UNIVERSITE D'AIX MARSEILLE
MINES ST-ETIENNE (EMSE)
ECOLE DOCTORALE 353

Institut Matériaux Microélectronique Nanosciences de Provence (IM2NP) / UMR7334

Thèse présentée pour obtenir le grade universitaire de docteur

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

Marc LACRUCHE

Caractérisation sécuritaire de circuits basse-consommation face aux attaques par laser

Soutenue le JJ/MM/AAAA devant le jury :

Jean-Luc AUTRAN	Université d'Aix Marseille	Président du Jury
Viktor FISCHER	Université de Saint-Etienne	Rapporteur
Luc HEBRARD	Université de Strasbourg	Rapporteur
Philippe MAURINE	LIRMM	Examinateur
Edith KUSSENER	Université d'Aix Marseille	Directrice de thèse
Jean-Baptiste RIGAUD	Mines Saint-Etienne	Encadrant de thèse
Jean-Max DUTERTRE	Mines Saint-Etienne	Encadrant de thèse
Mathieu LISART	ST Microelectronics	Invité

Résumé

L'augmentation de la complexité des circuits intégrés et l'explosion du nombre d'appareils mobiles aujourd'hui rend la minimisation de la consommation primordiale lors de la conception des circuits. Cependant, il est nécessaire de s'assurer que les techniques employées ne compromettent pas la sécurité des données lorsqu'elles sont utilisées dans des circuits sensibles. Et ce particulièrement face aux attaques physiques, les appareils mobiles déployés à grande échelle pouvant facilement tomber aux mains d'un attaquant, lui garantissant ainsi un accès direct au circuit.

Dans ce cadre, ce travail vise à évaleur l'impact des techniques de *body-biasing* sur la résistance des circuits aux attaques par injection de faute laser. Ces techniques permettent d'ajuster dynamiquement le compromis entre consommation et performance d'un circuit en modifiant la tension de polarisation des caissons. Le manuscrit se découpe en quatre chapitres. Il commence par un état de l'art couvrant la cryptographie, les attaques physiques (en particulier l'injection de faute laser), et les techniques de réduction de la consommation. Puis, le banc de test laser utilisé pendant ces travaux est présenté ainsi que le travail effectué pour permettre son automatisation et une première étude sur l'impact des impulsions laser de courte durée (30 ps) sur les mémoires SRAM. Le troisième chapitre rapporte les résultats d'une campagne d'injection de faute laser sur des mémoires soumises à des tensions d'alimentation réduites et au *body-biasing*. Celle-ci permet de mettre en évidence une augmentation de la sensibilité au laser des circuits lorsque leur tension d'alimentation est réduite et que le Forward Body Biasing est utilisé. A partir de ces résultats, le dernier chapitre propose une méthode utilisant les capacités basse-consommation d'un microcontrôleur pour durcir un AES matériel présent sur ce dernier.

Ces travaux permettent ainsi de montrer que les techniques de réduction de la consommation peuvent constituer un risque sécuritaire potentiel si elle ne sont pas prises en compte correctement. Cependant, les capacités apportées au circuit dans ce cadre peuvent être détournées pour améliorer sa résistance aux attaques.

Abstract

The increasing complexity of integrated circuits and the explosion of the number of mobile devices today makes power consumption minimisation a priority in circuit design. However, it is necessary to make sure that the design methods used to achieve low power consumption do not compromise the security of sensitive data processed by those circuits. In this regard, physical attacks are a particular concern, as mobile devices are often small and widely deployed and as such, can easily fall into an attackers hand, giving him physical access to the circuit.

In this context, this work aims at evaluating the impact of body-biasing methods on circuit vulnerability to laser fault injection. These methods allow to dynamically adjust the performance/consumption ratio of a circuit by modifying the bias voltage of the N-Wells and P-Wells in a triple well architecture. This paper is divided in four chapters. It begins by introducing cryptography, physical attacks (focusing on laser fault injection) and low power design methods. Then the test bench used during this thesis is described, as well as the developpement work done in order to allow its automation. Then an initial study of the impact of short duration laser pulses on SRAM memories is presented. The third chapter reports the results of a laser fault injection campaign on memories subjected to lowered supply voltage and Forward Body-Biasing. The results show a sensitivy increase of the circuits when supply voltage is lowered and FBB is activated. Based on these results, the last chapter introduces a method using the body-biasing and voltage scaling capabilities of a microcontroller to harden a hardware AES embedded on the latter.

In conclusion, this works shows that low-power design methods can induce additional security risks if they are not carefully taken into account. However the additional capabilities of the circuits intended for power consumption reduction can

be used in a different way to enhance device resillience to attacks.

Table des matières

Table des figures

Liste des tableaux

Introduction générale

L'apparition de la cryptographie remonte à l'antiquité, mais c'est à partir du XX^e siècle que l'usage de machines, d'abord électromécaniques puis électroniques, va permettre une sophistication des méthodes de chiffrement et une généralisation de son utilisation. Aujourd'hui, les communications sécurisées sont au cœur de nombreux aspect de notre société (téléphones mobiles, système bancaire, internet, PayTV, etc.). Par conséquent, de plus en plus d'objets contiennent des circuits intégrés permettant d'effectuer des opérations cryptographiques. La carte à puce est notamment un des exemples les plus communs de ce type d'objets. La robustesse des méthodes de chiffrement utilisées pour assurer la confidentialité, l'intégrité et l'authenticité des communications, est donc primordiale pour le bon fonctionnement de ces différents services.

Bien que les algorithmes de chiffrement utilisés de nos jours soient considérés comme étant sûrs d'un point de vue mathématique, les circuits permettant d'effectuer les calculs, eux, peuvent être sujets à des failles. Celles-ci ne sont pas liées à une implémentation incorrecte des algorithmes, mais simplement à la technologie de fabrication des composants. Lorsqu'un circuit intégré fonctionne, son activité transparaît au travers de différentes grandeurs physiques mesurables comme sa consommation d'énergie ou ses émissions électromagnétiques. D'une manière similaire, le circuit est affecté par son environnement, des facteurs comme la température ou sa tension d'alimentation pouvant impacter son fonctionnement.

