calculs et des éléments de connectivité sans-fil, nous voyons que, malgré les efforts faits depuis les années 90, l'écart entre la capacité de la batterie et la consommation des circuits tend à s'accroître. La conséquence directe est la diminution d'autonomie ressentie par les utilisateurs (Ah le bon vieux téléphone Nokia 3310 avec ses 4 jours d'autonomie !)

I. 1. 2. La consommation dynamique

La consommation dynamique représente l'énergie qui permet à nos circuits d'exécuter des calculs. C'est la consommation qui a, en premier lieu, fait l'objet de contre-mesures lors de la conception des circuits pour abaisser l'énergie consommée. Elle est la résultante de la somme de la puissance de commutation et de la puissance de court-circuit.

I. 1. 2. 1 Puissance de court-circuit

Lors du basculement de l'étage logique entre les états haut et bas, les plans de transistors N et P peuvent être conducteur simultanément pendant un court laps de temps ce qui crée un courant de court-circuit entre l'alimentation et la masse.



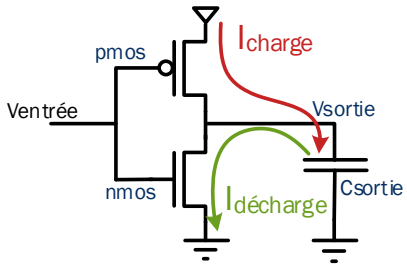
Ce courant est d'autant plus important que la transition en entrée de la porte logique est grande et peut s'exprimer en utilisant l'expression[4] :

$$P_{cc} = \frac{\beta}{12}.(V_{dd} - 2V_T)^3.\frac{\tau}{T}$$

dans cette équation V_{dd} est la tension d'alimentation, β représente le facteur de gain, Vt la tension de seuil des transistors N et P (considérée égale), τ la pente du signal et T sa période. Il existe peu de techniques de conception pour contrer cette source de consommation, l'idée est donc de la limitée en diminuant autant que faire se peut la pente à l'entrée de la porte. Une bonne pratique de conception est d'avoir un temps de transition à l'entrée de la porte au moins 2 fois inférieur à son temps de sortie. Ceci permet de limiter le ratio de la consommation dynamique autour de 10% de la consommation dynamique et donc de la faire passer au second ordre [5].

I. 1. 2. 2 La puissance de commutation

La puissance de commutation vient du mécanisme de charge et décharge des nœuds internes des circuits par la porte CMOS. L'énergie est consommée par la charge de la capacité de sortie par le transistor P. Cette charge est ensuite évacuée par le transistor N dans la masse. La charge accumulée par la capacité de sortie peut s'écrire sous la forme $C.V_{OUT}$ et par conséquent, l'énergie nécessaire fournit par l'alimentation $C.V^2$.



En intégrant cette énergie sur le temps, nous obtenons l'équation de la puissance dynamique [6] :

$$P_{dyn} = \alpha.f.C_{total}.V_{dd}^2$$

dans cette équation V_{dd} est la tension d'alimentation, α représente l'activité du circuit, C_{total} la capacité totale à charger vue par les portes du circuits et F est la fréquence de fonctionnement.

La capacité totale est la somme de la capacité des grilles d'entrées des portes chargées par notre transistor avec la capacité des fils de connections entre les transistors. Ces capacités parasites sont directement liées au choix de la technologie et est donc un paramètre sur lequel le concepteur a peu de solution à apporter. Nous nous attacherons donc au couple de paramètre fréquence/tension. Le principal constat est que cette puissance dépend linéairement de la fréquence et évolue de façon quadratique avec la tension. Nous privilégierons donc les techniques de réduction de la consommation qui s'applique à la tension car ce seront les plus efficaces pour réduire la consommation.

I. 1. 2. 1 Les Effets Secondaire

L'augmentation de la consommation et de la densité de puissance dans nos circuits électroniques génèrent plusieurs effets parasites. Parmi eux, deux ont retenu notre attention durant les travaux de thèse : la fiabilité et la compatibilité électromagnétique.

Le premier effet de la densité de puissance se voit sur la fiabilité des circuits. Le ratio du courant dynamique ramené à la surface du transistor tend à augmenter les effets du vieillissement [7]. Ce phénomène, même s'il n'est pas directement lié à l'efficacité énergétique peut être pris en compte grâce aux techniques et solutions proposées.

Nous retrouvons la même problématique sur les émissions électromagnétiques de nos circuits. Nous pouvons considérer que les émissions électromagnétiques sont générées au premier ordre par la commutation des portes logiques donc à la charge et à la décharge de la capacité. Un lien fort existe donc entre la compatibilité électromagnétique et la consommation de nos circuits ; aussi une technique qui pourrait réduire la consommation dynamique aurait mécaniquement un impact sur la compatibilité électromagnétique du circuit. L'inverse étant également vrai, nous pouvons ainsi combiner des techniques de contrôle du spectre avec des solutions de réduction de la consommation.

I. 1. 3. La consommation statique

Apparue avec le nœud technologique 65nm, la consommation statique est devenue une composante essentielle de la consommation de nos circuits intégrés actuels. Comme la consommation dynamique, la consommation statique est la résultante de la somme de huit contributions [8]

- I_{STH} : le courant de conduction sous le seuil
- I_{DIBL} : le courant dû à l'abaissement de la barrière de potentiel par le drain
- I_{GIDL} : le courant de fuite du drain induit par la grille
- I_R : le courant de fuite de la jonction p-n du drain polarisé en inverse
- I_{OX} : le courant tunnel à travers l'oxyde de grille
- I_{HCI} : le courant de grille dû à l'injection de porteurs chauds
- I_{PT} : le courant de perforation,
- I_{CS} : le courant de surface du canal dû à un effet de canal étroit.

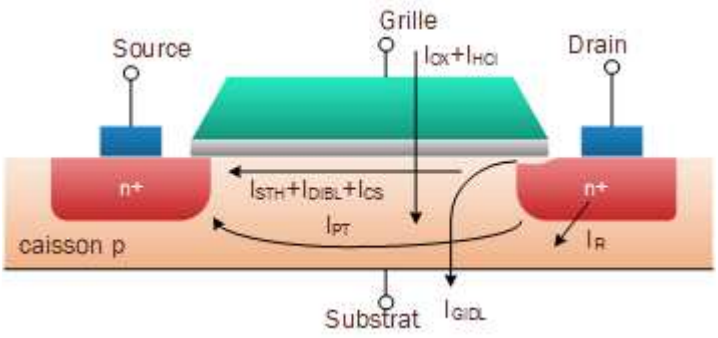


Figure 4 : Contributions des courant pour la consommation statique

Les courants de fuite qui nous intéressent sont les courants de fuite sous le seuil. Pour les autres courants, ils sont difficilement contrôlables au niveau conception et dépendent donc directement de la technologie cible choisie pour le circuit.

Le courant de fuite que nous appellerons courant statique se manifeste quand le circuit est dans un état stable. C'est un courant parasite qui traduit la consommation d'énergie nécessaire au maintien d'une donnée dans une mémoire par exemple, traduisant ainsi le fait qu'un transistor ne peut jamais être considéré comme un interrupteur parfait.

I. 1. 4. Les profils de consommation

Alors que les premiers circuits intégrés étaient essentiellement des microprocesseurs, de nouveaux besoins sont apparus et avec eux de nouveaux circuits. Il est ainsi apparu une hétérogénéité entre circuits et, en particulier, sur les modes de consommation. Des 2003, l'ITRS définit 3 modes de fonctionnement afin de classer les circuits [9].

Les 3 modes proposés sont les suivants :

- HP (*High Performance*): Haute performance. La vitesse de calcul de notre circuit devra être maximisée afin d'obtenir la plus haute performance. Les contraintes en consommation seront donc faibles voire inexistantes.
- LOP (*Low Operation Power*) : Operations à basses énergies. Les circuits LOP doivent avoir une performance raisonnable tout en essayant de limiter la consommation des opérations de calcul.
- LSTP (*Low STandby Power*) : Mode veille prédominant. Nous avons ici le monde de l'internet des objets avec des circuits le plus souvent en veille et disposant d'une puissance de calcul faible.

Ces trois modes vont nous permettent d'appliquer des solutions différentes en fonction du positionnement du produit sur le marché. Il semble en effet inutile d'appliquer une coûteuse solution de réduction de la consommation dynamique dans le cadre d'un produit HP et inversement, la réduction consommation dynamique d'un produit LSTP ne sera probablement pas la priorité.

I. 2. Quelques axes pour une meilleure gestion de l'énergie

I. 2. 1. Les premiers pas

Chronologiquement, la première source de consommation nécessitant des contre-mesures était la consommation dynamique dans les processeurs et les unités centrales. Le gain de performance venant quasi essentiellement de l'augmentation de fréquence, il s'accompagne linéairement d'une augmentation de la puissance que les boitiers n'arrivaient plus à évacuer.

Comme la consommation liée à l'horloge représente une part importante de la consommation des circuits (environ 30%) [10], il était donc naturel de réduire les évènements d'horloge au strict nécessaire. C'est ainsi que la technique de du vol d'horloge (*clock gating*) est apparue.

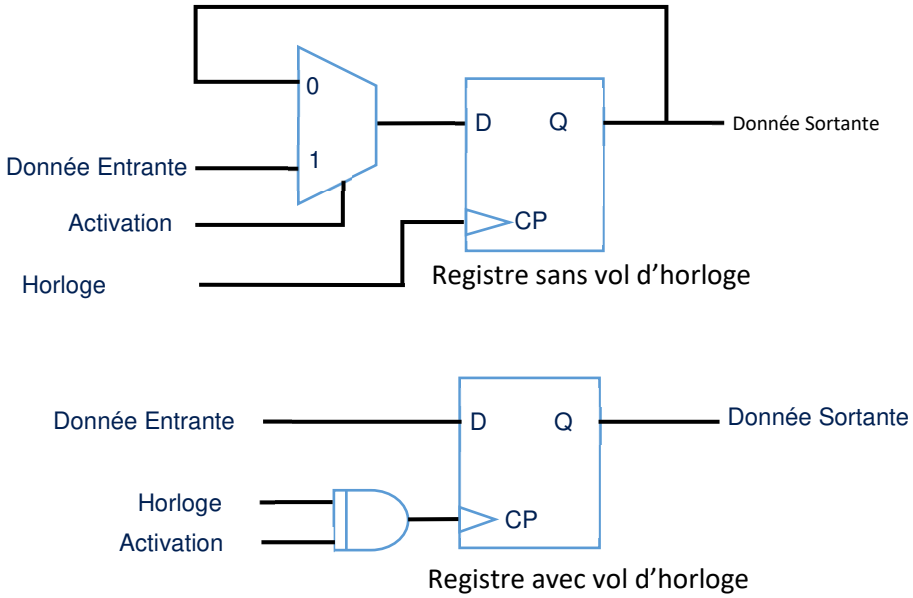


Figure 5 : Technique du "vol d'horloge"

Cette solution simple ne demande pas le développement de nouvelle cellule. L'idée est de réduire le nombre de transitions inutiles du signal d'horloge sur les registres afin de transmettre les données uniquement lorsque cela s'avère nécessaire. Bien que l'implémentation d'une telle fonction semble simple, elle est aujourd'hui essentiellement réalisée par les outils de CAO. Le concepteur décrit sa fonction sans le vol d'horloge et la fonctionnalité est ajoutée lors de la synthèse. L'outil de CAO est ainsi capable de reconnaitre une structure avec un multiplexeur de recirculation et de le remplacer par une cellule de vol d'horloge.

I. 2. 2. Diviser pour mieux régner

Vers les années 2000, l'apparition de systèmes hétérogènes alimentés sur batterie a mis en exergue pour les architectes de circuits une nouvelle problématique. Alors que les centres de calcul et les PC s'appuyaient alors sur un processeur central qui était soit allumé, soit éteint, l'architecture hétérogène s'appuie sur plusieurs unités de calcul spécialisées dans une tâche. On retrouve ainsi des accélérateurs audio, vidéo, graphique, etc. Ainsi, afin de réduire la consommation et augmenter l'autonomie des systèmes sur batterie, il devient important de contrôler l'extinction et l'allumage de ces accélérateurs. En effet, nous n'avons pas besoin d'un accélérateur graphique lors de l'écoute de notre chanson favorite. Cette notion de division des tâches et de contrôle requiert donc de nouvelles solutions afin de gérer l'énergie consommée par notre circuit. Deux choix s'offrent à nous : les îlots d'alimentation et les îlots de tension.

I. 2. 2. 1 Les îlots d'alimentation

Cette première solution est probablement la plus naturelle. Il s'agit de séparer les alimentations de chaque fonctionnalité et de les alimenter uniquement lorsque le besoin existe. Par analogie, on essaye d'allumer la lumière dans une pièce où il y a des personnes et nous utilisons un interrupteur pour cela.

La multiplication de régulateurs de tension sur le circuit imprimé est une solution qui semble simple mais qui se révèle économiquement non viable du fait de la multiplication des composants sur le circuit imprimé. Nous devons donc trouver un moyen d'embarquer ces interrupteurs directement dans le circuit et donc de créer des ilots de tension.

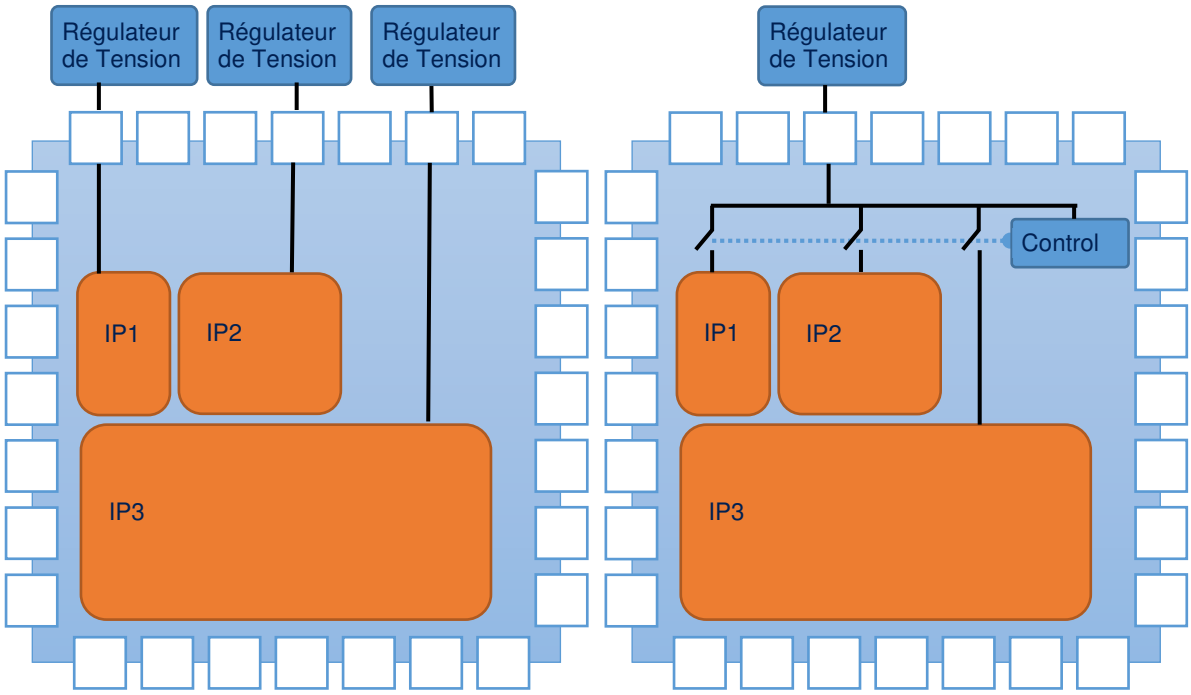


Figure 6 : Structure d'ilots d'alimentation

Comme on peut le voir sur la Figure 6, l'intérêt principal de ce mode de commande vient du fait que la logique de commande peut être directement intégrée au circuit, réduisant ainsi les besoins de composants externes. Il est intéressant de noter que ce circuit de commande devra être toujours alimenté sous peine de se retrouver dans l'impossibilité de réveiller le circuit.

I. 2. 2. 2 Les îlots de tensions

La séparation des fonctionnalités en îlots ouvre la voie à une autre technique de réduction de la consommation. Sous l'hypothèse que la performance du circuit est directement liée à sa tension d'alimentation, nous avons dans le cas d'un circuit hétérogène une tension d'alimentation suffisamment forte pour permettre au bloc IP le plus demandeur de performance de fonctionner. Cette tension étant distribuée sur tout le circuit, elle alimente aussi le bloc IP qui demande le moins de performance. Ce décalage entre la tension minimale requise et la tension appliquée engendre des pertes de puissance, donc de la consommation inutile. Afin de réduire ces pertes, nous pouvons alimenter nos différents blocs IP avec différents régulateurs. Comme pour les îlots d'alimentation, ajouter un régulateur a un coût qui doit être pris en compte par l'architecte afin de décider si le régulateur doit être intégré ou pas et si le surcoût lié au régulateur est acceptable.

Les niveaux de tensions de sortie de ces blocs IP seront donc différents rendant nécessaire le développement de convertisseurs de tension pour permettre la communication entre blocs IP. Ces convertisseurs deviennent ainsi obligatoires pour supporter cette technique.

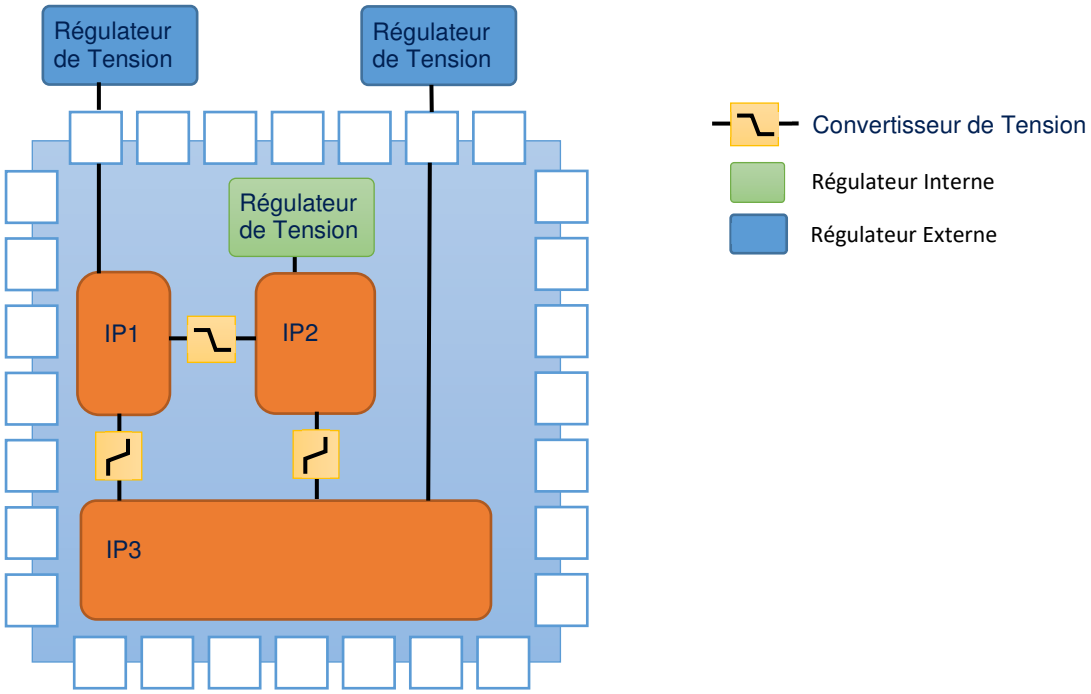


Figure 7 : Structure d'ilot de tension

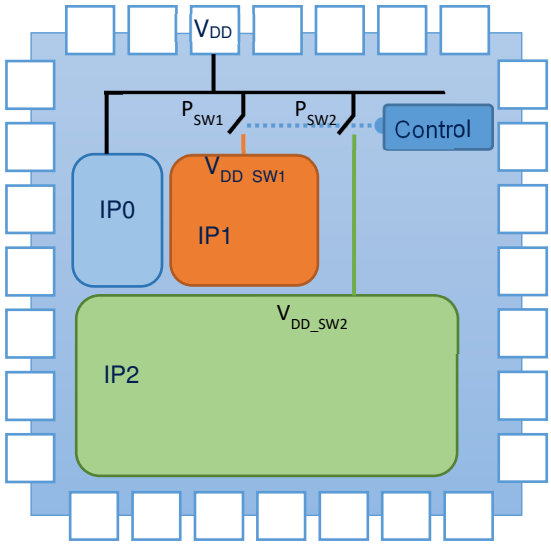
I. 2. 2. 3 Le support de la CAO

Lors de la conception d'un circuit, le concepteur utilise généralement un langage de description matériel de haut niveau (RTL ou C). Ces langages nés dans les années 1980 permettent de décrire le comportement du circuit. Malheureusement, ils n'ont pas été conçus pour prendre en compte les problématiques d'alimentations de nos circuits sur lesquels sont basées les solutions en îlots. Il est par exemple impossible de décrire un interrupteur ou un convertisseur de tension dans ces langages.

Les premiers essais d'intégration de ces solutions s'appuyaient donc sur des jeux de scripts utilisés par les outils CAO pour insérer les convertisseurs et/ou les interrupteurs aux bons endroits. Ces solutions ont servi de point de départ pour développer un nouveau moyen de description des structures d'alimentation de nos circuits.

Différentes initiatives ont vu le jours : CPF (*Common Power Format*), PFI (*Power Forward Initiative*) pour finalement converger vers le standard IEEE-1801 et son format maintenant bien connu l'UPF (*Unified Power Format*).

Ce fichier vient donc en complément du fichier RTL pour décrire la structure d'alimentation du circuit. Dès lors, les outils CAO auront besoin du couple RTL+UPF afin d'implémenter correctement le circuit.




```
set_scope TOP
create_power_domain PD_TOP -include_scope
create_power_domain PD_IP1 -element {IP1}
create_power_domain PD_IP2 -element {IP2}
create_supply_port  VDD -domain PD_TOP
create_supply_net   VDD -domain PD_TOP
connect_supply_net  VDD -ports VDD

create_power_switch SW1 -domain PD_IP1 \
          -output_supply_port {VDD_SW1} \
          -input_supply_port  {VDD} \
          -control_port       {Psw1}

create_power_switch SW3 -domain PD_IP2 \
          -output_supply_port {VDD_SW2} \
          -input_supply_port  {VDD} \
          -control_port       {Psw2}
```

Figure 8 : Exemple de fichier UPF

I. 2. 3. Le support du procédé de fabrication

I. 2. 3. 1 Evolution avec les nouveaux procédés de fabrication

La loi de Moore reste une loi d'intégration et toutes les avancées pour réduire les dimensions des procédés de fabrication ont été dictées par ce besoin d'intégration. Les

recherches pour proposer des technologies CMOS toujours plus fines cherchent à répondre à la question suivante : « comment intégrer plus de transistors dans une même surface de silicium ? ».

En analysant l'impact de cette réduction sur la consommation, nous pouvons constater deux phénomènes opposés.

Coté consommation dynamique, la réduction des procédés de fabrication a plutôt un effet positif. En effet, la réduction des dimensions s'accompagne généralement d'une diminution des tensions d'alimentation (même faible) et aussi d'une réduction des capacités de grille. Nous aurons donc pour une même fréquence donnée un circuit qui consommera moins sur une technologie plus avancée.

Cote consommation statique, les problématiques sont un peu différentes. Comme vu dans le chapitre I. 1. 3. , la réduction des dimensions s'accompagne cette foi d'une augmentation des courants de fuites donc de la consommation statique.

La connaissance et le choix du procédé de fabrication sont donc primordiaux lors de la définition de l'architecture d'un système-sur-puce car les techniques de réduction mises en œuvre dépendront fortement des caractéristiques du procédé choisi.

I. 2. 3. 2 L'émergence du FDSOI

À la suite de la fin du *happy scaling* et de la naissance du *More than Moore*, de nouvelles technologies sont apparues. Si nous ne parlerons pas des technologies exotiques dans cet ouvrage, nous allons plus particulièrement nous intéresser à l'émergence d'une d'entre elle à savoir le FDSOI.

L'idée du SOI est d'intercaler un isolant entre le substrat massif et le caisson qui servira à fabriquer le transistor.

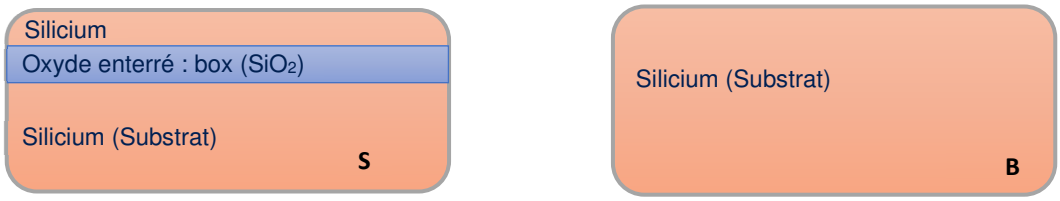


Figure 9 : Vue du substrat des technologies BULK(B) et SOI (S)

Ceci permet un meilleur contrôle du canal donc un gain en performance. Coté consommation, la réduction des capacités de jonction amène aussi un gain. Enfin, la possibilité d'enlever la prise substrat permet d'augmenter l'intégration.

Cette dernière possibilité est généralement peu utilisée. On lui préfèrera la polarisation du substrat sous l'isolant. Cette polarisation du substrat, permet de contrôler électriquement les performances du transistor. On peut donc ainsi en fonction de la tension appliquée obtenir un transistor rapide et consommant ou l'inverse.

Cette dernière possibilité ouvre ainsi la voie de solutions plus efficaces, à la compensation post-fabrication des variations de procédés de fabrication comme nous le verrons dans le chapitre III

I. 2. 4. Les solutions post fabrication

I. 2. 4. 1 Des solutions de tri

Le tri des circuits, en fonction de leur performance, est utilisé par les fabricants de processeurs pour PC. Le même circuit peut ainsi être vendu à des prix différents en fonction de sa performance intrinsèque en sortie de fabrication. Aujourd'hui, dans les architectures multicœur, le tri s'effectue plus généralement sur le nombre de cœurs disponibles. Par exemple, le circuit initial comprendra toujours 8-cœur mais pourra être vendu pour un circuit 6-cœurs si 2-cœurs sont défaillants à la suite des imperfections du procédé de fabrication [11].

Coté ASIC, la problématique est un peu différente car la performance de l'ASIC est taillée pour l'application finale. Le vendeur ne peut donc pas vendre des circuits avec des performances dégradées et est donc plutôt à la recherche de solutions qui lui permettent de récupérer les circuits qui seraient en dehors des spécifications initiales à la suite de variations mal contrôlées des procédés de fabrication.

I. 2. 4. 2 Des solutions de compensation

Les solutions de compensation sont donc une extension de la solution de tri qui permettent de rattraper les performances ou la consommation si le circuit est en dehors des spécifications. Sans cette solution, l'architecte doit prendre en compte la plage complète de variations des procédés pour définir les spécifications de son circuit. Si cette technique est appliquée, il peut utiliser une plage de variations plus petite sachant que les déviations par rapport à cette plage pourront être rattrapées. La réduction de la marge coté architecture amène une augmentation possible de la fréquence et une réduction de la consommation ou en résumé une meilleure efficacité énergétique.

Les paramètres disponibles pour compenser les variations des procédés de fabrication sont les tensions d'alimentation ou du substrat.

La principale limitation de cette solution est le coût induit par sa mise en place. En effet, sa mise en œuvre requiert l'ajout d'au moins 2 étapes supplémentaires. Il faut être capable de trier les circuits afin de connaitre leurs performances avant compensation et de les tester à nouveau une fois la compensation appliquée.

I. 2. 5. La logique asynchrone

I. 2. 5. 1 Disparition de l'horloge

Comme nous l'avons vu dans le chapitre I. 2. 1. , la contribution de l'horloge dans les circuits aujourd'hui oscille autour de 30%. Il est donc important de prendre en compte cette contrainte pour optimiser l'efficacité énergétique des circuits. Une des solutions les plus radicales à ce problème est d'éliminer la source du problème à savoir l'horloge. C'est ce que proposent de nombreux travaux de recherche issus de la communauté scientifique qui travaille sur les circuits asynchrones. Ces techniques proposent de remplacer le mécanisme de synchronisation global (l'horloge) par un mécanisme local de synchronisation de type rendez-vous (*handshake*). Contrairement donc à ce que le nom pourrait laissez penser, les circuits asynchrones sont donc parfaitement bien synchronisés.

I. 2. 5. 2 Alignement avec le calcul sur évènement

Ces techniques de synchronisation locale ouvrent des perspectives intéressantes dans la mesure où, par construction, l'activité est générée uniquement lorsque le calcul est nécessaire. L'approche synchrone qui permet à tous les registres d'être synchronisés en même temps peut être remplacée par une approche basée sur les évènements où le concepteur décrira les flots de données et de contrôle.

Du fait de l'élimination de la contribution de l'horloge, cette technique de conception apporte un gain de consommation direct. Elle ouvre aussi un chemin vers d'autres solutions grâce à l'élasticité des contraintes temporelles.

Chapitre II. Les premières techniques d'optimisation de l'efficacité énergétique

II. 1. Techniques de réduction de consommation dynamique

Réduire la consommation dynamique durant la conception n'est pas une mince affaire. En effet, la principale source de consommation vient des cycles de charge et de décharge des capacités du circuit et il est difficile de réduire ces cycles sans impacter la performance globale du circuit. Les premières solutions ont cherché à optimiser l'encodage des données afin de limiter au maximum la commutation des portes mais le gain reste limité [12].

En analysant la consommation dynamique dans les circuits, nous constatons qu'une grande partie de cette consommation vient de l'horloge (environ 30%). Cette constatation nous emmène directement vers la solution du vol d'horloge, présentée en introduction, qui reste aujourd'hui la solution principale pour réduire la consommation.

Si nous continuons à nous focaliser sur l'arbre d'horloge, nous constatons que la quasi-totalité des circuits utilise le front montant de l'horloge comme point de synchronisation. Le front descendant est donc inutilisé. L'énergie dissipée dans la décharge de la capacité globale de l'arbre d'horloge est ainsi perdue.

C'est cette problématique qui est adressée par l'utilisation des registres à double-front et que nous allons étudier dans la partie suivante.

II. 1. 1. Les registres à double-front

Dans un circuit traditionnel, la donnée se déplace de l'entrée vers la sortie sur un front d'horloge qui est généralement le front montant. L'utilisation des deux fronts est une solution pour réduire la consommation dynamique des circuits. Dans un registre double front, le front montant et le front descendant de l'horloge sont utilisés pour transférer les données de l'entrée vers la sortie. Cette solution permet de maintenir le débit d'un circuit tout en divisant sa fréquence par deux. La contrepartie de cette solution est l'utilisation d'un registre plus complexe et une gestion par les outils de CAO limitée, voire inexistante.

Plusieurs implémentations existent pour le registre à double front [13]. Nos travaux ne se focaliseront pas sur la définition d'un nouveau registre à double front mais plutôt sur la façon d'utiliser ces registres à double front dans un flot de CAO standard.

II. 1. 1. 1 Le circuit LDPC comme circuit témoin

Le circuit qui va nous servir d'exemple est un circuit de décodage de type « Vérificateur de parité à faible densité » [14] (*Low Density Parity Check* ou LDPC). Ce circuit est utilisé en code correcteur pour transmettre un message dans un canal bruité.

Selon la théorie de Shannon, l'absence de code correcteur empêche les erreurs pseudo-aléatoire générées lors de la traversée du canal bruité d'être corrigées.

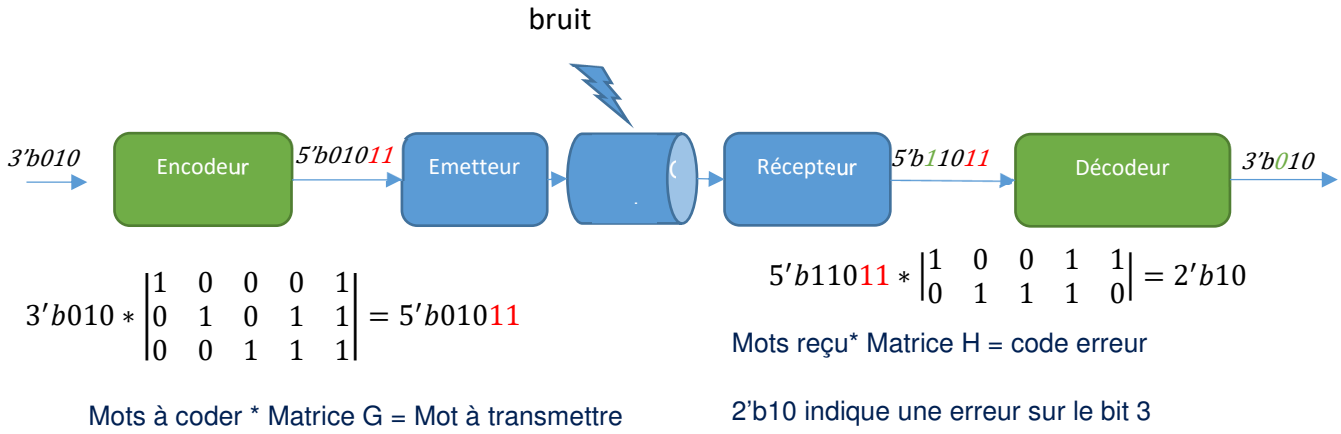


Figure 10 Principe de l'algorithme LDPC

Le principe du LDPC est donc d'introduire de la redondance lors de la mise en forme du message afin de pouvoir récupérer le signal utile dans le message après la traversée du canal. Cette redondance est générée par l'ajout d'un bit de contrôle créé à l'aide d'un algorithme imaginé par Robert Gallager. Le codage requiert l'utilisation de 2 matrices G et H. Alors que la première matrice de génération G sert à coder le signal, la matrice de parité H permet de retrouver l'information en sortie de canal.

Dans l'exemple de la Figure 10, notre message initial est codé sur 3 bits. Lors de son passage dans la matrice de génération, on retrouve un message codé sur 5 bits. Nous pouvons constater que les bits initiaux restent présents dans le message envoyé, le début de la matrice G étant la matrice identité. Dans notre exemple, le bit 3 est erroné et nous voyons bien qu'il peut être corrigé lors du calcul avec la matrice de parité H.

Ce principe a par la suite été amélioré pour arriver sur le LDPC qui nous sert d'exemple et qui nous permet de répondre aux contraintes du protocole Wifi au travers du standard IEEE 802.11n [15].

L'intérêt de ce décodeur pour notre exemple est qu'il contient des mémoires et peut être raccordé à un système sur puce via son interface ARM (AMBA). Nous allons donc nous servir de ce décodeur décrit dans la Figure 11 pour réaliser nos essais avec les registres à double front. Ce circuit de test contient toutes les problématiques que nous désirons adresser dans notre flot de conception à savoir :

1) Insertion et gain des registres double front.
2) Interface entre des IP simple front et double front
3) Utilisation des cellules de vol de cycle
4) Testabilité de la solution double front

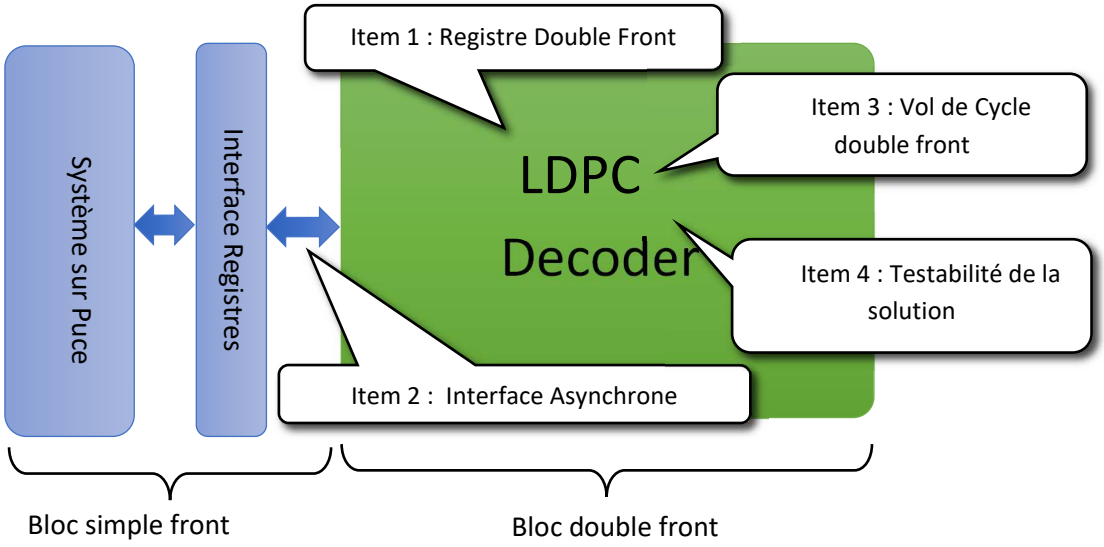


Figure 11 : LDPC Schéma bloc

II. 1. 1. 2 Analyse du Registre à double front

Au premier abord, il est facile de revendiquer un gain de 15% de la consommation dynamique. En effet, si nous estimons que la consommation de l'arbre d'horloge est de l'ordre de 30% de la consommation totale et si l'apport du double front amène la division par deux de la fréquence d'horloge, il est facile de décréter que le gain d'une solution double front est de 15% sur la consommation dynamique.

La réalité est malheureusement un peu différente. Par sa structure, la charge vue par l'arbre d'horloge sur le port d'entrée d'un registre double front est supérieur. Ce chiffre est de 20% dans le registre proposé dans ces travaux.

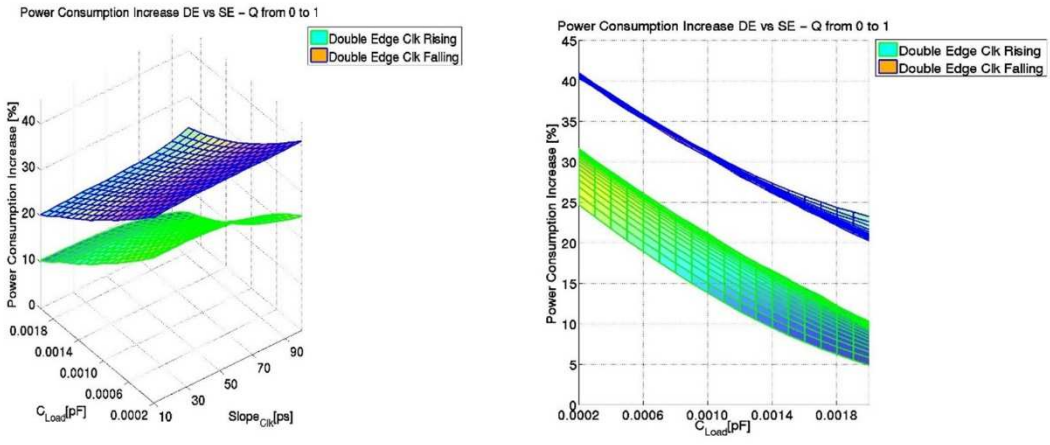


Figure 12 : Comparaison de la consommation d'un registre simple et double front

Comme on le voit sur la Figure 12, la consommation intrinsèque peut augmenter jusqu'à 30%. Dès lors, il est important de réserver ce type de solution pour des circuits ayant une consommation dynamique majoritairement lié à la fréquence d'horloge.

II. 1. 1. 3 Interface entre le simple et le double front

Dès lors qu'une solution innovante est proposée, il faut à un moment ou à un autre se poser la question de l'intégration avec l'existant. La solution double front n'échappe pas à la règle d'autant plus, comme nous l'avons vu précédemment, qu'elle ne peut pas être appliquée de façon globale. Les systèmes sur puce sont aujourd'hui composés d'un assemblage de blocs de propriété intellectuelle (IP). Parmi tous ces blocs, on peut imaginer qu'on utilisera une technique de conception standard basée sur front simple et que donc on devra communiquer avec un bloc utilisant une technique de double front.

Ainsi, nous devons réfléchir à une possibilité de passer du monde simple front au monde double front et vice-versa. L'avantage avec la solution double front est que cette synchronisation est naturelle car le front montant est commun aux deux blocs. Nous avons juste besoin d'utiliser un bloc de synchronisation standard basé sur un enchainement de 2 registres ou plus. Nous pourrons éventuellement perdre une demi-période d'horloge dans le cas où le signal de synchronisation est généré sur le front descendant mais ce délai n'est jamais critique dans le cas d'une interface asynchrone entre deux blocs.

II. 1. 1. 4 Le vol d'horloge dans le cas de double front.

Comme présenté en introduction, la technique du vol d'horloge est la technique la plus ancienne et probablement la plus efficace pour réduire la consommation dynamique. Il est donc primordial d'offrir au concepteur désirant utiliser les solutions de double front une solution équivalente. Le problème comme nous le voyons sur la Figure 13 est que l'utilisation de la technique standard peut induire un résultat erroné lors du vol de cycle.

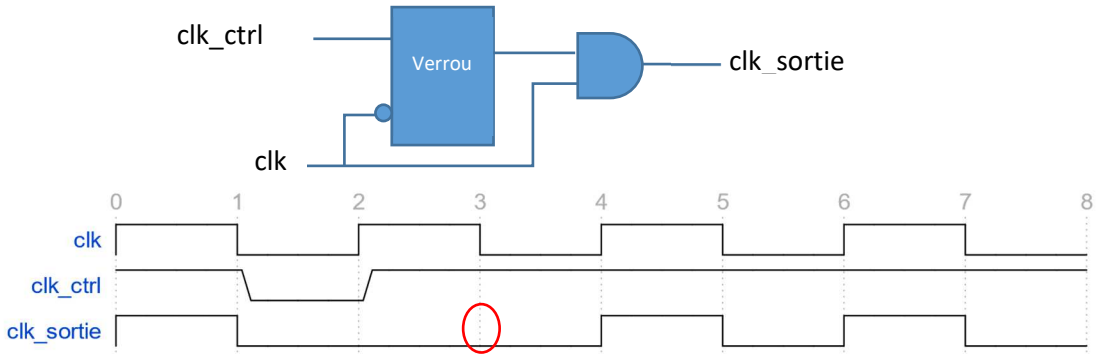


Figure 13 : Vol de cycle en logique double-front avec une cellule simple-front

Ici, le concepteur désire enlever le front 2 de l'horloge et uniquement le front 2. Il coupe donc le contrôle de l'horloge entre les cycles 1 et 2. Comme la cellule standard travaille uniquement sur front montant, il faut attendre le cycle 4 pour voir l'horloge réapparaitre. Le problème apparait pour le front 3 de l'horloge qui est un front actif en logique double-front mais que nous avons aussi volé sur l'horloge de sortie. Nous avons donc un comportement non attendu.

