Hybridation CMOS/STT-MRAM des circuits intégrés pour la sécurité matérielle de l'internet des Objets

Mounia Kharbouche-Harrari

AIX-MARSEILLE UNIVERSITÉ
ECOLE DOCTORALE 353
IM2NP - UMR CNRS 7334
En collaboration avec le CEA Tech et Spintec

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Sciences pour l'Ingénieur
Spécialité : Micro et Nanoélectronique

Mounia KHARBOUCHE-HARRARI

Hybridation CMOS/STT-MRAM des circuits intégrés pour la sécurité matérielle de l'Internet des Objets

Soutenue le 09/12/2019 devant le jury composé de :

Bruno Rouzeyre	Professeur des Universités, LIRMM	Président
Jacques-Olivier Klein	Professeur des Universités, Paris Sud	Rapporteur
Lorena Anghel	Professeur à l'INP, Grenoble	Rapporteure
Jean-Luc Danger	Directeur d'études, Telecom ParisTech	Examinateur
Jean-Michel Portal	Professeur des Universités, IM2NP	Directeur de thèse
Gregory Di Pendina	Ingénieur de recherche CNRS, Spintec	Encadrant
Romain Wacquez	Ingénieur-chercheur, CEA-Tech	Encadrant
Jérémy Postel-Pellerin	Maître de conférences, IM2NP	Encadrant
Driss Aboulkassimi	Ingénieur-chercheur, CEA-Tech	Encadrant

Numéro national de thèse/suffixe local : 2019AIXM0621/039ED353

Remerciements

Ces travaux de thèse ont été menés au sein de l'équipe commune CEA tech et École des Mines de Saint-Étienne 'SAS' située à Gardanne, en collaboration avec les laboratoires CNRS-SPINTEC (équipe 'Spintronics IC Design') à Grenoble et IM2NP (équipe 'Mémoires') à Marseille.

Mes premiers remerciements vont à mon encadrement de thèse, Jean-Michel, Jérémy, Gregory, Romain et Driss. Je vais regretter ces trois années qui ont été d'une extrême richesse scientifique grâce à vous et à vos différentes expertises. Nous avons pu explorer ensemble un nouveau sujet de recherche mêlant sécurité et hybridation CMOS/STT-MRAM.

Un immense merci à Lorena Anghel et Jacques-Olivier Klein pour leurs rapports et leurs conseils sur l'amélioration de la compréhension de certains points de ce manuscrit de thèse. Merci également à Bruno Rouzeyre et Jean-Luc Danger pour avoir examiné ces travaux et pour les discussions constructives que nous avons pu partager lors de la soutenance.

Je tiens évidemment à remercier toutes les personnes avec qui j'ai partagé un bureau au fil de ces trois années : Thomas, Paul, David, Raphaël, Alexis, Damien, Benjamin, Alexandre, Antoine, Meriem, Amina, Clément, Lina, Élise et Rémi. Cela a été un immense plaisir de partager avec vous au quotidien. Merci évidemment à tous les membres de l'équipe commune 'SAS' avec qui j'ai eu plaisir d'échanger et qui êtes restés à mon écoute en tout temps.

Je souhaite également remercier toutes les personnes qui ont facilité l'accomplissement de ces travaux : Anaïs, Michelle, Séverine et Bastien. Tous les membres de la plateforme Micro-PackS pour votre bonne humeur au quotidien ainsi que pour tout votre soutien lors de la mise en place du banc de test des circuits fabriqués lors du projet GREAT.

Toute l'équipe mémoire de l'IM2NP et en particulier Marc et Vincenzo qui m'ont été d'une grande aide lors de ma première année de thèse et avec qui j'ai eu un plaisir immense d'échanger. Je tiens également à remercier Guillaume, Rana, Lucian, François et Odilia de Spintec pour tout le contenu technique que nous avons partagé, pour votre sympathie et surtout votre accueil au sein de l'équipe ; malgré la distance, cela n'a jamais été un frein.

Pour finir, merci à ma famille et mes amis qui m'ont beaucoup soutenu au fil de ces années ; en particulier mon époux, Elias, qui m'a énormément soutenu pendant toutes les phases de la thèse, pendant les hauts et surtout les phases de stress. Encore merci à toi !

Bien à vous,
Mounia

Table des matières

Introduction générale

Depuis l'avènement de l'informatique, la quantité de données échangées entre différentes entités ou objets a littéralement explosé. En effet, 2,5 Exaoctets (10^{18}) sont créés chaque jour [1], [2]. À titre de comparaison, ce nombre correspond au nombre d'informations produites entre le début de l'humanité et l'année 2003. Ce chiffre ne cesse d'augmenter, mais la modification des usages de la société rend difficile une estimation à long terme.

Le développement rapide de l'Internet des Objets lors de cette dernière décennie n'a pas suffisamment pris en compte la sécurité des communications. En effet, bien que certains de ces objets traitent et transmettent des informations confidentielles, dont l'intégrité et l'authenticité doivent être garanties, la sécurité n'est pas une contrainte forte lors de leur développement. C'est ainsi qu'une personne *lambda* peut avoir un libre accès à des caméras de vidéosurveillances de lieux privés ou publics, à des webcams d'ordinateurs [3], aux données traitées par des bracelets connectés ou autres, en toute liberté. En effet, dans ce contexte où le nombre de fabricants d'objets connectés est important, le *time-to-market* occupe une place essentielle. Dès la fonctionnalité du composant, celui-ci est commercialisé sans études supplémentaires sur ses besoins sécuritaires.

Afin de sécuriser ces objets connectés, des algorithmes de cryptographie permettant le chiffrement des messages, peuvent être utilisés. Ces algorithmes permettent, le plus souvent, grâce à une clé confidentielle et personnelle de chiffrer des messages. Le déchiffrement ne peut être réalisé qu'en connaissant cette clé. Ainsi, deux entités peuvent communiquer en toute "sécurité". Toutefois, bien que ces chiffrements soient mathématiquement sûrs, ils sont tout de même sensibles à différents types d'attaques que nous pouvons désigner par attaques physiques. Ces méthodes visent à récupérer la clé de chiffrement utilisée par un algorithme de cryptographie, afin de pouvoir déchiffrer n'importe quel message.

Dans le cadre de cette thèse, nous ne traiterons que des attaques qui sont dites physiques car elles nécessitent l'accès au circuit. Ces attaques conduisent soit à des fuites (de courant par exemple) susceptibles d'être corrélées aux données que le circuit manipule, soit à la création d'une perturbation qui conduit à la modification des données stockées dans la mémoire, à la modification du programme en exécution, ou à l'injection d'une faute dans l'exécution d'un algorithme de chiffrement. Dans ce dernier cas, l'attaquant vise à retrouver la clé de chiffrement.

Bien avant l'aspect sécuritaire, la consommation des objets connectés a été un enjeu majeur pour leur déploiement. Pour cela, l'intégration de mémoires non-volatiles dans les circuits intégrés a démontré un avantage indéniable dans la consommation des applications normalement arrêtées ("*Normally-off*"). En effet, la consommation statique (ou de veille) d'un circuit qui est non-négligeable dans certains cas, est presque complètement inhibée dans le cas de l'hybridation des technologies CMOS et des mémoires non-volatiles émergentes. Les données sont au préalable stockées dans les mémoires avant la mise hors tension du circuit, ce dernier ne consommant quasiment plus d'énergie. Lorsque l'application doit reprendre son fonctionnement, les informations sont restaurées.

Dans ce contexte et dans ce travail, nous nous intéresserons à l'hybridation des algorithmes de cryptographie avec la technologie mémoire émergente STT-MRAM qui présente des propriétés

exceptionnelles pour une intégration dans les objets connectés (faible consommation, forte densité et rapidité de vitesse de fonctionnement). Pour ce faire, dans un premier temps, l'étude des éléments mémoires STT-MRAMs face aux injections de fautes par laser est réalisée afin de déterminer leur sensibilité ou robustesse face aux attaques. Ces éléments sont ensuite intégrés dans des algorithmes de cryptographie en vue du développement d'un chiffrement très faible consommation. L'algorithme de cryptographie légère "PRESENT" est privilégié. En effet, l'implémentation matérielle de cet algorithme est l'une des plus efficaces pour les objets connectés.

Afin de réduire les possibilités de manipulations frauduleuses de ces chiffrements, deux contre-mesures sont développées. Alors que la première désignée par DDHP (*Dual Detection of Heating and Photocurrent attacks*) se présente comme capteur d'attaques qui ne vise que la détection des perturbations ciblant les technologies CMOS et/ou STT-MRAM, sans les corriger; la seconde intitulée DTR (*Dual Technology Redundancy*) consiste au-delà de la détection, à pouvoir restituer un contexte sain dans les algorithmes de cryptographie après que ces derniers aient été ciblés par des perturbations physiques. En effet, cette seconde contre-mesure consiste à comparer deux chemins d'attaques, un chemin CMOS pur à un chemin hybride CMOS/STT-MRAM.

Chapitre 1

Sécurisation des circuits intégrés

> *If you say you understand quantum mechanics, then you don't understand quantum mechanics*
>
> **— Richard Feynman**

Presque tout objet peut être connecté à un réseau, constituant ainsi l'Internet des Objets. Ces objets connectés envahissent le marché pour répondre à différentes applications du quotidien. Ils ont des exigences accrues en terme de consommation et de surface. Toutefois, leur sécurité face aux attaques extérieures est encore peu étudiée, bien qu'elle soit un enjeu majeur de cette décennie et doit donc être garantie.

Ce premier chapitre est introduit par une description des différentes technologies mémoires existantes, traditionnelles et émergentes, dont la STT-MRAM qui est utilisée dans ces travaux. Puis, dans une seconde partie, les exigences des objets connectés en terme de sécurité sont développées. Enfin, ce chapitre est conclu par une dernière partie sur l'hybridation des technologies CMOS et STT-MRAM pour le développement d'architectures sécurisées en vue de leur intégration dans l'Internet des Objets.

Sommaire

1 Enjeux de la sécurité dans l'Internet des Objets

Lors de cette dernière décennie, les habitudes quotidiennes de la population française et à fortiori mondiale, ont été bouleversées. Les objets connectés sont devenus une réalité et ont pris une place considérable dans notre quotidien. Ces circuits se multiplient de façon exponentielle et apparaissent sous différentes formes, pour diverses applications. Les objets connectés ont été dénombrés à plus de 20 milliards, dont 7 milliards en ne considérant que les objets hors ordinateurs, tablettes et smartphones. L'estimation de la place de l'Internet des Objets dans les prochaines décennies indique qu'ils pourront atteindre plus de 70 milliards (dont 25 milliards d'objets connectés) en 2025 [4], [5]. Ceux-ci sont présents dans diverses applications grand public comme l'automobile, la domotique, la santé ou encore les applications militaires, illustrées Figure 1.1. Comme les ordinateurs, chaque objet possède sa propre adresse IP *(Internet Protocol)* permettant son identification dans un réseau informatique [6].

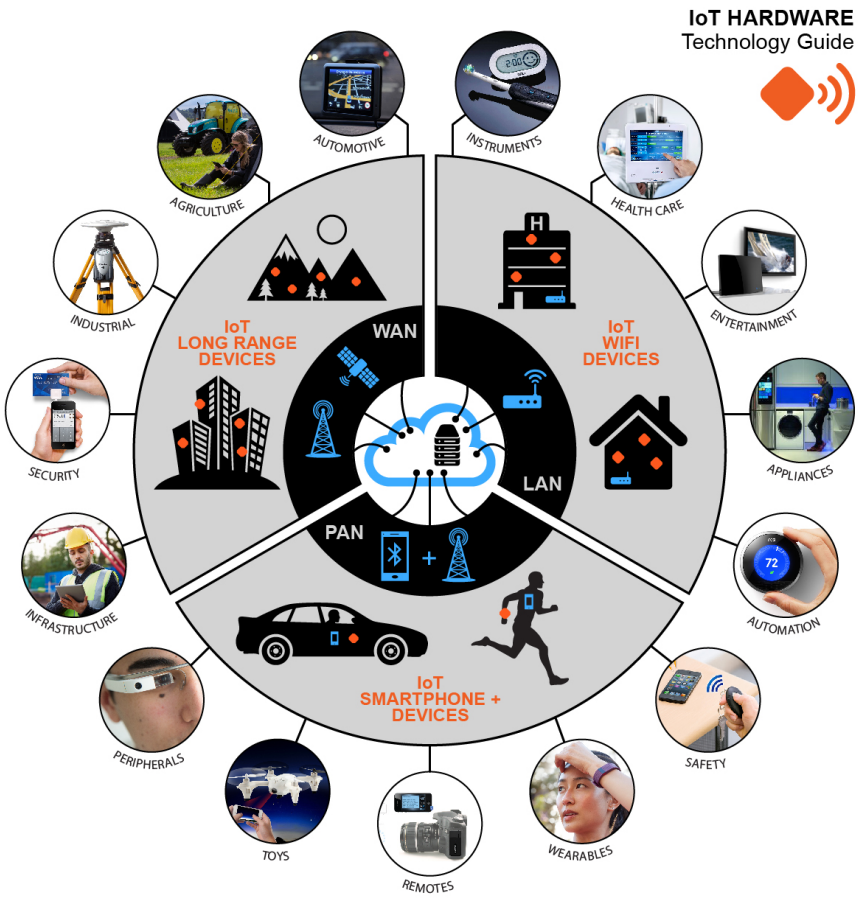


FIGURE 1.1 – Illustration des diverses applications de l'Internet des Objets [7].

Toutefois, lors du développement d'objets connectés communicants avec un environnement extérieur en mettant à disposition des informations sur le réseau, la caractéristique qui est encore peu prise en compte est la sécurité. En effet, ces dispositifs traitent des informations sensibles qui peuvent être publiques ou privées. Dans le second cas, la confidentialité, l'authenticité et l'intégrité des données traitées doivent être assurées à tout moment et inhiber toute attaque extérieure qui pourrait s'étendre et toucher l'ensemble du réseau d'objets connectés. Les attaques visant ces dispositifs peuvent être de deux natures différentes :

— Matérielle (ou *Hardware*) : cette catégorie d'attaques se caractérise par des attaques physiques qui peuvent soit perturber le fonctionnement physique du circuit, ces attaques sont

dites par perturbation, soit observer les émissions d'un circuit pour en déterminer les données manipulées, ces dernières sont dites par observation. Pour cela, il est nécessaire d'avoir un accès physique à l'objet. Les attaques par perturbation peuvent par exemple viser la mémoire où sont stockées les données sensibles sur le long terme (mémoire non-volatile de type Flash par exemple) ou les données en cours de traitement dans les registres ou dans les mémoires volatiles, de type mémoire statique ou – *Static Random Access Memory* – (SRAM) par exemple. Les attaques par observation ciblent les algorithmes de cryptographie afin d'en extraire la clé utilisée pour le chiffrement et donc pouvoir déchiffrer tout message crypté.

— Logicielle (ou *Software*) : cette seconde catégorie d'attaques exploite l'utilisation de protocoles de communication dans les objets connectés, afin de réaliser des attaques à distance, via le réseau. Une instanciation non-adaptée de ces protocoles peut induire des fuites et des intrusions dans les données ou les programmes qui sont traités par les objets. Ces attaques logicielles peuvent également être induites en exploitant des portes dérobées dont l'attaquant, contrairement à l'utilisateur, a connaissance. Contrairement aux attaques matérielles, l'accès physique à l'objet n'est pas forcément nécessaire dans ce cas.

Ces problématiques de sécurité sont aujourd'hui au centre des débats. Pour cela diverses solutions peuvent être étudiées. Dans le cadre de ces travaux, nous ne traiterons que des attaques matérielles qui peuvent viser les circuits intégrés et principalement viser les éléments mémoires où sont stockées les données sensibles de l'utilisateur. Effectivement, cette thèse traitera de l'intégration de la technologie mémoire émergente *Spin-Transfer Torque MRAM* – (STT-MRAM) dans l'Internet des Objets.

Dans un premier temps, son niveau de sécurité face à des perturbations extérieures est étudié. Puis, sa capacité d'hybridation avec la technologie CMOS est analysée afin d'obtenir des solutions cryptographiques plus performantes en terme de sécurité, tout en étudiant leur impact en terme de consommation et de surface. Avant de développer ces différentes études, ce premier chapitre met en avant un état de l'art non-exhaustif des technologies mémoires et de leur sécurité en vue de leur intégration et hybridation avec des procédés CMOS.

2 Les technologies mémoires

2.1 Les technologies mémoires traditionnelles

Les mémoires sont des composants essentiels de l'industrie de la microélectronique. Elles sont principalement utilisées pour le stockage de données ou de programmes. Le stockage de l'information dans ces composants, selon leur positionnement dans une architecture de calcul fait appel à différentes technologies mémoires comme représenté sur la Figure 1.2. Il existe deux familles principales de mémoires : les mémoires volatiles et mémoires non-volatiles.

2.1.1 Les mémoires volatiles

Les mémoires volatiles sont des dispositifs dont la préservation des données ne dépend que de son alimentation. Dès que celles-ci ne sont plus alimentées, l'information stockée est perdue. Cette catégorie de mémoires est principalement utilisée pour implémenter la mémoire cache ou les mémoires de travail ou mémoires primaires. Les mémoires statiques ou – *Static Random Access Memories* – (SRAMs), composent les mémoires dominantes embarquées dans un microcontrôleur alors que les mémoires dynamiques ou – *Dynamic Random Access Memories* – (DRAMs), représentent la plus importante part de marché en mémoires volatiles isolées (ou *stand-alone*).

Cette catégorie de mémoires présente des avantages majeurs facilitant son utilisation et intégration dans les circuits électroniques. Dans un monde connecté où les données communiquées

sont importantes, la rapidité de fonctionnement et la densité sont des atouts majeurs, des attraits mis en avant par les SRAMs et DRAMs.

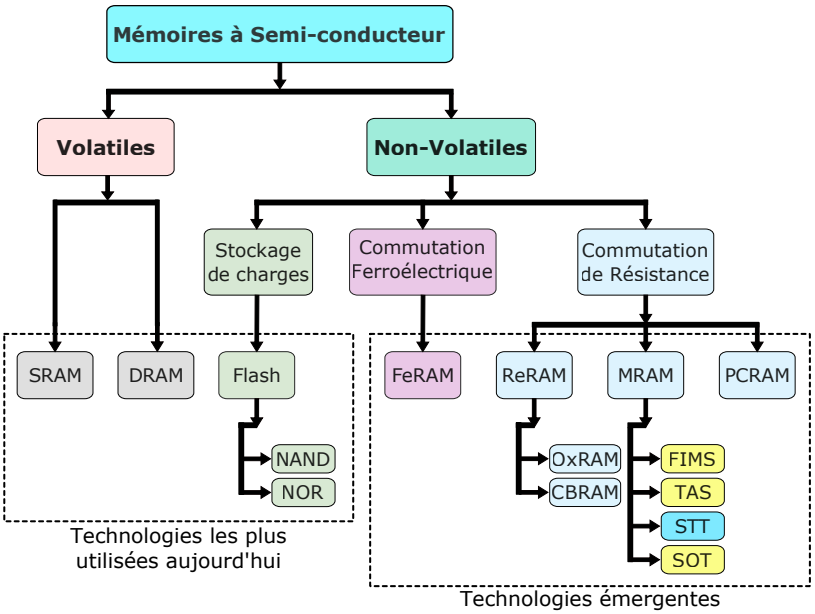


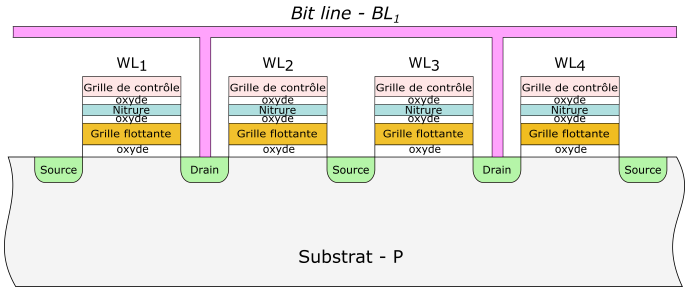
FIGURE 1.2 – Classification des différentes technologies de mémoires à semi-conducteur.

2.1.2 Les mémoires non-volatiles

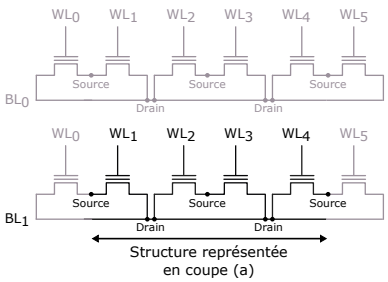
Dans les mémoires non-volatiles, l'information stockée est préservée même sans alimentation, comme par exemple dans le cas d'une clé USB. Tant que l'information écrite dans la mémoire n'est pas effacée ou modifiée, alors la donnée est invariable avec ou sans alimentation. Cette mémoire est principalement présente pour le stockage de masse. La Flash détient aujourd'hui la plus importante part de marché en terme de mémoires non-volatiles et notamment pour les mémoires isolées. La technologie Flash est la plus mature, elle est étudiée depuis 1984 [8]. Il existe deux types de mémoires Flash : la mémoire NOR et la mémoire NAND.

2.1.2.1 La NOR :

La technologie mémoire NOR est une architecture présentant une vitesse d'accès aux points mémoires importante puisque chaque point peut être adressé indépendamment des autres, ce qui est un atout pour le stockage d'instructions. En effet, chaque transistor est adressé par une *bitline* (BL) et une *wordline* (WL), comme illustré sur la Figure 1.3. Toutefois, cette caractéristique conduit également à la réduction de la densité de l'architecture. La Figure 1.3 illustre une vue en coupe et le schéma de fonctionnement d'une cellule mémoire NOR.



(a) Représentation en coupe de la mémoire NOR Flash.

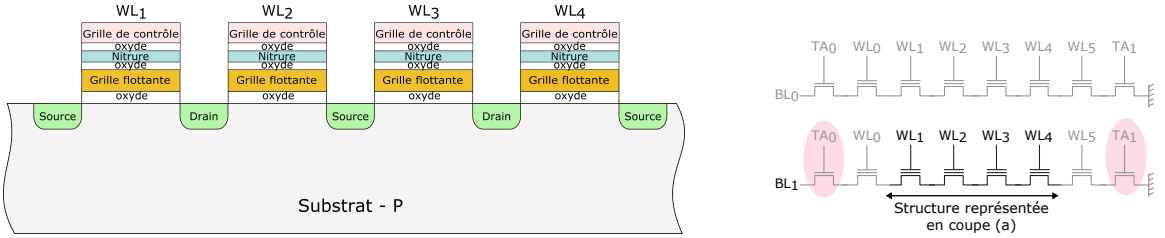


(b) Schéma de fonctionnement.

FIGURE 1.3 – La mémoire Flash NOR.

2.1.2.2 La NAND :

La catégorie de mémoires Flash de type NAND présente une plus forte densité que la technologie mémoire NOR. Toutefois, l'adressage d'un point mémoire précis ne peut être réalisé directement. En effet, dans une mémoire Flash de type NAND, toutes les mémoires sont connectées en série et la BL est connectée à un transistor d'accès TA_0, comme représenté sur la Figure 1.4. Ainsi, les temps d'écriture et de lecture d'une cellule particulière dans la matrice sont plus lents que pour son homologue NOR. D'autre part, afin de réaliser la lecture d'un plan mémoire, il est nécessaire d'activer toutes les WLs mises en jeu dans cette matrice. La Figure 1.4 représente une vue en coupe et le schéma de fonctionnement d'une mémoire Flash de type NAND.



(a) Représentation en coupe de la mémoire NAND Flash.

(b) Schéma de fonctionnement.

FIGURE 1.4 – La mémoire Flash NAND.

La Recherche et Développement (R&D) sur les mémoires non-volatiles émergentes est en perpétuelle évolution. Ces études sont réalisées afin d'atteindre une mémoire idéale dont les caractéristiques seraient les suivantes :

— Une consommation minimale déterminée par des tensions de fonctionnement de la mémoire proche des tensions utilisées pour le fonctionnement du CPU.

— Une endurance présentant au moins 10^{15} cycles d'écriture/effacement répondant aux contraintes de durée de vie des produits les plus exigeants.

— Un temps de lecture/écriture de la mémoire de l'ordre de la nanoseconde, permettant ainsi de réduire le gap entre la vitesse de traitement des cœurs de calcul et l'accès aux informations mémorisées.

— Une forte densité afin de réduire la surface silicium nécessaire pour fabriquer le produit et donc pouvoir plus facilement l'intégrer dans les applications embarquées et baisser en même temps les coûts.

— La non-volatilité pour conserver les informations même en l'absence d'alimentation.

Cette mémoire idéale serait définie comme une mémoire fonctionnant à la vitesse et au niveau de tension des mémoires volatiles de type SRAMs, tout en ayant la non-volatilité des mémoires Flash.

2.2 Les mémoires émergentes

Les différentes mémoires émergentes aujourd'hui étudiées se rapprochent par bien des points de la mémoire idéale, même si certains obstacles subsistent. Après la présentation de certaines de ces catégories de mémoires, il sera alors possible de comparer leurs caractéristiques, même si cela reste toujours subjectif.

2.2.1 Les mémoires résistives ReRAM

Les mémoires résistives *Resistive Random Access Memories* – (ReRAMs ou RRAMs) font partie des mémoires non-volatiles émergentes des plus prometteuses, présentant divers avantages comparé à la technologie Flash. Ces mémoires sont des dispositifs qui se positionnent physiquement

dans le *Back-End of Line* – (BEoL), contrairement à la mémoire Flash qui est définie directement sur le substrat. Cette caractéristique de positionnement des mémoires résistives sur les niveaux de métallisations supérieurs permet d'augmenter la densité des plans mémoires. Cette technologie émergente est aujourd'hui la plus utilisée dans les applications embarquées [9].

2.2.1.1 Oxide Random Access Memory - OxRAM :

La technologie mémoire *Oxide Random Access Memory* – (OxRAM) est une structure résistive de type Métal-Isolant-Métal qui est l'une des deux mémoires émergentes privilégiées par l'*International Technology Roadmap for Semiconductors* – (ITRS) [10] pour remplacer les Flashs. Elle combine les avantages de la mémoire Flash en terme de non-volatilité et de la mémoire SRAM en terme d'endurance et de consommation bien que son principal défaut reste la variabilité. Les deux états mémoires de Basse résistivité L_{RS} (*Low Resistive State*) et Haute résistivité H_{RS} (*High Resistive State*) sont induits respectivement par la création ou la destruction d'un filament conducteur d'anions d'Oxygène dans l'isolant compris entre les deux électrodes. Ce filament est créé par le passage d'un courant à travers la structure.

Nous pouvons distinguer deux catégories de mémoires OxRAMs; une version unipolaire (les commutations sont réalisées sous la même polarité mais à deux amplitudes distinctes) et bipolaire (qui est aujourd'hui celle qui est principalement étudiée).

Les commutations dans une OxRAM bipolaire sont réalisées pour deux polarités inverses. La tension permettant de réaliser l'opération de *Set* (commutation d'un état H_{RS} vers L_{RS}) a une polarité positive, inverse à la tension effectuant le *Reset* (L_{RS} vers H_{RS}), comme illustré Figure 1.5. Cette technologie mémoire est une candidate exceptionnelle dans le remplacement des mémoires conventionnelles pour les nœuds technologiques inférieurs à 20 nm [11], [12]. Toutefois, cette technologie peut induire des niveaux de courants élevés dans la structure pouvant conduire à sa destruction. C'est pourquoi il est nécessaire d'imposer lors des phases de programmation, des courants de *"compliance"* ou limite maximale à ne pas dépasser, comme illustré sur la Figure 1.5.

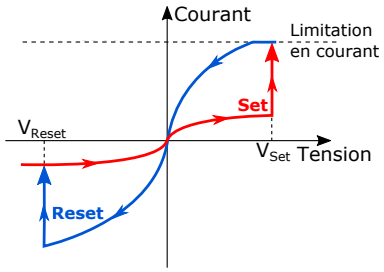


FIGURE 1.5 – Représentation du fonctionnement électrique d'une OxRAM bipolaire [12].

2.2.1.2 Conductive Bridging Random Access Memory - CBRAM :

La seconde technologie mémoire résistive décrite ici est la *Conductive Bridging Random Access Memory* – (CBRAM). Le changement d'état de cette architecture est induite par une commutation électrochimique. En effet, cette structure est composée de trois couches de matériaux différents, une électrode active et une électrode inerte séparées par un oxyde ou une électrolyte, comme représenté sur la Figure 1.6. L'état de faible résistivité L_{RS} (respectivement de forte résistivité H_{RS}) est alors atteint lors de la création (resp. destruction) d'un filament d'ions métalliques dans cet oxyde (comme par exemple d'Argent) [13], comme illustré Figure 1.6. Cette technologie est très attractive pour son utilisation de niveaux de tension de fonctionnement faibles.

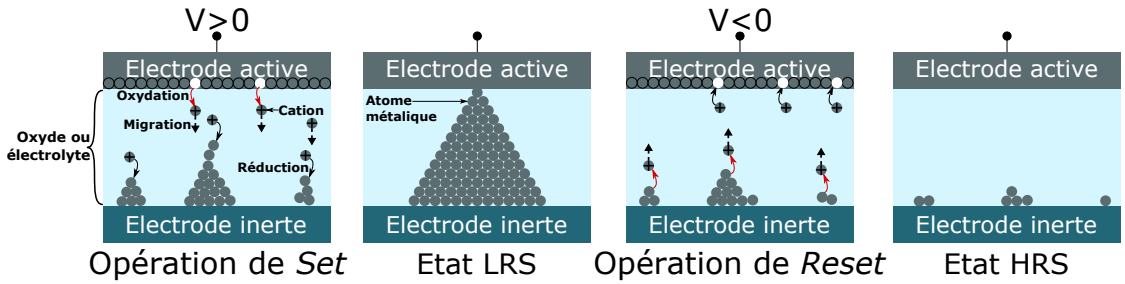


FIGURE 1.6 – Illustration du fonctionnement interne de la mémoire CBRAM [12].

2.2.2 Prérequis à la MRAM

Dans le contexte des mémoires non-volatiles émergentes, une seconde catégorie s'est illustrée par ses performances et caractéristiques intéressantes pour l'Internet des Objets, la technologie mémoire magnétique – ou *Magnetic Random Access Memory* – (MRAM), illustrée Figure 1.2. Afin d'introduire cette technologie mémoire, il est nécessaire de réaliser quelques rappels de physique et de magnétisme, principalement concernant les matériaux qui entrent en jeu dans la composition de ces structures. En effet, les MRAMs sont des technologies mémoires composées de matériaux magnétiques de diverses natures. La nature de ces matériaux est déterminée par les caractéristiques des moments magnétiques qui les composent.

2.2.2.1 Le moment magnétique :

La fonction principale du magnétisme est caractérisée par le moment magnétique des électrons, qui gravitent autour du nucléon. En effet, l'électron qui tourne autour du nucléon N génère un courant en sens inverse à son déplacement. Pour une boucle de courant d'intensité I, comme représentée sur la Figure 1.7, de surface *S*, alors le moment magnétique $\vec{m}$ exprimé en A.m^2 est défini par l'Équation 1.1 où $\vec{n}$ est le vecteur unitaire perpendiculaire à la surface de la boucle de courant.

$$\vec{m} = \text{I.S.}\vec{n} \tag{1.1}$$

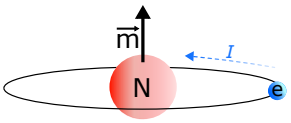


FIGURE 1.7 – Représentation du moment magnétique de spin $\vec{m}$.

Ainsi, ce moment magnétique peut être défini comme la force résultante induite par cette boucle de courant.

En outre, il est également défini par son moment magnétique angulaire et son moment magnétique intrinsèque ou spin. Alors que le premier est nul lorsque l'atome est immobile et sera donc négligé ici, le second quant à lui provient de la contribution des couches énergétiques non-pleines (les couches "3d" ou "4f" par exemple, comme développé en Annexe A). En effet, le moment magnétique total de l'atome correspond à la somme des moments magnétiques des électrons qui le forment. Ainsi, l'énergie des moments magnétiques de spins est non-nulle tant que les orbitales ne sont pas toutes remplies et donc tant que les moments magnétiques ne s'annulent pas.

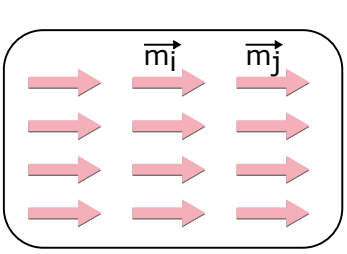
2.2.2.2 Les matériaux magnétiques :

Selon le tableau de D.I. Mendeleïev [14], la majorité des matériaux peuvent être décomposés en différentes catégories de matériaux : des matériaux ferromagnétiques, des matériaux antiferromagnétiques, des matériaux diamagnétiques et des matériaux paramagnétiques, dépendant des caractéristiques physiques des moments magnétiques intrinsèques de chaque composant.

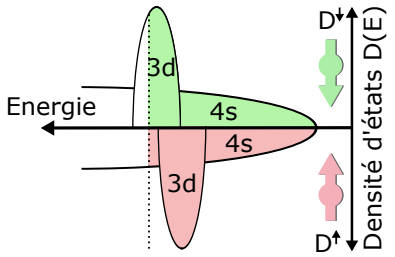
— Les matériaux ferromagnétiques sont des éléments pour lesquels tous les moments magnétiques ont la même direction (co-linéaires) et sont orientés dans le même sens de façon intrinsèque et spontanée, sans besoin d'application d'un champ magnétique extérieur [15], illustré Figure 1.8.a. Cet effet est induit par une énergie d'échange E_c qui est présente entre deux moments magnétiques directement voisins $\overrightarrow{m_i}$ et $\overrightarrow{m_j}$, tel que décrit en Équation 1.2 :

$$E_c = -J \sum_{i,\,j} \overrightarrow{m_i} \cdot \overrightarrow{m_j} \tag{1.2}$$

La constante de couplage d'échange J est positive afin d'assurer l'alignement des moments magnétiques. La Figure 1.8.b représente le remplissage réel de la couche "3d", où les *spins-down* (en vert) ne sont pas suffisants pour la remplir complètement.



(a) Représentation conventionnelle de l'organisation des moments $\overrightarrow{m}$.



(b) Densité d'états des *spins-up* et *spins-down*, organisation réelle des moments $\overrightarrow{m}$.

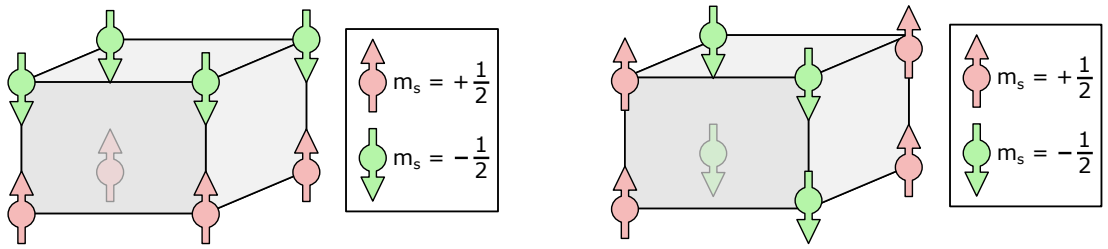
FIGURE 1.8 – Illustration du matériau ferromagnétique.

— Les matériaux antiferromagnétiques (AF) sont les composants pour lesquels tous les moments magnétiques sont co-linéaires bien que deux moments successifs ont un sens opposé (soit en x, y ou z). Ainsi, les matériaux antiferromagnétiques ont une aimantation globale nulle. Il existe un large nombre de classes de matériaux AFs, qui dépendent de l'arrangement des *spins-up* par rapport aux *spins-down* dans la matrice [15]. Pour exemple, soit les deux catégories d'antiferromagnétiques les plus représentatives : AF de type a et AF de type c, comme illustrés Figure 1.9.

- type a : Dans une structure cubique simple par exemple, les 4 coins de la face supérieure ont un spin inverse aux 4 coins de la face inférieure. Le spin est fixé selon le plan (100). Ce matériau est illustré par la Figure 1.9.a.
- type c : Dans la même structure cubique simple, ce type correspond au spins de même orientation magnétiques organisés selon la diagonale de cette structure (selon le plan (111)). Ce matériau est illustré par la Figure 1.9.b.

L'équation 1.2 représente également les matériaux antiferromagnétiques avec la seule différence que la constante de couplage d'échange J sera dans ce cas négative, imposant donc à deux moments magnétiques voisins d'être dans des sens inverses (bien qu'ils restent co-linéaires).

Les matériaux ferromagnétiques et antiferromagnétiques sont liés soit à la couche "3d" soit à la couche "4f" qui sont partiellement remplies.



(a) AF de type a.

(b) AF de type c.

FIGURE 1.9 – Deux modèles de matériaux antiferromagnétiques présentant la position des moments magnétiques (positifs et négatifs) dans une structure cubique.

— Les matériaux paramagnétiques ont des moments magnétiques dont le sens et la direction ne sont pas intrinsèquement prédéfinis. Ils s'orientent dans toutes les directions de l'espace de façon aléatoire, comme illustrés Figure 1.10.a. Cette catégorie de matériaux magnétiques met ainsi en jeu des moments magnétiques voisins qui peuvent être considérés comme indépendants les uns des autres et qui n'interagissent pas entre eux.

(a) Organisation des moments magnétiques sans application de champ.

(b) Ordonnancement des moments magnétiques $\vec{m}$ sous l'effet d'un champ $\vec{B}$.

FIGURE 1.10 – Les matériaux paramagnétiques sous différentes conditions extérieures.

Toutefois, lors de l'application d'un champ magnétique extérieur, ces moments magnétiques se polarisent et deviennent le temps de la polarisation magnétique des ferromagnétiques orientés dans le sens du champ magnétique appliqué [15], comme illustré Figure 1.10.b. D'autre part, une augmentation de la température a l'effet inverse sur ces moments et va davantage favoriser une orientation aléatoire.

— Les matériaux diamagnétiques sont des matériaux à faible susceptibilité magnétique négative. Contrairement aux matériaux paramagnétiques, les matériaux diamagnétiques, lorsqu'ils sont soumis à un champ magnétique, voient tous les électrons s'orienter dans le sens opposé au champ magnétique [15]. Ainsi, l'aimant et le matériau se repoussent.

2.2.3 Les technologies MRAMs

La découverte de M. Julliere [16] à l'Institut National des Sciences Appliquées de Rennes en 1975 de la variation de la conductance d'un empilement "magnétique - isolant - magnétique" selon l'orientation des couches ferromagnétiques conduit au développement de la Magnéto-Résistance Tunnel ou – *Tunnel MagnetoResistance* – (TMR). En 1988 et 1989, les équipes d'A. Fert [17] et P. Grünberg [18] ont découvert la Magnéto-Résistance Géante ou – *Giant MagnetoResistance* – (GMR). Ils ont été récompensés en 2007 du prix Nobel de physique pour cette conquête d'une nouvelle voie dans le stockage des données et ont ouvert de nouvelles voies à la recherche. Ce fut l'avènement de la spintronique. L'électron n'est plus considéré que par sa charge comme dans l'électronique traditionnelle mais également pour sa propriété quantique : le spin, d'où le nom spintronique.

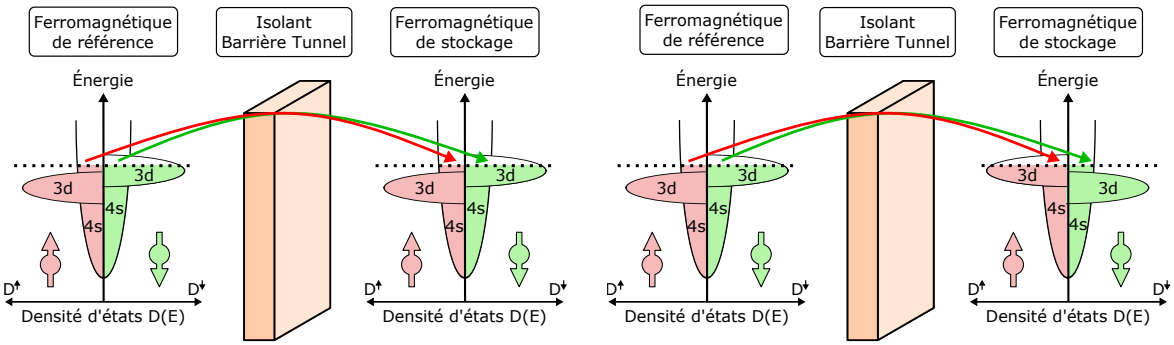
Ces découvertes conduirent au développement des mémoires MRAMs. Cette technologie est composée d'un isolant non-magnétique (ou barrière tunnel) positionné entre deux matériaux ferromagnétiques, constituant ainsi une hétérostructure magnétique appelée Jonction Tunnel Magnétique – ou *Magnetic Tunnel Junction* – (JTM) [19], illustrée sur la Figure 1.11.a.

(a) La Jonction Tunnel Magnétique - JTM. (b) État P. (c) État AP.

FIGURE 1.11 – Illustration de la Jonction Tunnel Magnétique (JTM) et de ses états.

L'orientation magnétique de l'une de ces couches ferromagnétiques est fixe, elle est définie comme couche de référence. La seconde couche ferromagnétique peut changer l'orientation de son aimantation, elle est désignée par couche libre ou couche de stockage. L'orientation magnétique de la couche libre par rapport à la couche de référence permet de définir l'information stockée.

Lorsque les deux ferromagnétiques composant cette jonction ont des aimantations qui sont dans une orientation commune, alors la JTM est dans un état parallèle P (faible résistivité notée R_P qui par convention correspond à un '0' logique), illustrée Figure 1.11.b. Dans ce cas, les électrons de spins majoritaires de l'une de ces couches (*spins-up*, illustrés sur la Figure 1.12.a) peuvent traverser la barrière tunnel jusqu'à la seconde couche ferromagnétique qui a ses électrons de spins majoritaires orientés dans la même aimantation magnétique.

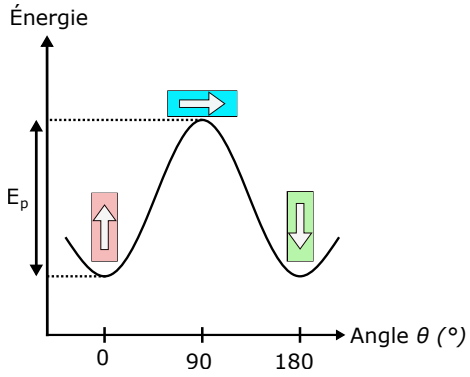


(a) Configuration parallèle des 2 ferromagnétiques composant la structure MRAM. (b) Configuration anti-parallèle des 2 ferromagnétiques composant la structure MRAM.

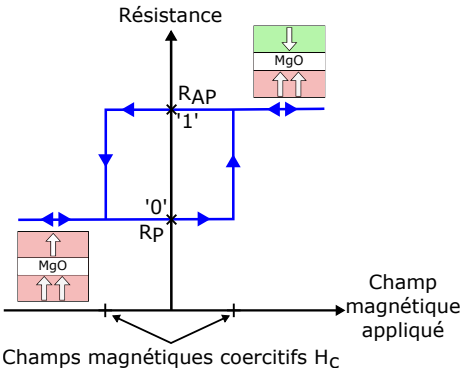
FIGURE 1.12 – Configurations possibles des deux ferromagnétiques composant la structure JTM.

Inversement, si les moments magnétiques de la couche de stockage sont co-linéaires à ceux de la couche de référence mais dans un sens opposé, alors la résistance est définie comme anti-parallèle AP (forte résistivité notée R_{AP} qui par convention correspond à un '1' logique), comme illustrée Figure 1.11.c. Dans ce cas, les électrons de spins majoritaires (*spins-up*, illustrés sur la Figure 1.12.b) qui doivent traverser l'isolant par effet tunnel sont minoritaires de l'autre côté (les *spins-down* sont majoritaires dans le second ferromagnétique). Le courant pouvant traverser la structure est faible et la résistance du point mémoire est importante.

Seuls ces deux états P et AP sont possibles dans les mémoires MRAMs. Tout autre état intermédiaire de la couche de stockage nécessiterait une énergie permanente E_p pour être stabilisé, comme représenté par la couche dont l'aimantation est horizontale (pour un angle de 90°) sur la Figure 1.13.a.



(a) Angle d'orientation de la couche de stockage par rapport à la référence.



(b) Évolution de la résistivité en fonction du champ magnétique.

FIGURE 1.13 – Orientation des deux couches ferromagnétiques dans une MRAM et évolution de la résistivité en fonction du champ magnétique.

Différentes générations de cette famille de mémoires ont été proposées. La classification de ces catégories est principalement due au moyen utilisé pour commuter la structure d'un état logique à l'autre. En effet, la commutation peut être provoquée par deux mécanismes : soit un champ magnétique (représenté sur la Figure 1.13.b) soit un courant. Le champ magnétique coercitif correspond à la force minimale à fournir à la structure afin de la commuter d'une aimantation à l'autre. De plus, il peut être noté que la stabilité thermique des composants peut modifier ce champ coercitif [20]. L'accroissement de la température peut donc modifier l'aimantation de la jonction.

La TMR exprime la proportion relative de la résistivité de l'état AP par rapport à l'état P. Cette grandeur est définie par l'Équation 1.3. L'évolution de la technologie mémoire MRAM est directement corrélée à l'augmentation significative de la TMR au fil de son développement et des matériaux utilisés, permettant ainsi de distinguer plus facilement les deux états de la mémoire.

$$TMR = \frac{R_{AP} - R_P}{R_P} \tag{1.3}$$

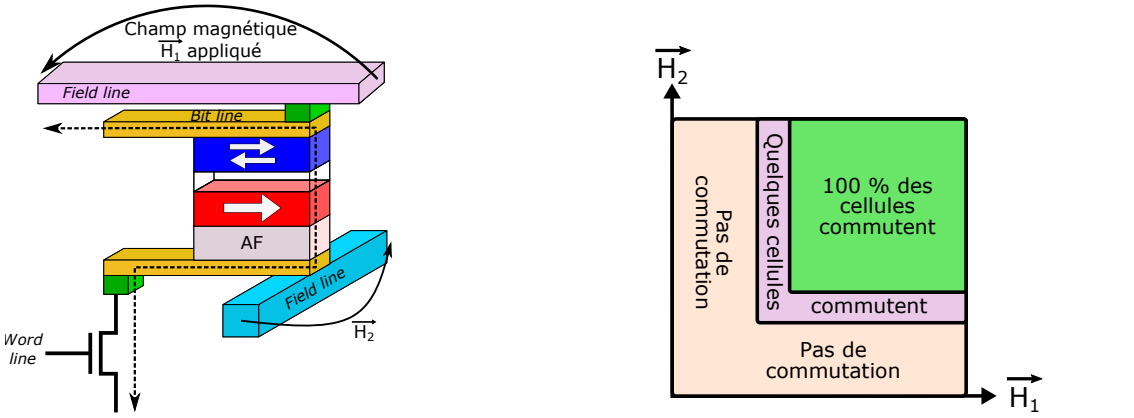
Bien que le matériau de prédilection de la barrière tunnel ait d'abord été l'Oxyde d'Aluminium (Al_2O_3) sous sa forme amorphe, il a ensuite été remplacé par de l'Oxyde de Magnésium (MgO) dans sa forme cristalline, le rendement de TMR de ce dernier étant bien supérieur. Alors que l'Oxyde d'Aluminium induit des TMRs observées inférieures à 100 % à température ambiante [21], [22], [23], l'Oxyde de Magnésium conduit à un passage cohérent des électrons dans l'empilement [24], [25], ce qui peut atteindre une TMR supérieure à 600 % à température ambiante et plus de 1000 % à basse température [26].

Dans cette section, nous décrirons quatre technologies mémoires MRAMs : la mémoire *Field Induced Magnetic Switching-Toggle* – (FIMS-Toggle), *Thermally Assisted Switching MRAM* – (TAS-MRAM), *Spin-Transfer Torque MRAM* (STT-MRAM) et *Spin-Orbit Torque MRAM* – (SOT-MRAM).

2.2.3.1 Field Induced Magnetic Switching-Toggle (FIMS-Toggle MRAM) :

La mémoire FIMS-Toggle utilise deux champs magnétiques simultanés pour commuter la JTM, via des lignes de cuivre qui passent en dessous (*word line*) et au-dessus (*bit line*) de l'élément mémoire, comme représenté Figure 1.14.a. En effet, ces deux champs permettent de fixer la position en x et en y de la jonction à écrire, dans la matrice de jonctions. Le développement et la commercialisation par Everspin de cette technologie se base sur la théorie de L. Savtchenko *et al.* sur la commutation d'une mémoire via la génération d'un champ magnétique induit par des lignes de courant [27]. Chacune des deux énergies fournies par les deux champs magnétiques doit être

suffisante pour faire commuter toutes les cellules visées, faute de quoi seules quelques cellules pourront commuter, comme illustré Figure 1.14.b.



(a) Représentation de l'architecture d'une mémoire FIMS-Toggle.

(b) Commutation en fonction des champs $\overrightarrow{H_1}$ et $\overrightarrow{H_2}$.

FIGURE 1.14 – La mémoire FIMS-Toggle.

Le principe d'écriture de la mémoire FIMS-Toggle consiste en l'inversion de l'orientation magnétique de la couche de stockage. Cette commutation est réalisée par un phénomène de nucléation et propagation correspondant à une rotation successive des moments magnétiques d'un angle de 45° [28].

Cette architecture est composée principalement d'un empilement antiferromagnétique (AF) - ferromagnétique - isolant - ferromagnétique. La couche AF a pour rôle de fixer l'aimantation de la couche de référence.

Cette méthode d'écriture souffre principalement de la précision de la sélectivité du point mémoire avec la miniaturisation, ainsi que de sa consommation relativement élevée.

2.2.3.2 Thermally Assisted Switching MRAM (TAS-MRAM) :

La commutation assistée par température ou *Thermally Assisted Switching MRAM* – (TAS-MRAM) combine un réchauffement localisé de la jonction mémoire sélectionnée et un champ magnétique unique appliqué pour modifier l'aimantation de la couche de stockage. L'une des différences majeures de cette technologie par rapport à la technologie FIMS-Toggle est la présence d'un second antiferromagnétique AF_2 au-dessus de la couche de stockage, comme illustré Figure 1.15. Ainsi, la JTM est composée d'un empilement : antiferromagnétique - ferromagnétique - isolant - ferromagnétique - antiferromagnétique. Ce second antiferromagnétique va conduire à fixer l'aimantation de la couche libre tant qu'aucun échauffement n'est induit dans la structure.

Lorsque la température résultante dans l'antiferromagnétique AF_2 est supérieure à sa température de Néel, celui-ci deviendra paramagnétique et n'appliquera plus de force (ou énergie d'échange) sur la couche ferromagnétique de stockage. L'écriture de la jonction est alors réalisée via l'activation de la ligne de champ (*Field line*, illustrée Figure 1.15), pendant que le matériau AF_2 est paramagnétique.

Cette méthode d'écriture résout en partie les difficultés de sélectivité observés sur la technologie FIMS-Toggle mais souffre également de difficultés, certes moindres, de miniaturisation et de consommation.

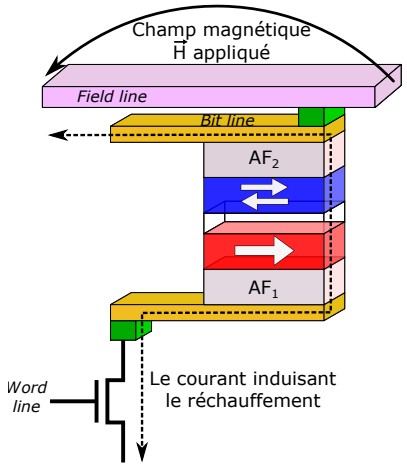


FIGURE 1.15 – Schéma de la mémoire TAS-MRAM.

2.2.3.3 Spin-Transfer Torque MRAM (STT-MRAM) :

La difficulté majeure des technologies MRAMs dont la commutation est réalisée par champ (FIMS-Toggle et TAS-MRAM) est la génération d'un champ localisé de manière extrêmement précise sur un nœud mémoire. En effet, le rayonnement du champ peut facilement conduire à la commutation des éléments voisins proches de la cellule visée. Dans le cadre du développement de cette catégorie de mémoires non-volatiles, de nouvelles générations utilisant des mécanismes de commutation différents sont apparues, principalement la technologie STT-MRAM.

Cette technologie ne nécessite qu'un **courant de spin polarisé** pour commuter d'un état de faible résistivité vers un état de forte résistivité et inversement. Un courant initialement non-polarisé devient polarisé en spins (soit en *spins-up* majoritaires soit en *spins-down* majoritaires) grâce à l'effet des moments magnétiques de la première couche ferromagnétique de la JTM qu'il traverse. L'analogie entre une JTM et un filtre peut être réalisée. Pour exemple, les électrons de *spins-down*, traversant un matériau ferromagnétique orienté majoritairement en *spins-up*, sont filtrés. Seuls les *spins-up* réussissent à traverser cette première couche ferromagnétique de la JTM.

L'injection d'un courant dans la JTM soit par la couche de stockage soit par la couche de référence conduit à l'écriture de l'un des deux états de la mémoire (P ou AP). Ce mécanisme est désigné par *Transfer Torque* [29], [30].

Considérons donc les deux phénomènes qui induisent la programmation des cellules STT-MRAMs :

— La commutation d'un état AP (forte résistivité) vers un état P (faible résistivité) :

  Considérons une JTM initialement dans un état AP (avec par exemple les *spins-up* majoritaires dans la référence et les *spins-down* majoritaires dans la couche de stockage, comme illustré étape (1) de la Figure 1.16) :

  Le courant (non-polarisé) injecté par la couche de référence conduit à sa polarisation en *spins-up*. Les électrons de *spins-up* peuvent alors traverser la barrière tunnel, étapes (2) et (3). Les électrons de *spins-down* sont eux réfléchis par la couche de référence, étape (2) de la Figure 1.16. Toutefois, la couche de stockage étant orientée en *spins-down*, celle-ci réfléchit quelques électrons de *spins-up*, étape (4). Lorsque la densité de courant qui traverse la STT-MRAM est suffisante, les électrons de *spins-up* transmis à la couche de stockage deviennent majoritaires aux électrons de *spins-down* présents, étape (5). La couche inverse donc son aimantation. L'état de la STT-MRAM bascule de AP vers P.

  Dans ce cas, la commutation est induite par la transmission des électrons majoritaires polarisés par la couche de référence.