Ces phénomènes peuvent être exploités par un adversaire pour attaquer un système cryptographique. Il est possible de distinguer deux types d'attaques physiques. Premièrement, les attaques par observation consistent à utiliser des mesures de l'activité du circuit pour retrouver des informations sensibles sur le système,

comme les clés de chiffrement utilisées. Deuxièmement, les attaques par injection de fautes, elles, visent à perturber le circuit via différents moyens (impulsions électromagnétiques, faisceau laser, etc.) pour provoquer des erreurs de calcul pouvant ensuite être utilisées pour remonter aux données sensibles.

En parallèle de la généralisation de l'utilisation des communications sécurisées, l'évolution de l'électronique est confrontée à un besoin toujours croissant d'une consommation d'énergie réduite. Deux facteurs contribuent à cela. D'une part, la consommation d'énergie est le premier obstacle pour augmenter la complexité des circuits face aux effets d'échauffement. D'autre part, la forte augmentation du nombre d'appareils mobiles rend la limitation consommation d'énergie cruciale pour augmenter la durée de vie des batteries.

Dans le cas de circuits basse consommation embarquant des capacités cryptographiques, il est donc primordial de s'assurer que les techniques utilisées pour réduire leur consommation ne compromettent pas la sécurité des opérations effectuées. De ce point de vue, les attaques physiques sont une préoccupation importante. En effet, la nature des appareils mobiles en fait des cibles idéales : ils sont souvent peu coûteux et déployés largement, ce qui permet à un attaquant d'obtenir facilement un accès physique aux circuits.

Dans ce contexte, le projet MAGE, dans lequel ces travaux s'inscrivent, a pour but de développer des microcontrôleurs basse consommation sécurisés en technologie CMOS 90 nm. MAGE comporte plusieurs volets, dont un un axé sur l'étude de l'impact sécuritaire des techniques de réduction de la consommation, et en particulier de celles utilisées dans les microcontrôleurs issus du projet, basées sur le *body-biasing*. Au de sein de ce volet, différents groupes de travail visent à évaluer ces techniques face à différents types d'attaques (injection électromagnétique, analyse de la consommation etc.). Dans ce cadre, l'objectif de cette thèse est de caractériser l'impact du *body-biasing* sur la résistance des circuits face à l'injection de fautes laser.

Le premier chapitre aborde les thématiques nécessaires à la compréhension de ce document. La cryptographie et les attaques physiques seront d'abord présentées. Puis l'injection de fautes laser sera approfondie : son principe de fonctionnement,

les différents paramètres entrant en jeu et les modèles permettant de la simuler. Ensuite les contre-mesures permettant de protéger les circuits face aux attaques en fautes seront présentées. Enfin, le chapitre se terminera par une section sur les différentes méthodes de réduction de la consommation, et en particulier les méthodes de *body-biasing* que ces travaux visent à évaluer.

Le second chapitre commence par introduire le banc de test utilisé durant cette thèse, et les travaux effectués pour permettre son automatisation. Puis une première campagne de tests visant à évaluer l'influence de l'utilisation d'impulsions laser de courte durée sur l'injection de fautes dans des mémoires de type SRAM est présentée, ainsi qu'une adaptation du modèle de simulation pour ces durées d'impulsion.

Le troisième chapitre rapporte les résultats d'une étude sur l'influence de la tension d'alimentation sur la sensibilité au laser de la mémoire RAM d'un microcontrôleur. Puis, une autre étude est effectuée, sur l'impact du *body-biasing* et de la tension d'alimentation sur la sensibilité d'un registre de l'accélérateur cryptographique matériel d'un autre microcontrôleur.

Le dernier chapitre s'appuie sur les résultats du chapitre précédent pour montrer comment les techniques de réduction de la consommation peuvent être détournées de leur usage initialement prévu pour améliorer la robustesse d'un circuit face à l'injection de fautes laser. Une caractérisation d'un AES matériel face à l'injection laser est d'abord menée. Le principe de la méthode de durcissement proposée est ensuite expliqué, puis celle-ci est appliquée à l'AES testé précedemment. Enfin, une campagne d'évaluation sur l'efficacité de la méthode est présentée.

CHAPITRE 1

Etat de l'art

1.1 Introduction

Comme mentionné dans l'introduction générale, au cours des dernières années, la consommation des circuits intégrés et leur sécurité sont devenues deux problématiques majeures à envisager lors des étapes de leur conception.

Dans ce contexte, le projet MAGE vise à développer des microcontrôleurs sécurisés et basse-consommation en technologie CMOS 90 nm. Le volet du projet traitant de la sécurité, se découpe en plusieurs parties visant à évaluer les méthodes de réduction de la consommation employées dans la conception des microcontrôleurs face à différents types d'attaques physique. Ces travaux ont ainsi pour but d'évaluer l'impact des techniques dites de *body-biasing* sur la résistances des circuits à l'injection de faute laser.

Ainsi, ce chapitre abordera les thématiques nécessaires à la compréhension de ce document : bases de la cryptographie, des attaques matérielles et des contre-mesures qui y sont associées, de l'injection de fautes par laser et des modèles permettant de la simuler, et enfin, des techniques de conception basse consommation à évaluer.

La première partie traite la cryptographie, et en particulier l'algorithme AES qui sera utilisé dans la suite de ces travaux.

Les attaques physiques seront ensuite présentées, en mettant l'accent sur les attaques par injection de fautes. L'injection laser sera alors approfondie, les différents

paramètres entrant en jeu seront détaillés et le modèle de simulation utilisé par la suite sera décrit. Les contre-mesures permettant de se prémunir des attaques physiques seront également évoquées.

Enfin, ce chapitre terminera la présentation des méthodes de conception basse-consommation, et en particulier celles de *body-biasing*, basées sur une polarisation du substrat variable, qui seront l'objet de l'étude ce manuscrit.

1.2 Notions de cryptographie

La cryptographie (du grec *kryptos* - caché et *graphein* - écrire) regroupe l'étude des communications sécurisées. Bien qu'on en trouve des traces remontant à l'antiquité, les méthodes de chiffrement utilisées jusqu'à la fin du XIX^e siècle se basaient principalement sur des permutations et la substitution des lettres d'un alphabet lors de la transmission d'un message.

La cryptographie moderne commence réellement à partir de 1883 avec l'essai "La Cryptographie Militaire" d'Auguste Kerckhoffs [16]. Celui-ci pose entre autres le principe suivant lequel la sécurité d'un système de chiffrement ne doit pas reposer sur le secret de son fonctionnement mais plutôt sur une clé secrète. A partir de la seconde guerre mondiale, les ordinateurs commencent à être utilisés à des fins cryptographiques, ouvrant la porte à une complexification des méthodes de chiffrement et permettant une utilisation plus répandue de la cryptographie.