Notre solution s'appuie sur l'observation suivante : en logique double front, l'horloge et l'horloge inversée ont exactement la même fonction. Ainsi, la polarité de l'horloge perd un peu de son sens car le front montant a exactement la même fonction que le front descendant. Dans la logique double front, seul l'instant d'occurrence d'un front est important et nous pouvons imaginer inverser l'horloge après le vol de front.

Dans une solution simple front, il est important de garder l'alignement des fronts montants avant et après le vol de cycle. Cette contrainte tombe dans le cas d'une solution double front et l'inversion d'horloge n'est plus un problème. Cette simple constatation, nous a permis d'imaginer la cellule présentée dans la Figure 14 qui a fait l'objet d'un dépôt de brevet [16]

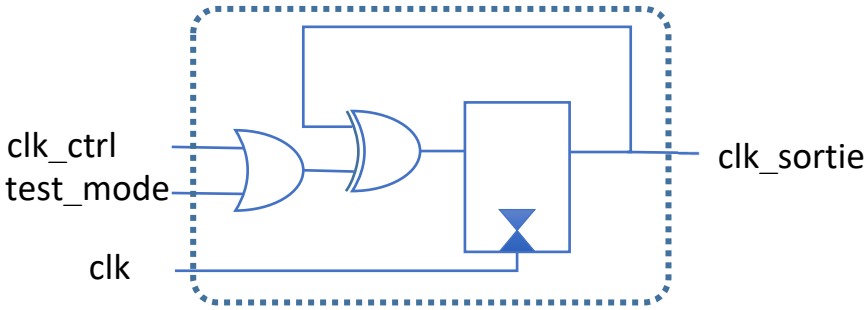
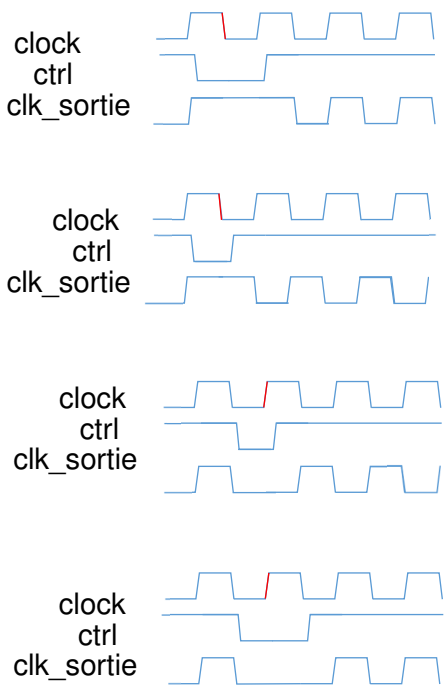


Figure 14 : Cellule de vol d'horloge double front

Cette cellule doit nous permettre de couvrir toutes les différentes possibilités de vol de cycle en fonction de l'arrivée du signal de contrôle de la cellule mais aussi de son extinction. Les 4 cas possibles sont décrits dans la Figure 15 dans laquelle nous pouvons vérifier que la séquence des cycles est bien respectée après le vol du cycle (en rouge)



Génération Control	Extinction Control	Polarité de l'horloge après le vol de cycle
Front Montant	Front Montant	Identique
Front Montant	Front Descendant	Inverse
Front Descendant	Front Montant	Inverse
Front Descendant	Front Descendant	Identique

Figure 15 : Chronogramme des vols de cycle en logique double-front

La pin `test_mode` (cf. Figure 14) est importante car elle permet d'« ouvrir » la cellule de vol d'horloge lors de la génération des *patterns* de test. Il est en effet primordial que la cellule ne s'active pas lorsque le *pattern* de test circule dans le registre à décalage constitué des chaines de scan. Le vol d'un cycle durant la phase de décalage entrainerait une détection d'erreur de la part de l'outil de test et donc le rejet d'un circuit potentiellement correct. Outre le contrôle de la cellule de vol de cycle, la testabilité de la solution double-front mérite d'être regardée de plus près et c'est ce que nous nous proposons de faire dans la partie suivante.

II. 1. 1. 5 La Testabilité en conception double front.

Si nous souhaitons amener la solution de conception double front a un niveau de qualité industrielle, il est crucial de penser à la testabilité de la solution et plus particulièrement à celle d'un registre à double front. Une solution de test fonctionnelle ou d'autotest semble intuitive mais l'industrie utilise maintenant une solution de génération de *pattern* automatique (ATPG) qui utilise les registres fonctionnels du circuit comme un « gigantissime » registre à décalage. Ce registre à décalage sert à contrôler les nœuds internes du circuit afin de vérifier les éventuels collages à 1 ou à 0.

Les outils d'ATPG actuel permettent de travailler sur front montant ou front descendant mais ils sont incapables d'analyser une structure de type registre qui réagit à la fois sur les fronts montants et descendants. Ils ne peuvent donc pas comprendre nos registres double-front et seront donc incapables de générer un jeu de *patterns* pour un circuit conçu en logique double front. Ne souhaitant pas développer un nouvel outil d'ATPG, nous nous sommes intéressés à la structure du registre double front afin de comprendre s'il serait possible de le rendre « analysable » par notre outil d'ATPG standard.

Si nous rentrons à l'intérieur d'un registre double front, nous pouvons constater que quel que soit son architecture interne, nous pourrons toujours le représenter par un empilement de deux registres simple front suivi d'un multiplexeur.

.

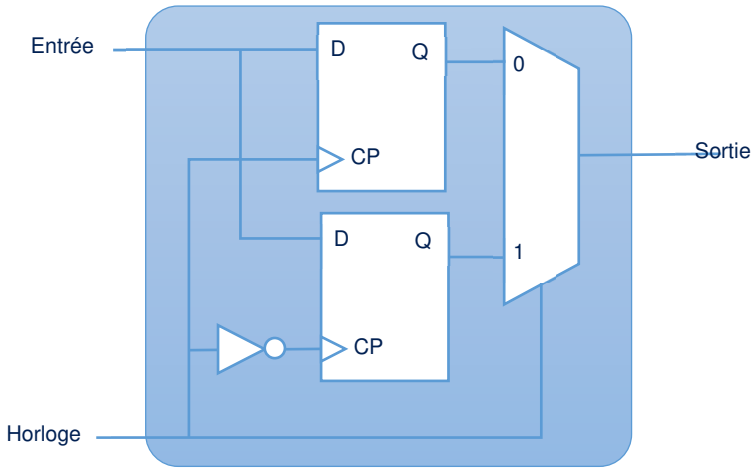


Figure 16 : Vue logique d'un registre double front sans testabilité

Nous avons maintenant deux registres simple front qui peuvent être analysés par l'outil d'ATPG. Nous pouvons maintenant appliquer la même technique que pour la testabilité des registres simple front en ajoutant un multiplexeur de testabilité qui nous permettra de créer le registre à décalage. Finalement, en connectant astucieusement les registres comme représenté sur la Figure 17, nous obtenons un registre qui s'insèrera correctement dans une chaine de scan et sera compris par l'outil d'ATPG. Il ne restera plus qu'à indiquer à l'outil que cette chaine comprend des registres sur front montant et d'autres sur front descendant. Cette topologie existe dans les circuits simple front donc est parfaitement supportée par les outils standard d'ATPG

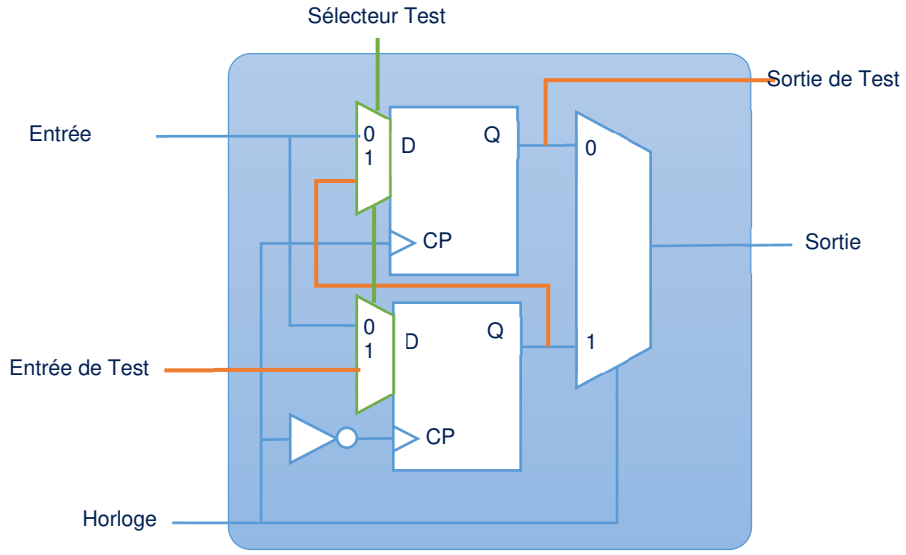


Figure 17 : Vue logique d'un registre double front avec testabilité.

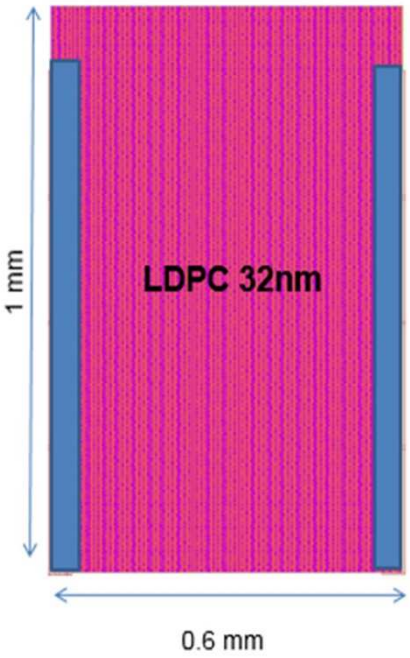
Cette structure a fait l'objet d'un dépôt de brevet [17] et nous permet d'avoir une solution complète et cohérente pour la testabilité des circuits double front.

II. 1. 1. 6 Conclusion sur le flot de conception double front.

Bien que plusieurs verrous aient été levés lors de ces travaux, il reste encore plusieurs travaux à réaliser pour arriver à une solution complète. Le lien entre les circuits simple front - double front et l'intégration de mémoires qui sont par nature simple front, constitue la prochaine problématique à résoudre.

Il serait aussi vraiment intéressant de se pencher sur les effets temporels dans des circuits à fréquence d'horloge élevée. Dans ces circuits, les problématiques de rapport cyclique et de métastabilité sont des sujets importants et une étude de ces phénomènes sur les registres double-front confirmeraient ou infirmeraient sûrement la pertinence de cette solution pour ce type de circuit qui reste aujourd'hui la cible principale de cette technique.

Finalement, lors de nos travaux sur le circuit témoin LDPC, nous avons pu démontrer un gain d'approximativement 15% par rapport à une implémentation standard en technologie 32nm. Cette comparaison a été réalisée en simulation sur une base de données placée-routée, la mise en place d'un silicium n'ayant pu être faite.



	Conso Totale	Conso Horloge	Rapport Horloge/Total
PC nple Front	31,9mW	9,27mW	29%
PC uble Font	27,27mW	4,63mW	17%
ximum Gain	**14,5%**	50%	

Figure 18 : Implémentation circuit témoin LDPC en logique double front

II. 2. Techniques de réduction de consommation statique

Les travaux sur la réduction de la consommation statique se sont appuyés sur la notion de diviser pour mieux régner. Contrairement au circuit de type processeur qui sont par leur nature monolithique, les systèmes sur puces embarquent plusieurs blocs de propriété intellectuelle spécialisés dans une fonctionnalité. Par exemple, nous pouvons avoir des circuits dédiés au son, à la vidéo ou autres. Ainsi nous pouvons facilement comparer par analogie notre circuit à un appartement avec différentes pièces de vie aux fonctionnalités différentes. Au même titre qu'il semble naturel d'éteindre la lumière lorsque l'on sort d'une pièce, on peut imaginer éteindre le bloc lorsque la fonction n'est plus requise. C'est ce mécanisme d'interrupteur que nous allons étudier dans cette partie.

Pour certaines fonctions, il n'est pas gênant de repartir de l'état initial lorsqu'il faut les rallumer. Cependant, d'autres fonctions nécessitent de garder les réglages lors de leur mise en veille afin d'assurer un réveil plus rapide et un fonctionnement rapide après le réveil. On peut imaginer ici la fonction audio qui ne doit pas avoir un nouveau calibrage à chaque rallumage sous peine de rendre l'expérience utilisateur pénible.

Cette contrainte de réveil nous permet de définir 1 mode actif et 3 modes d'extinction différents en fonction du temps de réveil demandée par l'application. Le temps de réveil peut être directement calculé comme le nombre de cycles d'horloge nécessaire entre la demande de réactivation du bloc et sa disponibilité dans le système

- ACTIF

  Le circuit est en activité. Les horloges sont présentes et le circuit réalise la fonction pour laquelle il a été conçu.
- STOP

  Le circuit est toujours alimenté, il n'est donc pas nécessaire de le réveiller. Des solutions de vol de cycle peuvent être mises en place afin de réduire la consommation dynamique mais il n'y a pas de gain en consommation statique dans ce mode. (Temps de réveil : <10 cycles.)
- VEILLE (ou ENDORMI)

  Dans ce mode, le périphérique est éteint mais les données stockées dans les registres sont sauvegardées. Cela permet un temps de réveil plus rapide car le système n'a pas besoin de reprogrammer les valeurs dans les registres avant d'accéder au périphérique (Temps de réveil : <100 cycles)
- ETEINT

  Dans ce mode, le périphérique est complètement éteint et aucune information n'est sauvegardée. Le réveil de ce bloc nécessitera donc une reprogrammation afin d'initialiser proprement son fonctionnement. (Temps de réveil : <10k cycles)

Nos différentes fonctions du système sur puce oscilleront entre ces 3 derniers modes d'extinction en fonction des requêtes de l'utilisateur et de la charge du système.

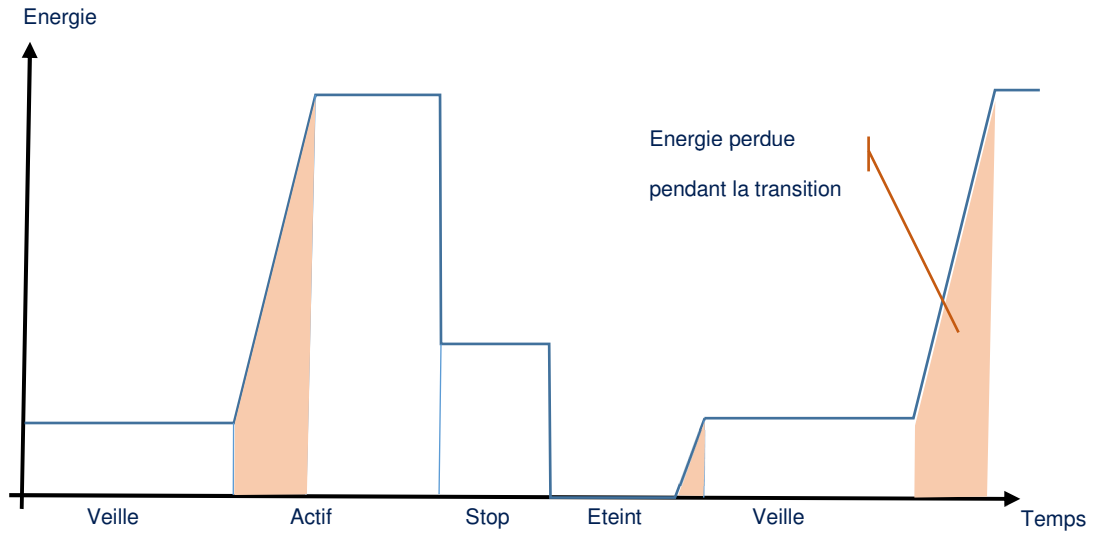


Figure 19 : Profil de consommation du système sur puce en fonction du mode

Il est intéressant de noter dans la Figure 19 que le passage d'un mode à un autre demande une certaine quantité d'énergie. Il est donc important pour le système de savoir si l'extinction d'une fonction durera suffisamment longtemps afin que le gain total soit positif. Ces

techniques d'analyse de charges ne font pas partie de nos travaux qui se focalisent sur les solutions efficaces d'extinction du bloc.

Pour supporter ces différents modes, nous avons besoin de 2 éléments. Le premier est un interrupteur qui nous permet de contrôler l'alimentation du domaine d'alimentation. Le deuxième est le registre qui permet de sauvegarder la donnée lorsque l'alimentation est éteinte. Les travaux sur la réduction de la consommation statique ont donc porté sur ces 2 éléments que nous allons décrire dans les 2 paragraphes suivants.

II. 2. 1. Commutateur d'alimentation

II. 2. 1. 1 Principe de l'interrupteur de puissance

Afin d'éteindre ou allumer un domaine d'alimentation, la technique la plus couramment utilisée est celle des interrupteurs de puissance ou MTCMOS (*Multi-Threshold CMOS*). Le principe est simple, nous utilisons un réseau de transistors que nous plaçons entre l'alimentation et le domaine alimenté. Afin de réduire la consommation statique lorsque le domaine est éteint, nous chercherons à utiliser des transistors avec une consommation statique minimale. Dans l'exemple [18], cette solution a permis de réduire d'un facteur 600 la consommation statique du périphérique.

Il existe deux positions pour le réseau d'interrupteurs de puissance. Les transistors peuvent être placés entre l'alimentation et le domaine alimenté, nous parlerons dans ce cas de type « Tête » et utiliserons un transistor PMOS. Ils peuvent être aussi placés entre la masse et le domaine alimenté et nous auront dans ce cas un type « Pied » avec utilisation d'un transistor NMOS.

Le choix peut se faire lors de la conception mais la caractéristique avantageuse du PMOS pour son courant de fuite réduit par rapport au NMOS amène généralement le concepteur à utiliser une solution de type « Tête ». Il est bien sur possible de combiner les deux mais le cout en surface devient alors trop important par rapport au gain en consommation.

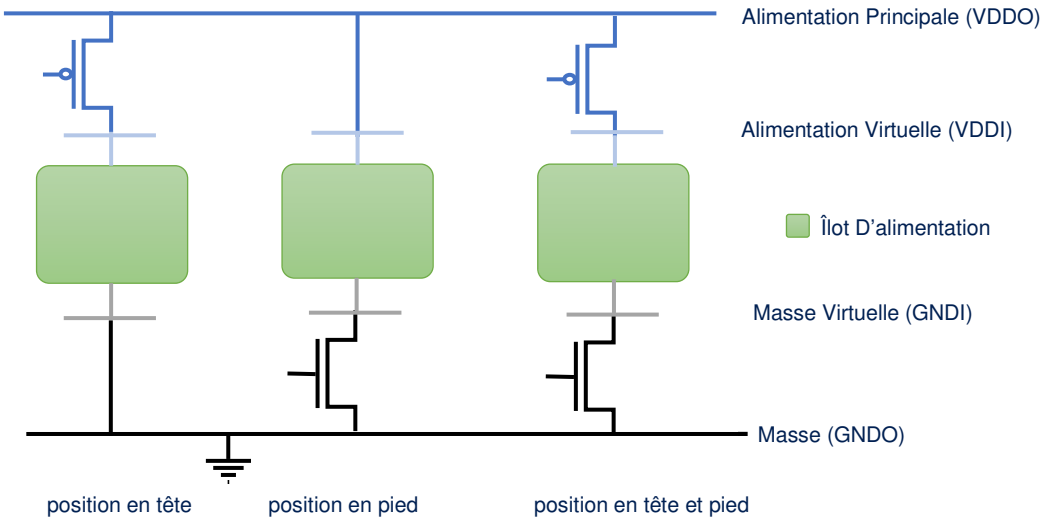


Figure 20 : Îlot d'alimentation avec ses positions d'interrupteurs

Intéressons-nous maintenant aux bonnes caractéristiques de ce transistor. Lorsque l'îlot est éteint, le courant de fuite doit être limité au maximum. Le bon transistor est donc celui qui a la plus grande résistance dans son état ouvert (R_{OFF}). C'est pour cela que le transistor de type P est généralement préféré au N car pour une taille identique, il possède une R_{OFF} plus élevée. D'un autre côté, quand l'îlot est actif, la résistance du transistor passant (R_{on}) induit une chute de tension aux bornes de l'îlot. Cette diminution de tension pourrait semblée intéressante dans le cas de la recherche de réduction de consommation mais n'oublions pas que dans ce cas, le circuit fonctionne et qu'il doit absolument assurer son niveau de performance. La chute de tension est donc vue comme quelque chose de pénalisant qu'il convient donc de minimiser au maximum.

Pour ce faire, il est primordial de réduire R_{on} et l'un des moyens utilisés pour réduire Ron est de connecter des transistors interrupteurs en parallèle. Ainsi, dans un réseau qui contient N interrupteur (Nsw), nous nous retrouvons avec l'équation suivante :

$$R_{totale} = \frac{R_{unitaire}}{N_{sw}}$$

Équation 1 : Resistance Totale d'un réseau d'interrupteur

Afin de dimensionner correctement notre réseau nous allons donc chercher à trouver le nombre d'interrupteurs qui fournira le meilleur ratio entre une forte résistance R_{OFF} et un faible résistance R_{on}. Une approche intéressante pour cette recherche d'optimum est présentée en [19] et considère la chute de tension induite par le R_{on} et le courant de fuite I_{OFF}. Nous pouvons ainsi définir les 2 métriques qui nous intéressent R_{ON} et R_{OFF} à l'aide des équations :

$$R_{on\_totale} = \frac{V_{drop}}{I_{on}}$$

$$R_{off\_totale} = \frac{V_{dd}}{I_{off}}$$

Équation 2 : Calcul de R_{ON} et R_{OFF}

Dans ces équations, V_{DROP} est la chute de tension vue sur l'alimentation virtuelle (VDDI) (voir Figure 20) et I_{ON}, I_{OFF} sont les courants consommés par l'îlot d'alimentation respectivement en mode ACTIF et VEILLE.

II. 2. 1. 2 Optimisation de l'interrupteur de puissance en 65nm PD-SOI.

L'arrivée de la technologie Silicium sur Isolant (SOI) partiellement désertée (PD-SOI) dans les techniques de conception d'interrupteur de puissance amène de nouvelles possibilités pour la conception d'interrupteurs de puissance. Une partie de nos travaux s'est donc orientée vers la recherche d'une solution optimale exploitant ce nouveau procédé de fabrication.

En 65nm PD-SOI, le transistor standard est de type body flottant (FB). C'est cette caractéristique qui permet les performances améliorées de la technologie silicium sur isolant par rapport au substrat massif (Bulk). Le problème est que l'effet lié au body flottant impacte négativement le courant I_{OFF} donc la performance de notre interrupteur. Pour retrouver de la performance en mode OFF, nous proposons de passer sur un transistor de type body contacté (BC) dans lequel le body du transistor est relié au substrat par un contact. Dans ce cas, le transistor aura des performances alignées sur un transistor utilisé en technologie sur substrat massif avec un courant I_{on} certes moins élevé mais avec un courant de fuite réellement diminué donc avec un ratio I_{ON}/I_{OFF} plus intéressant comme on peut le voir sur la figure suivante.

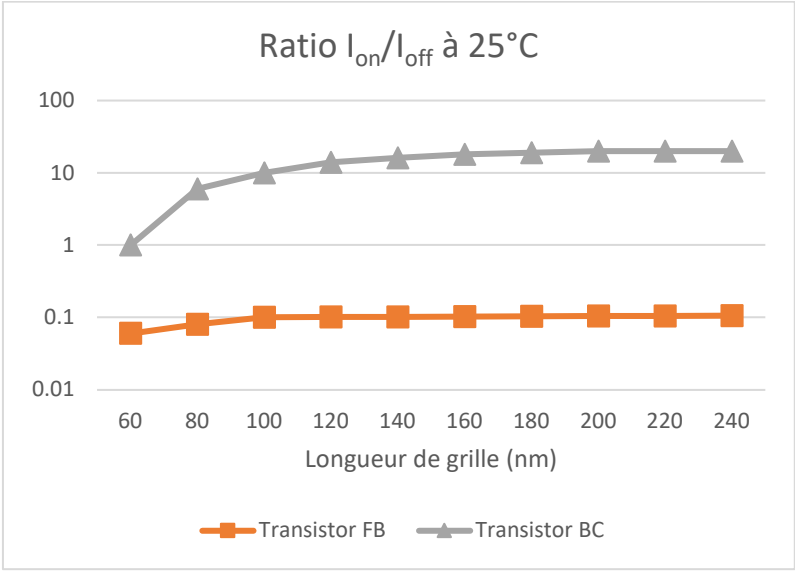


Figure 21 : Rapport I_{ON}/I_{OFF} des interrupteurs de puissance sous une tension de 1.2V

Nous avons maintenant choisi le type de transistors intéressant pour notre interrupteur et il nous reste à définir sa caractéristique à savoir sa tension de seuil pour laquelle nous proposons d'utiliser de manière astucieuse la prise de substrat (*body*). Les interrupteurs de puissance ont en général une tension de seuil qui ne peut pas être modulée par les

concepteurs et qui est choisie par les technologues. Une tension de seuil haute est souvent choisie car c'est elle qui permet d'obtenir le meilleur ration I_{ON}/I_{OFF}. Nos travaux autour de l'interrupteur ont cherché à optimiser plus ce ratio en réalisant une polarisation dynamique du *body* de l'interrupteur de puissance.

Il existe 2 chemins pour polariser le *body* du transistor. Nous pouvons appliquer une polarisation en direct (FBB ou *Forward Body Biasing*) qui nous permettra de réduire la tension de seuil V_{TH} et la résistance R_{on} du transistor. Ce double effet entrainera une augmentation du courant I_{on}. Inversement, une polarisation en inverse (RBB ou *Reverse Body Biasing*) aura pour effet de réduire le courant I_{OFF} par l'augmentation de la tension de seuil et de la résistance R_{OFF}. Il est usuel de considérer dans les technologies SOI et BULK une plage de polarisation de ±300mV. En effet, au-dessus de ces valeurs, le gain apporté par la polarisation est effacé par les effets parasites de courant inverse de jonction dans le cas du RBB et par la conduction en direct de la diode body/source pour le FBB. Le tableau suivant montre les gains obtenus avec des polarisations de 300mV.

Polarisation	FBB@300mV	RBB@300mV	
Gain du courant	I_{ON} (+9,1%)	I_{OFF} (-42%)	I_{ON}/I_{OFF} (+88%)

Tableau 1 : Gain en courant d'un interrupteur BC avec une grille de L=200nm

Idéalement, nous souhaiterions donc pouvoir polariser en direct lorsque l'îlot d'alimentation est allumé et en inverse lorsqu'il est éteint. C'est bien sur techniquement possible en ajoutant des sources de tension externes mais le prix en surface de ces sources et la complexité ajoutée pour contrôler ces sources de tension rendent cette proposition peu compatible avec les contraintes de coût de nos systèmes sur puce. Nous choisirons une approche plus pragmatique en sélectionnant uniquement le type de polarisation direct et en éliminant la source de tension.

La Figure 22 présente le principe général de la proposition qui consiste à utiliser un pseudo-pont diviseur de tension dont la valeur des résistances est contrôlée par la tension de grille. Cette polarisation est dite auto-adaptative car elle ne nécessite pas de générateur externe et la polarisation dépend de la valeur de la tension sur la grille donc de l'état de l'interrupteur.

Figure 22 : Principe général de la polarisation de l'interrupteur

La conception de ce circuit breveté [20] qui permet de polariser en direct le *body* du réseau d'interrupteurs de puissance se compose essentiellement de 2 transistors MOS utilisés pour réaliser la résistance variable. En mode OFF, le transistor P est ouvert et nous retrouvons un potentiel à VDD sur le *body* ce qui correspond au mode standard sans polarisation. Le montage maintient bien le R_{OFF} par rapport à la version standard. En mode ON, le transistor N est ouvert et la diode *Source/Body* (en gris sur la Figure 23) est polarisée en direct. Il est donc important de contrôler le courant circulant dans cette diode ce qui explique l'ajout de 2 diodes en direct entre la masse et la source du transistor N (en jaune sur la Figure 23). Ce montage nous permet de tirer vers le bas la polarisation du *body* ce qui nous permet de réduire le R_{on} de 20%. Ce montage est auto-adaptatif car plus le courant du body est élevé, plus la tension aux bornes des 2 diodes en directe est élevée et plus la différence de potentiel aux bornes de la diode *Source/Body* s'abaisse, réduisant d'autant le courant de *body*. C'est un système rebouclé et stable qui assure une polarisation de *body* en direct.

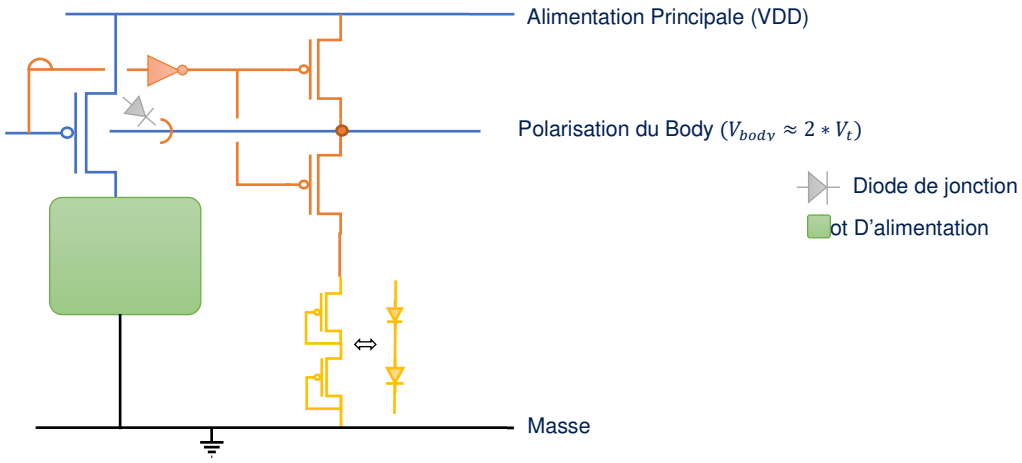


Figure 23 : Schéma du circuit auto-DTMOS

II. 2. 1. 3 Validation sur Silicium de l'interrupteur PD-SOI

Afin de vérifier la pertinence de nos choix, nous avons réalisé deux circuits de test en utilisant la technologie PD-SOI en 65nm de STMicroelectronics [21]. Pour ces circuits, nous avons réutilisé le LDPC présenté en II. 1. 1. 1 et pour lequel nous avons réalisé 2

implémentations différentes. La première est un portage en aveugle d'un LDPC réalisé sur une technologie BULK en 65nm. Dans cette implémentation, les interrupteurs de puissance deviendront des interrupteurs à corps flottant (FB) car aucune modification ne sera apportée pour ajouter le contact. Ce circuit nous servira de référence pour nos mesures. Les modifications des interrupteurs interviennent dans le deuxième circuit où les interrupteurs de puissance seront remplacés par des interrupteurs à corps contacté (BC). Nous ajouterons aussi dans cette deuxième implémentation notre solution d'auto-DTMOS. Le circuit contient donc 2 LDPC en parallèle qui peuvent être mesurés indépendamment grâce à une interface JTAG

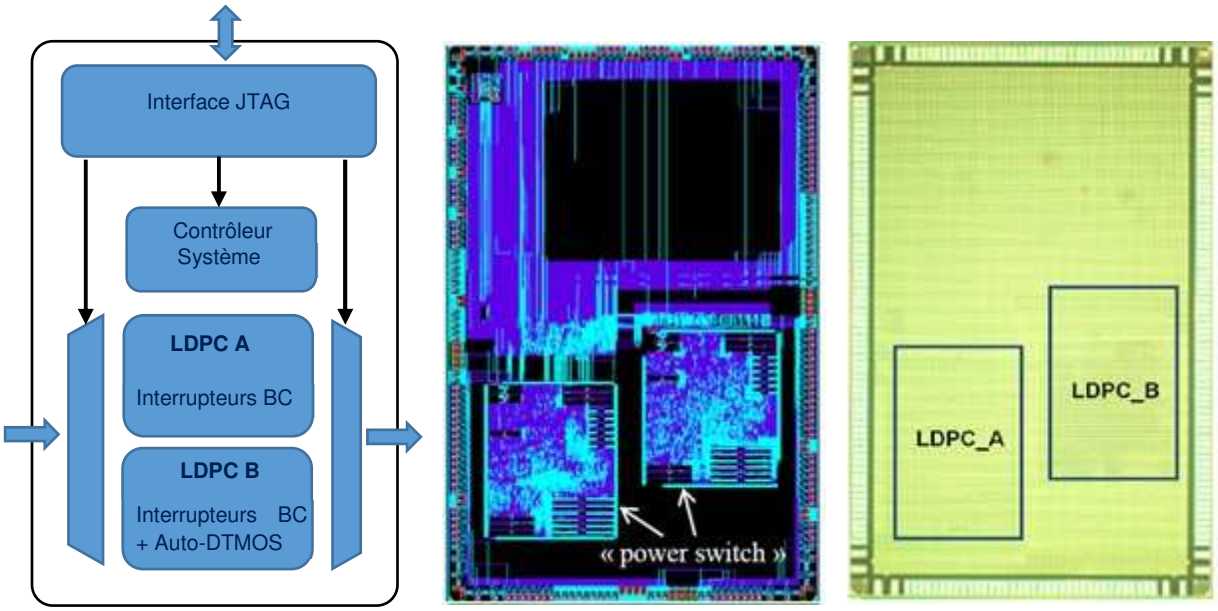


Figure 24 : Circuit LDPC en 65nm PD-SOI

Ce circuit doit nous permettre de mesurer l'avantage offert par notre solution qui devrait être une réduction du Ron en mode ACTIF et un maintien de la valeur du R_{OFF} en mode VEILLE. Ces mesures sont rassemblées dans la Figure 25 et ont été faites sur une solution de commutation contenant 3500 interrupteurs de largeur W=40µm et de longueur L=200nm respectivement.

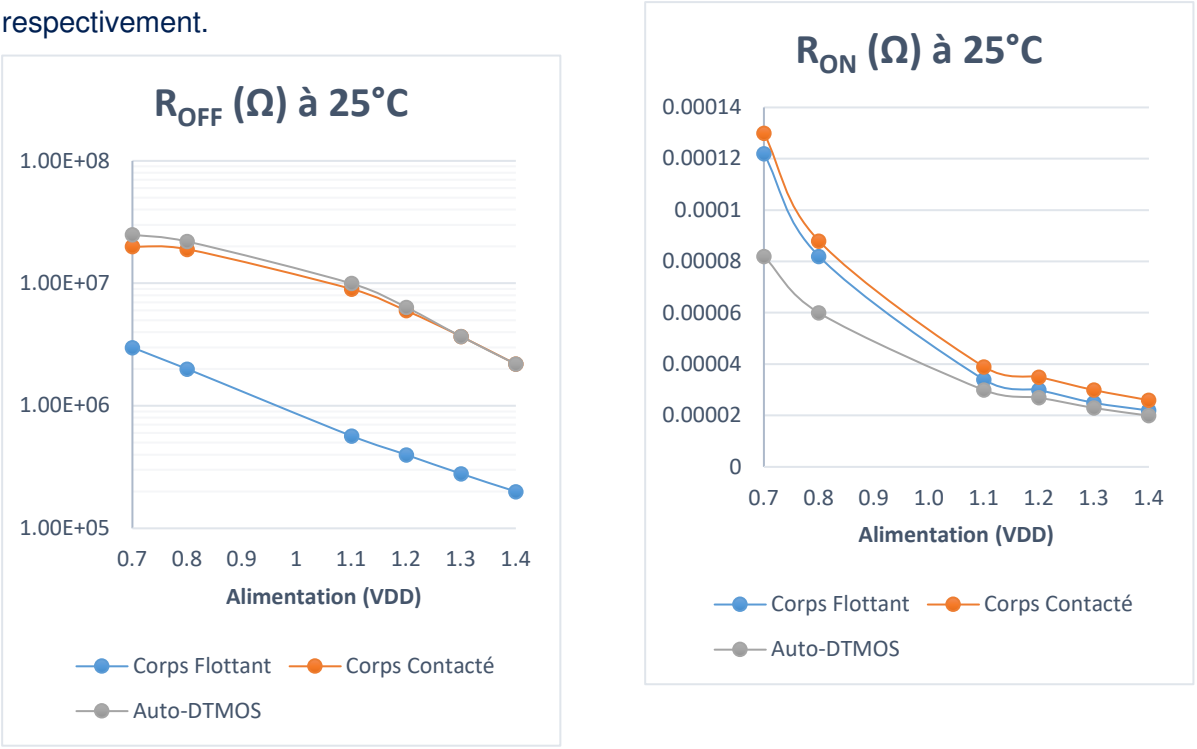


Figure 25 : Valeur de R_{ON} et R_{OFF} pour les différents types d'interrupteurs

Dans l'anneau dédié à la solution d'auto-polarisation (LDPC B), nous avons placé 4 circuits de polarisation (i.e. 1 circuit pour 1000 interrupteurs). Le R_{on} est mesuré suivant l'Équation 2 dans laquelle le courant I_{on} est fixé à 100mA et la tension V_{DROP} mesuré grâce à une pointe de test placé sur la grille. Grâce au circuit de polarisation du corps de l'auto-DTMOS, R_{ON} est 20% inférieur aux solutions SOI à corps contacté et nous gardons bien le R_{OFF} inchangé

II. 2. 1. 1 Contrôle de la Transition VEILLE vers ACTIF

Lors du changement de fonctionnement de l'îlot d'alimentation entre son mode VEILLE et son mode ACTIF, il existe un état de transition représenté en orange dans la Figure 19, durant lequel nous devons contrôler l'allumage de notre îlot. Ce contrôleur permet essentiellement de limiter l'appel de courant nécessaire pour charger l'îlot. Un allumage trop direct générerait une chute de tension sur l'alimentation principale, ce qui pourrait générer des défaillances sur d'autres périphériques connectés directement à cette alimentation en amenant la tension V_{DROP} hors des spécifications de fonctionnement. Ce pic de courant peut

atteindre des valeurs importantes comme les 53.8mA rapportés en [22] et incite à limiter le nombre de transistors contenus dans un îlot d'alimentation.

Il existe deux grandes familles de construction des îlots d'alimentation. La première, à gauche dans la Figure 26 est la structure que nous avons utilisée dans la partie précédente. Elle consiste à construire une structure en anneaux autour de l'îlot d'alimentation. Plus récemment, nous avons vu arriver une nouvelle structure dans laquelle les interrupteurs sont directement intégrés dans l'îlot (à droite dans la Figure 26). Les transistors utilisés pour construire l'interrupteur sont des transistors de type cellule standard ce qui ne permet pas à cette solution distribuée d'offrir les mêmes performances que la structure en anneau. Néanmoins, l'empreinte en surface de cette solution est beaucoup plus petite, ce qui peut la rendre intéressante pour certaines applications comme les microcontrôleurs utilisés dans l'internet des objets. Nous allons donc étudier quel moyen nous pouvons mettre en œuvre pour contrôler l'allumage de nos îlots en fonction de la structure utilisée.

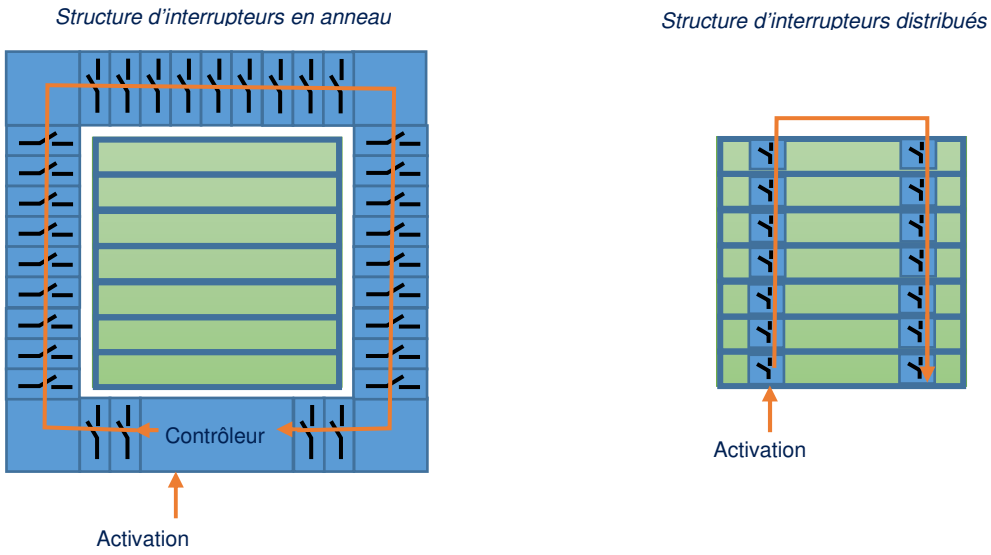


Figure 26 : Type de placement des interrupteurs de puissance

Dans le cas d'une structure périphérique, l'agencement des transistors ressemble au placement des plots d'entrée/sortie du circuit. La connexion est réalisée par aboutement et le nœud connecté aux grilles est facilement identifiable. Ainsi dans ce genre de structure, nous retrouvons généralement un contrôleur analogique qui pilote directement la grille du transistor [23]

Cette solution offre aussi l'avantage de pouvoir mesurer directement la tension sur la source de nos interrupteurs d'alimentation afin d'avoir une boucle de rétroaction. Ainsians le contrôleur programmable utilisée pour limiter le courant lors du réveil.

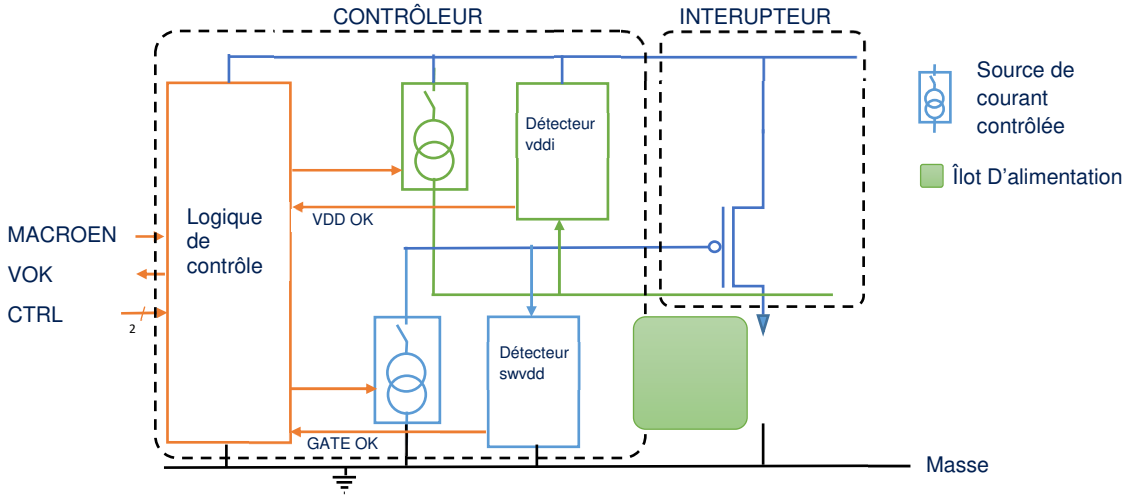


Figure 27 : Schéma du contrôleur et de l'interrupteur en anneau

Le mécanisme de gestion de la séquence de réveil agît en pré-alimentant l'îlot d'alimentation et en contrôlant la tension de grille des interrupteurs d'alimentation simultanément. Ce mécanisme est contrôlé par un registre de contrôle de deux bits autorisant un compromis entre le courant pic et le temps de réveil (broches CTRL de la figure 3). Le circuit de précharge intégré au contrôleur a été dimensionné de sorte qu'il fournisse 90 % de la charge d'énergie locale avant que l'interrupteur entre dans la région sous le seuil, les 10% restants sont injectés à travers l'interrupteur tout en gardant un contrôle sur la descente de la tension de grille grâce à une source de courant pilotés à l'aide d'un bus de contrôle sur 2 bits.