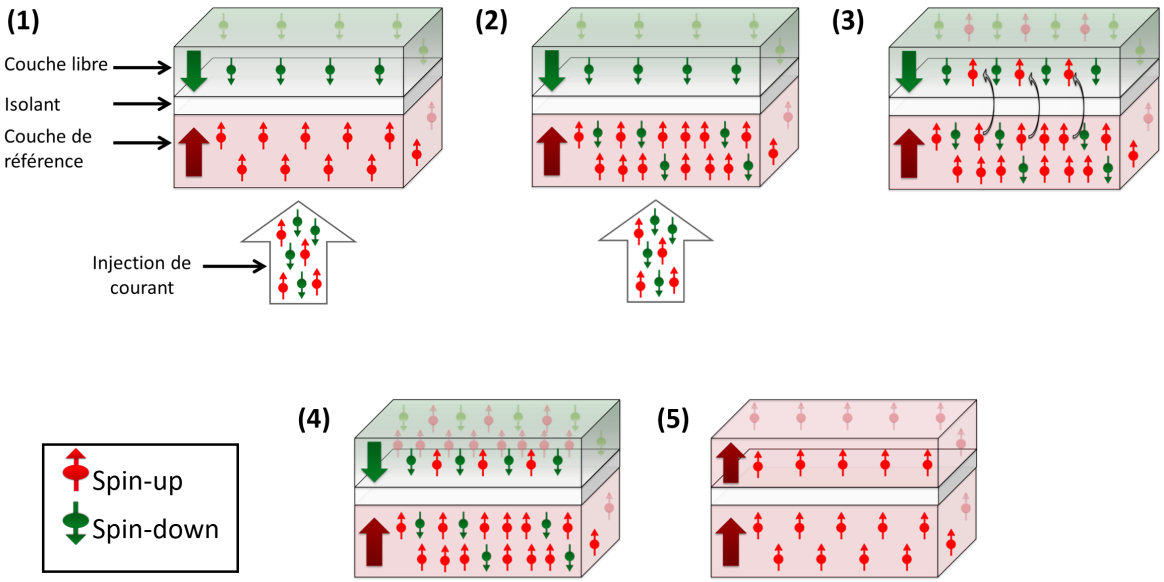


FIGURE 1.16 – Commutation d'une cellule STT-MRAM d'un état AP vers un état P.

— La commutation d'un état Parallèle P vers un état Anti-Parallèle AP est illustrée sur la Figure 1.17 :

Étape (1) : considérons la JTM initialement dans un état P, comme décrite Figure 1.17. Dans ce cas, les *spins-up* sont majoritaires dans les deux couches ferromagnétiques.

Étape (2) : le courant injecté par la couche de stockage est polarisé en *spins-up* par cette couche et traverse la JTM. Les *spins-down* sont eux réfléchis dans la couche de stockage.

Étape (3) : lorsqu'une densité de courant suffisante est atteinte, la population de *spins-up* et de *spins-down* (réfléchis) dans la couche de stockage s'inverse. Les *spins-down* deviennent majoritaires.

Étape (4) : La résistivité de la STT-MRAM bascule de l'état P vers l'état AP.

Dans ce cas, la couche de stockage joue le rôle de polariseur du courant et la commutation est induite par la réflexion des électrons minoritaires.

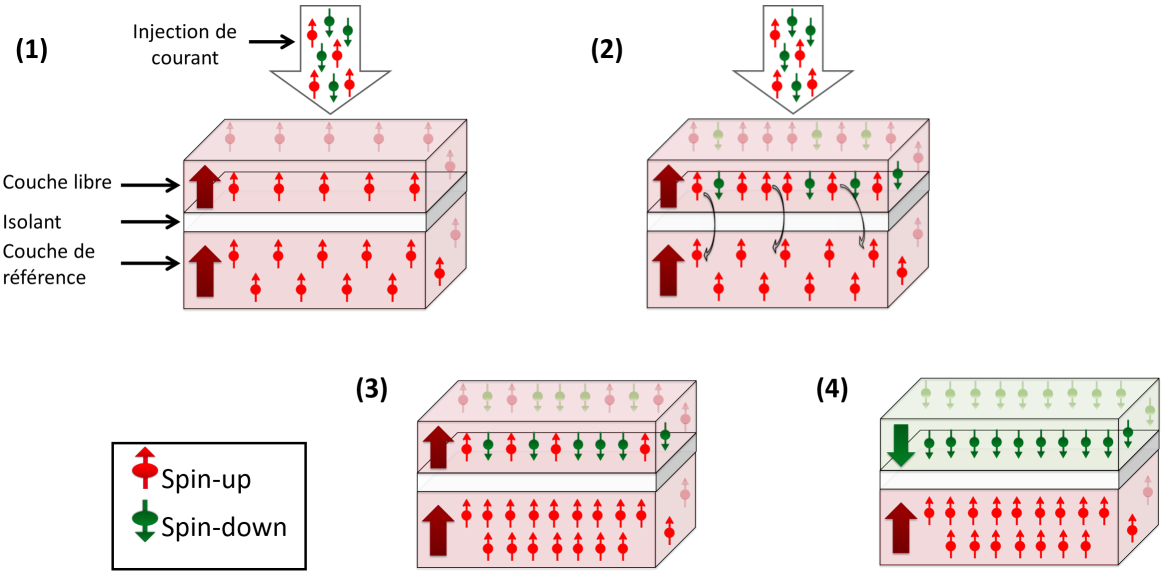
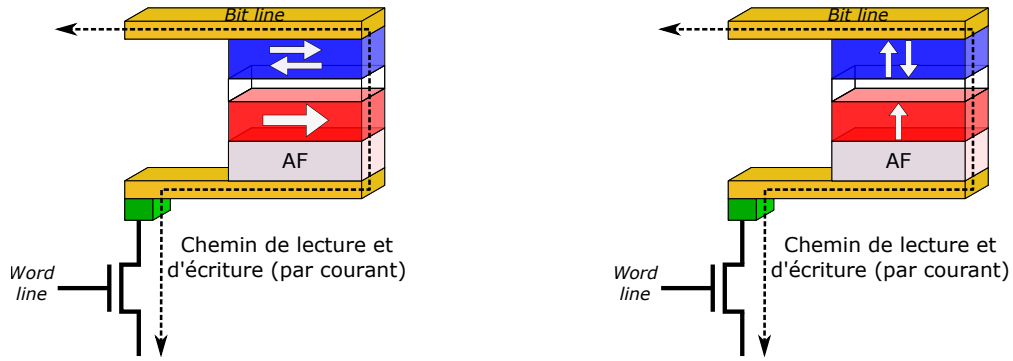


FIGURE 1.17 – Commutation d'une cellule STT-MRAM d'un état P vers un état AP.

— Lecture d'une cellule STT-MRAM :
Comme pour les technologies FIMS-Toggle et TAS-MRAM, cette structure est une architecture à deux terminaux, la lecture et l'écriture suivent le même chemin.
La lecture de l'état mémoire dans une STT-MRAM est réalisée par le passage d'un courant dans l'empilement de la JTM. Toutefois, la densité de courant doit être inférieure à la densité nécessaire pour réaliser une commutation de l'état mémoire, au risque d'écrire pendant une phase de lecture.

La STT-MRAM a été initialement développée comme une architecture à aimantation planaire. Les spins des couches ferromagnétiques étaient parallèles au plan, comme représenté Figure 1.18.a. Depuis, les architectures ont évolué en structures à aimantations perpendiculaires (p-STT-MRAM) dont les spins sont orthogonaux au plan de la JTM, comme illustré Figure 1.18.b.



(a) Schéma d'une STT-MRAM dont l'aimantation des ferromagnétiques est planaire au plan.

(b) Schéma d'une STT-MRAM dont l'aimantation des ferromagnétiques est perpendiculaire au plan.

FIGURE 1.18 – Illustration des deux architectures mémoires STT-MRAMs existantes.

Les p-STT-MRAMs sont plus efficaces en terme de consommation grâce à une plus forte anisotropie perpendiculaire et à la diminution des niveaux d'énergies nécessaires pour réaliser la commutation. L'énergie nécessaire pour commuter l'état de la cellule est exprimée par la stabilité thermique Δ_T (sans dimension). La densité de courant d'une architecture perpendiculaire $I_c^{perp.}$ est exprimée par l'Équation 1.4 [31].

$$I_c^{perp.} = (\frac{4e}{\hbar})\frac{\alpha k_B T}{\eta}\Delta_T \tag{1.4}$$

Cette expression est définie en fonction de : e la charge élémentaire de l'électron, $\hbar$ la constante de Planck réduite, α le coefficient d'atténuation de Gilbert, k_B la constante de Boltzmann, T la température et η la polarisation en spin du courant.

Cette technologie mémoire cumule différents avantages qui en font une structure de choix dans le panel des mémoires non-volatiles émergentes. En effet, sa compatibilité avec le CMOS, sa vitesse d'écriture et de lecture (de l'orde de la ns), sa densité et sa consommation réduite en font une mémoire de choix pour une utilisation dans diverses applications embarquées. C'est pourquoi, la STT-MRAM est aujourd'hui selon l'ITRS la mémoire susceptible de devenir la mémoire universelle [10], [32]. De plus, le rapport de *Yole development* de 2018 prévoit que cette technologie mémoire deviendra en 2023, la première technologie mémoire émergente embarquée [9].

2.2.3.4 Spin-Orbit Torque MRAM (SOT-MRAM) :

Contrairement à la mémoire STT-MRAM, la SOT-MRAM est une structure à trois terminaux [33] qui a été découverte par le laboratoire Spintec [34], [35] et [36]. Le principal avantage de cette génération de mémoires magnétiques est la séparation des chemins de lecture et d'écriture, qui se font par deux chemins distincts comme représentés sur la Figure 1.19. Cette caractéristique

limite la détérioration de l'isolant puisque seuls de faibles courants de lecture traversent cette barrière tunnel, augmentant donc leur endurance. De plus, compte tenu que le courant d'écriture ne traverse pas l'isolant, il est possible d'utiliser de forts courants pour atteindre des temps de commutation ultra-rapides jusqu'à 210 ps [37], [38].

Le fonctionnement de cette technologie mémoire a été démontré par I. M. Miron *et al.* [39], [40] puis par L. Liu *et al.* [41]. Ces travaux ont démontré la capacité de cet empilement à commuter grâce à une conduction de courant sur le niveau de métal inférieur pour inverser l'état stocké dans la couche libre, comme représenté sur la Figure 1.19. Les phénomènes mis en avant expliquant la commutation de cette couche de stockage sont l'effet *Rashba* [39] et l'effet de *Spin Hall* [42].

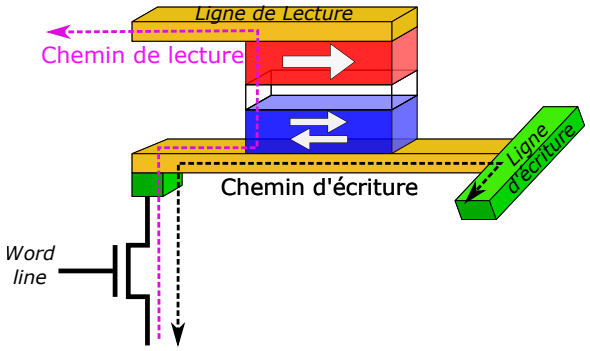


FIGURE 1.19 – Schéma de la technologie mémoire SOT-MRAM.

Le Tableau 1.1 conclut cette première section, en répertoriant les différentes propriétés des technologies mémoires MRAMs. Comme il peut être noté, les deux dernières générations STT-MRAM et SOT-MRAM présentent de nombreux avantages en terme de surface, consommation et vitesse de fonctionnement. Ces caractéristiques en font de bons candidats pour une intégration dans l'Internet des Objets. Dans le cadre de cette thèse, la technologie mémoire STT-MRAM est privilégiée pour sa faible surface et faible consommation.

	Commutation	Surface	Consommation	Vitesse	Terminaux
FIMS-toggle	Champ	Forte	Forte	Faible	2
TAS	Champ	Moyenne	Moyenne	Faible	2
STT	Courant	Faible	Faible	Moyenne	2
SOT	Courant	Moyenne	Faible	Forte	3

TABLEAU 1.1 – Caractéristiques des différentes technologies MRAMs présentées.

3 Sécurité des circuits intégrés

La cryptologie ou science du secret est le domaine qui permet d'assurer la sécurité des composants en rendant indéchiffrables, hormis pour l'expéditeur et le destinataire, les données échangées. Celle-ci englobe la **cryptographie** qui vise à chiffrer des messages personnels entre deux personnes grâce à des clés spécifiques. Sans ces clés, une personne extérieure est incapable de déchiffrer le message [43]. Afin de considérer la cryptographie comme opérationnelle et fonctionnelle, il est nécessaire de s'assurer que ces quatre points sont vérifiés :

— La confidentialité : les données sont accessibles uniquement aux sujets autorisés.
— L'intégrité : les données sont modifiables uniquement par les sujets autorisés.
— L'accessibilité : l'accès à un objet ne devrait pas être refusé à une personne autorisée.
— L'authenticité : Les données auxquelles l'utilisateur à accès sont vérifiées.

3.1 La cryptographie et ses enjeux

3.1.1 L'histoire de la cryptographie

La cryptographie est utilisée depuis l'antiquité pour chiffrer les messages. Le chiffrement de *César* est aujourd'hui le plus connu de cette période de l'histoire. Il consiste à chiffrer un message par un décalage des lettres de l'alphabet par un certain nombre fixe [44]. Ce type d'algorithme est simple à déchiffrer en réalisant une analyse fréquentielle selon l'utilisation des lettres dans la langue française. En effet, dans un message *M*, les lettres qui apparaissent le plus souvent peuvent être considérées comme les lettres qui apparaissent le plus dans la langue française. Par la suite, un choix est réalisé parmi les différentes possibilités.

La cryptographie fait également partie des moments les plus importants qui ont marqué l'histoire. Nous pouvons ainsi citer l'exemple de la machine *Enigma* utilisée par les allemands lors de la seconde guerre mondiale pour chiffrer leurs messages contenant leurs tactiques militaires [44], illustrée sur la Figure 1.20.

FIGURE 1.20 – Photographie d'Enigma [45].

Alan Turing a alors inventé ce qui peut être considéré comme un précepteur du premier ordinateur, qui permettait de déchiffrer ces messages et de garder une longueur d'avance sur les attaques planifiées par l'ennemi. Cette invention a joué un rôle majeur dans la victoire de la seconde guerre mondiale par les alliés.

Par la suite, la nécessité de protection des communications et des données par cryptographie en développpant des algorithmes de chiffrement fut décrétée par le *National Institute of Standards and Technology* – (NIST) anciennement *National Bureau of Standards* – (NBS) en 1973 [46]. Ces derniers doivent respecter les préceptes d'A. Kerckhoffs énoncés en 1883 [47] indiquant que :

— Le système de chiffrement doit être mathématiquement sécurisé.

— La sécurité du chiffrement ne doit dépendre que de la clé de chiffrement et non du secret de la méthode utilisée.

Il existe aujourd'hui un nombre important d'algorithmes de cryptographie qui peuvent être organisés en différentes catégories comme représenté sur la Figure 1.21.

3.1.2 La cryptographie asymétrique

L'algorithme de Cryptographie à Clé Publique ou – *Public Key Cryptography* – (PKC) est apparu en 1976 grâce à W. Diffie et M. Hellman qui ont ouvert une nouvelle voie dans la cryptographie [48].

Les PKCs effectuent des chiffrements asymétriques. Pour cela, pour deux entités communicantes, chacune nécessite deux clés : une clé privée et une clé publique. Alors que la clé publique est transmissible sans aucune restriction, la clé privée n'est connue que de son détenteur. Le chiffrement d'un message clair en utilisant ce protocole est facilement réalisable. Toutefois, le déchiffrement ne peut pas être réalisé en inversant simplement la fonction de chiffrement. Cette fonction n'est pas bijective.

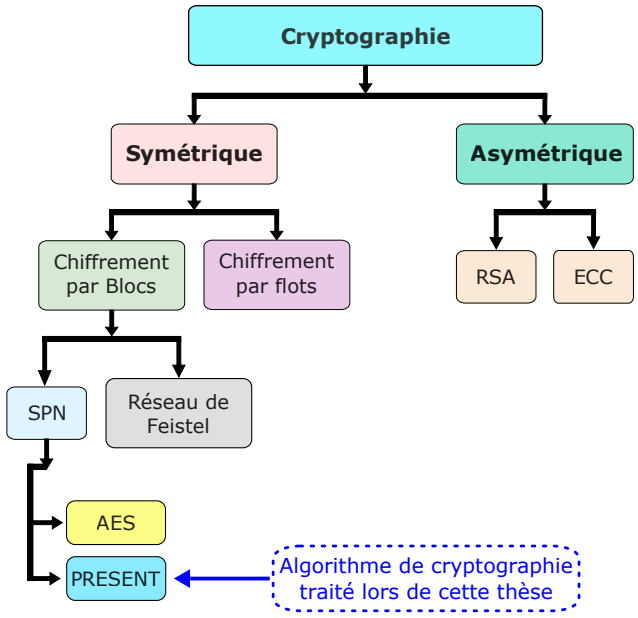


FIGURE 1.21 – Catégories des différents algorithmes de cryptographie.

Considérons le cas où deux identités *Alice* et *Bob* souhaitent communiquer des informations sensibles sur un canal non-sécurisé. Une personne malveillante, *Oscar*, pourrait alors observer leur conversation sur ce canal. C'est pourquoi *Alice* et *Bob* chiffrent leurs messages, comme illustré sur la Figure 1.22.

Pour réaliser le chiffrement de leurs messages, chaque partie a une paire de clés :

— Une clé publique connue de tous, C_A et C_B respectivement pour l'entité *Alice* et *Bob*.

— Une clé privée connue uniquement par l'identité, tel que CP_A la clé privée d'*Alice* et CP_B celle de *Bob*.

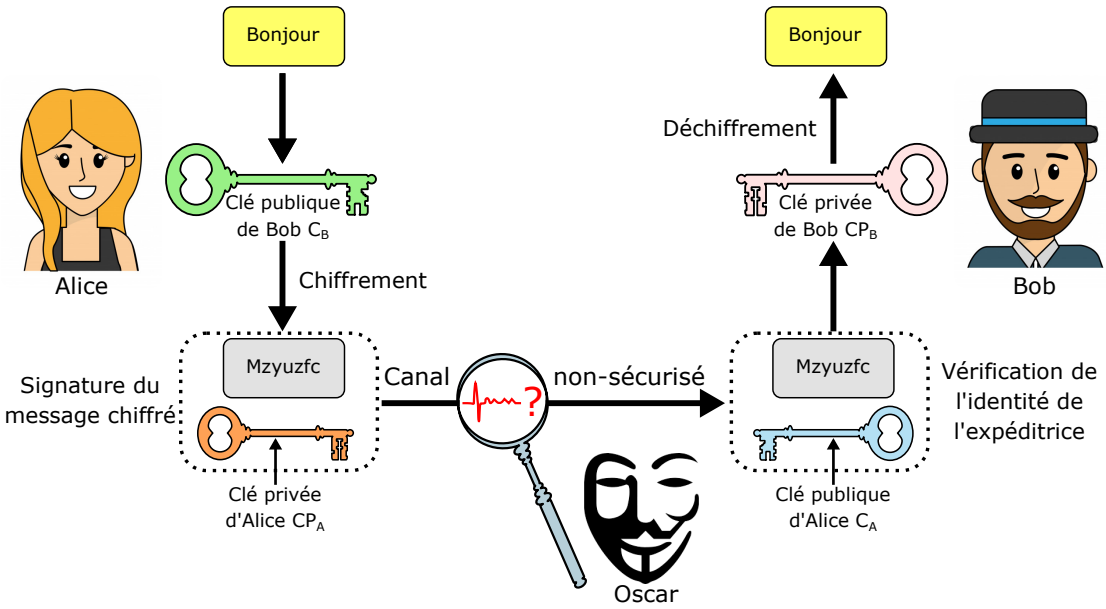


FIGURE 1.22 – Modèle de fonctionnement d'un algorithme de cryptographie asymétrique.

Le chiffrement et déchiffrement du message secret nécessitent l'utilisation de deux de ces quatre clés. En effet, lorsque *Alice* souhaite envoyer un message secret à *Bob*, le message est chiffré par C_B alors que le déchiffrement est réalisé par *Bob* grâce à CP_B, comme illustré sur la Figure 1.22.

De plus, la clé CP_A est utilisée par *Alice* pour signer son message alors que C_A permet à *Bob* de confirmer l'identité de l'expéditrice.

L'un des principaux algorithmes de la cryptographie asymétrique est le chiffrement de *Rivest Shamir Adleman* – (RSA), du nom de ses inventeurs, proposé en 1978 [49]. Le RSA est un protocole se basant sur l'exponentiation modulaire où le chiffrement est réalisé par la fonction : $C = M^e \ mod[N]$ tel que C est le texte chiffré, M le message clair, e et et N deux entiers. La fonction de déchiffrement d'autre part est réalisée par l'opération $M = C^d mod[N]$, tel que d un entier.

Cette catégorie d'algorithmes de cryptographie est généralement plus lente que les algorithmes de cryptographie symétriques [46], [6].

3.1.3 La cryptographie symétrique

Les algorithmes de cryptographie symétriques ou à clé secrète sont des schémas d'offuscation d'un message clair, en utilisant une même clé - privée et secrète - pour le chiffrement et le déchiffrement, comme illustré Figure 1.23. Cette clé est communiquée entre les deux identités *Alice* et *Bob* par un canal sécurisé au préalable. Cette application de chiffrement et de déchiffrement est dite bijective.

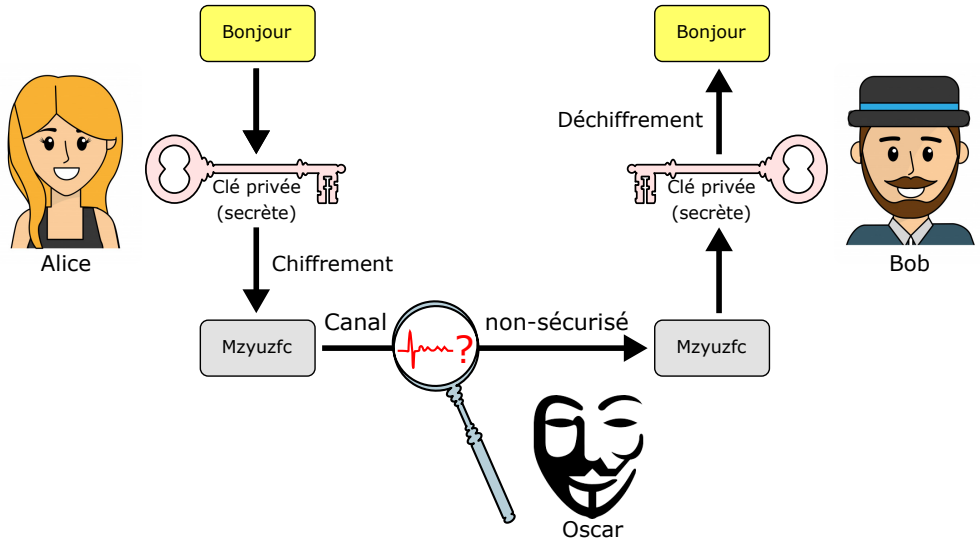


FIGURE 1.23 – Modèle de fonctionnement d'un algorithme de cryptographie symétrique.

Les algorithmes à clé secrète symétriques peuvent être décomposés en deux principales catégories de techniques de chiffrement : les chiffrements par flots et les chiffrements par blocs.

3.1.3.1 Les chiffrements par flots :

Les algorithmes de cryptographie par flots consistent à chiffrer des messages de toutes tailles [50], sans besoin de les découper au préalable [46]. Cette propriété permet de réaliser un chiffrement continu des données, sans nécessité d'attendre la transmission de tout le message. D'où, une utilisation accrue dans le domaine des télécommunications, où le transfert des données doit se faire de manière instantanée. Pour chaque message clair m de taille n $(m_{n-1} \cdots m_0)$ est associé une clé de taille identique n $(k_{n-1} \cdots k_0)$, afin de pouvoir réaliser le chiffrement bit à bit.

Les chiffrements par flots ont la particularité d'associer à chaque message un masque aléatoire, comme illustré sur la Figure 1.24. Il est nécessaire de modifier ce masque aléatoire après chaque chiffrement. Cette opération est appelée l'affectation d'un masque jetable. Ce masque est généralement généré par un générateur pseudo aléatoire. Ainsi, lors d'une première étape d'ini-

tialisation, la clé est combinée à un Vecteur d'Initialisation (*VI*) généré aléatoirement par un générateur pseudo aléatoire, comme illustré sur la Figure 1.24.

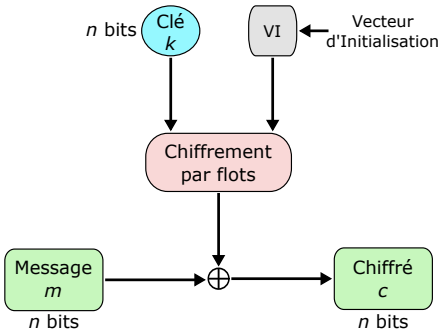


FIGURE 1.24 – Algorithme de chiffrement par flots : représentation des opérations réalisées sur un message clair m et une clé k pour obtenir le texte chiffré c, grâce à un Vecteur d'Initialisation *VI*.

Le résultat de ce masquage de la clé k permet le chiffrement du message m, bit à bit en utilisant majoritairement une fonction XOR, tel que développé en Équation 1.5.

$$\forall i \in [0,n], \quad c_i = m_i \oplus k_i \tag{1.5}$$

Il est possible de citer les algorithmes ChaCha [51], Mickey [52] ou encore Trivium [53] dans cette catégorie. Ces algorithmes sont recensés dans [54].

3.1.3.2 Les chiffrements par blocs :

La seconde catégorie des chiffrements symétriques est la famille des chiffrements par blocs. Contrairement aux chiffrements par flots, la taille des messages m à chiffrer dépend directement de l'algorithme de cryptographie utilisé. Pour chaque algorithme, la clé et le message clair ont des tailles fixes.

En outre, les chiffrements par blocs sont réalisés en plusieurs tours d'opérations contrairement aux chiffrements par flots. Ces algorithmes se décomposent en deux sous-blocs :

— Réseau de Feistel :
Les réseaux de Feistel tiennent leur nom d'H. Feistel, l'inventeur de l'algorithme Lucifer qui fut le premier de ce type de chiffrements à être développé [55].

Afin de réaliser le chiffrement d'un message clair m via les réseaux de Feistel, il est nécessaire de décomposer dans un premier temps ce mot, en deux mots de tailles équivalentes. Soit le mot m de taille $2r$ décomposé en deux mots D et G, de r bits chacun. Considérons par exemple le premier tour de chiffrement, le mot G_1 va être combiné à la clé alors que le mot D_1 est transmis directement au tour suivant. Lors du second tour, le chiffrement est réalisé sur D. Ainsi, lors de chaque tour une seule moitié du message est chiffrée (alors que la seconde moitié est transmise au tour suivant). Chaque tour utilise une clé intermédiaire calculée par l'algorithme via la fonction F.

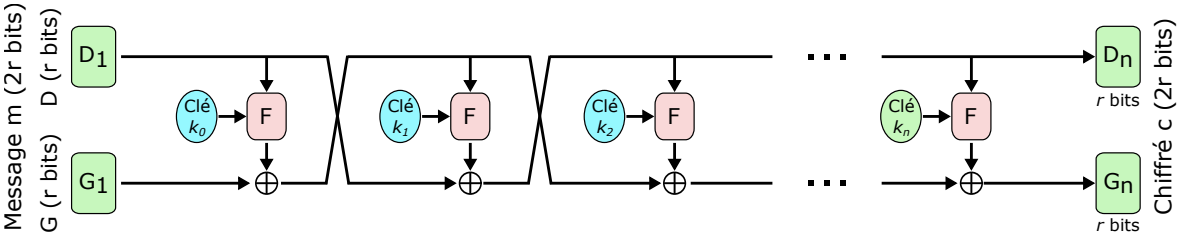


FIGURE 1.25 – Algorithme de type réseau de Feistel : représentation des différentes opérations réalisées sur un message clair m et une clé k_0 pour obtenir le texte chiffré c, après $(n+1)$ tours.

Dans cette catégorie de chiffrements, il est possible de citer : *Data Encryption Standard* – (DES) [56], triple-DES [57], Blowfish [58], SIMON et SPECK [59].

— Réseau de Substitution et de Permutation ou – *Substitution Permutation Network* – (SPN) : Ces algorithmes réalisent des opérations de substitution et de permutation à chaque tour du chiffrement. Cette technique de chiffrement se caractérise par trois opérations qui sont appliquées au message clair m afin d'obtenir le message chiffré c, comme illustré Figure 1.26.

- La première étape est l'exécution d'une fonction XOR entre la clé de chiffrement maître k_0 et le message clair m.
- Le résultat de cette opération est envoyé vers une fonction de substitution. Chaque partie du message est remplacée par un mot de taille identique. Cette fonction est induite par une table de substitution pré-définie.
- La dernière étape de ce premier tour de chiffrement est la permutation. Dans un mot m, la position p_i du bit b_i est modifiée (également en suivant une table de permutation pré-définie).

La succession de ces opérations est répétée le nombre de tours nécessaires pour le chiffrement du message par l'algorithme.

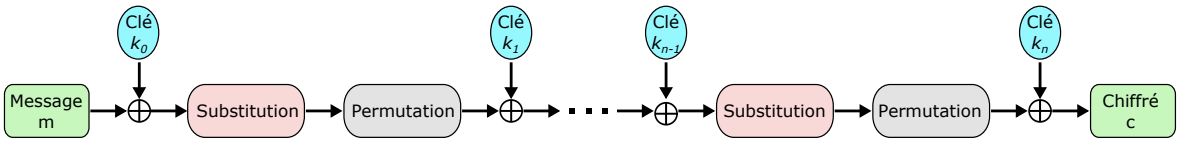


FIGURE 1.26 – Algorithme de type SPN : représentation des différentes opérations réalisées sur un message clair m et une clé k_0 pour obtenir le texte chiffré c, après $(n+1)$ tours.

Le chiffrement considéré aujourd'hui comme référence est le chiffrement standard avancé ou – *Advanced Encryption Standard* – (AES) (chiffrement de Rijndael) [60], [61]. Il est comme son nom l'indique la technique de chiffrement standard utilisée depuis 2007. Cet algorithme après avoir été publié en 1998, a démontré son intérêt en matière de performances (vitesse, consommation et sécurité) face à ses concurrents.

La cryptographie symétrique et la cryptographie asymétrique sont deux techniques complémentaires de chiffrement. Alors que la première présente des avantages majeurs en termes de performance et de vitesse, permettant ainsi le chiffrement à faible coût, d'un message par deux entités ayant un accès physique à la clé secrète, la seconde est nécessaire pour par exemple la communication, au préalable, de cette clé entre ces entités de manière sécurisée. De plus, la cryptographie asymétriques est utilisée lors de la confection de protocoles d'authentification et de signature des messages.

Les algorithmes de cryptographie (symétriques et asymétriques) se sont multipliés lors cette dernière décennie par la proposition de chiffrements dont les performances conviendrait à l'Internet des Objets (surface et consommation réduites et vitesse d'exécution importante). Ce fut l'avènement de la cryptographie légère ou – *Light Weight Cryptography* – (LWC). Cette famille d'algorithmes est extrêmement étudiée et présente un nombre important d'algorithmes développés, comme pour exemple : Camellia [62], Piccolo [63], SEA [64], KLEIN [65], PRINCE [66], PRIDE [67], ou encore PRESENT [68] qui est un standard de la cryptographie légère.

Dans le cadre des études réalisées lors de cette thèse, l'algorithme de cryptographie étudié est le chiffrement symétrique de type SPN : PRESENT [68], précédemment illustré sur la Figure 1.21.

3.2 Les différentes techniques d'attaques matérielles

Les attaques visant à exploiter des résultats erronés dans un calcul, de corrompre le fonctionnement du circuit et donc de récupérer des informations secrètes sont en constante évolution. Ces attaques peuvent être de différents types, logicielles ou matérielles. Les attaques qui sont présentées dans ce chapitre sont les attaques matérielles. Ces dernières peuvent être classées en deux catégories distinctes : les attaques par observation et les attaques par perturbation.

3.2.1 Les attaques par observation

Les attaques par observation correspondent aux attaques non-invasives qui peuvent être réalisées sur des circuits, sans altérer leur fonctionnement interne ou modifier leurs caractéristiques physiques. Cette catégorie d'attaques est constituée essentiellement des attaques par canaux cachés (ou canaux auxiliaires), des attaques qui exploitent des failles matérielles pour retrouver les données secrètes. En effet, le fonctionnement d'un circuit peut conduire à des fuites exploitables par l'attaquant. Ces fuites dépendent des données manipulées par le circuit, conduisant à des émissions électromagnétiques [69], une consommation en courant [70], des temps de calculs [71] ou encore des émissions acoustiques [72] qui dépendent des données manipulées, et donc de la clé de chiffrement.

Il existe différentes techniques d'analyse du courant de consommation pour recouvrer le message confidentiel traité par un algorithme de cryptographie. Les analyses simples, différentielles et de corrélation de la consommation sont les plus répondues.

3.2.1.1 Simple Power Analysis (SPA) :

L'Analyse Simple de Puissance ou – *Simple Power Analysis* – (SPA) est la plus simple technique d'attaques par canaux cachés. Afin de retrouver le message secret utilisé par un algorithme de cryptographie, l'attaquant doit uniquement observer la courbe de consommation du dispositif (ou toute autre grandeur en cours d'acquisition) afin de déterminer l'opération en cours de traitement.

La Figure 1.27 illustre la courbe de consommation d'un algorithme d'exponentiation modulaire (décrit en Algorithme 1). Cette opération fait partie de la fonction de déchiffrement de l'algorithme de cryptographie asymétrique RSA. Comme il peut être observé, les deux opérations de *MULTIPLY* et *SQUARE* nécessitent des niveaux distincts de courant pour pouvoir fonctionner. Il est alors possible de déterminer exactement quelle opération est en cours de calcul. Ainsi, comme précisé par l'algorithme 1, l'opération de *MULTIPLY* n'est réalisée que lorsque la donnée à traiter est égale à '1' (opération dans la boucle *Si*). Ainsi, grâce aux opérations réalisées, l'attaquant peut déterminer la donnée secrète traitée, comme précisé par la Figure 1.27. Cette méthode de cryptanalyse peut permettre de retrouver entièrement la clé secrète, uniquement par l'observation de la consommation du circuit.

Algorithme 1 Exponen. modulaire

Entrées : C, $d = (d_k, \cdots, d_0)_2, n$
Sortie : $M = C^d$ modulo n

```
M ← 1
tant que i = k : 0 faire
    M ← MOD(square(M), n)
    si d_i alors
        M ← MOD(multiply(M,C), n)
    fin si
fin tant que
Retourner M
```

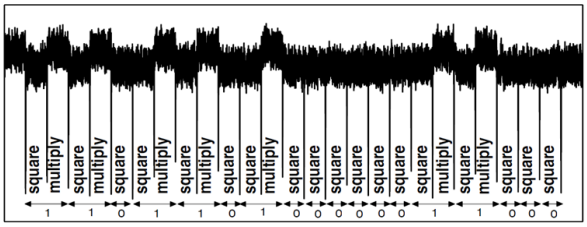


FIGURE 1.27 – Analyse Simple de Consommation d'une exponentiation modulaire [73].

Cette technique a également permis de réaliser différentes analyses sur l'algorithme de cryptographie avancé AES [74], [75], [76].

3.2.1.2 Differential Power Analysis (DPA) :

L'Analyse Différentielle de Puissance ou – *Differential Power Analysis* – (DPA) est une alternative à la SPA. Les pionniers de cette attaque par canaux cachés ont été P. Kocher *et al.* en 1999 [77].

Cette technique consiste à retrouver des sous-clés de la clé indépendamment les unes des autres. Cette sous-clé dépend directement d'une partie de l'entrée qui peut être soit connue soit calculable [46]. Ainsi, une attaque différentielle est menée pour des entrées différentes. Le même secret est utilisé à chaque exécution du programme. En réalisant plusieurs hypothèses sur l'état de chaque sous-clé et en analysant les résultats obtenus en sortie du chiffrement, il devient alors possible de réaliser une analyse statistique des résultats pour trouver la meilleure correspondance entre une sous-clé hypothétique et la sous-clé utilisée pour le calcul.

Ainsi, afin de réaliser cette attaque par canaux cachés, l'attaquant doit être en mesure de simuler plusieurs fois l'exécution de l'algorithme de cryptographie et connaître, si possible, la structure interne du dispositif attaqué.

L'analyse statistique peut également être instanciée grâce à la corrélation de la consommation du circuit, aux données manipulées [78]. L'Analyse de Corrélation de Puissance ou – *Correlation Power Analysis* – (CPA) consiste ainsi à étudier la corrélation entre des hypothèses de clé et les fuites effectivement mesurées par l'attaquant. Cette méthode utilise la corrélation de Pearson [79].

La Figure 1.28 illustre l'instanciation d'une analyse statistique basée sur la corrélation entre des massages et un octet de la clé. Comme il peut être noté sur le résultat en Figure 1.28.b, à partir de 250 mesures de la consommation du circuit, il est possible de réduire les possibilités de clés à 2. L'acquisition de 4000 courbes de mesure dans l'analyse statistique conduit à réduire le nombre de bonnes hypothèses de clés à une seule (la bonne clé utilisée pour le chiffrement).

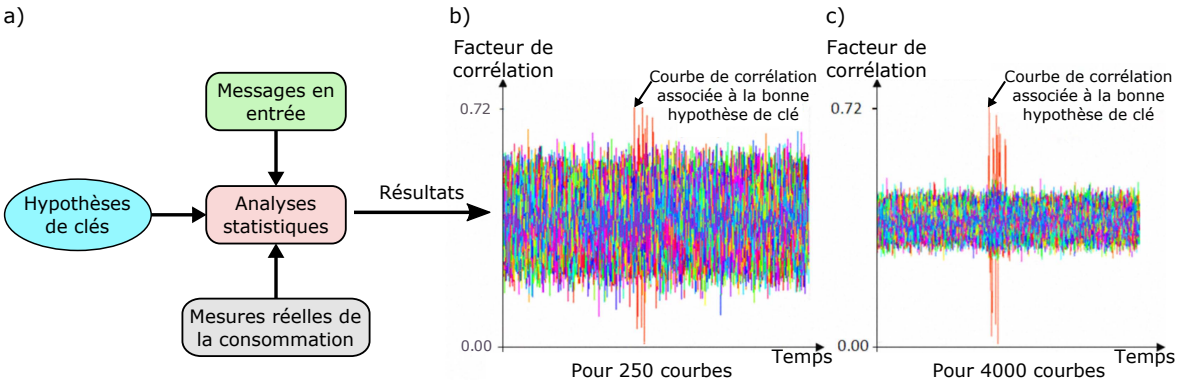


FIGURE 1.28 – Illustration des attaques par canaux cachés basées sur des analyses statistiques de corrélation. a) Illustration des paramètres d'entrée de l'analyse statistique. b) Courbes de corrélation entre les messages et les hypothèses de clés pour 250 courbes de mesures [80]. c) Courbes de corrélation entre les messages et les hypothèses de clés pour 4000 courbes de mesures [80].

3.2.2 Les attaques par perturbation

Les attaques par perturbation visent la modification et la perturbation du fonctionnement du circuit cible et des données ou des programmes qui y sont traités [81]. Ces techniques ciblent des zones sensibles connues de l'attaquant afin de modifier les réactions du circuit et de déterminer le secret en cours de traitement.

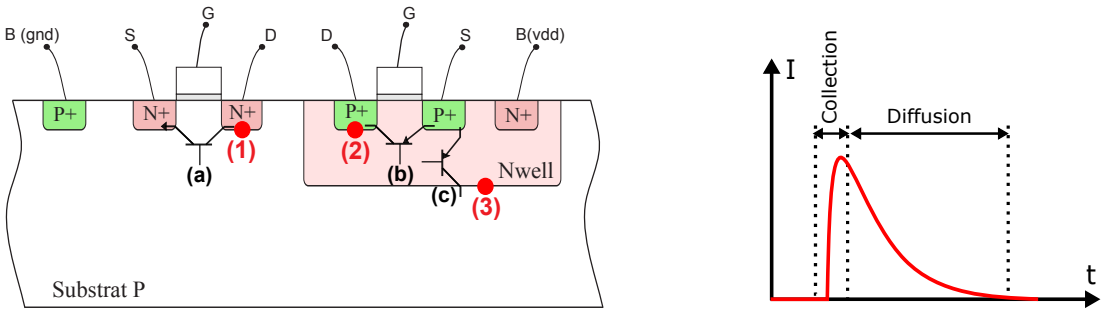
Les principales techniques d'attaques par perturbation appliquées aux circuits intégrés sont développées ci-dessous :

3.2.2.1 Injection par Amplification de la Lumière par Émission Stimulée de Radiation ou – *Light Amplification by Stimulated Emission of Radiation* (LASER) :

Cette méthode consiste à illuminer le circuit par un faisceau LASER afin d'y induire des effets photoélectriques [82] ou thermiques [83], via respectivement la génération d'un photo-courant dans la structure ou son échauffement.

Le photo-courant est généré dans l'une des zones sensibles aux illuminations des circuits CMOS, dans les jonctions PN polarisées en inverse notées (1), (2) et (3) sur la Figure 1.29. Ainsi, le faisceau va générer des paires électrons-trous sur son chemin. Ces paires sont collectées par les zones de charge d'espace de ces jonctions PN. C'est ainsi qu'un photocourant est généré, dont le profil est illustré Figure 1.29. Les illuminations LASERs peuvent également induire des effets thermiques dans les structures [83] [84], en modifiant la valeur en sortie d'une ou plusieurs portes logiques ou de points mémoires.

Cette technique est très utilisée dans le contexte des injections de fautes en raison de son excellente résolution spatiale et temporelle. En effet, contrairement aux techniques qui seront citées par la suite, l'injection LASER (ou d'un système équivalent tel que le faisceau d'ions focalisés ou – *Focused Ion Beam* – (FIB)) est la seule permettant de viser de façon très précise un transistor pour un nœud technologique mature ou une cellule pour des nœuds technologiques plus avancés. Toutefois, pour réaliser ces perturbations, l'équipement nécessaire est coûteux bien que de nouvelles études sont réalisées pour utiliser des équipements plus accessibles en préservant les avantages de cette technique d'attaque [85].



(a) Illustration des zones sensibles dans le CMOS.

(b) Profil d'une illumination.

FIGURE 1.29 – Injection Laser dans un circuit intégré.

3.2.2.2 Injection d'impulsions électromagnétiques :

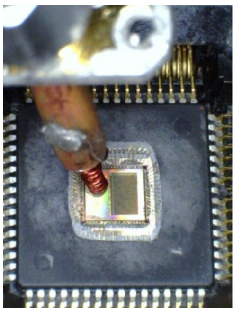
L'injection d'impulsions électromagnétiques consiste à générer localement de forts courants transitoires dans une bobine (l'antenne illustrée sur la Figure 1.30.a). Un champ électromagnétique [86] est alors créé au-dessus du circuit. Le couplage de cette sonde au circuit provoque un effet inductif sur le composant. Ce couplage est susceptible de modifier soit la polarisation et les rails d'alimentation du circuit, soit les arbres d'horloge.

Bien que spatialement moins précise que l'injection LASER, cette technique permet tout de même l'injection de fautes de manière relativement locale [81].

3.2.2.3 Altération de l'arbre d'horloge du circuit :

L'altération de l'arbre d'horloge a pour principal objectif de modifier les délais de propagation des signaux, entraînant la violation de contraintes temporelles. Cette altération de l'arbre d'horloge peut produire des erreurs, dans l'opération d'un algorithme de cryptographie par exemple.

(a) Exemple d'une sonde électromagnétique.



(b) Analyse d'un rayonnement électromagnétique via une sonde.

FIGURE 1.30 – Injection d'impulsions électromagnétiques [87].

En effet, lorsque les données ne parviennent par à traverser le chemin critique suite aux violations temporelles, alors le résultat de l'opération est une faute transitoire [88]. Cette attaque ne représente qu'un faible coût de mise en œuvre.

3.2.2.4 Variation de la température du circuit :

La plage de fonctionnement en température d'un dispositif dépend de la technologie utilisée et des applications qu'il vise, grand public ou militaire. L'altération de cette température hors des plages prévues par le constructeur conduit à la génération de fautes qui peuvent être irréversibles (permanentes ou même destructives).

3.2.2.5 Perturbation de l'alimentation ou de la tension de polarisation du substrat :

Cette dernière pratique consiste à modifier la tension d'opération du circuit soit par la modification de la tension des plots d'alimentation [89], soit par l'altération des conditions de polarisation du substrat du composant. Cette modification peut être accomplie de manière, plus ou moins importante (en diminuant ou en augmentant les niveaux de tension) et plus ou moins longtemps, pour compromettre le fonctionnement de la structure [88]. Alors que d'une part la modification de l'alimentation est peu coûteuse et non-localisée, l'altération du potentiel du substrat est plus coûteuse bien que l'erreur induite dans ce cas est localisée.

3.2.2.6 Comparaison des différentes attaques par perturbation :

Le tableau 1.2 illustre les avantages et inconvénients de ces différentes techniques de perturbation des circuits intégrés. Selon les coûts disponibles pour la caractérisation sécuritaire de la cible et les risques de destruction autorisés, une méthode de perturbation peut être préférée à une autre. Dans le cadre des travaux qui sont présentés dans ce manuscrit nous nous intéresserons principalement aux attaques par injection laser pour leur excellente résolution spatiale.

Technique	Précision spatiale	Précision temporelle	Coût	Risque de destruction
Laser	Forte	Forte	Élevé	Moyen
Impulsion EM	Moyen	Moyen	Moyen	Moyen
Horloge	Aucune	Forte	Faible	Faible
Alimentation	Aucune	Forte	Faible	Aucun
Substrat	Forte	Forte	Élevé	Moyen

TABLEAU 1.2 – Résumé des avantages et inconvénients de différentes techniques d'attaques matérielles par perturbation.

3.2.3 Les fautes et leurs modèles

Ces attaques par perturbation induisent des fautes de différentes natures dans un circuit intégré : principalement transitoires ou permanentes et peuvent répondre à différents modèles de fautes qui seront explicités ci-dessous.

3.2.3.1 Catégories de fautes :

— Les fautes transitoires ou provisoires : cette catégorie regroupe toutes les fautes dites temporaires qui sont induites dans les circuits intégrés suite aux perturbations. Lors d'une injection de fautes, un pic de courant est induit dans le circuit intégré et est suffisamment fort pour se propager dans le reste du circuit. Cette perturbation induit des erreurs sur son chemin de propagation puis se dissipe au bout d'un certain temps. La faute ainsi générée est réversible, en redémarrant le circuit par exemple.

— Les fautes permanentes : les fautes permanentes représentent la seconde catégorie de fautes qui peuvent être observées après une perturbation physique du circuit. Toutefois, ces fautes sont irréversibles et le nœud perturbé ne peut pas retrouver son fonctionnement initial, même après un redémarrage du circuit. Cette catégorie de fautes est moins intéressante pour un attaquant car elles rendent impossible la multiplication des attaques ciblées afin de reconstituer le message secret.

La modélisation de l'effet que peut induire une perturbation physique sur un circuit intégré permet par exemple à l'attaquant de retrouver la clé utilisée pour le chiffrement d'un message clair via un algorithme de cryptographie. La sensibilité des circuits intégrés à ces attaques augmente sensiblement avec la diminution de la taille des transistors dans le cadre de la loi de Moore.

3.2.3.2 Les modèles de fautes :

Dans le cadre de cette thèse, nous nous intéresserons principalement aux fautes qui peuvent être générées dans les éléments mémoires. Dans le cadre de la sécurité de ces circuits intégrés, ses fautes sont de type :

— Inversion de bit (ou *bit-flip*) correspondant à l'inversion de la valeur sauvegardée dans un bit. Ainsi, l'expression mathématique représentant ce modèle de fautes peut être défini tel que le bit inversé b_f est défini en fonction du bit initial b_i avec : $b_f = 1 - b_i$.
Il peut être décomposé en 2 sous-modèles usuellement utilisés dans le monde de la sécurité.

- *Bit-set* :
  Le *bit-set* correspond à la mise à '1' du bit visé par l'attaque, alors que celui-ci était initialement à '0'. Le bit induit b_f par ce modèle de fautes correspond à $b_f = 1$.
- *Bit-reset* :
  Le *bit-reset* correspond à la mise à '0' du bit visé par l'attaque, alors que celui-ci était initialement dans un état '1'. Le bit induit b_f par ce modèle de fautes correspond à $b_f = 0$.

— Collage : Cette technique consiste à coller un nœud du circuit, à une valeur non-modifiable permanente. Toute valeur susceptible d'y être affectée reste illisible et le nœud reste collé à cette même valeur. Le collage peut être considéré dans certains cas comme une erreur permanente qui n'a pas endommagé le composant physique. Toutefois, il peut aussi résulter d'une faute destructive qui a endommagé le nœud physique visé. Dans ce cas, le circuit devient inopérationnel et inutilisable. Cette catégorie de fautes relève davantage de la fiabilité et peut être donc moins intéressante à traiter dans le cadre de la sécurité des circuits intégrés. Le collage peut être décomposé en 2 sous-catégories :

- Collage à '0' :
  Ce modèle de fautes est représenté par le collage du bit b_i à '0' à un temps t. Alors, soit le bit b_f correspondant au bit b_i à un temps $t+\tau$, tel que $b_f = b_i = $ '0' .

- Collage à '1' :
  De manière complètement analogue et symétrique, le collage du bit b_i à '1' à un temps t induit la continuité de cette valeur ('1') à un temps $t+\tau$ pour le même bit b_i noté b_f. Ainsi, $b_f = b_i = $ '1' , $\forall t$.

4 Sécurité dans la logique hybride CMOS/STT-MRAM

La sécurisation des objets connectés est un enjeu primordial de cette décennie. Toutefois, cette recherche pour améliorer les niveaux de protection des circuits intégrés, ne doit pas se faire au détriment de leurs performances. En effet, l'Internet des Objets vise toujours les meilleures caractéristiques en terme de consommation, vitesse de fonctionnement et surface d'implémentation. Pour cela, des architecture hybrides CMOS/mémoires non-volatiles émergentes sont étudiées. Ces structures présentent des atouts majeurs pour le domaine de l'Internet des Objets en terme de sécurité, consommation et vitesse.

4.1 Les applications "normally-off/instant-on"

Dans le contexte de cette thèse, nous avons choisi de travailler sur la sécurité de la logique hybride qui a montré un intérêt majeur dans la réduction de la consommation des circuits intégrés [90] et donc des nouvelles caractéristiques qu'apporte ces nouvelles propriétés. En effet, l'intégration de la non-volatilé dans les composants permet sa mise hors tension lorsque ceux-ci ne sont pas utilisés et donc ainsi réduire la consommation statique du circuit sans perte de données, comme illustré Figure 1.31.

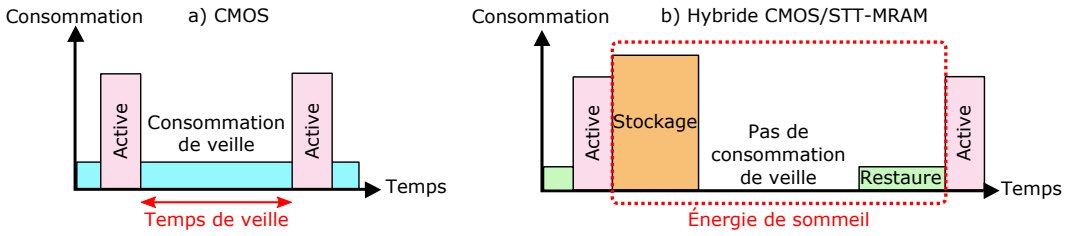


FIGURE 1.31 – Consommation de la logique CMOS et de la logique hybride.

Les premières bascules hybrides développées introduirent une circuiterie supplémentaire (par rapport aux bascules CMOS) pour l'écriture et la lecture des données dans les dispositifs mémoires émergents, permettant ainsi l'écriture d'un bit par bascule [91]. Ces bascules sont dites non-volatiles ou *Non-Volatile Flip-Flops* (NVFFs) .Toutefois, des structures plus avancées permettant désormais le stockage de plusieurs bits pour une même bascule sont développées. Ces dernières sont dites bascules multi-contextes non-volatiles (MC-NVFFs).

En outre, ces mémoires émergentes ont également permis le développement de structures visant la sécurité des objets, tels que la Fonction Physique Non-clonable ou – *Physically Unclonable Function* – (PUF) ou le Générateur de Nombres Aléatoires ou – *True Random Number Generator* – (TRNG).

4.2 Les structures hybrides CMOS/STT-MRAM

4.2.1 Les bascules non-volatiles

Les bascules hybrides utilisent des technologies émergentes afin de stocker les données en cours de traitement dans la bascule, dans des points mémoires non-volatiles. Cette méthodologie permet la mise hors tension des circuits inutilisés afin de limiter la consommation générale. Lorsque le système est éteint, le dispositif ne consomme plus d'énergie statique contrairement au cas où les données seraient complètement volatiles dans une bascule synchrone ou – *Flip-Flop* – (FF) standard.

Quelle que soit la technologie utilisée pour les points mémoires, la technique de réalisation de la bascule synchrone non-volatile ou – *Non-Volatile Flip-Flop* – (NVFF) reste toujours la même. Pour cela, l'un des deux verrous (ou *latch*) est remplacé par un verrou non-volatile intégrant les éléments mémoires, comme illustré sur la Figure 1.32.

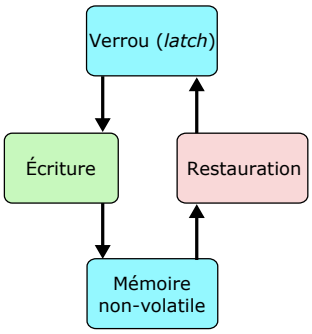


FIGURE 1.32 – La structure d'un verrou (ou *latch*) de la bascule non-volatile NVFF.

Avant la mise hors tension du circuit, les valeurs sont stockées dans les éléments mémoires. Après la réactivation du circuit, ces données sont restaurées dans les bascules. La structure peut alors poursuivre son fonctionnement standard [92]. Cette méthodologie est par exemple appliquée au cas des mémoires OxRAM dans [93].

Dans le cadre de cette thèse, les NVFFs qui nous intéresserons sont celles qui sont à base de la technologie mémoire STT-MRAM [94], [95], [96] et [97].

4.2.2 Les bascules multi-contextes CMOS/STT-MRAM

Ces avancées dans les bascules hybrides ont permis le développement d'une innovation majeure dans le monde des circuits intégrés : la Bascule Synchrone Non-Volatile Multi-Contexte ou – *Multi-Context Non-Volatile Flip-Flop* – (MC-NVFF). Cette nouvelle catégorie de bascules a tout d'abord été introduite par M. Hariyama *et al.* en 2008 [98] puis par W. Zhao *et al.* qui l'ont appliqué à la technologie mémoire STT-MRAM [99] puis a été fortement étudiée dans la littérature pour tous ces attraits préalablement cités [100], [101], [102]. Cette architecture est décrite sur la Figure 1.33 et a été proposée par D. Chabi *et al.* dans [91].