1.2.1 Les méthodes de chiffrement

Les méthodes de chiffrement modernes se divisent en deux catégories : les méthodes de chiffrement symétriques et asymétriques.

1.2.1.1 Le chiffrement symétrique

Un algorithme de chiffrement symétrique utilise une même clé pour le chiffrement et le déchiffrement du message. Cette clé doit être tenue secrète. Elle est partagée uniquement entre les personnes autorisées à accéder au contenu du message. La figure 1.1 montre le principe de fonctionnement d'un algorithme de chiffrement

symétrique : la personne A chiffre le message M à l'aide de la clé K et obtient le message chiffré C. Ce dernier peut alors être transmis sans risque à la personne B via un canal non sécurisé. B peut ensuite déchiffrer C à l'aide de la même clé K pour obtenir le message M original. Un adversaire interceptant C ne peut en revanche pas remonter à M s'il n'a pas connaissance de la clé K.

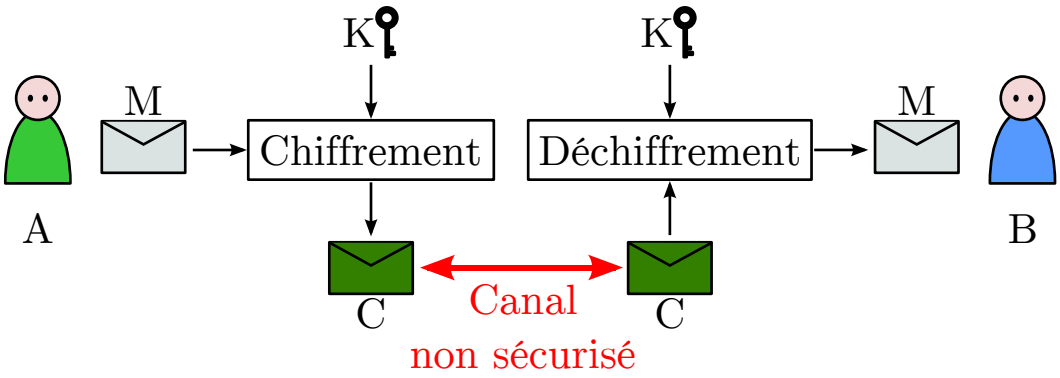


FIGURE 1.1 – Principe de fonctionnement d'un algorithme de chiffrement symétrique.

Deux types d'algorithmes symétriques existent : les algorithmes de chiffrement par blocs et les algorithmes de chiffrement par flot.

- Les algorithmes de chiffrement par blocs divisent les données en blocs de n bits (où n dépend de l'algorithme, généralement 32, 64, 128 etc.) puis chiffrent chaque bloc pour obtenir des blocs de n bits en sortie. Les algorithmes de chiffrement par bloc incluent notamment AES [17], Data Encryption Standard (DES) [18] et Blowfish [19].

- Les algorithmes de chiffrement par flot chiffrent les données de manière continue (généralement octet par octet). Ils incluent RC4, A5/1.

1.2.1.2 Le chiffrement asymétrique

Le chiffrement asymétrique utilise des clés différentes pour le chiffrement et le déchiffrement. Le principe a été introduit par Diffie et Hellman en 1976 [20], il utilise une clé publique pour le chiffrement qui peut être distribuée librement et une clé privée pour le déchiffrement, qui ne doit être connue que du destinataire du message. Comme sur la figure 1.2, pour envoyer un message à B, A doit utiliser la

clé publique de B qui est distribuée librement. Puis, pour déchiffrer le message, B devra utiliser sa clé privée, connue de lui seul.

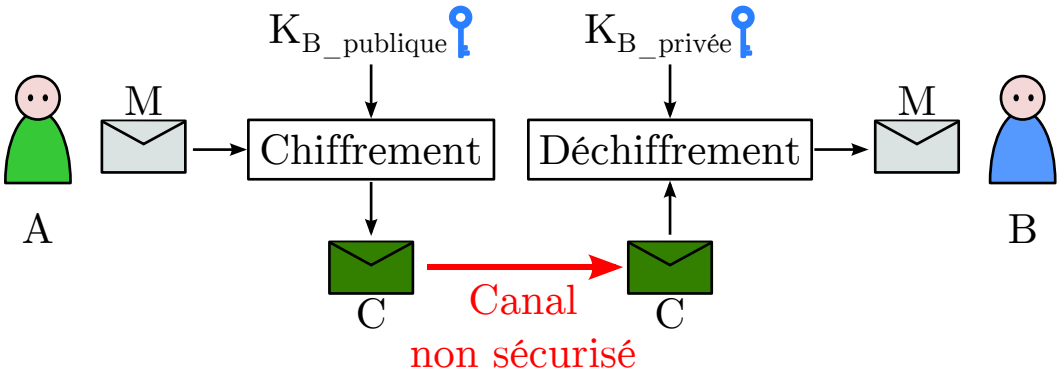


Figure 1.2 – Principe de fonctionnement d'un algorithme de chiffrement asymétrique.

L'algorithme de chiffrement asymétrique le plus couramment utilisé est le RSA, introduit en 1977 [21] par Rivest Shamir et Adleman. Plus récemment l'utilisation de courbes elliptiques pour effectuer du chiffrement asymmétrique a été introduite par Miller et Koblitz en 1985 [22].

Les algorithmes de chiffrement asymétriques ne seront pas étudiés dans la suite de ce manuscrit, et ne seront donc pas abordés plus en détail.

1.2.2 AES

L'AES [17] est l'algorithme de chiffrement qui sera utilisé comme exemple dans la suite de ces travaux. C'est un algorithme symétrique de chiffrement par blocs. A l'origine appelé Rijndael, il a été standardisé et renommé par le NIST en 2001 suite à un concours organisé en 1997 pour remplacer le DES, celui-ci n'étant alors plus considéré comme sûr face à l'augmentation de la puissance de calcul des ordinateurs (une attaque par force brute, ie. l'essai de toutes les clés possibles, prendrait un temps suffisamment court pour être envisageable).

L'AES est basé sur des réseaux de substitutions-permutations et utilise des blocs de 128 bits et des clés de 128,192 ou 256 bits.