L'activation de l'îlot est initiée par un front montant sur la broche de signal MACROEN. Les détecteurs intégrés dans le contrôleur surveillent les niveaux de tension de l'alimentation commutable localement et la puissance de commutation des transistors afin d'indiquer quand les interrupteurs sont complètement fermés et donc quand l'îlot est allumé. Le signal VOK est alors positionné à l'état haut pour signaler la fin de la séquence de réveil.

Dans le cadre d'interrupteurs distribués, les transistors sont généralement placés en colonne ou en damier dans le circuit. Il n'y a plus cette connexion par aboutement et le contrôle est réalisé de façon numérique en allumant plus ou moins de transistors. La solution couramment utilisée est une structure en chaine ou le temps de propagation du signal entre les transistors permet de limiter le courant d'allumage. La principale limite de cette technique est le manque de contrôle de ce courant.

Pour répondre à ce besoin de contrôle et essayer de retrouver un comportement proche de la structure en anneau, nous avons proposé une nouvelle solution brevetée [24] alliant un

contrôle numérique et une boucle de rétroaction constitué d'un capteur analogique.

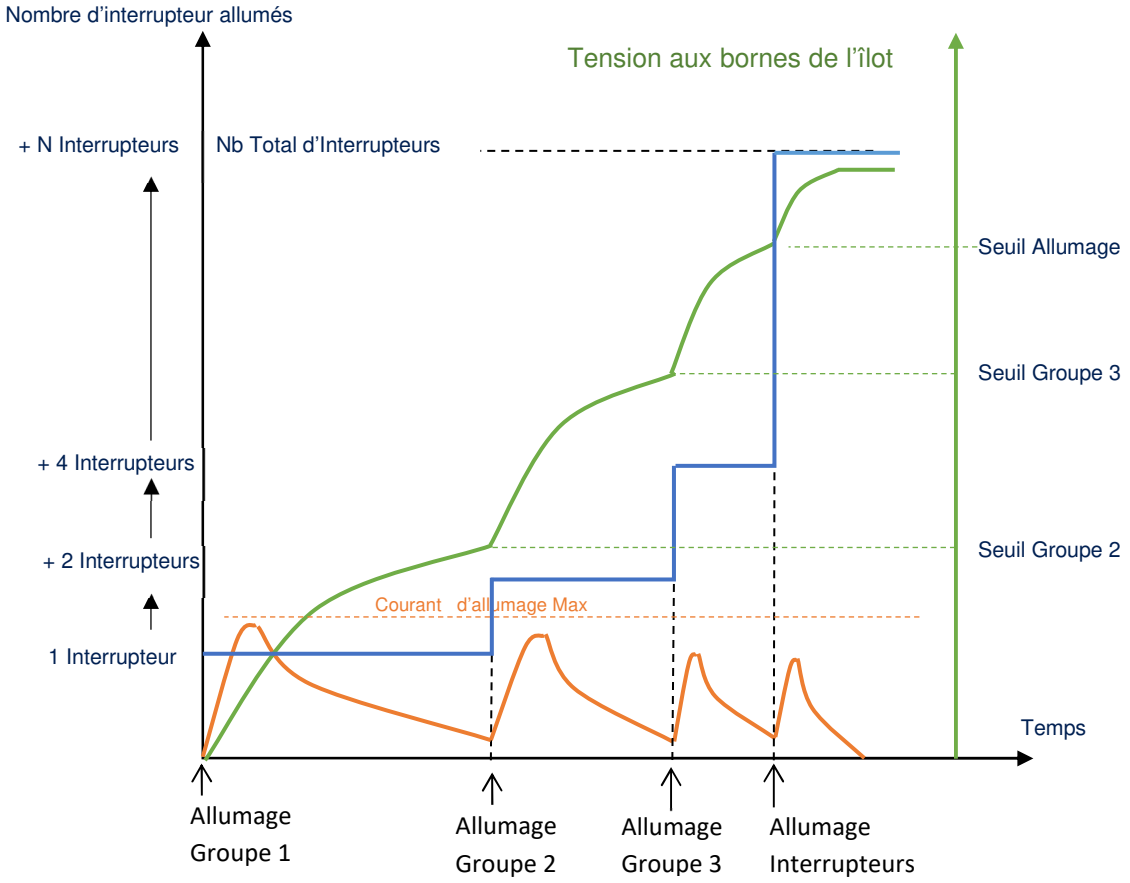


Figure 28 : Séquence d'allumage d'interrupteurs distribués

Les transistors de précharge sont construits en 3 groupes avec des ratios de largeur 1,2,4 entre les groupes. Le moment d'allumage des différents groupes est calculé par un comparateur analogique en fonction de seuil de détection. En réglant ces seuils, nous retrouvons un contrôle du courant d'allumage proche de celui de la solution en anneaux et surtout qui nous permet de rester en dessous d'un niveau de courant maximum défini par le concepteur.

II. 2. 2. Registre de rétention

II. 2. 2. 1 Le principe du registre à rétention

Le registre à rétention est le deuxième élément primordial des solutions faible consommation. On le retrouve dans les périphériques qui ont besoin de maintenir une information lorsqu'ils sont en mode VEILLE. Si nous prenons l'exemple d'un contrôleur de mémoire externe, il est important de garder les réglages de la mémoire connectée afin de pouvoir initier les transactions avec la mémoire dès que le périphérique sera réveillé et sans attendre une coûteuse reprogrammation de la configuration.

La solution classique d'un registre à rétention est l'ajout d'un verrou tampon dans lequel on vient stocker l'information en mode VEILLE. Comme on peut le voir sur la Figure 29, le tampon ajoute inévitablement de la surface supplémentaire à notre registre. Nous avons donc cherché dans ces travaux à supprimer ce verrou tampon ou plutôt à le fusionner avec l'étage esclave de notre registre.

La Figure 29 présente la bascule de rétention développée lors de ces travaux pour être utilisée dans les îlots d'alimentation commutables. C'est une paire de verrou maitre-esclave avec réutilisation du verrou esclave lors de la mise en rétention de la bascule. L'isolation est assurée par la porte de transmission RTG contrôlée par le signal SLEEP. Cette architecture requiert donc un simple signal d'entrée SLEEP afin de maintenir la donnée dans l'esclave pendant le mode VEILLE de l'îlot. Pendant cette phase de rétention, la logique 3-etats de l'esclave est court-circuitée afin de maintenir la donnée dans le point mémoire constitué des 2 inverseurs et d'isoler cet élément de mémorisation des variations sur les entrées de la bascule. Nous profitons aussi de la possibilité offerte par les technologies modernes d'avoir plusieurs tensions de seuil pour optimiser le registre. Ainsi, le registre sera majoritairement construit avec des transistors ayant une tension de seuil standard mais nous utiliserons des transistors à seuil haut pour le point mémoire de la partie esclave et de la porte de transmission RTG. Ceci nous permet de minimiser le courant de fuite lorsque le registre sera en rétention (et donc l'îlot en mode VEILLE). Finalement par rapport à un registre standard, i.e. sans possibilité de rétention, nous arrivons à maintenir la consommation dynamique (+2%) tandis que le temps de traversée du registre est dégradé seulement de 12% et l'augmentation en surface limitée à 30%

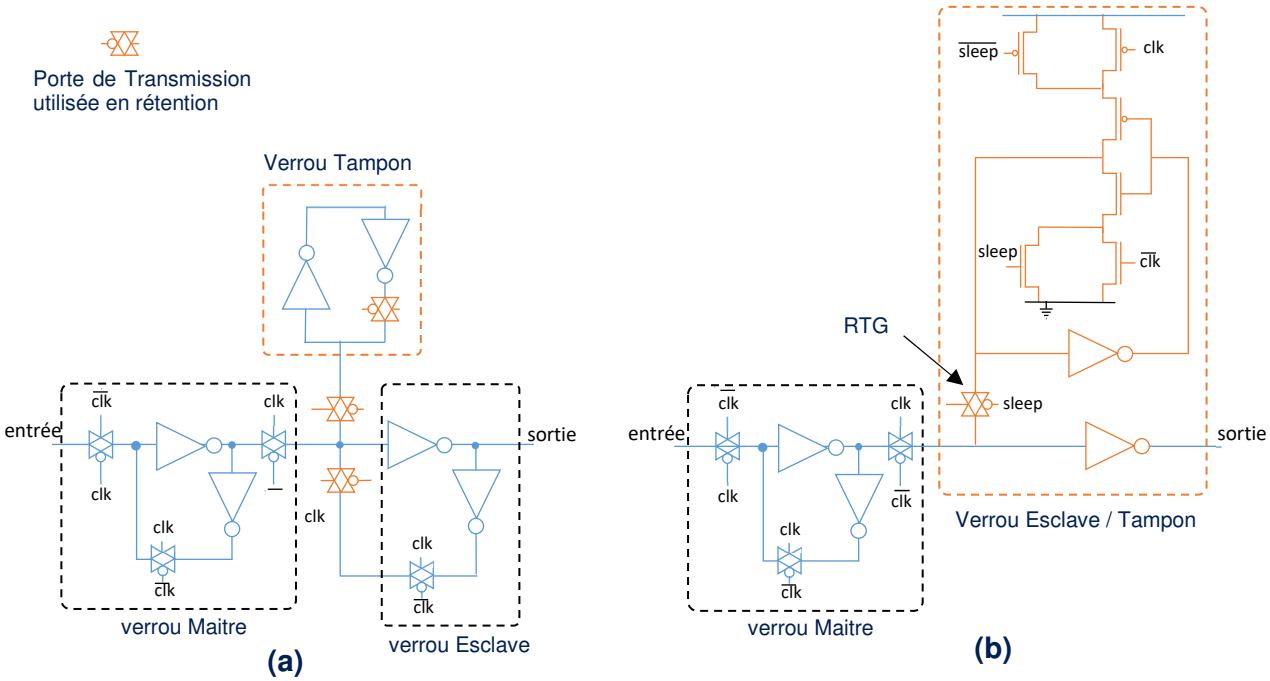


Figure 29 : Registre à rétention conventionnel (a) et optimisée (b)

Du fait de la fusion de 2 verrous, la solution propose un gain de surface de l'ordre de 30%. Néanmoins, cette optimisation ramène de nouvelles contraintes sur la polarité de l'horloge. Alors que dans un registre avec tampon, la dépendance de la rétention est uniquement sur le signal de veille, dans le registre proposé, l'horloge doit être à l'état bas pendant la veille sous peine de perdre la donnée. La raison est la mutualisation de la porte de transmission de l'horloge du verrou esclave avec celle du signal de mise en rétention du verrou tampon.

II. 2. 2. 2 Compatibilité avec le flot de conception

Dans le cas de circuits simples et travaillant uniquement sur des fronts montants, la contrainte de la polarité de l'horloge n'est pas un problème. Il suffit que le concepteur force son générateur d'horloge à l'état bas avant la mise en veille et s'assure que la remise en route de l'horloge s'effectue après la restauration des registres. Dans le cas de circuit plus complexes avec des parties travaillant sur front montant et d'autres sur front descendant, nous nous retrouvons avec le problème illustré dans le chronogramme de la Figure 30

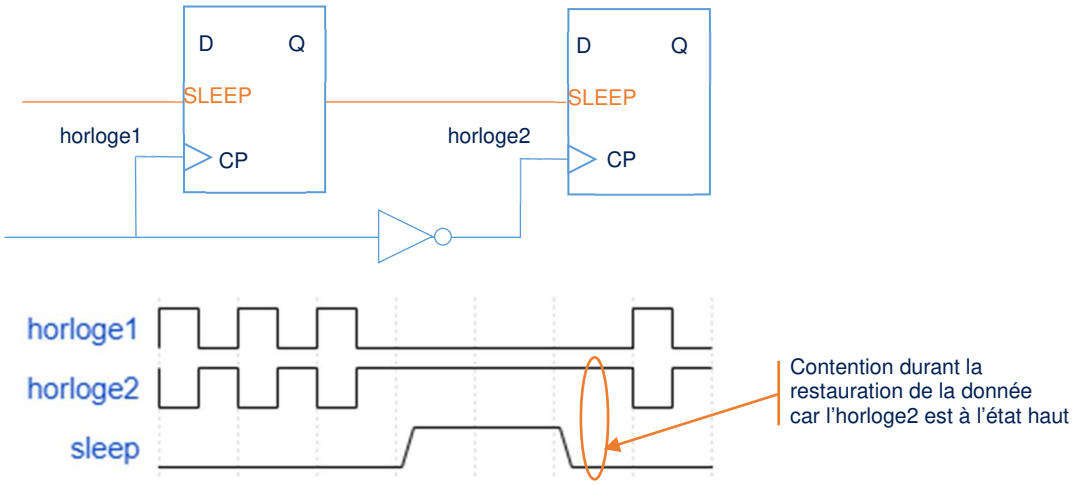


Figure 30 : Problématique de la rétention avec système travaillant sur front descendant

Un travail supplémentaire est donc nécessaire afin de rendre cette solution de rétention compatible avec la majorité des circuits. La solution que nous proposons dans le brevet [25] et présentée dans la Figure 31 reprend le principe de la cellule de vol d'horloge en l'adaptant afin de pouvoir contrôler l'horloge dans les différents cas qui posent problèmes.

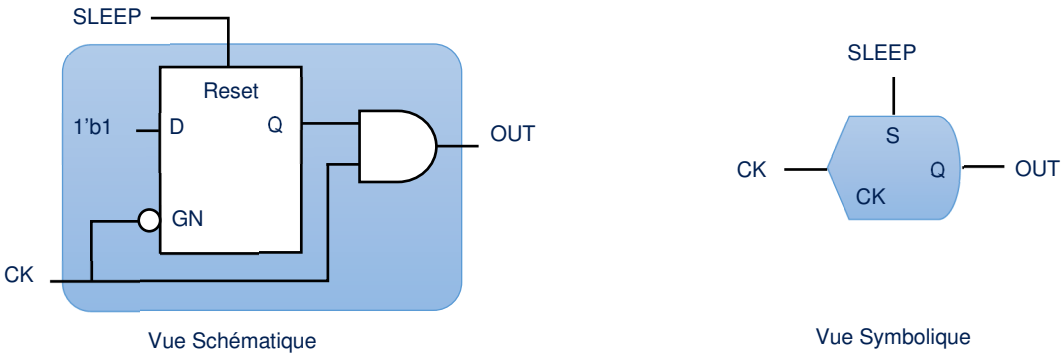


Figure 31 : Cellule de contrôle d'horloge en rétention

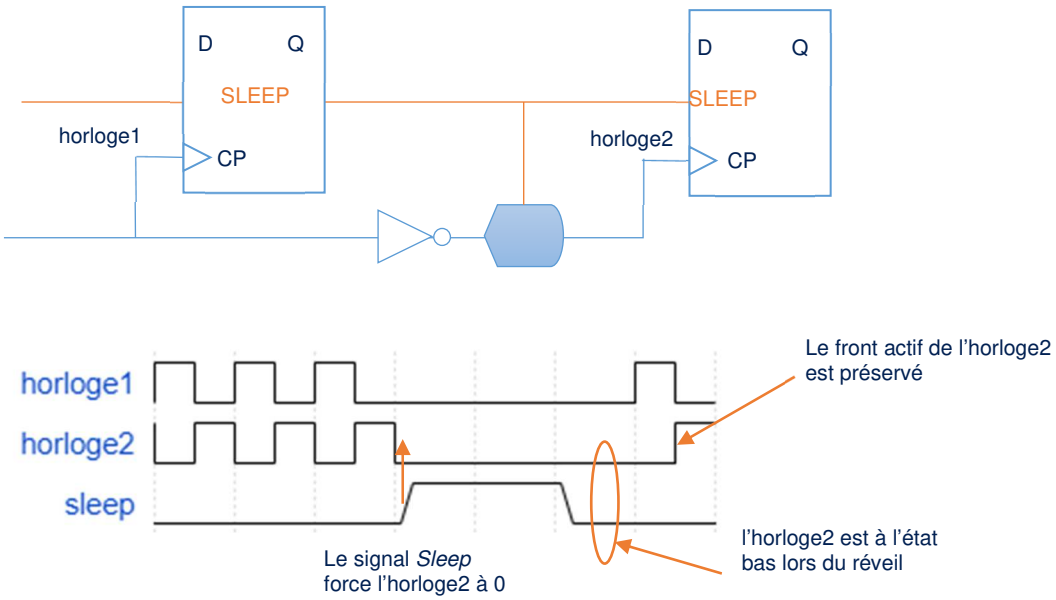


Figure 32 : Rétention sur des circuits travaillant avec des horloges inversées

En plus de régler le problème des doubles fronts, cette cellule permet de résoudre 2 autres cas de génération d'horloge rencontrés dans nos circuits synchrones.

1) **Périphérique sans contrôle d'horloge dédiée**

Nous retrouvons cette situation dans le cas de périphérique simple. L'horloge a besoin d'être coupée pour rentrer en mode de veille mais le périphérique ne dispose pas de contrôle dédié à cette horloge. L'ajout de la cellule de contrôle à l'entrée du signal d'horloge permet de résoudre ce problème en forçant un état bas sur cette horloge lors des phases d'endormissement de de réveil.

2) **Contrôle de l'horloge dans le cas d'un diviseur d'horloge interne**

Si le périphérique requiert une division d'horloge par 2, il est difficile d'assurer que l'horloge divisée est bien en position basse pendant la mise en veille. La cellule de contrôle permet d'assurer le positionnement de l'horloge divisée en position basse et assure aussi un redémarrage en phase afin de ne pas perdre la synchronisation entre les deux horloges.

II. 2. 2. 3 Testabilité de la solution de rétention

Comme pour les registres à double front, la testabilité de la solution de rétention est un élément à prendre en compte lors de l'utilisation de cette solution. *A contrario* du registre double front qui avait vu sa fonctionnalité altérée par l'utilisation du front descendant, le registre à rétention se comporte comme un registre simple front lorsque le circuit est en mode actif. Il s'agit donc d'ajouter à la suite de tests existants un nouveau stimulus qui permet de s'assurer

du bon comportement lors de la rétention de la donnée quand le circuit est en mode VEILLE. Le moyen proposé se rapproche des tests de rétention des mémoires embarquées. Dans un premier temps, nous utilisons la chaine de *scan* pour charger les registres avec un motif connu généré par l'outil d'ATPG. Ensuite, après avoir placé les registres en mode rétention, nous positionnons le circuit en mode de VEILLE en utilisant le régulateur externe ou les interrupteurs présentés dans le chapitre II. 2. 1. La dernière étape est de réveiller le circuit et de décharger les registres à travers la chaine de *scan*. La comparaison entre le motif généré et le motif reçu nous donnera le résultat du test sachant que nous attendons une égalité entre ces deux motifs pour déclarer le circuit correct.

II. 3. Le circuit de test ZEO

Afin de valider ces nouvelles solutions de réduction d'énergie, nous avons décidé de réaliser un circuit de test nommé ZEO (Zéro Energie Opération). Bien sûr, réaliser les calculs sans énergie est impossible mais l'idée était d'essayer de tendre vers cette limite en appliquant toutes les techniques basse consommation disponibles à cette période.

Ce circuit est donc le premier circuit de test conçu par les équipes STMicroelectronics qui utilise toutes les solutions de réduction de la consommation disponibles avec la technologie ST *65nm Low Power*. Le circuit peut être placé complètement en mode VEILLE et utilise donc une technique de rétention dite massive dans laquelle tous les registres sont des registres avec rétention. Cette technique est permise grâce à la réduction de surface des registres à rétention obtenue dans le chapitre II. 2. 2. Au niveau des interrupteurs, la solution choisit est une solution en anneaux. Le substrat utilisé n'étant pas un substrat type Silicium sur Isolant, nous n'avons pas pu utiliser les techniques d'optimisation de l'interrupteur présenté en II. 2. 1. 2 mais nous avons utilisé les solutions de contrôle de la transition présentées dans le chapitre II. 2. 1. 1

De plus, nous ajoutons autour du microprocesseur une alimentation dédiée pour le substrat. Les possibilités de gestion de l'énergie grâce à la polarisation du substrat entrevue lors des travaux sur l'interrupteur PD-SOI nous semble intéressante à approfondir pour les îlots d'alimentation.

II. 3. 1. Présentation du circuit

Il s'agit d'une plateforme multimédia contenant 1 cortex ARM926 et 2 accélérateurs pour le traitement audio et vidéo. On retrouve aussi un nombre de périphériques communs tel qu'un I^2C, un DMA ou une interface pour l'écran. Ce circuit avait typiquement sa place dans un téléphone portable du début des années 2000. C'est un système-sur-puce architecturé autour d'une interconnexion AMBA.

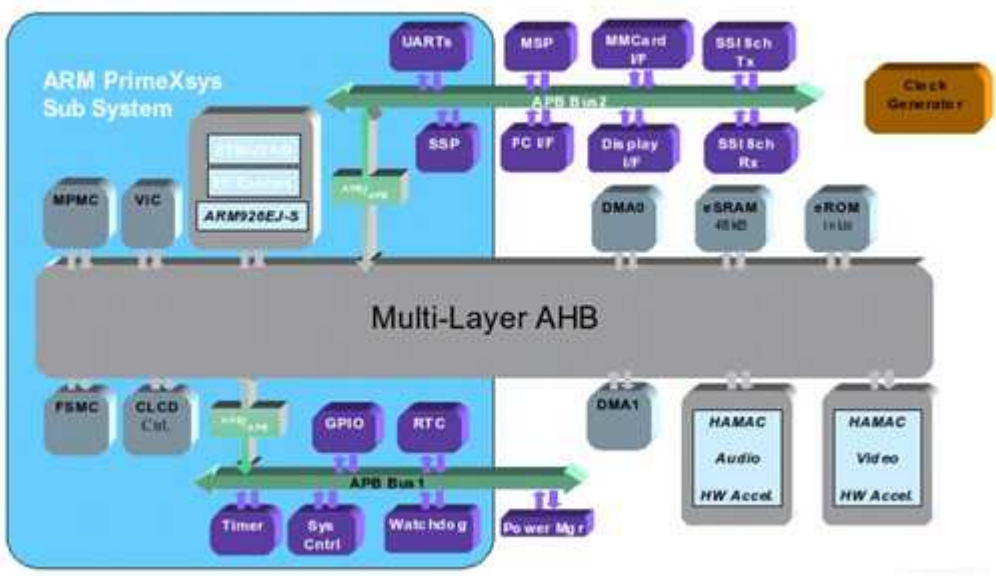


Figure 33 : Schéma Block du circuit ZEO

L'architecture des domaines d'alimentation est présentée en Figure 34. Le circuit se compose de 5 domaines d'alimentation différents

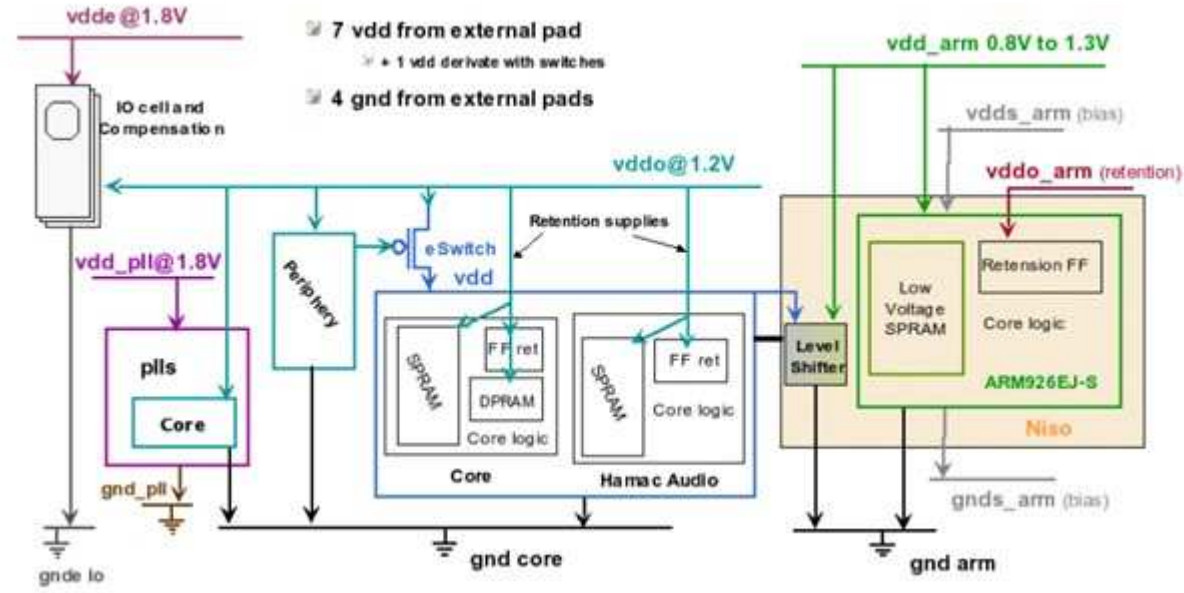


Figure 34 : Domain d'alimentation du circuit ZEO

Il y a 2 domaines haute tension pour les parties analogiques du circuit. Nos techniques de réduction de la consommation ne s'appliqueront pas à ces domaines. Les trois domaines qui nous intéresseront plus particulièrement seront :

1) Le domaine toujours alimenté

   Ce domaine contrôle le générateur d'horloge (PLL) et la partie du circuit qui pilote la mise en veille.

2) Le domaine du cœur de circuit

C'est le domaine principal de notre circuit. Il contient les accélérateurs audio et vidéo ainsi que tous les périphériques. Seul le cœur ARM est exclu de ce domaine car faisant partie du troisième domaine.

3) Le domaine à tension variable

Ce domaine permet de mesurer et d'évaluer les gains en énergie en faisant varier la tension du circuit ou sa tension de substrat. Par manque de temps, les mesures n'ont malheureusement pas pu être faite sur cette partie.

Le domaine du cœur est donc le domaine qui nous intéresse particulièrement. Il contient 94000 registres à rétention, avec une capacité totale en mémoire embarquée de 2.4Mbits et son périmètre est de 14mm.

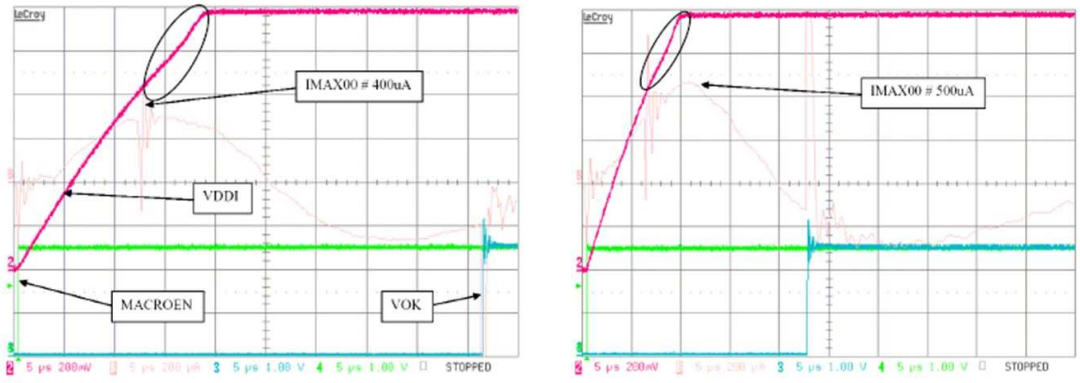
Pour contrôler la tension de ce domaine, nous utilisons une solution périphérique. Les interrupteurs utilisés sont de type « Tête » (voir Figure 20). La technologie n'étant pas PD-SOI, nous utilisons des transistors standards pour l'interrupteur et nous gardons le contrôle du réveil en analogique.

Concernant les registres à rétention, le faible impact en surface nous a permis d'appliquer une solution dite massive et 97% des registres sont devenus des registres à rétention. Après placement-routage, la pénalité en surface s'élevait à 0.336mm² pour les interrupteurs périphériques et à 0.12mm² pour l'introduction des registres à rétention. Au global, le coût en surface de la solution de réduction de la consommation est de 2.85% de la surface totale.

II. 3. 2. Mesures et conclusion du silicium ZEO

Les deux mesures qui nous ont particulièrement intéressé sont la mesure du temps de réveil avec le contrôle du courant lors de la mise sous tension et, bien sûr, le gain en consommation statique obtenu avec l'insertion des structures basse consommation.

Concernant le temps, nous obtenons bien un contrôle du courant avec un courant maximum qui est en dessous de 400µA au maximum avec le bus de contrôle de courant positionné à une valeur de 00. Cette configuration limite au maximum le courant de mise sous tension et nous donne un temps de réveil de l'ordre de 70µs. Nous mesurons aussi la possibilité d'accélération du temps de réveil grâce à ce bus de contrôle. Si nous positionnons ce bus à une valeur de 10, nous obtenons un temps de réveil de 25µs grâce à une injection de courant plus massive dans l'îlot d'alimentation. Le courant pic mesuré est alors de 500µA. Ainsi le concepteur pourra choisir une vitesse de réveil en fonction de la chute de tension supportée par l'îlot.



Bus de contrôle à 00 Bus de contrôle à 10

Figure 35 : Mesure du Temps de Réveil et du Courant Pic

Pour la mesure de consommation, nous obtenons finalement un gain de 9 grâce à l'addition des solutions d'efficacité énergétique introduite dans le circuit de test.

Îlot d'alimentation	Domaine Cœur	Domaine Variable Tension	Couronne d'E/S
Nombre de Registres	94000	19000	0
Consommation Mode VEILLE (sans rétention)	381µA	40µA	150µA
Consommation mode VEILLE (avec rétention)	43µA	3µA	150µA

Tableau 2 : Mesure de consommation des îlots

En conclusion, le circuit ZEO démontre un nouveau mode de réduction de la consommation. La rétention massive permet un arrêt/redémarrage en quelques cycles car toutes les informations du circuit restent maintenues dans les registres. Au niveau software, ce mode supplémentaire peut éviter la sauvegarde du contexte dans la mémoire permettant ainsi un réveil plus rapide.

Le circuit étant sorti de fonderie, le Figure 27 présente une photographie du circuit afin de se rendre compte de la taille du domaine cœur.

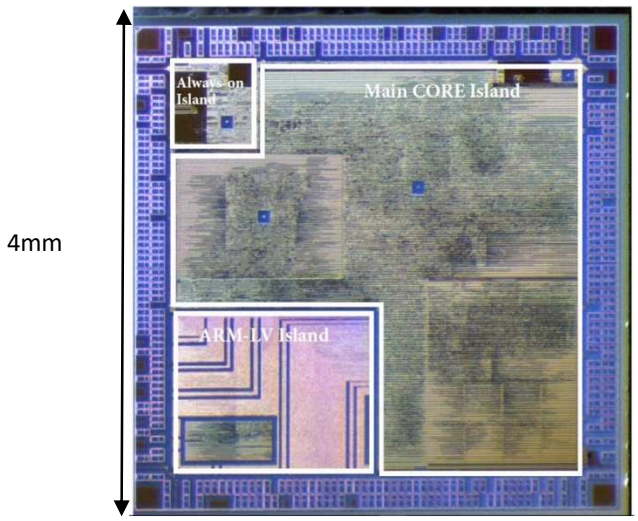


Figure 36 : Photographie du circuit ZEO

II. 4. Conclusion sur le chapitre

Dans ce chapitre, nous avons abordé les premières techniques proposées dans le but d'augmenter l'efficacité énergétique de nos circuits.

La première partie a été consacrée à la réduction de la consommation dynamique avec des travaux sur les registres à double-fronts. En utilisant le front descendant de l'horloge, nous avons montré que ce registre permet de maintenir une puissance de calcul équivalente tout en réduisant la fréquence de l'horloge - contributeur important à la consommation dynamique - d'un facteur 2. Afin de proposer une solution industrialisable, nous avons proposé une implémentation de ce registre supportant les solutions d'ATPG ainsi qu'une cellule permettant d'appliquer la technique du vol de cycles compatible avec ce registre double-fronts.

La deuxième partie s'est focalisée sur la consommation statique avec nos travaux sur les domaines d'alimentation. Nous avons dans un premier temps montrer l'importance d'optimiser l'interrupteur de puissance afin d'obtenir le meilleur compromis entre la minimisation des courants de fuite lorsque l'îlot est éteint et la maximisation du courant lorsqu'il est actif. Dans un second temps, nous avons proposé une solution de rétention qui permet au circuit d'entrer dans un mode de veille tout en conservant les informations importantes lors de son réveil. Ces solutions de rétention couplées aux interrupteurs de puissance ont été mises en œuvre sur un circuit de test précurseur des circuits que nous trouvons actuellement sur nos téléphones portables.

Chapitre III. Les solutions en lien avec le procédé de fabrication

.

III. 1. Prendre le contrôle de la tension

III. 1. 1. La gestion de la tension dynamique

Du fait de sa composante quadratique dans l'équation de la consommation dynamique, la tension des circuits fait l'objet d'une attention toute particulière et de nombreuses techniques existent pour contrôler la tension des circuits en fonction des performances requises. Nous pourrions parler ici des solutions DVFS ou AVS [26] où l'ajustement conjoint de la tension et de la fréquence permet de se rapprocher du point de fonctionnement optimal mais un nouveau levier est apparu avec l'arrivée des technologies ayant un substrat isolé.

En effet, l'arrivée de la technologie UTBB- FDSOI ouvre de nouvelles perspectives pour réduire la consommation et améliorer l'efficacité énergétique. Nous pouvons ainsi ajouter aux techniques dites traditionnelles, le contrôle de la tension de substrat qui va permettre d'avoir un levier supplémentaire sur le contrôle du transistor. Nous allons découvrir dans cette partie comment tirer bénéfice de ce nouveau levier.

III. 1. 2. Les premiers essais en technologie CMOS 45nm

Pour les essais, nous allons réutiliser le LDPC présenté dans le chapitre précédent. L'intérêt principal vient de la structure en sous ensemble du LDPC qui permet de valider plusieurs approches au sein du même circuit en appliquant différentes techniques sur les différents blocs. Le LDPC développé supporte les différents modes définis par le standard 802.11n (Wifi) et a été conçu en utilisant un flot de conception Matlab vers RTL. Comme nous pouvons le voir sur la Figure 37, chaque bloc supplémentaire allumé permet la gestion d'une trame wifi plus large

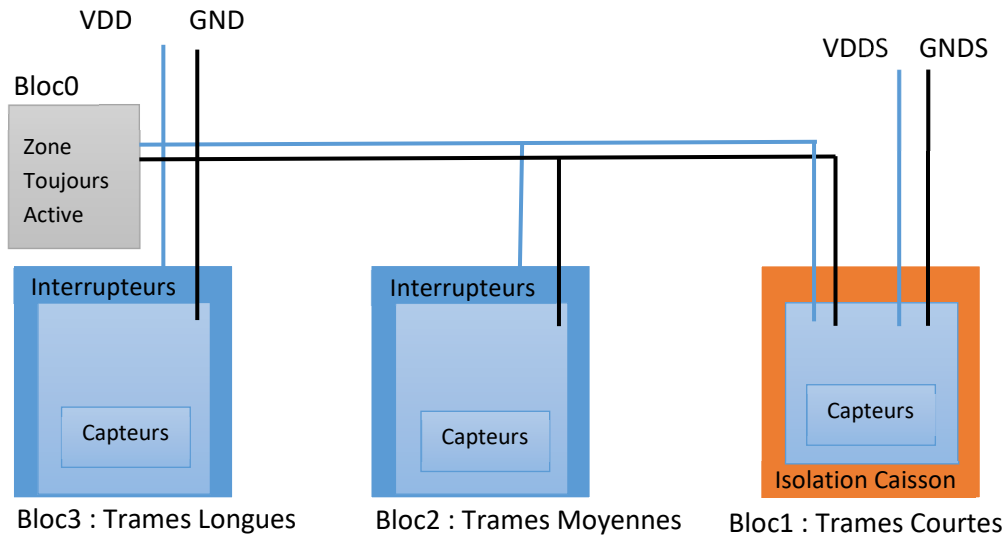


Figure 37 : Schéma du LDPC équipé de structures de réduction de consommation

Ce circuit a été fabriqué avec le procédé de fabrication ST 45nm Low Power et contient 860k Transistors, 29Kb de RAM et 67kB de ROM sur une surface totale de 1.47mm². De plus, ce circuit embarque des solutions avec des commutateurs d'alimentation afin de réduire sa consommation statique durant les périodes de veille. Dans cette implémentation, trois domaines d'alimentations sont ainsi définis pour gérer les trois tailles de trame supportées par notre encodeur/décodeur. Ainsi, nous pouvons éteindre les parties du circuit inutilisées dans le cadre de décodage de trame courte (blocs 2 et 3 éteints) ou moyenne (bloc 3 éteint). Enfin, les techniques de polarisation de substrat peuvent être appliquées dans le cadre des trames courtes donc uniquement sur le domaine principal.

L'implémentation physique de ce circuit fait appel à toutes les techniques connues pour réduire la consommation et une attention particulière a été portée sur l'îlot avec polarisation du substrat. Il faut en effet dans cet îlot définir une zone d'isolation du caisson N afin d'éviter la pollution des îlots adjacents lors de la polarisation.

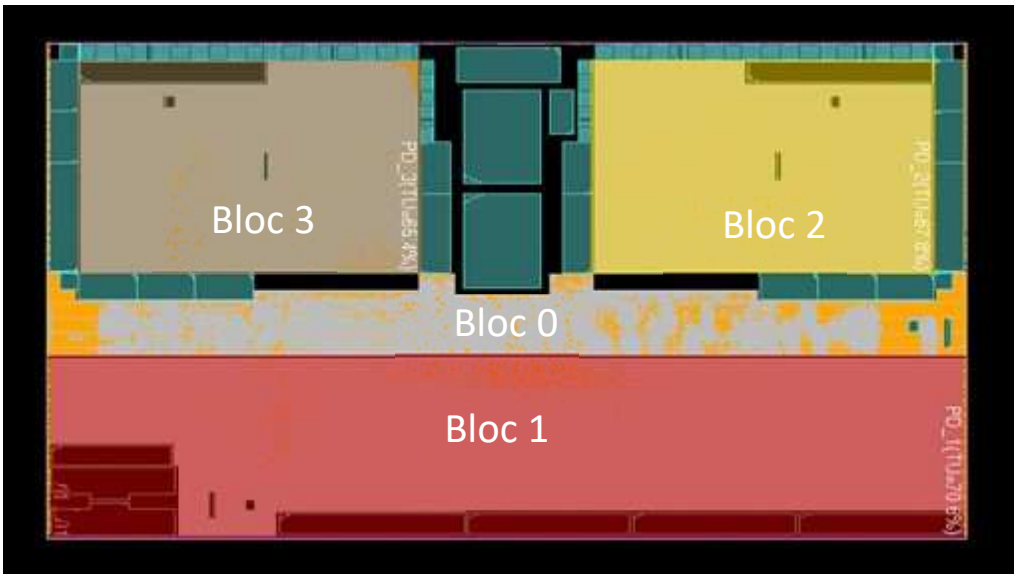


Figure 38 : Implémentation du LDPC en ST 45nm

Comme attendu et comme nous pouvons le constater sur les mesures de la Figure 39, cette répartition en îlots nous permet d'obtenir une granularité et un profil de consommation optimisé en fonction de la largeur de trame décodée.

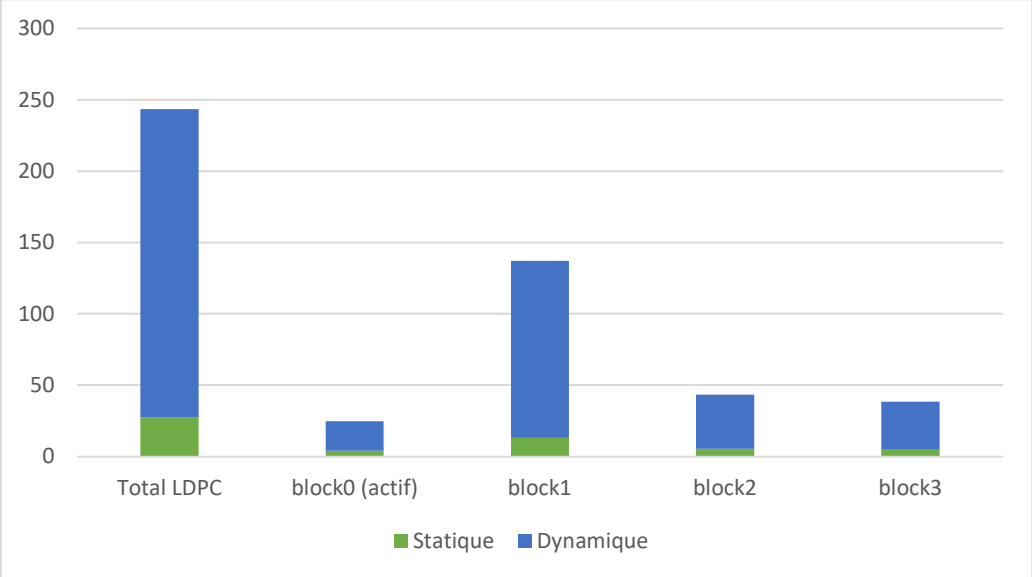


Figure 39 : Consommations (dynamique/statique) par bloc du LDPC en mW

Cette granularité s'obtient à l'aide d'interrupteurs qui permettent une variation discrète du profil de la consommation. Nous allons donc chercher maintenant à regarder s'il est possible d'optimiser ce profil de consommation en jouant de façon dynamique sur la tension ou plutôt sur les tensions car nous pouvons aussi jouer sur la tension de polarisation du substrat. Par simulation post-extraction, nous avons cherché à trouver le meilleur point de fonctionnement en fonction de la charge demandée. A savoir, sommes-nous capables avec le

même silicium de configurer un même produit dans les trois modes de fonctionnement définis dans le chapitre I à savoir LSTP, LOP et HP ?