Elle est composée de :

— Une bascule CMOS (FF) standard qui sert à transiter l'information à stocker ou à restaurer, bloc (a).

— Un circuit de lecture associé à un amplificateur (ou *sense amplifier*) pour restaurer le contexte, bloc (b). Ainsi, plusieurs jonctions en parallèles sont instanciées pour stocker différents contextes dans une même bascule.

— Un circuit d'écriture du contexte d'une jonction visée, blocs (c) et (d).

— Un circuit de désactivation de l'alimentation (*power gating*), bloc (e).

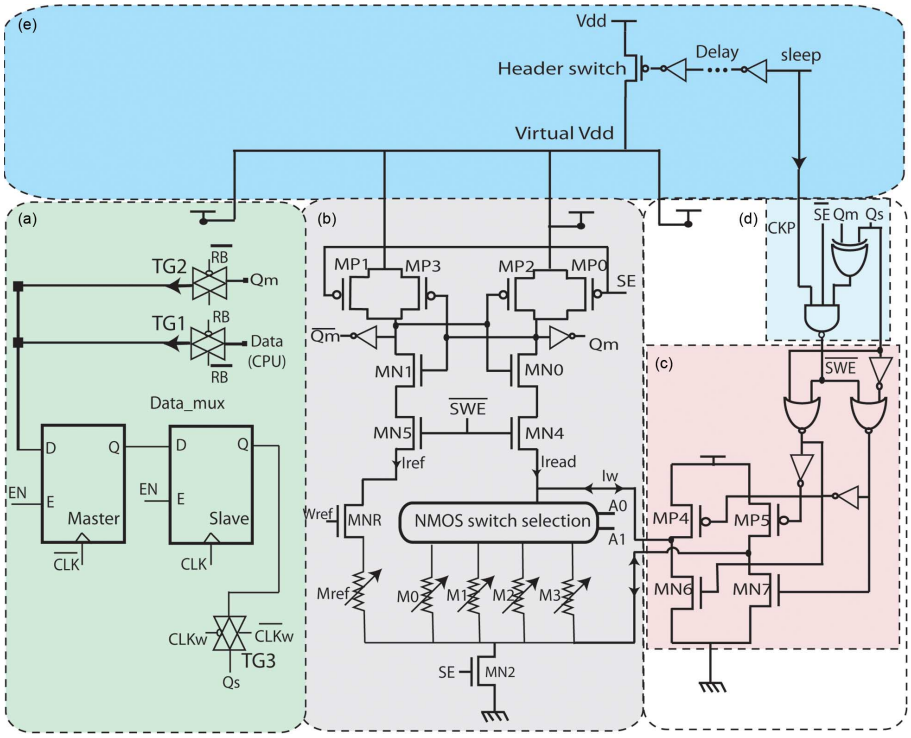


FIGURE 1.33 – Illustration de la bascule multicontexte telle qu'illustrée dans [91] décomposée selon ses différentes composantes.

Cette catégorie de bascules apporte un avantage majeur en terme d'hybridation et de réduction de la consommation des circuits intégrés. Ainsi, l'utilisation de ces composants dans des architectures de sécurité comme les algorithmes de cryptographie ou de protocoles de sécurité ouvre de nouvelles opportunités. Toutefois, cette intégration impose également de nouvelles contraintes et des risques qui feront l'objet d'une étude, réalisée pendant cette thèse.

4.3 Les architectures hybrides de sécurité

4.3.1 Les PUFs

La Fonction Physique Non-clonable ou – *Physically Unclonable Function* – (PUF) est utilisée pour l'implémentation de protocoles de sécurité, comme par exemple l'authentification du circuit [103]. Elle exploite les variabilités intrinsèques qui sont introduites par les procédés de fabrication. Ces variabilités conduisent à une signature unique et invariable dans le temps [104], pour chaque dispositif. Ainsi, cette structure se doit d'être robuste face aux variations environnementales. C'est pourquoi, ces signatures peuvent être vues comme les données biométriques du circuit.

Différentes architectures de PUFs ont été proposées : des PUFs arbitraires [105], des PUFs basés sur des oscillateurs en anneaux [106] ou encore sur les SRAMs [107]. Pour toutes ces structures, il est nécessaire de s'assurer que pour chaque *challenge* de taille n envoyé au PUF, une réponse de taille m est générée. Cette réponse dépend des caractéristiques physiques du circuit intégré [108]. Plus le nombre de paires *challenge-response* est important, plus le PUF est dit fort. Ainsi, des propriétés d'unicité, de réponse aléatoire et de stabilité sont recherchées pour chaque PUF.

Outre ces primitives de sécurité purement CMOS, de nouvelles structures exploitant la variabilité de la résistance de la JTM sont étudiées. Le courant de commutation de ces JTMs est un candidat prometteur pour l'identification de ces cellules [109]. En effet, la dépendance de la commutation des JTMs en fonction de la durée de l'impulsion appliquée est un enjeu majeur des JTM-

PUFs. De plus, les variations lors du procédé de fabrication, dans la taille des dispostifs et dans l'épaisseur des matériaux magnétiques utilisés dans les JTMs conduit à des caractéristiques de commutation distinctes [109]. Ces différentes particularités font de l'intégration de la STT-MRAM une opportunité pour ces primitives de sécurité.

La Figure 1.34.a illustre une architecture PUF.

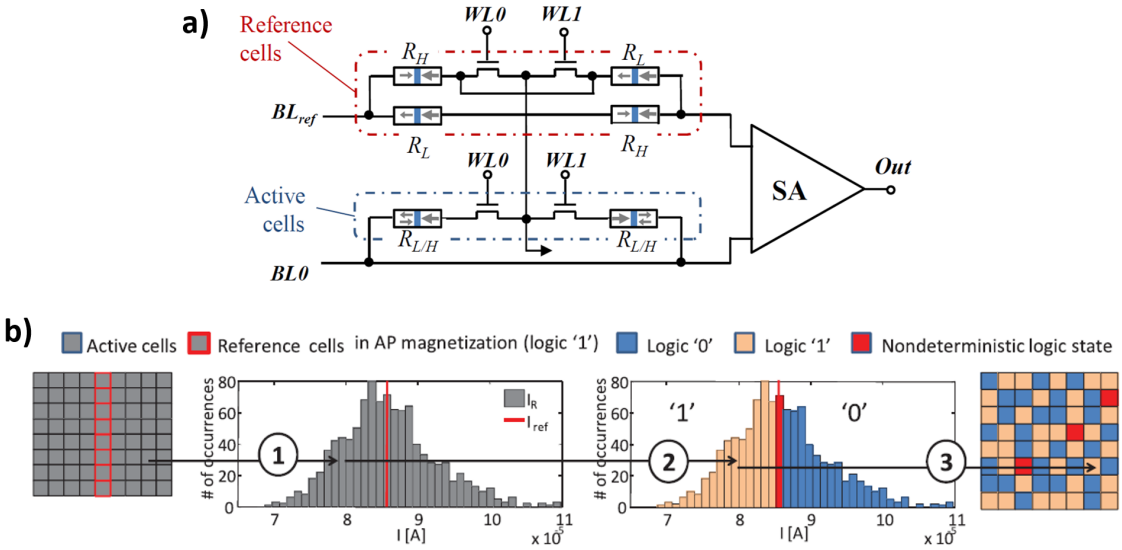


FIGURE 1.34 – Représentation d'une architecture PUF différentielle [103].

Dans cet exemple, des cellules actives (*active cells*) sont comparées à des cellules de référence (*reference cells*) (dont la résistivité est la moyenne de l'état de forte résistivité et de faible résistivité). L'état logique de chaque cellule active est positionné par rapport à cette référence, comme illustré Figure 1.34.b. Ainsi, une matrice unique de 64 bits (*response*) est obtenue dans cet exemple.

4.3.2 Les TRNGs

Outre le PUF, le Générateur de Nombres Aléatoires ou – *True Random Number Generator* – (TRNG) est un composant crucial des algorithmes cryptographiques [110]. Il peut par exemple être utilisé dans des chiffrements par flots pour générer les vecteurs d'initialisation aléatoires qui sont utilisés pour le calcul de la clé de chiffrement. Les architectures TRNGs utilisent principalement le bruit thermique pour générer ces nombres aléatoires [104].

Les défis de ces architectures stochastiques sont l'exploitation de nouvelles sources d'entropie et le développement de mécanismes consciencieux afin de réduire les erreurs de ces systèmes. En effet, pour chaque TRNG, des exigences de répétabilité, non-linéarité et résilience aux attaques doivent être vérifiées. De plus, compte tenu que les applications à ressources restreintes imposent des vitesses de fonctionnement importantes pour une consommation réduite, l'instanciation de l'élément mémoire non-volatile STT-MRAM peut devenir un réel atout pour ces structures [111].

Les structures hybrides STT-MRAM-TRNGs tiennent compte des phénomènes stochastiques et de variabilités induits par les JTMs. Afin de réaliser cette génération aléatoire, trois étapes sont nécessaires :

— Toutes les jonctions composant cette structure sont programmées dans l'état P.

— Puis, une impulsion pour écrire l'état AP est appliquée à ces composants. Toutefois, la durée et l'amplitude de cette impulsion sont fixées de façon à atteindre une probabilité de commutation de 50 % des cellules. Ainsi, après cette impulsion, 50 % des cellules commutent vers l'état AP alors que 50 % restent dans l'état P [110].

— La dernière étape de ce générateur de nombres aléatoires est alors de lire l'état des différentes jonctions, ce qui forme ce nombre aléatoire qui va par exemple entrer en jeu dans le calcul de la clé.

5 Conclusion

L'objectif de cette thèse consiste à réaliser une étude sur les capacités de l'hybridation de la technologie CMOS et de la technologie mémoire STT-MRAM, dans le cadre d'une intégration dans des solutions cryptographiques sécurisées pour l'Internet des Objets. C'est pourquoi ce premier chapitre illustre les différentes composantes multidisciplinaires complémentaires qui seront nécessaires afin de répondre à la problématique de cette thèse. En effet, le but de ce manuscrit est de développer une architecture hybride dont les performances en terme de sécurité (importante), de vitesse de fonctionnement (importante) et de consommation (faible) soient accordées aux besoins des applications comme l'Internet des Objets. Pour cela, ce chapitre a débuté par un état de l'art non-exhaustif des mémoires existantes en mettant en avant leurs potentiels et limites, grâce à cette étude le choix de la technologie mémoire STT-MRAM a été réalisé. Puis dans une seconde partie nous avons étudié les besoins en terme de sécurité des circuits intégrés et des avantages (et inconvénients) possibles de l'hybridation sur les performances des circuits.

Chapitre 2

Intégrité des mémoires STT-MRAMs

Vous ne trouverez jamais
ce que vous ne cherchez pas

— Confucius

L'hybridation de la technologie CMOS et de technologies mémoires émergentes pour le développement d'applications "Normally-Off" visant l'Internet des Objets nécessite de garantir la robustesse et la fiabilité aussi bien de la partie MOS que de la partie mémoire. La mémoire émergente étudiée est la mémoire STT-MRAM, elle a été choisie car elle présente des caractéristiques exceptionnelles pour ce type d'applications.

Une étude sécuritaire pour déterminer l'intégrité de jonctions unitaires face à des attaques par perturbations de type LASER est menée. En effet, les attaques LASERs sont réputées pour leur efficacité de par la finesse de leur résolution temporelle et spatiale.

Ainsi, la première section de ce chapitre traitera des phénomènes physiques induits par les injections LASERs et les propriétés de ces faisceaux. Puis, le protocole expérimental suivi afin de réaliser le test de l'intégrité des jonctions STT-MRAMs et les résultats observés seront décrits dans la seconde section. Enfin, la modélisation des phénomènes physiques induits par le tir LASER sont simulés dans la dernière partie.

Les travaux présentés dans ce chapitre ont été publiés à la conférence internationale "IOLTS 2018" [83].

Sommaire

1 Le LASER

L'Amplification de la Lumière par Émission Stimulée de Radiation ou – *Light Amplification by Stimulated Emission of Radiation* (LASER) est un système photonique qui produit une radiation électromagnétique cohérente et unidirectionnelle. Il est également un instrument qui présente deux intérêts majeurs : une résolution spatiale et une résolution temporelle extrêmement fines (respectivement de l'ordre du micromètre et de la picoseconde, et même femtoseconde à l'état de l'art industriel). En effet, pour des faisceaux gaussiens, la résolution spatiale atteignable est égale à la longueur d'onde du laser. Ces caractéristiques en font un outil de choix dans des études de fiabilités et/ou de sécurités des dispositifs microélectroniques.

1.1 Principe de fonctionnement d'une source LASER

Les différentes technologies LASERs peuvent fonctionner sous deux modes :

— En continu : l'énergie est envoyée de façon continue par la source de pompage. Dans ce cas, la puissance en sortie du LASER est modeste.

— Par impulsion : l'énergie du faisceau est envoyée pour une période définie. La durée de l'impulsion dépend de la précision de l'équipement et des demandes de l'application visée. Elle peut varier de la femtoseconde à la milliseconde.

La structure physique du LASER est composée comme illustré sur la Figure 2.1 par :

— Une cavité où le faisceau cohérent va naître. Elle est composée d'un matériau qui peut être soit sous forme gazeuse, liquide ou encore solide selon la catégorie du LASER. Cette cavité confine et amplifie les interactions atome-rayonnement dans cet espace pour obtenir un faisceau lumineux en sortie.

— Deux miroirs qui sont positionnés aux deux extrémités de la cavité. Alors que l'un de ces miroirs est complètement opaque et réfléchit tout le faisceau lumineux présent dans la cavité, le second transmet une partie du faisceau vers l'extérieur.

— Une source d'énergie qui est nécessaire pour initier le processus. Cette source de pompage peut être soit électrique soit optique selon le milieu amplificateur qu'elle doit activer. Elle fournit suffisamment d'énergie à la cavité pour pouvoir activer les interactions atome-rayonnement qui s'y produisent.

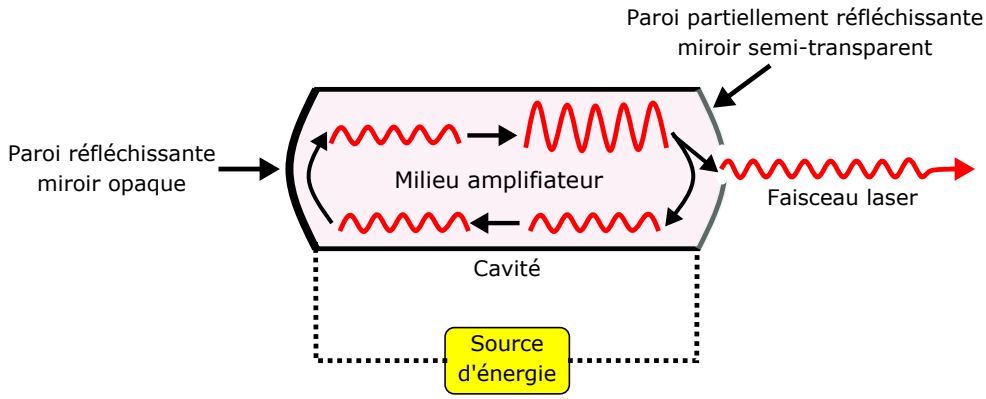


FIGURE 2.1 – Représentation de la composition d'une source LASER.

Ainsi, le faisceau lumineux laser créé par cet instrument est induit par des effets physiques, des interactions atome-rayonnement. Les interactions pouvant se produire à l'intérieur de la cavité sont diverses. En effet, comme il peut être noté sur la Figure 2.2, lorsque l'électron change de

couche électronique pour passer d'un niveau d'énergie supérieur E_n (tel que n le nombre quantique principal ou la dimension de l'orbitale) vers des niveaux d'énergies inférieurs, la différence d'énergie entre ces deux niveaux ΔE conduit à l'émission d'un photon par l'atome. Ce photon aura une longueur d'onde donnée (λ exprimée en nm) qui dépendra de la nature de l'atome et des changements d'états possibles pour ce même atome. Cette opération est la désexcitation de l'atome.

Si l'on considère par exemple l'atome d'Hydrogène, lorsque l'électron le composant est sur un niveau d'énergie supérieur ou égal à 2, alors l'atome aura tendance à vouloir perdre cet état d'excitation. Dans ce cas de figure, il pourra passer par exemple des orbitales 2 ou 3 vers le niveau fondamental en émettant un photon ultraviolet (de longueur d'onde respectivement : λ_{21}=122 nm et λ_{31}=103 nm), comme illustré Figure 2.2. Ces interactions vers le niveau fondamental composent la série de Lyman [112]. En outre, il est également représenté une interaction du niveau d'énergie n = 3 vers le niveau n = 2. Comme il peut être noté, ce phénomène conduit à l'émission d'un photon dans le domaine du visible (λ_{32} = 656 nm, rouge). Cette interaction fait quant à elle partie de la série de Balmer [113].

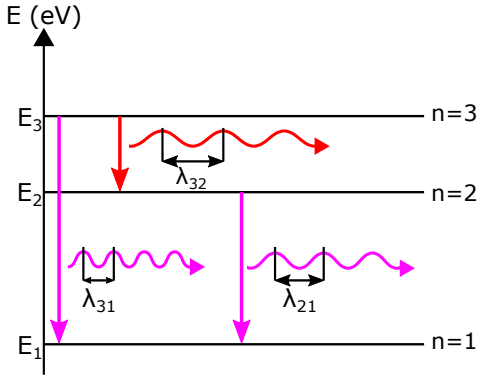


FIGURE 2.2 – Transitions énergétiques possibles de l'électron de l'atome d'Hydrogène.

La longueur d'onde émise par un électron lors de son changement de niveau électronique peut être déterminée en considérant : h la constante de Planck, c la célérité de la lumière dans le vide et ΔE la différence d'énergie entre deux niveaux d'orbitales. Cette expression est décrite par l'Équation 2.1.

$$\Delta E = h\frac{c}{\lambda} \tag{2.1}$$

Lors de la représentation du changement de couche électronique des électrons des atomes contenus dans la cavité du laser, il a été choisi pour des questions de simplification d'utiliser la représentation de E. Rutherford [114] pour décrire le fonctionnement intra-atome au lieu de la représentation de E. Shrödinger [115] où l'électron est assimilé à une fonction d'onde ψ qui est plus complète.

Les interactions atome-rayonnement qui peuvent régner dans l'enceinte de la cavité du LASER sont au nombre de trois et sont tout d'abord définies en fonction de la présence ou de l'absence d'un rayonnement incident [116].

— En présence d'un rayonnement incident :

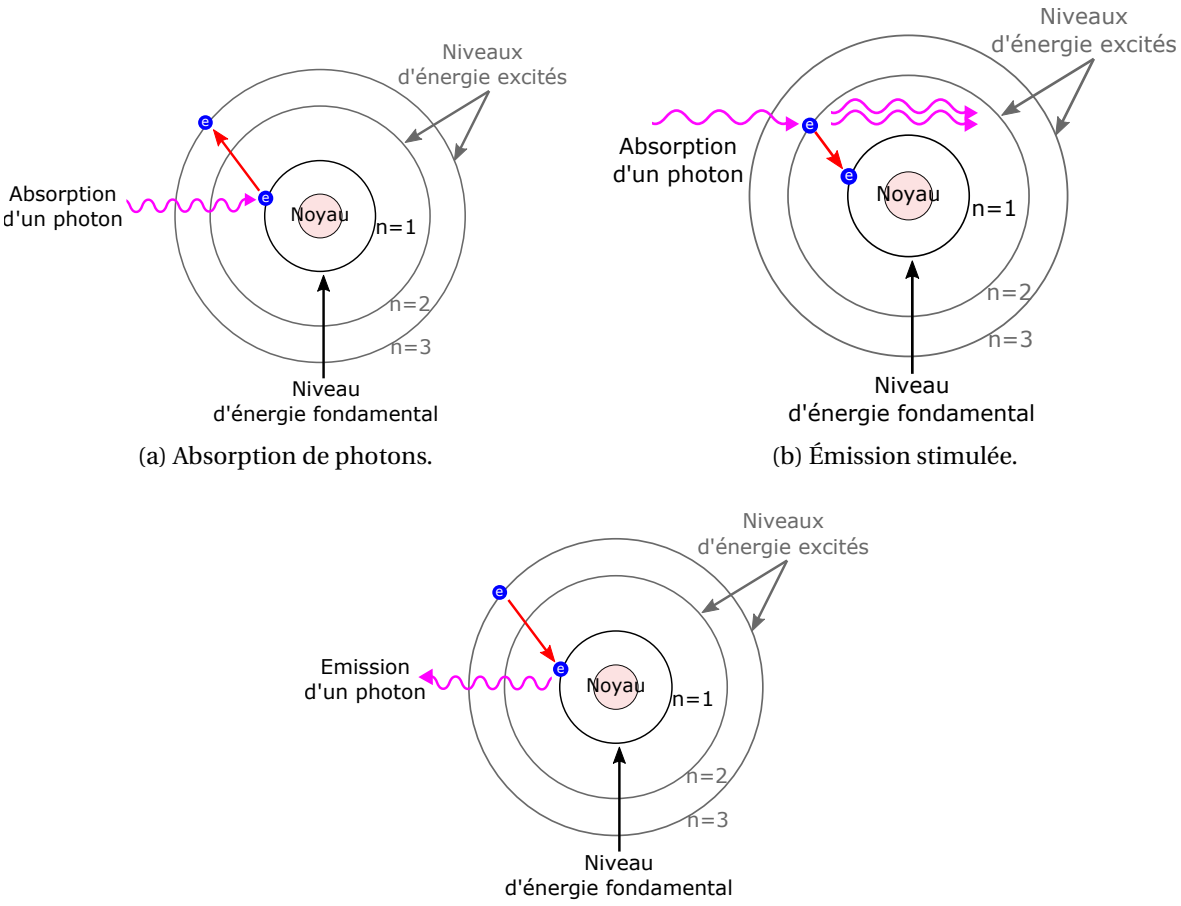
- Absorption :
  Cette première catégorie d'interactions est induite par un rayonnement externe incident sur un électron de l'atome. Cet électron va alors absorber l'énergie de ce photon et changer de niveau énergétique pour passer à un niveau supérieur. L'électron est alors

dans un état excité. En effet, comme représenté en exemple Figure 2.3.a, selon la longueur d'onde λ incidente, l'électron pourra passer de la couche d'énergie fondamentale (n = 1) à la couche n = 3.

- Émission stimulée :
  Le second phénomène qui peut se produire en présence d'un rayonnement incident sur l'atome est : l'émission stimulée. Cette interaction rayonnement - matière consiste pour l'électron à absorber le photon incident et émettre deux photons de même longueur d'onde, puisque son énergie est suffisante. Cette libération de photons et d'énergie se traduira alors par la régression de l'électron vers l'état fondamental comme représenté Figure 2.3.b.

— En l'absence d'un rayonnement incident :
Cette dernière catégorie d'émission est la seule n'étant pas le résultat de l'incidence d'un photon sur l'atome.

- Émission spontanée :
  L'émission spontanée est le résultat du changement de niveau d'énergie de l'électron de façon spontanée en émettant un photon dont la longueur d'onde dépend de l'état initial de l'électron et son niveau d'énergie d'arrivée comme décrit par l'Équation 2.1 et illustré Figure 2.3.c.



(a) Absorption de photons.

(b) Émission stimulée.

(c) Émission spontanée.

FIGURE 2.3 – Les interactions atome-rayonnement en présence (ou non) d'un faisceau incident.

Afin qu'un faisceau laser puisse fonctionner, il est nécessaire que l'émission stimulée soit prépondérante à l'intérieur de la cavité. D'où la présence du terme "Émission Stimulée" dans l'expression LASER (correspondant au "S" et au "E"). Ce phénomène est ainsi induit par l'utilisation

d'une source d'énergie qui pourra exciter les atomes mis en jeu dans la cavité. Les électrons des différents atomes interagiront en utilisant l'un des phénomènes explicités précédemment.

D'autre part, la longueur d'onde du faisceau lumineux est caractérisée par une lumière monochromatique qui peut se trouver dans le domaine du visible (vert, rouge ...), de l'infrarouge ou encore de l'ultraviolet.

1.2 Les différentes catégories de LASERs

Il existe différentes catégories de LASERs présentant des caractéristiques distinctes. Ils peuvent être décomposés selon la nature et l'état physique de la cavité (ou milieu amplificateur). Ainsi, il existe des LASERs :

1.2.1 Les lasers à gaz

Le mode d'excitaton des électrons et des atomes de cette famille est essentiellement induite par un pompage électrique de la cavité. Bien que fortement minoritaires, une activation thermique ou chimique peuvent également être développées [117]. Cet outil peut délivrer des radiations de longueurs d'ondes allant de l'ultraviolet à l'infrarouge, passant donc également par le visible. Il présente une forte mobilité des atomes et des ions dans la cavité du laser.

En outre, il peut également être noté une forte disparité entre les niveaux de puissance que peuvent fournir ces composants : du milliWatt au kiloWatt. Nous pouvons citer différentes espèces comme représentées sur la Figure 2.4. Dans cette gamme de LASERs, le Dioxyde de Carbone CO_2 est celui qui a le meilleur rendement, entre 15 % et 20 % [118][119]. En effet, le rendement de ces outils est une composante essentielle qui est déterminée par le rapport de la puissance optique délivrable par l'équipement sur la puissance fournie à l'amplificateur ou cavité [120]. Cette catégorie de LASERs pouvant développer de fortes puissances de l'ordre du kiloWatt est fortement utilisée dans l'industrie, comme par exemple pour de la découpe.

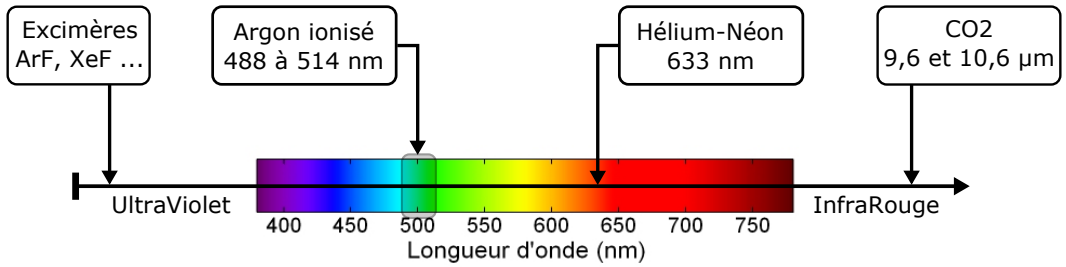


FIGURE 2.4 – Représentation de certains lasers à gaz selon leur domaine spectroscopique.

1.2.2 Les lasers à liquides (ou à colorants chimiques)

Cette catégorie de LASERs est comme son nom l'indique développée grâce à un milieu actif composé d'un colorant organique dans un solvant liquide. Contrairement aux LASERs à gaz, ces LASERs liquides fonctionnent grâce à un pompage optique [121]. En effet, l'activation de ce LASER est réalisée soit grâce à un second laser soit via des lampes à arc (minoritaires) [122]. Ces instruments peuvent fournir différentes longueurs d'onde dans la gamme du visible. Toutefois, ils sont peu utilisés compte tenu de leur faible durée de vie et de la toxicité de certains colorants employés.

1.2.3 Les lasers à solides

Cette dernière catégorie peut se décomposer en deux sous-parties délimitées par le moyen utilisé pour réaliser le pompage : les LASERs à semi-conducteurs et les LASERs solides cristallins.

— Les LASERs à semi-conducteurs ou LASERs à diode sont électriquement initiés par des recombinaisons de paires électrons-trous. Ils sont présentés comme les LASERs des plus compactes et efficaces [123] pouvant atteindre des rendements de 50 %. Dans cette famille de LASERs, il est possible de citer des équipements à base de Nitrure de Gallium (GaN) ou encore d'Arséniure de Gallium-Aluminium (AlGaAs). Toutefois, leur inconvénient majeur est leur faible résolution spatiale.

— Les LASERs solides cristallins sont définis par le cristal composant leurs milieux amplificateurs. Cette catégorie offre les meilleurs rendements (supérieurs à 50 %) et peut émettre de l'ultraviolet à l'infrarouge. En outre, ils sont activés optiquement par des lasers de type diodes ou encore par des lampes flash [121][123]. Nous pouvons citer le Grenat d'Yttrium-Aluminium dopé au Néodyme ou – *Neodymium doped Yttrium-Aluminum-Garnet* – (Nd-YAG) (Nd^{3+} : $Y_3Al_5O_{12}$) émettant principalement à 1064 nm. La bande d'émission la plus large est obtenue avec un Saphir dopé au Titane (Ti^{3+} : Al_2O_3), pouvant émettre entre 700 nm et 1100 nm [123].

  La structure fibrée est l'une des structures les plus développées pour des applications comme le traitement des matériaux ou de fiabilité des circuits intégrés à plus faible puissance par exemple. Son milieu amplificateur est composé d'une fibre optique. L'ordre de grandeur des puissances que peuvent fournir ces dispositifs va du milliWatt à la dizaine de Watts [124]. Ce faisceau laser est souvent dans l'infrarouge. Ces dispositifs ont des distances focales extrêmement petites permettant un tir laser très localisé (de l'ordre du micromètre) et de forte intensité.

2 Injection de fautes par laser sur jonctions unitaires STT-MRAMs

Dans le cadre de l'intégration de technologies hybrides CMOS/STT-MRAM dans diverses applications visant l'Internet des Objets, la vérification de la sécurité de cette technologie mémoire est primordiale. Ainsi, la première étape des travaux qui ont été entrepris lors de cette thèse a été de vérifier la sensibilité de dispositifs mémoires élémentaires STT-MRAMs, à aimantations perpendiculaires (dernière génération), à des perturbations extérieures de type LASER.

2.1 Protocole expérimental

L'intégrité des données contenues dans ces jonctions unitaires STT-MRAMs fournies par le laboratoire **Spintec** est caractérisée dans ce chapitre. Pour ce faire, le protocole expérimental qui a été mis en place spécifiquement pour cette étude est présenté dans ce chapitre.

Ainsi, les cellules ont été dans un premier temps testées électriquement. Compte tenu que ces dispositifs sont des composants de R&D, un conditionnement (ou cyclage) et une cartographie initiale vérifiant leur fonctionnalité (mesure de la résistivité) sont réalisés, comme représenté Figure 2.5 (étape 1). Puis, dans un second temps l'attaque LASER (étape 2) est effectuée en utilisant l'équipement LASER Nd-YAG qui fait partie de la familles des LASERs à solides. Ces dispositifs soumis aux perturbations LASERs sont ensuite électriquement testés (étape 3), afin de s'assurer de leur non-destruction et de leur comportement post-attaque. Ces différentes étapes du protocole suivi sont développées dans les sections suivantes, en mettant en avant les résultats obtenus.

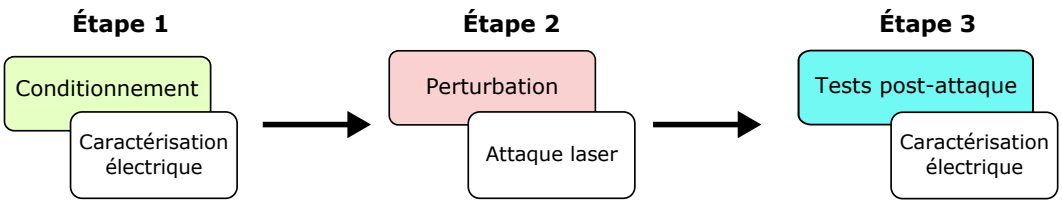


FIGURE 2.5 – Protocole expérimental suivi pour tester l'intégrité des jonctions STT-MRAMs.

2.1.1 Conditionnement initial des cellules STT-MRAMs

2.1.1.1 Banc de caractérisation électrique des dispositifs

La caractérisation et le conditionnement des dispositifs mémoires unitaires accessibles directement sur wafer sont réalisés grâce à une station sous-pointes, ou *prober*, équipée d'un Keysight B1530 photographié en Figure 2.6. La tension est injectée via les pointes de cet équipement directement dans les deux électrodes de la jonction puis le courant induit (et donc la résistivité) sont mesurés. Dans le cadre de ces expérimentations, deux familles de jonctions sont préparées avant attaque : des jonctions initialement dans l'état Parallèle P et des jonctions initialement dans l'état Anti-Parallèle AP.

FIGURE 2.6 – Station sous-pointes Cascade 200 mm utilisée pour réaliser la caractérisation électrique des jonctions mémoires unitaires STT-MRAMs.

Le bon fonctionnement du dispositif sous test est confirmé en vérifiant l'existence des deux valeurs de résistances stables et distinctes : R_P (faible résistivité) et R_{AP} (forte résistivité). Puis, chaque cellule est conditionnée électriquement afin de vérifier la variabilité des paramètres : $V_{Commutation}$ correspondant à la tension nécessaire pour commuter l'état de la jonction d'un état vers l'autre et $I_{Commutation}$ le courant correspondant nécessaire pour réaliser cette commutation. L'évolution de ces trois paramètres est mesurée (R, $V_{Commutation}$ et $I_{Commutation}$) sur 30 cycles. Chaque cycle est défini par les successions d'étapes notées (1), (2), (3) et (4) et représentées Figure 2.7 telles que :

(1) - Écriture de l'état P : commutation de AP vers P.

(2) - Opération de lecture : vérification de la valeur de résistance R_P.

(3) - Écriture de l'état AP : commutation de P vers AP.

(4) - Opération de lecture : vérification de la valeur de résistance R_{AP}.

La succession de ces 4 opérations est répétée 30 fois sur chaque cellule, toujours en suivant ce même schéma, afin de vérifier la stabilité de ces paramètres dans le temps ou de déterminer leur dérive.

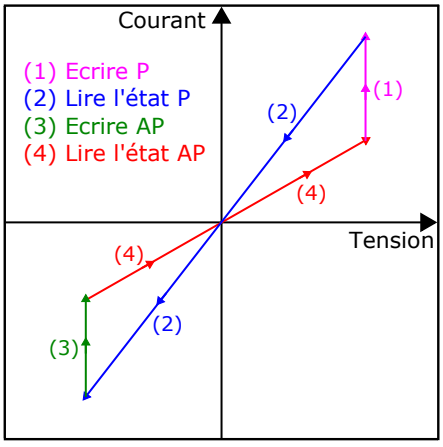


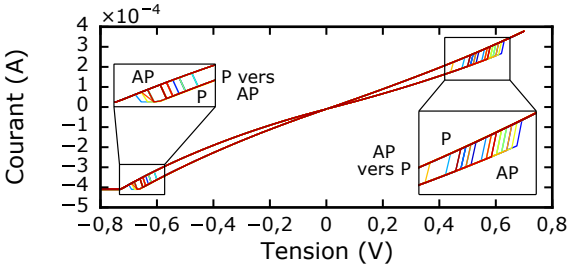
FIGURE 2.7 – Représentation I(V) des étapes d'**un cycle** de caractérisation.

La Figure 2.8 représente les caractérisations électriques du fonctionnement d'une cellule sur 30 cycles. Comme illustré, tous les paramètres évalués sont constants (très peu variables) et la

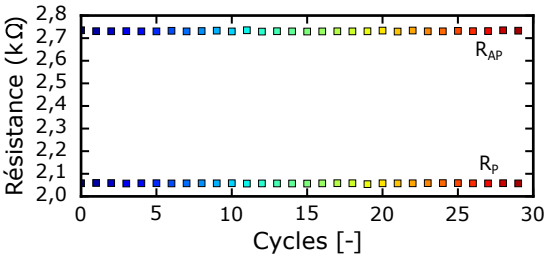
jonction reproduit le même fonctionnement électrique au fil des cycles.

En effet, les Figures 2.8.b, 2.8.c et 2.8.d démontrent respectivement l'invariance des valeurs de résistances R_P et R_{AP}, la constance des tensions de commutation de l'état AP vers P ($V_{Commutation\text{-}AP\text{-}vers\text{-}P}$) et inversement $V_{Commutation\text{-}P\text{-}vers\text{-}AP}$ et de la faible variation des courants de commutation $I_{Commutation\text{-}AP\text{-}vers\text{-}P}$ et $I_{Commutation\text{-}P\text{-}vers\text{-}AP}$.

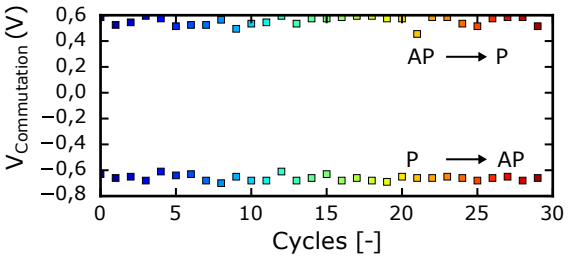
Cette jonction peut ainsi être considérée comme fonctionnelle.



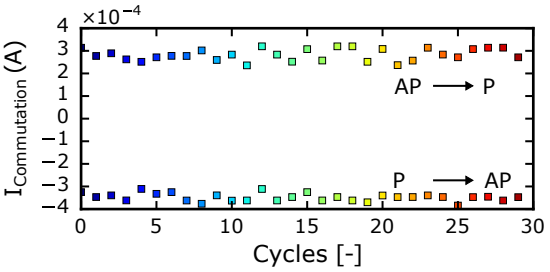
(a) Évolution du courant en fonction de la tension injectée dans la JTM.



(b) Évolution de la résistance en fonction du nombre de cycles.



(c) Évolution de la tension de commutation en fonction du nombre de cycles.



(d) Évolution du courant de commutation en fonction du nombre de cycles.

FIGURE 2.8 – Évolution de différentes grandeurs physiques mesurées sur une jonction unitaire STT-MRAM perpendiculaire avant attaque laser.

2.1.1.2 Vérification de la stabilité des paramètres physiques à 2 temps différents

Lorsque la fonctionnalité de ces dispositifs est confirmée sur 30 cycles, il est toutefois nécessaire de s'assurer de la stabilité de la valeur stockée dans la mémoire au fil du temps et de possibles variations des conditions extérieures du test, comme par exemple une modification de la température environnante lors de la caractérisation électrique des dispositifs.

Pour cela, la JTM doit présenter des distributions normales notées $\mathcal{N}$ invariables et qui ne divergent pas dans le temps, pour tous les paramètres électriques : R_P, R_{AP}, $V_{Commutation}$ et $I_{Commutation}$. Pour tous les échantillons testés, ces distributions sont calculées d'après l'Équation 2.2, exprimée en fonction de :

— x le paramètre électrique mesuré sous station sous-pointes (R, $I_{Commutation}$ ou $V_{Commutation}$), pour N cycles (dans ce cas 30 cycles).

— μ la valeur moyenne de la distribution tel que : $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$.

— σ son écart-type s'exprimant tel que : $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$.

La distribution normale de x est alors définie par l'équation :

$$\mathcal{N}(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2.2}$$

En effet, comme démontré sur la Figure 2.9, les courbes R_P à un temps t = 0 s et après 3 jours (de façon équivalente, R_{AP} à un temps t = 0 s et R_{AP} après 3 jours) démontrent que les variables R_P et R_{AP} restent quasiment constantes et valent : $R_P = 2,13$ kΩ et $R_{AP} = 2,8$ kΩ, avec un écart-type d'environ 1 Ω.

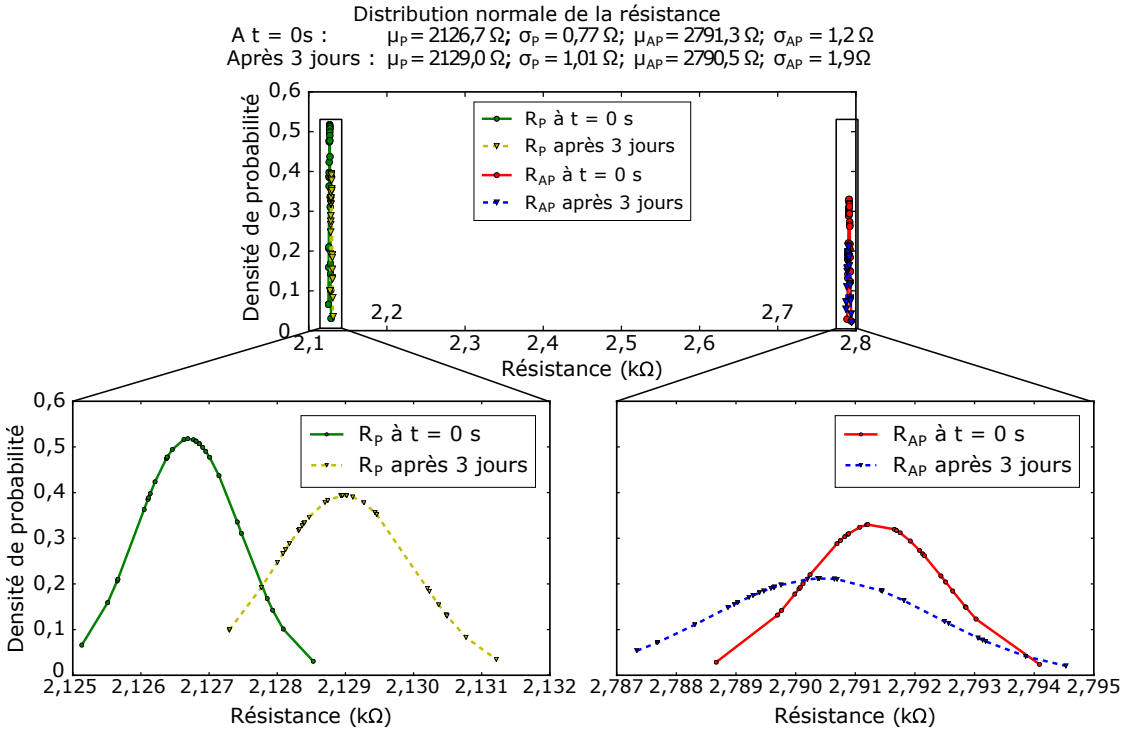


FIGURE 2.9 – Comparaison des valeurs de résistances à un temps t = 0 et la mesure des valeurs de résistance après 3 jours.

Ces vérifications permettent de s'assurer que les résultats qui seront obtenus ultérieurement après l'attaque sur ces jonctions unitaires ne sont pas dus à la précision de la mesure ni à une dérive temporelle. C'est pourquoi ces structures sont testées à un temps t puis la même mesure est réalisée 3 jours plus tard. Comme démontré sur la Figure 2.9, la valeur moyenne des valeurs de résistances peut varier de quelques ohms. Une variation plus importante de la valeur de résistivité du dispositif pré/post attaque sera alors attribuée à l'effet du LASER sur les jonctions.

La Figure 2.10 met en avant des résultats équivalents sur les courants nécessaires pour commuter cette même cellule à un temps t initial et 3 jours après. Comme il peut être noté, une faible variation de l'ordre de quelques microAmpères est affichée entre les deux mesures.

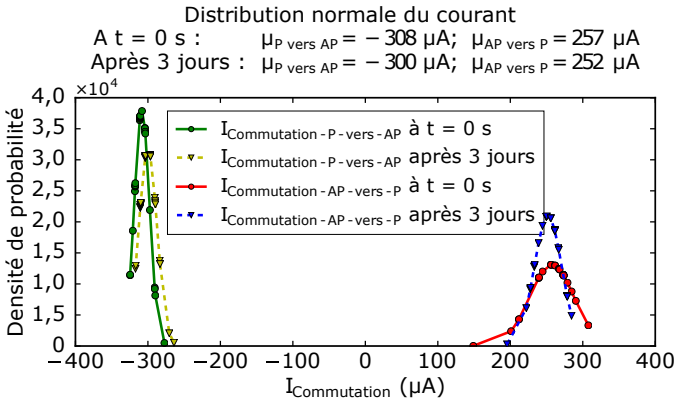


FIGURE 2.10 – Évolution du courant de commutation mesuré sur une jonction unitaire de type STT-MRAM perpendiculaire à un temps t = 0 s et après 3 jours.

Des résultats équivalents ont été obtenus sur tous les échantillons caractérisés et validés (fonctionnels) pour cette expérimentation. En effet, grâce à ces différentes vérifications et caractérisations électriques réalisées avant attaque, toutes les cellules opérationnelles ont été séparées en deux familles, des jonctions écrites vers l'aimantation P et d'autres vers l'état AP. La dernière étape de chaque caractérisation pré-attaque est ainsi d'écrire l'élément mémoire vers l'état souhaité (P ou AP) et de le lire (afin de confirmer sa bonne programmation).

2.1.2 Le banc de caractérisation physique : Laser Nd-YAG

Le LASER Nd-YAG fait partie des lasers solides à cavité cristalline (photographie en Figure 2.11). La matrice hôte est composée du cristal Grenat d'Yttrium-Aluminium ou – *Yttrium Aluminum Garnet* $Y_3Al_5O_{12}$ – (YAG) dopé aux ions Néodyme ou – *Neodymium* Nd^{3+} – (Nd). Il est un outil conçu pour le retrait de matières de natures diverses des circuits. En effet, à sa puissance maximale il peut délivrer une énergie de 500 nJ. Il peut alors être employé pour retirer de la passivation ou d'autre couches supérieures des circuits intégrés (des courts-circuits) par exemple, qu'il peut être nécessaire d'éliminer afin d'assurer une meilleure analyse de fiabilité du composant. Outre cette fonction principale, le LASER Nd-YAG utilisé à une énergie plus faible permet également la caractérisation sécuritaire et les analyses de défaillances des composants semi-conducteurs.

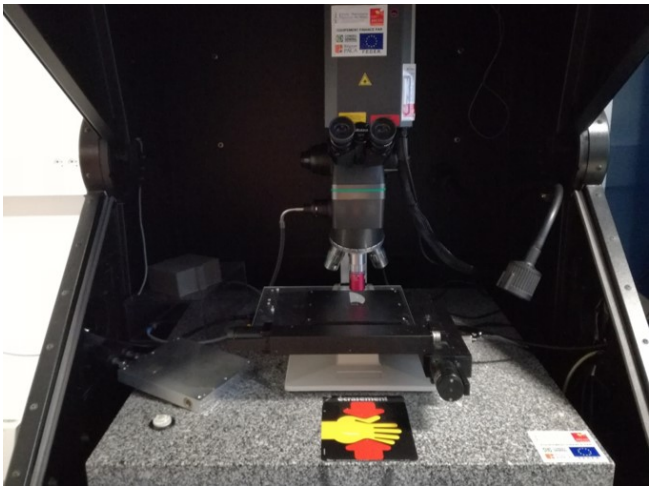


FIGURE 2.11 – Photographie du banc de caractérisation sécuritaire New Wave QuikLaze 50ST2 laser Nd-YAG de la plateforme Micro-Packs.

Le banc de caractérisation utilisé est disponible sur la plateforme Micro-PackS [125], une plateforme collaborative technologique de R&D pour innover dans le monde des objets connectés. Elle est basée au Centre Microélectronique de Provence à Gardanne. Cet équipement de la marque New Wave et de référence QuikLaze 50ST2 [126] dispose de 3 longueurs d'ondes : ultraviolet (λ = 355 nm), vert (λ = 532 nm) et infra-rouge (λ = 1064 nm). Ce banc permet de réaliser des tirs LASERs qui peuvent être uniques ou encore de tirer de façon impulsionnelle à une fréquence maximale de 50 Hz.

D'autre part, ce faisceau laser traverse un obturateur carré dont la projection sur le circuit correspond à une surface de 2,5 mm x 2,5 mm, sans objectif microscopique. Pour le réglage de la zone illuminée, cet obturateur peut être réglé grâce à des pourcentages d'ouvertures allant de 0 % (obturateur complètement fermé et donc aucune transmission du faisceau) à 100 % (obturateur complètement ouvert - transmission maximale du faisceau), avec un pas de 0,1 %. Ce paramètre permet ainsi un réglage précis de de la largeur du faisceau qui peut atteindre la jonction et de calculer la proportion de faisceau qui traverse cet obturateur. Ainsi, plus l'obturateur est fermé, plus le faisceau envoyé sur l'échantillon est fin et localisé. En outre, il est nécessaire de prendre en compte l'atténuation de ce faisceau sur le chemin optique pour déterminer l'énergie réelle

transmise à l'échantillon. En effet, nous pouvons considérer les trois coefficients de transmission t des faisceaux LASERs disponibles sur cet équipement en fonction de la longueur d'onde utilisée tel que précisés dans le Tableau 2.1.

Il est alors possible de déterminer l'énergie totale E_t déposée lors de l'attaque avec l'Équation 2.3. Cette énergie s'exprime en fonction de l'énergie de contrôle E_c envoyée par le laser, l'ouverture O_x en % de l'obturateur en x, l'ouverture O_y en y exprimée en % et le coefficient de transmission optique t du faisceau dans les objectifs (ou chemin optique).

$$E_t = E_c \times O_x \times O_y \times t \tag{2.3}$$

$\lambda(nm)$	355	532	1064
t	0,17	0,35	0,45

TABLEAU 2.1 – Coefficients de transmission optique des objectifs Mitutoyo.

2.1.3 Résultats expérimentaux de la caractérisation sécuritaire de STT-MRAMs

2.1.3.1 Paramètres de l'attaque laser

Après la vérification du fonctionnement des mémoires STT-MRAMs, les échantillons ont été attaqués par l'intermédiaire d'un faisceau LASER infrarouge dont la longueur d'onde λ est égale à 1064 nm. Ce faisceau irradie une surface d'environ 10 µm x 11 µm à la surface de l'échantillon comme illustré Figure 2.12. Cette zone correspond à la surface minimale observable au microscope délimitant la zone de la JTM.

Elle correspond à une ouverture de l'obturateur du laser de 20 % en x et de 22,5 % en y avec un objectif x50. Lorsque l'objectif x50 est enclenché, alors une ouverture de 100 % de l'obturateur correspond à une projection lumineuse d'environ 50 µm x 50 µm sur le circuit contre 2,5 mm x 2,5 mm sans objectif (comme précédemment précisé dans la partie 2.1.2).

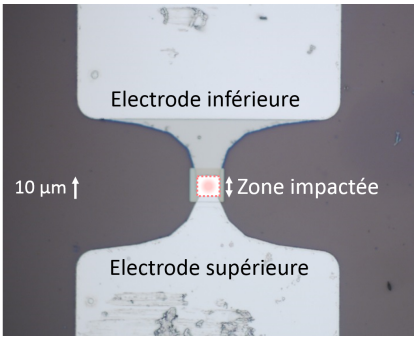


FIGURE 2.12 – Photographie d'un dispositif en mettant en avant la zone visée par le faisceau laser.

En outre, l'injection de fautes par LASER est réalisée à différents niveaux d'énergies afin de déterminer la sensibilité des jonctions par rapport au niveau d'énergie fourni. Pour ce faire, un faisceau laser d'une impulsion de 5 ns est envoyé sur l'échantillon par cycle de test : caractérisation électrique - attaque - caractérisation électrique.

2.1.3.2 Intégrité des jonctions après attaque

Pour rappel, lors de ces manipulations, deux groupes de mémoires sont distingués, des mémoires qui après conditionnement (ou cyclage) sont écrites dans un état AP et d'autres dans un état P.

Aucune modification n'a été observée sur les jonctions qui étaient dans un état initial pré-attaque P. La résistance et la TMR de la cellule sont restées identiques lors de la lecture post-attaque.

Toutefois, pour des jonctions qui étaient initialement dans l'état AP (avant l'attaque LASER), des modifications sont survenues sur la mémoire. Le protocole expérimental décrit précédemment a été suivi. Comme il peut être noté en Figure 2.13, un cyclage ou conditionnement initial de 30 cycles est réalisé avant l'attaque (agrandi dans les deux encarts insérés à gauche dans cette figure) afin de vérifier la dispersion des deux valeurs de résistances R_P et R_{AP}. Après ce conditionnement, la seconde étape consiste à lire la cellule afin de vérifier que son dernier état est l'état AP. La jonction est alors attaquée par le faisceau laser infrarouge, avec une énergie suffisante. Toutefois, la lecture post-attaque de cette structure démontre la modification de l'état de la mémoire d'un état AP (état logique '1') à un état P (état logique '0'). Ainsi, cette attaque a induit la modification de l'information stockée dans la mémoire. Le modèle de faute réalisé avec cette injection LASER est un *bit-flip* (inversion du bit stocké) de type *bit-reset*, comme illustré sur la Figure 2.13 pour l'une des cellules testées. La dernière étape de cette expérimentation est la confirmation de l'invariance des distributions des deux états résistifs post-attaque, distributions agrandies dans les deux encarts de droite de la Figure 2.13.

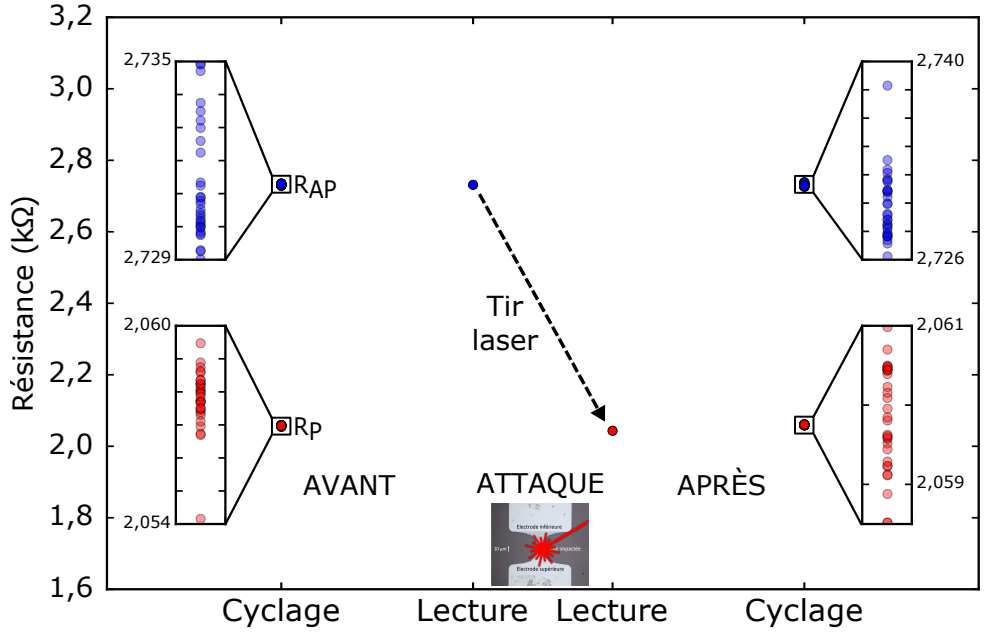


FIGURE 2.13 – Évolution des valeurs de résistances avant/après attaque laser avec une énergie de 530 nJ et cyclage de la cellule avant/après attaque laser.

2.1.3.3 Seuil d'énergie pour la détérioration de l'intégrité de jonctions STT-MRAM

La confirmation de cette commutation de l'état mémoire de la jonction suite à une attaque laser nous a conduit alors à la question : Pouvons-nous déterminer de façon fiable la quantité d'énergie minimale nécessaire à cette commutation?