Lors de l'exécution de l'algorithme, le bloc de 128 bits traité est considéré comme une matrice 4x4 de 16 octets que l'on nomme *state*. L'algorithme se divise en plu-

sieurs tours (aussi appelés rondes) dont le nombre dépend de la taille de la clé (10 pour 128 bits, 12 pour 196 bits et 14 pour 256 bits). La structure de l'algorithme est présentée sur la figure 1.3.

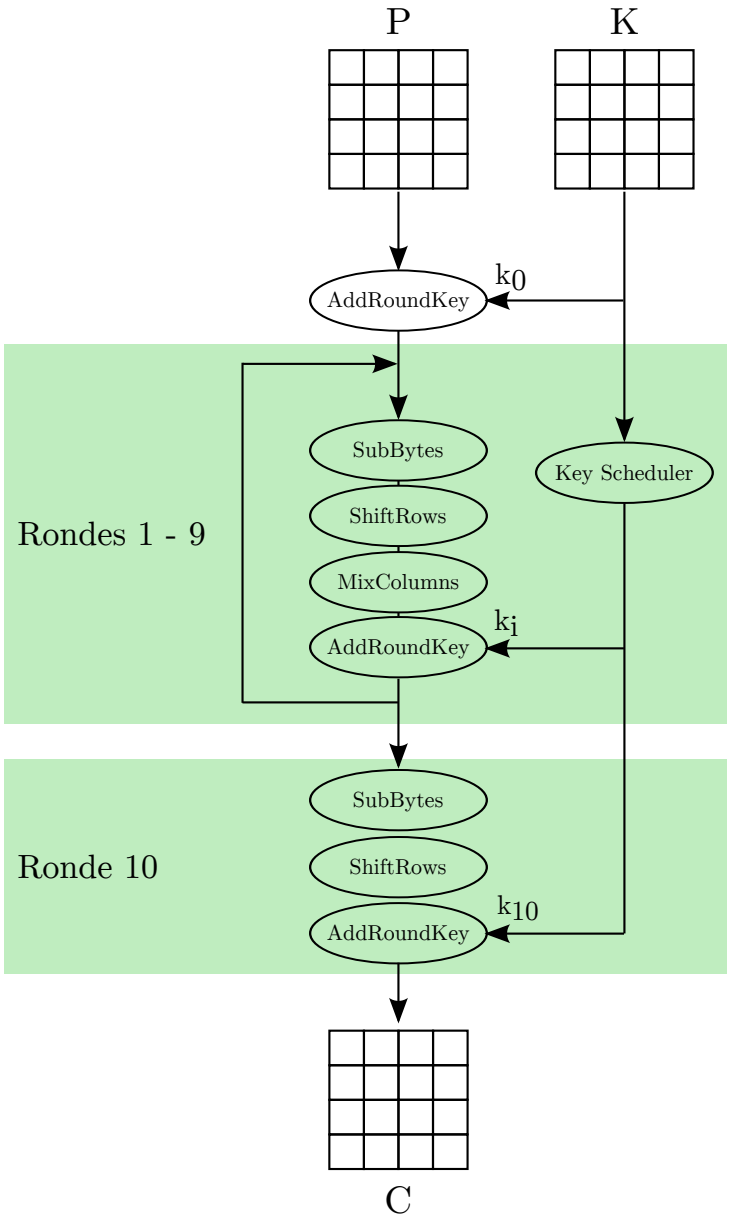


FIGURE 1.3 – Structure de l'algorithme AES pour une clé de 128bits

Chaque tour comporte jusqu'à 4 transformations :

- AddRoundKey : Un XOR bit à bit entre le *state* et une clé de ronde.
- SubBytes : Une opération de substitution non-linéaire où les octets du *state*

sont remplacés un à un à l'aide d'une table de substitution appelée S-Box comme sur la figure 1.4.

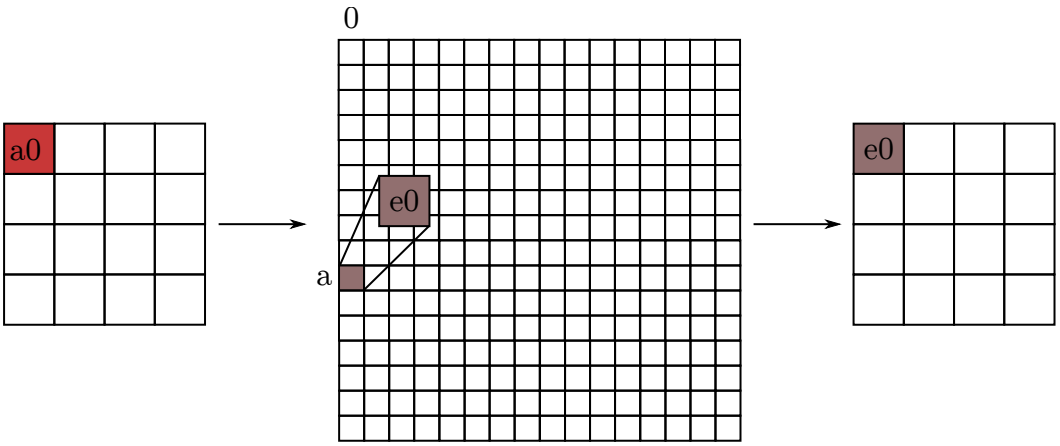


Figure 1.4 – Opération SubBytes de l'AES.

- ShiftRows : Les lignes du *state* sont décalées de manière cyclique de n octets vers la gauche, où n est l'indice de la ligne du *state*, comme sur la figure 1.5. La première ligne n'est pas décalée, la deuxième est décalée d'un octet, la troisième de deux octets et la quatrième de trois octets.

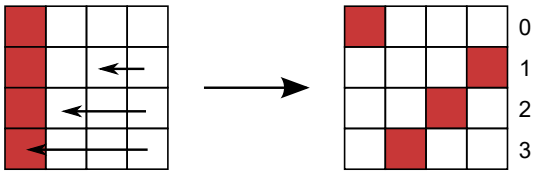


Figure 1.5 – Opération ShiftRows de l'AES.

- MixColumns : chaque colonne du *state* est considérée comme étant un polynôme dans $\mathbb{GF}_{2^8}$ et multipliée modulo $x^4 + 1$ par le polynôme $a(x) = 03x^3 + 01x^2 + 01x + 02$ ce qui peut s'écrire sous la forme d'une multiplication de matrices comme sur la figure 1.6.