Pour notre circuit LDPC, nous avons défini 3 algorithmes simples en fonction du positionnement du circuit. L'algorithme va chercher à s'approcher le plus possible de la cible tout en respectant la contrainte duale. Par exemple, dans le cas du mode HP, la cible sera la performance maximale et la contrainte sera le courant de fuite maximum que peut tolérer l'application. A l'inverse, dans le cas LSTP, on cherchera à réduire le courant de fuite tout en respectant une fréquence de fonctionnement minimale. Les leviers sont la tension d'alimentation de notre circuit (V_{DD}) et la tension de polarisation du substrat (V_{DDS}). Nous pouvons aussi définir la tension V_{BIAS} comme étant la différence entre V_{DD} et V_{DDS}. Si cette dernière est positive, nous parlerons de polarisation directe du substrat (FBB) et de polarisation inverse du substrat (RBB) dans le cas contraire.

La polarisation FBB permet en réduisant la tension V_{BS} d'accélérer le circuit alors que le RBB le ralentit [27]. Etant sur une technologie *bulk*, nous nous limiterons à une modulation de +/- 300mV afin d'éviter les effets indésirables liés à l'activation des jonctions S/D [28]

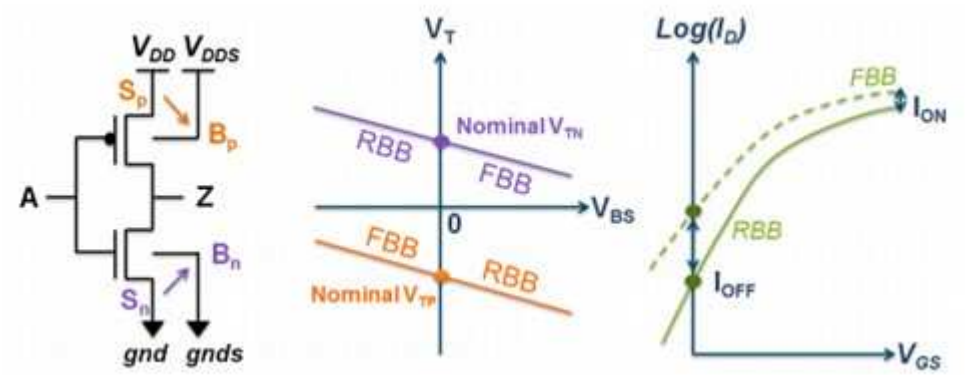


Figure 40 : Evolution des courants en fonction de la modulation de tension V_{BS}

En fonction du mode, l'algorithme va donc :

- Augmenter V_{DD} et V_{BIAS} dans le cas du mode HP afin d'obtenir la performance la plus élevée
- Augmenter uniquement V_{BIAS} dans le cas du mode LOP afin de ne pas augmenter la consommation dynamique de notre circuit.
- Diminuer V_{DD} et V_{BIAS} dans le cadre du mode LSTP afin de trouver la consommation de fuite minimale.

A chaque nouvelle étape, une analyse temporelle du circuit est effectuée ainsi qu'une estimation de sa consommation afin de s'assurer que les contraintes initiales sont toujours respectées [29].

Les limites que l'algorithme devra respecter sont résumées dans le Tableau 3. Les valeurs de tensions proviennent de la technologie 45nm Low Power et les valeurs de vitesse et de courant de fuite sont fixées pour notre application d'encodage LDPC.

Limite	Description	Valeur
V_{DDmin}	Tension minimum de fonctionnement	1.05V
V_{DDmax}	Tension maximum de fonctionnement	1.20V
$V_{BIASmin}$	Tension minimum pour la modulation du substrat	-300mV
$V_{BIASmax}$	Tension maximum pour la modulation du substrat	300mV
F_{MIN}	Fréquence minimum du circuit	200Mhz
$LEAK_{MAX}$	Courant de fuite maximum toléré par l'application	1500uA

Tableau 3 : Limites fixées pour l'algorithme de modulation de tension

Le Tableau 4 nous montre les points d'opération (V_{DD}, V_{BIAS}) obtenus pour chaque mode pour un procédé typique à une température de 30°C. La fréquence maximale en mode nominal, c'est-à-dire pour le couple (V_{DD}=1.10V, V_{BIAS}=0V) s'élèvent à 346Mhz. En appliquant le mode HP, nous obtenons pour le couple (V_{DD}=V_{DDmax}=1.20V, V_{BIAS}=100mV), une performance de 434Mhz. La tension V_{BIAS} n'atteint pas son maximum car nous sommes limités par le courant de fuite qui dépasserait la limite autorisée. Dans le mode LSTP, nous arrivons quasiment à réduire le courant de fuite d'un facteur 2 tout en gardant une fréquence au-dessus de la limite autorisée.

Mode	Default	HP	LOP	LSTP
Tension D'alimentation (V_{DD})	1.10V	1.20V	1.10V	1.05V
Tension Modulation du Substrat (V_{BIAS})	0mV	100mv	200mV	-300mV
Fréquence Max (gain en %)	346Mhz (réf.)	433Mhz (+25%)	382Mhz (+10.4%)	275Mhz (-20.5%)
Courant de fuite (ratio)	724µA (réf.)	1253µA (x1.73)	1203µA (x1.66)	421µA (x0.58)

Tableau 4 : Points d'opération optimum pour un procédé typique à 30°C

Nous entrevoyons ici l'intérêt des ajustements de tension sur la performance des circuits. Ces ajustements ont en outre l'avantage de pouvoir se réaliser après la fabrication du silicium ce qui les rend particulièrement intéressants pour ajouter de la souplesse aux caractéristiques des circuits électroniques. Ces solutions ont aussi été appliqués dans d'autres travaux présentés en [30]

III. 1. 3. La polarisation du substrat en 28nm FDSOI

Comme vu précédemment, la polarisation du substrat sur une technologie *bulk* se limite à une modulation de +/- 300mV afin d'éviter les effets indésirables liés à l'activation des jonctions S/D. Nous allons voir ici comment cette limite peut être levée sur une technologie FD-SOI et comment nous pouvons utiliser toute la plage offerte pour nos applications [30].

III. 1. 3. 1 La technologie UTBB FDSOI

La Technologie UTBB FDSOI est une technologie planaire à grille métallique utilisant un oxyde high-K (Figure 41). La source et le drain sont étendus pour réduire les résistances d'accès. Il n'y a pas de dopage de canal ni d'implantation de poches rendant le processus plus simple que son équivalent en substrat massif. L'épaisseur de la BOX est de 25nm menant à un bon compromis entre les capacités parasites entre drain/source et substrat d'un côté et l'effet de *body* de l'autre. Une face arrière, de type N ou P, est mise en œuvre sous la BOX pour améliorer l'effet de canal court (SCE) et ajuster la tension de seuil du transistor (Vt). La polarisation de la face arrière ou l'hybridation avec des transistors standards est faisable après avoir enlevé la BOX. L'isolation par des tranchées peu profondes (STI) est utilisée pour isoler électriquement les transistors entre eux.

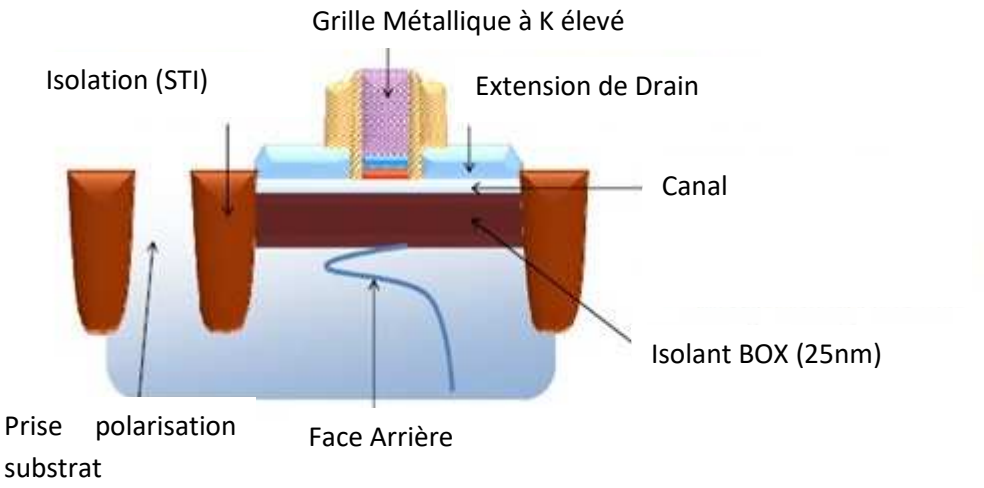


Figure 41 : Coupe d'un transistor en technologie UTBB-FDSOI

Dans la technologie SOI entièrement déplétée, le film de silicium, où la zone active du transistor est située, est amincie à environ un tiers de la valeur minimale de longueur de la porte. Ceci permet d'avoir un bon contrôle électrostatique du canal, c'est-à-dire un faible abaissement de la barrière induite par les drains (DIBL), en coupant les lignes de champs profondes provenant du drain. Dans un nœud UTBB FDSOI de 28nm, l'épaisseur du film de silicium T_{SI} est d'environ 8-9nm. Pour la technologie FDSOI de l'UTBB, le choix a été fait d'utiliser une BOX mince et cela apporte un certain nombre d'avantages : tout d'abord, il crée

une interface avec la face arrière, qui peut être considérée comme une grille arrière dont la valeur de polarisation modifie la tension seuil (V_T) du transistor. L'effet face arrière, représenter par la sensibilité du V_T à la tension de la grille arrière, est égale à 85mV/V pour une épaisseur de BOX de 25nm. *A contrario* d'une technologie sur substrat massif, la tension de la grille arrière également appelée polarisation de la face arrière ou *Back-Bias* (BB) peut varier sur une amplitude 2V. Cette polarisation permet donc un réglage dynamique de la valeur V_T, permettant aux concepteurs de circuits de définir plusieurs points de fonctionnement pour leurs circuits.

Le deuxième avantage de cette BOX mince est que, combinée avec une face arrière (la zone de dopage situé sous la BOX) de type N ou P, il est possible d'obtenir des transistors avec deux valeurs V_T différentes. Cela signifie que la technologie offre des stratégies de modulation différentes de la tension seuil, d'abord en fournissant une plate-forme multi-V_T et d'autre part en permettant un contrôle de la face arrière efficace au niveau du circuit.

III. 1. 3. 2 Amélioration du PPA à l'aide de la polarisation du substrat

La Figure 42 montre un compromis vitesse/fuite concurrentiel du FDSOI par rapport aux technologies CMOS *bulk* conventionnelles sous une tension nominale 3 types de technologies utilisant le nœud technologique 28nm sont représentés. La technologie standard ou *General Purpose* (28GP) utilise une tension nominale de 0,85V. La technologie FDSOI (28FD) utilise quant à elle une tension nominale de 1V alors que la technologie basse consommation ou *Low Power* (28LP) - qui utilise un transistor à épaisseur de grille supérieure - utilise une tension de 1,1V. Pour la consommation statique, le 28FD surpasse constamment à la fois les technologies 28LP et 28GP. Avec une application de la polarisation directe du substrat (FBB), le 28FD permet de pousser davantage les performances, évidemment au détriment d'une fuite de courant accrue, mais sans dégrader le rapport performance/fuite (ce qui en fait une solution intéressante pour augmenter l'efficacité énergique de circuit nécessitant de des calculs rapides suivie de temps d'endormissement long). En outre, avec une réduction de la tension de 200mV, le 28FD permet d'obtenir une performance équivalente au 28LP mais avec une consommation statique réduite.

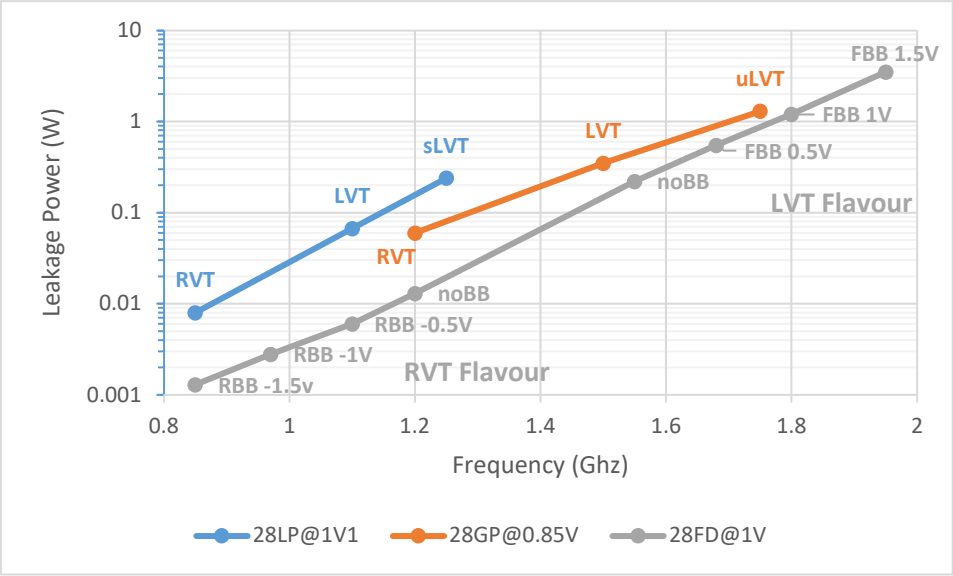


Figure 42 : Comparaison Performance/Puissance entre FDSOI et Bulk en 28nm

III. 1. 3. 3 La maximisation de l'efficacité en fonction des cas d'utilisation

Au-delà de la recherche de la meilleure performance possible pour une consommation statique, il est important d'avoir accès à la meilleure consommation totale du circuit, à savoir l'addition de la consommation dynamique à la consommation statique. Cette efficacité énergétique est d'autant plus intéressante en FDSOI qu'elle s'applique sur une large gamme de tension.

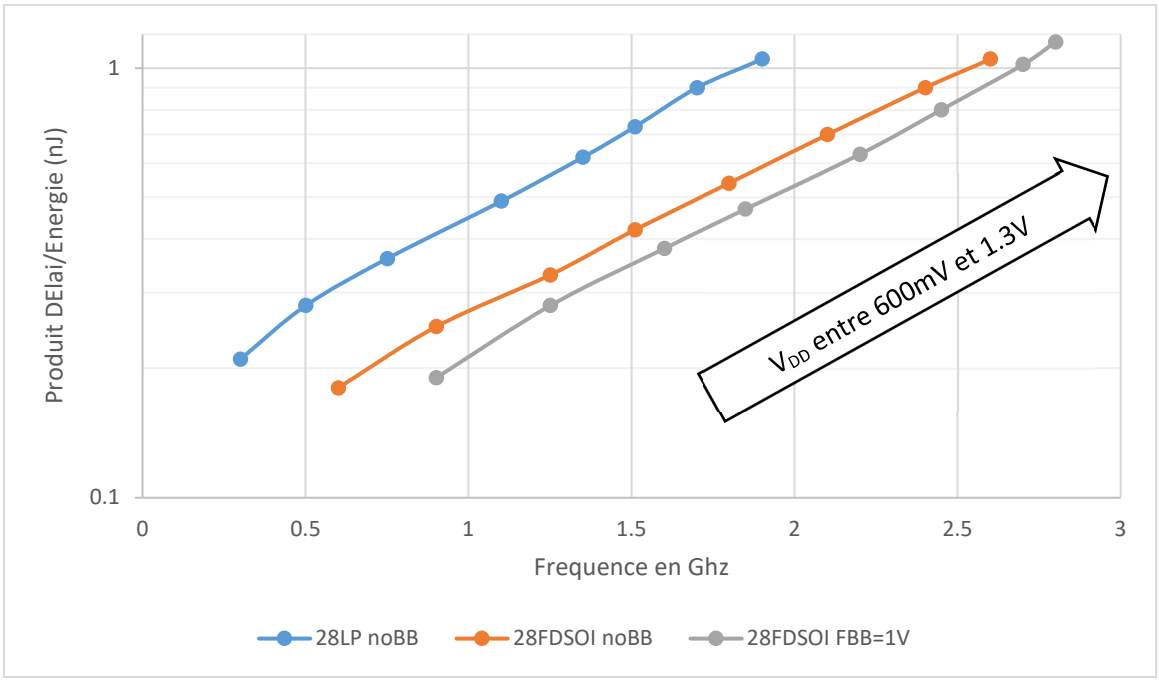


Figure 43 : Comparaison du produit délai/Energie entre FDSOI et Bulk en 28nm

Sur la Figure 43, le produit du délai par la consommation (PDP) ramené à la fréquence de fonctionnement pour différentes tensions entre 0,6 à 1,3 V montre une augmentation de vitesse de 40% à 200% sans coût énergétique supplémentaire. En outre, la technologie FDSOI démontre une efficacité énergétique attrayante avec une réduction de 50% de la consommation d'énergie pour une vitesse constante et ce quelle que soit la tension d'alimentation appliquée. Nous pouvons également observer que la technologie 28LP est pénalisée par une forte consommation d'énergie dynamique (V_{DD} est plus élevé), ce qui affecte négativement les chiffres de la puissance totale et, qu'en revanche, la technologie 28FDSOI est économe en énergie dans l'ensemble de la plage de tension V_{DD} et de la plage de fréquence cible.

En outre, avec le contrôle élargi des polarisations de substrat (FBB/RBB), il est possible soit d'améliorer l'efficacité énergétique ou d'atteindre des performances très élevées pour des activités en rafale ce qui offre de nouvelles opportunités.

III. 1. 3. 4 Les opportunités d'un contrôle de tension du substrat élargie

Dans la technologie planaire UTBB FD-SOI, la polarisation du substrat permet d'ajuster la tension de seuil des transistors. Ainsi, nous pouvons soit obtenir plus de courant (donc des performances plus élevées) au prix d'une l'augmentation du courant de fuite (FBB) ou inversement réduire les courants fuites au détriment d'une performance réduite (RBB).

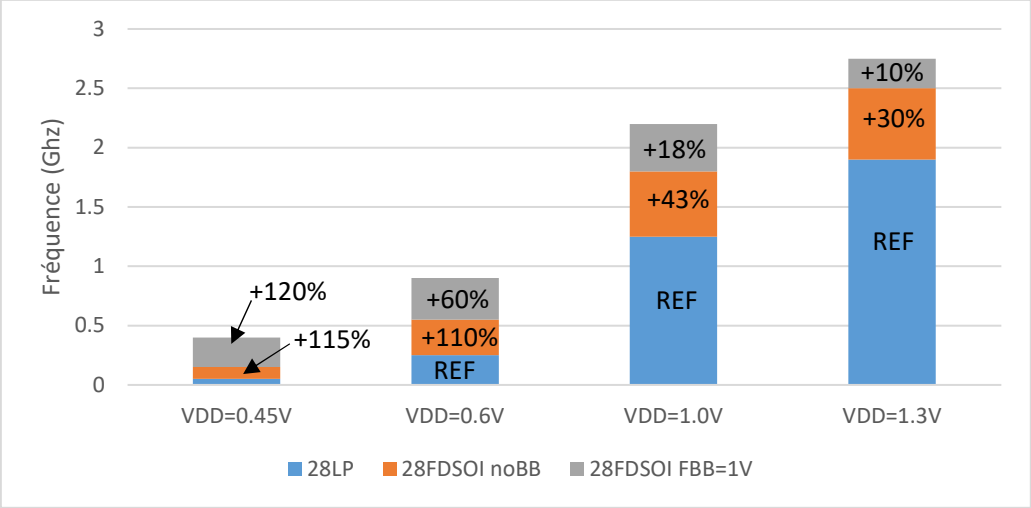


Figure 44 : Gain en performance en fonction de la polarisation avant du substrat (FBB)

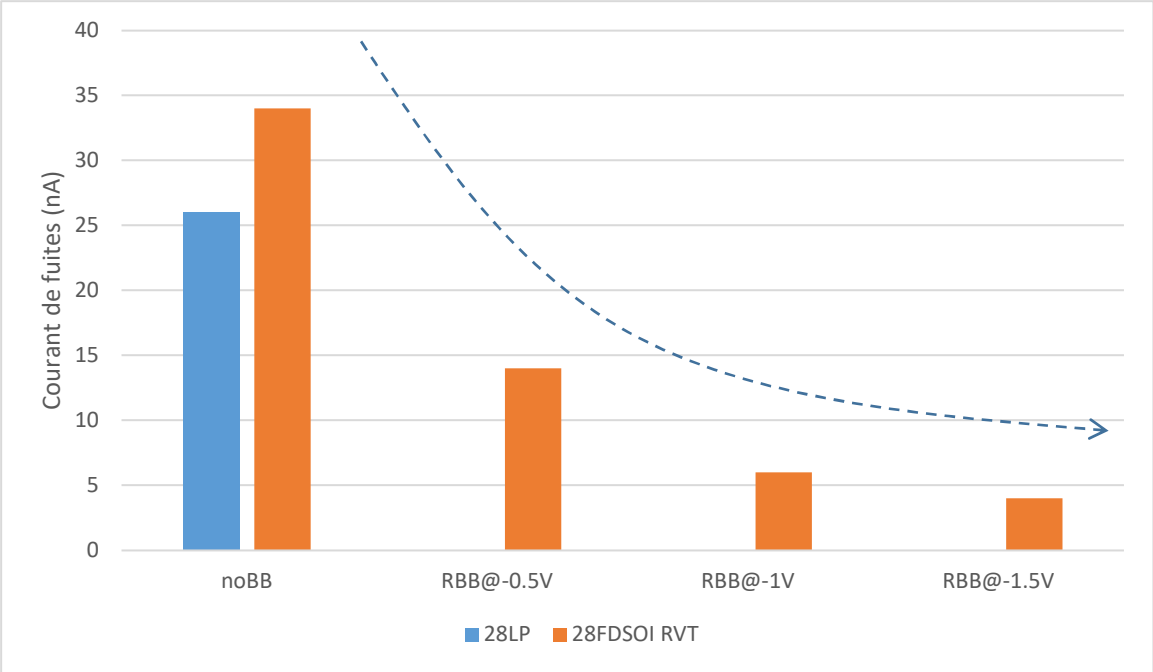


Figure 45 : Gain en consommation en fonction de la polarisation inverse du substrat (RBB) pour une tension de 600mV

La Figure 44 démontre une augmentation significative de la fréquence des transistors FDSOI par rapport à la technologie *bulk* pour toutes les tensions mesurées. Avec une polarisation directe de 1V, le FDSOI démontre un très fort gain de vitesse de 43% à 1,3V jusqu'à un facteur x5 à 0,45V. A l'inverse, lorsque la performance n'est pas requise, le FDSOI permet de réduire la consommation des fuites en appliquant des polarisations inverses (Figure 45). À 0,6 V (tension généralement utilisée pour la mise en veille), le courant de fuite est divisé par un facteur 2 à -0.5V mais nous pouvons gagner jusqu'à un facteur x10 en poussant la polarisation à -1.5V.

III. 1. 3. 5 La polarisation en complément de l'ultra-large variation de tension

Parmi les multiples utilisations du FDSOI, il en est une qui nous intéresse tout particulièrement. Il s'agit de la possibilité pour les circuits d'utiliser une large gamme de tension. Les systèmes sur puces utilisent aujourd'hui une technique dite DVFS (*Dynamic Voltage and Frequency Scaling*). Cette technique permet d'adapter l'alimentation du circuit en fonction de son besoin de puissance. Le même circuit pourra ainsi être utilisé pour effectuer un calcul lourd avec une tension élevée et sera capable d'effectuer des tâches plus légères en réduisant sa tension d'alimentation afin de réduire sa consommation. Cette technique

permet donc d'augmenter globalement l'efficacité énergétique de notre système et sera d'autant plus efficace que la plage de tension accessible par notre circuit est élevée.

Dans ce cadre, l'utilisation de la polarisation du substrat amène une extension du champ des possibles pour la tension d'alimentation en nous permettant de compenser les effets de perte de vitesse à très basse tension ou de contrôler l'énergie sur les tensions élevées.

Nous avons donc maintenant accès à un nouvel actionneur qu'il nous faut contrôler de la manière la plus efficace possible. Pour ce faire, il est crucial de connaitre le l'usage de notre circuit afin de lui appliquer la bonne tension et la bonne polarisation.

III. 2. Les techniques de compensation du procédé

Afin d'améliorer l'efficacité du contrôle du substrat et de la mise à l'échelle de la tension, nous pouvons ajouter à notre circuit certains dispositifs qui nous permettent de détecter l'emplacement des défauts liés aux procédés de fabrication. Cette indication de centrage des procédés de fabrication nous permet de mieux contrôler ces actuateurs vus dans la partie précédente. Ils nous permettront aussi dans les technologies les plus avancées de lutter contre la variation des procédés de fabrication en nous offrant un contrôle à grain fin, c'est-à-dire au niveau du circuit et non plus au niveau de la plaque complète (*wafer*).

III. 2. 1. Notre premier capteur de centrage : le PMB

Pour permettre la surveillance des performances sur puce, un ensemble d'oscillateurs spécifiques a été conçu afin de surveiller individuellement les performances des transistors de type N et P en termes de vitesse et de courant de fuite.

Chaque capteur est composé de trois structures oscillantes spécifiques appelées respectivement *'Speedometer NMOS'*, *'Speedometer PMOS'* et *'Leakometer'*. Ces structures en anneaux ont été conçues pour être aussi représentatives que possible des circuits réels, en termes de densité et de couches métalliques. Les signaux délivrés par ces structures sont des signaux carrés périodiques et c'est la période de ces signaux qui nous donne un aperçu de la qualité de la fabrication en termes de vitesse et de courant de fuite. Ces moniteurs ont été développés et validés sur différentes technologies successives et pour toutes les options des procédés disponibles afin de répondre à un large portefeuille de produits.

La validation de ces capteurs est une étape cruciale et requiert une grande quantité de données. L'analyse effectuée est d'ordre statistique et nécessite donc un grand nombre d'échantillons afin d'être représentative. Les capteurs permettent ainsi de capturer les variations globales des procédés de fabrication (entre plaques) mais aussi des variations plus fines entre les différents circuits d'une même plaque, voire entre différentes parties d'un même circuit.

À titre d'illustration de cette campagne de validation sur silicium, la Figure 46 représente les fréquences à la sortie du *'Speedometer NMOS'* et du *'Speedometer PMOS'*. Ces mesures ont été extraites sur cinq lots dits d'angle ou coins, intentionnellement traités pour obtenir des transistors N et P lents, typiques ou rapides.

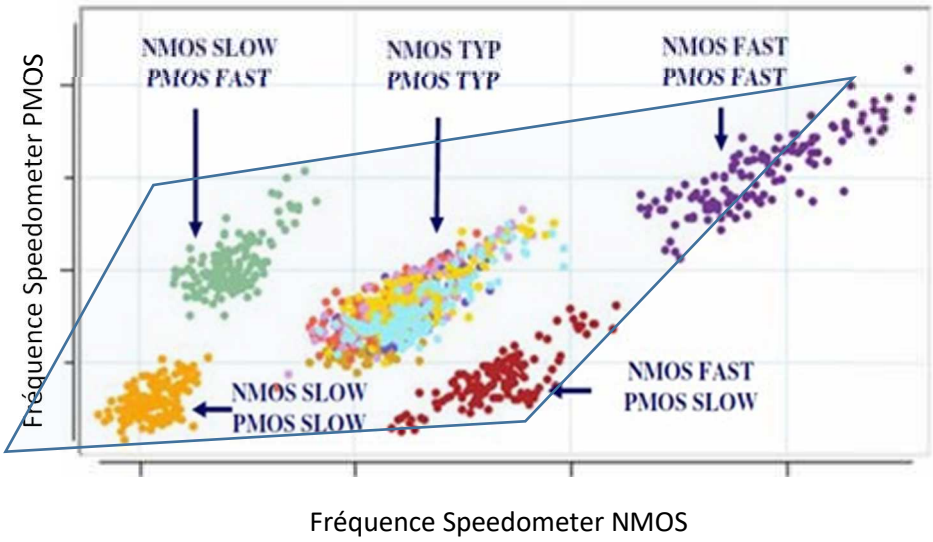


Figure 46 : Centrage des Transistors P et N sur un lot de 5 plaques coin

Comme prévu, les fréquences mesurées pour les lots coins Rapide/Rapide, Typique/Typique et Lent/Lent s'alignent presque parfaitement. Pour les lots de coins croisés (Rapide/Lent et Lent/Rapide) la mesure obtenue nous place la performance respectivement au-dessous et au-dessus de la ligne de coupe. Nous retrouvons ainsi le « trapèze » (en bleu sur la Figure 46) généralement utilisé pour caractériser les performances d'un procédé de fabrication. La précision du capteur est ainsi suffisante pour connaitre le centrage du procédé de fabrication sur chaque circuit.

Nous allons maintenant décrire comment utiliser de façon efficace ces capteurs pour compenser les variations des procédés de fabrication.

III. 2. 2. L'utilisation des capteurs pour compenser les variations

Une cible classique des techniques de réglage post-silicium est la conception de circuits à grande vitesse type processeur dans les technologies de pointe [31]. L'utilisation de ces techniques dans des circuits plus standards ou plus hétérogènes reste à développer. Dans ce cadre, nous devons définir un nouveau flot de conception qui cherche à optimiser au maximum l'efficacité énergétique en trouvant le meilleur point de fonctionnement tout en assurant un fonctionnement correct après adaptation [32]. Dans ces travaux, nous travaillerons

uniquement sur l'ajustement de la tension car elle réduit simultanément la consommation d'énergie statique et dynamique et cela pendant tout le cycle de vie du produit.

L'idée présentée dans la Figure 47 est ainsi de réduire la tension des pièces classées rapides afin de les aligner avec les performances des pièces lentes. Cette réduction de tension nous donnera un gain en consommation sur les pièces les plus énergivores.

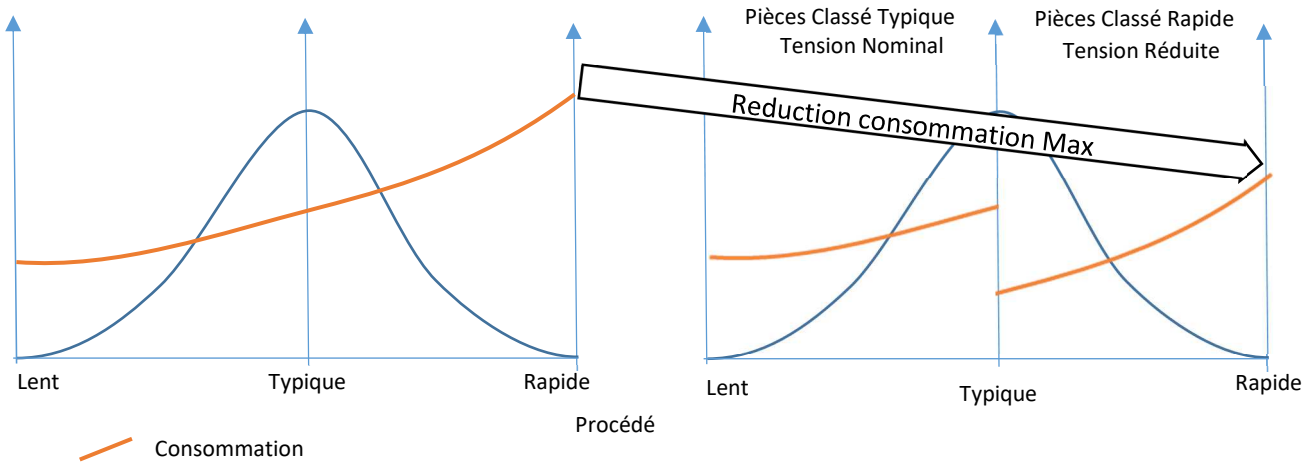


Figure 47 : Réduction de la consommation par ajustement de tension

III. 2. 2. 1 Conception de la solution

La première étape est d'ajouter les capteurs dans le circuit. Nous proposons d'utiliser une interface de type JTAG (Joint Test Action Group) afin de permettre l'accès à nos capteurs au travers d'un large éventail de solutions existantes. Nos capteurs seront donc vus comme une solution additionnelle de test ressemblant aux solutions BIST pour les mémoires.

Une fois cette intégration faite, nous devons nous assurer du bon fonctionnement du circuit dans sa nouvelle plage de tension. En effet, le coin le plus lent n'est plus le coin Lent mais le coin compensé. Nous devons donc avoir nos bibliothèques de conception caractérisées sur ce nouveau point et s'assurer que les délais sont corrects grâce à une analyse temporelle statique (ou STA).

III. 2. 2. 1 Utilisation de la solution

Afin de classifier correctement les circuits, une mesure des capteurs est réalisée au niveau du test électrique dit de tri des plaques silicium. La mesure pourra être répétée pour moyenner la valeur obtenue au détriment d'un temps et donc d'un coût de test augmenté. Cette mesure est ensuite comparée à une mesure de référence. Si la mesure du *Speedometer NMOS* (resp. PMOS) est supérieure à la mesure de référence, le transistor N (resp. P) sera considéré comme Rapide. Le classement de la pièce sera stocké dans une mémoire OTP afin d'informer le régulateur de la bonne tension à appliquer dans l'application.

III. 2. 2. 2 Résultats expérimentaux

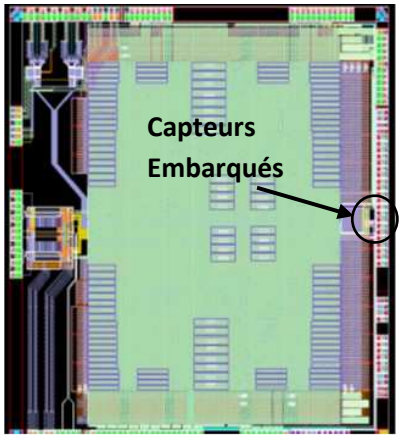


Figure 48 : Circuit de test pour la compensation en tension

Cette technique de compensation des variations du procédé a été appliqué à plusieurs produits au sein de STMicroelectronics. Pour ces travaux, nous avons intégré cette solution dans un circuit de test conçu dans une Technologie qui mélange les transistors LP et GP. Ce multiplexeur de canaux numériques pour satellite extérieur a une surface de silicium de 12,75 mm² et une puissance totale numérique de 950mW sous une tension nominale de 1V [33].

La stratégie de compensation retenue pour ce circuit a été d'appliquer une tension nominale sur les circuits classés Lent et de réduire la tension nominale de 80mV pour les circuits classés Rapide.

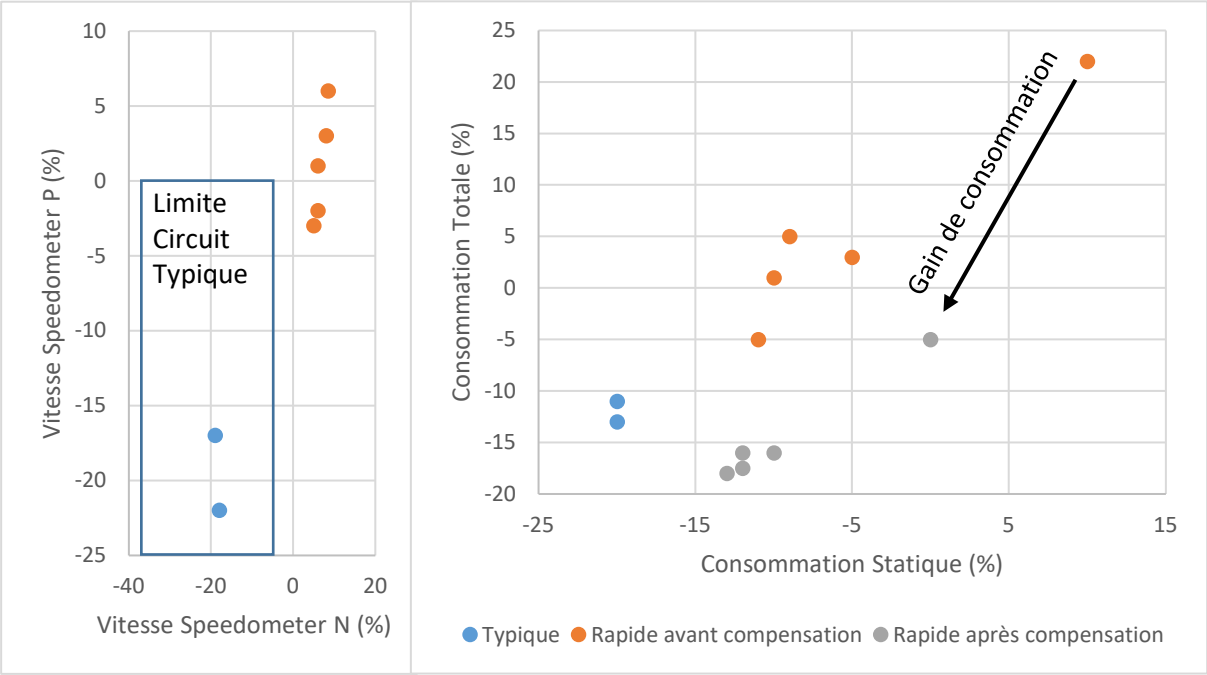


Figure 49 Performance du circuit de test avant et après compensation

Tel qu'il est présenté à la Figure 49, sans appliquer de compensation de tension, les circuits classés « Rapide » peuvent fonctionner à une fréquence maximale jusqu'à 8 % supérieure à la fréquence nominale. Cela s'explique principalement par la qualité des transistors NMOS et PMOS. Toutefois, leur consommation d'énergie statique augmente et peut atteindre une valeur jusqu'à 23% plus importante que celle des circuits classés « Typique » alors que le gain en performance de 8% ne peut être exploité. En effet, la

plateforme fonctionnant à une vitesse d'exploitation fixe sur toutes les pièces, elle est limitée par la fréquence des circuits classés « Lent ».

Comme prévu, lorsque la stratégie de compensation proposée est mise en œuvre, la réduction totale de la consommation d'énergie varie entre 10% et 20% tout en en maintenant la fréquence nominale d'exploitation.

Les résultats obtenus démontrent que la stratégie de compensation réduit la puissance totale des puces rapides sans compromettre les performances et peut éviter également l'utilisation de boitiers coûteux nécessaire à la dissipation de la puissance supplémentaire induite par les pièces classées « Rapides ».

Ces résultats démontrent également que les capteurs sur puce donnent des résultats précis au prix d'une légère augmentation de surface (0,04 mm²) et d'une augmentation du temps de test. Ces augmentations de surface et du temps de test doivent être mises en regard des gains obtenus sur chaque application afin d'estimer la pertinence économique de la solution.

III. 2. 3. Une nouvelle proposition de capteur : CODA

III. 2. 3. 1 Conception de la solution

L'une des limites du PMB est la corrélation entre les performances de l'oscillateur en anneau et les performances de la puce. Alors que l'oscillateur utilise une structure régulière basée sur des inverseurs, la performance du circuit est liée à son chemin critique qui est composé de cellules hétérogènes et complexes. Cette différence de structure génère des différences de mesures et réduit la corrélation entre notre capteur et le circuit. Afin de s'approcher au mieux du chemin critique, nous proposons une solution qui extrait ce chemin de la puce pour construire une boucle oscillante avec ses cellules. Nous l'appellerons CODA [34]

CODA est donc un capteur basé sur le circuit qui prend en compte les variations locales des chemins critiques impactant la fréquence maximale du circuit hôte (F_{MAX}). Il est indépendant du circuit logique hôte surveillé et peut fonctionner pendant que le circuit est actif. Il peut contenir jusqu'à 16 emplacements, dans lesquels différents clones de chemins critiques représentatifs (RCP) sont insérés à l'aide d'un mécanisme de « copier-coller ». Une duplication exacte (cellules, charge de connexion) d'un RCP (RCP-clone) est construite à proximité physique afin de capturer la variation locale du chemin critique. Le chemin direct (appelé *canari*) permet de capturer rapidement une erreur de timing et d'appliquer les techniques de « *replay* » présenté en [35]. Le chemin de retour (appelé *boucle*) permet de

générer une oscillation et d'utiliser les algorithmes de compensation présentés au chapitre précédent (Figure 50).

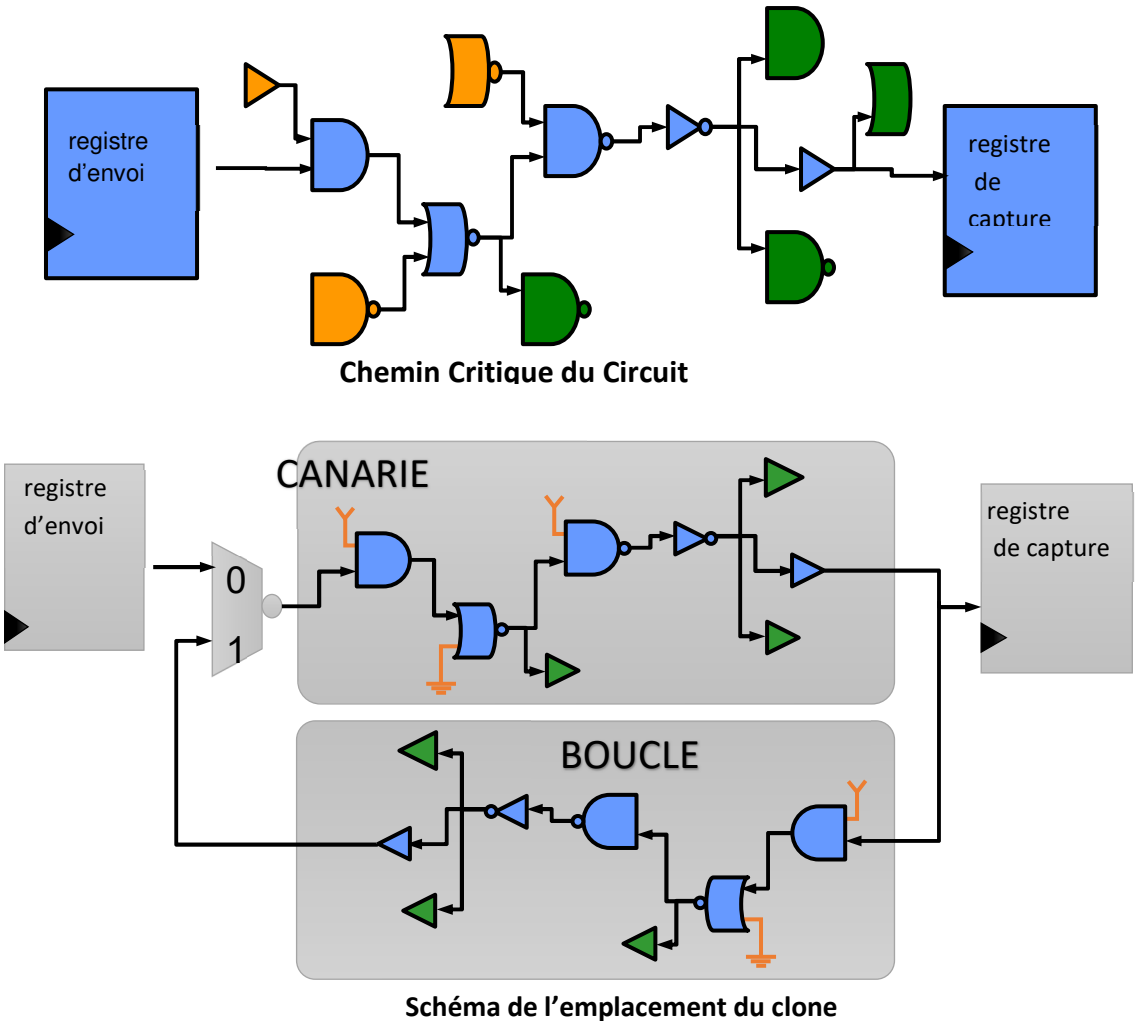


Figure 50 CODA Chemin critique et emplacement de clone

III. 2. 3. 2 Intégration de la solution

En s'appuyant sur un jeu de scripts TCL compatibles avec les outils de CAO, nous avons rendu le procédé de clonage entièrement compatible avec le flot de conception du circuit hôte. Comme nous le voyons sur la Figure 51, les cellules des chemins clonés sont placées au plus proche du chemin critique afin d'obtenir une bonne représentation des variations liées aux interconnections.