Afin de répondre à cette interrogation, 36 cellules unitaires fonctionnelles présentant des valeurs de TMR suffisantes (autour de 35 %) ont été choisies afin de déterminer la valeur de ce seuil. Pour cela, une injection de fautes LASER a été réalisée sur chacune de ces cellules à différents niveaux d'énergies, afin de déterminer à partir de quel niveau elles commutent de l'aimantation AP vers l'aimantation P.

Avant chaque injection LASER, chaque jonction est lue afin de vérifier qu'elle est dans un état AP puis l'attaque est réalisée, suivie d'une lecture du nouvel état. Quel que soit le nouvel état de la JTM, un cycle de lectures/écritures des deux états (P et AP) est réalisé afin de s'assurer de la non-détérioration des deux résistances de la mémoire. La jonction est alors laissée dans un état AP afin de réaliser l'attaque suivante, à un niveau d'énergie différent.

Une énergie de seuil pour laquelle 100 % des cellules ont inversé leur résistivité, pour passer d'un état AP vers un état P, est déterminée. C'est ainsi que le seuil de 547 nJ a pu être établi, comme représenté Figure 2.14. Ce seuil est déterminé en utilisant la formule 2.3 où l'énergie de commande du laser correspond à 27 µJ, avec O_x = 0,2, O_y = 0,225 et en considérant l'atténuation du chemin optique de 0,45 pour un faisceau infra-rouge.

Il est également intéressant de noter que plus de 80 % (28/36) des cellules commutent à partir de 509 nJ alors qu'elles n'étaient que 20 % (7/36) à commuter à une énergie de 475 nJ. Ainsi, sur ce faible intervalle d'énergie à fournir, la majorité des cellules ont commuté d'un état AP initial vers l'état P. À partir de 516 nJ, 90 % (32/36) des jonctions ont commuté. Le seuil de 547 nJ qui a été établi dans la cadre de ce travail est le seuil pour lequel les 36 cellules testées inversent leur aimantation. Toutefois, il peut être noté que selon les dispositifs testés et leurs propriétés intrinsèques, ce seuil pourra légèrement varier.

D'autre part, lors de ces expérimentations nous avons également pu confirmer qu'il n'existe pas d'effet cumulatif induit par cet effet LASER. Pour cela différentes attaques de niveaux d'énergie inférieurs au niveau de commutation établi (bien que leur somme soit supérieure à ce seuil) ne modifient pas l'état de la jonction et la lecture confirme le maintien de l'état AP.

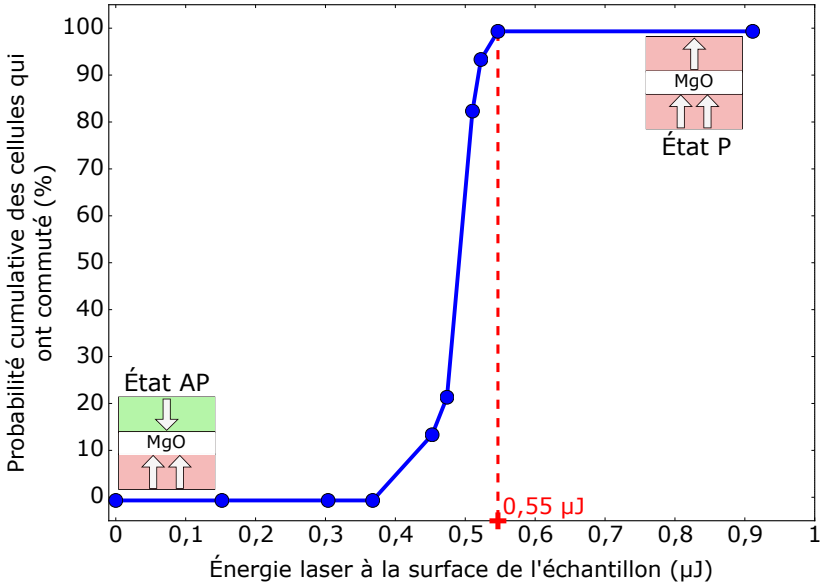


FIGURE 2.14 – Probabilité cumulative de la commutation de cellules STT-JTM après une attaque laser infra-rouge (mesure réalisée sur 36 cellules mémoires).

2.1.4 Fiabilité du point mémoire STT-MRAM post-attaque

Suite à l'attaque, le fonctionnement électrique des jonctions est revérifié. Pour cela, un nouveau conditionnement/cyclage de 30 cycles est réalisé en mesurant les paramètres électriques : R, $I_{Commutation}$, $V_{Commutation}$. Cette caractérisation électrique a conduit à des résultats équivalents (uniquement une faible variation due à la mesure est induite - de quelques ohms pour la résistance et de quelques microampères pour le courant de commutation -) à ceux obtenus lors de la phase de caractérisation électrique pré-attaque.

La Figure 2.15.a illustre un cycle de fonctionnement I(V) d'une cellule avant et après le tir LASER. Comme il peut être noté sur cette Figure, aucune variation ou détérioration n'a été induite par le faisceau LASER. La jonction réagit de façon équivalente avant et après attaque à la tension qui lui est envoyée en entrée sur ses deux électrodes.

D'autre part, cette figure met également en avant les distributions des résistivités R_P et R_{AP} et des courants nécessaires à la commutation de la cellule, comme illustrés Figure 2.15.b et Figure 2.15.c. Ces trois caractérisations électriques de l'opération de l'élément mémoire STT-MRAM post-attaque démontrent l'invariance du fonctionnement de la cellule pré/post-attaque.

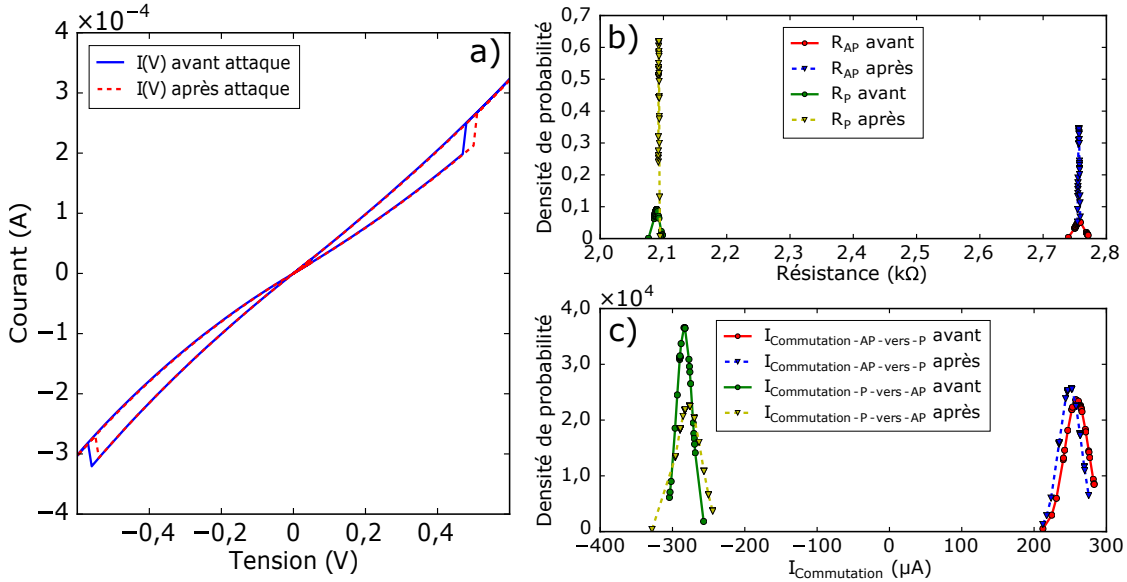


FIGURE 2.15 – Caractérisation électrique pré/post attaque d'une JTM. a) Représentation du courant en fonction de la tension avant et après attaque laser. b) Distribution des résistances avant et après attaque laser. c) Distribution des courants de commutation de la cellule avant et après attaque.

2.2 Discussion sur les phénomènes physiques induisant la commutation des cellules STT-MRAMs

Dans cette section, les phénomènes physiques qui ont conduit à l'inversion de la donnée stockée dans la jonction, de l'état logique '1' vers l'état '0' sont discutés. Pour cela, l'effet du tir LASER sur la JTM est expliqué, dans un premier temps au niveau d'abstraction circuit puis au niveau atomique.

2.2.1 Commutation de la JTM au niveau circuit

Le LASER Nd-YAG transmet de la chaleur [127] à la structure JTM. Les photons contenus dans le faisceau LASER conduisent à un accroissement de la température dans la JTM.

De plus, comme démontré dans [20] le champ magnétique coercitif H_c nécessaire à faire commuter la jonction d'un état à l'autre décroit lorsque la température de l'expérimentation augmente. En effet, H_c tend vers 0 Œ avec l'augmentation de la température, facilitant ainsi la commutation de la cellule d'un état vers l'autre, comme illustré Figure 2.16.

Ainsi, le faisceau LASER infrarouge induit par le LASER Nd-YAG conduit à l'accroissement de la température dans la JTM, à l'abaissement du champ magnétique coercitif de la structure et donc à la commutation de la structure. Cette inversion toujours dans le même sens, de l'aimantation des couches magnétiques de l'état AP vers l'orientation magnétique P est développée dans la partie suivante Section 2.2.2.

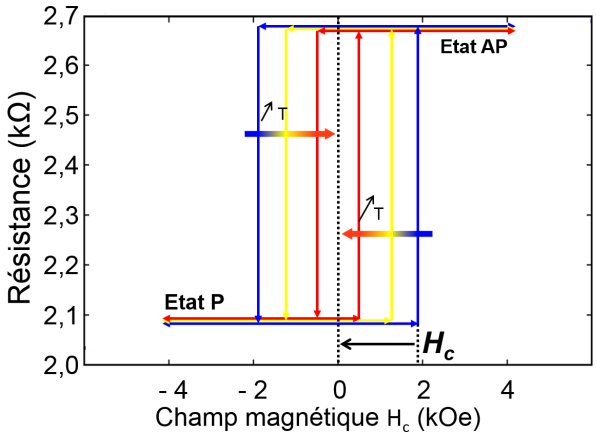


FIGURE 2.16 – Illustration de l'influence de la température sur le champ magnétique coercitif de la JTM.

2.2.2 Commutation de la JTM au niveau spins

Lorsque le faisceau LASER atteint la JTM, le réchauffement induit par la perturbation conduit à l'agitation et au déplacement, dans toutes les directions, des spins des électrons présents dans les couches ferromagnétiques. Nous mettons alors en avant deux hypothèses pour expliquer le phénomène induit par le tir LASER.

D'une part, considérons dans un premier temps le scénario où seule la couche de stockage perd son aimantation intrinsèque et devient paramagnétique le temps du réchauffement, comme illustré Figure 2.17.a.

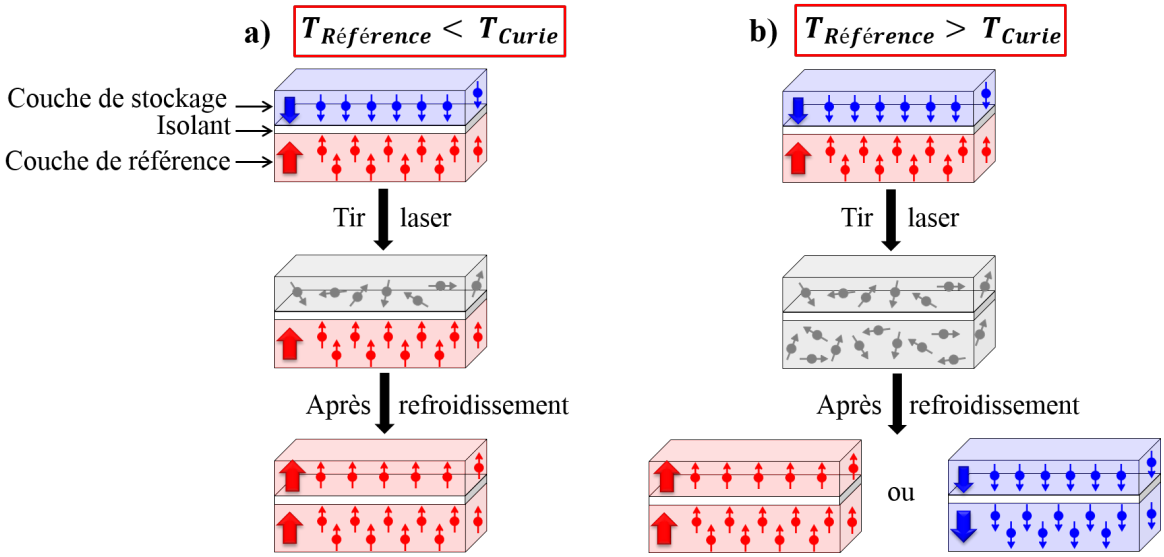


FIGURE 2.17 – Illustration de l'effet thermique induit sur les spins des deux couches ferromagnétiques de la JTM.

Dans ce cas, la température atteinte dans la couche de référence ne serait pas suffisante pour dépasser la température de Curie T_{Curie} du matériau ferromagnétique composant la JTM. T_{Curie} correspond à la température pour laquelle ce matériau perd son aimantation permanente le temps du réchauffement, en devenant un matériau paramagnétique. Lorsque la température diminue dans les couches ferromagnétiques, le champ magnétique terrestre couplé à la couche de référence, qui serait moins sensible à cette augmentation de température (dans ce scénario), modifient l'orientation de cette couche de stockage. Ainsi, les deux couches ferromagnétiques com-

posant la JTM s'orientent dans le même sens magnétique (état P) car l'énergie à fournir par les atomes est moins importante que pour l'état AP.

D'autre part, le second scénario consiste à avancer l'hypothèse que les deux couches ferromagnétiques composant la JTM peuvent être sensibles de la même manière au réchauffement qui les convertira en paramagnétiques, comme illustré Figure 2.17.b. Ce cas de figure illustre le scénario où la température de la couche de référence est supérieure à la température de Curie du matériau. Dans ce cas, après le refroidissement de la structure les spins s'orientent tous dans un état parallèle compte tenu que cela correspond à l'état de plus faible énergie. Ainsi, dans ce cas, l'état final de la structure peut être soit avec les spins des deux couches orientés vers le haut soit vers le bas, comme illustré sur la Figure 2.17.

3 Modèle thermique COMSOL de l'attaque LASER

L'hypothèse d'augmentation de la température dans les jonctions magnétiques a été confirmée par des simulations multi-physiques qui consistent à analyser le comportement d'un système dans différents domaines : thermique, électromagnétique, mécanique, etc. Dans le cadre de cette étude, une simulation des températures atteintes dans la jonction a été réalisée. En effet, comme démontré par D. Sands dans [127], le faisceau laser envoyé à la surface de l'échantillon est absorbé par les couches supérieures et se propage de couche en couche selon la direction verticale de la structure. Cette intensité transmise aux couches inférieures décroit de façon exponentielle. L'Équation 2.4 décrit ainsi la propagation de la radiation optique selon l'axe z.

$$S(z) = S(0)\exp(-\alpha z) = \alpha I_0(1-R)\exp(-\alpha z) \tag{2.4}$$

En considérant $S(0)$ l'absorption de la surface de l'échantillon tel que : α le coefficient d'absorption optique du matériau de surface, I_0 l'intensité initiale du faisceau laser avant d'atteindre la structure sous test, exprimée en $W.m^{-2}$ et R la réflectivité du matériau de surface (qui peut être calculée grâce à l'indice de réfraction du matériau).

3.1 Le transfert de chaleur dans les solides

Dans cette modélisation numérique, le LASER est considéré comme une boite noire dont l'énergie est répartie à la surface de l'échantillon en suivant un profil gaussien comme illustré Figure 2.18. Le centre de ce faisceau laser correspond au centre de l'élément mémoire magnétique.

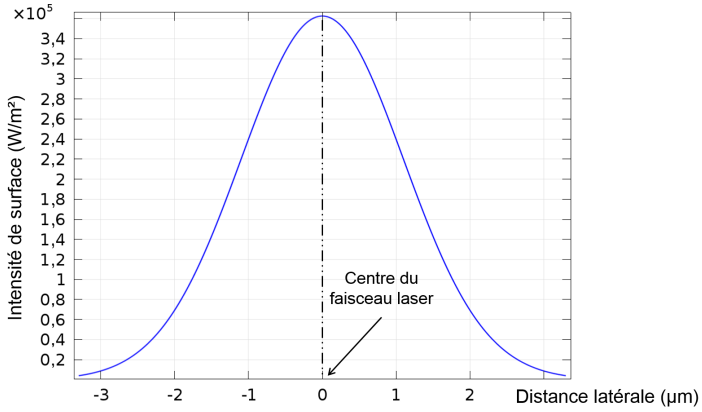


FIGURE 2.18 – Modélisation du faisceau Gaussien issu du laser Nd :YAG.

L'énergie transmise à la surface de l'échantillon se dissipe ensuite sous forme de chaleur, permettant de réchauffer cet empilement depuis la couche supérieure jusqu'à atteindre les couches

composant la JTM. Pour cela, il est nécessaire de déterminer l'énergie déposée en surface de l'échantillon. Cette intensité transmise est notée I_T. Elle est développée par l'Équation 2.5 et représentée sur la Figure 2.18.

$$I_T = I_0(1 - R) \tag{2.5}$$

Une partie de cette intensité est absorbée par la surface de la structure et s'exprime telle que définie en Équation 2.6.

$$S(0) = \alpha I_T \tag{2.6}$$

Cette énergie est ensuite transférée dans la structure sous forme de chaleur en suivant la loi empirique de Fourier, définie par l'Équation 2.7. La diffusion thermique de la chaleur Q dans l'empilement est alors exprimée en fonction du coefficient de conductivité thermique du matériau k et de la température T.

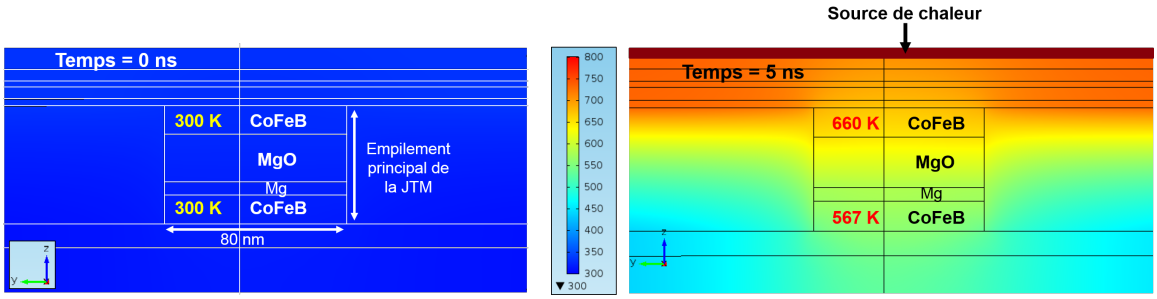
$$\frac{dQ}{dt} = -k\nabla T \tag{2.7}$$

Dans le contexte d'une impulsion LASER, la quantité de chaleur déposée par le faisceau optique à la surface de l'échantillon peut être modélisée en considérant cette diffusion de chaleur par l'énergie qu'elle apporte à la structure. Le changement de température dans la structure est alors équivalent à une modification de l'enthalpie H de l'empilement. En effet, ce paramètre thermodynamique est principalement intéressant lorsque la pression de l'expérimentation et de l'échantillon est constante. La variation d'enthalpie ΔH est alors exprimée en fonction de la masse m du matériau, de sa capacité calorifique c_p et de la variation de température ΔT observée dans le composant. Cette variation d'enthalpie est définie par l'Équation 2.8 et a également été modélisée par A. Krakovinsky dans le cadre du réchauffement de mémoires OxRAMs [128].

$$\Delta H = m \,.\, c_p \,.\, \Delta T \tag{2.8}$$

La Figure 2.19 présente les résultats de cette modélisation réalisée via l'outil COMSOL [129]. Comme défini par la Figure 2.19.a, les deux couches ferromagnétiques sont initialement considérées à température ambiante (à 300 Kelvins chacune). Ensuite, l'effet du tir LASER sur le réchauffement de l'architecture est simulé.

Ainsi, il peut être noté que la première couche de CoFeB (la couche libre) est à une température moyenne de 660 K alors que la seconde (la couche de référence) est à une température moyenne de 567 K, après 5 ns de simulation, comme illustré Figure 2.19.b. Cette augmentation de température dans la JTM conduit à l'abaissement de la coercivité magnétique [20] et au passage d'un état AP à un état P de la JTM, selon le procédé décrit en Section 2.2.2.



(a) Température au temps t = 0 ns.

(b) Température au temps t = 5 ns.

FIGURE 2.19 – Modélisations multi-physiques de l'évolution de la température dans l'empilement des jonctions unitaires.

3.2 Proposition de contre-mesures

Afin de limiter l'effet que peut avoir un tir LASER sur le basculement de l'état logique stocké dans une STT-MRAM (donc la commutation d'un état AP vers un état P), il est possible de réduire la température atteinte dans la jonction (Figure 2.19.b), en intégrant la jonction au plus proche du *Front-End of Line* – (FEoL), pour bénéficier de l'effet bouclier des couches d'interconnexion métalliques supérieures.

En effet, comme l'illustre la Figure 2.20, si par exemple l'épaisseur de la couche supérieure de Manganèse-Iridium (MnIr) de la structure est doublée (en l'augmentant de 150 nm à 300 nm), alors la température diminue drastiquement dans les deux couches ferromagnétiques au bout du même temps de simulation du tir LASER (au bout de 5 ns). Cette couche de MnIr correspond à la couche supérieure de l'empilement des jonctions qui nous ont été fournies et qui ont été testées dans le cadre de l'étude de l'intégrité de la technologie mémoire STT-MRAM.

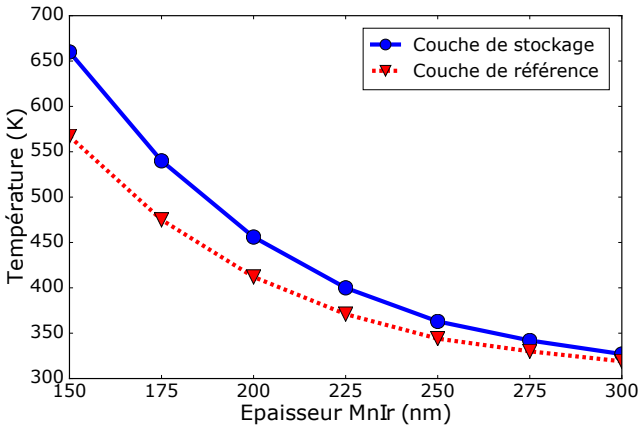


FIGURE 2.20 – Évolution de la température à t = 5 ns en fonction de l'épaisseur de MnIr.

Ainsi, l'augmentation de l'épaisseur de cette couche conduit à la diminution de l'intensité transmise (dû à la décroissance de façon exponentielle en suivant l'axe des z de l'intensité transmise), une énergie plus réduite atteint la JTM signifiant donc une faible variation d'enthalpie de la structure (non-modification de la température des couches ferromagnétiques composant l'empilement).

D'où l'obtention des résultats représentés sur la Figure 2.21. La température diminue dans la couche libre de CoFeB de la JTM pour tendre vers 327 K au lieu des 660 K obtenus pour la même structure et modélisée sur la Figure 2.19.b.

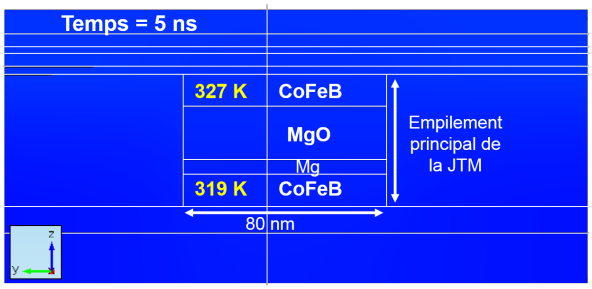


FIGURE 2.21 – Modélisation multi-physique de l'évolution de la température au temps t = 5 ns pour e_{MnIr} = 300 nm.

Toutefois, l'attaquant en réponse à cette contremesure peut augmenter le niveau d'énergie du tir LASER afin de fournir plus d'énergie et d'intensité à la JTM. Cependant, cette solution conduit plus rapidement à la détérioration des niveaux de métallisation supérieurs qui seront alors chauf-

fés de façon trop importante jusqu'à leur destruction. Ainsi, cela diminue fortement sa capacité à attaquer un élément mémoire précis afin d'en modifier la donnée.

4 Conclusion et perspectives

La technologie mémoire STT-MRAM présente des intérêts indéniables pour les objets connectés dans le paysage des mémoires émergentes grâce à sa vitesse de fonctionnement aussi bien en lecture qu'en écriture, sa forte densité et sa non-volatilité. Tous ces points en font une candidate sérieuse pour un remplacement de la technologie mémoire Flash dans les applications embarquées utilisées au quotidien.

Toutefois, une sécurité accrue de cette mémoire face aux attaques de différentes natures doit être assurée. Dans le cadre de ce chapitre, des attaques par injection de fautes via laser ont été réalisées sur des éléments mémoires STT-MRAMs unitaires d'aimantation perpendiculaire. Il a pu être démontré que lorsque l'énergie fournie par le faisceau LASER est suffisante, seuil déterminé égal à 547 nJ, alors toutes les cellules se retrouvent après attaque laser dans une aimantation parallèle. Donc cette attaque est d'autant plus dangereuse pour des cellules initialement dans l'aimantation anti-parallèle AP où elle induit un *bit-flip* ou *bit-reset* de l'état de la donnée stockée dans la JTM.

Cette commutation de l'état de la cellule a été simulée par des modélisations thermiques qui ont montré que l'inversion de l'orientation des couches ferromagnétiques est due à une augmentation de la température au sein de l'empilement, principalement des couches ferromagnétiques. Des simulations ont également montré l'avantage que peut avoir le rapprochement de cette technologie mémoire du FEoL. En effet, dans ce cas, la température atteinte dans les couches ferromagnétiques de la cellule peut être drastiquement réduite et ainsi éviter toute commutation inappropriée, due à une trop faible énergie atteignant la jonction.

En outre, cette contremesure obligerait l'attaquant à augmenter l'énergie du tir LASER. Ce qui lui serait désavantageux compte tenu qu'il atteindra plus rapidement les températures de fusion des niveaux de métallisation supérieurs et donc rendra l'attaque inutile puisque la cellule sera détruite sans accès à la donnée.

Outre l'augmentation de l'épaisseur de la couche supérieure composant l'empilement, il est également possible d'étudier l'ajout de couches supplémentaires basées sur d'autres matériaux compatibles avec le procédé magnétique tels que le Platine Pt, le Tungstène W ou encore le Tantale Ta. Ces derniers pourraient ainsi être intégrés au-dessus de l'empilement magnétique afin d'en assurer une meilleure stabilité face aux attaques de différentes natures. Il serait ainsi fortement intéressant de réaliser des injections de fautes sur des structures qui présenteraient cette contremesure technologique et ainsi de déterminer, le niveau de protection de cette contremesure et de la stabilité des paramètres intrinsèques pour le fonctionnement de la STT-MRAM.

Chapitre 3

Détecteur d'attaques thermiques et photoélectriques - DDHP

> *Le plus important étant de toujours y croire et de ne jamais abandonner*
>
> **— Stephen Hawking**

Dans le cadre de la protection des circuits intégrés face aux injections de fautes malicieuses, qu'elles soient de type LASER ou électromagnétique, de nombreux capteurs sont proposés. Le rôle de ces capteurs est de détecter toute faute ou toute anomalie qui peuvent être induites dans les circuits attaqués. Ce chapitre présentera dans un premier temps les différentes contre-mesures existantes pour la protection des circuits intégrés face aux attaques par perturbation et par observation. Puis, la solution proposée de détection d'illuminations et de perturbations thermiques intitulée double détection d'attaques induites par chauffage et effet photoélectrique ou – *Dual Detection of Heating and Photocurrent attacks* – (DDHP) sera présentée. Le fonctionnement et les performances de ce détecteur seront alors discutés dans ce chapitre.

Ces travaux ont fait l'objet d'une publication à la conférence internationale "IOLTS" en 2019 [130]

Sommaire

1 Stratégies de durcissement des circuits intégrés

La communauté radiative fut la première à utiliser le LASER pour émuler les *Single Event Latch-Up* – (SEL) et *Single Event Effect* – (SEE) qui sont introduits dans les circuits intégrés par des particules ionisantes. Ces phénomènes peuvent induire des effets de transistors bipolaires parasites dans la structure [131]. Depuis, le LASER est également utilisé pour générer des fautes dans les circuits de manière localisée (dans le temps et l'espace) [82], afin d'extraire des données confidentielles. Ainsi, nous décrirons les différentes solutions possibles pour améliorer la sécurité des circuits intégrés face aux injections de fautes.

Toutefois, outre les injections de fautes, les circuits intégrés peuvent également être sensibles aux attaques par observation et principalement aux attaques par canaux cachés. Ces dernières permettent de corréler les données traitées lors des mesures réelles au message secret. C'est pourquoi nous détaillerons les techniques communément utilisées pour protéger les composants face à ce type d'attaques.

1.1 Protections contre les attaques par canaux cachés

Comme développé dans le Chapitre 1, les attaques par canaux cachés (DPA et CPA par exemple) sont des attaques qui permettent de déterminer grâce à la consommation en courant d'un circuit (ou ses émanations électromagnétiques ou encore le temps nécessaire aux commutations) sa corrélation aux opérations et données exécutées. Afin de réduire la sensibilité des circuits intégrés face aux attaques par canaux cachés, il est nécessaire de réduire cette dépendance entre la consommation du circuit et les données manipulées. Pour cela, différentes solutions sont envisagées, elles sont principalement basées sur la dissimulation ou le masquage des opérations non-linéaires.

1.1.1 Principe de dissimulation

Cette technique utilisée dans le développement de contre-mesures vise comme son nom l'indique la dissimulation de la donnée qui est en cours de traitement, grâce à la modification du rapport signal sur bruit du système en cours d'analyse. En effet, cette technique vise à rendre l'attaque plus difficile, soit en réduisant l'intensité ou l'amplitude du signal mesuré, soit en augmentant le bruit du circuit grâce par exemple au générateur de bruit électromagnétique ou encore générateur aléatoire d'aléas.

— Stabilisation de la consommation des transitions '0' vers '1' et '1' vers '0' :
En effet, le redressement des émanations électromagnétiques et de la consommation du circuit pour que celle-ci reste stable quelle que soit la donnée manipulée, conduit à l'augmentation de la protection du circuit face aux attaques par canaux cachés. Pour ce faire, il est nécessaire que chaque transition de la donnée logique '0' vers '1' et '1' vers '0' soit équivalente en terme de spectre de consommation et d'émanations. La logique double rail à pré-charge [132], [133], [134] (ou encore triple rail [135]) a été introduite dans cette perspective. Cette méthode consiste pour le circuit à toujours manipuler une donnée et son inverse, impliquant donc une impossibilité pour l'attaquant de savoir quelle donnée est impliquée dans le calcul du texte chiffré par exemple.

— Augmentation du niveau de bruit :
La seconde solution pour la dissimulation des données est l'augmentation du bruit dans le système cryptographique, afin qu'il puisse être impossible à l'attaquant de distinguer le bruit du signal [136], et donc de retrouver les données. Afin de réaliser cette technique de dissimulation, l'une des méthodes pourrait être la parallélisation des opérations (beaucoup de transitions de '0' vers '1' et de '1' vers '0' qui se produisent simultanément), afin que le

calcul de corrélation soit erroné. L'augmentation du bruit du circuit peut également être envisagée du point de vue temporel [137], en multipliant par exemple le nombre d'horloges internes, complexifiant la synchronisation du système par l'attaquant ou encore en rajoutant des instructions aléatoires lors de l'exécution de l'algorithme, bien que celles-ci ne rentrent pas en compte dans le calcul du texte chiffré final. Toutefois, cette offuscation de la donnée est réalisée au prix de l'augmentation de la consommation du circuit.

1.1.2 Principe de masquage

Le masquage [138] consiste à affecter un ou plusieurs masques aléatoires sur les données. Le nombre de masques utilisés fixe l'ordre du masquage. Le système cryptographique combinera alors les valeurs affectées par ce masque pour les compiler avec le message clair ou la clé de chiffrement bit à bit, modifiant ainsi le résultat du chiffrement. Lors de chaque exécution de l'algorithme, le masque doit être calculé indépendamment des valeurs des autres masques. Les valeurs intermédiaires calculées par les chiffrements ne dépendent donc pas directement des valeurs sensibles.

— Masquage logique ou booléen : cette catégorie de masquage consiste à réaliser une opération de "XOR" bit à bit entre le masque généré aléatoirement et le message clair ou la clé de chiffrement. Il est alors possible de citer l'exemple des vecteurs d'initialisations utilisés dans les chiffrements par flots.

— Masquage arithmétique : cette technique repose sur la réalisation d'une fonction d'addition sur plusieurs bits entre la composante à masquer et le masque.

Cette étape de masquage est ensuite retirée au texte chiffré final afin d'obtenir le message chiffré correspondant au message clair et à la clé utilisés. Toutefois, il reste possible d'attaquer des algorithmes de cryptographie présentant cette contre-mesure en augmentant l'ordre de l'attaque. En effet, pour un masquage d'ordre n, l'attaque doit être d'ordre supérieur ou égal à n [139].

1.2 Protections contre les injections de fautes

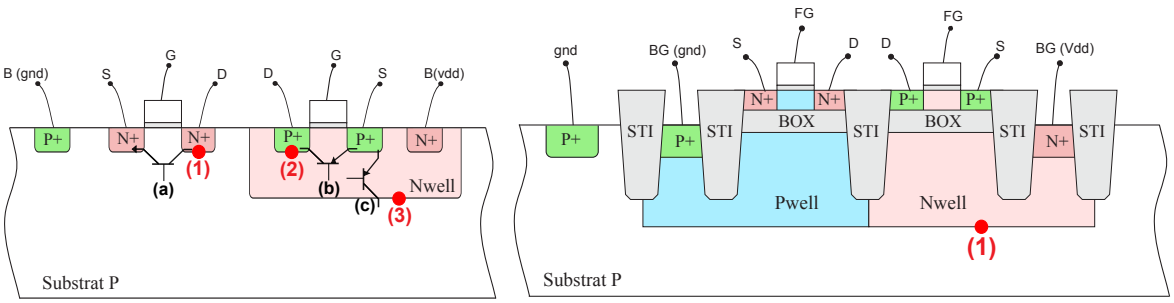
Afin de protéger les circuits contre différents types d'injections de fautes, il est nécessaire de réaliser une étude complète des risques qu'encourt le dispositif. Selon la sensibilité de l'application, des solutions pourraient alors être proposées en prenant en compte le coût de la solution en terme de surface et d'énergie par rapport à ses apports sécuritaires. Dans cette première section, nous développerons différentes solutions technologiques et architecturales visant à réduire la sensibilité des circuits intégrés face aux injections de fautes.

1.2.1 Diminution de la sensibilité du circuit face aux attaques par perturbation

1.2.1.1 Durcissement technologique :

Bien que la technologie CMOS *Fully Depleted Silicon On Insulator* – (FD-SOI) ait été initialement développée dans une optique de réduction des courants de fuites des composants CMOS bulk, celle-ci s'est vu attribuer par la suite l'ambition d'être l'atout technologique qui permet de réduire la sensibilité des circuits intégrés face aux injections de fautes.

Comme représenté sur la Figure 3.1, les zones PN sensibles aux attaques sont réduites à une pour la technologie FD-SOI contre trois pour la technologie bulk CMOS. Ces zones sont notées (1), (2) et (3) sur la Figure 3.1.a et (1) sur la Figure 3.1.b. En effet, le volume de silicium où le faisceau LASER peut générer des paires électrons-trous est drastiquement réduit sur la technologie FD-SOI grâce à la couche de diélectrique enterrée sur la face arrière du composant. Ainsi, dans le cadre du durcissement du substrat, la technologie CMOS FD-SOI réduit la sensibilité du composant aux injections de fautes comparée à une technologie CMOS standard de type Bulk [140].



(a) La technologie CMOS bulk.

(b) La technologie CMOS FD-SOI.

FIGURE 3.1 – Représentation des zones sensibles des technologies CMOS et FD-SOI.

1.2.1.2 Développement de boucliers :

L'intégration de boucliers, ou *shields*, sur le circuit permet également de se prémunir d'attaques physiques. Pour cela, une protection métallique est ajoutée sur les couches supérieures du circuit, prévenant toutes les intrusions invasives telles que le micro-sondage, comme illustré Figure 3.2. Ces couches sont constituées de lignes métalliques enchevêtrées. Ces boucliers peuvent être actifs [141] ou passifs [142]. Ces boucliers couvrant le circuit présentent des propriétés et des avantages non-négligeables comme la dissipation de chaleur qui peut être induite suite à une attaque ou encore la réflexion d'un faisceau laser lors de l'injection de fautes.

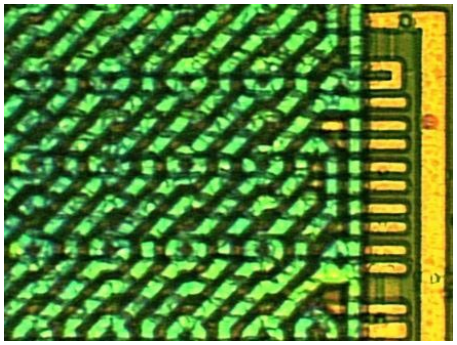


FIGURE 3.2 – Vue au microscope d'un bouclier métallique (extraite de [143]).

1.2.2 Détection de perturbations physiques et de fautes

La détection d'une attaque par perturbation peut être réalisée en détectant, soit la perturbation physique qui conduira à une faute (une surtension par exemple), soit l'erreur générée par cette perturbation. Cette seconde famille de détecteurs repose principalement sur de la redondance.

1.2.2.1 Redondance Double (DMR) :

La redondance spatiale est l'une des méthodes de sécurisation des circuits intégrés des plus étudiées comme par exemple dans [144], [145], [146] ou encore [147]. Elle consiste à dupliquer un composant maître sensible et de réaliser une comparaison en sortie des deux fonctions identiques après chaque opération [148]. Cette technique appelée redondance double ou – *Double Modular Redundancy* – (DMR) durcit les architectures et informe de toutes les modifications susceptibles d'atteindre l'une des composantes visées. Cette méthodologie est conceptuellement la plus simple puisque la seule étape nécessaire est la duplication de la fonction désignée comme sensible vis-à-vis d'une attaque en faute. Toutefois, cette solution augmente considérablement la surface et la consommation, quasiment d'un facteur deux par rapport au circuit de référence.

Considérons par exemple les tables de substitution (ou *SBox*) comme les zones sensibles d'un circuit donné, alors la duplication est réalisée sur ce même composant, comme conceptuellement illustrée Figure 3.3. Les sorties de ces deux blocs (initialement identiques et qui devraient diffuser la même donnée sur les deux branches) sont alors transmises à une porte logique de type "XOR" qui déterminera si les données traitées par les deux fonctions sont bien identiques ou distinctes. Tant que la sortie "Err" de cette porte logique reste à '0', aucune erreur n'est induite dans les *SBox*. Toutefois, dès le basculement de cette valeur à '1', il est possible que l'un des chemins a été contrefait ou modifié, comme dans le cas d'une attaque extérieure visant à retrouver le message confidentiel du chiffrement. Cette méthode DMR permet de détecter une erreur mais non de la corriger.

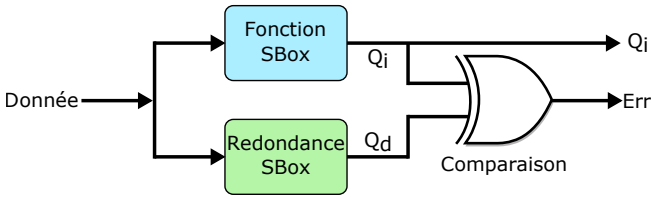


FIGURE 3.3 – Modalités d'implémentation d'une solution de type DMR.

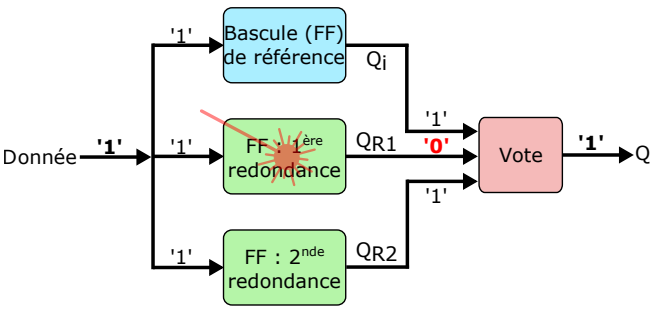
1.2.2.2 Redondance Triple (TMR) :

Cette technique de détection d'attaques (DMR) peut être enrichie pour devenir une technique de protection de circuits intégrés (détection et correction d'erreurs). En effet, elle est basée sur le principe de triplement des composants sensibles permettant au-delà de la détection, la correction de toute erreur qui surviendrait sur l'un de ces chemins [149]. Ainsi, puisque trois chemins équivalents sont comparés, comme illustrés sur la Figure 3.4.a, l'erreur est révélée et corrigée en procédant à un vote majoritaire sur la donnée. Ce vote est réalisé en diffusant la donnée qui se trouve sur une majorité de chemins (au moins 2/3), comme représenté sur la Figure 3.4.b.

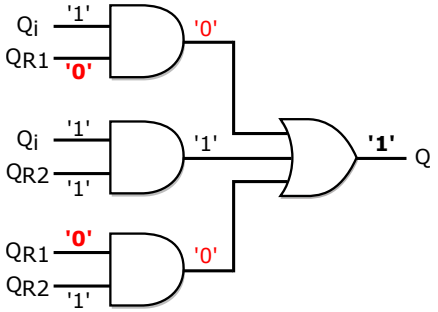
Pour exemple, prenons le cas où une attaque vise la 1ère redondance du circuit initial, c'est-à-dire le 2nd chemin d'attaque pour modifier la sortie saine Q_{R1} de '1' vers '0'.

Lors du vote représenté sur la Figure 3.4.b soit : $Q_i = Q_{R2}$ = '1' et Q_{R1} = '0'. Alors, les sorties des trois portes logiques "ET" sont respectivement, '0', '1' et '0'. La sortie Q de la porte "OU" conduit la valeur logique '1', qui correspond à la valeur saine initialisée dans le circuit (Donnée = '1').

De façon analogue, si la valeur à transmettre est '0' avec au moins deux chemins qui diffusent cette donnée, alors les trois portes logiques "ET" conduisent la valeur '0' qui est transmise au nœud Q par la porte logique "OU". Cette méthodologie de correction d'erreur est appelée Redondance Triple ou – *Triple Modular Redundancy* – (TMR) [150].



(a) Représentation de l'implémentation d'une solution de type TMR.



(b) Modélisation du circuit de vote majoritaire.

FIGURE 3.4 – Modalités d'implémentation d'une solution de type TMR.

Cette technique de correction d'erreurs est efficace car il est fortement improbable qu'une faute soit induite, en même temps, sur deux chemins distincts. Toutefois, l'inconvénient majeur de cette méthode est le triplement (voire plus avec les circuits de vote) de la surface silicium nécessaire pour réaliser l'implémentation de l'opération sensible, et donc une augmentation non-négligeable de la consommation et du coût.

1.2.2.3 Redondance Temporelle :

De façon équivalente à la technique de duplication précédemment décrite, la redondance temporelle [151] repose sur la duplication du circuit sensible pour créer deux chemins pour la transmission de la donnée. Toutefois, des retards sont intégrés sur le chemin dupliqué. Ces derniers ne modifient l'état que de l'une des bascules au front-montant et conduisent ainsi à deux valeurs distinctes en sortie des bascules. Cette différence conduit à la détection d'une attaque potentielle sur les chemins.

En outre, afin d'assurer la non-détection de faux-positifs, le délai introduit doit avoir une durée inférieure à la durée de l'impulsion laser qui peut être générée dans le circuit.

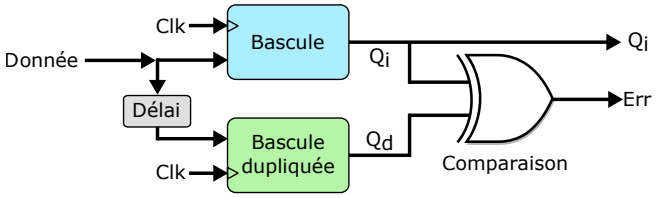


FIGURE 3.5 – Modalités d'implémentation d'une solution de type redondance temporelle.

1.2.2.4 Les capteurs de perturbations physiques :

Outre les redondances spatiales et temporelles permettant la détection de fautes induites dans un circuit intégré, les capteurs de perturbations physiques permettent de détecter toute anomalie générée dans le système. Ces anomalies ou perturbations peuvent être dues à une variation de la tension d'alimentation (détecteur à base d'un comparateur de tension par exemple), de la fréquence d'horloge, de la température (capteur à base d'une diode) ou encore d'une illumination photoélectrique (détection à base d'une photodiode polarisée en inverse). La Figure 3.6 présente différents capteurs, résumés dans [152]. Ces capteurs sont conçus pour monitorer des fautes distinctes.

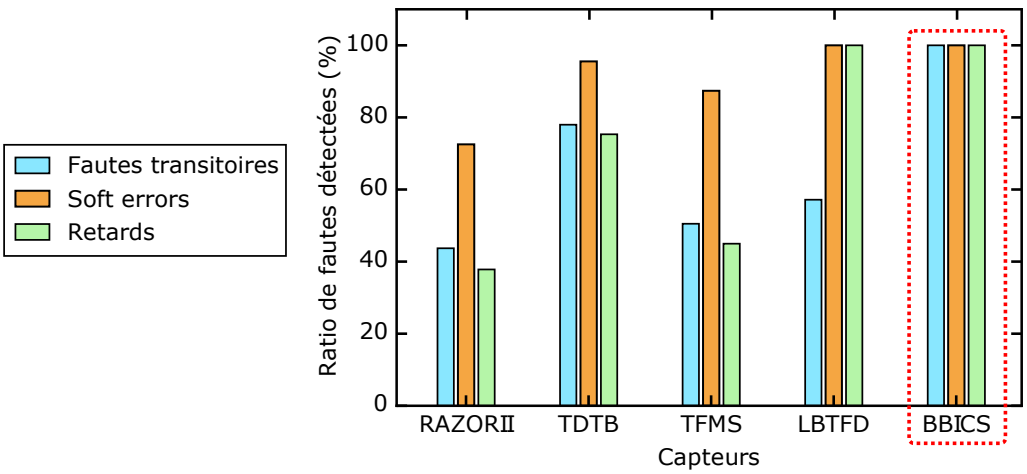


FIGURE 3.6 – Performances de détection de différents capteurs de la littérature.

Le BBICS présente ainsi des caractéristiques excellentes pour la détection de différentes catégories de fautes comme les fautes transitoires. Ce capteur est présenté plus longuement dans la Section 1.3 et dans la proposition que nous faisons d'un détecteur performant contre différentes catégories d'attaques (l'objectif de ce chapitre).

Toutefois, ces capteurs ne sont capables que de déterminer l'occurrence d'une perturbation physique ou d'une faute mais non de la corriger. C'est pourquoi, cette opération doit être relayée au système qui doit déterminer la marche à suivre, à savoir corriger la faute ou supprimer toutes les données en cours de traitement et réinitialiser la structure. Ce choix dépendra essentiellement du niveau de confidentialité des données traitées et de l'application. Les capteurs de perturbations peuvent être favorisés aux redondances spatiales dans le cadre d'une intégration dans les objets connectés, compte tenu que la pénalité en surface est plus réduite dans ce cas.

1.3 Détecteur de courants de substrats BBICS

Le détecteur de courants de substrat ou – *Bulk Built-In Current Sensor* – (BBICS) est l'une des architectures les plus attrayantes du paysage des détecteurs de perturbations physiques et des solutions proposées pour détecter une faute induite par effet photoélectrique [153]. Le BBICS est principalement captivant pour son fort potentiel de détection des invasions ciblant le substrat du circuit intégré.

1.3.1 Description du capteur BBICS

Chaque injection de fautes induisant un courant soit dans le substrat de type P "Pwell" soit dans le puits de type N "Nwell" implique le déclenchement d'un drapeau *"Flag"* informant l'utilisateur de cette attaque (ou tentative). Ainsi, "Pwell" et "Nwell" sont simultanément surveillés respectivement par le *pull-up* et le *pull-down* des réseaux CMOS [154] comme représentés schématiquement sur la Figure 3.7. Ainsi, cette structure est un capteur embarqué qui surveille le courant transitoire induit par une source laser au niveau du substrat. Depuis que cette architecture a été proposée par E. H. Neto *et al.* dans [155], [156], le capteur de type BBICS est l'un des détecteurs les plus étudiés par les experts de la sécurité [157], [158] ou encore [159]. En effet, ces différentes études ont considérablement amélioré le fonctionnement du capteur et développé son éventail de facultés (comme par exemple la détection simultanée de toute attaque sur le "Pwell" et sur le "Nwell" et l'amélioration des niveaux de sensibilité du détecteur).

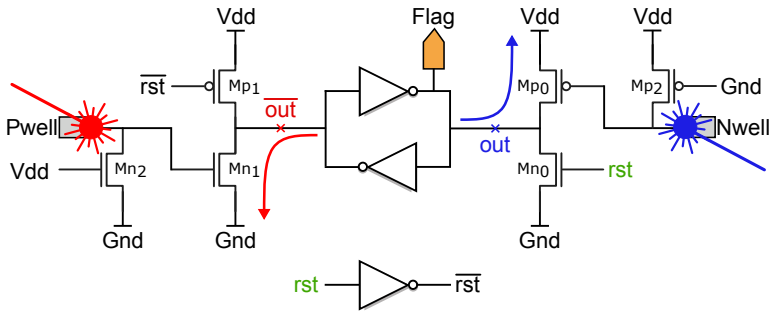


FIGURE 3.7 – Schéma électrique du capteur BBICS.

1.3.2 Fonctionnement du BBICS

Le BBICS est une architecture qui permet la détection de photo-courants ultra-rapides (de l'ordre de la centaine de picoseconde au moins) générés dans le substrat. Lorsque cette injection est réalisée, le basculement du verrou (ou *latch*, correspondant aux deux inverseurs rebouclés) composant cette structure d'un état à son inverse, permet de détecter la perturbation physique.

Dans son état de fonctionnement stationnaire (hors attaque), les deux inverseurs du verrou dont les entrées/sorties ("out" et $\overline{\text{out}}$") sont rebouclées se trouvent dans un état tel que la sortie "out" est à l'état bas '0' (via le transistor M_{n0}) alors que "$\overline{\text{out}}$" est fixé à '1' (via le transistor M_{p1}). Ces états sont déterminés lors de l'initialisation de l'architecture via le signal de contrôle "rst". Après l'initialisation, le verrou (ou *latch*) maintient cet état tant qu'aucune perturbation n'est générée sur le substrat. Les nœuds "out" et "$\overline{\text{out}}$" basculent respectivement vers les états logiques '1' et '0' suite à la faute transitoire induite dans la structure.

Considérons le cas où l'attaque perturbe le substrat de type P "Pwell" (illustré Figure 3.7), le photo-courant induit dans le dispositif est forcé par le transistor M_{n2} à la masse. Ce courant va ainsi conduire à l'augmentation de la tension V_{ds} de ce transistor jusqu'à devenir suffisante pour activer le transistor M_{n1}, puisque sa tension de contrôle V_{gs} augmente jusqu'à dépasser la tension de seuil du transistor. Par conséquent, l'activation du transistor M_{n1} conduit le nœud "$\overline{\text{out}}$" à la masse. Par rétroaction, le nœud "out" est forcé à Vdd, déclenchant alors le signal de détection d'attaques *"Flag"*.

Dans le cas de l'établissement d'un courant transitoire dans le caisson de type N "Nwell" du circuit, de manière analogue, les transistors M_{p0} et M_{p2} forcent le nœud "out" à l'état logique '1', induisant le basculement de *"Flag"* à '1', comme illustré Figure 3.7.

La sensibilité de ce capteur aux faibles injections de courants est déterminée par le dimensionnement des deux inverseurs rebouclés formant un verrou. Cette caractéristique est à prendre en compte lors de la conception afin de détecter de petites variations aux nœuds "out" et "$\overline{\text{out}}$". La précision de détection est d'autant plus performante selon le dimensionnement des transistors M_{p0} et M_{n1}. En effet, plus le rapport W/L de ces transistors est élevé, plus le verrou est sensible.

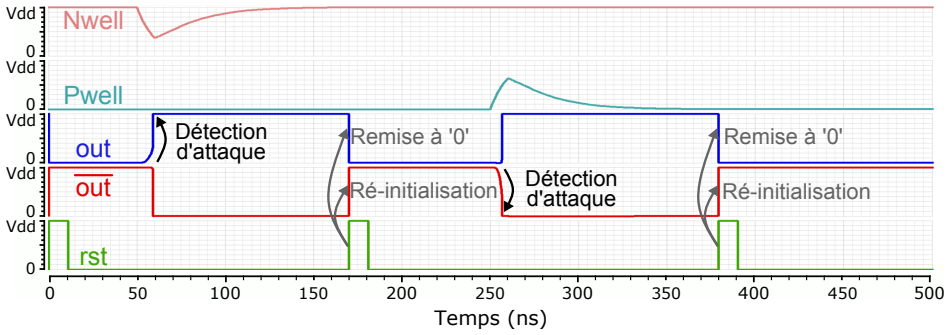


FIGURE 3.8 – Simulation électrique du fonctionnement du capteur BBICS.

Il convient de noter qu'une fois que le signal de détection d'attaques *"Flag"* est activé, le verrou reste bloqué en mode **perturbation détectée** (signal *"Flag"* à '1'). Il est donc nécessaire de réinitialiser l'architecture après chaque détection par le biais du signal de remise à zéro "rst", comme illustré sur la Figure 3.8.

L'association de ce capteur BBICS dédié à la détection d'illuminations LASERs, au travail effectué sur la sensibilité des jonctions de type STT-MRAMs face à des perturbations thermiques, a conduit au développement d'une architecture performante pouvant détecter différentes catégories d'attaques. Ce capteur est nommé *Dual Detection of Heating and Photocurrent attacks* (DDHP) et est présenté en section 2.

2 Détecteur d'attaques thermiques et photoélectriques

Le travail présenté dans ce chapitre a pour but de détecter des perturbations :

— Photoélectriques : induites sur les jonctions PN polarisées en inverse.

— Thermiques : générées sur les dispositifs mémoires émergents positionnés dans le BEoL, comme par exemple les mémoires OxRAMs [160] ou STT-MRAMs [83].

Pour cela, un nouveau capteur associant la solution de détection de courants de bulk anormaux BBICS et l'instanciation de la technologie STT-MRAM de type perpendiculaire au plan est développé. Cette architecture innovante apporte un haut niveau de sécurité permettant la détection simultanée de fautes de natures différentes : thermiques et/ou photoélectriques. Cette solution matérielle proposée dans ce chapitre est désignée par *Dual Detection of Heating and Photocurrent attacks* (DDHP).