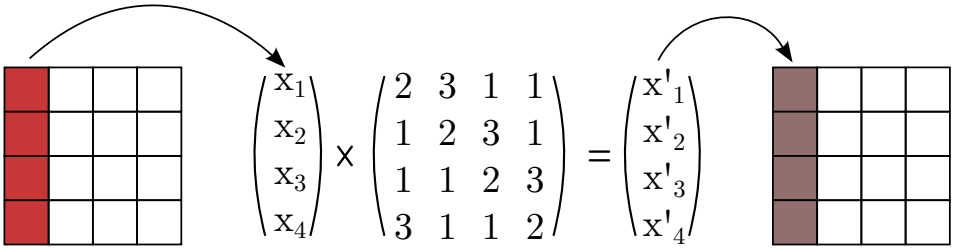


FIGURE 1.6 – Opération MixColumn de l'AES.

- La dernière ronde ne comporte pas d'opération MixColumns.

KeyScheduler : Cette opération s'effectue en parallèle du reste du calcul et permet de calculer les sous-clés de ronde utilisées pour l'opération AddRoundKey à partir de la clé de chiffrement.

Une description détaillée de l'algorithme est donnée par le NIST dans [23], avec notamment différents détails de calcul de chaque transformation.

1.3 Attaques physiques

Bien que les algorithmes de chiffrement standards soient considérés comme étant sûrs en l'état des connaissances mathématiques actuelles et des puissances de calcul disponibles, leurs implémentations matérielles peuvent néanmoins être sujettes à des faiblesses. Ces faiblesses peuvent être exploitées par des attaques sur l'implantation physique de l'algorithme. On peut classer les attaques possibles en deux catégories :

- **Les attaques passives ou par observation :** lors du fonctionnement d'un circuit manipulant des données secrètes, différentes grandeurs physiques observables du circuit (consommation électrique [24], émissions électro-magnétiques [25] temps de calcul [26], etc.) peuvent contenir des informations corrélées aux données manipulées. Il est ainsi possible de remonter aux données sensibles via l'observation de ces émissions.

- **Les attaques actives ou par injection de fautes :** lors du fonctionnement d'un circuit, il est possible de le perturber, induisant ainsi des erreurs dans les calculs. Celles-ci peuvent ensuite être exploitées pour remonter aux données protégées. Ces attaques seront détaillées dans la section 1.3.1.

Chaque catégorie peut également être divisée en trois sous-catégories :

- **Les attaques non-invasives** ne requièrent aucune altération du circuit ou du boitier le contenant.
- **Les attaques semi-invasives** requièrent une étape de préparation sur le boîtier contenant le circuit (souvent une ouverture mécanique ou chimique). Un exemple de circuit ouvert est montré figure 1.7.
- **Les attaques invasives** requièrent d'accéder directement au composant afin d'y apporter des modifications pour en altérer le fonctionnement ou brancher une sonde afin de lire les données directement sur le circuit.

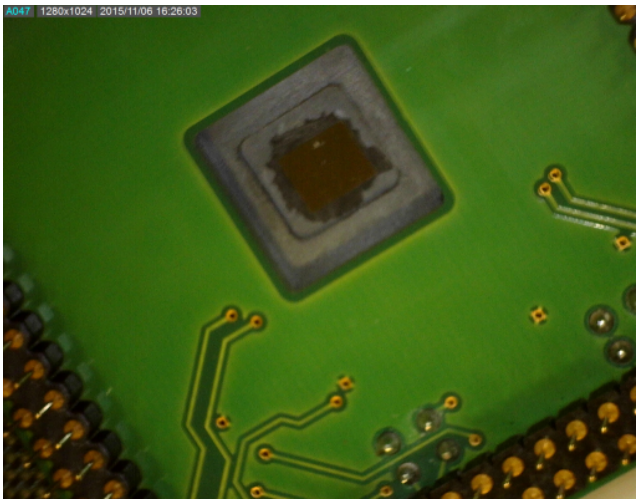


Figure 1.7 – Un microcontrôleur ouvert (via la face arrière).

L'injection de faute par laser est une méthode d'attaque par injection de fautes semi-invasive, celle-ci nécéssitant l'ouverture du boîtier des circuits cibles. La suite de ce chapitre se concentrera donc sur les attaques par injection de fautes.

1.3.1 Attaques par injection de fautes

Plusieurs méthodes de perturbation ont été introduites dans la littérature pour injecter des fautes dans le fonctionnement d'un circuit.

- **Perturbation de l'alimentation** [27] : En envoyant un pic de tension ou de courant sur l'alimentation d'un circuit, il est possible de créer des erreurs de

fonctionnement dans celui-ci. Cette méthode d'injection de fautes ne permet pas de choisir l'emplacement de l'attaque (tout le circuit est perturbé). Il est toutefois possible de modifier la durée l'amplitude et l'instant d'injection pour ajuster l'effet obtenu.

- **Perturbation de l'horloge** [28] : En altérant la fréquence d'horloge du circuit, il est possible de créer des erreurs par violation de temps de chemin critique. Comme pour les attaques par perturbation de l'alimentation, ces attaques impactent l'intégralité du circuit.

- **Impulsion électromagnétiques** [29] : En envoyant une forte impulsion électromagnétique via une antenne placée généralement au dessus du circuit, il est possible de perturber le circuit localement, contrairement aux méthodes précédentes. Les antennes utilisées sont des bobines reliées à un générateur d'impulsion. La figure 1.8 montre des exemples de sondes utilisées.

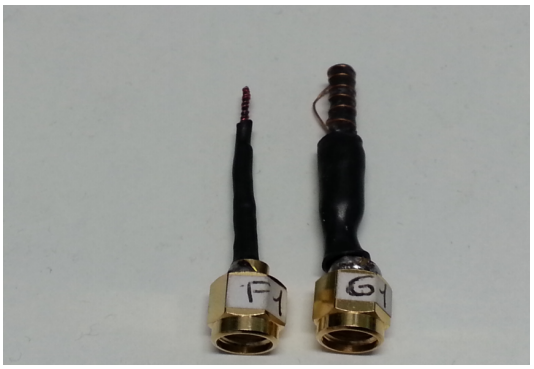


FIGURE 1.8 – Sondes utilisées pour l'injection électromagnétique. [1]

- **Impulsion laser** [30] : En illuminant un circuit à l'aide d'un faisceau laser, il est possbible d'y injecter des fautes via l'effet photoélectrique. Cette méthode est semi-invasive et généralement plus coûteuse que les précédentes, mais permet d'obtenir une excellente précision spatiale et temporelle. Elle est présentée plus en détail dans la suite de ce chapitre.