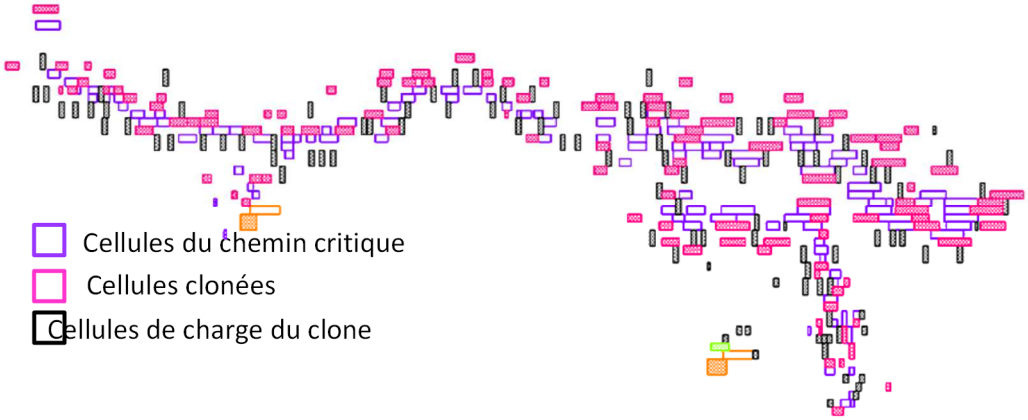


Figure 51 Implémentation de la solution de clonage

Deux modes de mesure sont disponibles pour chaque emplacement, canarie (moniteur de retard) et boucle (fréquence d'oscillation). Le retard de propagation du RCP ($D_{RCP)}$ est déterminé en mode canari par le retard de RCP-clone ($D_{RCP\text{-}clone)}$ - un avertissement émis apparaît quand $F_{CLK}=1/D_{RCP\text{-}clone}$ ($< 1/D_{RCP)}$ - et peut être corrélé en mode boucle par la mesure directe de la fréquence oscillante ($F_{RCP\text{-}clone).}$

III. 2. 3. 3 Résultats expérimentaux

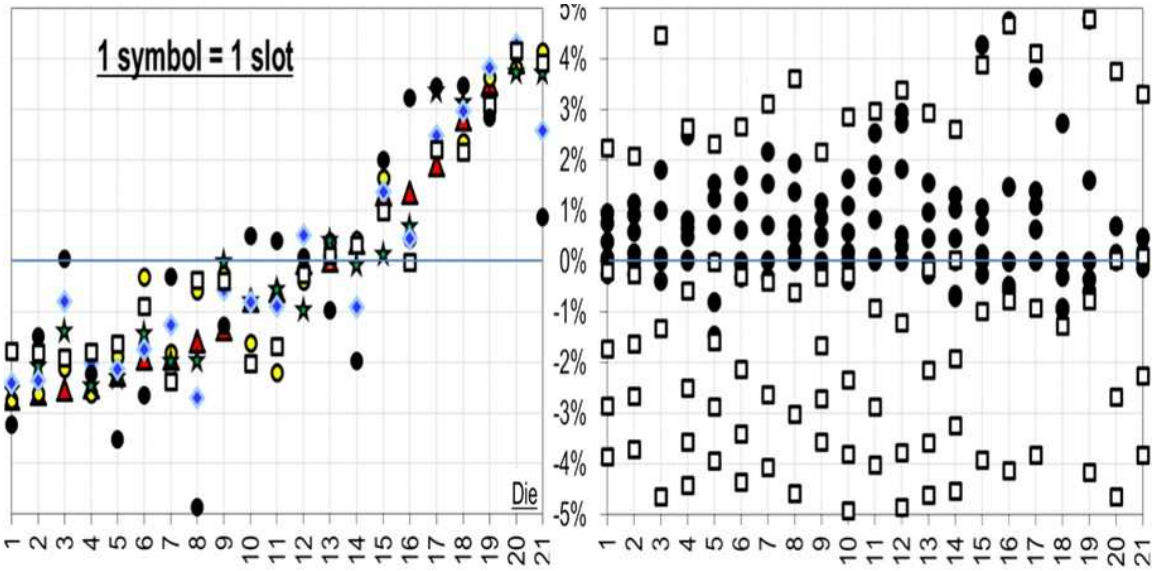
Cette solution a été embarquée dans le circuit de test BHAG-DSP [36]. La mesure sur silicium de BHAG-DSP a été faite pour déterminer F_{MAX} et effectuer une corrélation avec la prédiction de fréquence de CODA. Les mesures de deux chemins cloné (*slot*) présentées dans la Figure 52 montrent que la fréquence F_{MAX} peut être prédite avec une précision de 2,3% en utilisant le mode *canari* sur la plage de tensions [1V-1.4V]. L'analyse de plusieurs circuits en mode boucle montre que le rapport entre $F_{RCP\text{-}clone}$ et F_{MAX} varie de 5% à 1V pour un seul circuit. Ce ratio varie de moins de 2% sur la gamme complète de « tension large » [600mV-1.4V].

Les travaux présentés en [37] et [38] montrent une extension possible avec une approche astucieuse de l'utilisation du chemin *canari*. Dans ces travaux, nous utilisons le CODA non pas pour mesurer la performance du circuit mais plutôt pour estimer son vieillissement. Le chemin *canari* est ainsi régulièrement activé en parallèle du chemin critique lors de la vie du circuit. Le déclenchement du drapeau de détection d'erreur nous indique le vieillissement du *canari* et par conséquent le vieillissement du circuit.

Comparaison Fréquence du chemin cloné / Fréquence Max du circuit

Mesure d'1 chemin cloné sur plusieurs puces
Vdd [1.0V] Température 25°C

Mesure de plusieurs chemin clonés sur une même puce
Vdd [0.8V-1.3V] Température 25°C



Puce n°9/25°C/Vdd	0.8	0.9	1	1.1	1.2	1.3	Erreur
Slot •: Clone vs Fmax	-6.4%	-2.7%	-0.7%	0.7%	2.0%	2.6%	±4.5%
Slot : Clone vs Fmax	-1.1%	-0.2%	0.2%	0.5%	-0.1%	0.1%	±0.6%

Figure 52 : Mesures Silicium de la solution CODA

III. 3. Conclusion du chapitre

Au travers d'une recherche de lien entre les procédés de fabrication et la conception de circuits intégrés, ce chapitre nous propose deux solutions nous permettant d'aller chercher un optimum en efficacité énergétique.

Comme nous l'avions abordé dans le premier chapitre, la tension joue un rôle clef dans la réduction de la consommation dynamique. Savoir comment contrôler cette tension est donc capital dans notre recherche d'efficacité énergétique. La première partie de ce chapitre propose ainsi des premières solutions pour contrôler la tension principale et commence à s'intéresser à la possibilité de contrôler la tension de la face arrière en polarisant le substrat. Cette solution de polarisation a été essayée avec un circuit de test sur une technologie 45nm qui a permis d'ouvrir la voie à de plus amples recherches.

Par la suite, nous avons vu comment cette solution de polarisation de substrat entre en symbiose avec la technologie FDSOI. Cette technologie permet d'obtenir une gamme plus

large de tensions possibles pour polariser le substrat. Cet élargissement permet de multiples combinaisons comme nous l'avons montré dans un nouveau circuit de test qui exploite cette possibilité de polarisation pour élargir la plage de tension principale du circuit tout en conservant des performances acceptables.

Dans la deuxième partie, nous avons cherché à lutter contre la dispersion des procédés de fabrication avancés. Nous avons vu que cette dispersion avait un impact direct sur l'efficacité énergétique et que connaitre le positionnement du circuit après fabrication permettait de compenser ces variations. Cette première solution a été suivie d'une deuxième proposition dans laquelle nous avons conçu un capteur qui clone le chemin critique du circuit. Cette réplication permet d'obtenir une précision de mesure extrêmement fine et ouvre ainsi la possibilité d'un contrôle précis de l'efficacité énergétique des circuits embarquant cette solution.

Chapitre IV. Les innovations de rupture pour demain

IV. 1. Du contrôle de l'horloge …

Dès le commencement des travaux sur l'optimisation de l'efficacité énergétique des circuits, l'horloge et le système d'horlogerie associé ont été identifiés comme un contributeur majeur de la consommation énergétique de nos circuits. Le contrôle de l'horloge est ainsi déterminant pour obtenir des circuits qui respectent finalement les contraintes de vitesse mais aussi de consommation.

Il existe une grande variété de travaux sur le contrôle de l'horloge et les travaux sur les systèmes utilisant les deux fronts présentés dans le Chapitre II intègre complètement cette problématique. Nous nous sommes aussi intéressés à d'autres techniques nous permettant de contrôler plus finement l'horloge.

IV. 1. 1. L'impact de la température sur l'horloge.

Travailler sur le couple tension/fréquence pour améliorer l'efficacité énergétique est assez répandu avec les techniques DVFS et leurs variantes auto-adaptative type AVS, ABB etc… Dans le cadre des travaux menés, nous nous sommes davantage intéressés à l'impact de la température sur la fréquence de nos circuits.

IV. 1. 1. 1 Le phénomène d'inversion de température

Traditionnellement, le courant d'un transistor MOSFET (I_D) diminue avec l'augmentation de la température. Par conséquent, la performance la plus faible est obtenue lorsque le transistor est soumis à une température élevée. Avec la réduction des dimensions du transistor, V_{DD} et V_T diminuent, mais pas aussi rapidement que le reste des paramètres (comme l'épaisseur de l'oxyde de porte, la longueur du canal, etc.). Regardons l'équation actuelle du courant de drain du transistor.

$$I_D = \frac{1}{2}.\mu.C_{ox}.\frac{W}{L}(V_{GS} - V_T)^2$$

I_D varie linéairement en fonction de μ (mobilité) et $[V_{GS}\text{-}V_T]^2$ (tension d'overdrive). La mobilité et le Vt diminuent avec l'augmentation de la température et vice versa ; de même, nous pouvons remarquer que le courant dépend de la différence entre V_{GS} et V_T. Il y a donc un phénomène inverse entre la mobilité et le terme (V_{GS}-V_T), et celui qui aura le plus d'impact sur le courant final déterminera si le courant augmente ou diminue avec l'augmentation de la température. Comme nous le voyons sur la figure suivante, pour les technologies nanométriques, la tension d'alimentation est réduite à une valeur proche de 0,9V alors que la tension de seuil (V_T) ne varie pas aussi agressivement en atteignant des valeurs de 0,3~0,4V.

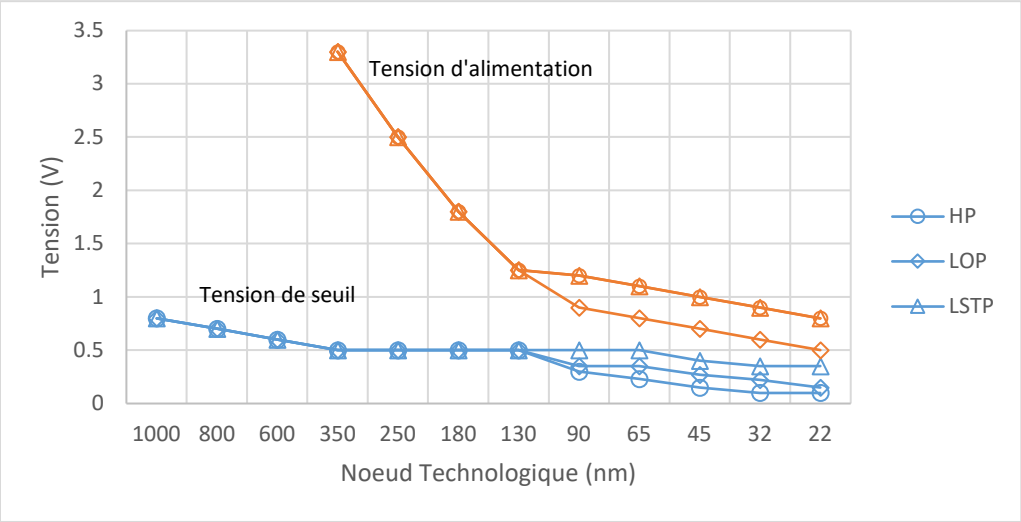


Figure 53 : Tension de seuil et d'alimentation en fonction des nœuds technologiques (Source ITRS 2004)

Finalement, même si la mobilité s'améliore à basse température, l'augmentation du Vt et donc la réduction (V_{GS}-V_T) a un plus grand impact sur le courant, ce qui entraîne moins de courant à basse température qu'à une température plus élevée [39]. Ce phénomène est aussi visible sur la Figure 54 sur laquelle nous voyons le comportement s'inverser si le transistor travaille majoritairement dans la zone I.

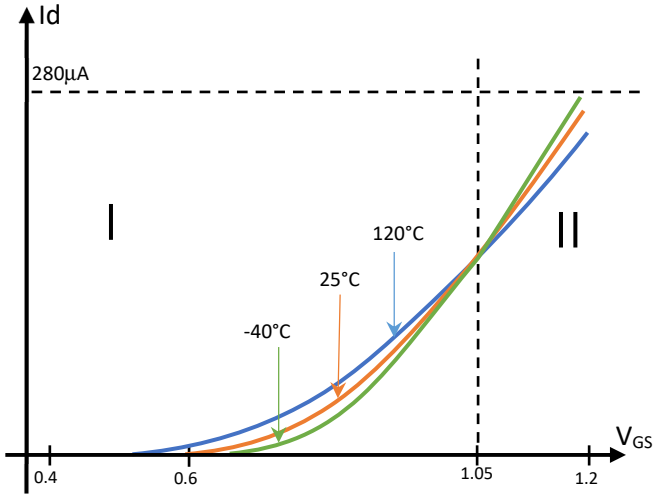


Figure 54 : Evolution du courant d'un transistor N (source : LIRMM)

En conclusion, à l'échelle nanométrique, le coin le plus lent est le coin à basse température et non plus à température élevée. Ce phénomène est amplifié avec des dispositifs à fort V_T utilisés pour réduire les courants de fuite des circuits submicroniques.

IV. 1. 1. 1 Une utilisation du phénomène d'inversion de température

Pour les technologies les plus avancées, la recherche du meilleur compromis entre la performance et la réduction des courants de fuite ont amené à se concentrer sur des techniques de conception proche du seuil. En déplaçant la tension d'alimentation à une valeur légèrement supérieure à la tension de seuil, les effets de courant de fuite peuvent être considérablement réduits. En supposant une activité logique où 15% des portes sont actives à chaque cycle d'horloge, nous observons sur la Figure 55 un croisement entre la puissance dynamique et la consommation d'énergie statique lorsque le circuit fonctionne à une tensions proche de la tension de seuil [40]. Le point de fonctionnement optimal est celui qui maximise l'efficacité énergétique de nos circuits.

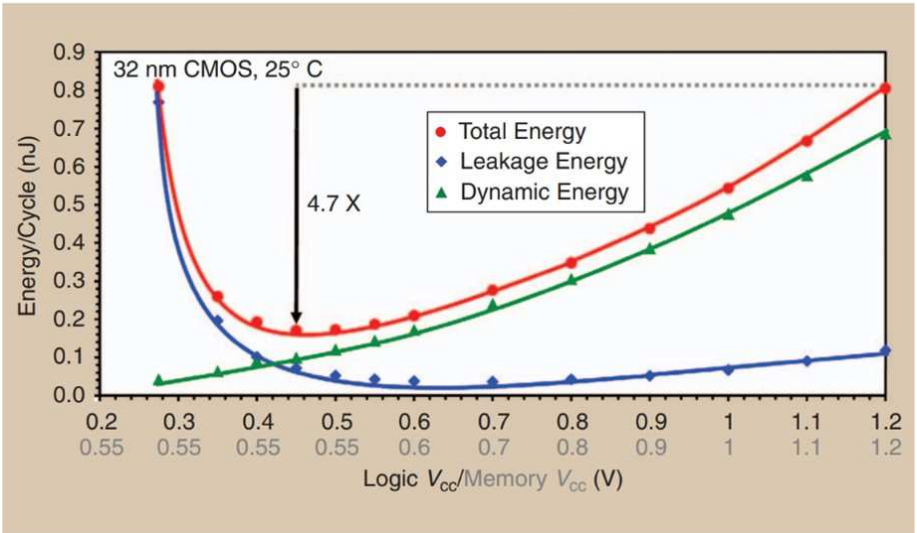


Figure 55 : Point de fonctionnement optimal en 32nm (Source : University of Michigan)

Par construction, la différence entre la tension d'alimentation et la tension de seuil des circuits conçus avec cette technique est faible. Ils ont donc une dépendance inversée par rapport à la température que nous voulons exploiter.

Quand la tension d'alimentation est largement supérieure à la tension de seuil, les circuits ont un comportement classique et proche du graphique de gauche dans la Figure 56. La fréquence nominale doit être garantie à toutes les températures. Le concepteur doit ainsi s'assurer que le circuit atteint sa fréquence de fonctionnement lorsqu'il est soumis à une température haute. Il correspond en général à la température maximum supportée par le boitier ou l'application.

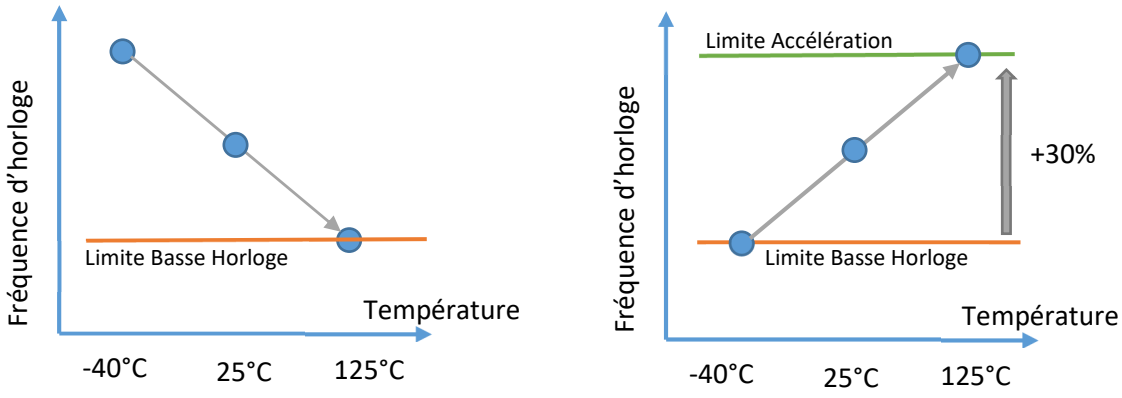


Figure 56: Lien entre fréquence et température pour 2 types de circuits

Dans le cadre de circuits fonctionnant sous le régime d'inversion de température, nous devons toujours garantir la fréquence nominale ; aussi les optimisations seront faites à basse température. A l'aide d'un capteur de température et d'un système asservi de l'horloge type DLL, nous pouvons augmenter la fréquence lors de l'échauffement du circuit. Ce système ne fonctionne pas, dans le cas où la tension d'alimentation est largement supérieure à la tension de seuil, car une augmentation de température implique un ralentissement du circuit.

Ce système breveté [41] permet d'obtenir un gain de 30% entre la fréquence nominale et la fréquence accélérée. Il est particulièrement intéressant dans les systèmes fonctionnant en rafale (voir chapitre I) où l'activité doit être réduite sur un temps minimum afin de rendormir le système le plus rapidement possible.

IV. 1. 2. Le contrôle de la gigue globale

La gigue est un élément clef à prendre en compte lors de la construction de l'arbre d'horloge. Il s'agit de la différence de temps d'arrivée des horloges entre deux registres communiquant entre eux. Si les registres sont dans le même bloc de propriété intellectuelle, nous parlerons de gigue locale. Dans le cas où les registres sont dans 2 blocs différents, nous parlerons de gigue globale [42]. A l'inverse de la gigue locale qui est quasi parfaitement maitrisée par les outils de CAO, la gigue globale est plus difficile à appréhender.

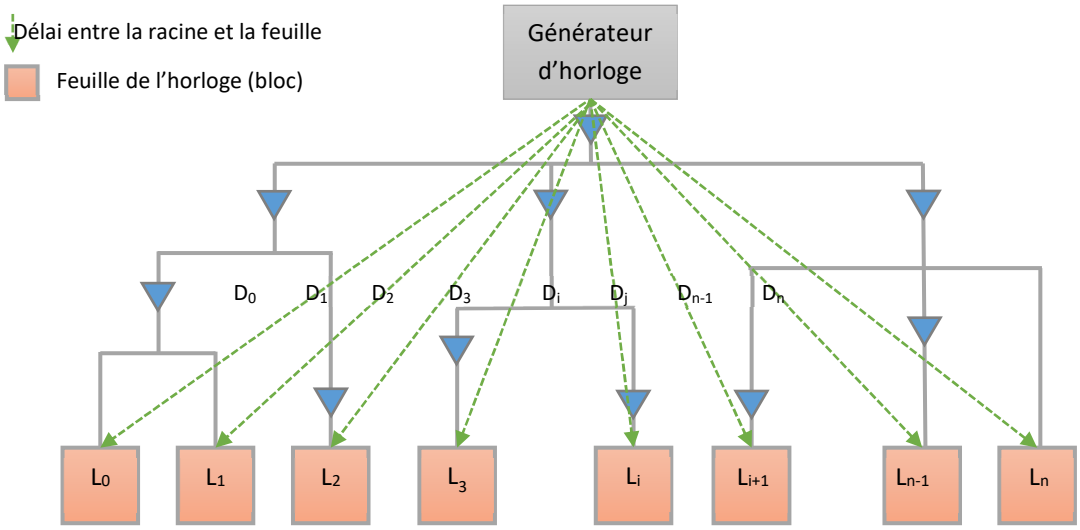


Figure 57 : Structure arbre d'horloge

Sur la Figure 57, la gigue globale est la différence maximum entre le délai de 2 blocs de propriété intellectuelle (PI) et peut donc s'exprimer avec l'équation suivante :

$$Gigue\ Globale = \max\left(\left|D_i - D_j\right|\right)$$

Une des solutions connues pour contrer les effets de la gigue globale est l'utilisation d'architecture de type Globalement Asynchrone Localement Synchrone (GALS) où les blocs de PI disjoints communiquent de façon asynchrone entre eux. Cette solution fonctionne plutôt bien pour des blocs hétérogènes mais ne s'applique pas dans le cas de blocs homogènes qui ont besoin d'échanger un grand nombre de données entre eux avec une latence réduite. Nous retrouvons par exemple dans cette catégorie les unités de calcul des processeurs multicœurs ou les commutateurs rapide internet. Pour ces circuits, l'impact d'un mauvais contrôle de la gigue globale se traduit par une perte de vitesse et une augmentation de la surface pour compenser les violations de maintien induite par cette gigue. Il est donc important de contrôler cette gigue d'autant plus que ces types de circuits sont en général suffisamment complexes pour que des effets de variation de procédé, de tension ou température apparaissent entre les blocs rendant l'impact de la gigue globale encore plus important.

Des solutions globales existent s'appuyant sur une homogénéisation des structures d'arbre d'horloges afin de combattre toutes les irrégularités responsables d'une augmentation de la gigue globale. La plus connue de ces solutions est l'arbre en H qui permet d'obtenir une régularité quasi parfaite de l'arbre d'horloge [43]. L'inconvénient principal de cette solution reste le coût en surface et en efficacité énergétique. En effet, pour la surface, l'arbre doit couvrir tout le circuit et utilise donc de précieuses ressources de routage. Coté efficacité énergétique, la capacité intrinsèque d'une telle structure est aussi très grande augmentant ainsi l'énergie nécessaire à sa commutation.

Une solution alternative est d'utiliser une construction de type arbre synthétisé dite CTS (Figure 57) afin de relier les blocs entre eux. Cette structure doit être dynamique afin de compenser les variations de procédés, de tension et température entre les blocs. La solution hibiskew [44] permet de répondre à ce besoin. Dans cette solution, l'horloge est reliée par un simple fil entre le générateur central et le bloc IP. La structure est donc fortement hétérogène et doit être compensée afin d'atteindre les objectifs de gigue demandés. Cette compensation s'effectue par un contrôleur additionnel placé en sortie du générateur d'horloge. Un mécanisme de rétroaction est implémenté afin de fournir au contrôleur la date d'arrivée de l'horloge. Le contrôleur a la charge d'aligner les dates d'arrivée en mesurant le décalage de phase des horloges de retour et en ajustant les lignes à retard (µPDL) positionnées sur le fil d'horloge.

Dans un premier temps, il est nécessaire de calibrer le chemin de retour. Nous utilisons pour cela des lignes à retard programmables qui nous permettent d'équilibrer les chemins de retour entre les 2 blocs. Une horloge basse fréquence est générée par le contrôleur de gigue (fil orange de la Figure 58) et il s'assure que les fronts d'horloge arrivent en phase et simultanément après avoir traversé les deux blocs en vert dans la Figure 58.

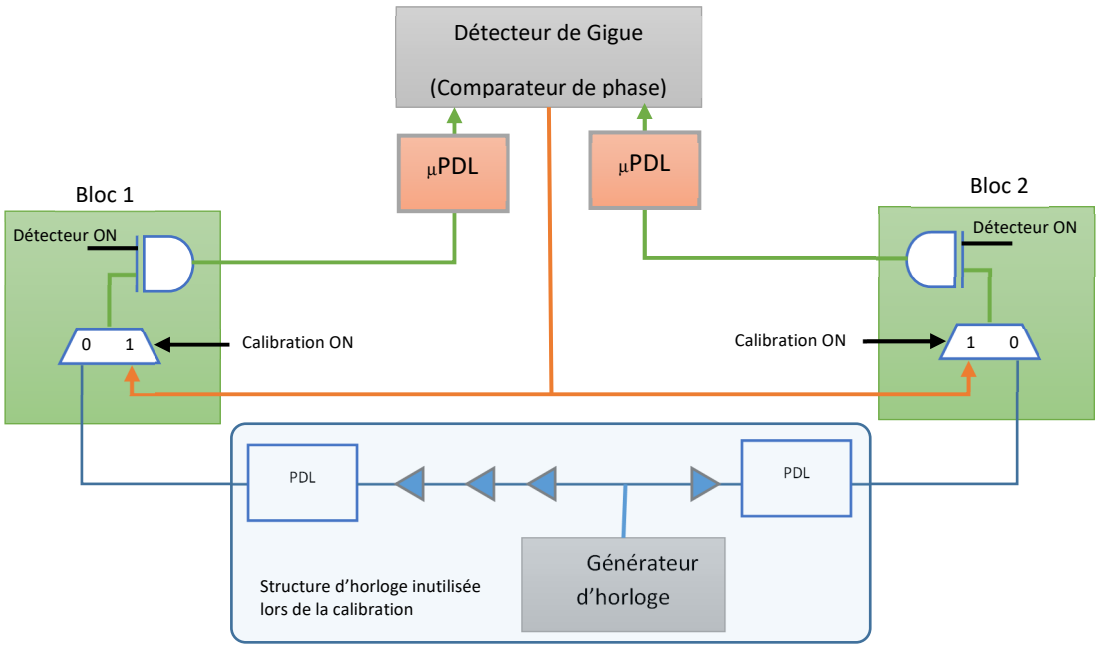


Figure 58 : Calibration de la gigue

Un fois la calibration terminée, nous pouvons garantir que la mesure du front d'horloge à l'entrée de chaque bloc sera cohérente. Nous pouvons donc basculer en mode fonctionnel pour mesurer le décalage de la structure d'horloge et essayer de le compenser.

La mesure du décalage se fait par paire de blocs. Afin d'aligner tous les blocs entre eux, il est donc nécessaire de définir un bloc IP de référence. Par la suite, chaque bloc IP sera aligné par rapport à la référence.

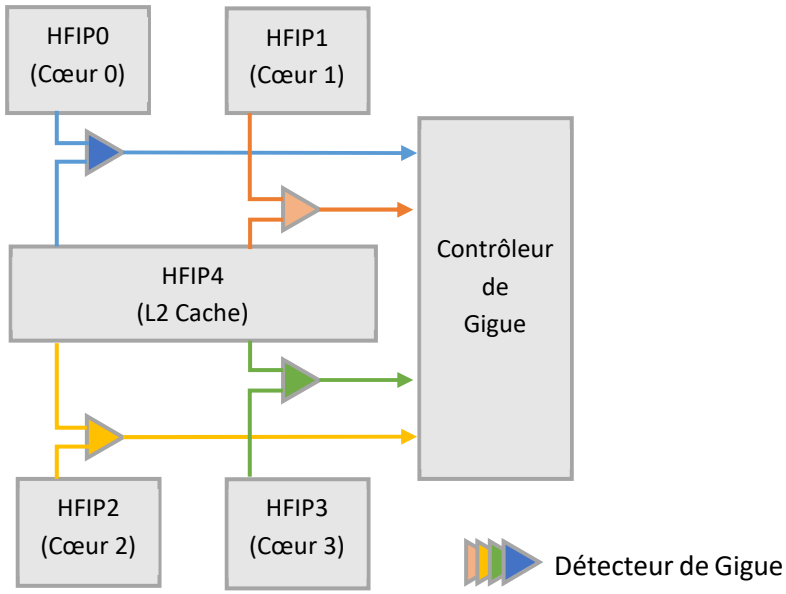


Figure 59 : Exemple de contrôle de gigue pour un processeur 4-coeurs

Nous avons implémenté cette solution sur un processeur 4 cœurs. La mémoire cache de niveau 2 est utilisée comme bloc IP de référence. Le contrôleur de gigue alignera donc successivement les cœurs 1 à 4 en gardant la mémoire cache comme référence. La solution simulée utilise une technologie CMOS032LP de STMicroelectronics. Elle nous montre un gain de gigue mais surtout une gigue globale proche de 10ps. Cette solution est proche des solutions avec un arbre en H mais avec une efficacité énergétique bien meilleure grâce à la structure synthétisée [45].

Le tableau suivant représente les valeurs de gigue extraites lors de la simulation rétro-annotée de notre circuit sous une tension de 1V et dans le pire cas des procédés de fabrication. Les résultats montrent un alignement des gigues entre les différentes IPs. Par exemple, la gigue entre le coeur1 et le coeur2 augmente mais reste dans la spécification initiale des 10ps.

Pair	Gigue Initiale	Gigue après compensation
CORE0 vs CORE1	26ps	11ps
CORE1 vs CORE2	7ps	10ps
CORE2 vs CORE3	33ps	4ps
CORE3 vs CORE4	38ps	10ps
CORE4 vs CACHE	17ps	7ps

Tableau 5 : Gigue avant et après compensation

IV. 2. …vers la disparition de l'horloge

Le calcul synchrone par construction produit un événement par cycle d'horloge. A contrario, la conception basée sur événements réduit considérablement ce nombre. Dans ce cas, seul l'événement utile au calcul est généré ce qui permet de minimiser leur nombre. Cette réduction est importante pour l'efficacité énergétique et la disparition de l'arbre d'horloge induite par cette technique de conception apporte un gain supplémentaire.

Les premiers travaux, présentés dans les sections suivantes, se sont portés sur la réalisation d'une plateforme asynchrone et la création d'un lien série asynchrone (ou basé sur événements) permettant de relier des blocs IP distants entre eux. Ces travaux sont restés en lien avec les capteurs présentés dans le Chapitre III. Nous pensons en effet que le nombre de capteurs présents sur un circuit ainsi que leur granularité vont augmenter rendant indispensable la création d'un réseau permettant de relier les capteurs et d'analyser leurs données.

IV. 2. 1. Une plateforme de caractérisation de la technologie

Lors du développement de plateformes de conception sur un nouveau nœud technologique, il est important de caractériser les paramètres clés des éléments de la plateforme. Ces paramètres de performance clés sont généralement la surface, la fréquence maximale et la consommation dynamique/statique pour des procédés de fabrication donnés. La mesure de ces paramètres est donc nécessaire pour valider l'alignement théorique des performances fournies par les cartes modèles et la réalité du silicium. Actuellement, il existe un certain nombre d'approches différentes permettant d'évaluer cet alignement, chacune ayant ses propres avantages et limitations. Une telle caractérisation sur silicium est possible grâce à une grande variété de circuits et de mesures *in silico*. Par exemple, la caractérisation de la fréquence maximale peut être évaluée à l'aide de l'approche classique des oscillateurs d'anneau, d'un bloc de conception numérique de référence ou d'un bloc IP complet au niveau du produit (système-sur-puce). L'approche classique de l'oscillateur en anneau reste un bon moyen pour comparer l'alignement entre la CAO et le silicium. Elle ne permet pas de caractériser un circuit en conditions réelles où la nature aléatoire du placement et du routage peut impacter fortement la performance. La solution retenue alors est d'utiliser des blocs IP de référence comme un cœur de processeur mais la montée en fréquence de ce type de bloc rend l'intégration dans un circuit de test complexe et coûteuse. Cette complexité est également accrue compte tenu des besoins de mener une corrélation entre outils CAO et mesures sur silicium sur une très large plage de tensions et de températures.

Lors de ces travaux [46], nous avons introduit et validé une nouvelle méthode de corrélation des paramètres des procédés de fabrication, basée sur un circuit de test MTAM16 utilisant une technologie de conception asynchrone QDI (*Quasi-Delay Insensitive*).

IV. 2. 1. 1 Le circuit de mesures MTAM16

L'architecture de la puce de surveillance s'articule autour de microcontrôleur asynchrone TAM16 de la société Tiempo associé à une RAM, une ROM, un lien série RS232 pour communiquer avec un ordinateur hôte et un GPIO utilisé pour fournir des informations d'état et sélectionner les exécutions du programme de surveillance exécuté sur le contrôleur.

Afin de préparer les travaux sur les capteurs embarqués vue au chapitre III, nous avons aussi équipé le circuit d'une liaison en série insensible aux délais pour communiquer avec ces capteurs de surveillance.

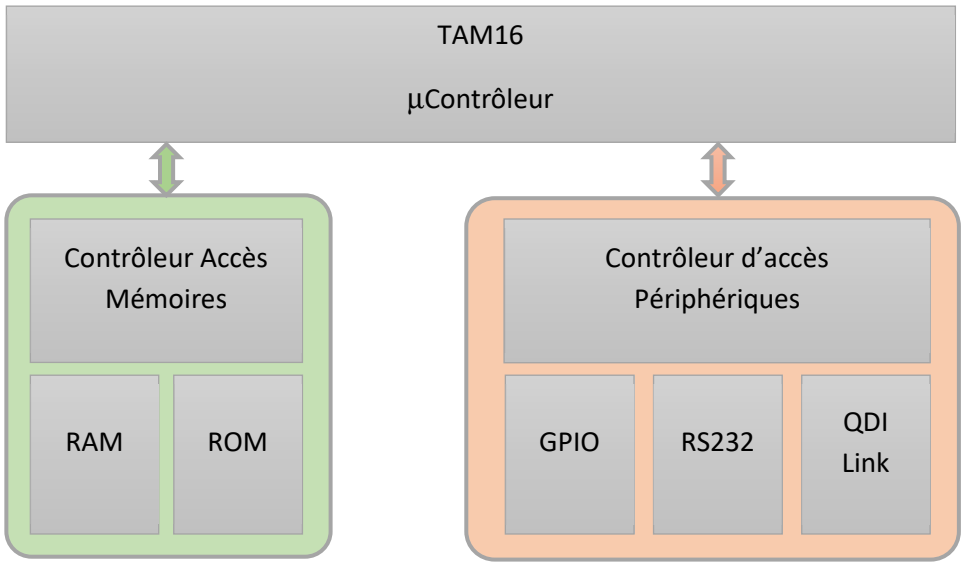


Figure 60 : Schéma Bloc du circuit de mesures MTAM16

Tous les composants du circuit de test, y compris les deux mémoires, utilisent une conception quasi-insensible aux délais. Ce choix est motivé par la nécessité d'utiliser exclusivement des cellules standards et de ne pas s'appuyer sur des blocs qui ne sont pas encore disponibles dans un nouveau procédé de fabrication. La cellule standard étant généralement le premier élément de la plateforme technologique, il nous semblait important d'utiliser uniquement ce type de cellules afin de bénéficier d'un accès rapide au silicium. Notre circuit de test représente environ 300 K portes logiques équivalentes.

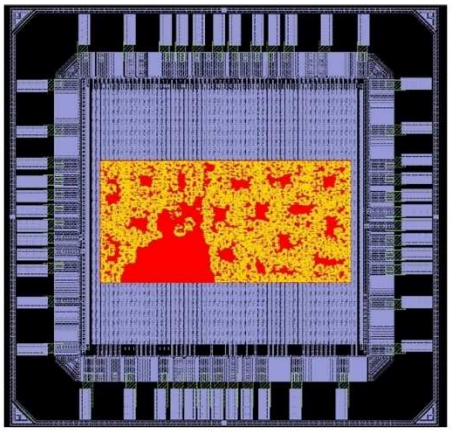


Figure 61 : Implémentation du circuit de mesures MTAM16

Les capteurs implémentés dans cette application sont exclusivement faits de programmes courts exécutant quelques instructions sur le TAM16 en boucle et qui sont représentatifs d'une activité numérique. Étant donné la variété des instructions qui sont utilisées, de nombreux chemins logiques différents sont exercés au cours de la caractérisation des performances, donnant une large couverture des chemins possibles. Ces mesures fournissent donc un moyen de calculer la fréquence et la consommation caractéristique d'un bloc de logique numérique.

IV. 2. 1. 1 Résultats de la caractérisation

Afin de valider notre approche, nous avons fabriqué le circuit sur deux technologies de STMicroelectronics à savoir le 32nm et sa variante rétrécie en 28nm. L'approche quasi insensible aux délais nous a permis de concevoir le circuit en 32nm et de le fabriquer directement sur les 2 nœuds technologiques sans avoir à retoucher l'implémentation de notre circuit.

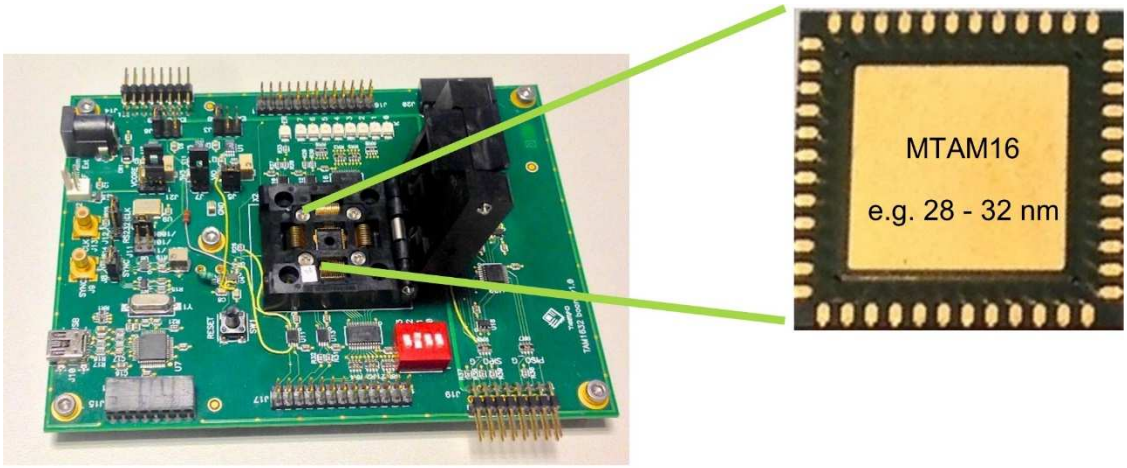


Figure 62 : Carte de caractérisation du circuit de mesures MTAM16

La mémoire ROM intégrée à la puce nous a permis d'intégrer un programme de test (BIST) permettant l'autotest de toutes les parties du dispositif de surveillance. Le circuit contient aussi un mécanisme simple pour charger les programmes de test écrit en C dans la mémoire de la puce via une interface RS232. De cette façon, nous avons pu charger les programmes de caractérisation de silicium directement à partir d'un PC connecté via l'USB au port RS232 de la puce. Une fois le circuit disponible, les caractéristiques fondamentales de la conception sans horloge facilitent grandement la caractérisation des performances du silicium. Tout d'abord, la robustesse de la conception sur une très large plage de tensions permet des mesures de performance pour des algorithmes de calcul complexes sur cette large plage de tension. Nous retrouvons ici la simplicité de la caractérisation de l'oscillateur d'anneau mais avec un circuit plus représentatif des blocs IP utilisés dans les produits. Cette robustesse permet également des mesures de courant et de performance sur toute la plage de tensions ce qui n'est pas le cas dans un circuit synchrone ou le couple tension/fréquence doit correspondre à un point de validation en analyse temporelle finale.

Afin de vérifier la corrélation de notre programme de test avec les performances du silicium, nous avons comparé les mesures de silicium par rapport aux moniteurs de performance connus disponibles sur les mêmes plaquettes de silicium. Comme nous le voyons sur la Figure 63, les résultats nous montrent que les performances du MTAM16 mesurés à l'aide d'une solution de test embarquée BIST (*Build in Self-Test*) sont conformes aux attentes de centrage du silicium extraites grâce au capteur, ce qui confirme la validité de notre approche.