La simulation du capteur est réalisée en utilisant le nœud technologique 28 nm CMOS FD-SOI associé à une technologie STT-MRAM perpendiculaire au plan dont le diamètre de jonction est fixé à 40 nm. L'efficacité du capteur proposé est équivalente pour un procédé CMOS bulk, comme indiqué dans [140]. Même si la technologie CMOS FD-SOI est moins sensible que la technologie CMOS bulk, les deux processus impliquent des mécanismes similaires face aux attaques [140].

2.1 Fonctionnement du capteur DDHP

Dans cette section, le détecteur DDHP sera décomposé en deux structures qui seront décrites successivement :

— L'architecture permettant la détection d'une perturbation extérieure de type photoélectrique et/ou thermique.

— Le circuit d'initialisation/programmation des jonctions STT-MRAMs et de la circuiterie CMOS.

2.1.1 Détection d'attaques externes

L'architecture proposée est basée sur l'opération de détection du BBICS pour les injections de fautes sur les substrats de type P ou celles dans les caissons de type N. Elle est associée à une structure permettant la détection de l'altération et de la modification de la résistivité de la jonction tunnel magnétique MTJ_{sense} (d'un état AP vers l'état P, comme présenté Chapitre 2, Section 2.1.3 en page 45). Le schéma de la phase de détection de cette nouvelle structure est présenté sur la Figure 3.9.

D'une part, cette opération de détection est réalisée en comparant les résistivités de deux points mémoires : la jonction de détection MTJ_{sense} à une référence notée MTJ_{ref}. Cette référence est choisie comme valeur moyenne des résistances R_{AP} et R_P. Elle doit garder une résistivité $R_{MTJ\ ref}$ telle que : $R_{MTJ\ ref} = \frac{R_{AP}+R_P}{2}$. Pour cela, quatres jonctions STT-MRAMs sont positionnées de manière série/parallèle deux à deux. Deux résistances (de faible et forte résistivité) sont instanciées en série. Cette branche est en parallèle d'une seconde branche composée de deux résistances de résistivités R_{AP} et R_P, comme illustrés sur la Figure 3.9.

D'autre part, étant donné que l'état initial de la boucle de rétroaction du BBICS est donné pour "out" forcé au niveau logique bas '0' et "$\overline{out}$" forcé au niveau logique haut '1', alors la jonction MTJ_{sense} (de détection, programmée initialement dans un état logique '1') est connectée au nœud "$\overline{out}$". La référence est quant à elle connectée au nœud "out". En effet, tant que le circuit est en mode veille, la résistivité de la jonction MTJ_{sense} est restituée dans le verrou et ainsi le nœud "$\overline{out}$" reste à l'état logique '1' (MTJ_{sense} toujours dans l'aimantation AP), alors que "out" est maintenu au niveau bas. Ainsi, le verrou et le détecteur restent dans un état stationnaire.

L'altération de la résistivité de cette mémoire de détection permet la mise en évidence d'une attaque de type thermique. Cette détection est effectuée par les deux transistors M_{nRd1} et M_{nRd2}. Ces transistors sont commandés par le signal de lecture "Rd", comme illustré Figure 3.9. En fonction de l'état logique de la jonction MTJ_{sense}, le '1' indiquant l'absence de fautes sur la technologie STT-MRAM et le '0' signalant une attaque ou une tentative d'attaque sur le circuit, la réponse du circuit diffère.

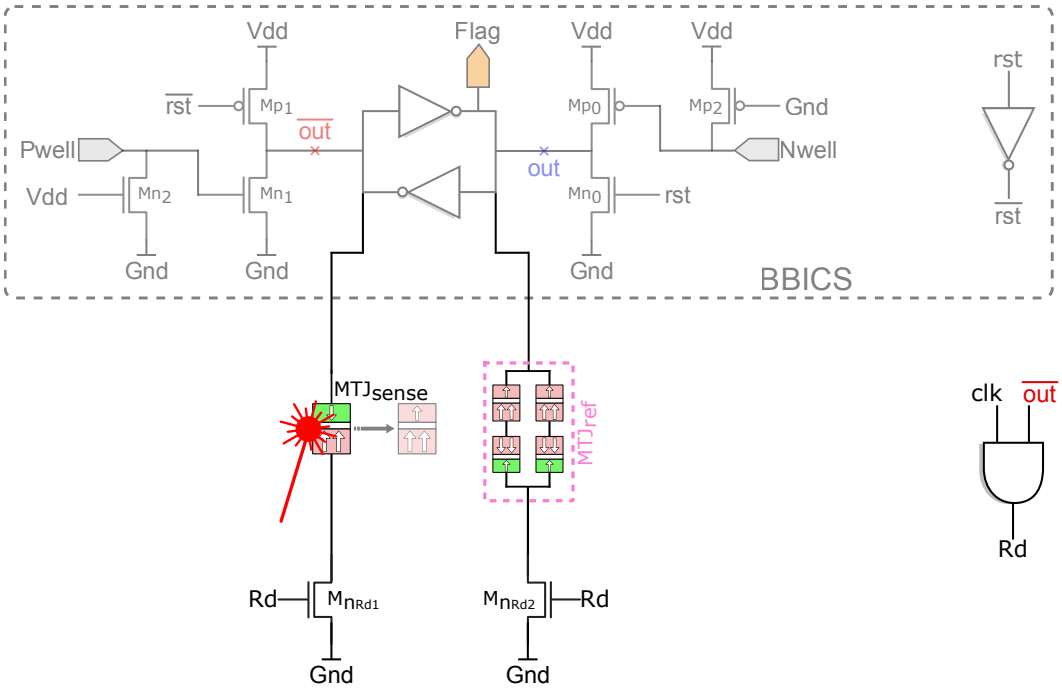


FIGURE 3.9 – Fonctionnement du capteur DDHP en mode détection.

— D'une part, si la valeur stockée dans la jonction MTJ_{sense} correspond à la valeur logique '1' (forte résistivité et une aimantation magnétique AP), alors cette valeur est comparée à la référence (de résistivité moyenne). Ces deux points mémoires se déchargent respectivement à travers les transistors M_{nRd1} et M_{nRd2}. La branche de plus faible résistance (la référence) se décharge plus rapidement. Par conséquent, l'état logique '0' est restitué au nœud "out" et par rétroaction '1' au nœud "$\overline{out}$". Dans ce cas de figure, le système reste en mode **"pas de détection"** d'attaques.

— D'autre part, lorsque la MTJ_{sense} commute de la polarisation AP à l'aimantation magnétique P (de '1' vers '0'), la résistivité de cette branche de détection devient plus faible que celle de la branche de référence. Ainsi, alors que "$\overline{out}$" est forcé à '0', la *latch* conduit à l'écriture d'un '1' au nœud "out", activant le signalement de l'attaque.

Comme illustré sur la Figure 3.9, la détection d'une modification de la résistivité sur la jonction MTJ_{sense} est limitée à l'addition de 2 transistors (M_{nRd1} et M_{nRd2}) et d'une porte logique de type "ET", par rapport à l'architecture de base du BBICS présentée précédemment. Cette porte "ET" réalise la lecture de l'état des jonctions mises en jeu dans ce capteur, à une fréquence donnée. Cette étape implique que le fonctionnement de ce capteur est réalisé de manière dynamique (contrairement au BBICS dont le fonctionnement est statique). Dans le cadre d'une forte exigence de sécurité de l'application cible, la fréquence de fonctionnement du détecteur peut être élevée, tandis que pour un niveau de sécurité standard, l'horloge de fonctionnement interne du circuit pourrait être suffisante.

En outre, cette opération de lecture n'est effectuée que si aucune faute (photoélectrique et/ou thermique) n'a déjà été induite dans le détecteur DDHP. En effet, lorsqu'une détection d'attaque survient au nœud *"Flag"*, alors le nœud "$\overline{out}$" est forcé à '0'. Ainsi, tant que *"Flag"* est à '1', le signal de contrôle "Rd" des transistors M_{nRd1} et M_{nRd2} (en sortie de la porte logique "ET") reste à '0'. Les jonctions ne sont plus lues. Pour redémarrer le capteur, l'architecture doit être réinitialisée via le signal "rst" et les jonctions reprogrammées vers leurs états initiaux (MTJ_{sense} vers l'aimantation magnétique AP et MTJ_{ref} vers une résistivité moyenne).

2.1.2 Phase de reprogrammation

Ainsi, dans le cadre du DDHP, pour toute faute induite sur la MTJ_{sense}, cette jonction doit être reprogrammée vers son état de référence, vers l'état AP. Pour cela, un schéma de programmation permettant également la réécriture des 4 jonctions de référence vers les états AP et P (deux par deux) a été instancié.

Afin de garantir dans la référence que deux jonctions sont écrites dans l'état P et les deux autres dans l'état AP, deux jonctions en série sont inter-connectées par leurs couches de référence, comme illustré dans l'encart de la Figure 3.10. En effet, dans un premier temps, sur les 2 branches parallèles, le courant est injecté par la couche de stockage des JTMs (les jonctions R_{1a} et R_{2a} écrites vers l'état AP). Ce courant est ensuite injecté dans les jonctions R_{1b} et R_{2b} par leurs couches de référence (écrivant leur état vers P). Cette disposition des jonctions de références permet de toujours comparer la jonction de détection (MTJ_{sense} dont l'état déterminera si l'attaque a eu lieu) à une référence (MTJ_{ref}) de résistivité moyenne.

La Figure 3.10 illustre les éléments supplémentaires nécessaires à intégrer au circuit afin de réaliser la première initialisation des JTMs et leur réécriture après chaque attaque. Comme indiqué, la programmation des différentes jonctions est permise par l'ajout de 3 transistors (M_{nWrE0}, M_{nWrE1} et M_{pWrE1}) et d'une porte logique de type "inverseur".

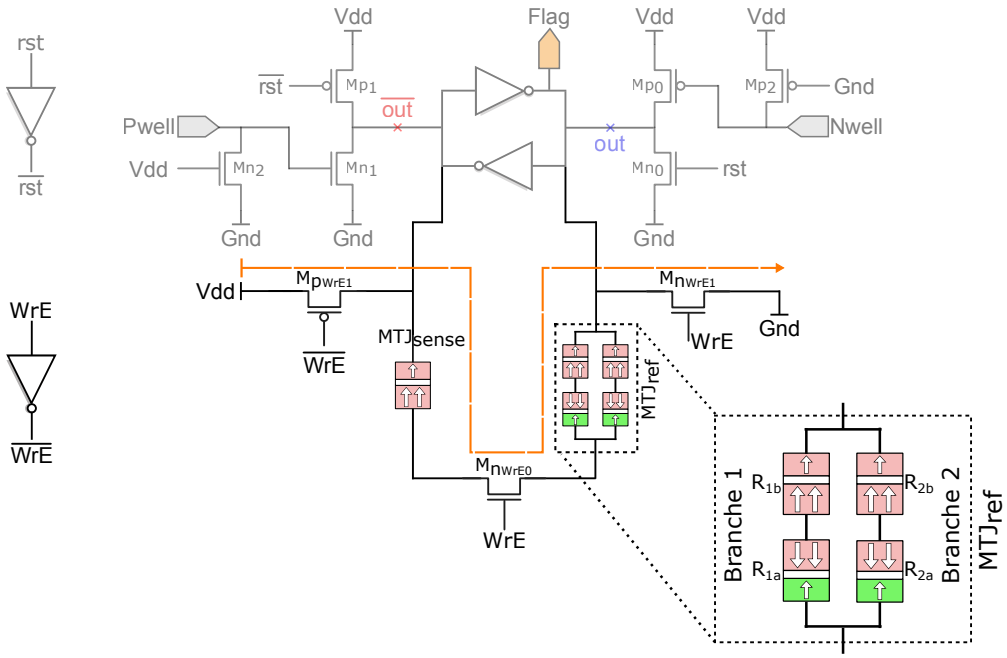


FIGURE 3.10 – Fonctionnement du DDHP en phase d'initialisation/programmation. Encart : illustration de la disposition des jonctions dans la référence.

La programmation des jonctions est réalisée par le passage d'un courant à travers les transistors M_{nWrE0}, M_{nWrE1} et M_{pWrE1}. Le chemin parcouru par ce courant est illustré sur la Figure 3.10. En effet, ce courant commence par traverser le transistor M_{pWrE1}, puis l'aimantation AP est programmée dans la jonction MTJ_{sense} (courant injecté par la couche de stockage). Ce courant est ensuite conduit par le transistor M_{nWrE0} et est divisé en deux composantes afin de réaliser la programmation de la référence, d'après le protocole décrit précédemment. Il est alors transmis au dernier transistor M_{nWrE1}.

À ce circuit d'écriture des JTMs, la porte logique "inverseur" est ajoutée, contrôlée par le signal d'écriture "WrE". La sortie de cette porte active le transistor PMOS tandis que les transistors NMOS sont quant à eux directement activés par le signal d'écriture "WrE".

2.2 Attaque visant les jonctions de référence

Cette architecture reste efficace contre les attaques par injection de fautes, même si un attaquant expérimenté cible les jonctions de référence et non la jonction introduite pour la détection. Ce second chemin d'attaque doit être pris en compte lors de la réalisation du dessin des masques (ou *layout*) du circuit. En effet, l'une des solutions les plus simples à implanter pour compromettre cette attaque, consiste à fortement rapprocher la jonction de détection et les jonctions de référence qui ont une aimantation AP, afin qu'elles soient dans le diamètre du tir LASER, environ 1 µm. Par conséquent, il serait impossible de changer l'orientation magnétique de la référence sans commuter la jonction de détection. Ainsi, la structure détectera toujours toute attaque visant ces jonctions.

2.3 Simulations électriques du capteur

L'architecture générale du capteur DDHP est représentée Figure 3.11. Cette structure est composée des différentes sous-architectures décrites précédemment (sous-sections 2.1.1 et 2.1.2) et peut détecter différents modèles d'attaques : des perturbations photoélectriques à travers le BBICS et des effets thermiques qui peuvent être induits sur les STT-MRAMs.

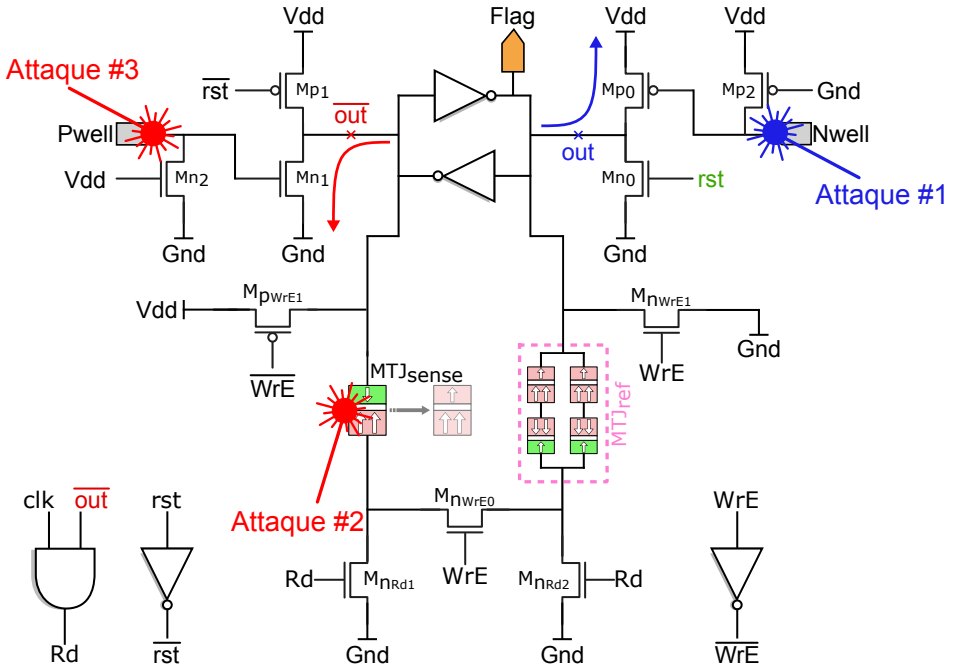


FIGURE 3.11 – Schéma de fonctionnement général du détecteur DDHP.

2.3.1 Vérification du fonctionnement électrique du DDHP

Comme le montre la Figure 3.12, la double détection photoélectrique et thermique est efficace pour différentes techniques d'injection de fautes. Alors que le BBICS détecte un courant de substrat anormal dans le "Nwell" et dans le "Pwell", correspondant respectivement aux détections 1 et 3, la 2nde est induite par la lecture dynamique de la MTJ_{sense} ayant commuté d'un état AP vers l'aimantation magnétique P, par exemple sous l'effet d'une irradiation de type laser.

D'une part, après chaque détection, la structure doit être réinitialisée par le signal "rst" et la forte résistivité de la jonction MTJ_{sense} doit être rétablie grâce au signal "WrE" de période t_{write}. Les JTMs de référence sont également re-programmées. Après la ré-initialisation du circuit, MTJ_{sense} est re-programmée vers le niveau logique '1' (résistivité élevée - R_{AP}) et la référence vers une résistivité moyenne.

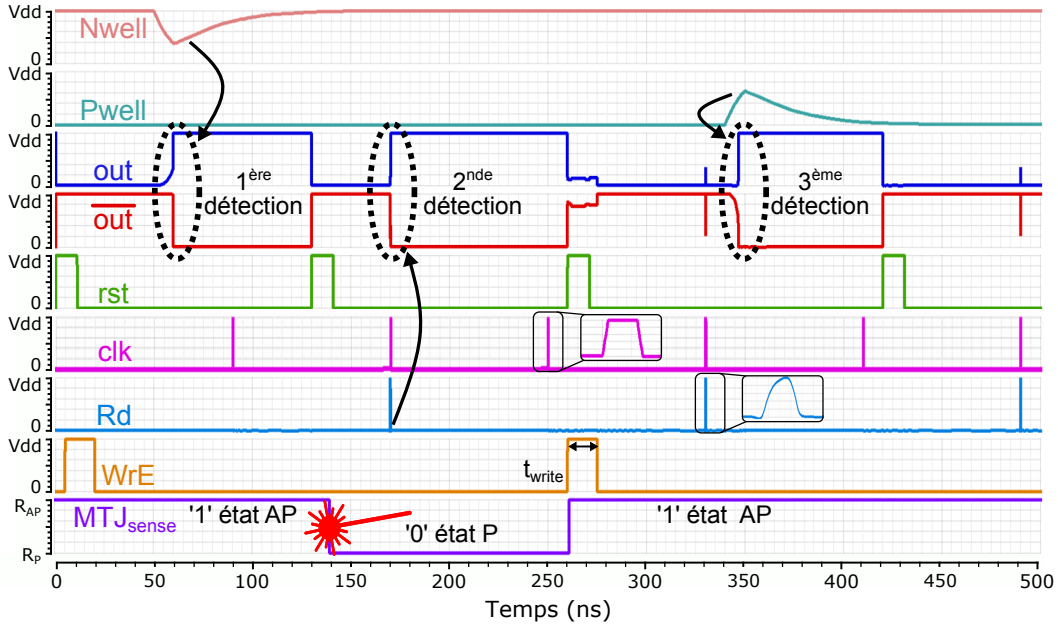


FIGURE 3.12 – Simulation électrique du capteur DDHP. Mise en évidence de la bonne détection des trois catégories d'attaques, respectivement : photoélectrique sur le Pwell (attaque #1), thermique sur la STT-MRAM (attaque #2) et photoélectrique sur le Nwell (attaque #3).

D'autre part, aucune modification n'est observée sur les nœuds "out" et "$\overline{\text{out}}$" tant qu'aucune erreur n'est survenue. Ces vérifications sont effectuées à t = 330 ns et t = 490 ns, comme illustré Figure 3.13. Ainsi, cette architecture n'introduit pas de faux positifs (le signal *"Flag"* levé alors qu'aucune attaque n'est en cours), dans le cas idéal de cette simulation. Sur cette simulation, représentée sur la Figure 3.12, la durée d'impulsion d'écriture choisie pour les JTMs est égale à 15 ns pour écrire la jonction de détection et celles de référence. Cette durée peut être réduite à 1,1 ns pour ne réaliser que l'écriture de la MTJ_{sense} vers l'état AP. Le signal de lecture des jonctions est induit quant à lui par une impulsion de 1 V et d'une durée de 80 ps.

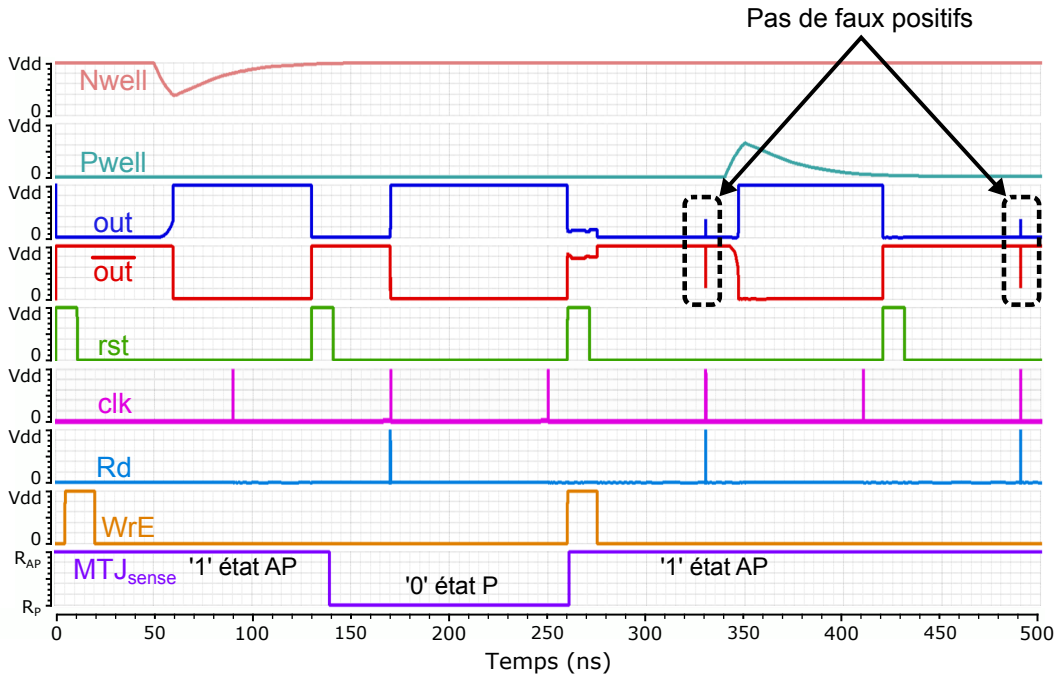


FIGURE 3.13 – Absence de faux-positifs lors de la simulation électrique dans le cas idéal du DDHP.

2.3.2 Simulations Monte Carlo du détecteur

L'architecture DDHP a été simulée à l'aide de la méthode Monte Carlo (MC) en réalisant une simulation de 1000 itérations (ou tirages) pour vérifier le fonctionnement de la structure. Cette simulation a été réalisée pour une TMR fixée à 150 % et en tenant compte de la distribution gaussienne des résistances R_{AP} et R_P des jonctions, avec une variabilité maximale de 10 %.

De plus, la simulation MC nous a permis d'optimiser le dimensionnement des transistors pour que le nombre de faux positifs soit réduit au maximum (voire nul), ce qui sera démontré dans cette section.

2.3.2.1 Fonctionnement du circuit sur 1000 simulations parallèles (avant optimisation) :

Dans le cadre de la simulation MC de ce détecteur, le même dimensionnement des transistors que celui de la simulation électrique idéale est utilisé (les jonctions sont considérées comme ayant deux résistivités distinctes et une TMR suffisante). Comme il peut être noté sur la Figure 3.14, bien que les trois détections (photoélectriques sur le *Pwell* et *Nwell* et thermique sur la STT-MRAM) aient été réalisées, des faux-positifs ont été observés sur 5 % des simulations.

Lors de cette implémentation tous les transistors du DDHP ont une longueur de grille de 30 nm (longueur minimale -*length L*- autorisée par le *Process Design Kit* – (PDK)). La largeur de ces transistors (*width W*) est fixée selon le tableau 3.1.

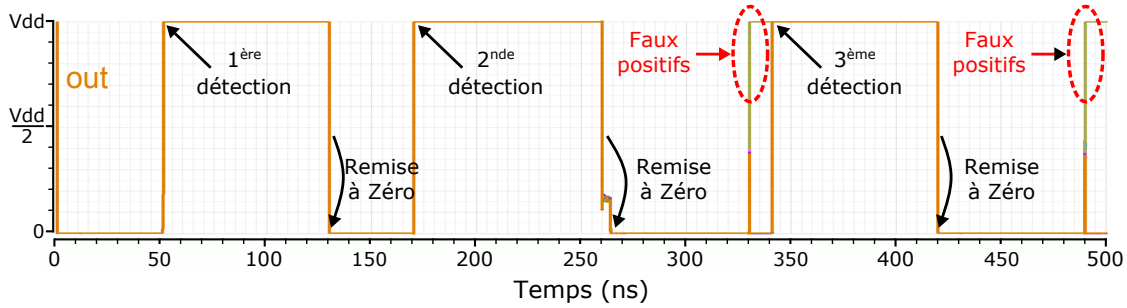


FIGURE 3.14 – Simulation Monte Carlo de 1000 itérations du détecteur DDHP. Simulation sans optimisation de la taille des transistors de lecture : 5 % de faux-positifs lorsque les deux transistors de lecture sont équilibrés.

Transistors	M_{n0}	M_{p0}	M_{n1}	M_{p1}	M_{n2}	M_{p2}
W (nm)	80	850	80	80	80	80

Transistors	M_{nRd1}	M_{nRd2}	M_{nWrE0}	M_{nWrE1}	M_{pWrE1}
W (nm)	160	160	480	80	400

TABLEAU 3.1 – Dimensionnement des largeurs des transistors du capteur DDHP avant son optimisation.

Il est donc nécessaire d'optimiser le dimensionnement des transistors composant cette structure afin de réduire le nombre de faux-positifs.

2.3.2.2 Optimisation du circuit :

La structure du DDHP est optimisée en déséquilibrant la taille des transistors M_{nRd1} et M_{nRd2}, comme représenté dans le tableau 3.2. Tous les transistors utilisent toujours la même longueur de grille de 30 nm. Ainsi, pour un rapport W/L du transistor M_{nRd1} supérieur au rapport W/L du transistor M_{nRd2}, la capacité de détection de cette architecture est optimisée. Une précision et une fiabilité de détection de 100 % sur le nœud de sortie "out" est atteinte, aucun faux positif n'est détecté, comme illustré sur la Figure 3.15. Ainsi, les paramètres qui ont une influence majeure

sur le fonctionnement de ce capteur sont les dimensionnements des transistors de lecture (ou de détection).

Transistors	M_{n0}	M_{p0}	M_{n1}	M_{p1}	M_{n2}	M_{p2}
W (nm)	80	850	80	80	80	80

Transistors	M_{nRd1}	M_{nRd2}	M_{nWrE0}	M_{nWrE1}	M_{pWrE1}
W (nm)	140	80	480	80	400

TABLEAU 3.2 – Dimensionnement des transistors (largeur de grille) du capteur DDHP après optimisation.

De plus, ce déséquilibre permet une meilleure détection des attaques qui sont induites sur la référence. En effet, si toutes les jonctions commutent d'une aimantation AP vers l'aimantation P lors d'une attaque (les jonctions de détection et de référence), alors la branche de détection, dont le ratio W/L du transistor est plus important, commutera toujours plus rapidement que la branche de référence, puisque le courant fourni sera plus élevé.

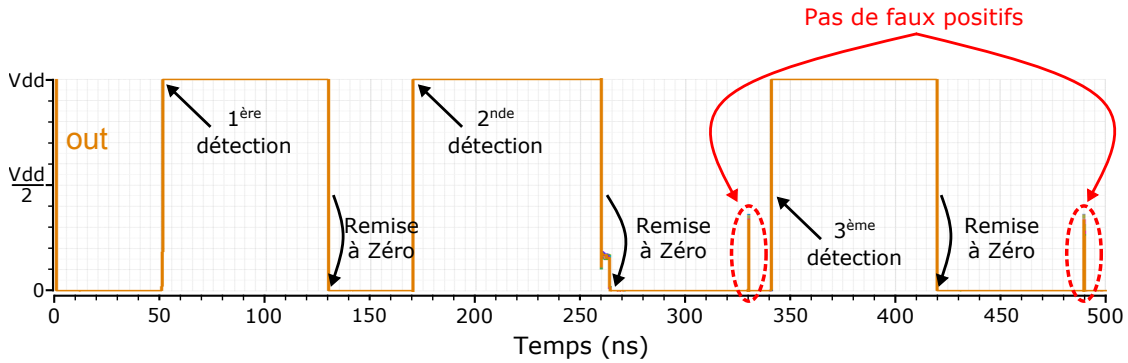


FIGURE 3.15 – Simulation Monte Carlo de 1000 itérations du détecteur DDHP. Optimisation du dimensionnement des transistors : pas de faux positifs lorsque les deux transistors de lecture sont déséquilibrés.

2.3.3 Superficie du capteur DDHP

L'architecture proposée DDHP est comparée au détecteur BBICS en termes de surface. Comme représenté sur la Figure 3.16, le capteur DDHP proposé demande une certaine surface supplémentaire estimée à environ 6,4 portes équivalentes ou – *Gate Equivalents* – (GE) par rapport à une solution purement BBICS.

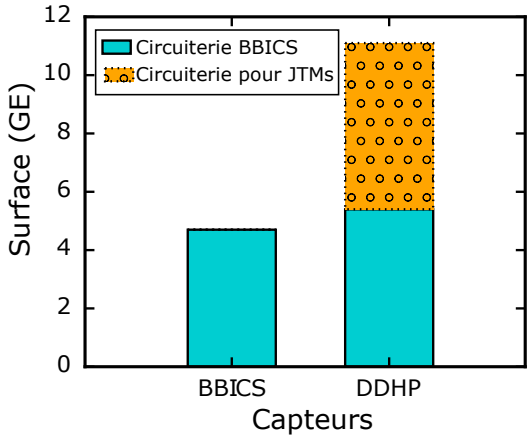


FIGURE 3.16 – Illustration de la surface de l'architecture DDHP, comparée à la structure BBICS.

Toutefois, cette contre-mesure intègre un panel de détection d'attaques plus important que le BBICS compte tenu qu'il permet la détection de perturbations qui visent le substrat et de mémoires non-volatiles émergentes.

3 Conclusion et perspectives

Dans le contexte de la sécurisation des objets connectés, le développement de solutions physiques réduisant la sensibilité des circuits aux injections de fautes et aux attaques par observation sont décrites dans ce chapitre. Le capteur proposé permet de détecter des perturbations physiques, aussi bien les attaques par injection de fautes réalisées sur le substrat, que les perturbations thermiques qui peuvent être induites sur les mémoires émergentes STT-MRAMs. Ce détecteur est appelé double détection d'attaques induites par chauffage et effet photoélectrique ou – *Dual Detection of Heating and Photocurrent attacks* – (DDHP). Alors que le BBICS procède à une détection statique, la détection des perturbations induites sur les jonctions du DDHP est dynamique. Ce capteur a démontré grâce à des simulations électriques et des simulations de Monte Carlo une excellente efficacité de détection et une très bonne robustesse face aux variations de procédés.

Il est tout de même nécessaire de réaliser l'implémentation du dessin des masques de cette contre-mesure afin la faire fabriquer et d'analyser la surface du circuit intégré où ce capteur permettra la détection. Il sera alors possible de proposer une version améliorée prenant en compte ces résultats afin d'obtenir une solution susceptible de *protéger* toute la surface du circuit. Pour cela, une méthode serait alors d'intégrer une matrice de jonctions 1R (pour la détection) où toute commutation pourrait être relevée.

Chapitre 4

Implémentation hybride de l'algorithme PRESENT

Il n'y a qu'une façon d'échouer, c'est d'abandonner avant d'avoir réussi

— Georges Clemenceau

Dans le cadre de ce chapitre, nous nous intéresserons dans un premier temps à à l'implémentation de l'algorithme de cryptographie par bloc PRESENT via l'hybridation des technologies CMOS et STT-MRAM. Cette étude vise à en optimiser les performances pour une utilisation dans les applications à ressources restreintes. Puis, l'apport de l'hybridation pour le niveau de sécurité sera discuté. Pour cela, les JTMs intégrées dans ce chiffrement hybride sont utilisées pour distinguer deux chemins d'attaques dans l'algorithme, un chemin CMOS pur et un chemin hybride. Cette redondance intitulée Double Redondance Technologique ou – *Dual Technology Redundancy* – (DTR) permet de détecter toute attaque qui serait induite sur la technologie CMOS ou sur la technologie STT-MRAM.

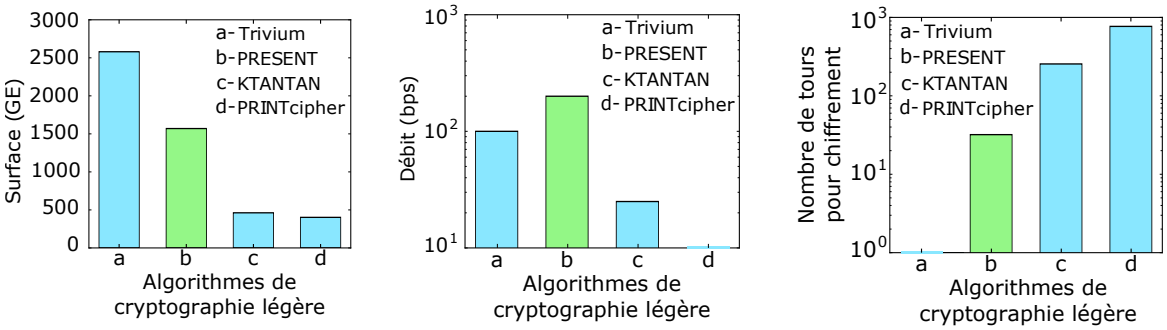
Les travaux présentés dans ce chapitre ont été publiés aux conférences internationales "DCIS 2018" [161] et "ISCAS 2019" [162].

Sommaire

1 Implémentations matérielles de la cryptographie légère

La sécurité est un enjeu majeur du développement d'applications dont les ressources sont restreintes, comme dans les applications embarquées de l'Internet des Objets. Pour cela, différents algorithmes de cryptographie légère ont été développés. Ces chiffrements répondent aux divers besoins de ces applications. Alors que certains visent une implémentation matérielle, d'autres visent une instanciation logicielle [163].

Pour exemple, la Figure 4.1 illustre l'implémentation matérielle de quatre algorithmes de cryptographie légère : Trivium [53], PRESENT [68], KTANTAN [164] et PRINTcipher [165]. Lors du choix de l'utilisation d'un algorithme de cryptographie, différents paramètres doivent être pris en compte : la surface silicium nécessaire pour l'implémenter (représentée sur la Figure 4.1.a), son efficacité de chiffrement (représentée sur la Figure 4.1.b) ainsi que le nombre de tours nécessaires pour l'obtention de ce texte chiffré (représentée sur la Figure 4.1.c).



(a) Surface de l'implémentation physique d'algorithmes de cryptographie légère.

(b) Efficacité (débit) d'algorithmes de cryptographie légère.

(c) Nombre de tours nécessaires pour chiffrer un message, pour différents algorithmes.

FIGURE 4.1 – Comparaison de l'implémentation physique de différents algorithmes de cryptographie légère : Trivium [53], PRESENT [68], KTANTAN [164] et PRINTcipher [165].

Afin de pouvoir comparer différents algorithmes de cryptographie implémentés dans des nœuds technologiques distincts, la métrique GE est utilisée. Cette métrique est déterminée en divisant la surface silicium d'un circuit intégré, par la taille d'une porte NAND à deux entrées dans la même technologie, afin d'obtenir la surface de ce circuit exprimée en GE. Cette métrique est donc indépendante du nœud technologique. La surface doit être la plus faible possible pour une intégration dans les objets connectés. Le débit quant à lui est exprimé en bps (bits par seconde) et montre l'efficacité de cet algorithme (recherchée la plus élevée possible). Enfin, la vitesse est déterminée par le nombre de tours que dois effectuer chaque bloc de l'algorithme afin d'obtenir le texte chiffré final.

Le meilleur compromis entre ces différents paramètres doit être trouvé. Comme illustré sur la Figure 4.1, le chiffrement PRESENT se présente comme un bon candidat pour ces paramètres. De plus, l'algorithme de cryptographie PRESENT a été standardisé en 2012 par la norme ISO/IEC 29192-2 :2012 [166]. C'est pourquoi, nous avons décidé dans le cadre de ce chapitre de considérer l'hybridation de cet algorithme de cryptographie avec la technologie mémoire STT-MRAM pour une amélioration potentielle des performances de ce chiffrement en terme de consommation principalement, tout en modifiant ses caractéristiques sécuritaires.

1.1 Algorithme de cryptographie légère PRESENT

L'algorithme de cryptographie PRESENT a été proposé en 2007 par Bogdanov *et al.*, lors de la conférence internationale *Cryptographic Hardware and Embedded Systems* – (CHES) [68] comme alternative à l'algorithme de cryptographie utilisé jusqu'à aujourd'hui comme standard, l'AES.

PRESENT fait partie des algorithmes de chiffrements par blocs et plus précisément de la sous-catégorie SPN. La Figure 4.2 illustre le fonctionnement général du chiffrement PRESENT. La première étape de cet algorithme cryptographique consiste à réaliser une fonction "XOR" entre le message clair à chiffrer M de 64 bits et la clé maître k de 80 bits. Puis, une succession de substitutions-permutations sont associées aux mises à jour de la clé afin d'obtenir, au $32^{ème}$ tour, le texte chiffré C de 64 bits.

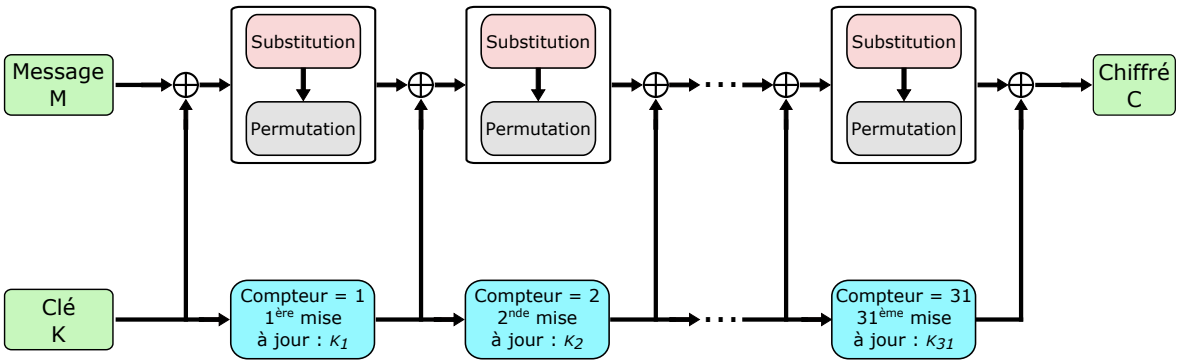


FIGURE 4.2 – Représentation du fonctionnement général de l'algorithme de cryptographie PRESENT.

1.2 Fonctionnement de l'algorithme PRESENT

La Figure 4.3 illustre l'architecture complète de l'algorithme de cryptographie PRESENT. Lors du début du chiffrement, le message clair M et la clé k sont respectivement envoyés vers les multiplexeurs, "Mux_chiffrement" et "Mux_clé". Ainsi, lors du premier coup d'horloge, ces multiplexeurs transmettent respectivement aux deux bascules "FF_chiffrement et "FF_clé", le message M et la clé k.

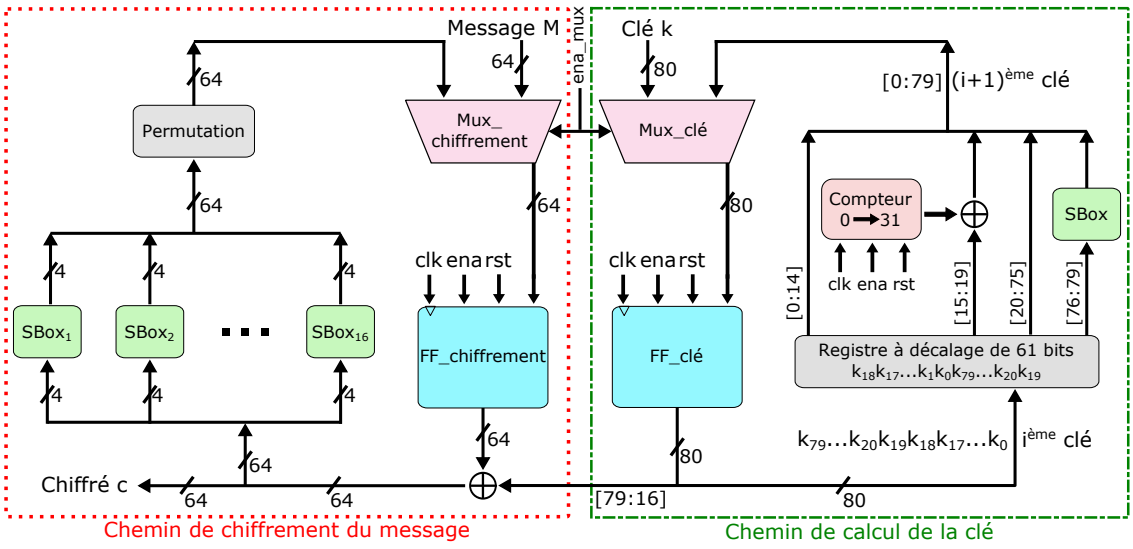


FIGURE 4.3 – Architecture de l'algorithme de cryptographie PRESENT.

Lors des coups d'horloge suivants, le texte en cours de chiffrement et la clé mise à jour sont envoyés aux bascules via les multiplexeurs. La première étape de chiffrement du message est réalisée par une fonction "XOR" entre les sorties des bascules "FF_chiffrement" (tous les bits) et "FF_clé" (les bits de 16 à 79). Le résultat de l'opération "XOR" est découpé en 16 mots de 4 bits, comme illustré sur la Figure 4.3. Chaque mot de 4 bits est ensuite envoyé à une table de substitution notée *SBox*. Cette table de substitution remplace chaque mot x de 4 bits par un nouveau mot $S(x)$ de 4 bits [68], tel qu'illustré par le Tableau 4.1.

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
S(x)	C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2

TABLEAU 4.1 – Table de substitution de 4 bits vers 4 bits instanciée dans PRESENT [68].

Les 16 mots résultants de cette substitution sont ensuite concaténés pour former un mot de 64 bits. Les positions des bits dans le mot sont ensuite permutées grâce à une couche de permutation, comme indiqué dans [68]. Cette table de permutation est définie par le Tableau 4.2. Elle consiste à modifier les positions de 60/64 bits de ce mot. Les bits initialement à la position i sont translatés à la position P(i) dans le message. Les bits aux positions $i = 0$, $i = 21$, $i = 42$ et $i = 63$ restent inchangés.

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
P(i)	0	16	32	48	1	17	33	49	2	18	34	50	3	19	35	51
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
P(i)	4	20	36	52	5	21	37	53	6	22	38	54	7	23	39	55
i	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
P(i)	8	24	40	56	9	25	41	57	10	26	42	58	11	27	43	59
i	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
P(i)	12	28	44	60	13	29	45	61	14	30	46	62	15	31	47	63

TABLEAU 4.2 – Table de permutation du chiffrement PRESENT, tel que définie dans [68].

A chaque cycle d'horloge, pendant le calcul du texte en cours de chiffrement (substitution et permutation), la clé est mise à jour. L'actualisation de cette clé (80 bits) est réalisée en suivant le processus définit ci-dessous :

— Décalage des bits de la clé maître envoyée en entrée de l'algorithme, par 61 bits vers la droite. Ainsi, lorsqu'au premier tour cette clé est considérée comme $k_n = [k_{79} \cdots k_{20} k_{19} k_{18} k_{17} \cdots k_0]$ avec k_i le i^{ème} bit de la clé, alors en sortie de ce registre à décalage la clé devient : $k_d = [k_{18} k_{17} \cdots k_1 k_0 k_{79} \cdots k_{20} k_{19}]$, (comme illustré Figure 4.3).

— Altération de la clé k_d, comme décrit ci-dessous :

- Les bits de 0 à 14 restent inchangés.
- Une fonction "XOR" est réalisée entre le mot formé par les bits 15 à 19 et la valeur contenue dans le compteur interne de l'algorithme de cryptographie. Ce compteur détermine à quel tour du chiffrement se trouve l'algorithme de cryptographie.
- Les bits 20 à 75 restent inchangés.
- Les bits de 76 à 79 sont substitués en utilisant la même table de substitution *SBox* précédemment décrite et illustrée sur la Figure 4.3.

La nouvelle clé obtenue et le message en cours de chiffrement sont ensuite renvoyés aux multiplexeurs et toutes les opérations précédemment décrites sont réitérées. Cet algorithme nécessite 32 coups d'horloge pour réaliser le chiffrement du message clair M et de la clé k pour l'obtention du texte chiffré C.

2 PRESENT hybride CMOS/STT-MRAM

L'algorithme de cryptographie PRESENT a été étudié dans le cadre de cette thèse en vue de son intégration dans le monde des objets connectés. Ainsi, ce chapitre propose une version hybride CMOS/STT-MRAM de ce chiffrement afin d'en déterminer les performances en terme de

consommation et de surface par rapport à une architecture de référence purement CMOS. Ce chapitre met également en avant les attraits que peut apporter cette architecture en terme de sécurité, aux circuits intégrés. En effet, l'intégrité des données traitées, leur confidentialité et leur authenticité doivent être garanties à tout moment.

2.1 Architecture hybride CMOS/STT-MRAM de l'algorithme PRESENT

L'architecture hybride CMOS/STT-MRAM de l'algorithme de cryptographie PRESENT est basée sur le chiffrement PRESENT mais en modifiant certains blocs stratégiques. En effet, certaines FFs ont été remplacées par des NVFFs définies dans [167].

2.1.1 Algorithme PRESENT hybride en technologie CMOS/STT-MRAM

La modification des registres "FF_chiffrement", "FF_clé" et "Compteur" (illustrés Figure 4.3), de volatiles à non-volatiles, conduit à stocker respectivement les données, la clé et le compteur en cours de traitement afin de les restaurer ultérieurement, comme représenté sur la Figure 4.4.

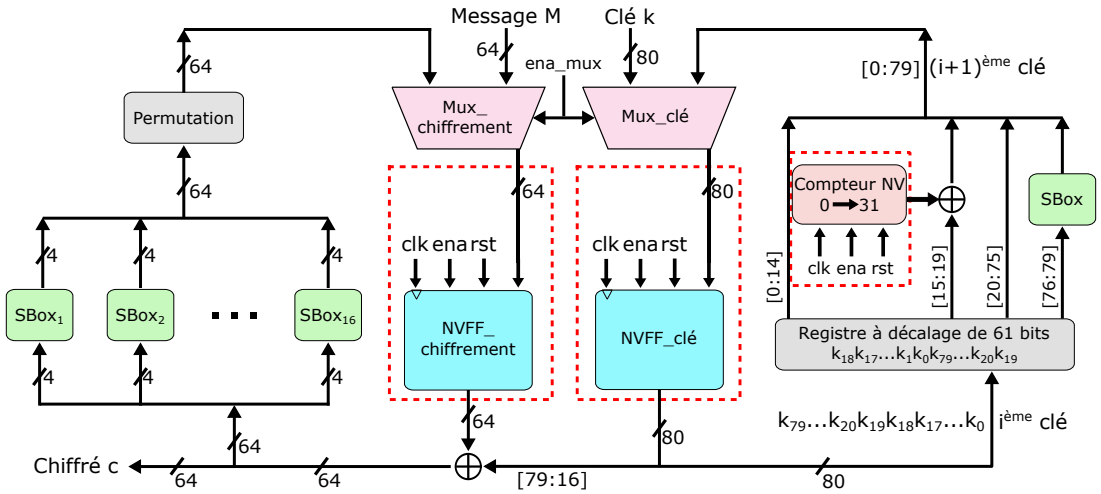


FIGURE 4.4 – Représentation de l'algorithme PRESENT hybride CMOS/STT-MRAM.

Ce nouveau type de chiffrement hybride CMOS/STT-MRAM permet d'avoir un système de stockage avant la mise hors tension du système. Ces données stockées sont ensuite restaurées, lors de la ré-activation du système. Il n'y a pas de consommation de veille lorsque le circuit est hors tension, les données sauvegardées étant non-volatiles. Il est important de noter que les circuits hybrides ouvrent de nouvelles opportunités à la sécurité telle que la restauration du contexte approprié lorsque des attaques sont détectées, par exemple.

La contribution principale de ce chapitre consiste en l'évaluation de quatre circuits hybrides du chiffrement PRESENT basés sur des NVFFs, pour des nœuds technologiques matures (CMOS 180 nm et STT-MRAM de 200 nm de diamètre) et des nœuds avancés (CMOS 28 nm et STT-MRAM de 40 nm). Ces circuits sont comparés à une référence du chiffrement PRESENT développée en pur CMOS en utilisant le nœud 180 nm.

Cette référence et le circuit hybride utilisant les nœuds CMOS 180 nm et STT-MRAM de 200 nm de diamètre ont été fabriqués dans le cadre du projet européen GREAT [168], dont Spintec est membre.

2.1.2 Propriétés de l'hybridation des technologies CMOS et STT-MRAM

L'approbation de cette solution cryptographique pour une intégration dans les objets connectés nécessite l'évaluation de sa consommation comparée à une structure CMOS pure. Pour cela,

une comparaison des énergies des deux circuits lorsqu'ils sont inutilisés est réalisée. La quantification de cette grandeur dans le cas de l'architecture hybride requiert l'évaluation des énergies de stockage et de restauration (respectivement $E_{stockage}$ et $E_{restauration}$) du circuit. Le résultat de cette énergie est ensuite comparé à l'énergie de veille (ou statique) $E_{veille\ CMOS}$ de la structure implémentée en CMOS pur, comme représenté sur la Figure 4.5. Ainsi, l'énergie de repos ($E_{repos\ hybride}$) de cet algorithme cryptographique hybride peut être définie comme dans l'Equation 4.1 et sur la Figure 4.5 avec :

$$E_{repos\,hybride} = E_{stockage} + E_{restauration} \tag{4.1}$$

Par conséquent, afin de déterminer le temps de veille minimal t_{veille} pour lequel le chiffrement hybride est énergétiquement plus efficace que l'algorithme de chiffrement PRESENT CMOS pur, la comparaison de la consommation lorsque les deux algorithmes sont inactifs est effectuée. Ainsi, l'algorithme de cryptographie PRESENT hybride CMOS/STT-MRAM est plus efficace pour :

$E_{repos\ hybride} < E_{veille\ CMOS}$ avec : $E_{veille\ CMOS} = P_{statique\ CMOS} . t_{veille}$.

$P_{statique\ CMOS}$ correspond à la puissance statique consommée par l'implémentation pure CMOS du chiffrement PRESENT, lors de son état de veille.

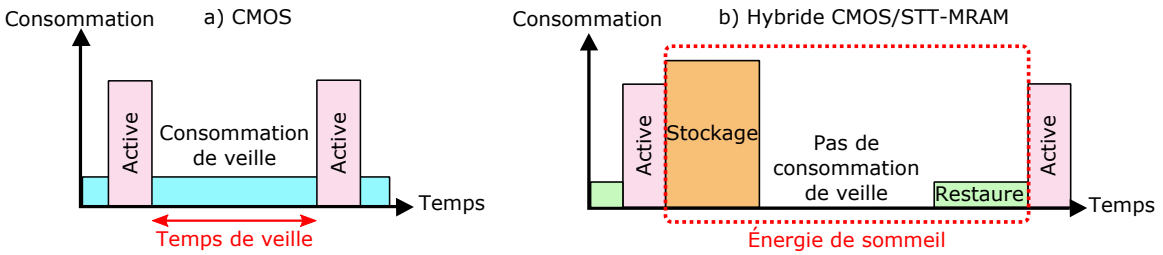


FIGURE 4.5 – a) Représentation de la consommation de l'algorithme de cryptographie PRESENT en mode actif et en mode veille. b) Illustration de la consommation de l'algorithme PRESENT hybride CMOS/STT-MRAM pendant sa phase active et de repos.

2.2 Implémentation physique du chiffrement PRESENT

Cet algorithme de cryptographie est dans un premier temps instancié en utilisant une technologie CMOS bulk 180 nm et une taille de jonction de 200 nm (diamètre de la JTM), pour deux scénarios distincts qui sont notés #1 et #2. En considérant que la version hybride de l'algorithme de cryptographie PRESENT est composée de 151 NVFFs, alors le schéma de programmation de ces jonctions doit être étudié. En effet, le courant nécessaire pour réaliser une commutation de cette taille de jonction (200 nm) est de l'ordre de grandeur du milliampère. L'écriture des jonctions doit donc être réalisée en série, soit en programmant une NVFF par coup d'horloge (scénario #1), soit en parallélisant partiellement les écritures, avec 5 NVFFs écrites par coup d'horloge (scénario #2). Ces deux scénarios sont comparés à une architecture pure CMOS afin d'illustrer les atouts et les inconvénients de l'hybridation sur les performances d'un algorithme de cryptographie de type PRESENT.

2.2.1 Flot de conception standard de l'algorithme PRESENT en technologie pure CMOS

Avant de réaliser l'implémentation physique d'une architecture hybride de l'algorithme de chiffrement PRESENT sur *Application-Specific Integrated Circuit* – (ASIC), la première étape nécessaire est de réaliser l'implémentation de l'algorithme PRESENT en utilisant le procédé pur CMOS bulk 180 nm.

Pour une réduction conséquente du nombre de plots d'entrées/sorties du circuit, nous avons choisi d'intégrer trois registres à décalage en entrées du message clair et de la clé, et en sortie du message chiffré, permettant l'envoi et la récupération des données de façon série. En effet, le message clair et la clé sont envoyés en entrées de l'algorithme bit par bit. Le message chiffré est récupéré en sortie de l'algorithme également bit par bit. Ainsi, le message clair, la clé et le message chiffré ne nécessitent plus que 3 plots contre 208 plots normalement. L'implémentation de cette architecture a été réalisée en suivant le flot de conception décrit ci-dessous.

2.2.1.1 Modélisation et simulation comportementale :

La description des circuits au niveau *Register Transfer Level* – (RTL) ainsi que la simulation comportementale du circuit PRESENT CMOS sont réalisées via ModelSim [169]. Le niveau hiérarchique RTL a été défini en utilisant le langage de programmation VHDL. La Figure 4.6 illustre le fonctionnement du chiffrement PRESENT. En effet, comme il peut être noté sur cette figure, le chiffrement d'un texte clair égal à "FFFF FFFF FFFF FFFF" et d'une clé "0000 0000 0000 0000 0000" conduit à l'obtention d'un texte chiffré "A112 FFC7 2F68 417B" (ou en binaire "1010 0001 0001 0010 1111 1111 1100 0111 0010 1111 0110 1000 0100 0001 0111 1011"), dont la lecture est réalisée de droite à gauche (bit par bit, du MSB vers le LSB, comme illustré sur la Figure 4.6).