- **Perturbation de la tension de polarisation du substrat** [31,32] : Egalement semi-invasive, cette méthode assez récente, utilise une pointe posée sur la face arrière du substrat du composant pour y envoyer un pic de tension.

1.3.2 Aspect mathématique des attaques par injection de fautes

Les attaques par injection de fautes ont été introduites par Boneh et al. dans [33], où une attaque sur le RSA est présentée. Celle-ci permet de retrouver la clé de chiffrement privée en utilisant une faille dans les implémentations utilisant le théorème des restes chinois. L'attaque ne requiert qu'une seule faute pour extraire la clé : il est nécessaire de connaître un message M, son chiffré C et un chiffré fauté C' du même message pour la mener à terme.

La première attaque sur l'AES a été présentée en 2003 par Piret et al. dans [34]. Depuis, de multiples attaques ont été présentées et améliorées telles que celles présentées dans [35–39].

Un facteur déterminant dans le choix de l'attaque est le modèle de faute requis par l'attaque et celui que l'attaquant est capable d'obtenir sur le circuit cible.

1.3.2.1 Exemple de l'attaque de Piret et al.

L'attaque sur l'AES de [34] requiert l'obtention de couples chiffré correct/chiffré fauté correspondant à des fautes mono-octet étant survenues juste avant le MixColumn de la ronde 9. A partir de deux couples issus de fautes injectées sur la même colonne du *state*, il est possible de retrouver 4 octets de k_{10} (la sous-clé de la dernière ronde). Ainsi, avec 8 chiffrés fautés (deux par colonne), il est ainsi possible de retrouver l'intégralité de k_{10}. Une fois k_{10} connue, il est possible de retrouver la clé de chiffrement initiale.

On note que comme montré sur la figure 1.9, lorsqu'une opération ShiftRows, SubBytes ou AddRoundKey est appliquée à un *state* fauté, le nombre d'octets fautés ne varie pas. En revanche l'opération MixColumns a pour effet de propager la faute sur les 4 octets de la colonne de l'octet fauté.

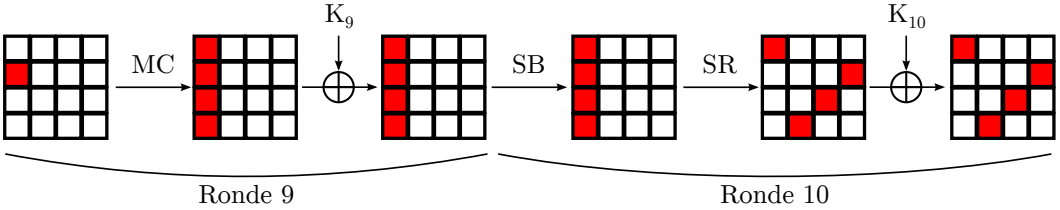


Figure 1.9 – Propagation des fautes dans l'AES

L'attaque se divise en deux parties. La première étape est de construire une liste des différences possibles entre le *state* correct et le *state* fauté à la sortie de l'opération MixColumns de la ronde 9, notée D. La position de l'octet fauté sur sa colonne étant inconnue, il y a donc $255 \times 4 = 1020$ éléments de 4 octets dans cette liste.

On construit ensuite la liste des possibilités pour les quatre octets de k_{10}, ce qui donne une liste L de 2^{32} éléments.

A partir d'une paire d'un chiffré correct et un chiffré fauté correspondant à un même message, il est alors possible de calculer la différence entre le *state* correct et le *state* fauté à la sortie du MixColumns de la ronde 9 pour chaque hypothèse de 4 octets de k_{10} contenue dans L.

Une nouvelle liste $\mathcal{L}$ contenant les valeurs canditates de 4 octets pour lesquels la différence est contenue dans D, peut alors être construite. Celle-ci contient environ 2^{10} éléments. En répétant ensuite la même opération en utilisant une seconde paire chiffré correct/chiffré fauté correspondant à une faute sur la même colonne, il y a alors 98% de chances de retrouver les 4 octets de k_{10} correspondants.

En répétant ces étapes avec des chiffrés fautés provenant de fautes sur les autres colonnes, il est alors possible de retrouver l'intégralité de k_{10}, à partir de laquelle la clé de chiffrement initale K peut être calculée en utilisant l'opération KeyScheduler inverse.

Une variante de l'attaque présentée dans le même article permet également de retrouver l'intégralité de la clé à partir d'un unique couple chiffré fauté/chiffré correct à l'aide de fautes survenues avant le MixColumns de la ronde 8, mais nécéssite un temps de calcul plus important.

1.4 Injection de fautes laser

L'injection de fautes laser utilise un faisceau laser pour éclairer un circuit, créant ainsi un courant via l'effet photoélectrique, permettant de modifier l'état du circuit. Avant d'être introduite dans le domaine de la sécurité en 2001 par Skorobogatov et Anderson [30], l'injection de fautes par laser était déja utilisée en fiabilité depuis 1965 pour simuler les effets des radiations sur les circuits intégrés [40]. Dans

ce domaine, l'injection laser permet d'obtenir une excellente précision de test spatialement et temporellement pour un coût réduit par rapport aux autres méthodes utilisées (faisceaux d'ions) pour simuler ces phénomènes [41]. En sécurité, l'intérêt de l'injection laser est également sa précision spatiale et temporelle, mais le coût est généralement plus élevé que pour des méthodes moins précises comme l'injection électromagnétique. En effet, l'injection de fautes laser nécessite l'utilisation d'équipements complexes et coûteux qui ne sont généralement pas présents dans les laboratoires de caractérisation électrique. Ce travail de thèse se concentrera sur l'analyse des effets des attaques par injection de fautes laser.

1.4.1 L'effet photoélectrique

L'injection de fautes laser se base sur l'effet photoélectrique pour créer localement des courants transitoires dans des circuits et ainsi perturber leur fonctionnement. Lorsqu'un composant est éclairé par un faisceau laser, des paires électron-trou se forment par effet photoélectrique si l'énergie des photons est supérieure à celle de la bande interdite du silicium. La plupart du temps les paires électron-trou ainsi créées se recombinent sans avoir d'impact sur le circuit.