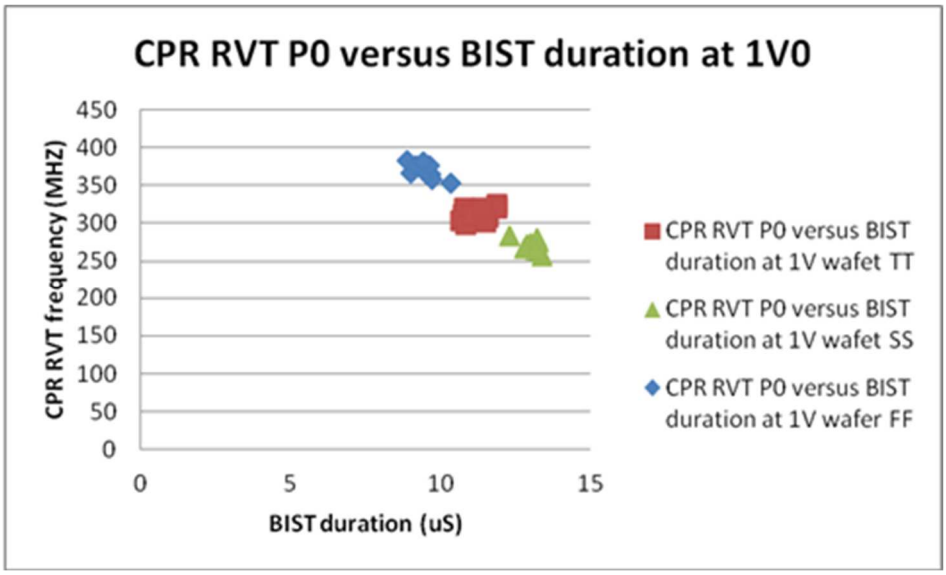


Figure 63 : Corrélation entre TAM16 et les capteurs de performance silicium

Nous pouvons ainsi nous servir du MTAM16 pour mesurer les principales caractéristiques du procédé de fabrication. Nous présentons ici deux mesures classiques de performance (Figure 64) et de consommation dynamique (Figure 65) mais bien d'autres caractéristiques peuvent être extraites à l'aide de ce circuit. Dans ces exemples, il est intéressant de noter l'effet du rétrécissement du procédé 32nm vers du 28nm.

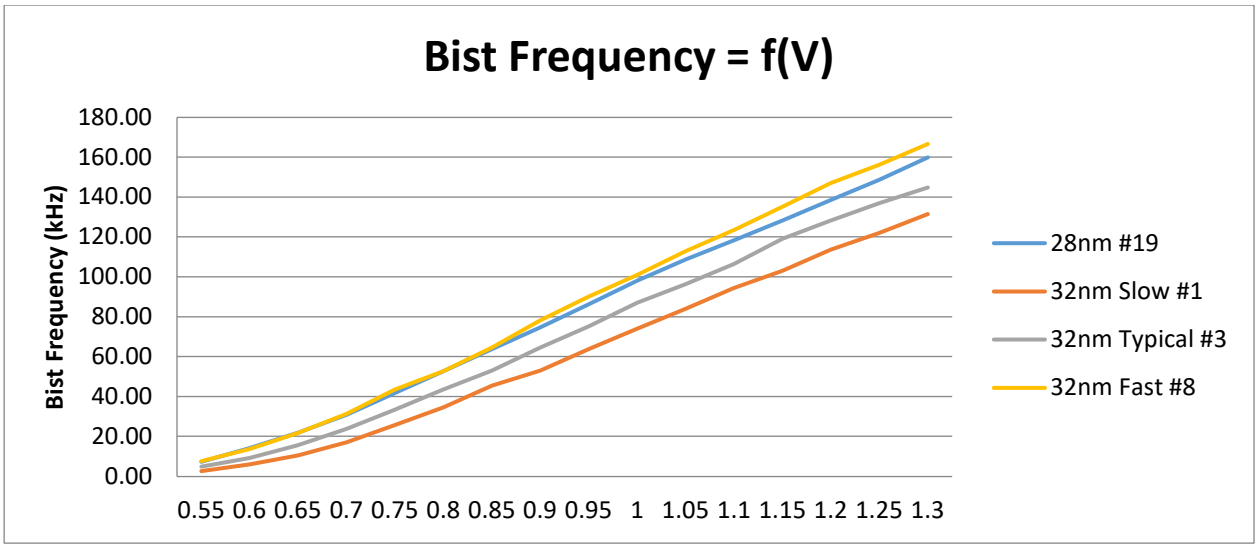


Figure 64 : mesure de la performance du Silicium à l'aide du MTAM16

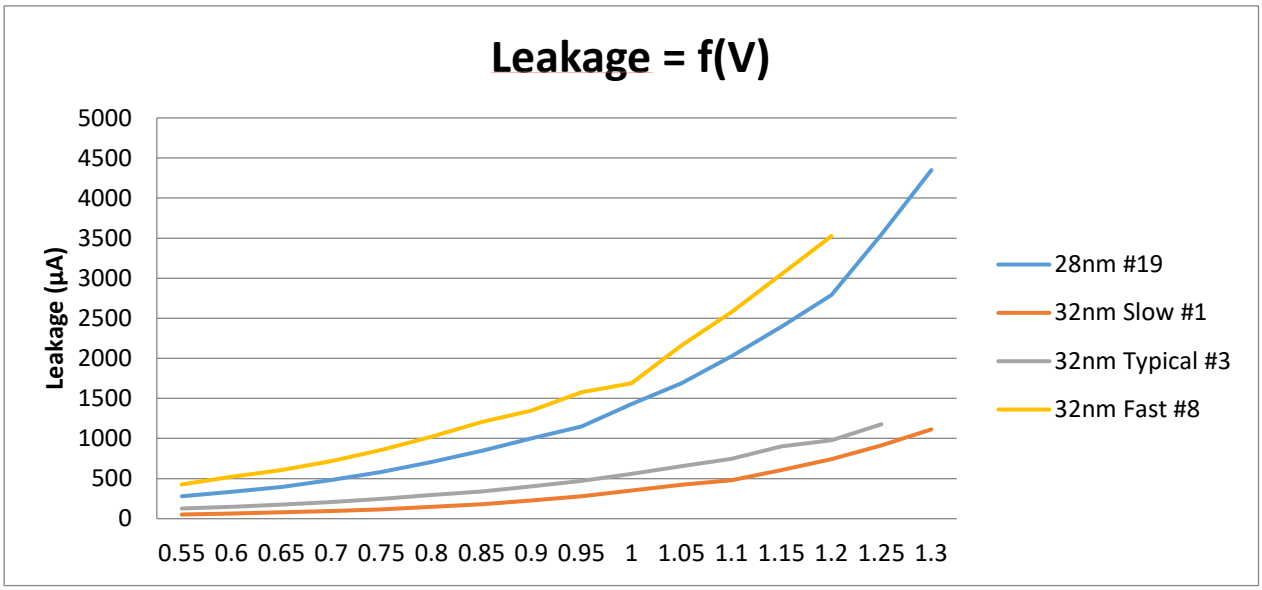


Figure 65 : Mesure de la consommation statique à l'aide du MTAM16

IV. 2. 2. Un réseau de capteurs distribués

L'amélioration de l'efficacité énergétique passe aussi par une connaissance plus fine des conditions extérieures. Il est important de connaitre le centrage du procédé de fabrication, la température ou la tension de fonctionnement afin de s'adapter au mieux. Fournir aux architectes SOC les moyens d'intégrer des capteurs sur la puce [47] avec une simplification des exigences d'intégration et de vérification de la conception représente un vrai avantage compétitif de la solution. Ces travaux se sont donc focalisés sur la réalisation d'un réseau d'accès aux capteurs tolérants aux processus tension/température et fonctionnant sans horloge. Ce réseau sera appelé ASPIC dans les parties suivantes.

IV. 2. 2. 1 Architecture du réseau de capteurs

La construction de l'architecture a été pilotée par deux contraintes : une intégration aisée aux système standards existants et la nécessité d'avoir une solution simple qui permette de connecter les capteurs entres eux et ce quel que soit la localisation du capteur dans la puce. Le capteur doit pouvoir être implémenté de la même façon qu'il soit proche du processeur ou dans un autre domaine d'alimentation, voire sur une autre puce dans le cas de l'intégration 3D.

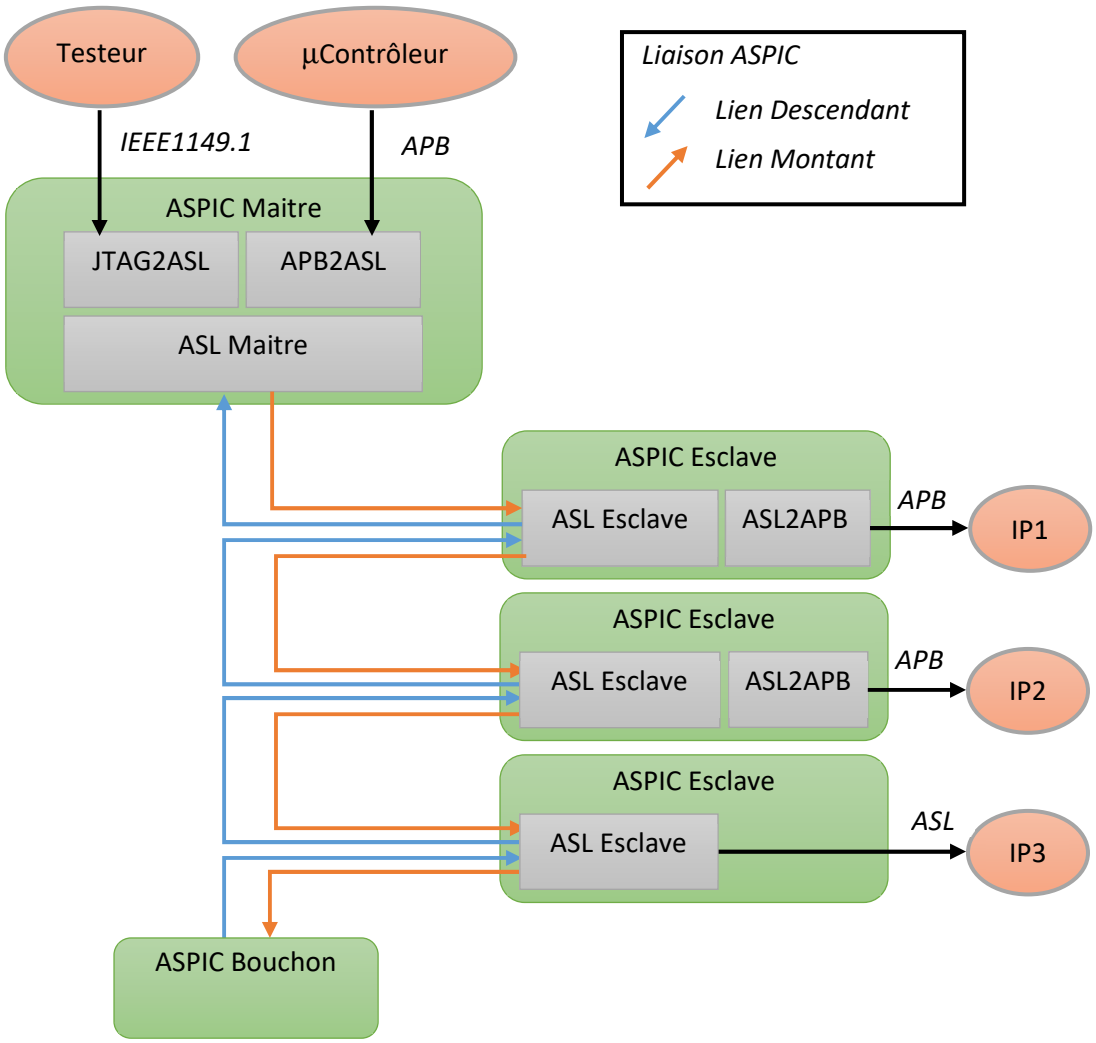


Figure 66 : Schéma Bloc du réseau ASPIC

Tel que décrit dans la Figure 66, l'architecture est composée de plusieurs blocs chargés de connecter le contrôleur maître (ASPIC Maitre) via le lien asynchrone série aux contrôleurs esclaves *(*ASPIC Esclave*)* qui sont eux-mêmes connectés aux différents capteurs (IP asynchrone ou synchrone).Afin de répondre à la première contrainte sur la facilité d'intégration, le bus APB standard (*ARM: Advanced Peripheral*) [48] est adopté aux limites du réseau série asynchrone ainsi qu'une interface IEEE1149.1 pour accéder aux capteurs en mode de test.

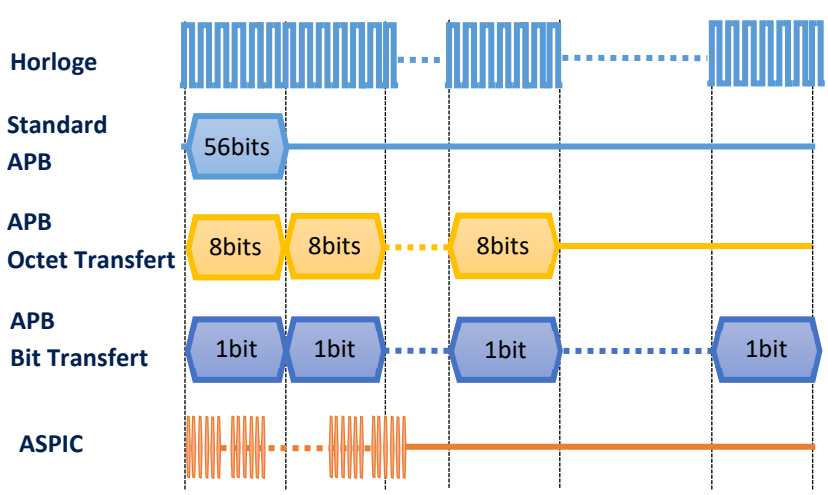
Le fonctionnement du réseau est donc le suivant. Le bloc ASPIC Maitre reçoit les commandes envoyées par le microcontrôleur par l'intermédiaire du bus APB et les transmet au lien de communication en série. Ce bloc est composé de deux parties : une interface esclave APB qui qui décode les signaux APB et fournit les commandes à un deuxième block ASL Maitre asynchrone qui communique avec le lien sériel. Du côté du capteur, un bloc ASPIC esclave reçoit les commandes via la communication asynchrone et les transmet au(x) capteur(s) par l'intermédiaire du bus APB. A nouveau, ce composant est composé de deux

parties : une interface ASL esclave qui décode les signaux asynchrones et pilote le deuxième bloc constitué d'une interface Maitre APB pour piloter le capteur synchrone. Si le capteur est asynchrone, le bloc ASPIC esclave peut se réduire à un seul composant Async esclave (comme illustré dans la Figure 66 avec IP3). Le bloc final de la Figure 66 est connecté au lien de communication série qui communique avec un circuit bouchon, afin de fermer le lien série et d'empêcher de bloquer le système dans le cas où un message traverserait tous les capteurs connectés sans être traité. Ce composant spécifique absorbe donc les messages non traités et renvoie les messages d'erreurs associés.

La structure de la liaison ASPIC est une structure simple en guirlande utilisant une liaison descendante maître pour transmettre des informations aux capteurs et une liaison montante piloté par les esclaves qui permet de récupérer les informations des capteurs et, également, de remonter des messages d'erreurs. Ce choix de liaison est motivé par la seconde contrainte d'intégration du réseau dans d'un système complet. Il autorise l'adoption d'un protocole de communication inspiré du slogan « branche et profite ». Cette sérialisation offre un autre avantage pour les systèmes sur puce avec plusieurs domaines de tension. Elle permet de réduire jusqu'à 10 fois le nombre de signaux nécessitant une adaptation de tension par rapport au bus standard avec transferts de données parallèles [4]. Notre protocole de série à bits unique (ASPIC) simplifie donc grandement la gestion complexe des contraintes de synchronisation requises lors de l'intégration de la conception pour les signaux traversant les domaines physiques et de tension.

Un inconvénient potentiel de la réduction de la bande passante (due à la sérialisation et à la resynchronisation) est surmonté en rendant la solution ASPIC indépendante aux délais et temps de propagation (dans les portes et les interconnexions) grâce à l'utilisation de la logique asynchrone QDI. La nature auto-séquencée de la solution permet au lien d'exécuter une communication série à grande vitesse indépendamment de la fréquence d'horloge des différents blocs IP qui contrôlent l'échange parallèle de données aux interfaces APB. Ceci est illustré par l'exemple (Figure 67) d'un lien série ASPIC qui supporte une bande passante équivalente à un protocole APB4 de 50 Mhz.

Fréquence : Source = Destination = 50Mhz
Largeur du bus de données : 56bits (APB4 transaction)



Bande Passante	#Conv. de tension	Analyse Temporelle
466Mb/s	58	Yes
66Mb/s	10	Yes
8Mb/s	3	No
761Mb/s @ 1.1V 233Mb/s @ 0.6V	3	No

Figure 67 : ASPIC vs APB Performance en bande passante.

IV. 2. 2. 2 Le Protocole ASPIC

Au niveau physique, le lien asynchrone utilise un protocole de communication à quatre phases, insensible aux délais. Les données sont codées en double rails (deux fils) permettant d'envoyer à la fois l'encodage binaire et l'information de requête. Un signal supplémentaire d'acquittement assure le séquençage de la communication (voir Figure 68). Par conséquent, six fils sont utilisés pour soutenir la communication dans les deux directions, maître à esclave et esclave à maitre.

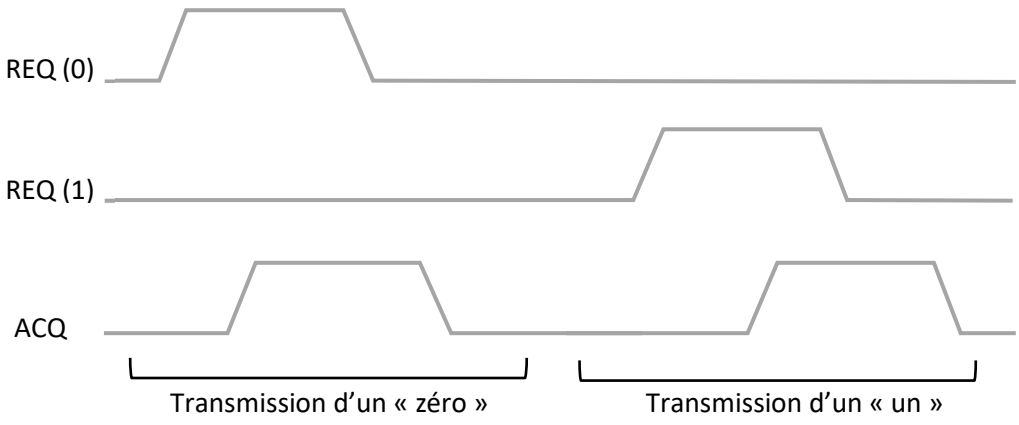


Figure 68 : Lien Physique de communication ASPIC.

Au niveau logique, le réseau sériel asynchrone prend en charge deux types de messages. Le contrôleur peut demander à écrire dans le registre d'un capteur (envoi d'un message d'écriture) ou il peut demander à lire le registre d'un capteur (envoi d'un message de lecture). Les messages sont constitués d'une série de bits qui ont un format prédéfini et fixe, organisé en octets

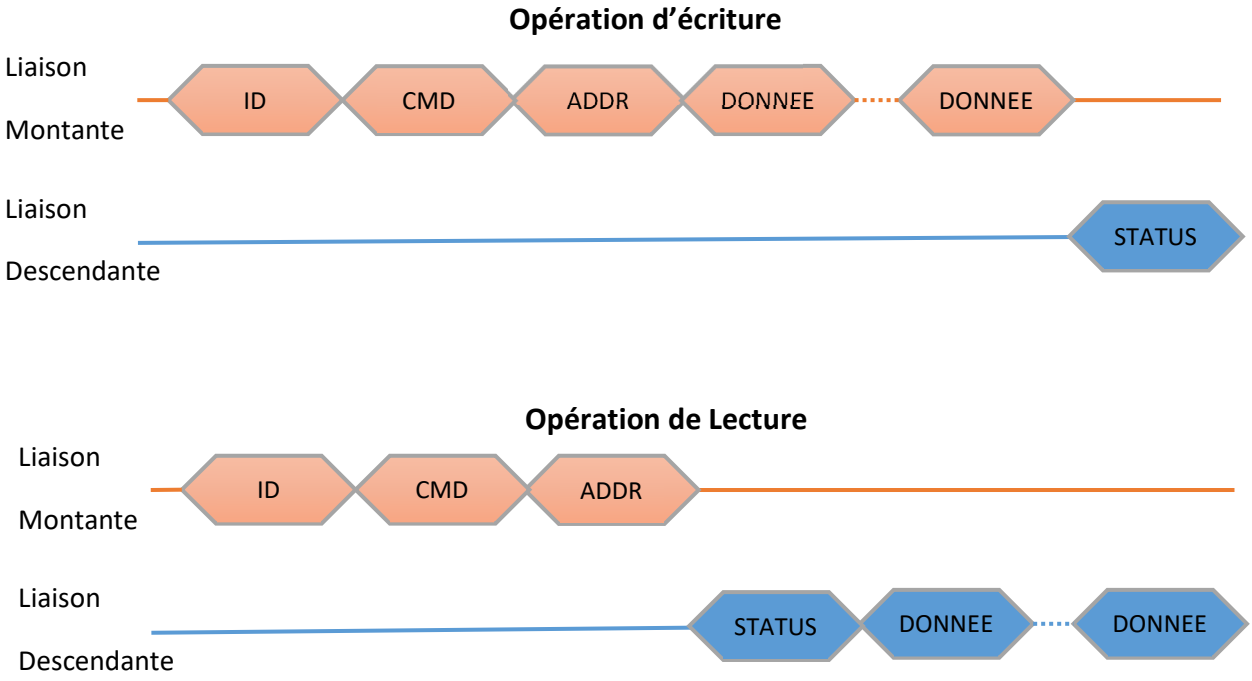


Figure 69 : Protocol De communication ASPIC.

Tel que décrit dans la Figure 69, le message d'écriture est composé d'un octet d'identification qui spécifie la position du capteur dans la chaine, d'un octets CMD fournissant les paramètres de commande, d'un octet ADDR fournissant l'adresse des données à accéder dans le capteur, et enfin d'un à quatre octets qui sont les données à écrire. En réponse à ce message, un octet de statut est renvoyé par le bloc esclave pour informer le maitre de la bonne réception du message. L'octets CMD est également utilisé pour spécifier le nombre d'octets de données contenu dans le message. Si nous nous intéressons maintenant au message de lecture de la Figure 69, nous retrouvons plus ou moins la même structure. Il est composé d'un octet d'identification spécifiant la position du capteur, suivi d'un octet CMD pour les paramètres de commande, et d'un octet ADDR fournissant l'adresse à lire dans les registres du capteur. En réponse à ce message, l'esclave commence par renvoyer l'octet de STATUS pour informer si la commande est réussie ou non. Si la commande est réussie, les octets des données lues sont renvoyés par l'Esclave. Si la commande n'est pas réussie, aucune donnée ne suit et la communication s'arrête.

Cette structure de message, nous permet de définir l'algorithme suivant pour les stations esclave du lien ASPIC. Étant donné que l'octet d'identification informe sur la position du bloc IP esclave dans la chaîne, les esclaves comparent l'ID à zéro. Si l'identifiant est nul, l'esclave traite le message et renvoi en fonction de la commande l'état et/ou les données. Si l'identifiant est non- nul, l'esclaves décrémente l'identifiant de 1 et transmet le message avec l'identifiant décrémenté. Il attend ensuite que le statut et les données reviennent d'un esclave plus bas dans la chaine afin de les transmettre au maître. Par conséquent, pour atteindre le premier esclave dans la chaine, l'ID de champ doit contenir la valeur « 0 ». Ce mécanisme nous permet de rendre le matériel des esclaves identique et permet d'ajouter, enlever, échanger des esclaves dans la chaîne en guirlande très facilement et sans impact sur le matériel du système-sur-puce car seuls les identifiants auront besoin d'être mis à jour dans le micrologiciel.

IV. 2. 2. 3 La station ASPIC Esclave

Une caractéristique clé de la solution proposée est qu'il tente de répondre au slogan « branchez et profitez ». Pour nos capteurs, cela signifie qu'ils peuvent être situés n'importe où dans le Système sur Puce, puis « branché » sur le lien et il suffit pour eux d'être connectés au lien pour être fonctionnel sans avoir à retoucher ou optimiser le lien pour supporter ce nouveau capteur. Ceci a été rendu possible grâce aux choix techniques suivants utilisés pour concevoir ce capteur (Figure 70).

- Le lien série est sans horloge, il n'est donc pas nécessaire de gérer la distribution d'horloge ou le franchissement de domaine d'horloge au niveau supérieur, et les passages de domaine de tension sont facilement implémentés à l'aide uniquement de cellules de décalages de tension (*level shifters*).
- Le lien de série est insensible aux délais, il n'y a donc pas de contrainte de temps spécifique pour les stations durant les étapes de placement et de routage.
- Les interfaces des capteurs sont des blocs logiques asynchrones insensibles aux délais, appelés stations asynchrones, offrant une solution simple pour leur intégration puisqu'il n'est pas nécessaire de vérifier leurs contraintes temporelles durant l'intégration.
- Les capteurs sont connectés aux stations asynchrones à l'aide d'une interface APB standard qui permet aux architectes des système-sur-puce de réutiliser les capteurs existants.

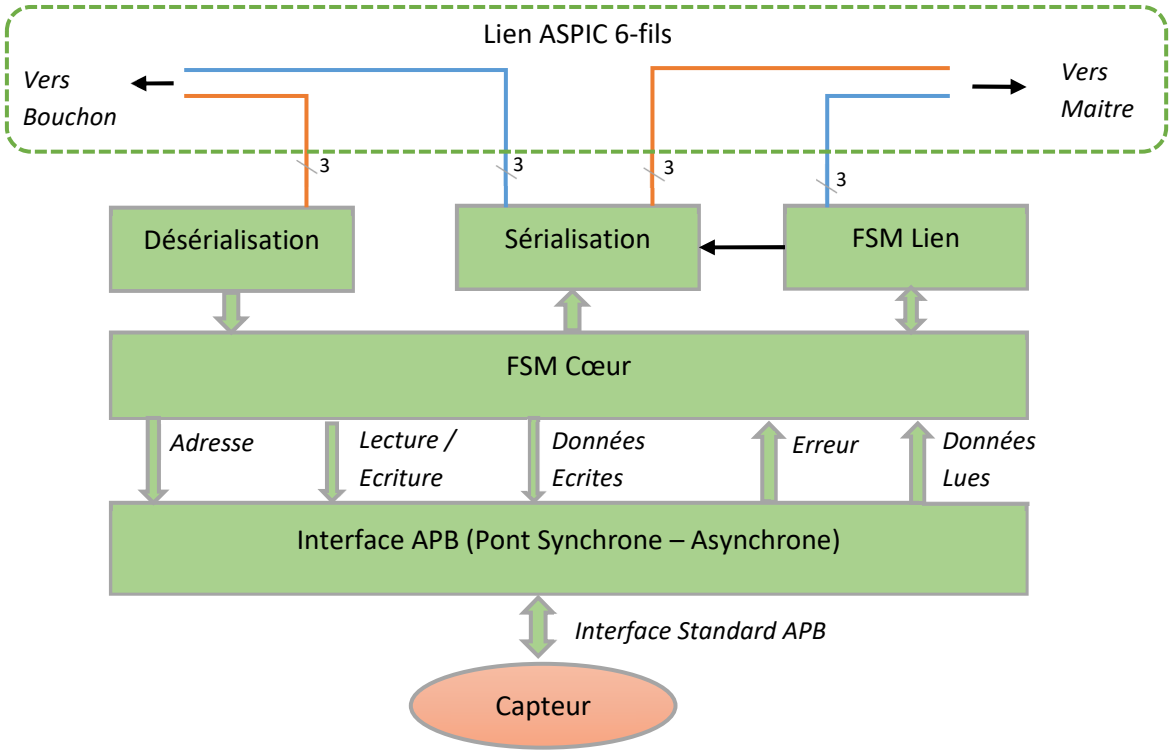


Figure 70 : Schéma bloc de la station ASPIC Esclave.

Les stations ASPIC esclaves sont constituées d'interfaces APB ainsi que d'interfaces synchrones/asynchrones qui assurent une intégration et une réutilisation facile des adresses IP existantes. En effet, le lien entre les différents capteurs peut être vu comme une passerelle sur laquelle nous pouvons insérer de manière transparente un nouveau capteur.

IV. 2. 2. 4 Le Circuit de Test TACO

La puce TACO (Figure 71) est une extension du circuit MTAM16 présenté dans la section précédente. Nous sommes repartis de la base du MTAM16 pour construire le microcontrôleur qui nous servira à piloter le réseau. L'intérêt principal réside dans le périphérique série qui nous permet de connecter directement le MTAM16 au réseau. Le circuit TACO se compose donc d'un sous-système de microcontrôleur asynchrone, ainsi que d'un réseau de capteurs sur puce. Le réseau de capteurs contient 15 blocs de capteurs monitorant des procédés de fabrication différents (Figure 72), basés sur une architecture extensible, capable de répondre à n'importe quel nombre ou classe de capteurs (tension/thermique/procédé de fabrication). Le réseau ASPIC, le microcontrôleur et les différents capteurs ont été globalement conçus à l'aide des méthodologies standards de conception de SoC, mais nécessite le recours à un outil de synthèse dédié à l'asynchrone de la société Tiempo et à une bibliothèque de cellules standards spécifiques pour la conception de circuits sans horloge [49]. Afin de permettre l'accès à l'équipement de test en production, une interface JTAG IEEE1149.1 [50] est utilisée en plus de l'interface RS232 du microcontrôleur.

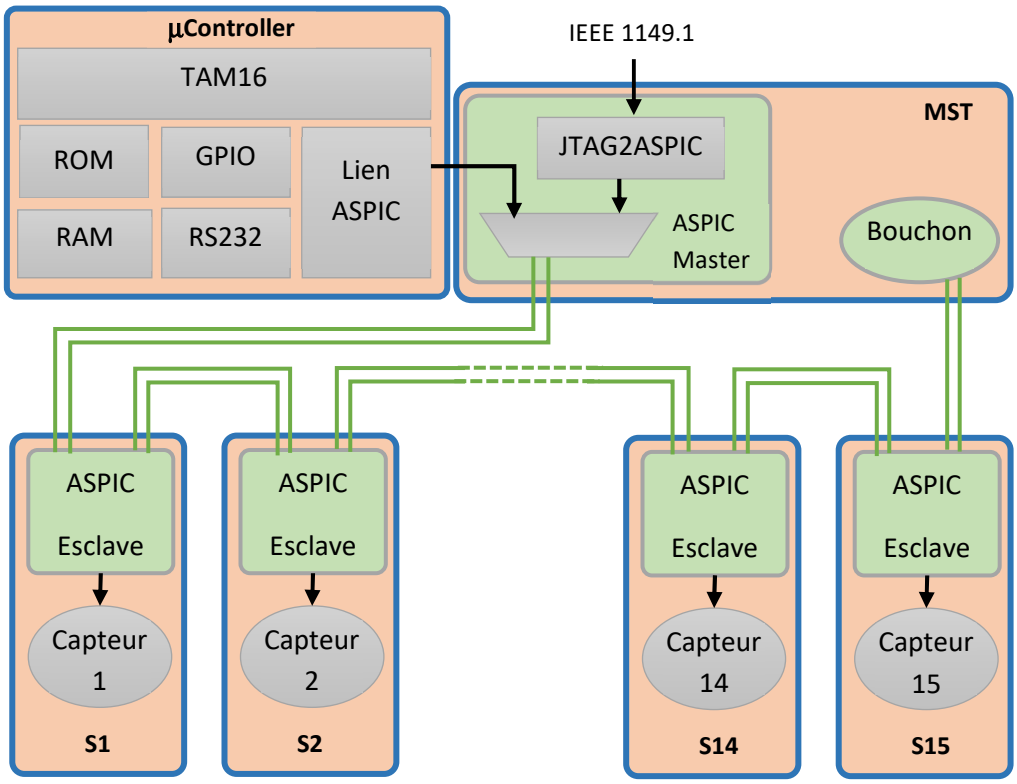
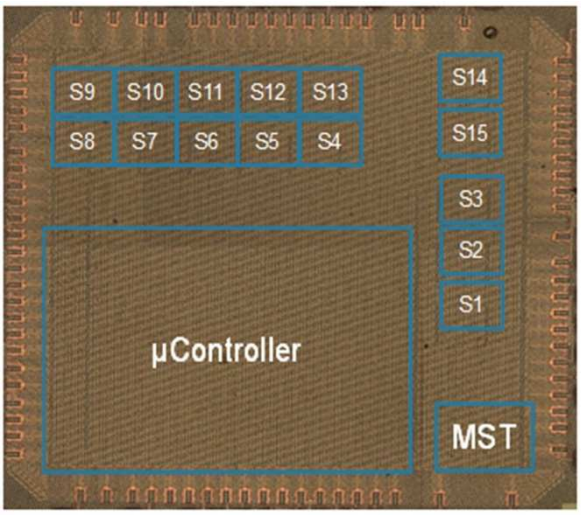


Figure 71 : Schéma block du circuit TACO et du réseau ASPIC (en vert).

Le circuit a été conçu et fabriqué en technologie STMicroelectronics 28nm UTBB FDSOI (voir chapitre III). Les cadres bleus de la Figure 71 et Figure 72 représentent les blocs qui ont été implémentés hiérarchiquement afin de valider notre hypothèse de branchement. Une fois implémentés, nos différents blocs ont juste été assemblés dans le circuit TACO sans aucune vérification temporelle. Seule une vérification des règles de dessin a été effectuée afin de s'assurer de la qualité de fabrication de notre circuit.



	Détail Circuit
Taille	**1,5 mm²**
Nb de portes	**450K**
Bande Passante	**233Mb/s @ 0.6V** **761Mb/s @ 0.6V**
Fréquence	**100Mhz**

Figure 72 : Photographie du circuit et ses caractéristiques.

IV. 2. 2. 5 Test et Validation du Circuit TACO

La société Tiempo a conçu une carte qui intègre les ressources nécessaires pour exécuter les routines de validation. Cette carte intègre les alimentations électriques, les potentiomètres pour varier les tensions d'alimentation, une interface hôte basée sur un pont de communication USB à RS232, des commutateurs pour conduire les ports d'entrée, des LED ou des broches pour surveiller les ports de sortie, et une prise pour placer les puces prototypes. La Figure 73 montre une photo de la carte réalisée.

Cette carte nous a permis de réaliser une première campagne de mesures en utilisant les solutions de BIST intégrées (voir IV. 2. 1. 1) du microcontrôleur MTAM16. Ce BIST est mis en œuvre par le biais de programmes de test intégrés qui exercent le microcontrôleur, les mémoires et les périphériques. En plus de tester le microcontrôleur, la RAM, la ROM et la communication sérielle RS232 à l'aide des programmes BIST, des programmes dédiés ont été écrits en C et téléchargés dans le microcontrôleur afin de tester et de caractériser également le réseau de capteurs.

Figure 73 : Carte de test pour le circuit TACO.

Une seconde campagne a été réalisée en en utilisant de l'équipement de test automatique industriel (ATE) connecté à interface JTAG IEEE1149.1. Aucune modification n'a été apportée par rapport au flot existant grâce à l'utilisation de cette interface standard. Les

résultats silicium (Figure 74) montrent une réponse en fréquence en fonction de la tension pour deux oscillateurs en anneau différents (nous avons mesuré les blocs avec les capteurs 1 et 3, mais tous les capteurs étaient accessibles). Les mesures ont indiqué que pour une plage de tensions allant de 0,6 V à 1,3 V et pour trois températures différentes, -40°C, +25°C et +125°C. Ces résultats démontrent la validité de la solution sans horloge pour accéder aux capteurs distribués par le lien asynchrone.

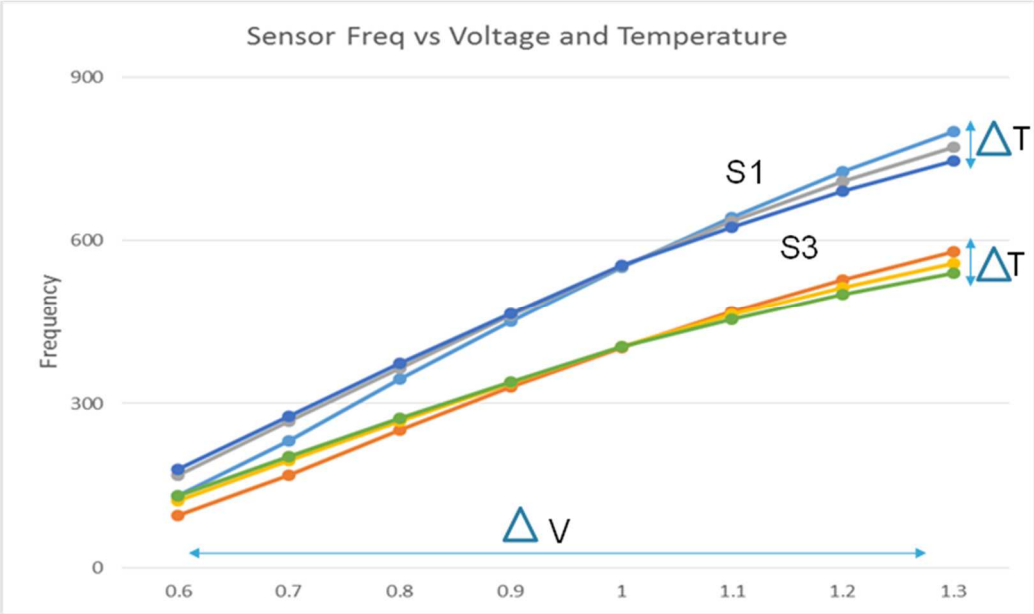


Figure 74 : Mesure silicium des capteurs S1 et S3 pour 3 température -40°C, +25°C and +125°C.

IV. 2. 2. 6 Résultats de la caractérisation

Plusieurs campagnes de mesures et de caractérisation ont été menées. La Figure 75 donne la fréquence par rapport à la tension d'un programme exécutant une boucle sur le microcontrôleur. Comme le microcontrôleur est sans horloge et fonctionne toujours à sa vitesse maximale, ce graphique informe sur la dispersion des trente échantillons testés. En effet, le microcontrôleur sans horloge peut lui-même être considéré comme un moniteur des procédés de fabrication. Nous pouvons le confirmer grâce à la Figure 76 où une corrélation forte entre la fréquence du capteur et la fréquence du microcontrôleur est mise en évidence.

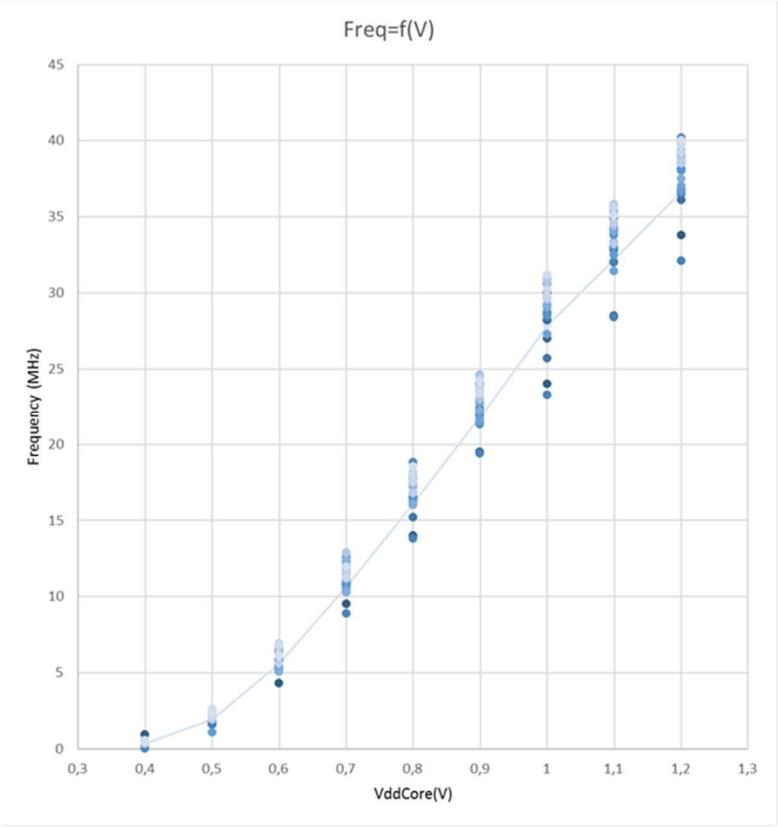


Figure 75 : Mesures Silicium fréquence vs tension pour un programme exécuté en boucle (30 circuits).

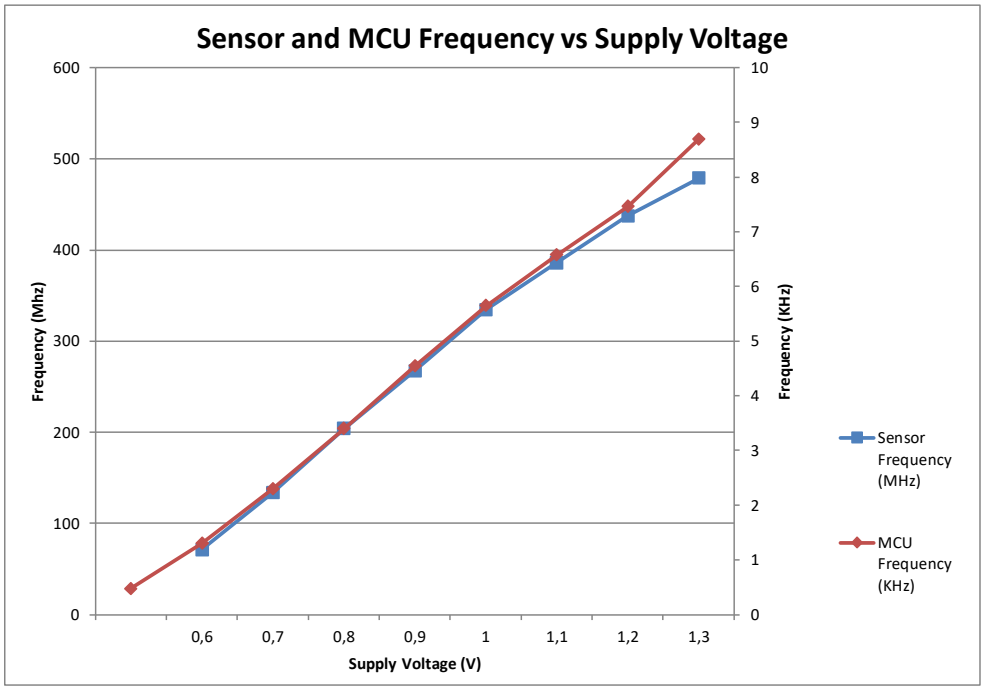


Figure 76 : Comparaison de fréquence entre un capteur et le CPU.

La Figure 77 donne la fréquence du microcontrôleur par rapport au courant statique mesuré à 0,4 Volt illustrant une autre caractéristique intéressante, à savoir la corrélation entre

la vitesse du circuit et son courant de fuite. On soulignera également que la logique asynchrone et le réseau ASPIC sont opérationnels jusqu'à 0,4 Volt.