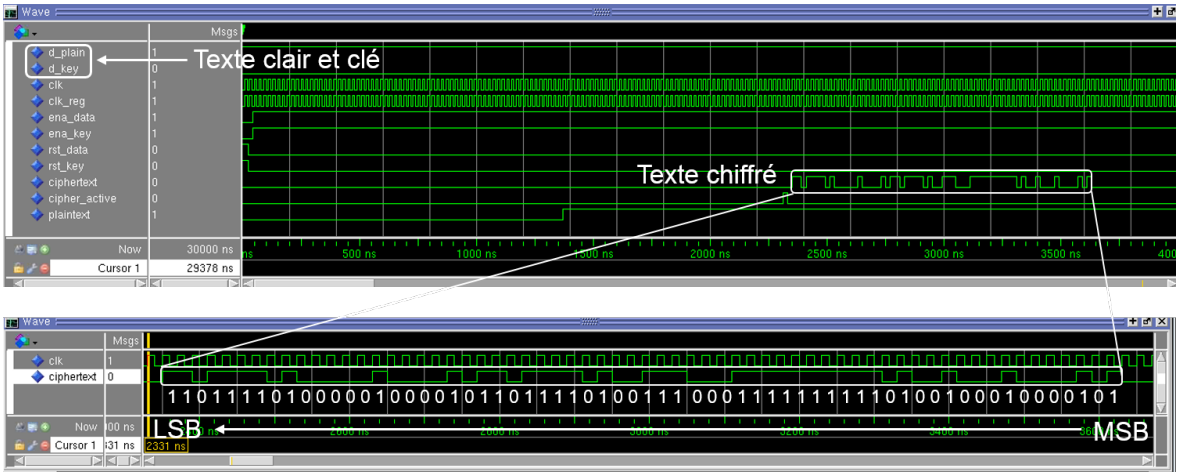


FIGURE 4.6 – Fonctionnement au niveau RTL de l'algorithme de cryptographie PRESENT.

Cette vérification a également été réalisée sur les trois autres vecteurs de test donnés dans [68]. De plus, la mise à zéro de l'algorithme entre deux chiffrements a été validée, permettant ainsi de chiffrer un nouveau message après chaque chiffrement.

2.2.1.2 Synthèse de l'algorithme de cryptographie PRESENT :

Ce fichier VHDL est ensuite synthétisé via Design Vision [170]. Une netlist verilog définissant le circuit au niveau portes logiques est alors créée. De plus, un fichier "*.sdf" ou "Standard Delay Format" est généré. Ce fichier détermine les délais et retards qui sont instanciés entre les différentes portes logiques inter-connectées.

2.2.1.3 Simulation post-synthèse comportementale :

Afin de réaliser les simulations post-synthèse de l'implémentation développée, le test-bench décrit lors de la simulation comportementale du circuit est ré-utilisé sous ModelSim. Cette simulation utilise la netlist synthétisée et le fichier "*.sdf" pour vérifier la fonctionnalité du circuit. De faibles retards sont introduits (environ 290 ps entre l'horloge et le texte chiffré en sortie). Outre les retards introduits par les portes logiques, le résultat de cette simulation est identique à celui illustré sur la Figure 4.6.

2.2.1.4 Simulation électrique du circuit synthétisé :

Le même résultat est obtenu lors de la simulation électrique de cet algorithme de cryptographie PRESENT en technologie CMOS pure, comme représenté sur la Figure 4.7. En effet, de manière analogue à la simulation comportementale, le même texte chiffré résulte du chiffrement des messages clairs et de la clé pour cet algorithme de cryptographie PRESENT.

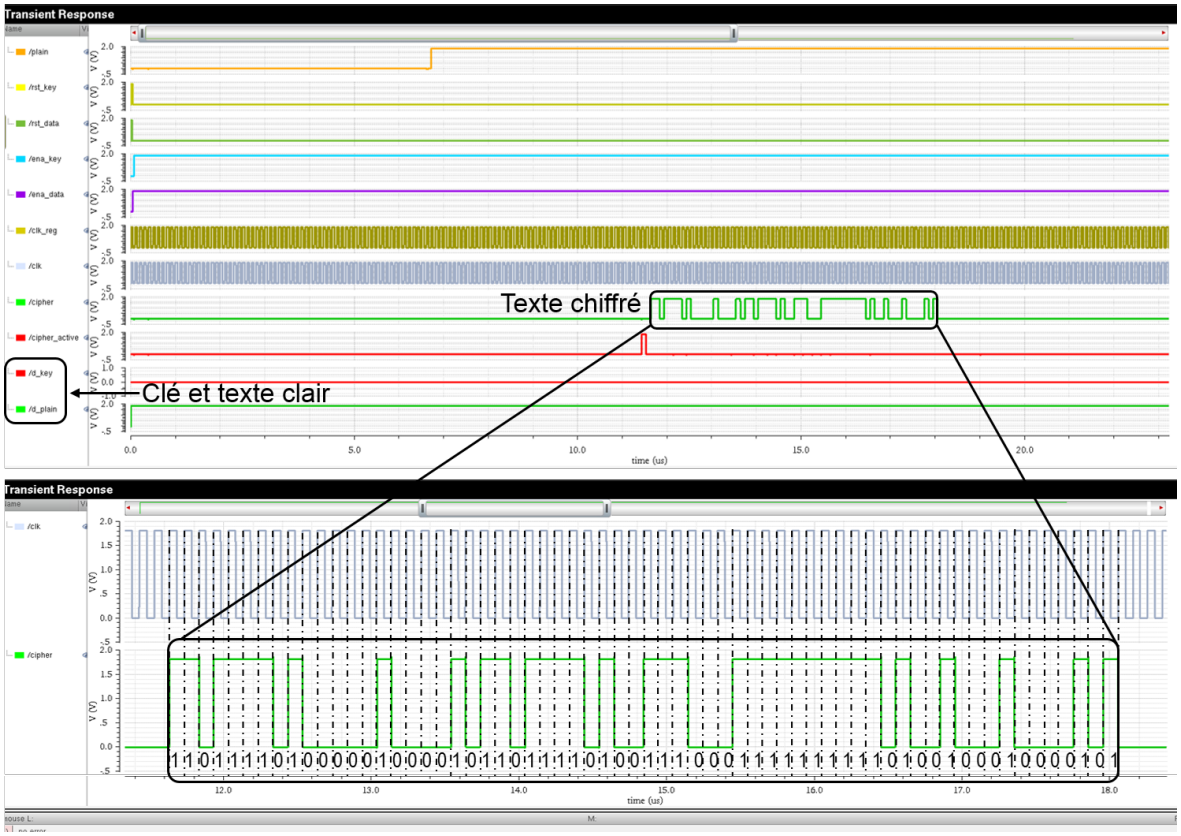


FIGURE 4.7 – Fonctionnement électrique de l'algorithme de cryptographie PRESENT.

2.2.1.5 Placement - Routage :

Après la simulation post-synthèse comportementale et électrique de l'algorithme, la netlist synthétisée est utilisée dans SoC Encounter afin de réaliser l'étape de placement et de routage. Cette étape génère le dessin des masques (ou *layout*) du circuit.

D'une part, le placement consiste à placer de façon optimale les différentes cellules standards utilisées dans la conception du circuit dans le cœur. Cette étape vise à réduire au maximum les connexions entre les cellules standards et ainsi réduire les erreurs et parasites qui peuvent survenir lors du fonctionnement du circuit. D'autre part, le routage connecte les différentes composants.

2.2.1.6 Vérification des règles de dessin :

La phase de vérification physique des circuits intégrés est primordiale afin d'envoyer l'algorithme de cryptographie PRESENT en fabrication. Il existe différents outils de Conception Assistée par Ordinateur – (CAO) comme ASSURA [171] ou encore DIVA, facilitant à l'utilisateur sa phase de vérification. Souvent, les fonderies fournissent au concepteur un fichier technologique permettant de mettre en évidence toutes les règles nécessaires à respecter pour le PDK utilisé. Le fichier fourni peut être compatible avec plusieurs outils de conception.

— Design Rule Check (DRC) : La première vérification à apporter au circuit implémenté est de vérifier que les règles de dessin sont respectées sur le dessin des masques. Cette règle est

nommée le *Design Rule Check* – (DRC). En fonction des machines utilisées lors du processus de fabrication des cellules standards et selon le nœud technologique utilisé, les règles de dessin sont différentes. Comme par exemple, lors de l'étape de gravure primordiale dans le dessin des masques, il est nécessaire de s'assurer qu'aucun court-circuit ou circuit ouvert n'est introduit. C'est pourquoi, ces vérifications sont essentielles. Il existe différents types de règles de dessin comme par exemple : la largeur minimale d'un polygone d'une certaine nature, de l'espacement entre plusieurs rectangles du même niveau de métallisation (les rectangles bleus sur la Figure 4.8.a), surfaces minimales d'une couche (Figure 4.8.b), des violations de type court-circuit (2 lignes de métallisation de même niveau qui se recouvrent, représenté sur la Figure 4.8.c).

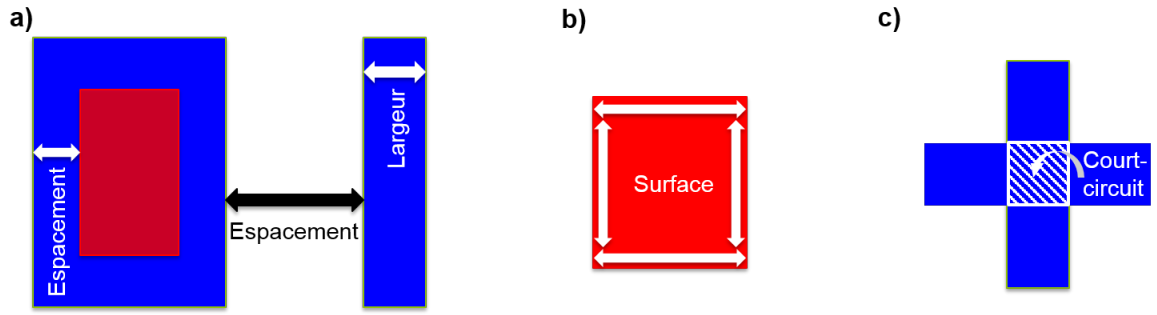


FIGURE 4.8 – Règles de DRC à vérifier sur le dessin des masques. a) espacement et largeur des couches. b) Surface d'une couche. c) Illustration d'un court-circuit dû à eux couches de métallisations qui se couvrent.

— Layout Versus Schematic (LVS) : Le *Layout Versus Schematic* – (LVS) quand à lui consiste à vérifier la correspondance entre le *layout* et le *schematic* réalisé. Cette étape consiste à vérifier que tous les nœuds, instances et pins décrits dans le*schematic* correspondent par la nomenclature et caractéristiques à ceux qui sont décrits sur le *layout*.

— Antennes : Les procédés de fabrication des circuits intégrés peuvent produire des accumulations de charges électriques sur les circuits. Ces charges électriques peuvent détruire les composants mis en jeu. On parle d'effet d'antenne. Le respect de ces règles de dessin permet d'éviter ces effets ou du moins les réduire.

2.2.1.7 Tapeout :

Après le placement/routage, les vérifications physiques du circuit et des simulations électriques de cette primitive de sécurité, nous générons un fichier GDSII (*Graphic Data System*) qui retranscrit le dessin des masques pour l'envoyer au fabricant. Ce format GDSII est un format binaire qui ne peut être lu qu'en utilisant les fichiers de la technologie. Ensuite, le fondeur réalise la vérification physique au niveau masque, *Design For Manufacturing* – (DFM), sur ce fichier. La fabrication des masques peut alors être entreprise et les puces fabriquées.

2.2.2 Résumé du flot de conception

Le flot de conception suivi afin de réaliser l'implémentation matérielle de l'algorithme de cryptographie PRESENT en technologie CMOS pure utilisant le nœud technologique 180 nm est résumé sur la Figure 4.9.

Il est absolument primordial de valider chaque étape de ce flot lors de la conception du circuit avant de passer à l'étape suivante, au risque de concevoir un circuit non-fonctionnel. Ce n'est que lorsque toutes ces étapes sont valides qu'il est possible d'envoyer le circuit en fabrication.

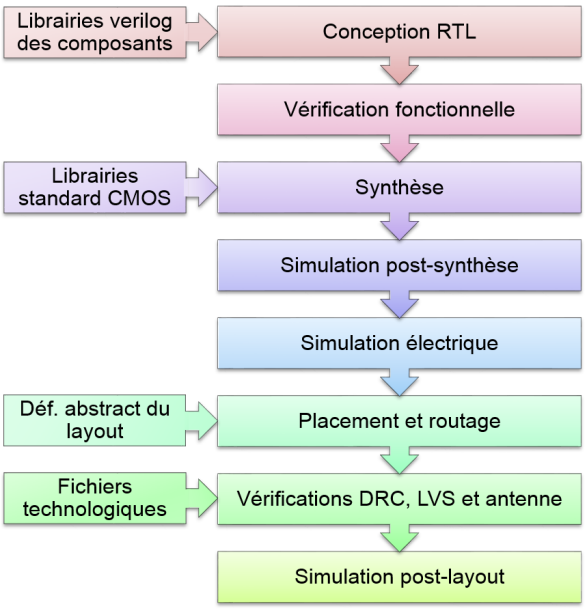


FIGURE 4.9 – Illustration du flot de conception suivi.

2.2.3 Scénario de référence : PRESENT pur CMOS en technologie 180 nm

Cette implémentation de l'algorithme de cryptographie PRESENT en technologie CMOS pure (180 nm) compte 1922 GE pour sa superficie et est illustrée sur la Figure 4.10. La taille de ce chiffrement est plus importante que celle annoncée dans [68] (1500 GE), suite aux trois registres à décalage ajoutés en entrées/sortie du chiffrement pour réduire le nombre de plots.

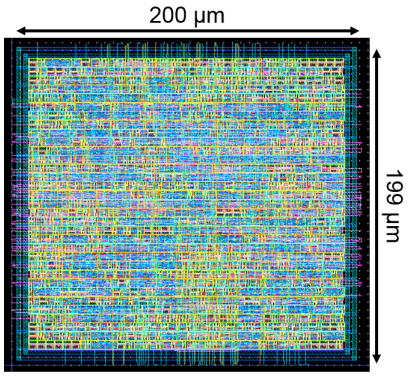


FIGURE 4.10 – Layout de l'implémentation en pur CMOS de l'algorithme de cryptographie PRESENT.

En terme de consommation de veille $P_{\text{statique CMOS}}$, celle-ci est mesurée égale à 78,27 µW et 32,9 nW pour une version avec coupure de l'horloge (*clock gated*). Le principe de *clock gating* consiste à réduire la consommation d'un circuit intégré en désactivant (grâce à un circuit supplémentaire) les arbres d'horloges du circuit, lorsque celui-ci est inactif.

Après l'implémentation de cette version de référence du chiffrement PRESENT, nous nous sommes intéressés au développement des versions hybrides discutées précédemment.

2.3 Flot de conception hybride CMOS/STT-MRAM de l'algorithme de cryptographie PRESENT

Afin de réaliser l'implémentation hybride CMOS/STT-MRAM sur ASIC de l'algorithme de cryptographie PRESENT, le même flot de conception est suivi que celui décrit dans la Section 2.2.1 et

représenté sur la Figure 4.9. La seule modification apportée au flot est la modification des cellules standards appelées lors de la réalisation de la synthèse de l'architecture. En effet, au lieu d'utiliser la librairie standard développée par le fondeur pour les bascules CMOS, nous utilisons la librairie décrivant des bascules non-volatiles aux différents niveaux d'abstraction (verilog, *schematic* et *layout*). Cette librairie non-volatile a été développée dans le cadre du projet GREAT par les différents partenaires.

Pour cela, chaque composante non-volatile est synthétisée indépendamment les unes des autres ("compteur", "FF_chiffrement" et "FF_clé"). Par exemple, le fichier synthétisé du compteur, grâce à la librairie de NVFFs, intègre des connections supplémentaires par rapport à une FF CMOS standard. Ces connections rajoutés sont "Rd" : lecture de l'état des jonctions, "Wr" : permettant leur écriture et "PowerOff" autorisant la désactivation de la bascule. La netlist obtenue lors de la synthèse est alors modifiée pour renommer et connecter convenablement ces différentes connections dans le circuit. Cette manipulation peut être réalisée manuellement ou à l'aide d'un script. Cette deuxième solution a été privilégiée dans le cadre de la conception de ces circuits. L'entité synthétisée du composant "compteur NV" est alors intégrée à la description globale du chiffrement PRESENT. Il en est de même pour les autres structures "NVFF_chiffrement" et "NVFF_clé".

De plus, le placement et routage des algorithmes hybrides sont réalisés grâce à l'appel du fichier définissant le *layout* des composants, dont les composants non-volatiles. Il en est de même pour les fichiers utilisés pour les vérifications du circuit intégré.

C'est ainsi que les architectures hybrides de ce chiffrement PRESENT ont été réalisées.

2.3.1 Scénario #1 : Schéma d'écriture série des NVFFs

2.3.1.1 Besoins en surface :

Compte tenu des besoins en courant pour commuter l'état d'une NVFF pour une STT-MRAM de 200 nm (au moins 1 mA pour programmer une jonction), ce premier scénario illustre le cas où une NVFF est programmée à la fois, c'est-à-dire que les JTMs sont programmées en série deux par deux via 151 coups d'horloges, comme illustré sur la Figure 4.11. En effet, la consommation d'énergie de 151 NVFFs instanciées et programmées en parallèles serait bien trop importante pour une intégration dans des applications à ressources restreintes comme l'Internet des Objets. Ainsi, en plus des registres à décalage ajoutés pour réduire le nombre d'entrées/sorties, il est alors également nécessaire de rajouter une circuiterie pour réaliser cette programmation série des NVFFs.

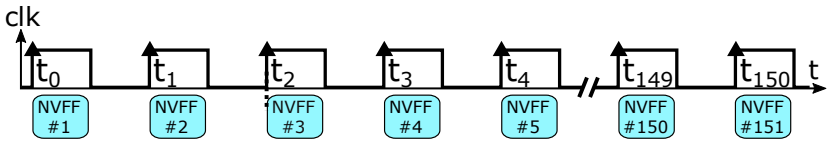


FIGURE 4.11 – Schéma de programmation du premier scénario. Écriture totalement série des NVFFs de l'algorithme de cryptographie PRESENT hybride CMOS/STT-MRAM.

La surface de cette première architecture hybride du chiffrement PRESENT nécessite 18264 GE, ce qui correspond environ à 9 fois la surface de l'implémentation CMOS pure (la référence décrite en sous-section 2.2.3). Cette augmentation importante dans la taille du circuit est dûe aux circuits ajoutés pour réaliser la programmation série et à la taille des bascules NVFFs qui sont plus conséquentes que des FFs standards. La FF standard utilisée dans notre étude équivaut à environ 7 GE. L'évolution de cette FF de volatile à non-volatile induit dans le pire cas, pour une JTM de 200 nm de diamètre, une structure NVFF [167] d'environ 99 GE. En effet, cette bascule intègre de la circuiterie supplémentaire pour la réduction de sa consommation (*power-gating*). De plus, compte tenu de l'énergie importante nécessaire à fournir pour écrire des JTMs de cette taille, l'instanciation de transistors de tailles conséquentes est nécessaire.

2.3.1.2 Consommation :

Outre la surface, le défi majeur des algorithmes de cryptographie est leur consommation, qui doit être réduite dans le cas d'applications embarquées.

Ainsi, l'énergie $E_{repos\ hybride}$ est déterminée égale à 22,22 nJ, pour une fréquence d'horloge de 10 MHz. Le temps de veille minimal t_{veille} pour lequel ce scénario est efficace énergétiquement est de 675 ms. Si le circuit est éteint pendant une période plus longue que cette valeur t_{veille}, l'algorithme hybride CMOS/STT-MRAM est énergétiquement plus efficace que la version CMOS pure du chiffrement.

Le scénario de référence de l'algorithme PRESENT en technologie CMOS pure ainsi que ce premier scénario en technologie hybride CMOS/STT-MRAM ont tous les deux été fabriqués dans le cadre du projet GREAT.

2.3.2 Scénario #2 : schéma d'écriture partiellement série des NVFFs

Le scénario #1 est légèrement amélioré pour de meilleures performances en terme de vitesse et d'efficacité énergétique, cela correspond au scénario #2.

2.3.2.1 Besoins en surface :

Le scénario #2 est basé sur un processus de stockage partiellement parallèle des JTMs composant les NVFFs. Dans ce cas, les NVFFs sont programmées cinq par cinq, comme illustré sur la Figure 4.12. Pour cela, 30 cycles d'horloge sont nécessaires pour la programmation des 151 NVFFs.

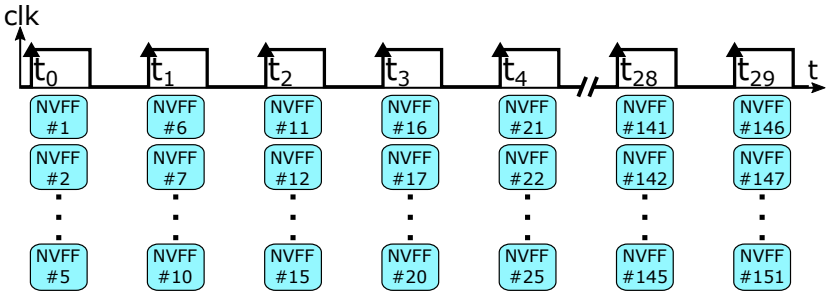


FIGURE 4.12 – Schéma de programmation du second scénario. Écriture partiellement parallèle des NVFFs de l'algorithme de cryptographie PRESENT hybride en technologie CMOS/STT-MRAM.

Cette implémentation hybride CMOS/STT-MRAM du chiffrement PRESENT nécessite l'utilisation de 16382 GE. Ainsi, les besoins en silicium de cette architecture hybride sont 8 fois plus importants que pour une version CMOS pure. Ce second scénario est toutefois plus léger que le premier scénario car les registres nécessaires pour réaliser l'écriture série des NVFFs sont moins nombreux (30 contre 151 registres à décalage dans le premier cas). Les mêmes bascules NVFFs sont utilisées pour réaliser cette fonctionnalité du chiffrement.

2.3.2.2 Consommation :

En terme de consommation, le scénario #2 présente une énergie $E_{repos\ hybride}$ déterminée égale à 1,61 nJ. Ainsi, cette implémentation est 13 fois plus efficace énergétiquement comparée au scénario #1, une réduction due à la diminution du nombre de registres à décalage de 151 à 30.

Pour la même fréquence d'horloge de 10 MHz, le temps de veille minimal t_{veille} pour lequel cette structure hybride de l'algorithme de cryptographie PRESENT est efficace est 49,1 ms (comparé à une structure CMOS pure). Ce temps de veille est 13 fois plus faible que celui déterminé pour le scénario #1.

2.3.3 Résumé des résultats des implémentations réalisées

Cette évaluation des caractéristiques des circuits PRESENT hybrides implémentés en utilisant des technologies matures (technologie CMOS 180 nm et une jonction JTM de 200 nm de diamètre) présente une bonne efficacité énergétique qui doit être compensée par une grande surface silicium, comme résumé dans le tableau 4.3.

Scénarios	Surface du layout en µm²	Surface en GE	t_{veille} minimal en ms
Référence CMOS	39 800	1 922	
#1 (série)	346 896	18 264	675
#2 (partiellement parallèle)	294 294	16 384	49,1

TABLEAU 4.3 – Surface du layout (en μm^2) et en GE de différentes implémentations du chiffrement PRESENT et illustration du temps de veille à partir duquel la structure hybride est énergétiquement plus efficace.

La Figure 4.13 illustre les layouts des trois architectures de l'algorithme de cryptographie PRESENT réalisées en utilisant ce procédé. Les architectures (a), (b) et (c) correspondent respectivement à l'algorithme de cryptographie implémenté en utilisant la technologie pure CMOS bulk 180 nm, la première architecture hybride de l'algorithme PRESENT utilisant un schéma de programmation totalement série (scénario #1) et la seconde version hybride de cet algorithme dont la programmation des NVFFs est réalisée de manière partiellement parallèle (scénario #2). La partie (d) de cette figure met en avant la superposition des trois layouts précédemment décrits afin de pouvoir comparer visuellement la superficie des trois architectures du chiffrement PRESENT réalisées.

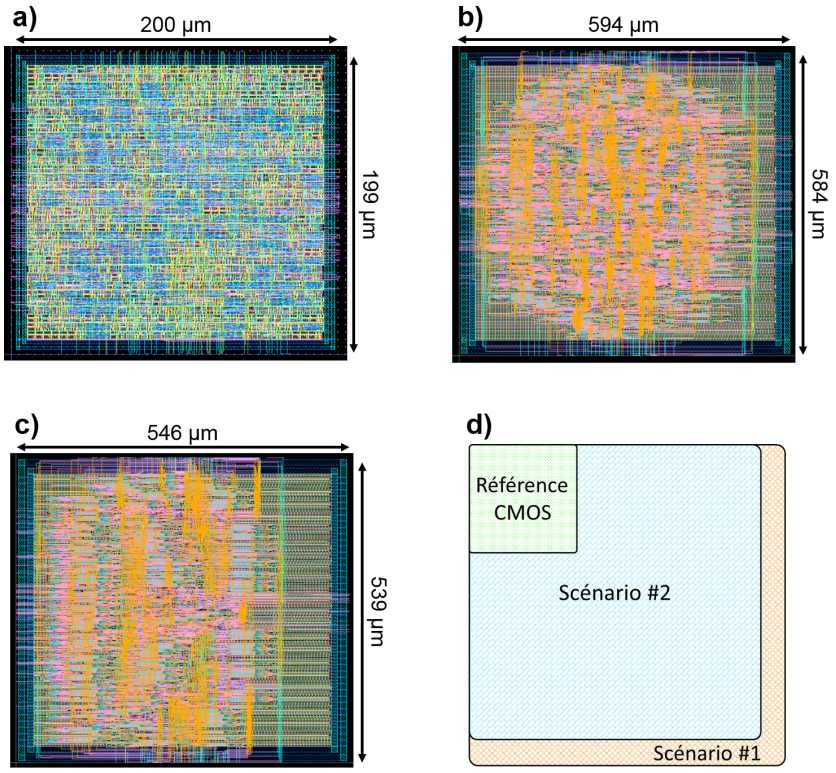


FIGURE 4.13 – Les layouts de différentes implémentations de l'algorithme de cryptographie PRESENT basées sur la technologie CMOS 180 nm et STT-MRAM 200 nm. a) Implémentation en pur CMOS. b) 1er scénario hybride basé sur une programmation totalement série des JTMs. c) 2nde configuration de l'algorithme pour une écriture des jonctions partiellement parallèle. d) Superposition des trois layouts réalisés.

Le Tableau 4.4 valide la métrique GE telle que précédemment définie. Pour cela, les surfaces (en μm^2) des layouts des structures hybrides sont comparées à la surface (en μm^2) du layout du chiffrement pur CMOS implémenté. Il en est de même pour les surfaces GE des algorithmes hybrides réalisés, ils sont comparés à la même référence pure CMOS en GE. Comme il peut être noté ici, les ratios surface du layout et en GE sont très proches, même si le décompte du GE est plus pessimiste que le layout réel de la structure. Cette métrique GE sera utilisée pour l'estimation de la surface d'algorithmes de cryptographie hybrides CMOS/STT-MRAM en utilisant des nœuds technologiques avancés.

Scénarios	Comparaison implémentations hybrides par rapport à la référence CMOS	
	Rapport ($\mu m^2{}_{hybride}/\mu m^2{}_{CMOS}$)	Rapport ($GE_{hybride}/GE_{CMOS}$)
#1 (série)	8,7	9,5
#2 (partiellement parallèle)	7,4	8,5

TABLEAU 4.4 – Rapport de la surface des implémentations hybrides CMOS/STT-MRAM de l'algorithme de cryptographie PRESENT (scénarios #1 et #2) par rapport à son implémentation pure CMOS (le scénario de référence).

Toutefois, avant cette instanciation du chiffrement PRESENT en technologie avancée hybride CMOS/STT-MRAM, il est essentiel de vérifier l'évolution du courant de commutation d'une jonction mémoire en fonction du nœud technologique.

2.3.4 Évolution des densités de courants en fonction du nœud technologique de la STT-MRAM

L'équation de Landau-Lifshitz-Gilbert-Slonczewski (LLG) [172], [173] décrit l'évolution du mouvement de précession d'un moment magnétique $\vec{m}$ en fonction d'un courant (ou tension) qui est appliqué sur l'empilement. La Figure 4.14 illustre les niveaux de courant nécessaires pour commuter une cellule d'un état AP vers l'état P et inversement, en fonction de la taille de la JTM. Ainsi, plus le diamètre de la jonction est réduit, plus le courant nécessaire pour réaliser la commutation est faible. Par exemple, pour réaliser la commutation d'une jonction de 200 nm de AP vers P, un courant minimal de 1 mA est essentiel. Toutefois, pour une JTM de 40 nm, un courant de 46 µA peut être suffisant pour réaliser la commutation de l'orientation magnétique AP vers l'état P [172], [173].

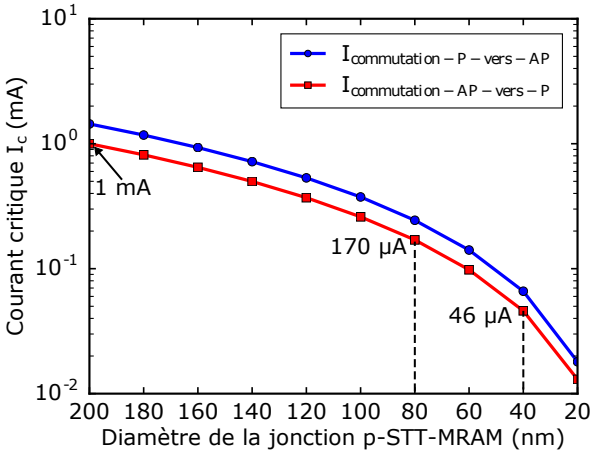


FIGURE 4.14 – Évolution du courant de commutation d'une STT-MRAM en fonction du diamètre de la JTM.

La réduction importante du courant nécessaire à générer dans la JTM pour des nœuds technologiques avancés (40 nm par rapport à 200 nm par exemple) conduit à une réduction drastique de la consommation de la mémoire STT-MRAM. Nous nous sommes alors intéressés aux gains qui

peuvent être atteints sur la consommation et la surface au niveau d'abstraction supérieur, pour une NVFF. Pour ce faire, des simulations de la NVFF proposée dans [97], en utilisant la technologie CMOS bulk 180 nm pour différentes tailles de jonctions, ont été réalisées.

La Figure 4.15 représente les résultats obtenus pour différentes tailles de jonctions. Ces différentes NVFFs ont été développées au niveau *schematic* afin d'en déterminer la consommation et d'estimer la taille de chaque NVFF grâce à la taille des transistors utilisés (en multipliant les W et L des différents transistors). Ainsi, ces résultats suivent la même tendance que ceux obtenus via les équations de LLG. La Figure 4.15.a illustre la réduction de la consommation en écriture d'une bascule hybride avec la réduction du diamètre de jonction. En effet, entre 200 nm et 40 nm de diamètre de jonction, l'énergie nécessaire pour le stockage des données est divisée par un facteur 27.

De plus, la diminution du courant de programmation avec la taille de la jonction, permet de réduire aussi les tailles (le rapport W/L) des transistors. D'où, comme illustré sur la Figure 4.15.b, un gain intéressant en terme de surface consommée par une NVFF. En effet, nous considérons par exemple une jonction de diamètre 200 nm comme référence afin d'observer les gains en surface qui peuvent être atteints. Comme attesté par cette figure, un gain en surface de 87 % peut être satisfait uniquement en réduisant la taille de la jonction de 200 nm à 40 nm. Ainsi, la miniaturisation des procédés de fabrication est essentielle dans le cadre de l'intégration de cette technologie mémoire dans les applications à ressources restreintes.

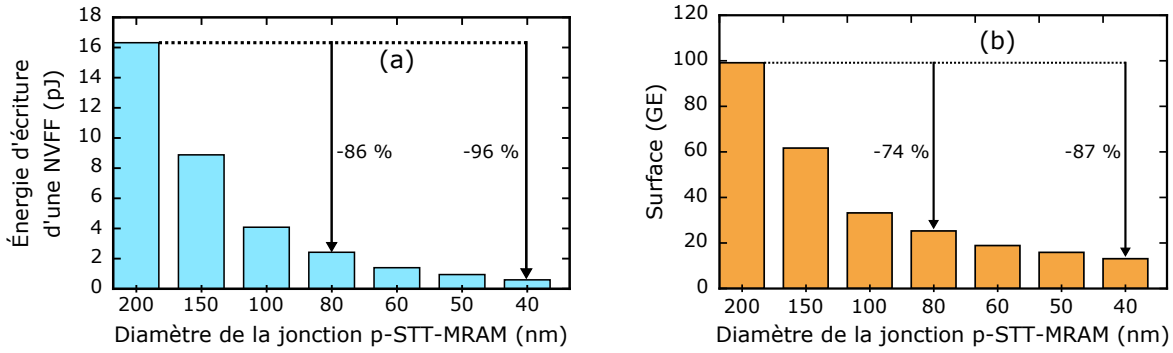


FIGURE 4.15 – Illustration des performances des NVFFs en fonction du diamètre de la JTM. a) Évolution de l'énergie d'écriture de la NVFF en fonction du diamètre de la JTM. b) Surface de la NVFF en fonction du diamètre de la JTM.

2.4 Estimation des performances du chiffrement PRESENT pour les nœuds technologiques avancés

Les équations de LLG ont permis de mettre en évidence une réduction importante de la consommation des STT-MRAMs lors de la miniaturisation du diamètre des JTMs. Des simulations réalisées sur des NVFFs ont permis de souligner des gains intéressants sur la consommation et la taille des circuits suivant la taille des jonctions. En conséquence, il est intéressant d'étudier l'évolution des métriques (consommation/surface) au niveau de l'algorithme de cryptographie PRESENT complet.

Pour cela, les scénarios #3 et #4 de l'algorithme de chiffrement PRESENT utilisent respectivement une technologie CMOS bulk 180 nm et CMOS FD-SOI 28 nm. Pour ces deux scénarios, le diamètre des STT-MRAMs est fixé à 40 nm, ce qui correspond à un nœud technologique atteignable à ce jour.

Avec des JTMs de 40 nm, le courant de programmation d'une jonction est faible (environ une cinquantaine de microampères). C'est pourquoi, contrairement aux scénarios #1 et #2, les sché-

mas de programmation des JTMs sont complètement parallèles pour les scénarios #3 et #4. Les mêmes métriques GE et t_{veille} sont estimées pour les scénarios #3 et #4, afin de les positionner par rapport aux deux scénarios #1 et #2 développés précédemment.

— En terme de surface, d'après les précédentes implémentations, le circuit pur CMOS de l'algorithme cryptographique PRESENT requiert 1922 GE dont 865 GE pour les registres à décalage, les tables de substitution, table de permutation et multiplexeurs (sans comptabiliser les registres : "FF_chiffrement", "FF_clé" et "Compteur"). La surface exigée par les scénarios #3 et #4 est alors estimée en ajoutant aux 865 GE (des circuits qui restent volatiles), la surface des bascules qui sont devenues non-volatiles ("NVFF_chiffrement", "NVFF_clé" et "Compteur NV").

  Cette estimation est complétée en considérant cette fois, la NVFF développée dans [97]. En effet, cette architecture est compatible avec les niveaux de courants de programmation des jonctions de faible diamètre. Dans ce cas, elle est équivalente à environ 13 GE pour la technologie CMOS bulk 180 nm et STT-MRAM de diamètre 40 nm. La taille de cette NVFF peut être réduite à 9,9 GE pour la technologie CMOS FD-SOI 28 nm et STT-MRAM de diamètre 40 nm. Ces deux métriques sont déterminées grâce à la taille des transistors instanciés dans les *schematics* des deux bascules. Ainsi, afin de déterminer la surface de cet algorithme de cryptographie, la surface totale des 151 NVFFs est ajoutée à la surface du circuit qui est maintenue dans un état pur CMOS, soit aux 865 GE composant les couches de substitutions, permutations, registres à décalage et multiplexeurs.

— La consommation d'énergie au cours du processus de stockage est estimée au niveau NVFF, puis calculée en fonction de la configuration de la programmation des jonctions (entièrement parallèle pour les scénarios #3 et #4). Ainsi, l'énergie totale obtenue correspond à l'énergie d'écriture d'une NVFF multipliée par le nombre de NVFFs composant les blocs non-volatiles ("NVFF_chiffrement", "NVFF_clé" et "compteur NV"), c'est-à-dire les 151 NVFFs mises en jeu.

2.4.1 Scénario #3 : technologies bulk CMOS 180 nm et STT-MRAM de diamètre 40 nm

Le scénario #3 représente l'estimation de l'implémentation hybride de l'algorithme de cryptographie en utilisant la technologie CMOS bulk 180 nm et la technologie STT-MRAM de diamètre 40 nm.

2.4.1.1 Besoins en surface :

Comme développé précédemment, la réduction drastique du courant d'écriture des JTMs de diamètre 40 nm (par rapport aux JTMs de 200 nm) permet un processus de programmation complètement parallèle pour stocker et restaurer des données dans les NVFFs. Ainsi, la miniaturisation des STT-MRAMs (et donc des NVFFs) réduit considérablement la surface silicium nécessaire pour réaliser l'implémentation matérielle de ce chiffrement. De plus, les registres à décalage introduits pour les écritures entièrement séries ou partiellement parallèles des JTMs ne sont plus nécessaires. Ainsi, ce scénario ne nécessite plus que 2843 GE. Les besoins en silicium de cette architecture hybride sont donc 1,48 fois plus importants que ceux d'une version CMOS pure de 1922 GE.

2.4.1.2 Consommation :

L'énergie de stockage/restauration de ce troisième scénario du chiffrement hybride CMOS/STT-MRAM de l'algorithme de cryptographie PRESENT est réduite à 90 pJ. Cette métrique est 246 fois plus faible que l'énergie relevée pour le scénario #1.

Ainsi, le temps d'attente minimal t_{veille} pour lequel ce scénario est efficace énergétiquement est de 2,7 ms, soit 250 fois plus petit que le scénario #1. Si le circuit est éteint pendant une période

plus longue que cette valeur t_{veille}, l'algorithme hybride CMOS/STT-MRAM est plus économe en énergie que la version CMOS pure du chiffrement.

2.4.2 Scénario #4 : technologies FD-SOI 28 nm et STT-MRAM de diamètre 40 nm

Le scénario #4 représente l'estimation de l'implémentation hybride de l'algorithme de cryptographie PRESENT en utilisant la technologie CMOS FD-SOI 28 nm et la technologie STT-MRAM de diamètre 40 nm.

2.4.2.1 Besoins en surface :

Comme pour le scénario #3, la miniaturisation des jonctions introduites dans les NVFFs permet de réaliser un processus de stockage et de restauration des données entièrement parallèle.

Ainsi, la superficie de cette version du chiffrement est limitée à 2367 GE, ce qui est compatible avec la barre des 2500 GE recherchée par la cryptographie légère. Les besoins en silicium de cette dernière architecture hybride sont donc 1,23 fois plus importants que ceux d'une version CMOS pure.

2.4.2.2 Consommation :

Outre la superficie réduite de ce quatrième scénario de l'implémentation de l'algorithme PRESENT, l'énergie d'écriture des jonctions dans les NVFFs est également drastiquement réduite. En effet, une énergie de 6,1 pJ est suffisante pour l'écriture des 151 NVFFs. Cette énergie correspond à un temps d'attente minimal de 185 µs à partir duquel la version hybride du chiffrement est énergétiquement plus efficace que son homologue CMOS pur, soit 3648 fois plus réduite que le scénario #1.

2.4.3 Résumé des performances des différents scénarios évalués

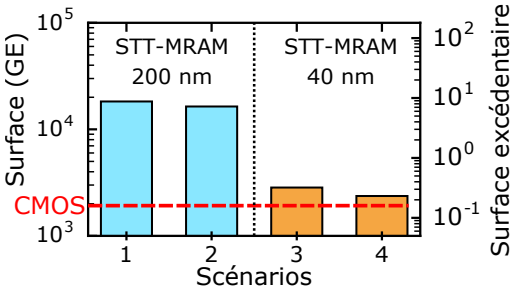
Dans le cadre de l'implémentation matérielle de l'algorithme de cryptographie PRESENT, nous avons proposé dans ce travail quatre structures hybrides CMOS/STT-MRAM du chiffrement. Chaque structure utilise des nœuds technologiques et/ou un schéma de programmation des NVFFs distinct, comme rappelé dans le Tableau 4.5.

	Schémas de programmation des NVFFs			
	Série	Partiellement Parallèle	Parallèles	
Scenarios	**#1**	**#2**	**#3**	**#4**
CMOS (nm)	180	180	180	28
STT-MRAM (nm)	200	200	40	40

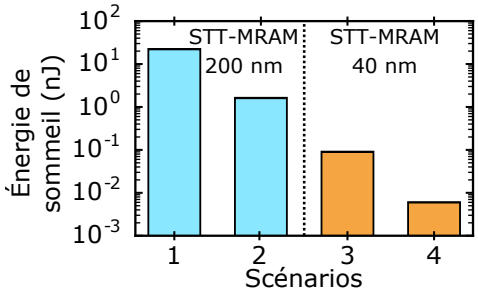
TABLEAU 4.5 – Nœuds technologiques CMOS et STT-MRAM utilisés pour les différents scénarios du chiffrement PRESENT.

Les performances illustrées par ces différents scénarios sont représentées Figure 4.16. La pénalité en surface des scénarios #1 et #2 pour les noeuds technologiques matures est importante où respectivement des facteurs x9 et x8 sont observés. Toutefois, ces derniers restent intéressants en terme de consommation pour des applications "Normally-Off". En effet, ces algorithmes sont plus efficaces en énergie que la version purement CMOS après des temps d'inactivité respectivement de 675 ms et 49 ms, comme illustré sur la Figure 4.16.c. En outre, les scénarios #3 et #4 pour lesquels les diamètres de jonctions sont de 40 nm présentent une pénalité en surface limitée par rapport à l'architecture CMOS pure. Ces structures sont également plus efficaces énergétiquement.

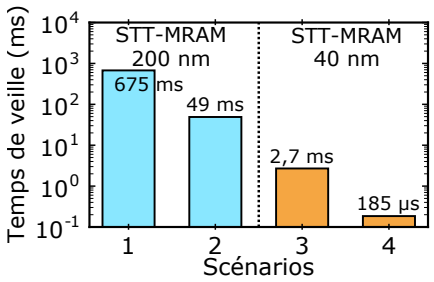
Les temps nécessaires pour réaliser la phase de stockage dans les bascules non-volatiles sont de 15 µs pour le scénario #1 (programmation série), 3 µs pour le scénario #2 (programmation partiellement parallèle) et 100 ns pour les scénarios #3 et #4 (programmation parallèle). Ainsi, les temps de stockage et de restauration des données dans les NVFFs sont inférieurs de plusieurs ordres de grandeur aux temps de veille minimaux à respecter pour atteindre l'efficacité énergétique de l'architecture hybride par rapport à la version CMOS pure.



(a) Évolution de la surface en fonction des différents scénarios développés.



(b) Évolution de l'énergie de sommeil des NVFFs.



(c) Temps de veille à partir duquel la version hybride est énergétiquement plus efficace que la version CMOS pure.

FIGURE 4.16 – Évaluation des performances de différents scénarios de l'algorithme de cryptographie PRESENT.

Ainsi, la conception d'un algorithme de cryptographie PRESENT basé sur des NVFFs peut être une véritable alternative en terme de puissance et de surface pour des nœuds technologiques CMOS et STT-MRAM avancés, par rapport à une architecture CMOS pure. Pour une optimisation maximale de la structure, il est nécessaire d'envisager un système de stockage et restauration entièrement parallèle (diamètre des JTMs inférieur à 80 nm). Toutefois, malgré l'amélioration des performances de ce chiffrement (principalement en terme de consommation), il reste nécessaire de s'assurer du niveau de sécurité de cette nouvelle architecture et des points de vulnérabilités quelle peut présenter face aux attaques physiques de différentes natures : par perturbation ou encore par observation.

3 Durcissement du chiffrement PRESENT hybride

Le durcissement de l'algorithme PRESENT peut être envisagé par exemple par l'intégration d'un capteur capable de détecter les fautes photoélectriques induites sur le CMOS ou thermiques sur la technologie STT-MRAM, comme le DDHP proposé dans le chapitre 3.

Dans cette section, nous proposons de durcir l'algorithme de cryptographie PRESENT de façon intrinsèque à son fonctionnement. Pour cela, une seconde contre-mesure est proposée dans ce chapitre.

3.1 Bascules multi-contextes (MC-NVFF)

La MC-NVFF est un composant mémoire permettant de stocker dans une même bascule plusieurs données dans des jonctions STT-MRAMs instanciées en parallèles [98]. Cette architecture, largement étudiée dans la littérature, a démontré de réels avantages en terme de consommation et de surface, comme rappelé dans le Chapitre 1, Section 4.2.2 (page 30). Dans le cadre du développement de l'algorithme de cryptographie PRESENT en technologie hybride CMOS et STT-MRAM, cette structure peut induire un intérêt majeur dans la sécurité des objets connectés.

Dans ce travail, nous proposons deux structures de bascules compactes MC-NVFFs : une version asymétrique et une version symétrique, afin de les intégrer dans l'implémentation du chiffrement PRESENT. La première architecture est une MC-NVFF asymétrique qui vise à augmenter la résistance de cet algorithme de cryptographie face aux injections de fautes par exemple de type LASER ou électromagnétique. La seconde architecture permet, outre l'injection de fautes, d'inhiber les émissions par canaux cachés du circuit, et ainsi le protéger à la fois des attaques par perturbations et des attaques par observations.

3.1.1 MC-NVFF Asymétrique

La première architecture développée dans ce paragraphe est une MC-NVFF asymétrique, représentée sur la Figure 4.17.

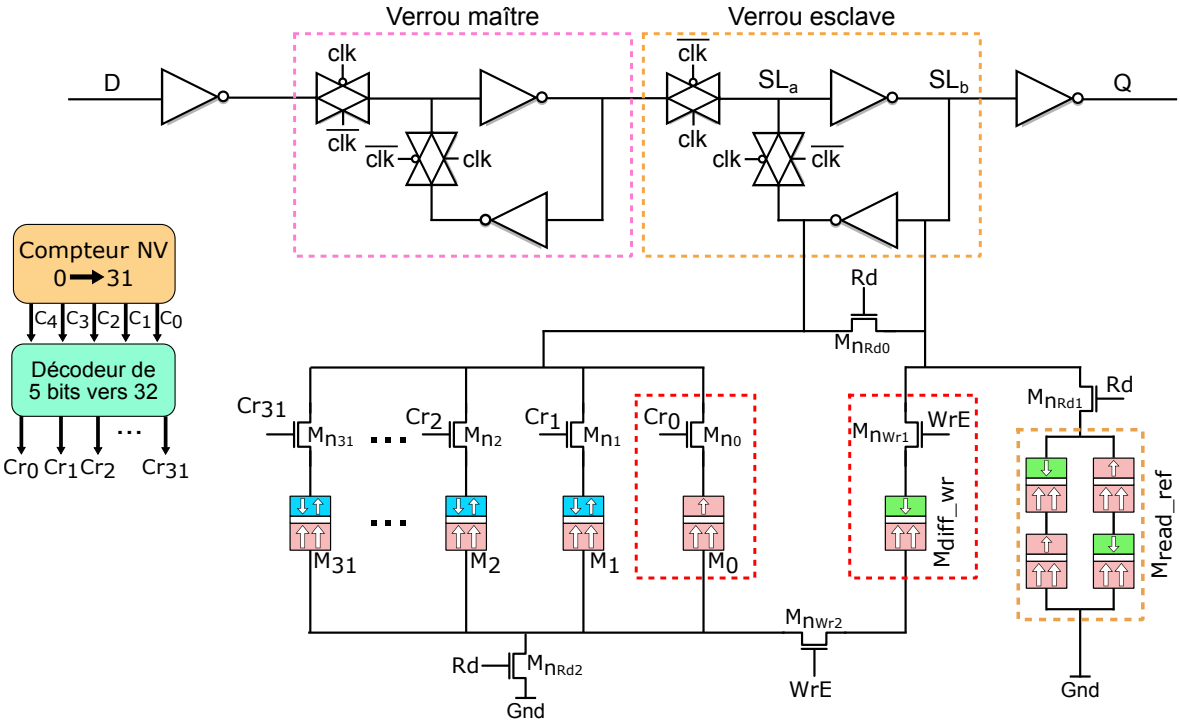


FIGURE 4.17 – Description de la Flip-Flop Multi-contexte MC-NVFF asymétrique.

Cette bascule est basée sur une structure FF CMOS standard, à laquelle est connectée une structure de sauvegarde et de restauration de jonctions de type STT-MRAM. Ainsi, les données en cours de traitement dans le verrou esclave sont stockées dans les JTMs et ces données peuvent être restaurées dans le verrou ultérieurement (après une remise à zéro par exemple). Compte tenu que cette bascule a été réalisée pour une intégration dans le chiffrement PRESENT, alors 32 bascules positionnées en parallèles permettent le stockage des 32 clés et sous-clés nécessaires pour le chiffrement d'un message clair. Cette MC-NVFF fonctionne comme suit :

— Le stockage :
La première phase de fonctionnement de cette bascule consiste à stocker les données en cours de traitement dans la partie CMOS, dans les jonctions magnétiques dédiées. Cette opération est réalisée en activant les transistors contrôlés par le signal d'écriture "WrE", c'est-à-dire les transistors Mn_{nWr1}, Mn_{nWr2}. De plus, les transistors M_{ni} avec i le $(i+1)^{ème}$ tour de chiffrement sont activés via un des signaux de sortie du décodeur 5 bits vers 32 bits, n'activant que la cellule visée, comme illustré sur la Figure 4.17. Afin d'écrire les 32 tours du compteur, 32 jonctions en parallèles notées M_i sont nécessaires.

L'écriture est réalisée en faisant passer un courant dans la jonction à écrire M_i et la jonction d'écriture M_{diff_wr}. Le courant passe pour la première jonction par la couche de stockage (écrivant la JTM dans un état AP) et pour la seconde jonction par la couche de référence (écrivant la JTM dans un état P), ou inversement. Par exemple lors du premier tour de fonctionnement de la bascule et/ou de l'algorithme, lorsque le nœud "SL_b" est au niveau haut 'V_{dd}', alors cette valeur AP est écrite dans la jonction M_{diff_wr} et P dans la jonction M_0, comme illustré sur la Figure 4.17. En effet, un courant est induit respectivement dans le transistor Mn_{nWr1}, dans la M_{diff_wr} par sa couche de stockage, dans le transistor Mn_{nWr2}, dans la jonction M_0 par sa couche de référence et enfin dans le transistor M_{n0}.

— La restauration :
Cette phase consiste à restaurer dans la *latch* esclave les valeurs qui ont été stockées à l'étape précédente. Cette phase restaure au nœud "SL_a" la valeur sauvegardée dans l'une des jonctions M_i. En effet, pour exemple, lors de la restauration du premier tour (M_0) de l'algorithme de cryptographie, le '0' précédemment stocké dans la MC-NVFF est restauré au nœud 'SL_a' et '1' au nœud 'SL_b'. Cette opération est réalisée en comparant la dite jonction (M_i) à une jonction de référence M_{read_ref}. Cette référence est choisie comme la valeur moyenne des deux résistivités R_P et R_{AP}. Effectivement, la résistivité R_{read_ref} de cette référence M_{read_ref} est exprimée par : $(R_{AP} + R_P)/2$. Cette jonction de comparaison est assurée par la définition de 4 jonctions, tel que deux jonctions en série sont dans une aimantation inverse (P et AP), et parallélisées à deux autres jonctions qui sont également dans des aimantations inverses, comme illustré sur la Figure 4.18.

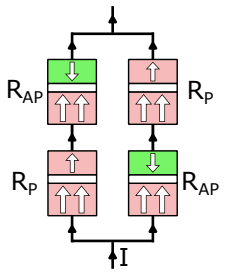


FIGURE 4.18 – Illustration de l'état des jonctions dans la référence.

3.1.2 MC-NVFF Symétrique

La Figure 4.19 représente la déclinaison symétrique de la MC-NVFF. Cette dernière vise un stockage des données dans une paire de jonctions. Cette architecture met en avant une robustesse plus importante face aux attaques par canaux cachés. En effet, le couple d'informations stocké aux nœuds "SL_a" et "SL_b" qui peut être soit (0, 1) soit (1, 0) induira toujours la même signature électromagnétique lors la programmation ou de la lecture des jonctions et rendra ainsi robuste toute structure qui est observée.

D'une part, l'écriture de cette bascule multi-contexte symétrique est réalisée en recopiant dans la jonction activée par le tour du chiffrement, les valeurs contenues dans le verrou esclave. Soit

par exemple, le 3ème cycle de l'algorithme de cryptographie, alors les transistors M_{n2} et M_{n2x} sont activés et les valeurs opposées du verrou "SL_a" et "SL_b" sont stockées respectivement dans les jonctions M_2 et M_{2x}, comme illustré sur la Figure 4.19.

D'autre part, la lecture est simplifiée dans ce cas compte tenu que la comparaison est réalisée entre deux valeurs opposées sauvegardées dans les jonctions visées. Les deux branches du verrou esclave se déchargent plus ou moins rapidement selon la résistivité (l'état magnétique) de la jonction, induisant une restauration du couple de données dans la *latch*.