En revanche, lorsque le faisceau laser est ciblé à proximité d'une jonction PN polarisée en inverse, les paires électron-trou sont déplacées dans des directions opposées par le champ électrique existant dans la zone de charge d'espace de la jonction, créant ainsi un courant, comme illustré sur la figure 1.10.

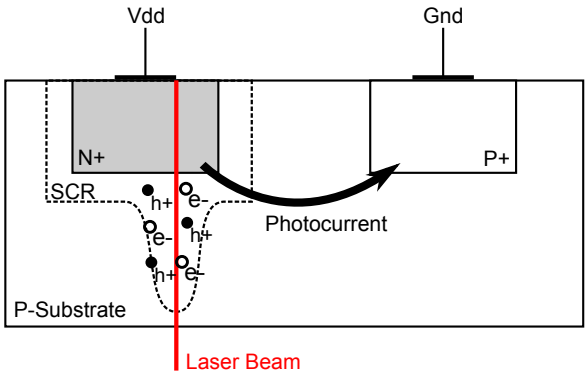
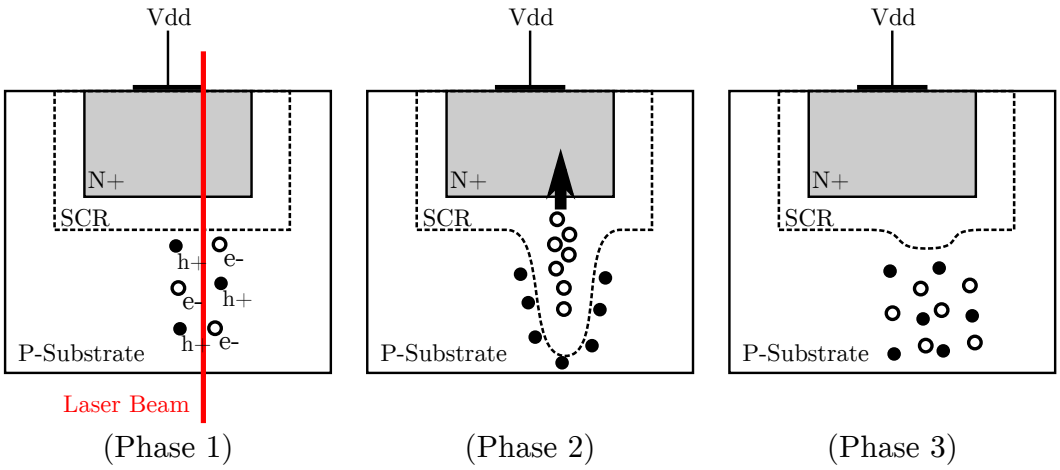


Figure 1.10 – L'effet photoélectrique dans une jonction PN polarisée en inverse.

Le phénomène se déroule en trois phases comme montré sur la figure 1.11 :



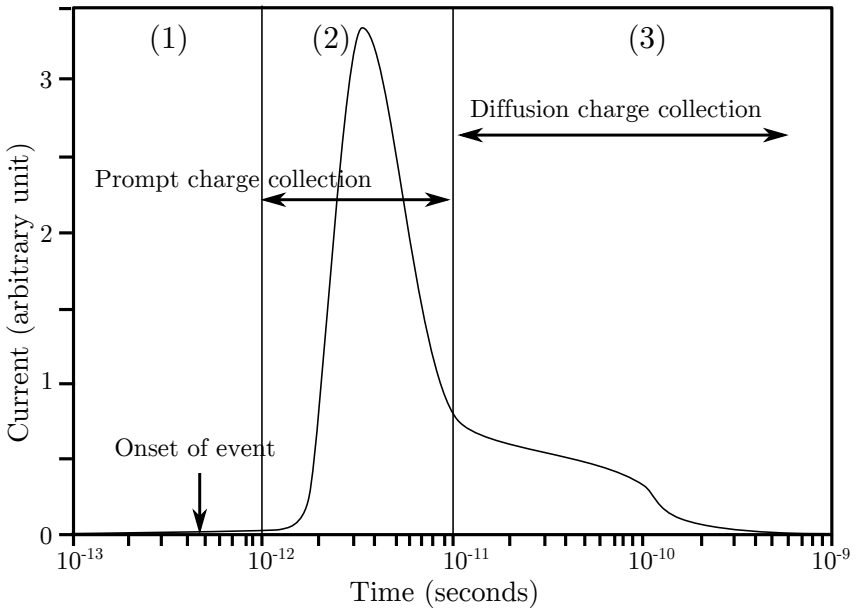


FIGURE 1.11 – Les trois étapes de l'effet photoélectrique [2]

- **Phase 1 :** Création des porteurs le long du chemin du faisceau laser.
- **Phase 2 :** Les porteurs de charge sont entraînés par le champ électrique de la jonction. Cela a pour effet d'étirer la zone de charge d'espace le long du faisceau. Les charges proches sont alors collectées, ce qui entraîne un pic de courant.
- **Phase 3 :** Lorsque le faisceau laser est coupé, les charges restantes sont ensuite collectées lentement durant une phase de diffusion, l'intensité du courant

diminue graduellement jusqu'à ce que toutes les charges aient été collectées ou recombinées.

Le profil temporel du photocourant est donc constitué d'une montée rapide liée à la phase de collection des charges, suivie d'une retombée plus lente après la disparition du faisceau liée au phénomène de diffusion.

1.4.2 De l'effet photoélectrique à la faute logique

La création d'un photocourant n'entraîne pas nécessairement un changement d'état du circuit et la création d'une faute. L'injection ou non d'une faute dépend de la position ciblée sur le circuit et de son état. La section suivante détaille comment un pic de tension peut être créé dans la logique d'un circuit en prenant pour exemple la porte la plus simple de la logique CMOS, l'inverseur.