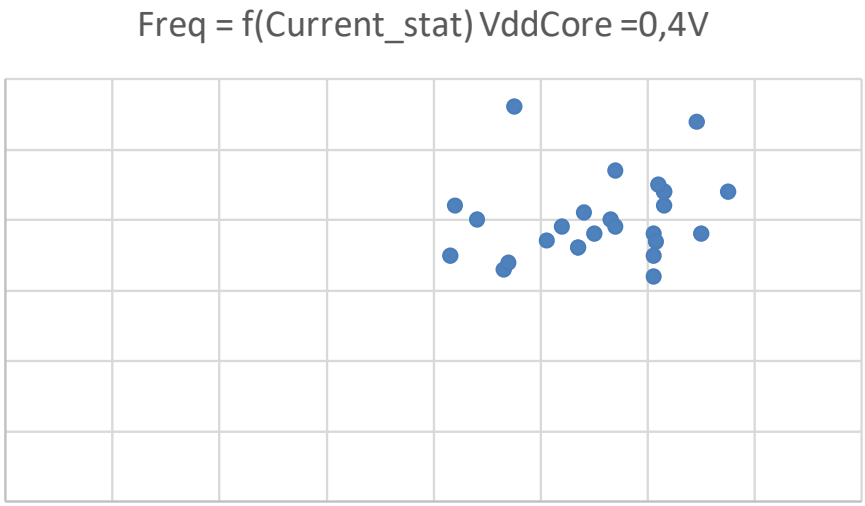


Figure 77 : Comparaison fréquence vs courant statique à 0,4V (Unité arbitraire).

La carte d'application nous a permis de travailler plus localement sur le lien ASPIC pour lequel nous avons cherché à caractériser sa consommation et sa bande passante. Pour ce faire, un programme C a été développé et exécuté dans le microcontrôleur pour accéder aux capteurs 4 et 14. L'énergie mesurée comprend l'exécution des routines logicielles de bas niveau par le microcontrôleur ainsi que l'énergie consommée par la liaison sérielle asynchrone. Il n'a malheureusement pas été possible de mesurer uniquement l'énergie consommée par le lien asynchrone, sans inclure l'énergie consommée par la routine contrôlant le lien et exécutée par le microcontrôleur, ce dernier étant le principal contributeur en termes de consommation. Par conséquent, une transaction de lecture, qui consiste à échanger huit octets, consomme 2,27nJ et 2,74nJ lors de l'accès aux capteurs 4 et 14 respectivement. Cela représente 284pJ et 343pJ par octet, y compris l'activité de microcontrôleur. Pour une transaction d'écriture, échangeant huit octets, la consommation est de 1,92nJ et 2,44nJ pour les capteurs 4 et 14 respectivement, soit 240pJ et 305pJ par octet. Pour toutes ces transactions, la vitesse est d'environ un mégaoctet par seconde à 0,9 Volt (Figure 78) et la consommation actuelle moyenne est d'environ 700µA.

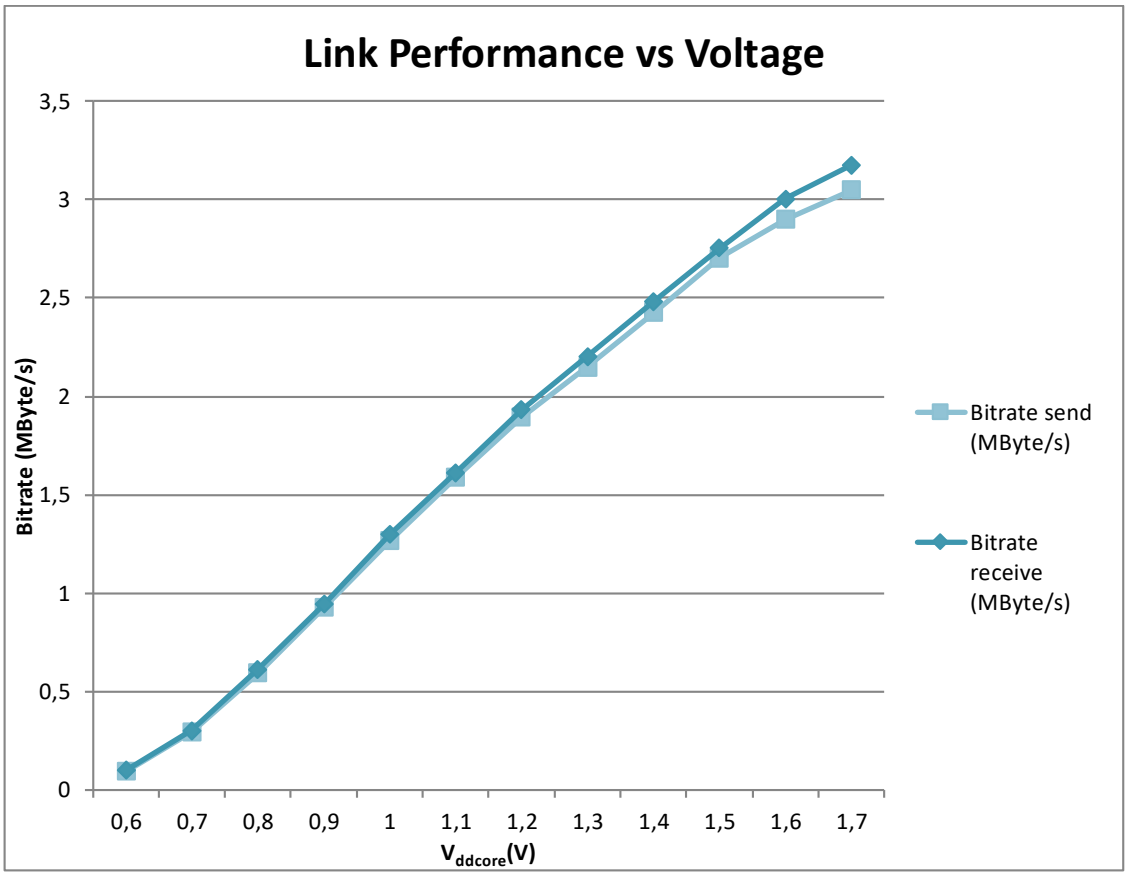


Figure 78 : Performance du lien ASPIC en fonction de la tension d'alimentation.

IV. 2. 2. 7 Conclusion

Ces travaux nous ont permis de développer un réseau nous permettant d'accéder facilement aux capteurs. Cette information est un élément fondamental des solutions auto-adaptatives car elle permet de fermer aisément la boucle de rétroaction. Par rapport au réseau asynchrone série sur puce précédemment proposé comme CHAIN [51], notre solution est parfaitement intégrée dans les flots de conception standards, et entièrement « branchez et profitez » pour faciliter son intégration physique dans les System-sur-Puce. De plus, son interface multistandard APB/JTAG facilite également la connexion avec les interconnexions existantes. Ainsi, ce lien a pu être réutilisé lors des travaux pour le projet européen THINGS2DO dans lequel le réseau a permis de connecter facilement entre eux différentes blocs IP de différents groupes .[52].

L'avantage de la conception sans horloge pour un tel réseau est la simplification de l'intégration lors de la conception et une robustesse accrue par conception vis-à-vis des variations de procédés de fabrication, de tension ou de température [49], [53]. La simplification de la conception inclut le fait que des blocs de capteurs peuvent être ajoutés (position physique, nombre et commande) sans aucun effort de conception supplémentaire (contraintes

de synchronisation, validation de l'interface, validation temporelle). Finalement, l'immunité à la variabilité temporelle (processus, température et tension) lié à l'utilisation de solutions quasi-insensibles aux délais permet l'utilisation de ce réseau sur de larges plages de tension et de température sans ajouter de marges de synchronisation pénalisantes.

En ce qui concerne les perspectives de ces travaux, l'empreinte du lien pourrait être diminuée en utilisant un nombre réduit de fils. En effet, en raison de la nature séquentielle du protocole de liaison, l'utilisation d'un lien bidirectionnel réduirait le nombre de fils à trois au prix de l'ajout d'un contrôleur de direction dans les esclaves. Une autre orientation liée à ce travail qui est actuellement en cours d'analyse [51] est la vérification formelle du lien à l'aide de la méthodologie et du flot développés dans [54]. En effet, la vérification formelle de la justesse fonctionnelle du lien compléterait avantageusement sa robustesse en termes de mise en œuvre, conduisant ainsi à une solution hautement fiable.

IV. 3. Vers un contrôle du spectre électromagnétique

IV. 3. 1. Le spectre électromagnétique

L'élasticité du calcul sur évènements offre une possibilité intéressante lorsque l'on s'intéresse à l'émission électromagnétique de nos circuits. Bien que non directement liée à la consommation, c'est une contrainte importante pour des applications comme l'internet des objets, pour lesquelles l'efficacité énergétique est primordiale.

La possibilité de contrôler finement l'instant du calcul permet d'avoir un effet sur le spectre électromagnétique émis par le circuit. Cette spécificité est particulièrement intéressante dans le cadre de la compatibilité avec les circuits de communication RF car elle permet de positionner les émissions électromagnétiques dans une plage de fréquences qui ne perturbent pas les communications Radiofréquence. Nous allons voir dans les chapitres suivants la solution proposée exploitant les propriétés évènementielles d'une structure micropipeline.

IV. 3. 1. 1 La problématique de la CEM

Les problèmes électromagnétiques sont bien connus des concepteurs de circuits lors de la mise en œuvre d'un bloc analogique et d'un bloc numérique sur le même circuit. Par exemple, un récepteur RF et son contrôleur numérique de bande de base sont souvent associés sur le même circuit. La partie numérique est un contributeur majeur aux émissions électromagnétiques qui peut perturber le fonctionnement de la partie RF [55]. L'une des causes principales est l'horloge des circuits synchrones qui en produisant de fortes impulsions périodiques sur l'alimentation électrique, génère un grand nombre d'harmoniques dans le spectre. Ces émissions, étant capables de rendre un bloc analogique sensible inopérant [56],

ont fait apparaître la compatibilité électromagnétique (CEM) des circuits comme étant un problème supplémentaire à résoudre pour les concepteurs [57]. Les émissions électromagnétiques de différents circuits dans un même environnement créent des interférences électromagnétiques (EMI). En effet, un circuit est généralement classé en deux catégories lorsqu'on parle d'IME. Si les émissions électromagnétiques du circuit polluent l'environnement, le circuit est qualifié d'agresseur. Si le circuit est sensible au champ électromagnétique et susceptible de dysfonctionnement ou d'être détruit, le circuit est classé comme victime.

Ainsi, les concepteurs ont développé différentes techniques pour augmenter l'immunité contre les agressions électromagnétiques. Par exemple, ils protègent les parties sensibles des circuits [58], ou rendent le circuit plus robuste par conception [59]. Néanmoins, à notre connaissance, il n'existe pas de stratégies efficaces de conception pour contrôler ou façonner les émissions électromagnétiques.

IV. 3. 1. 2 La classe de circuits de micropipeline

Dans ce qui suit, nous présentons une méthode permettant de mettre en forme et de contraindre le rayonnement électromagnétique d'un circuit numérique asynchrone. Afin d'évaluer notre méthode, nous avons utilisé une classe de circuits asynchrones connue sous le nom de micropipeline [60]. Toutes les autres classes de circuits asynchrones peuvent également être utilisées. L'avantage du micropipeline est qu'il est conçu avec une séparation entre le chemin de données et le chemin de contrôle. Le chemin de données utilise donc une structure similaire aux circuits synchrones mais la synchronisation globale implémentée par l'arbre d'horloge est remplacée par un contrôleur faisant progresser les données individuellement et pas à pas dans le circuit. Ainsi, la synchronisation est assurée par un chemin de contrôle qui est composé de petits contrôleurs distribués implémentant un protocole « poignée de main » [60] synchronisant localement le transfert des données.

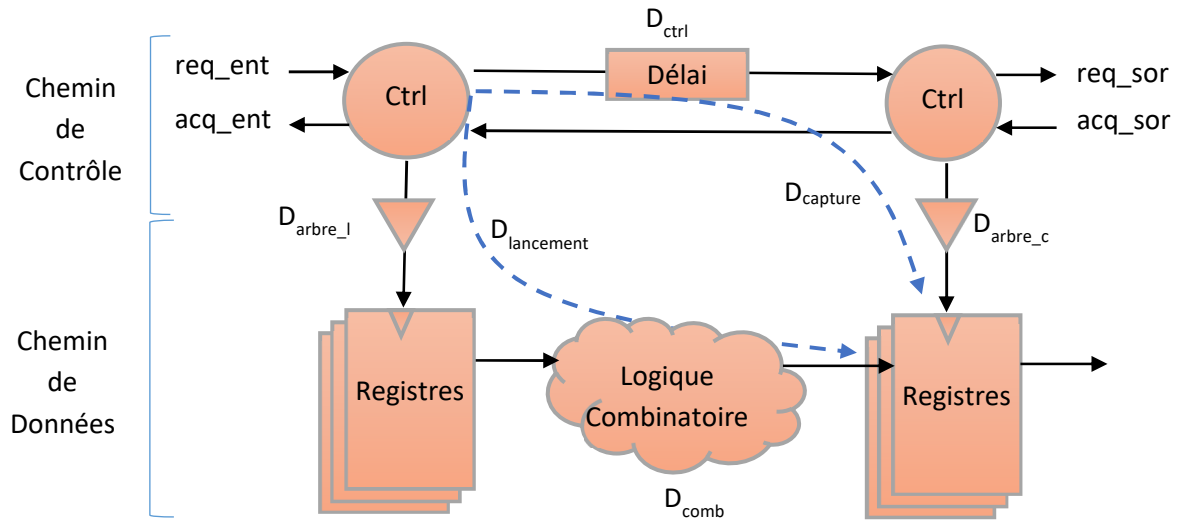


Figure 79 : Architecture d'un circuit micropipeline

Afin d'assurer le bon fonctionnement de l'architecture micropipeline, certaines contraintes temporelles doivent être respectées [61] afin d'assurer la synchronisation locale des données. De façon similaire aux circuits synchrones, une contrainte de positionnement doit être respectée afin de s'assurer que la donnée arrive dans le registre avant l'horloge émise par le chemin de contrôle. Elle peut se résumer avec l'Equation 1

$$D_{capture} > D_{lancement}$$

Equation 1 : Contrainte temporelle de positionnement

Le signal de capture traverse l'élément de délai (D_{ctrl}) et le retard lié à la construction de l'arbre d'horloge de capture ($D_{horloge\_c}$). Pour le signal de lancement, il traverse le délai lié à l'arbre d'horloge de lancement ($D_{horloge\_l}$) suivi du délai du chemin critique de la logique combinatoire (D_{comb}). Nous pouvons ainsi déduire de l'Equation 1, l'Equation 2 qui nous donne la valeur minimale du retard que nous devons insérer pour respecter la contrainte temporelle de positionnement. La valeur maximale de ce retard est quant à elle déterminée par la performance attendue du circuit.

$$D_{ctrl} > D_{horloge\_l} + D_{comb} - D_{horloge\_c}$$

Equation 2 : Equation du retard minimum

Cette approche micropipeline offre donc intrinsèquement un ensemble de retards, D_{ctrl}, qui peuvent être utilisés pour la propagation et le contrôle du spectre électromagnétique émis d'un circuit numérique. Nous allons voir maintenant comment calculer ce retard.

IV. 3. 1. 3 Le modèle de courant comme source des émissions EM

Dans les circuits synchrones, la consommation dynamique peut être modélisée en fonction des considérations suivantes :

- Dans les circuits CMOS, la commutation des portes produit la consommation dynamique.
- La majeure partie de l'activité de commutation des portes est localisée juste après un front d'horloge.
- L'activité de commutation de l'horloge et de son arbre associé ainsi que les registres sont les contributeurs majoritaires de la consommation dynamique.

Par conséquent, la consommation dynamique d'un circuit CMOS numérique peut au premier ordre être modélisée grâce aux impulsions de courant (voir la Figure 80).

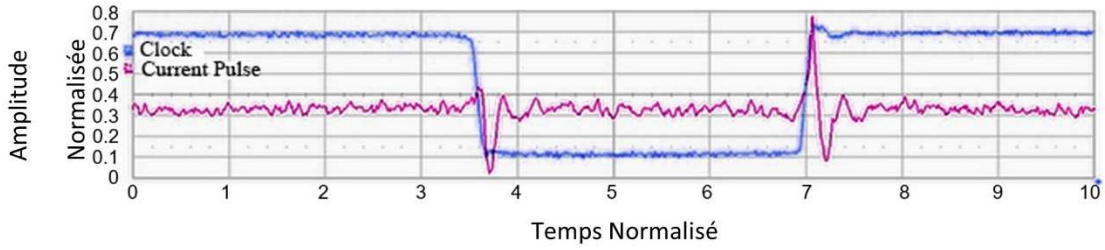


Figure 80 : Consommation de Courant CMOS

L'horloge globale ayant disparu dans les circuits asynchrones, l'activité de commutation des registres est maintenant synchronisée par des contrôleurs locaux. Par conséquent, la distribution des impulsions de courant n'est plus uniforme en temps. Il apparait aussi que nous pouvons modéliser la courbe de consommation de courant comme la somme des pics de courants aux instants de commutation des bascules.

Il est donc important de définir un modèle précis du pic de courant. Dans le cas général, un modèle triangulaire est utilisé [62]. Nous lui avons préféré un modèle basé sur deux exponentielles dos à dos qui s'inspire de la distribution asymétrique de Laplace et définit par l'Equation 3 et représenté sur la Figure 81.

$$f(x) = A.\begin{cases} e^{b(x-t_{com})}, & x < t_{com} \\ e^{-c(x-t_{com})}, & x \geq t_{com} \end{cases}$$

Equation 3 : Exponentielle dos à dos

Dans cette équation, A représente l'amplitude du pic de courant et t_{com} l'instant de commutation. b et c représentent respectivement les coefficients de croissance et de décroissance des exponentielles. Les paramètres de cette équation sont calculés pour chaque étage du chemin de données grâce à une extraction temporelle statique. Le temps de montée de la première exponentielle (le paramètre a) vient de la propagation du signal dans le chemin

de requête jusqu'au registre. Il dépend donc principalement de la profondeur de l'arbre d'horloge représenté par le délai D_{arbre_I} dans la Figure 79. La hauteur du pic (paramètre A) est lié au nombre du registres en considérant que tous les registres basculent en même temps (ce qui est évidement un pire cas). Enfin le temps de descente (paramètre b) vient de la longueur du chemin critique dans la logique combinatoire notée D_{comb} la Figure 79.

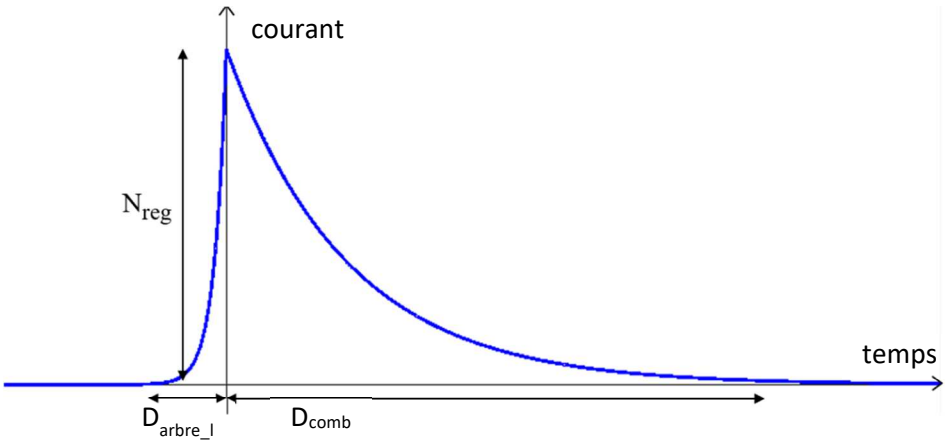


Figure 81 : Forme d'une impulsion de courant à l'aide du modèle exponentielle dos à dos

Afin de valider la pertinence de ce modèle, un étage simple constitué de registres et d'un chemin combinatoire a été simulé avec Cadence Spectre®. La courbe bleue de la Figure 82 représente la consommation de ce circuit après le front montant du signal de requête (le signal rouge). La courbe verte représente le pic de courant modélisé avec un triangle et la courbe noire correspond à notre modèle du pic de courant modélisé avec deux exponentielles dos à dos. Nous constatons que la consommation du circuit correspond bien à notre modèle et que le modèle triangulaire surévalue la consommation.

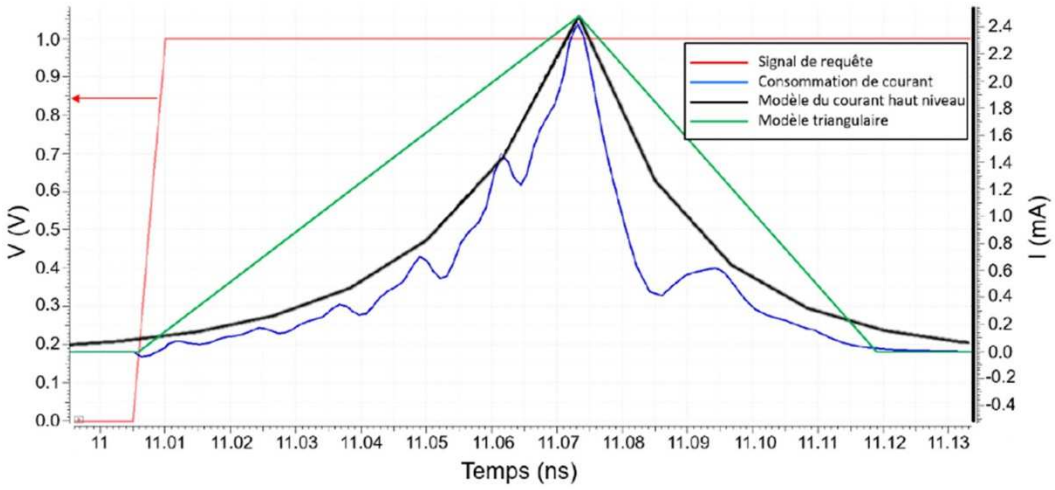


Figure 82 : Simulation transistor d'un étage du chemin de données

IV. 3. 1. 4 Flot de conception pour contrôler le rayonnement EM

Le flot de conception mis en place s'appuie sur un simulateur de circuit rapide dédié basé sur un modèle comportemental du chemin de contrôle en SystemVerilog et sur une estimation du courant s'appuyant sur le modèle décrit dans la section précédente (voir la Figure 81Figure 83). Ce simulateur de réseau de pétri temporelle s'appuie sur l'outil de Cadence Xcelium® et nous permet donc d'obtenir l'instant de déclenchement du pic de courant ainsi que son gabarit. La première étape consiste à concevoir le circuit au niveau transfert de registres (RTL) et d'effectuer une analyse temporelle du circuit afin de déterminer la valeur minimale des retards insérés dans le chemin de contrôle (D_{ctrl}) Cette valeur sera utilisée comme la référence minimale absolue lors de l'ajustement des retards pour façonner le spectre. Selon la vitesse cible du circuit, une valeur maximale sera également fixée.

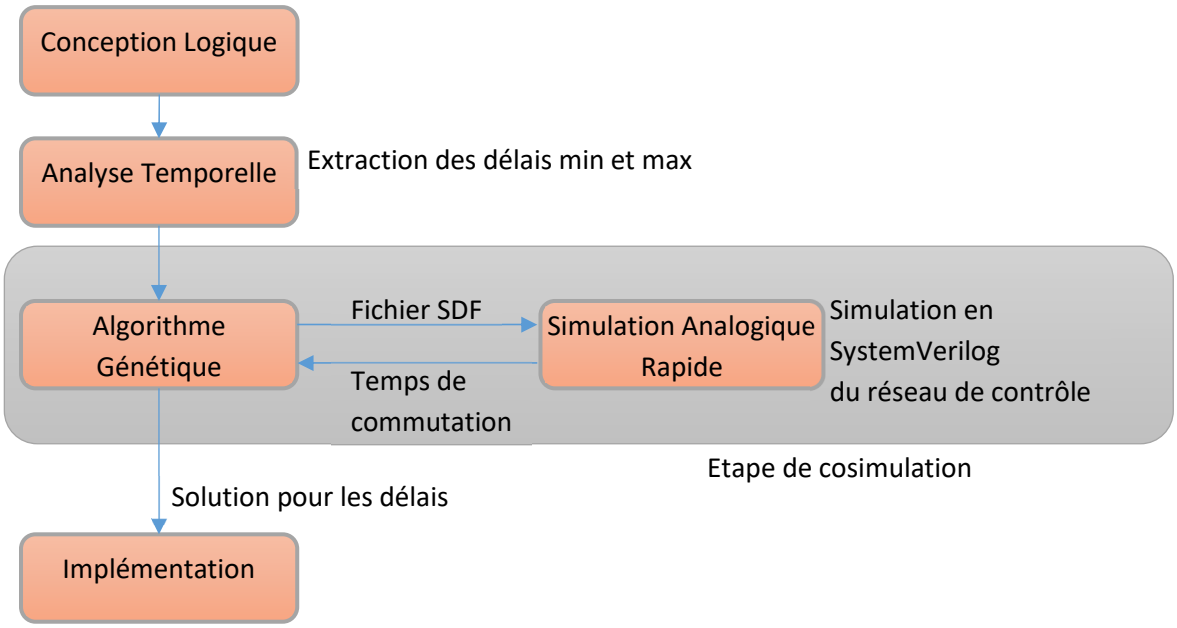


Figure 83 : Flot de conception pour contrôler le rayonnement EM

Après cette première étape vient ensuite l'étape de mise en forme (la boîte grise de la Figure 83). Nous faisons appel à un algorithme génétique (AG) pour calculer les nouveaux retards du chemin de contrôle produit à chaque génération et les fournissons à notre simulateur analogique rapide via un fichier SDF (*Standard Delay File*). Chaque combinaison de délai est ainsi simulée avec le simulateur de réseau de Petri temporel afin d'obtenir la courbe de consommation. Ensuite, le flot effectue une transformée de fourrier rapide sur la courbe de consommation afin de comparer le spectre électromagnétique avec le masque spectral défini par le concepteur. Une solution est trouvée lorsqu'une combinaison de retard

donne un spectre électromagnétique qui s'insère dans le masque spectral. Lorsque le spectre souhaité est obtenu, les retards appropriés sont alors transmis pour la prochaine étape qui permet de construire les délais nécessaires à la mise en œuvre finale du circuit.

IV. 3. 1. 5 Etape de cosimulation : Algorithme Génétique

Introduite par le professeur Holland de l'Université du Michigan [63], l' Algorithme Génétique (AG) s'inspire de la biologie et de phénomènes naturels. Cet algorithme manipule des populations d'individus, où chacune représente une solution potentielle à un problème. Chaque individu est composé de gènes qui sont les paramètres du problème. L'AG (Voir Figure 84) démarre avec une population initiale composée de N individus constituant une solution au problème posé *(a priori* non satisfaisante sinon nous n'aurions pas besoin de l'AG !). Ensuite, chaque individu *est* associé à une fonction de coût ou fonction *fitness* qui permet d'évaluer les individus par rapport au problème posé. Dans notre cas, la fonction de coût sera une mesure de distance entre le spectre obtenu et le gabarit visé. L'idée de l'AG est de sélectionner, parmi les individus, les parents qui permettront de construire la prochaine génération en trois étapes.

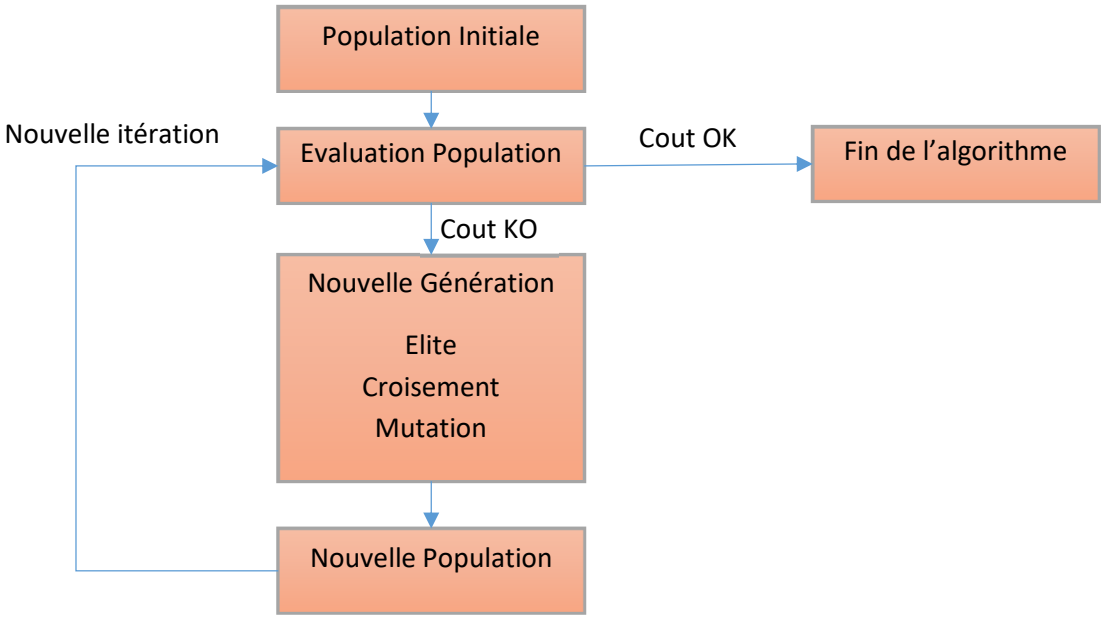


Figure 84 : Processus de l'Algorithme Génétique

Tout d'abord, à partir d'une génération, les enfants de l'élite, les individus avec les meilleurs coûts, sont automatiquement sélectionnés pour la prochaine génération. Comme il n'y a aucune garantie que les enfants soient meilleurs que leurs parents, une partie de la population est conservée pour la prochaine génération. Puis, comme pour la biologie, des enfants croisés sont créés en combinant les gènes de deux parents. Enfin, des enfants

mutants résultant de changements aléatoires dans les gènes des individus sont générés. Cette étape de mutation évite d'être enfermé dans une solution.

Enfin, la nouvelle génération est évaluée avec la fonction *fitness*. Si une solution est trouvée, l'algorithme s'arrête, sinon il est répété jusqu'à qu'une solution (acceptable) soit trouvée. Pour éviter la mort de la population, toutes les générations doivent avoir la même taille. Cet algorithme permet de trouver une solution pour s'adapter au masque spectral. Il est à noter que le spectre doit juste être compatible avec le gabarit et que la notion d'optimum spectral n'a pas aucun sens ici.

Si nous reprenons notre problématique, les gènes sont les retards dans le chemin de contrôle. L'algorithme commence par une population initiale aléatoire d'une centaine d'individus qui respectent les contraintes minimales et maximales pour les gènes. L'AG effectue une transformation de Fourier rapide (FFT) de la consommation en courant calculée par le simulateur rapide afin de déterminer le spectre en fréquences des individus. Le spectre de fréquences est ensuite comparé au masque spectral défini par le concepteur. Tous les points du spectre qui sont au-dessus du masque spectral, sont ajoutés pour calculer la fonction de coût de l'individu. L'AG s'arrête lorsqu'un individu ayant une fonction de coût de 0 est trouvée.

IV. 3. 1. 6 Etape de cosimulation : Simulation de Courant Rapide

Pour accompagner et nourrir l'algorithme génétique nous avons besoin de connaitre les temps de commutation de nos registres. Ce calcul est effectué à l'aide d'un simulateur analogique rapide s'appuyant sur une description SystemVerilog du réseau de contrôle. Ce choix se justifie par la facilité de simuler un modèle SystemVerilog en utilisant un simulateur basé sur des événements.

Grâce au banc d'essai en SystemVerilog, les instants d'activation des registres sont calculés et mémorisés dans un fichier. Une impulsion est alors associée à chaque temps d'activation pour obtenir la consommation totale du circuit micropipeline en sommant tous ces évènements

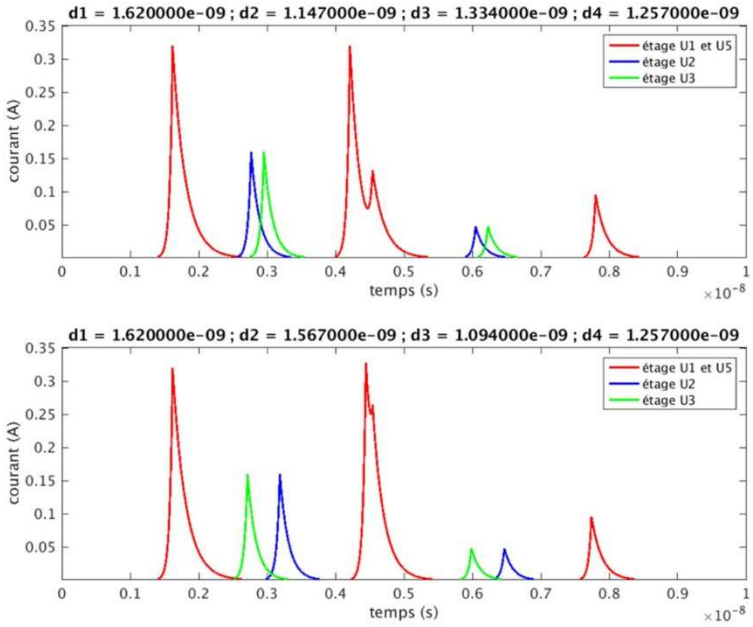


Figure 85 : Consommation de courant avec deux populations différentes

Sur la Figure 85 nous observons la consommation de courant pour 2 tirages différents de délais venu de l'algorithme génétique. Les différentes hauteur et largeur de pic viennent de l'application du modèle de courant sur les événements. Ce profil de consommation peut être transféré dans le domaine fréquentiel à l'aide d'une transformée de Fourier rapide afin d'être comparé au gabarit et renvoyé à l'algorithme génétique pour un nouvel essai si la fonction de coût n'est pas nulle.

IV. 3. 1. 7 Etape de cosimulation : L'obtention d'une solution

Afin de disposer de circuit de teste, la norme pour le chiffrement avancé des données (AES – *Advanced Encryption Standard*) a été implémentée dans le style micropipeline avec le flot proposé afin de façonner son spectre électromagnétique. La Figure 86 montre le schéma du fonctionnement l'AES.

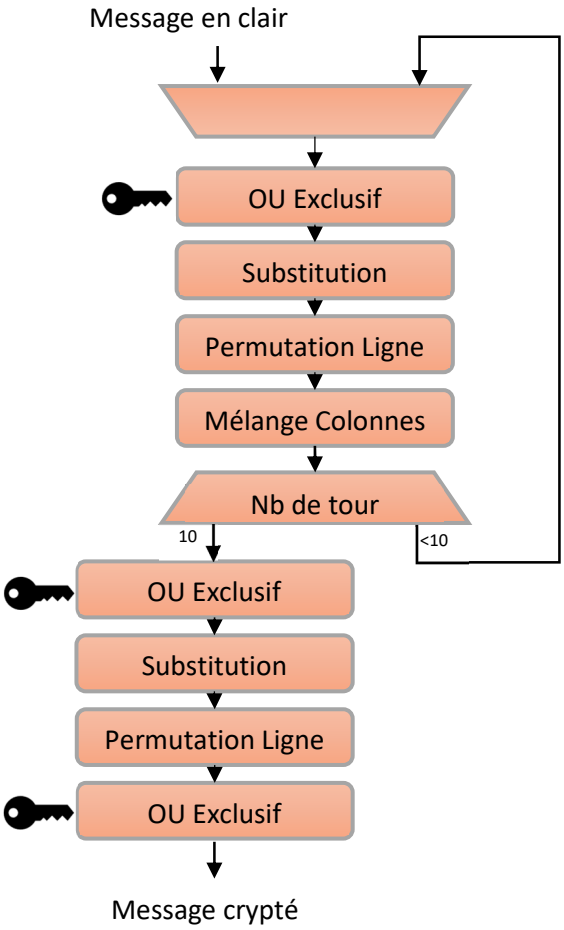


Figure 86 : Schéma block d'un AES 128bits

Une fois la conception RTL et la synthèse logique réalisées, nous avons lancé notre flot de cosimulation associant l'algorithme génétique et le simulateur rapide afin de trouver une solution adaptée aux exigences spectrales. Une fois que le spectre d'un individu est sous le masque spectral défini nous arrêtons les boucles d'optimisation.

Au début du processus de mise en forme, la Figure 87 est obtenue en appliquant une FFT sur le courant consommé par l'AES. Le spectre résultant d'une simulation au niveau portes, la courbe bleue, est comparé au spectre de la cosimulation, la courbe rouge. Nous observons que le spectre des modèles de haut niveau est plus pessimiste que le spectre de simulation des portes. La courbe noire représente le masque spectral choisi pour cette analyse. Dans notre exemple, l'idée principale est de réduire les harmoniques et nous voyons que le spectre actuel est au-dessus du masque et que donc il ne correspond pas aux spécifications du circuit.

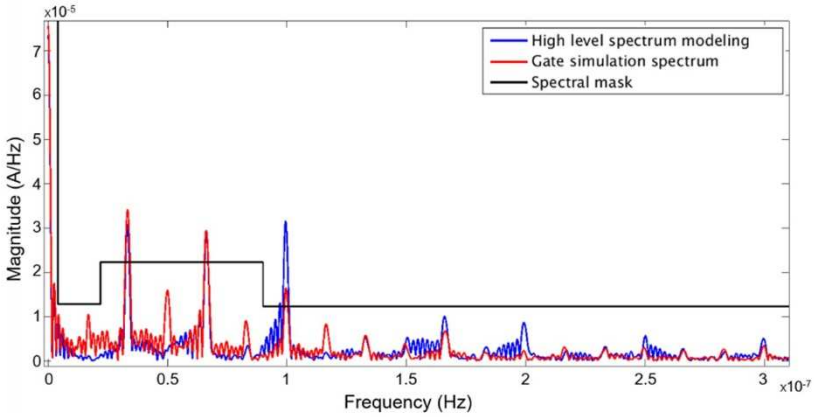


Figure 87 : Spectre de l'AES obtenu avant optimisation

Sur la Figure 88, la modélisation du spectre de haut niveau de l'ensemble de l'AES est comparée au spectre de simulation au niveau portes logiques après l'application de notre flot de conception L'ensemble des retards nécessaire à l'ajustement du masque spectral a été trouvé grâce à l'algorithme génétique après environ six heures d'exécution sur une ferme de calcul.

Nous observons que les deux spectres (le spectre du modèle de haut niveau en bleu et le spectre de simulation des portes logiques en rouge) sont toujours inférieurs au masque spectral (courbe noire). Nous notons que le spectre de simulation des portes est généralement sous la modélisation du spectre de haut niveau. Notre modèle ne tient pas compte des effets parasitaires qui se produisent pendant le changement de la logique combinée (comme les commutations indésirables ou les effets de diaphonie [64]. Notre modèle ne prenant pas en compte ces effets qui contribuent à étaler le spectre du circuit, il est donc un peu pessimiste par rapport à celui de la simulation au niveau portes.

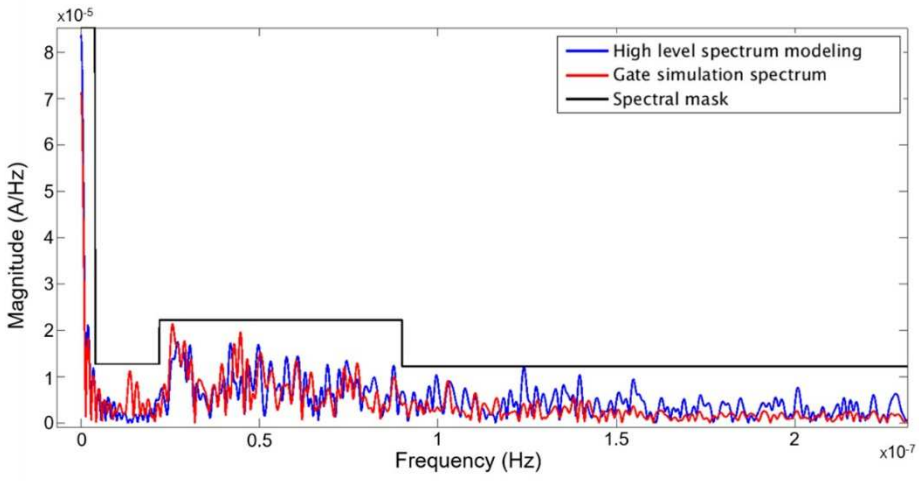


Figure 88 Spectre de l'AES obtenu après optimisation

IV. 3. 2. Le circuit de test pour les mesures EM

Afin de pouvoir valider nos résultats sur silicium, un circuit de test a été réalisé en technologie 40nm de STMicroelectronics. Nous avons décidé d'implémenter l'AES décrit dans le chapitre précédent. Nous avons pu bénéficier d'une plateforme de test pour intégrer notre AES qui est un bloc IP parmi d'autres. La Figure 89 nous montre le circuit de test complet ainsi que l'AES représenté en bleu.

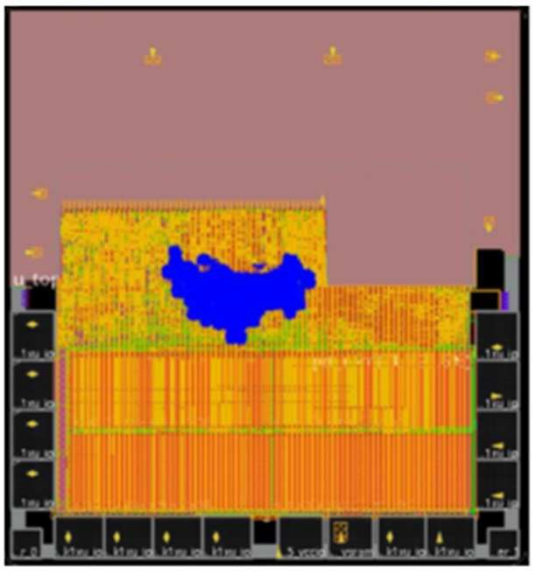


Figure 89 : Circuit de Test incluant l'AES

L'implémentation de ce circuit nous a montré quelques limites dans la structure des contrôleurs. Le choix d'utiliser des lignes à retards programmables construites en cellules standard apporte un vrai avantage pour la rapidité de mise en œuvre mais comme nos pouvons le constater sur la Figure 90, l'impact en surface est non négligeable et peut largement être optimisé avec des lignes à retards customisées.