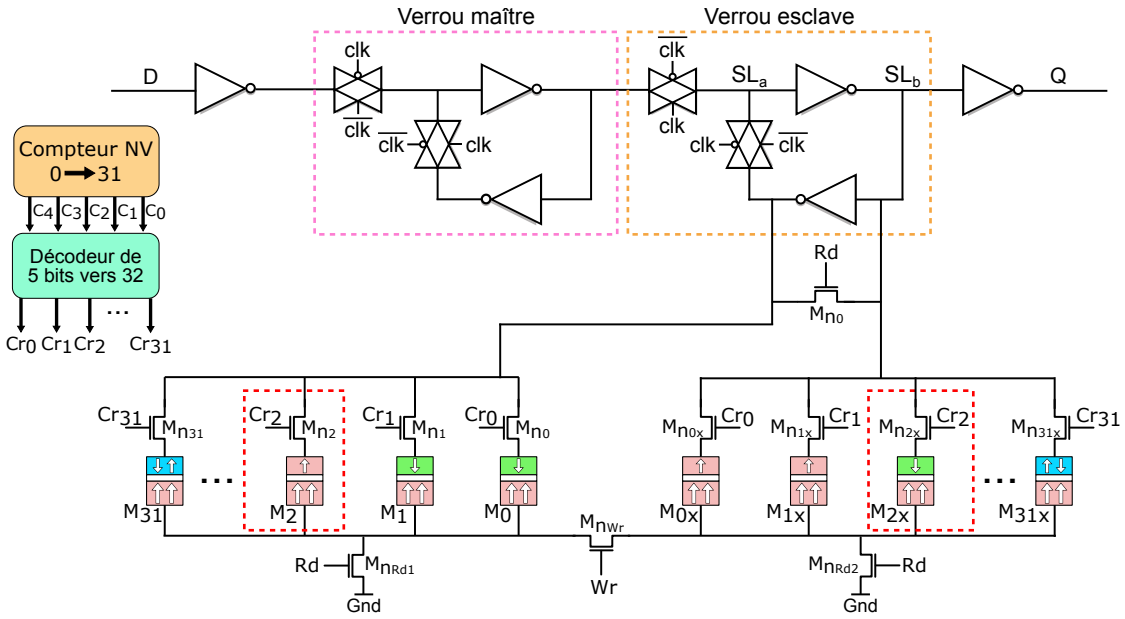


FIGURE 4.19 – Description de la Flip-Flop Multi-contexte symétrique.

3.2 Redondance Technologique Double : DTR

3.2.1 Caractéristiques de la solution DTR

Afin de tenir compte des propriétés qu'apportent les STT-MRAMs aux circuits hybrides, cette section s'intéresse au développement d'une solution matérielle que nous avons proposée, permettant de se prémunir de ces attaques. Cette solution, basée sur de la redondance, est désignée par Double Redondance Technologique ou – *Dual Technology Redundancy* – DTR. Cette contre-mesure consiste à comparer deux chemins d'attaques différents afin de répondre à trois caractéristiques principales :

— C_1 : La première caractéristique consiste à discriminer si une attaque s'est produite sur la partie CMOS ou sur la jonction mémoire STT-MRAM du chiffrement hybride PRESENT.

— C_2 : La seconde phase introduite par cette contre-mesure permet de comparer les résultats stockés dans les STT-MRAMs, aux données en cours de calcul. Ces vérifications sont réalisées aussi bien dans la zone de chiffrement du message que de calcul de la clé. Elles permettent de détecter toute erreur et selon les demandes de l'application, soit restaurer un contexte sain, soit réinitialiser le circuit complet.

— C_3 : La dernière fonctionnalité introduite par la solution proposée consiste à récupérer le dernier état sain sauvegardé, lorsque la caractéristique C_1 a défini que la région attaquée est le CMOS.

Ainsi, la structure durcie proposée dans ce travail afin de diminuer la sensibilité de l'algorithme de cryptographie PRESENT face aux injections de fautes est représentée sur la Figure 4.20.

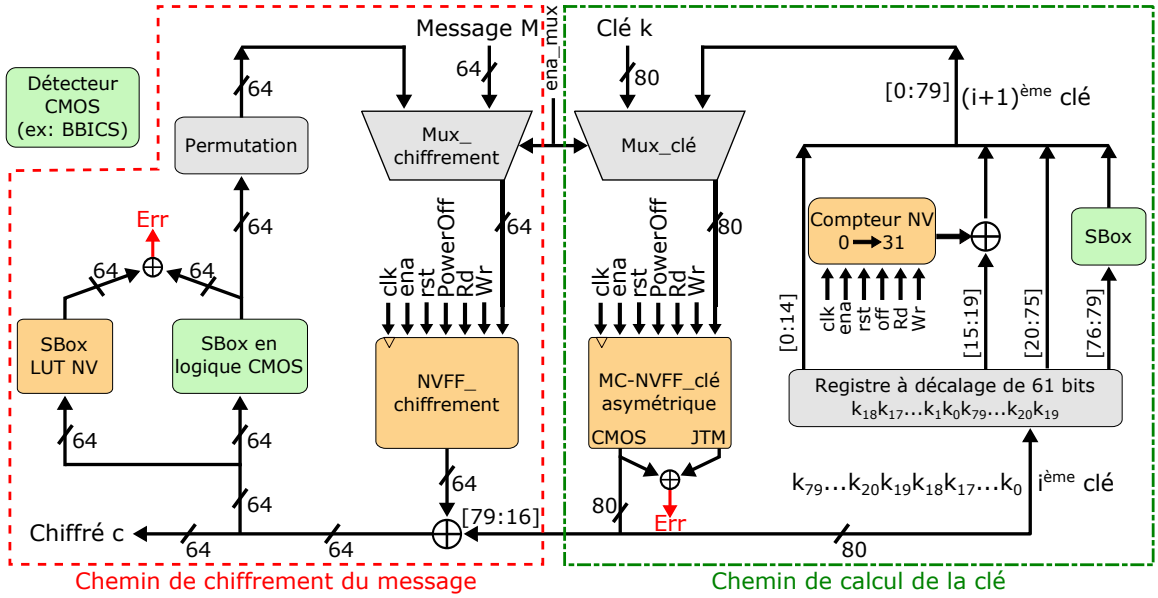


FIGURE 4.20 – Schéma durci de l'algorithme PRESENT hybride pour contrecarrer les injections de fautes.

3.2.2 Mise en œuvre de la contre-mesure DTR

Afin de répondre à la première caractéristique C_1, cette architecture nécessite l'intégration d'un détecteur d'attaques thermiques ou photoélectriques afin de discriminer la cible de l'attaque (CMOS ou STT-MRAM). Dans le cadre de notre étude, le détecteur choisi est le BBICS qui permet la détection des courants injectés dans les substrats de type P et dans les caissons de type N [159] (présenté dans le Chapitre 3, Section 1.3, page 61).

L'exigence C_2 est satisfaite en modifiant la bascule du chemin de calcul des sous-clés ("FF_clé") en une MC-NVFF. Ces MC-NVFFs introduites permettent de comparer localement la clé courante à chaque tour (dans le CMOS) avec celle qui a été stockée dans les STT-MRAMs au préalable, lors de l'initialisation de l'algorithme, comme illustré Figure 4.20. Ainsi, les 80 MC-NVFFs en parallèle permettent le stockage des 32 clés de chiffrement de 80 bits. La version asymétrique de cette bascule est favorisée pour une implémentation dans le chiffrement PRESENT afin de limiter la pénalité en surface silicium de l'algorithme. En effet, alors que la version symétrique nécessite 68 transistors pour réaliser la programmation et la lecture des jonctions, le version asymétrique ne nécessite que 37 transistors pour réaliser ces opérations.

Dans le chemin de chiffrement du texte, des tables de substitution hybrides basées sur des tables de correspondance non-volatiles ou – *Non-Volatile Look-Up Tables* – (NV-LUTs) de STT-MRAMs [174] sont introduites, comme illustré Figure 4.20. Ces NV-LUTs ont pour principe de sauvegarder une table de vérité dans des STT-MRAMs et de lire les données selon l'entrée sélectionnée. Ainsi, 16 tableaux non-volatiles de 4 bits sont intégrés dans le fonctionnement de cet algorithme. Pour chaque tableau, 16 JTMs sont instanciées et chaque élément mémoire est adressé grâce à un décodeur et un arbre de transistors NMOS, comme illustré Figure 4.21. Les valeurs de la table de substitution sont stockées dans ces NV-LUTs pendant la phase d'initialisation.

Cette opération d'initialisation consiste en l'écriture des tableaux de correspondance des NV-LUTs et du stockage des 32 tours de clés, avant le chiffrement du premier message clair. Lors du chiffrement, les comparaisons sont réalisées sur les deux points :

— Dans le chemin du chiffrement du message, les valeurs stockées dans les NV-LUTs sont comparées aux valeurs en cours de traitement dans les tables de substitution qui sont implémentées en CMOS pur (basées sur des portes logiques).

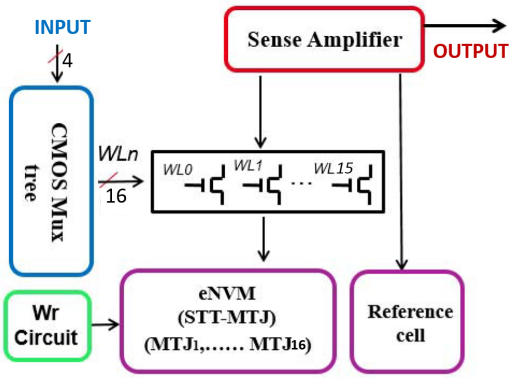


FIGURE 4.21 – Représentation d'une NV-LUT 4 bits [174].

— Dans le chemin de calcul de la clé, les clés préalablement stockées dans les STT-MRAMs sont comparées aux clés en cours de calcul.

Afin de satisfaire l'exigence C_3, la fonctionnalité de certains composants est modifiée, par leur évolution de volatiles à non-volatiles. En effet, afin de répondre à cette exigence de restauration d'un contexte sain, l'emploi de NVFFs est nécessaire. Ces dites bascules non-volatiles remplacent les bascules de chiffrement notées "FF_chiffrement" et du compteur utilisé pour la mise à jour de la clé.

3.2.3 Fonctionnement de la solution DTR

Pour résumer, lors du fonctionnement normal de l'algorithme de cryptographie, des comparaisons sont réalisées entre le chemin CMOS où le chiffrement est en cours de calcul et les données stockées dans les STT-MRAMs. La Figure 4.22 représente les opérations qui sont réalisées en fonction des différents cas (attaque ou non).

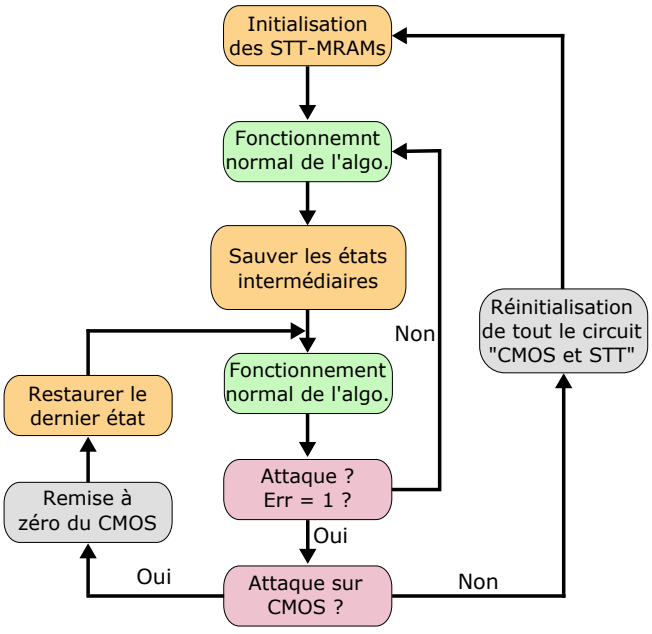


FIGURE 4.22 – Représentation des opérations réalisées sur cette architecture hybride durcie en cas d'attaque.

Tant que les données sur les deux chemins sont identiques, aucune erreur n'est induite sur les portes "XOR" de comparaison des deux chemins CMOS et hybride. Le circuit poursuit son opération. Toutefois, dès la modification de l'un de ces chemins, l'erreur est relevée. L'algorithme se doit alors de déterminer quel chemin a été corrompu.

Le CMOS est alors vérifié grâce au détecteur BBICS. Si c'est bien le CMOS qui a été attaqué, alors la circuiterie CMOS est réinitialisée et les derniers états sains stockés dans les STT-MRAMs sont restaurés dans les bascules. Dans le cas inverse si l'attaque a été induite sur la technologie STT-MRAM, alors tout le circuit est réinitialisé (le CMOS et les mémoires). Une nouvelle initialisation des jonctions est alors nécessaire (des NV-LUTs et des MC-NVFFs).

3.2.4 Injection de fautes

Dans le cadre de l'intégration de la MC-NVFF pour l'achèvement de la caractéristique C_2 dans la contremesure DTR, la MC-NVFF asymétrique est modifiée afin de pouvoir réaliser la comparaison des deux chemins CMOS et hybride de façon intrinsèque à la bascule. Pour cela, seul le verrou esclave a été dupliqué afin de limiter la surface nécessaire à l'implémentation de cette bascule, comme illustré sur la Figure 4.23.

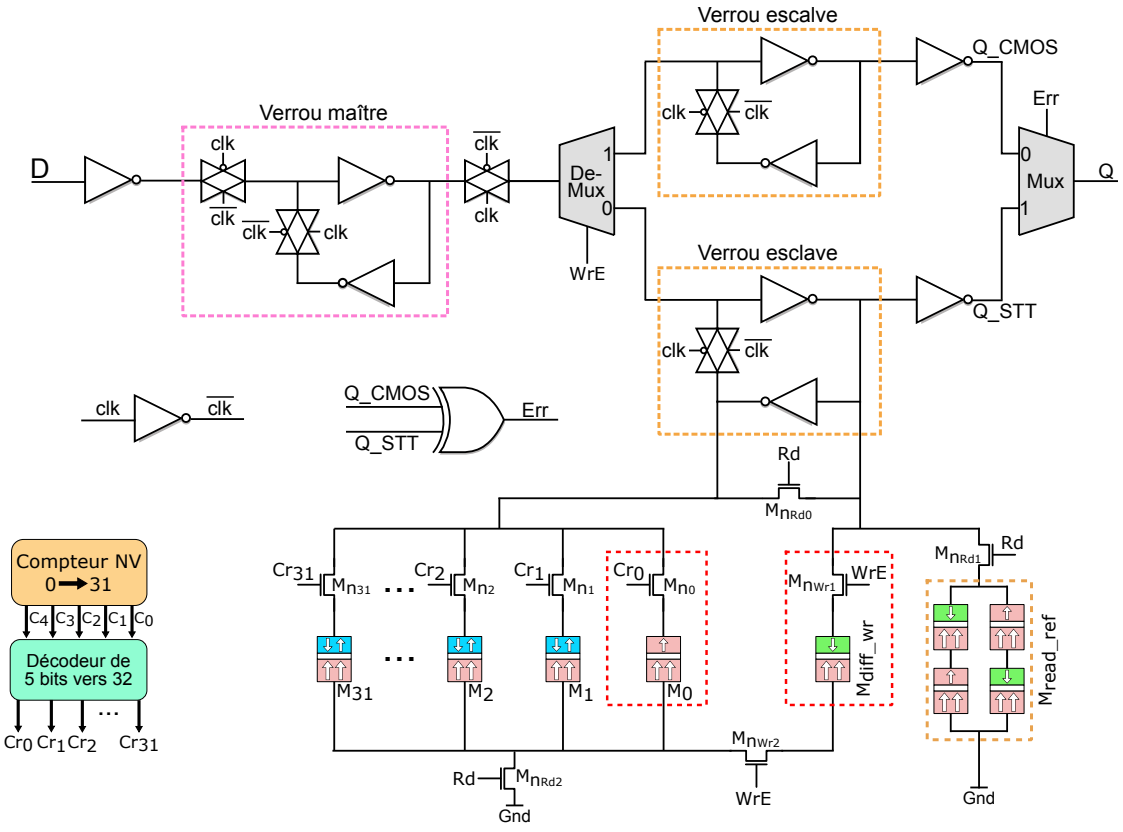


FIGURE 4.23 – Description de la Flip-Flop Multi-contexte MC-NVFF asymétrique développée pour contrecarrer les injections de fautes.

Deux composants ont été ajoutés à cette architecture, un démultiplexeur et un multiplexeur. Pendant la phase de programmation (définie par l'activation du signal 'WrE') et d'initialisation des STT-MRAMs, les données dans le verrou maître sont transmises directement au verrou esclave hybride. En dehors de ces phases de programmation ou d'initialisation, cette MC-NVFF fonctionne de façon conventionnelle en technologie CMOS pure.

La sortie de ce verrou esclave CMOS pur est comparée à celle du verrou esclave hybride. Tant que ces valeurs sont identiques, le nœud "Err" reste à '0', la valeur de sortie du verrou CMOS est

transmise aux étapes suivantes. Si les deux sorties des verrous esclaves ("Q_CMOS" et "Q_STT") sont différentes, le nœud "Err" bascule à '1', alors les valeurs générées en sortie de la MC-NVFF sont les valeurs en sortie du verrou hybride, tant que la caractéristique C_1 démontre que l'attaque est survenue sur le CMOS et non sur les STT-MRAMs.

Cette structure durcie de MC-NVFF est simulée afin de vérifier le bon fonctionnement des différentes étapes de la bascule, la sauvegarde des différents tours de compteurs et leur restauration dans le verrou esclave. Soient les signaux c_4, c_3, c_2, c_1 et c_0 correspondants aux 5 bits du compteur de cycle de l'algorithme, tel que c_4 soit le bit de poids fort et c_0 le bit de poids faible. Comme illustré sur la Figure 4.24, lors de l'activation du signal 'WrE' pour l'écriture des 32 jonctions de la MC-NVFF, la valeur traitée par le verrou esclave est sauvegardée dans la jonction visée. En effet, tant que la sortie 'Q_STT' est à '0', l'activation successive des jonctions M_0 à M_{19} les programme dans un état P, correspondant à l'état logique '0'. Après le basculement en sortie du verrou esclave du nœud 'Q_STT', les dernières jonctions (de M_{20} à M_{31}) sont programmées dans un état AP, état logique '1'. Après cette phase de vérification de la bonne écriture des 32 jonctions de cette bascule, sont réalisées des lectures occasionnelles pour restaurer les valeurs préalablement stockées. Cette lecture est réalisées pour les $16^{ème}$ (compteur à 15 et M_{15} dans un état '0'), $25^{ème}$ (compteur à 24 et M_{24} dans un état '1') et $32^{ème}$ (compteur à 31 et M_{31} dans un état '0') cycles. Ils sont bien rétablis aux nœuds du verrou esclave et donc à fortiori sur la sortie de cette bascule non-volatile 'Q_STT'.

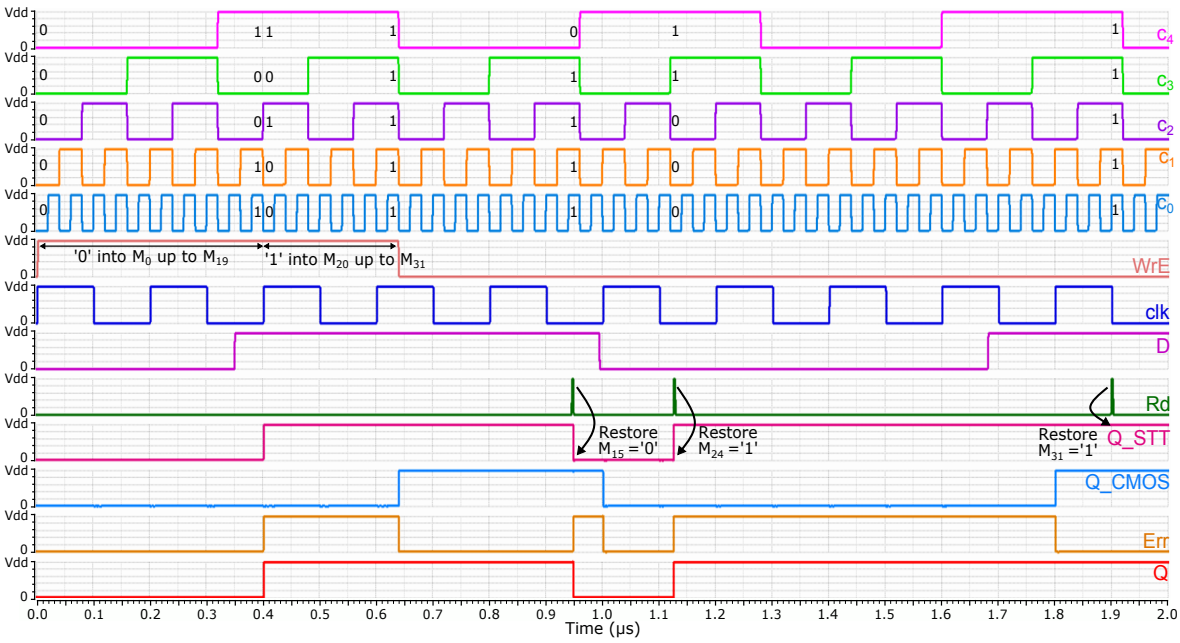


FIGURE 4.24 – Description de la Flip-Flop Multi-contexte MC-NVFF asymétrique développée pour contrecarrer les injections de fautes.

De plus, la Figure 4.24 met en avant le bon fonctionnement de la comparaison des deux sorties des bascules CMOS et hybride, notée 'Err', qui se déclenche dès que les données traitées par les deux verrous sont distinctes. De plus, la sortie "réelle" de la MC-NVFF désignée par 'Q' correspond bien aux valeurs contenues dans le verrou CMOS tant que le signal "Err" ne se déclenche pas et à celles contenues dans le verrou hybride CMOS/STT-MRAM dès que le drapeau détecte une attaque. Ce signal "Err" est couplé au signal "*Flag*" du BBICS.

Outre les injections de fautes, l'utilisation de bascules multi-contextes symétriques permettrait également de se prémunir des attaques par canaux cachés, puisque le circuit traitera toujours une donnée logique et son inverse sur un même verrou dans les jonctions mémoires (lors du stockage et de la restauration). Ainsi, cette architecture serait aussi robuste contre les injections de fautes que les attaques par canaux auxiliaires.

3.2.5 Proposition complémentaire pour une architecture PRESENT durcie

Afin de réduire la taille de l'implémentation de l'algorithme de cryptographie PRESENT hybride CMOS/STT-MRAM durci en utilisant la méthode DTR, nous proposons de remplacer le chemin des NV-LUTs par un chemin basé sur une *SBox* inverse, comme représenté Figure 4.25. Cette dernière est déjà présente dans le système afin de réaliser le décryptage des textes chiffrés . Ainsi, cette ressource serait utilisée aussi bien pour le chiffrement du texte clair par la clé de chiffrement, que pour le décryptage de ce message. Cette solution permettrait donc de réduire la surface d'implémentation de la structure en préservant son durcissement compte tenu que la redondance est maintenue sur les deux zones sensibles, le chemin de chiffrement du message et du chemin de calcul de la clé, comme illustré sur la Figure 4.25.

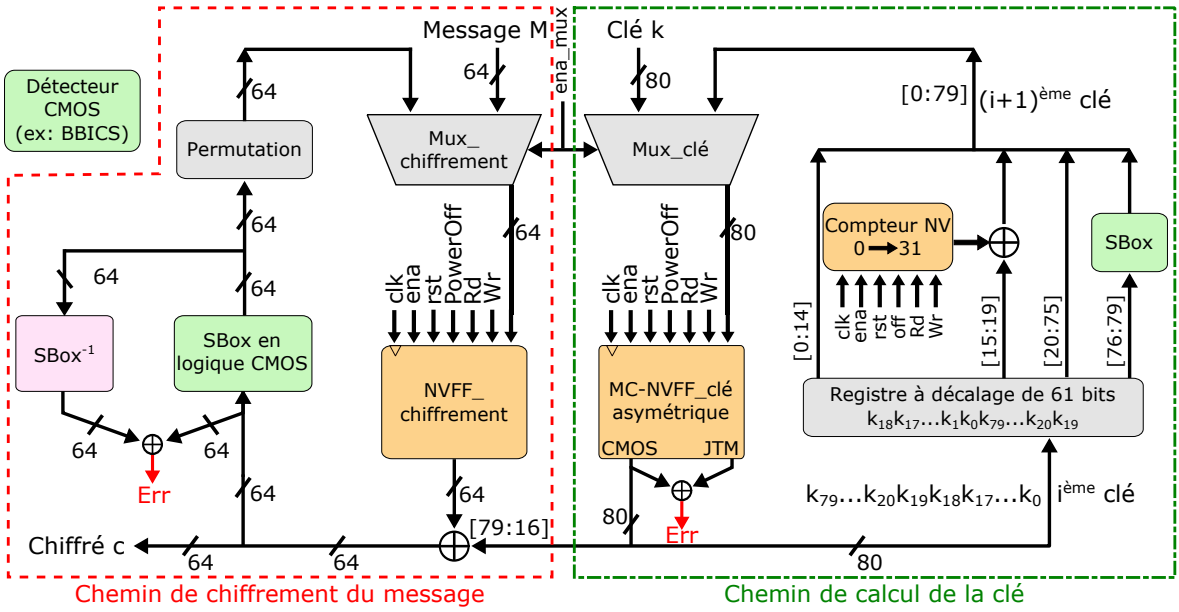


FIGURE 4.25 – Représentation de l'algorithme de cryptographie PRESENT durci en utilisant la redondance.

4 Mise en place du banc de test des circuits intégrés

Dans le but de valider les circuits développés, le chiffrement PRESENT en technologie pure CMOS, le chiffrement CMOS/STT-MRAM de PRESENT (scénario #1) ainsi que le capteur DDHP ont été fabriqués sur silicium. Dans ce chapitre, la mise en place d'une plateforme de test pour mener des attaques physiques sur les dispositifs sera présentée.

4.1 Puce fabriquée

Sur les 25 mm^2 de silicium de la puce électronique, le projet lié aux travaux de thèse utilise environ 6,25 mm^2. Pour l'instanciation des signaux nécessaires au fonctionnement de ces trois circuits, 38 pads sont utilisés sur les 44 disponibles (divisés en deux barrettes de 22 plots), comme illustré sur la Figure 4.26. Le PRESENT CMOS (#1 sur la Figure 4.26) utilise 12 pads, le PRESENT hybride (#2 sur la Figure 4.26) occupe 16 pads et le capteur DDHP (#3 sur la Figure 4.26) utilise 10 pads. Le projet a été implémenté dans le coin haut droit du circuit.

Ces circuits sont conçus en utilisant le nœud technologique CMOS 180 nm ce qui implique des niveaux de tensions de 1,8 V afin de préserver l'oxyde des transistors MOS et des MRAMs.

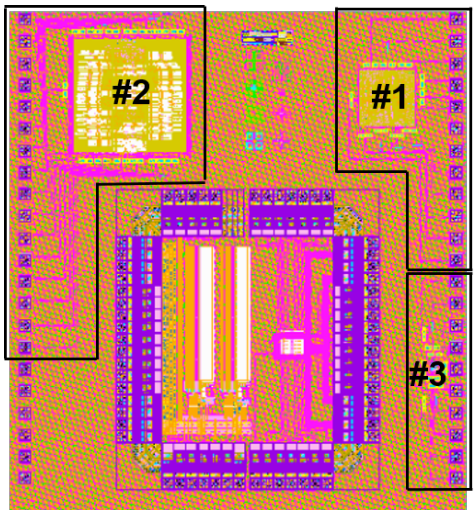
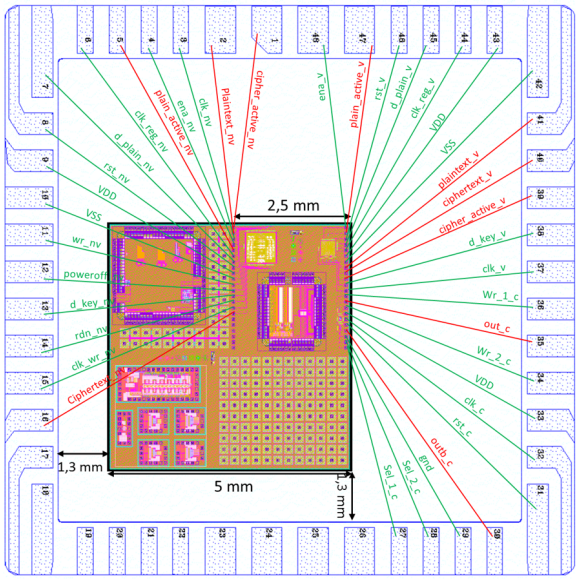


FIGURE 4.26 – Photographie des trois architectures (PRESENT CMOS #1, PRESENT hybride #2 et capteur #3) implémentées dans la puce produite sur le projet européen GREAT.

4.2 Mise en boîtier

Le choix a été fait de mettre les puces électroniques en boîtier pour permettre les attaques physiques par illuminations laser ou radiations électromagnétiques. L'association Micro-PackS, qui dispose du laser Nd-YAG utilisé durant la thèse, a également les ressources nécessaires pour la mise en boîtier des puces électroniques. Nous avons donc utilisé les boîtiers disponibles dans leur catalogue selon les spécifications liées au projet. En considérant le nombre de pads, la surface de la puce et la nécessité de centrer le projet par rapport au boîtier, le boîtier *Dual In-Line* – (DIL) 48 céramique a été retenu, comme illustré sur la Figure 4.27.



(a) Illustration de l'emplacement du circuit dans la cavité du DIL 48 .

(b) Photographie d'un circuit mis en boitier.

FIGURE 4.27 – Vue de la cavité du DIL 48 avec le circuit connecté par des fils de câblage en Or de 25 µm.

La mise en boîtier a été réalisée manuellement. Une fois la colle conductrice déposée au fond de la cavité du DIL 48, la puce est placée délicatement dessus en respectant le centrage visible sur la Figure 4.27.a. Pour fixer la colle, le boîtier monté passe 25 minutes à 170 °C dans une étuve.

Finalement, les 38 pads sont reliés aux pins respectives du boîtier DIL 48 par l'intermédiaire de fils de câblage d'or de 25 µm de diamètre. 10 pins sont laissées libres et seront reliées à la masse ultérieurement. Le résultat de cette mise en boitier est illustré sur la Figure 4.27.b.

4.3 Plateforme de test

Les trois composants implémentés ont de nombreux signaux entrants et sortants. L'objectif de la plateforme de test est donc de pouvoir interpréter tous ces signaux de contrôle (en vert sur la Figure 4.27.a), de façon synchronisée. Les signaux de sortie (en rouge sur la Figure 4.27.a) sont lus sur un oscilloscope afin d'analyser le fonctionnement de chaque composant.

L'architecture matérielle proposée est basée sur une carte de prototypage équipée d'un microcontrôleur et un circuit imprimé faisant l'interface entre la carte mère et le dispositif encapsulé en DIL 48. La première carte est étiquetée "carte mère" et la seconde "carte fille".

4.3.1 Carte mère

Une étude de marché a été réalisée afin de déterminer la carte mère à instancier pour le test des puces développées. Pour cela, différentes solutions qui se trouvent dans la gamme des besoins du projet ont été envisagées : l'Arduino Due [175], l'Arduino Mega 2560 [176], la carte de prototypage STM32F446RE Nucleo-64 [177] et un FPGA Zybo Z7 [178]. Comme il peut être noté sur le Tableau 4.6, les deux meilleurs candidats sont l'Arduino Due et la STM32F446 Nucleo-64. Toutefois, afin de minimiser le temps de développement de la plateforme, le choix de la carte mère a été orienté vers la famille des cartes de prototypages Arduino [179] (illustré sur la Figure 4.28). Ces outils matériels sont très puissants. Ils embarquent des microcontrôleurs Atmel Microchip et proposent un environnement de développement simplifié. Les entrées/sorties du microcontrôleur sont accessibles facilement par l'intermédiaire de broches femelles sur lesquelles il est possible de connecter tout type de composants ou de circuits imprimés.

	Exigences du projet	Cartes de prototypage candidates			
		Arduino Due	Arduino Mega 2560	STM32F446 Nucleo-64	FPGA Zybo Z7
Sorties PWM	**7**	12	15	18	40
Entrées/Sorties logiques	**28**	40	37	28	
Sorties analogiques	**1**	2	0	2	0
Fréquence d'horloge	**~ 10 MHz**	Jusqu'à 84 MHz	Jusqu'à 16 MHz	Jusqu'à 180 MHz	Jusqu'à 125 MHz
Niveau logique	**1,8 V**	3,3 V	5 V	3,3 V	3,3 V
Encombrement spatial	**Le plus faible possible**	101,52 mm x 53,3 mm	101,52 mm x 53,3 mm	80 mm x 70 mm	142 mm x 122 mm

TABLEAU 4.6 – Illustration des cartes de prototypage candidates et des exigences du projet.

4.3.2 Carte fille

Il n'est pas possible de connecter directement le dispositif encapsulé en boîtier DIL 48 sur la carte mère. C'est pourquoi, il est nécessaire de fabriquer une carte intermédiaire qui permettra l'adaptation électrique et mécanique entre les deux.

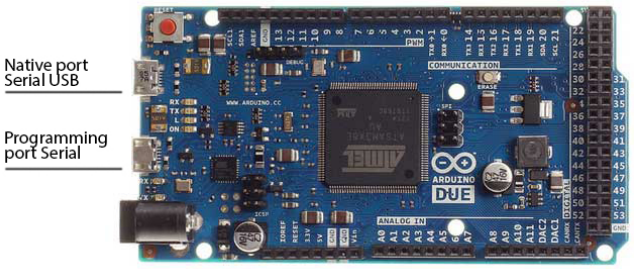


FIGURE 4.28 – Illustration d'une carte Arduino Due [175].

4.3.2.1 Adaptation mécanique :

La plateforme sera utilisée pour tester plusieurs échantillons, il faut donc un système mécanique qui permette le retrait rapide des échantillons sous test, sans les endommager. Il existe pour cela des supports *Zero Insertion Force* – (ZIF) adaptés aux boîtiers DIL 48. Le support ZIF lui-même est soudé sur la carte fille qui se connecte à la carte mère par les broches prévues à cet effet.

4.3.2.2 Adaptation électrique :

Le microcontrôleur SAM3X8E [180] de l'Arduino Due est alimenté en 3,3 V. Ce niveau de tension n'est pas compatible avec le niveau de tension de la logique utilisée, qui est de 1,8 V. Pour réaliser cette conversion, un régulateur de tension linéaire STMicroelectronics LD1117AS18TR [181] est utilisé. Il existe également des régulateurs à découpage qui ont un meilleur rendement mais du fait de leur fonctionnement en commutation, rejettent beaucoup d'ondes électromagnétiques parasites. Ainsi, le 3,3 V est fourni par la carte mère (Arduino Due) et le 1,8 V fourni par le régulateur de tension LD1117AS18TR. Toutefois, il reste à convier ce niveau de tension au DIL 48.

Pour réaliser cette opération, un émetteur-récepteur (*transceiver*) logique est utilisé. Le SN74AVC20T245 de Texas Instruments [182] est instancié dans la carte fille. Ce dernier dispose de deux rails d'alimentation, l'un en 3,3 V et l'autre en 1,8 V. De façon matérielle, il est possible de configurer les voies dans un sens ou dans l'autre. De cette manière les sorties du DIL 48 en 1,8 V sont rehaussées en 3,3 V avant d'être envoyées au microcontrôleur. Inversement, les sorties du microcontrôleur en 3,3 V sont abaissées à 1,8 V avant d'être envoyées au DIL 48. Cet émetteur-récepteur peut gérer un maximum de 20 voies simultanément. Cette dernière considération oblige l'utilisation de deux de ces composants afin d'augmenter ou diminuer les niveaux de tension des 34 signaux digitaux à contrôler pour les 3 circuits.

4.3.2.3 Conception et fabrication :

La carte fille reprend le layout de la carte mère pour la forme de la carte et l'emplacement des broches femelles tout autour du microcontrôleur. Ainsi la carte fille sera équipée de broches mâles pour connecter les deux cartes ensemble. Une fois les composants choisis et le schéma électrique validé, le layout est réalisé à l'aide du logiciel de CAO DesignSpark PCB [183]. Le routage est réalisé manuellement sur un circuit imprimé quatre couches de cuivre afin de respecter les règles de Compatibilité ÉlectroMagnétique – (CEM). La première couche est routée horizontalement, la couche inférieure est réservée au plan de masse et ne contient aucune piste. La troisième couche est réservée aux alimentations (3,3 V et 1,8 V) et enfin la couche la plus basse est routée verticalement. Cet empilement limite les couplages capacitifs et inductifs qui peuvent induire des courants parasites.

La fabrication de la carte fille a été sous-traitée à la société Beta Layout. Le montage des composants a été réalisé au sein du laboratoire. Le résultat est photographié sur la Figure 4.29.

La Figure 4.30 illustre le résultat de la carte fille montée sur la carte mère.

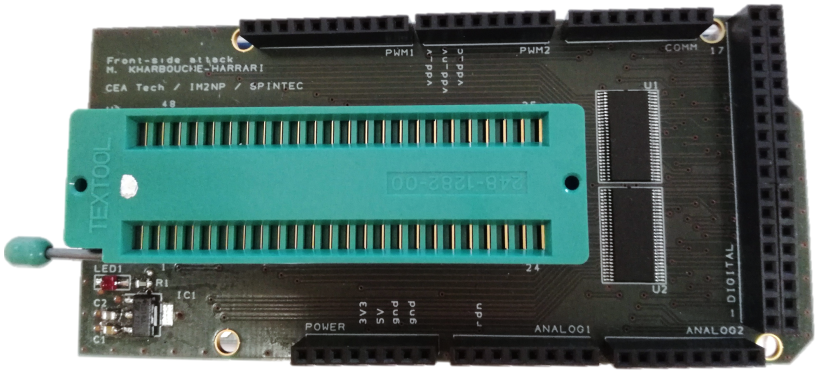


FIGURE 4.29 – Vue de la partie supérieure de la carte fille montée.

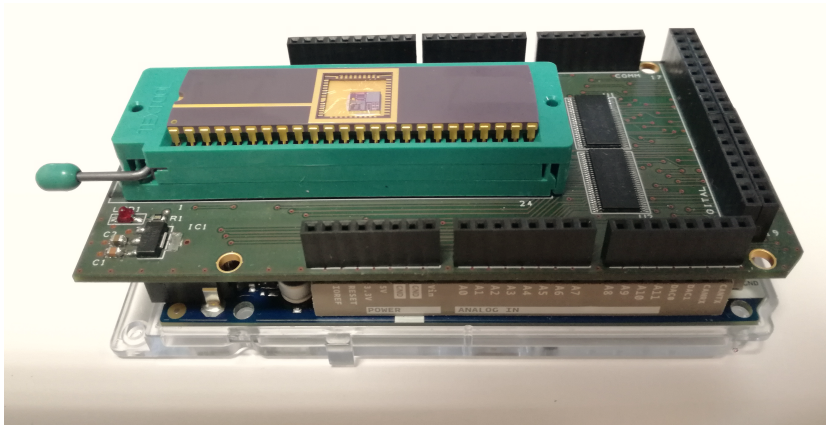


FIGURE 4.30 – Photographie de la plateforme réalisée et montée.

5 Conclusion et perspectives

Dans un premier temps, ce chapitre a mis en avant une architecture hybride du chiffrement PRESENT en utilisant les technologies CMOS et STT-MRAM. Cette hybridation a démontré un avantage majeur dans la consommation du chiffrement PRESENT pour des nœuds technologiques avancés avec une faible contrepartie en terme de surface. L'utilisation de technologies hybrides dans l'Internet des Objets pour des applications "*Normally-off / Instant-On*" peut devenir essentielle.

Dans un second temps, la proposition d'une solution de détection d'attaques a été proposée. Cette solution est intitulée Double Redondance Technologique ou – *Dual Technology Redundancy* – (DTR). Cette solution propose d'instancier deux chemins de données dans les algorithmes de cryptographie par exemple, un chemin pur CMOS et un chemin hybride basé sur la technologie mémoire STT-MRAM. Une comparaison entre ces deux chemins est réalisée lors de l'exécution de la donnée afin de détecter toute attaque qui aurait pu viser le CMOS ou la technologie mémoire non-volatile STT-MRAM. Pour cela, des bascules non-volatiles NVFFs et des bascules multi-contextes MC-NVFFs sont introduites dans le fonctionnement du chiffrement.

Toutefois, cette dernière architecture du chiffrement reste à être fabriquée sur silicium afin d'en déterminer ses performances en terme de consommation et de surface. De plus, des attaques de différentes natures doivent être réalisées afin d'en préciser les limites sécuritaires. Enfin, les puces fabriquées n'ont pas pu être testées dans le temps imparti. La carte de test a été validée et est opérationnelle ce qui devrait permettre la validation expérimentale dans un futur proche.

Conclusions et perspectives

Conclusion générale

Lors de cette thèse, nous nous sommes intéressés à différents aspects de l'hybridation de la logique CMOS et de la technologie mémoire STT-MRAM qui est pressentie pour remplacer la mémoire Flash embarquée. Ainsi, l'un des enjeux majeurs du développement de ces architectures hybrides est leur niveau de sécurité face aux attaques physiques, dites de perturbation essentiellement.

C'est pourquoi dans un premier temps l'étude de l'intégrité des mémoires STT-MRAMs a été réalisée. Comme il a pu être démontré dans le chapitre 2, toutes les jonctions magnétiques qui ont initialement une aimantation magnétique antiparallèle basculent vers l'état parallèle lors d'une illumination sous faisceau laser suffisamment énergétique. Aucun effet cumulatif de l'énergie laser délivrée n'est observé pour induire cette inversion du bit stocké dans la mémoire de l'état logique '1' vers l'état logique '0'. Ainsi le modèle de fautes induit sur ces jonctions mémoires est un modèle de type *bit-reset*. L'état final des jonctions qui basculent est toujours l'état parallèle, ou de faible résistivité. Le seuil d'énergie minimal nécessaire à fournir par laser est d'environ 547 nJ mais dépendra fortement de l'anisotropie des échantillons. Cette commutation a de plus été illustrée par des simulations physiques en utilisant l'outil COMSOL. Celles-ci ont démontré que ce *bit-flip* est dû à une augmentation de température dans la JTM. Ainsi, la solution technologique pour réduire la sensibilité de ces structures face aux attaques laser en face avant serait de rapprocher cette structure du FEoL.

Suite à ces observations, le Chapitre 3 présente un détecteur susceptible de tirer avantage de la sensibilité des JTMs aux attaques thermiques pour en faire un atout en tant que capteur d'attaques. Ce capteur est conçu pour détecter à la fois des attaques photoélectriques qui sont induites dans le substrat, que des attaques thermiques qui peuvent viser les niveaux d'interconnexion. Cette double capacité de détection a été nommée « *Dual Detection of Heating and Photocurrent attacks - (DDHP)* ». Des simulations électriques ont démontré l'efficacité de ce capteur.

Outre l'utilisation de ces jonctions pour la détection d'attaques, nous nous sommes également intéressés à leur intégration dans des algorithmes de cryptographie légère de type "PRESENT" dans le dernier chapitre de ce manuscrit (Chapitre 4). Le développement du chiffrement hybride CMOS/STT-MRAM de cet algorithme de cryptographie "PRESENT" est réalisé grâce à la modification de certains composants stratégiques, passant de volatiles à non-volatiles, permettant ainsi le stockage des données dans les jonctions avant la mise hors tension du circuit et leur restauration lors de sa réactivation. Cette méthode permet de réduire à quasi nulle l'énergie statique qui est usuellement consommée lorsque le chiffrement n'est pas utilisé et le bloc en veille. Ainsi, dans un premier temps nous avons pu démontrer l'efficacité énergétique de l'architecture hybride de l'algorithme PRESENT par rapport à une structure CMOS pure pour des nœuds technologiques avancés. En effet, pour une technologie FD-SOI 28 nm et un diamètre de jonction de 40 nm, la solution devient énergétiquement efficace après seulement 185 µs de veille. En contrepartie, une surface silicium supplémentaire de 23 % est nécessaire pour implémenter ce chiffrement comparé à une architecture CMOS pure. Cette solution cryptographique hybride, énergétiquement inté-

ressante, peut être durcie en intégrant le capteur DDHP. Pour durcir davantage ce chiffrement, une contre-mesure intégrée à son fonctionnement interne est proposée. Celle-ci est désignée par Double Redondance Technologique ou – *Dual Technology Redundancy* – (DTR). Cette technologie est basée sur la comparaison de deux chemins de données, un chemin CMOS et un chemin hybride CMOS/STT-MRAM, en partant du principe que les deux chemins ne présentent pas la même sensibilité aux mêmes attaques.

Perspectives

Les résultats mis en avant par ces travaux de thèse ouvrent la voie à de nombreuses perspectives à court, moyen et long terme.

Dans un premier temps, il est nécessaire de réaliser le test physique, électrique et sécuritaire, des primitives de sécurité qui ont été conçues et fabriquées, mais qui n'ont malheureusement pu être testées avant la fin de cette thèse. La perturbation de ces dispositifs face à différentes catégories d'attaques, telles que les injections de fautes de type LASER (photoélectrique et thermique) ou électromagnétique ou encore des attaques par observation doivent être entreprises. Cette caractérisation sécuritaire complète permettra, à moyen terme, de compléter les résultats encourageants obtenus en simulation.

Selon les résultats des analyses qui pourront être réalisées sur ces circuits, il sera possible de proposer de nouvelles solutions de conception visant la réduction de leur sensibilité aux attaques grâce à l'apprentissage préalablement réalisé sur ces structures, principalement sur le capteur DDHP. Suite à cette étude, le capteur pourra alors être amélioré en vue d'une intégration dans les systèmes visant différentes applications dans les domaines de l'Internet des Objets ou encore de l'automobile. Pour cela, il serait nécessaire d'envisager l'utilisation de matrices de jonctions qui pourront protéger toute la surface du circuit intégré, pour de faibles surfaces de circuits, ou de multiplier le nombre de détecteurs dans un système. Il est donc primordial de déterminer les capacités de détection de ce capteur en termes de surface "*protégée*".

Une autre perspective qui peut être envisagée pour compléter ces travaux de thèse, serait l'évaluation de l'architecture hybride de l'algorithme de cryptographie PRESENT intégrant la proposition de redondance DTR, pour différencier deux chemins d'attaques sensibles à deux modèles de fautes distincts. Il sera tout d'abord nécessaire de réaliser les simulations électriques de l'architecture complète afin d'en évaluer les performances, principalement en termes de consommation. Cette architecture pourrait alors être fabriquée sur silicium afin d'en établir les atouts et les inconvénients comparativement à la structure CMOS pure et la structure hybride CMOS/STT-MRAM, en termes de performances et de sécurité.

Publications

— **Kharbouche-Harrari M.**, Postel-Pellerin J., Di Pendina G., Wacquez R., Aboulkassimi D., Bocquet M., Sousa R., Delattre R. and Portal J.-M. : "Impact of a Laser Pulse on a STT-MRAM Bitcell : Security and Reliability Issues", In : *2018 IEEE 24th International Symposium on On-Line Testing And Robust System Design (IOLTS)*, pp. 243-4, 2018.

— **Kharbouche-Harrari M.**, Alhalabi R., Postel-Pellerin J., Wacquez R., Aboulkassimi D, Nowak E, Prejbeanu I.-L., Prenat G. and Di Pendina G. : "MRAM : from STT to SOT, for security and memory", In : *2018 Conference on Design of Circuits and Integrated Systems (DCIS)*, pp. 1-6, 2018.

— **Kharbouche-Harrari M.**, Di Pendina G., Wacquez R., Dieny B., Aboulkassimi D., Postel-Pellerin J. and Portal J.-M. : "Light-Weight Cipher Based on Hybrid CMOS/STT-MRAM : Power/Area Analysis", In : *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-5, 2019.

— **Kharbouche-Harrari M.**, Wacquez R., Di Pendina G., Dutertre J.–M., Postel-Pellerin, J. Aboulkassimi D. and Portal J.-M. : "Dual Detection of Heating and Photocurrent attacks (DDHP) Sensor using Hybrid CMOS/STT-MRAM", In : *2019 IEEE 25th International Symposium on On-Line Testing And Robust System Design (IOLTS)*, pp. 322-327, 2019.

Annexe A

De la physique au magnétisme

Pour la compréhension du fonctionnement de ces mémoires magnétiques, il est nécessaire d'introduire le domaine de la spintronique. Pour cela, ce paragraphe débute par quelques rappels de physique et de magnétisme qui permettent d'expliquer le fonctionnement de cette famille de mémoires et de mettre en avant leurs avantages et inconvénients.

Soit, chaque électron de l'atome considéré par ses quatre propriétés magnétiques définies par :

— Le nombre quantique principal n correspondant à la dimension de l'orbitale, c'est-à-dire à la couche sur laquelle l'électron se trouve, tel que $n \in \mathbb{N}^*$. Ce nombre n est déterminé par l'énergie de l'électron.

— Le nombre quantique orbital l, également appelé nombre quantique secondaire ou azimutal définit la forme de l'orbitale $0 < l < n-1$. Il existe différentes formes d'orbitales : s (sphère), p (2 lobes), d (4 lobes) ou encore f (8 lobes) [184], correspondant respectivement à $l = 0, 1, 2$ et 3.

— Le nombre quantique magnétique m_l fixe l'orientation spatiale de l'orbitale et est lié au nombre quantique orbital par la relation : $-l \leqslant m_l \leqslant +l$. Il fixe la direction du vecteur de moment angulaire de l'électron.

— Le nombre quantique de spin m_s précise le sens de rotation de l'électron sur lui-même, horaire (*spin-up* correspondant à $m_s = +\frac{1}{2}$) ou anti-horaire (*spin-down* correspondant à $m_s = -\frac{1}{2}$). Cette propriété magnétique définit le moment cinétique angulaire de l'électron. La spintronique se base sur cette propriété et non pas uniquement sur la charge de l'électron, comme pour l'électronique conventionnelle.

La détermination des quatre nombres quantiques caractérisant les propriétés de chaque électron de l'atome sont définies par différentes règles : le principe d'exclusion de Pauli [185], la règle de Hund [15] et de remplissage de Klechkowski [186].

En effet, le principe d'exclusion de Pauli tel qu'énoncé en 1925 indique que deux électrons ne peuvent partager les 4 mêmes nombres quantiques dans un même système (n, l, m_l, m_s). Au moins l'un de ces nombres doit être distinct, comme représenté sur la Figure A.1.

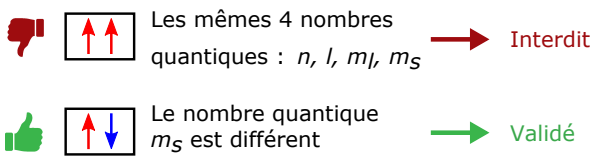


FIGURE A.1 – Représentation du principe d'exclusion de Pauli.

La règle de F. Hund instaurée vers 1927 indique qu'il est nécessaire de remplir deux électrons de spins opposés par direction d'orbitale, uniquement après le remplissage par au moins un électron des autres orbitales, comme illustré Figure A.2.

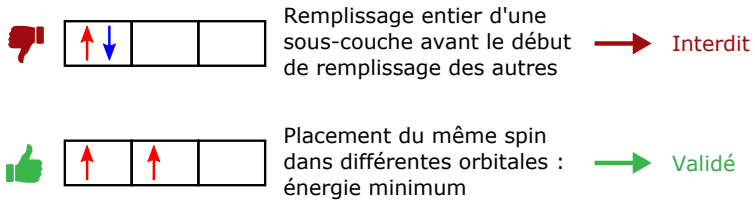
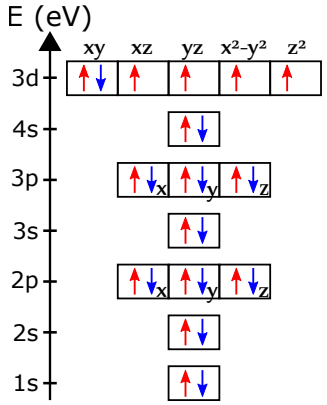


FIGURE A.2 – Représentation de la règle d'Hund.

La dernière règle que doivent suivre les électrons est la règle de Klechkowski. Les couches sont remplies de sorte que "$n+l$" soit toujours croissant. En effet, il est nécessaire de débuter le remplissage des orbitales par $1s$ correspondant à $n+l=1$, avec $n=1$ et $l=0$. Les orbitales suivantes sont $2s$ (correspondant à $n+l=2$) puis $2p$ (correspondant à $n+l=3$), $3s$, $3p$, $4s$ etc, tel que décrit sur la Figure A.3.a. L'énergie nécessaire à l'électron pour se fixer sur une orbitale augmente selon l'expression $n+l$, comme illustré Figure A.3.b.

n \ l	0	1	2	3
1	1s			
2	2s	2p		
3	3s	3p	3d	
4	4s	4p	4d	4f
5	5s	5p	5d	5f
6	6s	6p	6d	
7	7s	7p		

(a) Règle de remplissage des orbitales selon Klechkowski.



(b) Représentation des orbitales en fonction de leur niveau énergétique.

FIGURE A.3 – Règle de remplissage des orbitales selon Klechkowski.

Ces caractéristiques magnétiques des électrons et des atomes entrent en jeu dans la description des matériaux magnétiques utilisés pour le développement de technologies mémoires magnétiques MRAMs. La description de ces matériaux et de leurs propriétés intrinsèques seront développées dans la partie suivante.