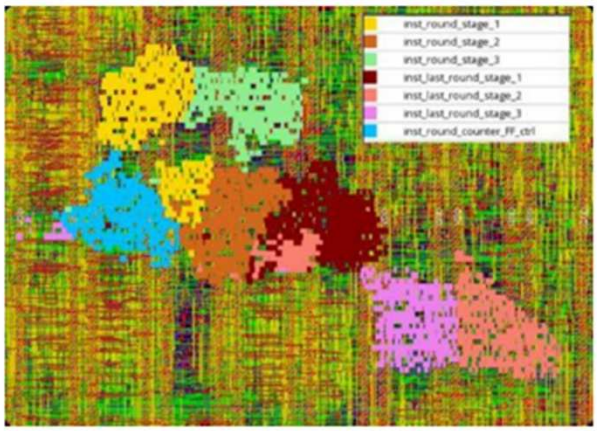
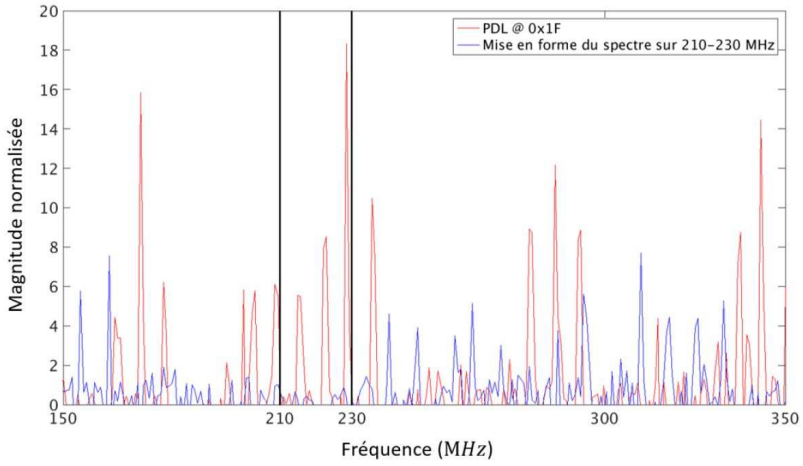


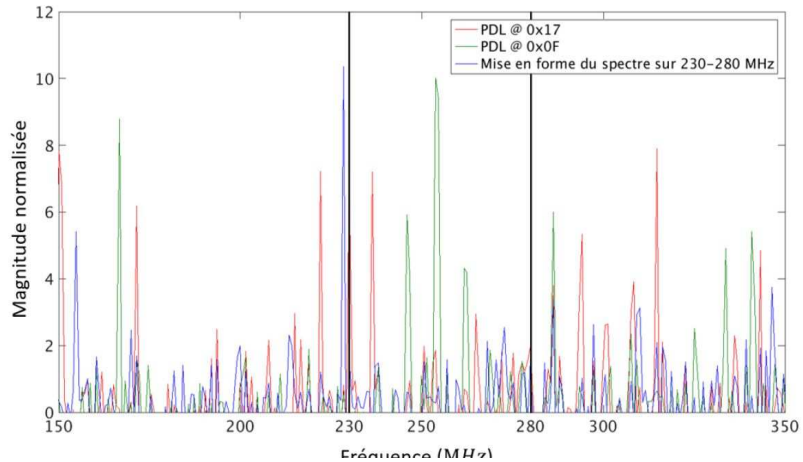
Figure 90 : Implémentation Physique des contrôleurs

Nos premiers essais de mesure dans une mini-cage de faraday ou dans une cellule TEM se sont révélés infructueux. La raison principale est que le rayonnement de notre AES est relativement faible. Nous avons ensuite utilisé un scanner de surface pour mesurer les faibles émissions électromagnétiques de notre circuit. Le code des lignes à retards a été calculé en utilisant la méthode présentée au chapitre IV. 3. 1. 4 . Dans cet exemple, nous cherchons à réduire les émissions spectrales dans une plage donnée. La Figure 91 nous montre le déplacement de spectre pour trois configurations différentes. Les raies rouges et vertes représentent le spectre avec des codes par défaut pour les lignes à retards et les raies bleus représente le spectre lorsque le code est calculé avec l'outil de cosimulation. Dans les configurations A, B et C, nous cherchons à réduire les émissions dans la plage de fréquence 210Mhz-230Mhz, 230Mhz-280Mhz et 370Mhz-400Mhz respectivement.

Configuration A



Configuration B



Configuration C

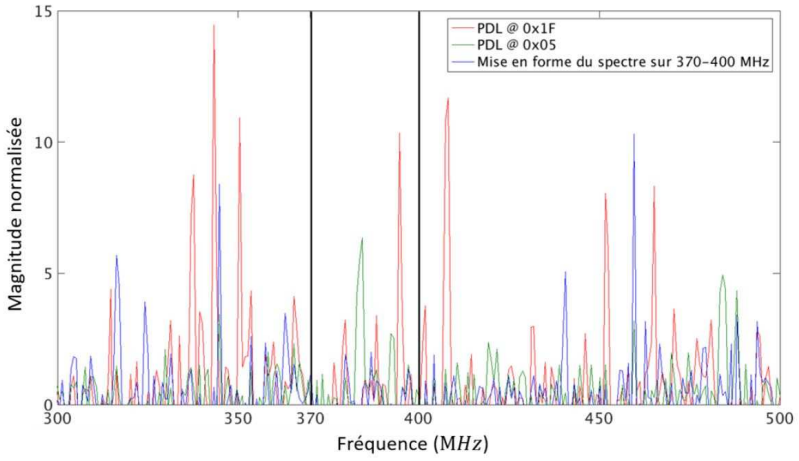


Figure 91 : Déplacement du spectre en fonction du code de la ligna à retard

Grâce à la cosimulation, nous obtenons pour les 3 configurations des intervalles sans raies bleues ; ce qui est en ligne avec le résultat souhaité. Nous arrivons ainsi avec le même matériel à obtenir 3 configurations de spectre différentes. Ces travaux sur la modification du rayonnement électromagnétique ont donné lieu à un dépôt de brevet [65].

IV. 4. Conclusion sur le chapitre

Ce chapitre se présente comme une ouverture vers de nouvelles solutions et pistes pour continuer à améliorer l'efficacité énergétique de nos circuits.

L'horloge restant un contributeur important de la consommation dynamique, nous avons cherché dans un premier temps à développer des systèmes de contrôle de cette horloge. La gestion de l'effet d'inversion de température ou l'introduction d'un contrôleur de gigue sont des solutions envisageables pour réduire la consommation dynamique. Cela étant dit, la possibilité apportée par la logique asynchrone d'enlever l'horloge offre une perspective intéressante. Cette perspective est ainsi explorée à travers la réalisation d'un réseau de capteurs. Nous avons utilisé la solution du capteur du chapitre précédent pour l'intégrer dans un circuit asynchrone contenant un microcontrôleur. La conception et les mesures de ce circuit ont confirmé la souplesse d'intégration et la robustesse aux variations de tension d'une telle solution.

Dans une deuxième partie, nous avons exploré un nouveau champ avec l'introduction de technique permettant de contrôler le spectre électromagnétique. Contrôler le spectre présente des intérêts multiples, qu'il s'agisse de respecter une régulation tel que la directive relative à la compatibilité électromagnétique des appareils électroniques (2014/30/UE) ou d'éviter les perturbations. Nous avons donc vu une méthode nous permettant d'estimer l'émission électromagnétique d'un circuit mais, surtout, de la contrôler en utilisant une solution de calcul basée sur évènements s'appuyant sur des circuits micropipelines. Cette méthode complète a également été validée sur un circuit de test avec lequel nous avons démontré la capacité de la solution à positionner les évènements de calcul pour déplacer le spectre électromagnétique. Cette nouvelle capacité de contrôle du calcul semble offrir un nouveau levier de conception pour des circuits basés sur évènements à haute efficacité énergétique.

Conclusion et perspectives

Les travaux présentés dans ce manuscrit ont accompagné une quinzaine d'années d'évolution des circuits électroniques vers un besoin accru en efficacité énergétique. Que ce soit pour répondre à une problématique de portabilité, de durabilité ou pour lutter contre l'augmentation des courants de fuites liés à la réduction des dimensions des procédés de fabrication, les communautés académique et industrielle ont dû créer un nouvel axe de recherche pour contrôler la consommation des circuits. Ces travaux en sont le fruit et montrent la variété des possibles pour améliorer l'efficacité énergétique de nos circuits.

Dans un premier temps, nous avons proposé des solutions basées sur des îlots d'alimentation couplés à une logique de rétention nous permettant de maintenir l'information utile à un coût énergétique très faible. Ces solutions sont aujourd'hui utilisées de façon commune dans les objets connectés et plus particulièrement dans les téléphones portables. Cette solution est surtout efficace sur des objets majoritairement en veille. Pour des objets majoritairement actifs tels que les routeurs internet où la consommation dynamique est le facteur prépondérant, nous présentons une solution s'appuyant sur des registres double-front. Nous parlons ici de solutions qui vont au-delà du schéma, car cela nécessite d'être couplé à un flot de conception industrialisable. La solution utilise donc le flot de conception standard et présente les adaptations nécessaires pour intégrer ce registre à moindre coût.

Par la suite, nous nous sommes intéressés aux limites rencontrées par les transistors sur les procédés de fabrication avancés. Les effets de variation de la tension de seuil et les problématiques de courant de fuite sous le seuil sont devenus nos deux principales cibles. En parallèle, les technologues nous ont proposé une nouvelle technologie planaire à film mince, le FD-SOI. Cette technologie offre au concepteur un nouveau levier de contrôle des performances des transistors avec la polarisation du substrat. Nous avons donc cherché à optimiser l'usage de ce nouveau levier pour lutter contre les deux effets visés. Pour le courant de fuite, nous avons à travers un circuit montrer comment la polarisation du substrat permettait de positionner le fonctionnement d'un circuit sur son optimum d'efficacité énergétique. Pour la variation de la tension de seuil, nous avons développé des capteurs performants permettant de mesurer la position du circuit par rapport aux procédés de fabrication afin de lui appliquer un ajustement de la tension permettant de le recentrer.

Alors que les deux premières parties présentent des solutions directement industrialisables, La troisième partie ouvre déjà quelques perspectives pour le futur. L'utilisation de la logique asynchrone, longtemps mise de côté pour sa complexité supposée, retrouve un regain d'intérêt pour repousser les limites de l'efficacité énergétique. Sa capacité intrinsèque à calculer uniquement lorsque cela s'avère nécessaire offre une alternative intéressante à la logique synchrone pour les circuits travaillant sur évènements.

Nous avons ainsi travaillé sur deux solutions s'appuyant sur cette logique asynchrone. Dans un premier temps, un réseau de capteurs conçu en logique asynchrone quasi insensible aux délais (QDI) a été intégré dans un circuit de test afin de vérifier la pertinence de l'approche. Les performances et surtout la robustesse du circuit par rapport aux variations des procédés de fabrication et de tensions, ont ouvert de nouvelles perspectives. Il pourrait en effet être utiliser dans les circuits autonomes avec récupération d'énergie dans lesquels la régulation de tension est un problème complexe qui pourrait donc être simplifié, voire supprimé. La deuxième solution s'appuie sur la logique asynchrone à données groupées (Bundle Data). Ces travaux innovants proposent un flot complet de conception qui permet de contrôler avec précision le moment où le circuit exécute un calcul. Un exemple d'utilisation de cette technique est le positionnement du spectre électromagnétique dès la phase de conception, ce qui reste à notre connaissance une exclusivité mondiale.

Perspectives

Comment repousser les limites de l'efficacité énergétique ? C'est sûrement une question qui animera les débats et qui pilotera les choix stratégiques de la filière électronique durant les prochaines décennies. Ce besoin n'impacte pas uniquement les objets connectés comme on pourrait l'imaginer naïvement car l'omniprésence de l'électronique dans quasiment toutes les applications nécessite d'avoir des circuits efficaces dans de multiples conditions. Comme il semble impossible d'obtenir une solution unique, une hétérogénéité de techniques, circuits et procédés de fabrication est en train d'apparaitre. Avec une telle richesse, nos travaux sur la logique asynchrone trouveront probablement un écho dans le monde de l'internet des objets où les besoins en puissance de calcul ne sont pas si importants, où l'autonomie en veille et la possibilité de se réveiller sur un évènement sont, quant à eux, les deux éléments clefs pour ce type d'application.

Il semblerait aussi pertinent de se pencher sur la problématique en plein essor de l'intelligence artificielle. Les principaux réseaux de neurones sont actuellement conçus pour être exploités dans des centres de calcul mais un nouveau besoin se fait sentir pour rapprocher le réseau de l'application. Dans ce cas, le réseau se trouvera obligatoirement confronter à une

limitation de l'énergie disponible et devra donc être en mesure de s'y adapter. La capacité de la logique asynchrone à tolérer les variations de tension nous laisse aussi à penser qu'elle pourra amener un gain significatif sur le bilan énergétique de ces réseaux.

Enfin, nous avons, grâce aux nombreuses réalisations de différents circuits intégrés, confronté les solutions synchrone et asynchrone. Ce débat est je pense derrière nous car il est difficilement imaginable que la logique asynchrone remplace la logique synchrone. Nous prônerons donc une approche plus en adéquation avec la logique du « *more than Moore* » ou la logique asynchrone trouvera sa place à l'intérieur de systèmes synchrones afin d'en améliorer les capacités. Il y a donc un vrai besoin de proposer un flot de conception en accord avec les principes de la conception synchrone. La méthode proposée pour le contrôle du spectre mérite donc d'être étendue afin de faire profiter nos futurs circuits des capacités remarquables de cette approche.

Production Scientifique

Brevets

[P11] S. Engels, A. Aurand, E. Maurin « Flip-flop with a metal programmable initialization logic state ». N° US10505522B1 (2019)

[P10] S. Engels, L. Fesquet, S Germain, “System and Method for Managing Requests in an Asynchronous Pipeline”, N° US2020184110A1 (2018)

[P9] P. Galy, S. Athanasiou, J. Le-Coz, S. Engels “Electronic device for heating an integrated structure, for example an MOS transistor ». N° US2016370815A1 (2015)

[P8] N. L’Hostis, S. Engels, F. Blisson, and C.-M. Lachaud, “Method and apparatus for controlling power supply,”. N° US2013043936A1 (2014)

[P7] J. Le Coz, S. Engels, and A. Tournier, “Electronic circuit design method,”. N° US8819615B2 (2014)

[P6] T. Lehuche and S. Engels, “Clock Signal Synchronization Circuit,”. N° US2013300458A1 (2013)

[P5] J. Le Coz, A. Valentian, P. Flatresse, and S. Engels, “Transistor substrate dynamic biasing circuit,” N° US2012062313A1 (2013)

[P4] S. Engels, “Dual-edge register and the monitoring thereof on the basis of a clock,”. N° US8436652B2 (2013)

[P3] S. Engels, “Clock control circuit,”. N° FR2925243A1 (2012)

[P2] R. Wilson, S. Engels, and E. Balossier, “Method and device to control the frequency of a clock signal of an integrated circuit,”. N° US2011199149A1; (2011)

[P1] S. Engels, D. Jacquet, and N. L’Hostis, “Electronic device for generating gated clock signal for register of integrated circuit,” N° FR2964478A1 (2009)

Articles de revue

[J8] S. Germain, S. Engels, L. Fesquet, “A High-Level Current Modeling for Shaping Electromagnetic Emissions in Micropipeline Circuits”, Journal of Low Power Electronics and Applications, vol. 9, no 1, 2019.

[J7] M. Renaudin, A. Bouzafour, S. Engels, R. Wilson. “A 6-Wire Plug and Play Clockless Distributed On-Chip-Sensor Network in 28 nm UTBB FD-SOI”, Journal of Low Power Electronics 2018/9/1, 2018

[J6] P. Vivet, S. Lesecq, D. Puschini, A. Molnos, F. Thabet, B. Tain, K. Ben Chehida, S. Engels, R. Wilson, and D. Fuin. “A Fine-Grain Variation-Aware Dynamic Vdd-Hopping AVFS Architecture on a 32 nm GALS MPSoC”. - IEEE Journal of Solid-State Circuits, 2014

[J5] N. Moubdi, P. Maurine, R Wilson, N Azemard, V. Dumettier, A. Bansal, S. Barasinski, A. Tournier, G. Durieu, D. Meyer, P. Busson, S. Verhaeren, S. Engels “On-Chip Process Variability Monitoring Flow” Journal of Low Power Electronics (JOLPE), Dec. 2010.

[J4] S. Engels, R. Wilson, N. Azemard, P Maurine, V Migairou, “Timing Margin Evaluation with a Simple Statistical Timing Analysis Flow.” Journal of Embedded Computing (2009-07)

[J3] B. Foret, L. Rolindez, C. Adobati, S. Engels, « Unified Environment for Mixed-signal Top-level SoC Verification.” IEEE Journal of Solid-state Circuits, 42(5), 992-1002.

[J2] S. Engels; R. Wilson; N. Azémard; P. Maurine “A comprehensive performance macro-modeling of on-chip RC interconnects considering line shielding effects,” Integration, the VLSI Journal, vol. 39, no. 4, pp. 433–456, Jul. 2006.

[J1] B. Lasbouygues; S. Engels; R. Wilson; P. Maurine; N. Azemard; D. Auvergne, “Logical effort model extension to propagation delay representation,” Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, vol. 25, no. 9, pp. 1677–1684, Sep. 2006.

Conférences internationales avec comité de lecture et publication des actes

[C20] R. Jadue, S. Engels, L. Fesquet. “An Event-Based Strategy for ASK demodulation.” in 5th International Conference on Event-Based Control, Communication, and Signal Processing, (EBCCSP 2019), Vienna, Austria.

[C19] R. Jadue, S. Engels, L. Fesquet 27th IFIP/IEEE, “A Digital Event-Based Strategy for ASK demodulation”; International Conference on Very Large Scale Integration (VLSI-SoC 2019), Oct 2019, Cuzco, Peru

[C18] L. Fesquet, Y. Decoudu, R. Iga, T. Ferreira de Paiva Leite, O. Rolloff, M. Diallo, R. Possamai Bastos, K. Morin-Allory, S. Engels . “A Distributed Body-Biasing Strategy for Asynchronous Circuits.” 27th IFIP/IEEE International Conference on Very Large-Scale Integration (VLSI-SoC 2019), Oct 2019, Cuzco, Peru.

[C17] T. Ferreira de Paiva Leite, R. Jadue, S. Engels, R. Possamai Bastos, L. Fesquet. « Fine Grain Body-Biasing: A strategy for asynchronous circuits.” Application, Design and Technology Conference (ADTC 2018), Jun 2018, Grenoble, France

[C16] S. Germain, S. Engels, L. Fesquet, « Shaping Electromagnetic Emissions of Event-Driven Circuits Thanks to Genetic Algorithms », in Third International Conference on Advances in Signal, Image, and Video Processing (SIGNAL 2018), Nice, France, 2018.

[C15] S. Germain, S. Engels, L. Fesquet, « A Design Flow for Shaping Electromagnetic Emissions in Micropipeline Circuits », in 24th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC 2018), Vienna, Austria, 2018.

[C14] S. Germain, S. Engels, L. Fesquet, « Event-Based Design Strategy for Circuit Electromagnetic Compatibility », in 3rd International Conference on Event-Based Control, Communication and Signal Processing (EBCCSP 2017), Funchal, Portugal, 2017, p. 1-7.

[C13] M. Renaudin, A. Buhrig, C. Guillemet, R. Wilson, S. Engels, “Clockless Design Performance Monitoring for Nanometer Technologies,” in Asynchronous Circuits and Systems (ASYNC 2014), 20h IEEE International Symposium on, 2014

[C12] P. Flatresse; B. Giraud; J. Noel; B. Pelloux-Prayer; F. Giner; D. Arora; F. Arnaud, N. Planes, J. Le Coz; O. Thomas; S. Engels; G. Cesana; R. Wilson; P. Urard. “Ultra-wide body-bias range LDPC decoder in 28nm UTBB FDSOI technology,” in Solid-State Circuits Conference Digest of Technical Papers (ISSCC 2013).

[C11] E. Beigne; I. Miro-Panades; Y. Thonnart; L. Alacoque; P. Vivet, S. Lesecq, D. Puschini, F. Thabet, B. Tain, K. Benchehida, S. Engels, R. Wilson, D. Fuin, “A fine grain variation-aware dynamic Vdd-hopping AVFS architecture on a 32nm GALS MPSoC,” in ESSCIRC (ESSCIRC 2013), Proceedings of the, 2013.

[C10] E. Beigne, A. Valentian, B. Giraud, O. Thomas, T. Benoist, Y. Thonnart, S. Bernard, G. Moritz, O. Billoint, Y. Maneglia, P. Flatresse, J.P. Noel, F. Abouzeid, B. Pelloux-Prayer, A. Grover, S. Clerc, P. Roche, J. Le Coz, S. Engels, R. Wilson, “Ultra-Wide Voltage Range designs in Fully-Depleted Silicon-On-Insulator FETs,” in Design, Automation & Test in Europe Conference & Exhibition (DATE 2013), 2013.

[C9] V. Huard, E. Pion, F. Cacho, D. Croain, V. Robert, R. Delater, P. Mergault, S. Engels, P. Flatresse, N.R. Amador, L. Anghel, “A predictive bottom-up hierarchical approach to digital system reliability,” in Reliability Physics Symposium (IRPS 2012), IEEE International, 2012.

[C8] J. Le Coz, P. Flatresse, S. Engels, A. Valentian, M. Belleville, C. Raynaud, D. Croain, P. Urard, “Comparison of 65nm LP bulk and LP PD-SOI with adaptive power gate body bias for an LDPC codec,” in Solid-State Circuits Conference Digest of Technical Papers (ISSCC 2011)

[C7] N.R. Amador, V. Huard, E. Pion, F. Cacho, D. Croain, V. Robert, S. Engels, P. Flatresse, L. Anghel, “Bottom-up digital system-level reliability modeling,” in Custom Integrated Circuits Conference (CICC 2011), IEEE, 2011, pp. 1–4.

[C6] N. Moubdi, P. Maurine, R. Wilson, N. Azemard, S. Engels, L. Rolindez, V. Heinrich, “Voltage scaling and body biasing methodology for high performance hardwired LDPC,” in IC Design and Technology (ICICDT 2010), IEEE International Conference on, 2010, pp. 82–85.

[C5] N. Moubdi, P. Maurine, R Wilson, N Azemard, V. Dumettier, A. Bansal, S. Barasinski, A. Tournier, G. Durieu, D. Meyer, P. Busson, S. Verhaeren, S. Engels “Product On-Chip Process Compensation for Low Power and Yield Enhancement,” in Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation (PATMOS 2009), vol. 5953

[C4] N. Moubdi, R. Wilson, S. Engels, N. Azemard, P. Maurine.” On-Chip Process Variability Monitoring.”: Design, Automation and Test in Europe, (DATE 2009), Nice, France.

[C3] V. Migairou, R. Wilson, S. Engels, Z. Wu, N. Azemard, P. Maurine “A Simple Statistical Timing Analysis Flow and Its Application to Timing Margin Evaluation,” in Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation (PATMOS 2007), vol. 4644

[C2] B. Lasbouygues, J. Schindler, S. Engels, P. Maurine, X. Michel, N. Azemard, D. Auvergne.” Timing Performance Representation of a CMOS Standard Cell Library.” In XVIII Design of Circuits and Integrated Systems Conference (DCIS 2003) (pp. 83-88).

[C1] D. Subiela, S. Engels, L. Dugoujon, R. Esteve-Bosch, B. Mota, L. Musa, A. Jimenez-de-Parga. “A low-power 16-channel AD converter and digital processor ASIC,” in Solid-State Circuits Conference, 2002. (ESSCIRC 2002). Proceedings of the 28th European, 2002, pp. 259–262.

Communications Orales (invitées)

[I1] L. Fesquet, S. Germain, J. Simatic, A. Cherkaoui, T. Le Pelleter, S. Engels, « Event-based processing: a new paradigm for low-power », in 19th IEEE Mediterranean Electrotechnical Conference (IEEE Melecon’18), Marrakesh, Morocco, 2018.

[I2] L. Fesquet, J. Simatic, A. Darwish, A. Cherkaoui, S. Engels, S. Germain, « From events to data-driven processing », in 3rd International Conference on Event-Based Control, Communication and Signal Processing (EBCCSP 2017), Funchal, Portugal, 2017.

Bibliographie

[1] G. E. Moore, « Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. », *IEEE Solid-State Circuits Society Newsletter*, vol. 11, n° 3, p. 33-35, sept. 2006, doi: 10.1109/N-SSC.2006.4785860.

[2] A. K. Yadav, K. Upadhyay, P. Gandhi, et Vaishali, « Various Issues and considerations for the Static Power Consumption in NANO-CMOS: Design Perspective », *Materials Today: Proceedings*, vol. 10, p. 136-141, 2019, doi: https://doi.org/10.1016/j.matpr.2019.02.198.

[3] M. Niroomand et H. R. Foroughi, « A rotary electromagnetic microgenerator for energy harvesting from human motions », *Journal of applied research and technology*, vol. 14, n° 4, p. 259-267, 2016.

[4] H. J. M. Veendrick, « Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits », *IEEE Journal of Solid-State Circuits*, vol. 19, n° 4, p. 468-473, 1984, doi: 10.1109/JSSC.1984.1052168.

[5] A. P. Chandrakasan, S. Sheng, et R. W. Brodersen, « Low-power CMOS digital design », *IEEE Journal of Solid-State Circuits*, vol. 27, n° 4, p. 473-484, 1992, doi: 10.1109/4.126534.

[6] H. Veendrick, *Deep-submicron CMOS ICs: from basics to ASICs*. Springer Science & Business Media, 2000.

[7] J. Henkel, « The triangle of power density, circuit degradation and reliability », in *2017 30th IEEE International System-on-Chip Conference (SOCC)*, sept. 2017, p. 1-2. doi: 10.1109/SOCC.2017.8226039.

[8] A. Keshavarzi, K. Roy, et C. F. Hawkins, « Intrinsic leakage in deep submicron CMOS ICs-measurement-based test solutions », *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, vol. 8, n° 6, p. 717-723, déc. 2000, doi: 10.1109/92.902266.

[9] « Semiconductor Industry Association (SIA), International Roadmap for Semiconductors », 2003. http://www.itrs2.net/itrs-reports.html

[10] D. Duarte, V. Narayanan, et M. J. Irwin, « Impact of technology scaling in the clock system power », in *Proceedings IEEE Computer Society Annual Symposium on VLSI. New Paradigms for VLSI Systems Design. ISVLSI 2002*, avr. 2002, p. 59-64. doi: 10.1109/ISVLSI.2002.1016875.

[11] N. Evanson, « Explainer: What is Chip Binning », 15 juin 2020. https://www.techspot.com/article/2039-chip-binning

[12] Y. Aghaghiri, F. Fallah, et M. Pedram, « Irredundant address bus encoding for low power », in *Proceedings of the 2001 international symposium on Low power electronics and design*, 2001, p. 182-187.

[13] J. Tschanz, S. Narendra, Z. Chen, S. Borkar, M. Sachdev, et V. De, « Comparative delay and energy of single edge-triggered & dual edge-triggered pulsed flip-flops for high-performance microprocessors », in *Proceedings of the 2001 international symposium on Low power electronics and design*, 2001, p. 147-152.

[14] R. Gallager, « Low-density parity-check codes », *IRE Transactions on information theory*, vol. 8, n° 1, p. 21-28, 1962.

[15] IEEE, « IEEE Standard for Information Technology “IEEE Std 802.11n” », 2009. www.standard.ieee.org

[16] S. Engels, D. Jacquet, et N. L’Hostis, « Electronic device for generating gated clock signal for register of integrated circuit », FR 2964478 A1, 19 juin 2009

[17] S. Engels, « Dual edge register and the monitoring thereof on the basis of a clock », 2013 [En ligne]. Disponible sur: https://www.google.com/patents/US8436652

[18] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, et J. Yamada, « 1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS », *IEEE Journal of Solid-state Circuits*, vol. 30, p. 847-854, 1995.

[19] C. Hwang, C. Kang, et M. Pedram, « Gate sizing and replication to minimize the effects of virtual ground parasitic resistances in MTCMOS designs », in *7th International Symposium on Quality Electronic Design (ISQED'06)*, 2006, p. 6-pp.

[20] J. Le Coz, A. Valentian, P. Flatresse, et S. Engels, « Transistor substrate dynamic biasing circuit », US8570096, 2013 [En ligne]. Disponible sur: https://www.google.com/patents/US8570096

[21] J. Le Coz *et al.*, « Comparison of 65nm LP bulk and LP PD-SOI with adaptive power gate body bias for an LDPC codec », in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, févr. 2011, p. 336-337. doi: 10.1109/ISSCC.2011.5746343.

[22] Y. Kanno *et al.*, « Hierarchical power distribution with 20 power domains in 90-nm low-power multi-CPU processor », in *2007 IEEE International Conference on Integrated Circuit Design and Technology*, 2007, p. 1-4.

[23] Z. Liu et V. Kursun, « Characterization of wake-up delay versus sleep mode power consumption and sleep/active mode transition energy overhead tradeoffs in MTCMOS circuits », in *2008 51st Midwest Symposium on Circuits and Systems*, 2008, p. 362-365.

[24] N. L'Hostis, S. Engels, F. Blisson, et C.-M. Lachaud, « Method and apparatus for controlling power supply », US8710917, 2014 [En ligne]. Disponible sur : https://www.google.com/patents/US8710917

[25] S. Engels, « Clock control circuit », FR 2925243 A1, 9 mars 2012

[26] K. Shi, « Low-power SOC implementation: What you need to know. », in *SoCC*, 2010, p. 95.

[27] B. Pelloux-Prayer, « Optimisation de l'efficacité énergétique des applications numériques en technologie FD-SOI 28-14nm », Theses, Université de Grenoble, 2014. [En ligne]. Disponible sur : https://tel.archives-ouvertes.fr/tel-01130989

[28] T. Kuroda *et al.*, « A 0.9-V, 150-MHz, 10-mW, 4 mm/sup 2/, 2-D discrete cosine transform core processor with variable threshold-voltage (VT) scheme », *IEEE Journal of Solid-State Circuits*, vol. 31, n° 11, p. 1770-1779, 1996.

[29] N. Moubdi *et al.*, « Voltage scaling and body biasing methodology for high performance hardwired LDPC », in *IC Design and Technology (ICICDT), 2010 IEEE International Conference on*, juin 2010, p. 82-85. doi: 10.1109/ICICDT.2010.5510289.

[30] E. Beigne *et al.*, « Ultra-Wide Voltage Range designs in Fully Depleted Silicon-On-Insulator FETs », in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2013*, mars 2013, p. 613-618. doi: 10.7873/DATE.2013.135.

[31] J. Sartori, A. Pant, R. Kumar, et P. Gupta, « Variation-aware speed binning of multi-core processors », in *2010 11th International Symposium on Quality Electronic Design (ISQED)*, 2010, p. 307-314.

[32] N. Moubdi *et al.*, « On-Chip Process Variability Monitoring », *Journal of Low Power Electronics (JOLPE)*, vol. 6, n° 4, p. N/A, déc. 2010, doi: http://dx.doi.org/10.1166/jolpe.2010.1109.

[33] N. Moubdi *et al.*, « Product On-Chip Process Compensation for Low Power and Yield Enhancement », in *Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation*, vol. 5953, J. Monteiro et R. van Leuken, Éd. Springer Berlin Heidelberg, 2010, p. 247-255. [En ligne]. Disponible sur : http://dx.doi.org/10.1007/978-3-642-11802-9_29

[34] J. Le Coz, S. Engels, et A. Tournier, « Electronic circuit design method », 2014 [En ligne]. Disponible sur: https://www.google.com/patents/US20140089885

[35] D. Blaauw *et al.*, « Razor II: In situ error detection and correction for PVT and SER tolerance », in *2008 IEEE International Solid-State Circuits Conference-Digest of Technical Papers*, 2008, p. 400-622.

[36] R. Wilson *et al.*, « A 460mhz at 397mv, 2.6 ghz at 1.3 v, 32b vliw dsp, embedding f max tracking », in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, p. 452-453.

[37] N. R. Amador *et al.*, « Bottom-up digital system-level reliability modeling », in *Custom Integrated Circuits Conference (CICC), 2011 IEEE*, sept. 2011, p. 1-4. doi: 10.1109/CICC.2011.6055343.

[38] V. Huard *et al.*, « A predictive bottom-up hierarchical approach to digital system reliability », in *Reliability Physics Symposium (IRPS), 2012 IEEE International*, avr. 2012, p. 4B.1.1-4B.1.10. doi: 10.1109/IRPS.2012.6241830.

[39] Y. Zu, W. Huang, I. Paul, et V. J. Reddi, « Ti-states: Processor power management in the temperature inversion region », in *2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2016, p. 1-13. doi: 10.1109/MICRO.2016.7783758.

[40] N. Pinckney, D. Blaauw, et D. Sylvester, « Low-Power Near-Threshold Design: Techniques to Improve Energy Efficiency, *IEEE Solid-State Circuits Magazine*, vol. 7, n° 2, p. 49-57, Spring 2015, doi: 10.1109/MSSC.2015.2418151.

[41] R. Wilson, S. Engels, et E. Balossier, « Method and device to control the frequency of a clock signal of an integrated circuit », EP2345948A1, 2011 [En ligne]. Disponible sur : https://www.google.com/patents/EP2345948A1?cl=en

[42] « Clock Definition ». https://asic-soc.blogspot.com/2009/01/clock-definitions.html

[43] W.-K. Loo, K.-S. Tan, et Y.-K. Teh, « A study and design of CMOS H-Tree clock distribution network in system-on-chip », in *2009 IEEE 8th International Conference on ASIC*, 2009, p. 411-414. doi: 10.1109/ASICON.2009.5351254.

[44] T. Lehuche et S. Engels, « Clock Signal Synchronization Circuit », US 20130300458, 2013 [En ligne]. Disponible sur: https://www.google.com/patents/US20130300458

[45] V. G. Srivatsa, A. P. Chavan, et D. Mourya, « Design of Low power amp; High Performance Multi Source H-Tree Clock Distribution Network », in *2020 IEEE VLSI DEVICE CIRCUIT AND SYSTEM (VLSI DCS)*, 2020, p. 468-473. doi: 10.1109/VLSIDCS47293.2020.9179954.

[46] M. Renaudin, A. Buhrig, C. Guillemet, R. Wilson, et S. Engels, « Clockless Design Performance Monitoring for Nanometer Technologies », in *Asynchronous Circuits and Systems (ASYNC), 2014 20th IEEE International Symposium on*, mai 2014, p. 108-109. doi: 10.1109/ASYNC.2014.24.

[47] W. Liu, Y. Wang, X. Wang, J. Xu, et H. Yang, « On-Chip Sensor Network for Efficient Management of Power Gating-Induced Power/Ground Noise in Multiprocessor System on Chip », *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, n° 4, p. 767-777, avr. 2013, doi: 10.1109/TPDS.2012.193.

[48] ARM, « ARM AMBA Specification ». http://www.arm.com/products/system-ip/amba-specifications.php

[49] M. Renaudin et A. Fonkoua, « Tiempo asynchronous circuits system verilog modeling language », in *2012 IEEE 18th International Symposium on Asynchronous Circuits and Systems*, 2012, p. 105-112. doi: 10.1109/ASYNC.2012.22.

[50] V. Szekely, C. Marta, Z. Kohari, et M. Rencz, « CMOS sensors for on-line thermal monitoring of VLSI circuits », *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, vol. 5, n° 3, p. 270-276, 1997.

[51] J. Bainbridge et S. Furber, « Chain: a delay-insensitive chip area interconnect », *IEEE Micro*, vol. 22, n° 5, p. 16-23, 2002.

[52] L. Fesquet *et al.*, « A Distributed Body-Bias Strategy for Asynchronous Circuits », présenté à 27th IFIP/IEEE International Conference on Very Large-Scale Integration (VLSI-SOC 2019), Cuzco, oct. 2019.

[53] J. Sparsø et S. Furber, *Principles of asynchronous circuit design: a systems perspective*. Boston: Kluwer, 2010.

[54] A. Bouzafour, M. Renaudin, H. Garavel, R. Mateescu, et W. Serwe, « Model-checking synthesizable systemverilog descriptions of asynchronous circuits », in *2018 24th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, 2018, p. 34-42.

[55] C. M. Hung et K. Muhammad, « RF/Analog and Digital Faceoff-Friends or Enemies in an RF SoC », *VLSI Technology Systems and Applications (VLSI-TSA), 2010 International Symposium on*, p. 19-20, 2010.

[56] M. Cazzaniga, P. Joubert Doriol, A. Sanna, E. Blanc, V. Liberali, et D. Pandini, « Evaluating the impact of substrate noise on conducted EMI in automotive microcontrollers », *Electromagnetic Compatibility of Integrated Circuits (EMC Compo), 2013 9th Intl Workshop on*, p. 129-133, 2013.

[57] M. Ramdani *et al.*, « The Electromagnetic Compatibility of Integrated Circuits—Past, Present, and Future », *IEEE Transactions on Electromagnetic Compatibility*, vol. 51, n° 1, p. 78-100, 2009, doi: 10.1109/TEMC.2008.2008907.

[58] R. Rossi, G. Torelli, et V. Liberali, « Model and verification of triple-well shielding on substrate noise in mixed-signal CMOS ICs », *ESSCIRC 2004 - 29th European Solid-State Circuits Conference (IEEE Cat. No.03EX705)*, p. 643-646, 2003.

[59] M. Mardiguian, *Controlling Radiated Emissions by Design*, 3e éd. Norwell, Massachusetts: Kluwer Academic Publishers, 2014.

[60] I.E. Sutherland, « Micropipelines », *Communication of the ACM*, vol. 32, n° 6, p. 720-738, 1989.

[61] G. Gimenez, A. Cherkaoui, G. Cogniard, et L. Fesquet, « Static Timing Analysis of Asynchronous Bundled-Data Circuits », in *2018 24th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, mai 2018, p. 110-118. doi: 10.1109/ASYNC.2018.00036.

[62] S. Germain, « Contrôle du spectre électromagnétique d'un circuit numérique asynchrone », UGA, 2019.

[63] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor: University of Michigan Press, 1975.

[64] J. Monteiro, S. Devadas, A. Ghosh, K. Keutzer, et J. White, « Estimation of average switching activity in combinational logic circuits using symbolic simulation », *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 16, n° 1, p. 121-127, 1997, doi: 10.1109/43.559336.

[65] S. Engels, L. Fesquet, et S. Germain, « System and Method for Managing Requests in an Asynchronous Pipeline », US2020184110 (A1)., 2020

Liste des Acronymes

ABB	*Adaptative Body Biasing*. Adaptation de la polarisation du substrat en fonction des performances du circuit
ASIC	*Application Specific Integrated Circuit.* Circuit dédié à une application spécifique (ex : routeur IP)
AVS	*Adaptative Voltage Scaling*. Technique d'adaptation de la tension en fonction des performances du circuit
BD	*Bundle Data*. Classe de circuit asynchrone utilisant des données groupées
BIST	*Build in Self-Test*. Solution de test intégrée directement dans le circuit.
CAO	Conception assisté par Ordinateur
CEM	Compatibilité Electro Magnétique
CTS	*Clock Tree synthesis*. Technique de construction en arbre d'un circuit d'horloge
DLL	*Delay Locked Line*. Ligne à retard avec verrouillage du délai
EM	Electro Magnétisme
FBB	*Forward Body Bias*. Polarisation du substrat visant une accélération du circuit
FDSOI	*Fully Depleted*. Technologie SOI entièrement déserté.
GALS	*Globally Asynchronous Locally Synchronous*. Circuit comprenant des blocks fonctionnement de manière synchrone mais échangeant les données de façon asynchrone entre eux.
HP	*High Performance*. Type de circuit majoritairement en activité
IP	*Intellectual Property*. Circuit décrivant une fonctionnalité (lien série, décodeur vidéo, …)
LOP	*Low Operating Power*. Type de circuit alternant périodes de veilles et d'activités
LSTP	*Low Standby Power*. Type de Circuit majoritairement en veille
PDL	*Programmable delay line*. Ligne a retard programmable généralement couplé à une DLL
PDSOI	*Partially Depleted*. Technologie SOI partiellement désertée
PPA	*Power Performance Area*. Critère de mesure de la qualité de conception d'un circuit.
QDI	*Quasi Delay Insensitive* Classe de circuit asynchrone quasiment insensible au délai.
RBB	*Reverse Body Bias*. Polarisation du substrat visant une réduction de la consommation du circuit.
SOI	Silicon-On-Insulator. Procédé utilisant un isolant sur le substrat

Titre : Conception pour l'efficacité énergétique des circuits intégrés

Résumé : Durant de nombreuses décennies, l'industrie de la microélectronique s'est développée selon la loi de Moore qui exprime la réduction de la surface des puces et l'augmentation de leur performance en fonction du temps. Cette loi se combine à l'échelle de Dennard qui stipule que la densité énergétique des transistors doit rester constante. Jusqu'au début des années 2000, les technologues ont pu ajuster la tension d'alimentation des circuits intégrés pour respecter cette mise à l'échelle. Mais, l'arrivée des technologies nanométriques avec leur capacité limitée de réduction de la tension a bousculé ce paradigme. Il est devenu dès lors essentiel pour suivre la loi de Moore de proposer des solutions au niveau de la conception des circuits intégrés afin de contrôler leur densité énergétique donc leur consommation. Les travaux proposés dans ce manuscrit essayent de répondre à ce besoin en partant des principaux contributeurs aux consommations statiques et dynamiques. Nous démarrons avec des solutions intuitives pour la réduction de ces consommations comme les registres à double fronts ou, encore, la mise en place d'interrupteurs de puissance. Nous explorons ensuite des voies pour optimiser l'efficacité énergétique et converger vers le point de consommation minimum qui préserve la capacité de calcul du circuit intégré. Enfin, nos derniers travaux autour du calcul sur évènements proposent une alternative intéressante à la logique synchrone par sa capacité intrinsèque à utiliser l'énergie de façon parcimonieuse.

Abstract: For many decades, the microelectronics industry has grown following the Moore's Law, which expresses the chip area reduction and performance increase over time. This law is combined with Dennard's scaling, which stipulates that the energy density of transistors must remain constant. Until the early 2000s, technologists were able to adjust the integrated circuit supply voltage for respecting this scaling rule. However, the arrival of nanometric technologies with their limited capacity to reduce voltage has shaken up this paradigm. Therefore, it is becoming essential to follow the Moore's law to propose design solutions to control the integrated circuit energy density and power consumption. The proposed work in this manuscript tries to answer these needs by mitigating the main contributors to static and dynamic power consumption. We first start with intuitive solutions reducing these consumptions such as dual-edge registers or power switches. We then explore different ways to optimize energy efficiency and converge towards the minimal power consumption while preserving the integrated circuit computing capabilities. Finally, our latest works on event-driven computing propose an interesting alternative to synchronous logic because of its intrinsic ability to sparingly use energy.