Liste des acronymes

AES	chiffrement standard avancé ou – *Advanced Encryption Standard – p. 23*
ASIC	*Application-Specific Integrated Circuit – p. 76*
BBICS	détecteur de courants de substrat ou – *Bulk Built-In Current Sensor – p. 61*
BEoL	*Back-End of Line – p. 8*
CAO	Conception Assistée par Ordinateur – *p. 78*
CBRAM	*Conductive Bridging Random Access Memory – p. 8*
CEM	Compatibilité ÉlectroMagnétique – *p. 99*
CHES	*Cryptographic Hardware and Embedded Systems – p. 72*
CPA	Analyse de Corrélation de Puissance ou – *Correlation Power Analysis – p. 25*
DDHP	double détection d'attaques induites par chauffage et effet photoélectrique ou – *Dual Detection of Heating and Photocurrent attacks – p. 55*
DES	*Data Encryption Standard – p. 23*
DFM	*Design For Manufacturing – p. 79*
DIL	*Dual In-Line – p. 97*
DMR	redondance double ou – *Double Modular Redundancy – p. 58*
DPA	Analyse Différentielle de Puissance ou – *Differential Power Analysis – p. 25*
DRAMs	mémoires dynamiques ou – *Dynamic Random Access Memories – p. 5*
DRC	*Design Rule Check – p. 79*
DTR	Double Redondance Technologique ou – *Dual Technology Redundancy – p. 71*
FD-SOI	*Fully Depleted Silicon On Insulator – p. 57*
FEoL	*Front-End of Line – p. 52*
FF	bascule synchrone ou – *Flip-Flop – p. 30*
FIB	faisceau d'ions focalisés ou – *Focused Ion Beam – p. 26*
FIMS-Toggle	*Field Induced Magnetic Switching-Toggle – p. 13*
GE	portes équivalentes ou – *Gate Equivalents – p. 69*
GMR	Magnéto-Résistance Géante ou – *Giant MagnetoResistance – p. 11*
ITRS	*International Technology Roadmap for Semiconductors – p. 8*
JTM	Jonction Tunnel Magnétique – ou *Magnetic Tunnel Junction – p. 12*
LASER	Amplification de la Lumière par Émission Stimulée de Radiation ou – *Light Amplification by Stimulated Emission of Radiation p. 26*
LVS	*Layout Versus Schematic – p. 79*
LWC	cryptographie légère ou – *Light Weight Cryptography – p. 23*

MC-NVFF	Bascule Synchrone Non-Volatile Multi-Contexte ou – *Multi-Context Non-Volatile Flip-Flop – p. 30*
MRAM	mémoire magnétique – ou *Magnetic Random Access Memory – p. 9*
NBS	*National Bureau of Standards – p. 19*
Nd	Néodyme ou – *Neodymium* Nd^{3+} *– p. 44*
Nd-YAG	Grenat d'Yttrium-Aluminium dopé au Néodyme ou – *Neodymium doped Yttrium-Aluminum-Garnet – p. 40*
NIST	*National Institute of Standards and Technology – p. 19*
NVFF	bascule synchrone non-volatile ou – *Non-Volatile Flip-Flop – p. 30*
NV-LUTs	tables de correspondance non-volatiles ou – *Non-Volatile Look-Up Tables – p. 92*
OxRAM	*Oxide Random Access Memory – p. 8*
PDK	*Process Design Kit – p. 68*
PKC	Cryptographie à Clé Publique ou – *Public Key Cryptography – p. 19*
PUF	Fonction Physique Non-clonable ou – *Physically Unclonable Function – p. 29*
ReRAMs ou RRAMs	*Resistive Random Access Memories – p. 7*
RSA	*Rivest Shamir Adleman – p. 21*
RTL	*Register Transfer Level – p. 77*
SEE	*Single Event Effect – p. 56*
SEL	*Single Event Latch-Up – p. 56*
SOT-MRAM	*Spin-Orbit Torque MRAM – p. 13*
SPA	Analyse Simple de Puissance ou – *Simple Power Analysis – p. 24*
SPN	Réseau de Substitution et de Permutation ou – *Substitution Permutation Network – p. 23*
SRAM	mémoire statique ou – *Static Random Access Memory – p. 5*
SRAMs	mémoires statiques ou – *Static Random Access Memories – p. 5*
STT-MRAM	*Spin-Transfer Torque MRAM – p. 5*
TAS-MRAM	*Thermally Assisted Switching MRAM – p. 13*
TMR	Magnéto-Résistance Tunnel ou – *Tunnel MagnetoResistance – p. 11, 59*
TRNG	Générateur de Nombres Aléatoires ou – *True Random Number Generator – p. 29*
YAG	Grenat d'Yttrium-Aluminium ou – *Yttrium Aluminum Garnet* $Y_3Al_5O_{12}$ *– p. 44*
ZIF	*Zero Insertion Force – p. 99*

Bibliographie

[1] G. Siméon, "Données le vertige," Dec. 2012. [Online]. Available : https://www.liberation.fr/futurs/2012/12/03/donnees-le-vertige_864585

[2] "Planetoscope - Statistiques : Informations publiées dans le monde sur le net (en Gigaoctets)." [Online]. Available : https://www.planetoscope.com/Internet-/1523-informations-publiees-dans-le-monde-sur-le-net-en-gigaoctets-.html

[3] "Insecam - World biggest online cameras directory." [Online]. Available : http://www.insecam.org/en/

[4] "State of the IoT 2018 : Number of IoT devices now at 7b – Market accelerating." [Online]. Available : https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/

[5] "IoT : number of connected devices worldwide 2012-2025." [Online]. Available : https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

[6] B. Lac, "Cryptographie légère intrinsèquement résistante aux attaques physiques pour l'Internet des Objets," Manuscrit de thèse, Université de Lyon opérée au sein de l'Ecole des Mines de Saint-Etienne, Oct. 2018.

[7] D. Barton, "Unsecured IoT - A Dangerous Gambit!" 2017. [Online]. Available : https://www.a2n.net/2017/03/20/unsecured-iot/

[8] F. Masuoka, M. Asano, H. Iwahashi, T. Komuro, and S. Tanaka, "A new flash E2prom cell using triple polysilicon technology," in *1984 International Electron Devices Meeting*, Dec. 1984, pp. 464–467.

[9] "Emerging Non-Volatile Memory," Yole Development, Tech. Rep., Dec. 2018. [Online]. Available : http://www.yole.fr/Emerging_NVM_Memory_Activities_Webcasts.aspx#.XMbGzzAzapo

[10] "2013 International Technology Roadmap for Semiconductors (ITRS)." [Online]. Available : https://www.semiconductors.org/resources/2013-international-technology-roadmap-for-semiconductors-itrs/

[11] S. Hong, "Memory technology trend and future challenges," in *2010 International Electron Devices Meeting*, Dec. 2010, pp. 12.4.1–12.4.4.

[12] M. Bocquet, "Caractérisation et modélisation compacte de mémoires émergentes," Habilitation à diriger des recherches, Aix Marseille Université, Jun. 2017. [Online]. Available : https://hal.archives-ouvertes.fr/tel-01737675

[13] R. Waser, R. Dittmann, G. Staikov, and K. Szot, "Redox-Based Resistive Switching Memories – Nanoionic Mechanisms, Prospects, and Challenges," *Advanced Materials*, vol. 21, no. 25-26, pp. 2632–2663, 2009. [Online]. Available : https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.200900375

[14] "Periodic Table of Elements." [Online]. Available : https://iupac.org/what-we-do/periodic-table-of-elements/

[15] S. Blundell, *Magnetism in condensed matter*. Oxford master series in condensed matter physics, 2001.

[16] M. Julliere, "Tunneling between ferromagnetic films," *Physics Letters A*, vol. 54, no. 3, pp. 225–226, Sep. 1975. [Online]. Available : http://www.sciencedirect.com/science/article/pii/0375960175901747

[17] M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices," *Phys. Rev. Lett.*, vol. 61, no. 21, pp. 2472–2475, Nov. 1988. [Online]. Available : https://link.aps.org/doi/10.1103/PhysRevLett.61.2472

[18] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, "Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange," *Phys. Rev. B*, vol. 39, no. 7, pp. 4828–4830, Mar. 1989. [Online]. Available : https://link.aps.org/doi/10.1103/PhysRevB.39.4828

[19] E. Kültürsay, M. Kandemir, A. Sivasubramaniam, and O. Mutlu, "Evaluating STT-RAM as an energy-efficient main memory alternative," in *2013 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, Apr. 2013, pp. 256–267.

[20] C. Park, J. J. Kan, C. Ching, J. Ahn, L. Xue, R. Wang, A. Kontos, S. Liang, M. Bangar, H. Chen, S. Hassan, S. Kim, M. Pakala, and S. H. Kang, "Temperature Dependence of Critical Device Parameters in 1 Gb Perpendicular Magnetic Tunnel Junction Arrays for STT-MRAM," *IEEE Transactions on Magnetics*, vol. 53, no. 2, pp. 1–4, Feb. 2017.

[21] T. Miyazaki and N. Tezuka, "Giant magnetic tunneling effect in Fe/Al2o3/Fe junction," *Journal of Magnetism and Magnetic Materials*, vol. 139, no. 3, pp. L231–L234, Jan. 1995. [Online]. Available : http://www.sciencedirect.com/science/article/pii/0304885395900012

[22] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, "Large Magnetoresistance at Room Temperature in Ferromagnetic Thin Film Tunnel Junctions," *Phys. Rev. Lett.*, vol. 74, no. 16, pp. 3273–3276, Apr. 1995. [Online]. Available : https://link.aps.org/doi/10.1103/PhysRevLett.74.3273

[23] D. Apalkov, A. Khvalkovskiy, S. Watts, V. Nikitin, X. Tang, D. Lottis, K. Moon, X. Luo, E. Chen, A. Ong, A. Driskill-Smith, and M. Krounbi, "Spin-transfer Torque Magnetic Random Access Memory (STT-MRAM)," *J. Emerg. Technol. Comput. Syst.*, vol. 9, no. 2, pp. 13 :1–13 :35, May 2013. [Online]. Available : http://doi.acm.org/10.1145/2463585.2463589

[24] S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S.-H. Yang, "Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers," *Nature Materials*, vol. 3, no. 12, p. 862, Dec. 2004. [Online]. Available : https://www.nature.com/articles/nmat1256

[25] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, "Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions," *Nature Materials*, vol. 3, no. 12, p. 868, Dec. 2004. [Online]. Available : https://www.nature.com/articles/nmat1257

[26] Y. M. Lee, J. Hayakawa, S. Ikeda, F. Matsukura, and H. Ohno, "Effect of electrode composition on the tunnel magnetoresistance of pseudo-spin-valve magnetic tunnel junction with a MgO tunnel barrier," *Appl. Phys. Lett.*, vol. 90, no. 21, p. 212507, May 2007. [Online]. Available : https://aip.scitation.org/doi/full/10.1063/1.2742576

[27] L. Savtchenko, B. N. Engel, N. D. Rizzo, M. F. Deherrera, and J. A. Janesky, "Method of writing to scalable magnetoresistance random access memory element," US Patent US6 545 906B1, Apr., 2003. [Online]. Available : https://patents.google.com/patent/US6545906/en

[28] G. Di Pendina, "Conception innovante et développement d'outils de conception d'ASIC pour Technologie Hybride CMOS/Magnétique," Manuscrit de thèse, Université Grenoble Alpes, Oct. 2012. [Online]. Available : https://hal-cea.archives-ouvertes.fr/tel-00750121v2

[29] J. C. Slonczewski, “Current-driven excitation of magnetic multilayers,” *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 1, pp. L1–L7, Jun. 1996. [Online]. Available : http://www.sciencedirect.com/science/article/pii/0304885396000625

[30] L. Berger, “Emission of spin waves by a magnetic multilayer traversed by a current,” *Phys. Rev. B*, vol. 54, no. 13, pp. 9353–9358, Oct. 1996. [Online]. Available : https://link.aps.org/doi/10.1103/PhysRevB.54.9353

[31] N. Perrissin, G. Gregoire, S. Lequeux, L. Tillie, N. Strelkov, S. Auffret, L. D. Buda-Prejbeanu, R. C. Sousa, L. Vila, B. Dieny, and I. L. Prejbeanu, “Perpendicular shape anisotropy spin transfer torque magnetic random-access memory : towards sub-10\hspace0.167emnm devices,” *J. Phys. D : Appl. Phys.*, vol. 52, no. 23, p. 234001, Mar. 2019. [Online]. Available : https://doi.org/10.1088%2F1361-6463%2Fab0de4

[32] “ITRS 2.0 2015 Edition - IEEE Electronics Packaging Society.” [Online]. Available : https://eps.ieee.org/itrs-2-0-2015-edition.html

[33] K. Cao, H. Zhao, M. Wang, and W. Zhao, “Spin orbit torques for ultra-low power computing,” in *2015 IEEE 11th International Conference on ASIC (ASICON)*, Nov. 2015, pp. 1–4.

[34] G. Gaudin, I. M. Miron, P. Gambardella, and A. Schuhl, “Writable Magnetic Memory Element,” US Patent US20 120 020 152A1, Jan., 2012. [Online]. Available : https://patents.google.com/patent/US20120020152A1/en

[35] G. Gaudin, I. M. Miron, P. Gambardella, and A. Schuhl, “Writable Magnetic Element,” US Patent US20 120 098 077A1, Apr., 2012. [Online]. Available : https://patents.google.com/patent/US20120098077A1/en

[36] G. Gaudin, I. M. Miron, P. Gambardella, and A. Schuhl, “Writable Magnetic Element,” US Patent US20 120 018 822A1, Jan., 2012. [Online]. Available : https://patents.google.com/patent/US20120018822/un

[37] M. Cubukcu, O. Boulle, N. Mikuszeit, C. Hamelin, T. Brächer, N. Lamard, M.-C. Cyrille, L. Buda-Prejbeanu, K. Garello, I. M. Miron, O. Klein, G. de Loubens, V. V. Naletov, J. Langer, B. Ocker, P. Gambardella, and G. Gaudin, “Ultra-Fast Perpendicular Spin–Orbit Torque MRAM,” *IEEE Transactions on Magnetics*, vol. 54, no. 4, pp. 1–4, Apr. 2018.

[38] K. Garello, C. O. Avci, I. M. Miron, M. Baumgartner, A. Ghosh, S. Auffret, O. Boulle, G. Gaudin, and P. Gambardella, “Ultrafast magnetization switching by spin-orbit torques,” *Appl. Phys. Lett.*, vol. 105, no. 21, p. 212402, Nov. 2014. [Online]. Available : https://aip.scitation.org/doi/full/10.1063/1.4902443

[39] I. M. Miron, G. Gaudin, S. Auffret, B. Rodmacq, A. Schuhl, S. Pizzini, J. Vogel, and P. Gambardella, “Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer,” *Nat Mater*, vol. 9, no. 3, pp. 230–234, Mar. 2010. [Online]. Available : http://www.nature.com/nmat/journal/v9/n3/full/nmat2613.html

[40] I. M. Miron, K. Garello, G. Gaudin, P.-J. Zermatten, M. V. Costache, S. Auffret, S. Bandiera, B. Rodmacq, A. Schuhl, and P. Gambardella, “Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection,” *Nature*, vol. 476, no. 7359, pp. 189–193, Aug. 2011. [Online]. Available : http://www.nature.com/nature/journal/v476/n7359/full/nature10309.html

[41] L. Liu, O. J. Lee, T. J. Gudmundsen, D. C. Ralph, and R. A. Buhrman, “Current-Induced Switching of Perpendicularly Magnetized Magnetic Layers Using Spin Torque from the Spin Hall Effect,” *Phys. Rev. Lett.*, vol. 109, no. 9, p. 096602, Aug. 2012. [Online]. Available : http://link.aps.org/doi/10.1103/PhysRevLett.109.096602

[42] A. Hoffmann, “Spin Hall Effects in Metals,” *IEEE Transactions on Magnetics*, vol. 49, no. 10, pp. 5172–5193, Oct. 2013.

[43] A.-P. Mirbaha, “Etude de la vulnérabilité des circuits cryptographiques l’injection de fautes par laser.” Manuscrit de thèse, École Nationale Supérieure des Mines de Saint-Étienne, Dec. 2011. [Online]. Available : https://tel.archives-ouvertes.fr/tel-00844751

[44] “Cryptologie : art ou science du secret?” [Online]. Available : https://www.ssi.gouv.fr/particulier/bonnes-pratiques/crypto-le-webdoc/cryptologie-art-ou-science-du-secret/

[45] Rama, “Enigma-K machine of the Swiss Army. Cryptography collection of the Swiss Army headquarters.” [Online]. Available : https://commons.wikimedia.org/wiki/File:Enigma-IMG_0486-black.jpg

[46] D. Jauvart, “Sécurisation des algorithmes de couplages contre les attaques physiques,” Manuscrit de thèse, Université Paris-Saclay préparée à l’Université de Versailles Saint-Quentin, Sep. 2017.

[47] A. Kerckhoffs, *La cryptographie militaire*. Journal des sciences militaires, 1883. [Online]. Available : http://archive.org/details/117Kerckhoffs

[48] W. Diffie and M. E. Hellman, “New directions in cryptography,” *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, Nov. 1976.

[49] R. L. Rivest, A. Shamir, and L. Adleman, “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems,” *Communications of the ACM*, vol. 21, pp. 120–126, 1978.

[50] I. Bernard, “Cryptanalyse de chiffrement par flot,” 2015, supinfo, École Supérieure d’Informatique. [Online]. Available : https://www.supinfo.com/articles/single/1290-cryptanalyse-chiffrement-flot

[51] D. J. Bernstein, “ChaCha, a variant of Salsa20,” 2008. [Online]. Available : https://cr.yp.to/chacha.html

[52] S. Babbage and M. Dodd, “The MICKEY Stream Ciphers,” in *New Stream Cipher Designs : The eSTREAM Finalists*, ser. Lecture Notes in Computer Science, M. Robshaw and O. Billet, Eds. Berlin, Heidelberg : Springer Berlin Heidelberg, 2008, pp. 191–209. [Online]. Available : https://doi.org/10.1007/978-3-540-68351-3_15

[53] C. De Cannière and B. Preneel, “Trivium,” in *New Stream Cipher Designs : The eSTREAM Finalists*, ser. Lecture Notes in Computer Science, M. Robshaw and O. Billet, Eds. Berlin, Heidelberg : Springer Berlin Heidelberg, 2008, pp. 244–266. [Online]. Available : https://doi.org/10.1007/978-3-540-68351-3_18

[54] “Lightweight Stream Ciphers,” université du Luxembourg. [Online]. Available : https://www.cryptolux.org/index.php/Lightweight_Stream_Ciphers#cite_note-BD08-13

[55] H. Feistel, “Block Cipher Cryptographic System,” Patent US3 798 359 (A), Mar., 1974, cIB : H04L9/06; (IPC1-7) : H04L9/00. [Online]. Available : https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=19740319&DB=&locale=fr_EP&CC=US&NR=3798359A&KC=A&ND=1

[56] NIST, “Data Encryption Standard (DES),” U.S. Department of Commerce, Tech. Rep. Federal Information Processing Standard (FIPS) 46-3 (Withdrawn), Oct. 1999. [Online]. Available : https://csrc.nist.gov/publications/detail/fips/46/3/archive/1999-10-25

[57] E. Barker and N. Mouha, “Recommendation for the Triple Data Encryption Algorithm (TDEA) Block Cipher,” National Institute of Standards and Technology, Tech. Rep. NIST Special Publication (SP) 800-67 Rev. 2, Nov. 2017. [Online]. Available : https://csrc.nist.gov/publications/detail/sp/800-67/rev-2/final

[58] B. Schneier, “Description of a new variable-length key, 64-bit block cipher (Blowfish),” in *Fast Software Encryption*, ser. Lecture Notes in Computer Science, R. Anderson, Ed. Springer Berlin Heidelberg, 1994, pp. 191–204.

[59] R. Beaulieu, S. Treatman-Clark, D. Shors, B. Weeks, J. Smith, and L. Wingers, “The SIMON and SPECK lightweight block ciphers,” Tech. Rep., Jun. 2015.

[60] J. Daemen and V. Rijmen, "The Block Cipher Rijndael," in *Lecture Notes in Computer Science - LNCS*, vol. 1820, Jan. 1998, pp. 277–284.

[61] NIST, "AES Development - Cryptographic Standards and Guidelines | CSRC." [Online]. Available : https://csrc.nist.gov/projects/cryptographic-standards-and-guidelines/archived-crypto-projects/aes-development

[62] K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, and T. Tokita, "Camellia : A 128-Bit Block Cipher Suitable for Multiple Platforms — Design andAnalysis," in *Selected Areas in Cryptography*, ser. Lecture Notes in Computer Science, D. R. Stinson and S. Tavares, Eds. Springer Berlin Heidelberg, 2001, pp. 39–56.

[63] K. Shibutani, T. Isobe, H. Hiwatari, A. Mitsuda, T. Akishita, and T. Shirai, "Piccolo : An Ultra-Lightweight Blockcipher," in *Cryptographic Hardware and Embedded Systems – CHES 2011*, ser. Lecture Notes in Computer Science, B. Preneel and T. Takagi, Eds. Springer Berlin Heidelberg, 2011, pp. 342–357.

[64] F.-X. Standaert, G. Piret, N. Gershenfeld, and J.-J. Quisquater, "SEA : A Scalable Encryption Algorithm for Small Embedded Applications," in *Smart Card Research and Advanced Applications*, ser. Lecture Notes in Computer Science, J. Domingo-Ferrer, J. Posegga, and D. Schreckling, Eds. Springer Berlin Heidelberg, 2006, pp. 222–236.

[65] Z. Gong, S. Nikova, and Y. W. Law, "KLEIN : A New Family of Lightweight Block Ciphers," in *RFID. Security and Privacy*, ser. Lecture Notes in Computer Science, A. Juels and C. Paar, Eds. Springer Berlin Heidelberg, 2012, pp. 1–18.

[66] J. Borghoff, A. Canteaut, T. Güneysu, E. B. Kavun, M. Knezevic, L. R. Knudsen, G. Leander, V. Nikov, C. Paar, C. Rechberger, P. Rombouts, S. S. Thomsen, and T. Yalçın, "PRINCE – A Low-Latency Block Cipher for Pervasive Computing Applications," in *Advances in Cryptology – ASIACRYPT 2012*, ser. Lecture Notes in Computer Science, X. Wang and K. Sako, Eds. Springer Berlin Heidelberg, 2012, pp. 208–225.

[67] M. R. Albrecht, B. Driessen, E. B. Kavun, G. Leander, C. Paar, and T. Yalçın, "Block Ciphers – Focus on the Linear Layer (feat. PRIDE)," in *Advances in Cryptology – CRYPTO 2014*, ser. Lecture Notes in Computer Science, J. A. Garay and R. Gennaro, Eds. Springer Berlin Heidelberg, 2014, pp. 57–76.

[68] A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw, Y. Seurin, and C. Vikkelsoe, "PRESENT : An Ultra-Lightweight Block Cipher," in *Cryptographic Hardware and Embedded Systems - CHES 2007*, ser. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2007, pp. 450–466. [Online]. Available : https://link.springer.com/chapter/10.1007/978-3-540-74735-2_31

[69] J.-J. Quisquater and D. Samyde, "ElectroMagnetic Analysis (EMA) : Measures and Countermeasures for Smart Cards," in *Smart Card Programming and Security*. Springer Berlin Heidelberg, 2001, pp. 200–210. [Online]. Available : http://link.springer.com/chapter/10.1007/3-540-45418-7_17

[70] P. Kocher, J. Jaffe, B. Jun, and P. Rohatgi, "Introduction to differential power analysis," *J Cryptogr Eng*, vol. 1, no. 1, pp. 5–27, Apr. 2011. [Online]. Available : https://doi.org/10.1007/s13389-011-0006-y

[71] P. C. Kocher, "Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems," in *Advances in Cryptology — CRYPTO '96*, ser. Lecture Notes in Computer Science, N. Koblitz, Ed. Springer Berlin Heidelberg, 1996, pp. 104–113.

[72] D. Genkin, A. Shamir, and E. Tromer, "RSA Key Extraction via Low-Bandwidth Acoustic Cryptanalysis," in *Advances in Cryptology – CRYPTO 2014*, ser. Lecture Notes in Computer Science, J. A. Garay and R. Gennaro, Eds. Springer Berlin Heidelberg, 2014, pp. 444–461.

[73] Y. Oren, "Information Security – Theory vs. Reality, Power Analysis," 2011. [Online]. Available : https://slideplayer.com/slide/3377741/

[74] S. Mangard, "A Simple Power-Analysis (SPA) Attack on Implementations of the AES Key Expansion," in *Information Security and Cryptology — ICISC 2002*, ser. Lecture Notes in Computer Science, P. J. Lee and C. H. Lim, Eds. Springer Berlin Heidelberg, 2003, pp. 343–358.

[75] D. J. Bernstein, "Cache-timing attacks on AES," in *preprint*, 2005. [Online]. Available : https://cr.yp.to/antiforgery/cachetiming-20050414.pdf

[76] D. A. Osvik, A. Shamir, and E. Tromer, "Cache Attacks and Countermeasures : The Case of AES," in *Topics in Cryptology – CT-RSA 2006*, ser. Lecture Notes in Computer Science, D. Pointcheval, Ed. Springer Berlin Heidelberg, 2006, pp. 1–20.

[77] P. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," in *Advances in Cryptology — CRYPTO' 99*, ser. Lecture Notes in Computer Science, M. Wiener, Ed. Springer Berlin Heidelberg, 1999, pp. 388–397.

[78] E. Brier, C. Clavier, and F. Olivier, "Correlation Power Analysis with a Leakage Model," in *Cryptographic Hardware and Embedded Systems - CHES 2004*, ser. Lecture Notes in Computer Science, M. Joye and J.-J. Quisquater, Eds. Springer Berlin Heidelberg, 2004, pp. 16–29.

[79] S. B. Ors, F. Gurkaynak, E. Oswald, and B. Preneel, "Power-analysis attack on an ASIC AES implementation," in *International Conference on Information Technology : Coding and Computing, 2004. Proceedings. ITCC 2004.*, vol. 2, Apr. 2004, pp. 546–552 Vol.2.

[80] D. Aboulkassimi, M. Agoyan, L. Freund, J. Fournier, B. Robisson, and A. Tria, "ElectroMagnetic analysis (EMA) of software AES on Java mobile phones," in *2011 IEEE International Workshop on Information Forensics and Security*, Nov. 2011, pp. 1–6.

[81] M. Lecomte, "Système embarqué de mesure de la tension pour la détection de contrefaçons et de chevaux de troie matériels," Manuscrit de thèse, Université de Lyon opérée au sein de l'Ecole des Mines de Saint-Etienne, 2016.

[82] S. P. Skorobogatov and R. J. Anderson, "Optical Fault Induction Attacks," in *Cryptographic Hardware and Embedded Systems - CHES 2002*, ser. Lecture Notes in Computer Science, B. S. Kaliski, e. K. Koç, and C. Paar, Eds. Springer Berlin Heidelberg, 2003, pp. 2–12.

[83] M. Kharbouche-Harrari, J. Postel-Pellerin, G. Di Pendina, R. Wacquez, D. Aboulkassimi, M. Bocquet, R. Sousa, R. Delattre, and J.-M. Portal, "Impact of a Laser Pulse on a STT-MRAM Bitcell : Security and Reliability Issues," in *2018 IEEE 24th International Symposium on On-Line Testing And Robust System Design (IOLTS)*, Jul. 2018, pp. 243–244.

[84] A. Krakovinsky, M. Bocquet, R. Wacquez, J. Coignus, D. Deleruyelle, C. Djaou, G. Reimbold, and J.-M. Portal, "Impact of a laser pulse on HfO2-based RRAM cells reliability and integrity," in *2016 International Conference on Microelectronic Test Structures (ICMTS)*, Mar. 2016, pp. 152–156.

[85] S. Skorobogatov, "Local heating attacks on Flash memory devices," in *2009 IEEE International Workshop on Hardware-Oriented Security and Trust*, Jul. 2009, pp. 1–6.

[86] C. H. Kim and J.-J. Quisquater, "Faults, Injection Methods, and Fault Attacks," *IEEE Design Test of Computers*, vol. 24, no. 6, pp. 544–545, Nov. 2007.

[87] N. Borrel, "Évaluation d'injection de fautes Laser et conception de contre-mesures sur une architecture à faible consommation," Manuscrit de thèse, Université d'Aix-Marseille, Marseille, 2015.

[88] D. El-Baze, "Conception et Évaluation d'un Détecteur d'Attaque par Injection de Fautes," Manuscrit de thèse, Université de Lyon opérée au sein de l'École des Mines de Saint-Étienne, 2017.

[89] A. Barenghi, G. M. Bertoni, L. Breveglieri, M. Pellicioli, and G. Pelosi, "Low voltage fault attacks to AES," in *2010 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*, Jun. 2010, pp. 7–12.

[90] C. Layer, K. Jabeur, L. Becker, B. Diény, S. Gros, V. Javerliac, P. Paoli, and F. Bernard-Granger, "Hybrid STT/CMOS Design of an Interrupt Based Instant On/Off Mechanism for Low-Power SoC," in *2015 IEEE Computer Society Annual Symposium on VLSI*, Jul. 2015, pp. 315–320.

[91] D. Chabi, W. Zhao, E. Deng, Y. Zhang, N. B. Romdhane, J.-O. Klein, and C. Chappert, "Ultra Low Power Magnetic Flip-Flop Based on Checkpointing/Power Gating and Self-Enable Mechanisms," *IEEE Transactions on Circuits and Systems I : Regular Papers*, vol. 61, no. 6, pp. 1755–1765, Jun. 2014.

[92] J.-M. Portal, M. Bocquet, M. Moreau, H. Aziza, D. Deleruyelle, Y. Zhang, W. Kang, J.-O. Klein, Y.-G. Zhang, C. Chappert, and W.-S. Zhao, "An Overview of Non-Volatile Flip-Flops Based on Emerging Memory Technologies," *Journal of Electronics Science and Technology*, vol. 12, pp. 173–181, Jun. 2014.

[93] M. Nataraj, A. Levisse, B. Giraud, J.-P. Noel, P. Meinerzhagen, J.-M. Portal, and P.-E. Gaillardon, "Design methodology for area and energy efficient OxRAM-based non-volatile flip-flop," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2017, pp. 1–4.

[94] G. Di Pendina, K. Torki, G. Prenat, Y. Guillemenet, and L. Torres, "Ultra Compact Non-volatile Flip-Flop for Low Power Digital Circuits Based on Hybrid CMOS/Magnetic Technology," in *Integrated Circuit and System Design. Power and Timing Modeling, Optimization, and Simulation*, ser. Lecture Notes in Computer Science, J. L. Ayala, B. García-Cámara, M. Prieto, M. Ruggiero, and G. Sicard, Eds. Springer Berlin Heidelberg, 2011, pp. 83–91.

[95] K. Ryu, J. Kim, J. Jung, J. P. Kim, S. H. Kang, and S.-O. Jung, "A Magnetic Tunnel Junction Based Zero Standby Leakage Current Retention Flip-Flop," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 11, pp. 2044–2053, Nov. 2012.

[96] Y. Lakys, W. Zhao, J.-O. Klein, and C. Chappert, "Hardening Techniques for MRAM-Based Nonvolatile Latches and Logic," *IEEE Transactions on Nuclear Science*, vol. 59, pp. 1136–1141, Aug. 2012.

[97] K. Ali, F. Li, S. Y. H. Lua, and C.-H. Heng, "Compact spin transfer torque non-volatile flip flop design for power-gating architecture," in *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Oct. 2016, pp. 119–122.

[98] M. Hariyama, S. Ishihara, N. Idobata, and M. Kameyama, "Non-Volatile Multi-Context FPGAs Using Hybrid Multiple-Valued/Binary Context Switching Signals," in *International Conference on Engineering of Reconfigurable Systems & Algorithms (ERSA)*, 2008.

[99] W. S. Zhao, T. Devolder, Y. Lakys, J.-O. Klein, C. Chappert, and P. Mazoyer, "Design considerations and strategies for high-reliable STT-MRAM," *Microelectronics Reliability*, vol. 51, no. 9, pp. 1454–1458, Sep. 2011. [Online]. Available : http://www.sciencedirect.com/science/article/pii/S0026271411002642

[100] E. Deng, W. Kang, Y. Zhang, J.-O. Klein, C. Chappert, and W. Zhao, "Design Optimization and Analysis of Multicontext STT-MTJ/CMOS Logic Circuits," *IEEE Transactions on Nanotechnology*, vol. 14, no. 1, pp. 169–177, Jan. 2015.

[101] E. Deng, L. Anghel, G. Prenat, and W. Zhao, "Multi-context non-volatile content addressable memory using magnetic tunnel junctions," in *2016 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, Jul. 2016, pp. 103–108.

[102] C. Münch, R. Bishnoi, and M. B. Tahoori, "Multi-bit non-volatile spintronic flip-flop," in *2018 Design, Automation Test in Europe Conference Exhibition (DATE)*, Mar. 2018, pp. 1229–1234.

[103] E. I. Vatajelu, G. Di Natale, M. Barbareschi, L. Torres, M. Indaco, and P. Prinetto, "STT-MRAM-Based PUF Architecture Exploiting Magnetic Tunnel Junction Fabrication-Induced Variability," *J. Emerg. Technol. Comput. Syst.*, vol. 13, no. 1, pp. 5 :1–5 :21, May 2016. [Online]. Available : http://doi.acm.org/10.1145/2790302

[104] E. I. Vatajelu, G. Di Natale, and P. Prinetto, "Security primitives (PUF and TRNG) with STT-MRAM," in *2016 IEEE 34th VLSI Test Symposium (VTS)*, Apr. 2016, pp. 1–4.

[105] Z. Cherif Jouini, J.-L. Danger, and L. Bossuet, "Performance evaluation of Physically Unclonable Function by delay statistics," in *2011 IEEE 9th International New Circuits and systems conference*, Jun. 2011, pp. 482–485.

[106] M. Al-Haidary and Q. Nasir, "Physically Unclonable Functions (PUFs) : A Systematic Literature Review," in *2019 Advances in Science and Engineering Technology International Conferences (ASET)*, Mar. 2019, pp. 1–6.

[107] J. Trujillo, C. Merino, and P. Zarkesh-Ha, "SRAM Physically Unclonable Functions Implemented on Silicon Germanium," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–4.

[108] S. Sankaran, S. Shivshankar, and K. Nimmy, "LHPUF : Lightweight Hybrid PUF for Enhanced Security in Internet of Things," in *2018 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)*, Dec. 2018, pp. 275–278.

[109] T. Marukame, T. Tanamoto, and Y. Mitani, "Extracting Physically Unclonable Function From Spin Transfer Switching Characteristics in Magnetic Tunnel Junctions," *IEEE Transactions on Magnetics*, vol. 50, no. 11, pp. 1–4, Nov. 2014.

[110] S. Ghosh, "Spintronics and Security : Prospects, Vulnerabilities, Attack Models, and Preventions," *Proceedings of the IEEE*, vol. 104, no. 10, pp. 1864–1893, Oct. 2016.

[111] Y. Qu, J. Han, B. F. Cockburn, W. Pedrycz, Y. Zhang, and W. Zhao, "A true random number generator based on parallel STT-MTJs," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2017*, Mar. 2017, pp. 606–609.

[112] "e.m.c.2 - Niveaux d'énergie de l'atome d'hydrogène - Emission et absorption de lumière." [Online]. Available : http://e.m.c.2.free.fr/niveaux-energie-hydrogene-emission-absorption.htm

[113] "Modèle de Bohr de l'atome d'hydrogène." [Online]. Available : https://fr.khanacademy.org/science/physics/quantum-physics/atoms-and-electrons/a/bohrs-model-of-hydrogen

[114] E. Rutherford, "The scattering of and particles by matter and the structure of the atom," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 21, no. 125, pp. 669–688, May 1911. [Online]. Available : https://www.tandfonline.com/doi/full/10.1080/14786440508637080

[115] E. Schrödinger, "An Undulatory Theory of the Mechanics of Atoms and Molecules," *Phys. Rev.*, vol. 28, no. 6, pp. 1049–1070, Dec. 1926. [Online]. Available : https://link.aps.org/doi/10.1103/PhysRev.28.1049

[116] C. Schwob and L. Julien, "Le laser : principe de fonctionnement," *Reflets phys.*, no. 21, pp. 12–16, Oct. 2010. [Online]. Available : https://www.refletsdelaphysique.fr/articles/refdp/abs/2010/04/refdp201021p12/refdp201021p12.html

[117] R. Joecklé, "Lasers à gaz," Jan. 2000. [Online]. Available : https://www.techniques-ingenieur.fr/base-documentaire/42452210-sources-laser/download/af3271/lasers-a-gaz.html

[118] "Laser : Fundamentals - Gas lasers." [Online]. Available : http://www.optique-ingenieur.org/en/courses/OPI_ang_M01_C01/co/Contenu_15.html

[119] "Le principe de fonctionnement des différents types de laser." [Online]. Available : http://buthod-adrien.e-monsite.com/pages/partie-1/le-principe-de-fonctionnement-du-laser/

[120] Futura, "Laser." [Online]. Available : https://www.futura-sciences.com/sciences/definitions/physique-laser-1989/

[121] S. Forget, "Les lasers et leurs applications," laboratoire de Physique des Lasers, Université Paris-Nord.

[122] “Laser : Fundamentals - Dye lasers.” [Online]. Available : http://www.optique-ingenieur.org/en/courses/OPI_ang_M01_C01/co/Contenu_16.html

[123] “Laser : Fundamentals - Solid state lasers.” [Online]. Available : http://www.optique-ingenieur.org/en/courses/OPI_ang_M01_C01/co/Contenu_17.html

[124] D. Pureur and A. Biasi, “Les lasers à fibre,” *Photoniques*, no. 51, pp. 47–48, Jan. 2011. [Online]. Available : https://www.photoniques.com/articles/photon/abs/2011/01/photon201151p47/photon201151p47.html

[125] “Micro-PackS : Une plateforme technologique.” [Online]. Available : https://www.pf-micropacks.org/fr/micro-packs/la-plate-forme/

[126] “QuikLaze 50st2 - New Wave Research - PDF Catalogue | Technical Documentation | Brochure.” [Online]. Available : http://pdf.directindustry.com/pdf/new-wave-research/quiklaze-50st2/24449-73341.html

[127] D. Sands, “Pulsed Laser Heating and Melting,” in *Heat Transfer - Engineering Applications, Prof. Vyacheslav Vikhrenko (Ed.)*, 2011. [Online]. Available : http://www.intechopen.com/books/heat-transfer-engineering-applications/pulsed-laser-heating-and-melting

[128] A. Krakovinsky, “Caractérisation sécuritaire des OxRRAM,” Manuscrit de thèse, Université d’Aix-Marseille, Marseille, 2017.

[129] “Logiciel de modélisation COMSOL Multiphysics®.” [Online]. Available : https://www.comsol.fr/

[130] M. Kharbouche-Harrari, R. Wacquez, G. Di Pendina, J. M. Dutertre, J. Postel-Pellerin, D. Aboulkassimi, and J.-M. Portal, “Dual Detection of Heating and Photocurrent attacks (DDHP) Sensor using Hybrid CMOS/STT-MRAM,” in *2019 IEEE 25th International Symposium on On-Line Testing And Robust System Design (IOLTS)*, Jul. 2019.

[131] F. Faccio, “Design Hardening Methodologies for ASICs,” in *Radiation Effects on Embedded Systems*, R. VELAZCO, P. FOUILLAT, and R. REIS, Eds. Dordrecht : Springer Netherlands, 2007, pp. 143–160. [Online]. Available : https://doi.org/10.1007/978-1-4020-5646-8_7

[132] K. Tiri and I. Verbauwhede, “Securing Encryption Algorithms against DPA at the Logic Level : Next Generation Smart Card Technology,” in *Cryptographic Hardware and Embedded Systems - CHES 2003*, ser. Lecture Notes in Computer Science, C. D. Walter, e. K. Koç, and C. Paar, Eds. Springer Berlin Heidelberg, 2003, pp. 125–136.

[133] A. Razafindraibe, M. Robert, and P. Maurine, “Improvement of dual rail logic as a countermeasure against DPA,” in *2007 IFIP International Conference on Very Large Scale Integration*, Oct. 2007, pp. 270–275.

[134] A. Razafindraibe, M. Robert, and P. Maurine, “Formal Evaluation of the Robustness of Dual-Rail Logic Against DPA Attacks,” in *Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation*, ser. Lecture Notes in Computer Science, J. Vounckx, N. Azemard, and P. Maurine, Eds. Springer Berlin Heidelberg, 2006, pp. 634–644.

[135] R. Soares, N. Calazans, V. Lomné, P. Maurine, L. Torres, and M. Robert, “Evaluating the Robustness of Secure Triple Track Logic Through Prototyping,” in *Proceedings of the 21st Annual Symposium on Integrated Circuits and System Design*, ser. SBCCI ’08. New York, NY, USA : ACM, 2008, pp. 193–198, event-place : Gramado, Brazil. [Online]. Available : http://doi.acm.org/10.1145/1404371.1404425

[136] H. Li, G. Ma, G. Li, G. Wang, and T. Zhou, “A New Protect Cryptographic Circuit Approach Using Dynamic Current Model Logic Circuit,” in *2007 International Conference on Mechatronics and Automation*, Aug. 2007, pp. 2221–2225.

[137] C. Clavier, J.-S. Coron, and N. Dabbous, “Differential Power Analysis in the Presence of Hardware Countermeasures,” in *Cryptographic Hardware and Embedded Systems — CHES 2000*, ser. Lecture Notes in Computer Science, e. K. Koç and C. Paar, Eds. Springer Berlin Heidelberg, 2000, pp. 252–263.

[138] T. S. Messerges, "Securing the AES Finalists Against Power Analysis Attacks," in *Fast Software Encryption*, ser. Lecture Notes in Computer Science, G. Goos, J. Hartmanis, J. van Leeuwen, and B. Schneier, Eds. Springer Berlin Heidelberg, 2001, pp. 150–164.

[139] M. Rivain and E. Prouff, "Provably Secure Higher-Order Masking of AES," in *Cryptographic Hardware and Embedded Systems, CHES 2010*, ser. Lecture Notes in Computer Science, S. Mangard and F.-X. Standaert, Eds. Springer Berlin Heidelberg, 2010, pp. 413–427.

[140] J.-M. Dutertre, V. Beroulle, P. Candelier, L.-B. Faber, M.-L. Flottes, P. Gendrier, D. Hély, R. Leveugle, P. Maistri, G. Di Natale, A. Papadimitriou, and B. Rouzeyre, "The case of using CMOS FD-SOI rather than CMOS bulk to harden ICs against laser attacks," in *2018 IEEE 24th International Symposium on On-Line Testing And Robust System Design (IOLTS)*, Jul. 2018, pp. 214–219.

[141] A. Beit-Grogger and J. Riegebauer, "Integrated circuit having an active shield," US Patent US6 962 294B2, Nov., 2005. [Online]. Available : https://patents.google.com/patent/US6962294/en

[142] P. Laackmann and H. Taddiken, "Apparatus for protecting an integrated circuit formed in a substrate and method for protecting the circuit against reverse engineering," US Patent US6 798 234B2, Sep., 2004. [Online]. Available : https://patents.google.com/patent/US6798234/en

[143] O. Kömmerling and M. G. Kuhn, "Design Principles for Tamper-resistant Smartcard Processors," in *Proceedings of the USENIX Workshop on Smartcard Technology on USENIX Workshop on Smartcard Technology*, ser. WOST'99. Berkeley, CA, USA : USENIX Association, 1999, pp. 2–2, event-place : Chicago, Illinois. [Online]. Available : http://dl.acm.org/citation.cfm?id=1267115.1267117

[144] J. Teifel, "Self-Voting Dual-Modular-Redundancy Circuits for Single-Event-Transient Mitigation," *IEEE Transactions on Nuclear Science*, vol. 55, no. 6, pp. 3435–3439, Dec. 2008.

[145] V. Petrović, G. Schoof, and Z. Stamenković, "Redundant circuits with latchup protection," in *2013 IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS)*, Dec. 2013, pp. 117–120.

[146] S. Okumura, Y. Nakata, K. Yanagida, Y. Kagiyama, S. Yoshimoto, H. Kawaguchi, and M. Yoshimoto, "Low-energy block-level instantaneous comparison 7t SRAM for dual modular redundancy," *IEICE Electron. Express*, vol. 9, no. 6, pp. 470–476, 2012. [Online]. Available : https://www.jstage.jst.go.jp/article/elex/9/6/9_6_470/_article

[147] R. Gong, W. Chen, F. Liu, K. Dai, and Z. Wang, "A New Approach to Single Event Effect Tolerance Based on Asynchronous Circuit Technique," *J Electron Test*, vol. 24, no. 1, pp. 57–65, Jun. 2008. [Online]. Available : https://doi.org/10.1007/s10836-007-5029-z

[148] F. Sellers, M. Xiao, and L. Bearnson, *Error Detecting Logic for Digital Computers*, ser. McGraw-Hill, 1968.

[149] R. E. Lyons and W. Vanderkulk, "The Use of Triple-Modular Redundancy to Improve Computer Reliability," *IBM Journal of Research and Development*, vol. 6, no. 2, pp. 200–209, Apr. 1962.

[150] P. Reviriego, C. J. Bleakley, and J. A. Maestro, "Diverse Double Modular Redundancy : A New Direction for Soft-Error Detection and Correction," *IEEE Design Test*, vol. 30, no. 2, pp. 87–95, Apr. 2013.

[151] M. Nicolaidis, "Time Redundancy Based Soft-Error Tolerance to Rescue Nanometer Technologies," in *Proceedings of the 1999 17TH IEEE VLSI Test Symposium*, ser. VTS '99. Washington, DC, USA : IEEE Computer Society, 1999, p. 86. [Online]. Available : http://dl.acm.org/citation.cfm?id=832299.836499

[152] R. A. Camponogara Viera, R. P. Bastos, J. M. Dutertre, P. Maurine, and R. I. Jadue, "Method for evaluation of transient-fault detection techniques," *Microelectronics Reliability*, vol. 76-77, pp. 68–74, Sep. 2017. [Online]. Available : http://www.sciencedirect.com/science/article/pii/S0026271417302913

[153] J. M. Dutertre, R. Possamai Bastos, O. Potin, M. L. Flottes, B. Rouzeyre, and G. Di Natale, "Sensitivity tuning of a bulk built-in current sensor for optimal transient-fault detection," *Microelectronics Reliability*, vol. 53, no. 9, pp. 1320–1324, Sep. 2013. [Online]. Available : http://www.sciencedirect.com/science/article/pii/S0026271413002448

[154] C. Champeix, N. Borrel, J.-M. Dutertre, B. Robisson, M. Lisart, and A. Sarafianos, "Experimental validation of a Bulk Built-In Current Sensor for detecting laser-induced currents," in *2015 IEEE 21st International On-Line Testing Symposium (IOLTS)*, Jul. 2015, pp. 150–155.

[155] E. H. Neto, I. Ribeiro, M. Vieira, G. Wirth, and F. L. Kastensmidt, "Evaluating Fault Coverage of Bulk Built-in Current Sensor for Soft Errors in Combinational and Sequential Logic," in *2005 18th Symposium on Integrated Circuits and Systems Design*, Sep. 2005, pp. 62–67.

[156] E. H. Neto, I. Ribeiro, M. Vieira, G. Wirth, and F. L. Kastensmidt, "Using Bulk Built-in Current Sensors to Detect Soft Errors," *IEEE Micro*, vol. 26, no. 5, pp. 10–18, Sep. 2006.

[157] G. Wirth, "Bulk built in current sensors for single event transient detection in deep-submicron technologies," *Microelectronics Reliability*, vol. 48, no. 5, pp. 710–715, May 2008. [Online]. Available : http://www.sciencedirect.com/science/article/pii/S0026271408000085

[158] A. Simionovski and G. Wirth, "Simulation Evaluation of an Implemented Set of Complementary Bulk Built-In Current Sensors With Dynamic Storage Cell," *IEEE Transactions on Device and Materials Reliability*, vol. 14, no. 1, pp. 255–261, Mar. 2014.

[159] R. Possamai Bastos, L. A. Guimarães, F. Sill Torres, and L. Fesquet, "Architectures of bulk built-in current sensors for detection of transient faults in integrated circuits," *Microelectronics Journal*, vol. 71, pp. 70–79, Jan. 2018. [Online]. Available : http://www.sciencedirect.com/science/article/pii/S0026269217306882

[160] A. Krakovinsky, M. Bocquet, R. Wacquez, J. Coignus, and J.-M. Portal, "Thermal laser attack and high temperature heating on HfO2-based OxRAM cells," in *2017 IEEE 23rd International Symposium on On-Line Testing and Robust System Design (IOLTS)*, Jul. 2017, pp. 85–89.

[161] M. Kharbouche-Harrari, R. Alhalabi, J. Postel-Pellerin, R. Wacquez, D. Aboulkassimi, E. Nowak, I. L. Prejbeanu, G. Prenat, and G. Di Pendina, "MRAM : from STT to SOT, for security and memory," in *2018 Conference on Design of Circuits and Integrated Systems (DCIS)*, Nov. 2018, pp. 1–6.

[162] M. Kharbouche-Harrari, G. Di Pendina, R. Wacquez, B. Dieny, D. Aboulkassimi, J. Postel-Pellerin, and J. Portal, "Light-Weight Cipher Based on Hybrid CMOS/STT-MRAM : Power/Area Analysis," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–5.

[163] J. Hosseinzadeh and M. Hosseinzadeh, "A Comprehensive Survey on Evaluation of Lightweight Symmetric Ciphers : Hardware and Software Implementation," pp. 31–41, Jul. 2016.

[164] C. De Cannière, O. Dunkelman, and M. Knežević, "KATAN and KTANTAN — A Family of Small and Efficient Hardware-Oriented Block Ciphers," in *Cryptographic Hardware and Embedded Systems - CHES 2009*, ser. Lecture Notes in Computer Science, C. Clavier and K. Gaj, Eds. Springer Berlin Heidelberg, 2009, pp. 272–288.

[165] L. Knudsen, G. Leander, A. Poschmann, and M. J. B. Robshaw, "PRINTcipher : A Block Cipher for IC-Printing," in *Cryptographic Hardware and Embedded Systems, CHES 2010*, ser. Lecture Notes in Computer Science, S. Mangard and F.-X. Standaert, Eds. Springer Berlin Heidelberg, 2010, pp. 16–32.

[166] "ISO/IEC 29192-2 :2012 Information technology – Security techniques – Lightweight cryptography – Part 2 : Block ciphers," Jan. 2012. [Online]. Available : https://www.iso.org/standard/56552.html

[167] M. Tahoori, S. M. Nair, R. Bishnoi, S. Senni, J. Mohdad, F. Mailly, L. Torres, P. Benoit, A. Gamatie, P. Nouet, F. Ouattara, G. Sassatelli, K. Jabeur, P. Vanhauwaert, A. Atitoaie, I. Firastrau, G. Di Pendina, and G. Prenat, "Using multifunctional standardized stack as universal spintronic technology for IoT," in *2018 Design, Automation Test in Europe Conference Exhibition (DATE)*, Mar. 2018, pp. 931–936.

[168] "GREAT project : heteroGeneous integRated magnetic tEchnology using multifonctionnal stAndardized sTack (MSS)." [Online]. Available : http://www.great-research.eu/

[169] "ModelSim®." [Online]. Available : https://www.mentor.com/products/fv/modelsim/

[170] "Design Compiler Graphical." [Online]. Available : https://www.synopsys.com/implementation-and-signoff/rtl-synthesis-test/design-compiler-graphical.html

[171] "Assura Physical Verification." [Online]. Available : https://www.cadence.com/content/cadence-www/global/en_US/home/tools/digital-design-and-signoff/silicon-signoff/assura-physical-verification.html

[172] L. Landau and E. Lifshitz, "On the theory of the dispersion of magnetic permeability in ferromagnetic bodies," *Phys. Z. Sowjetunion*, vol. 8, no. 153, pp. 101–114, 1935.

[173] J. C. Slonczewski, "Currents, torques, and polarization factors in magnetic tunnel junctions," *Phys. Rev. B*, vol. 71, no. 2, p. 024411, Jan. 2005. [Online]. Available : https://link.aps.org/doi/10.1103/PhysRevB.71.024411

[174] R. Alhalabi, G. Di Pendina, I. L. Prejbeanu, and E. Nowak, "High speed and high-area efficiency non-volatile look-up table design based on magnetic tunnel junction," in *2017 17th Non-Volatile Memory Technology Symposium (NVMTS)*, Aug. 2017, pp. 1–4.

[175] "Arduino Due specifications." [Online]. Available : https://store.arduino.cc/due

[176] "Arduino Mega 2560 Rev3." [Online]. Available : https://store.arduino.cc/mega-2560-r3

[177] "STM32f446re." [Online]. Available : https://www.st.com/content/st_com/en/products/microcontrollers-microprocessors/stm32-32-bit-arm-cortex-mcus/stm32-high-performance-mcus/stm32f4-series/stm32f446/stm32f446re.html

[178] Digilent, "Zybo Z7 Reference Manual." [Online]. Available : https://reference.digilentinc.com/reference/programmable-logic/zybo-z7/reference-manual

[179] "Arduino website." [Online]. Available : https://www.arduino.cc/

[180] "ATSAM3x8e - 32-bit SAM Microcontrollers." [Online]. Available : https://www.microchip.com/wwwproducts/en/ATsam3x8e

[181] "LD1117a." [Online]. Available : https://www.st.com/en/power-management/ld1117a.html

[182] "SN74avc20t245 20-Bit Dual Supply Bus Transceiver with Configurable Voltage Translation and 3-State Outputs | TI.com." [Online]. Available : http://www.ti.com/product/SN74AVC20T245#

[183] "DesignSpark PCB Software." [Online]. Available : https://www.rs-online.com/designspark/pcb-software

[184] L. Miseur, "Nombres quantiques et orbitales." [Online]. Available : http://www.lachimie.net

[185] W. Pauli, "Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren," *Z. Physik*, vol. 31, no. 1, pp. 765–783, Feb. 1925. [Online]. Available : https://doi.org/10.1007/BF02980631

[186] D. P. Wong, "Theoretical justification of Madelung's rule," *J. Chem. Educ.*, vol. 56, no. 11, p. 714, Nov. 1979. [Online]. Available : https://doi.org/10.1021/ed056p714

Résumé

Cette dernière décennie a été le théâtre du développement rapide de l'Internet des Objets. Cette expansion s'accompagne du renforcement des besoins et contraintes des circuits intégrés : une consommation faible et une surface silicium maîtrisée. Toutefois, cet engouement récent pour les objets connectés pousse souvent les fabricants à précipiter la mise sur le marché de leurs produits, parfois au détriment de la sécurité.

Dans le cadre des travaux entrepris lors de cette thèse, nous nous sommes particulièrement intéressés aux atouts et inconvénients que peut apporter l'hybridation de la technologie CMOS avec la technologie mémoire non-volatile émergente STT-MRAM. Ces architectures innovantes doivent permettre le développement d'applications faible consommation visant la sécurité des objets connectés. Pour cela, la conception d'un algorithme de cryptographie légère hybride CMOS/STT-MRAM basé sur le chiffrement PRESENT a été réalisée.

Ainsi, la première étude menée a consisté à étudier la robustesse de jonctions mémoires STT-MRAMs unitaires face aux attaques physiques de type perturbation, avant leur intégration dans le chiffrement. Pour ce faire, des injections de fautes Laser ont été effectuées afin d'évaluer l'intégrité des données stockées.

Suite aux observations des expérimentations réalisées sur ces mémoires de type STT-MRAM perpendiculaires, un nouveau capteur d'attaques physiques basé sur cette technologie mémoire a été proposé, le DDHP. Ce détecteur permet la détection simultanée d'attaques photoélectriques et d'attaques thermiques qui peuvent viser les circuits intégrés.

Mots clés : STT-MRAM, hybridation, cryptographie légère, sécurité matérielle, capteur DDHP.

Abstract

In the last decade, the Internet of Things deployment highlighted new needs and constraints in terms of consumption and area for integrated circuits. However, the recent craze for connected objects and due to the extremely pressing time-to-market demand, the manufacturers commercialize their products, sometimes at the expense of their security.

The main focus of the work undertook during this thesis consists in the hybridization of the CMOS technology with the emerging non-volatile memory technology STT-MRAM. This study aims to determine the assets and drawbacks of this hybridization. These innovating architectures must allow the development of low power applications and support the growth of secured connected objects. Thus, the design of a hybrid CMOS/STT-MRAM lightweight cryptographic algorithm based on the PRESENT cipher is realised.

This is how the first study carried out consisted in investigating the robustness of STT-MRAM junctions facing physical attacks, before their integration in the cryptographic algorithm. To do this, laser fault injections were performed in order to evaluate the integrity of the sensitive data stored in the cells.

Following the experiments carried out on perpendicular STT-MRAM memories, a new physical attack detector based on this memory technology is proposed, designated by DDHP. This sensor allows simultaneous detection of photoelectrical and thermal attacks that can target integrated circuits.

Keywords : STT-MRAM, hybridization, lightweight cryptography, hardware security, DDHP sensor.