1.4.2.1 Injection d'une faute dans la logique combinatoire

Un inverseur est constitué de deux transistors, un NMOS et un PMOS, celui-ci possède donc quatre jonctions PN correspondant aux drains et sources de ces derniers. Dans le cas représenté sur la figure 1.12, où l'entrée de l'inverseur est dans un état "0" le transitor PMOS est passant et le NMOS est bloqué. En conséquence, l'état de sortie de l'inverseur est à "1". Dans le cas où une impulsion laser cible le drain du transistor NMOS, un photocourant est créé allant du drain vers le substrat. Cela a pour effet de décharger la capacité de charge, abaissant de manière transitoire le niveau logique en sortie de l'inverseur. Une fois le tir laser terminé, le photocourant disparaît et la capacité de charge se recharge, ramenant la sortie de l'inverseur à sa valeur normale. Ce transitoire de tension est appelée Single Event Transient (SET) [42]. S'il est suffisamment important, il permet de passer sous le seuil logique et d'inverser la sortie de l'inverseur. La faute peut ensuite se propager dans la logique en aval, et éventuellement entraîner une faute dans le circuit.

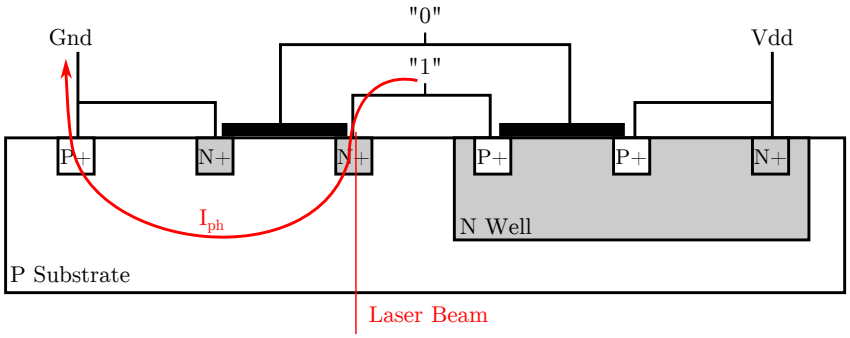


FIGURE 1.12 – Tir sur le drain du NMOS OFF d'un inverseur.

Un tir laser ciblé sur les autres jonctions PN de l'inverseur dans cet état de fonctionnement n'entrainerait en revanche pas de SET. En effet, les courants transitoires créés dans les drains et sources du PMOS fuiraient vers le NWell qui est polarisé à Vdd, ce qui ne déchargerait donc pas la capacité de charge. Enfin, comme la différence de potentiel est quasi nulle entre la source du NMOS et le substrat, une impulsion laser ciblée sur cette dernière ne créerait pas de photocourant significatif.

En suivant un raisonnement similaire détaillé sur la figure 1.13, on peut montrer que lorsque l'entrée de l'inverseur est à "1", seul un tir laser ciblé sur le drain du transistor bloqué (dans ce cas le PMOS) aurait un effet sur la sortie de l'inverseur : en créant un courant transitoire allant du NWell vers la sortie, la capacité de sortie se charge, créant ainsi un pic de tension positif. Une fois le phénomène terminé, la capacité se décharge et la sortie retrouve son état bas.

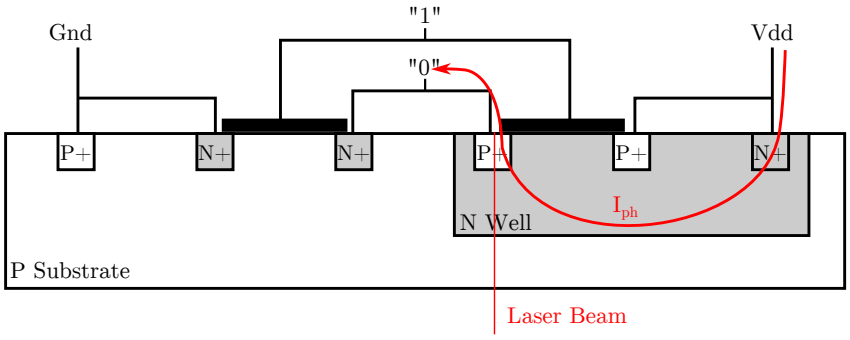


FIGURE 1.13 – Tir sur le drain du PMOS OFF d'un inverseur.

Dans les deux cas, la jonction PN pouvant avoir un effet sur la sortie du transistor est le drain du transistor bloqué comme résumé sur la figure 1.14.

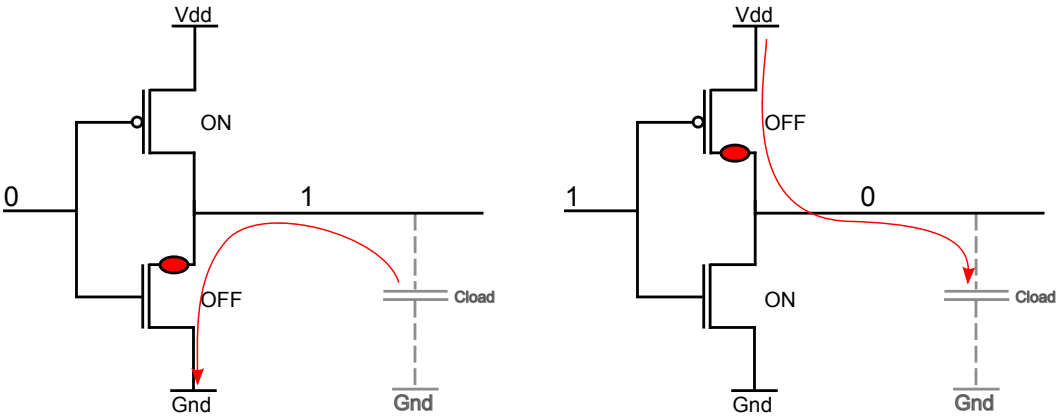


Figure 1.14 – Injection laser dans un inverseur.

Ce transitoire de tension peut ensuite se propager en aval à travers la logique combinatoire du circuit, ce qui peut par suite entraîner une faute. Pour qu'une faute résulte du transitoire de tension, il faut que celui-ci parvienne à l'entrée d'un élément mémoire au moment où celui-ci échantillonne la donnée. Ce phénomène est détaillé sur la figure 1.15. Dans le premier cas, le transitoire de tension arrive à l'entrée de la bascule D après que celle-ci a échantillonné la donnée, le transitoire de tension est alors sans effet. En revanche dans le deuxième cas, la valeur de D faussée par le transitoire de tension est échantillonnée par la bascule D, une faute est alors injectée dans le fonctionnement du circuit.

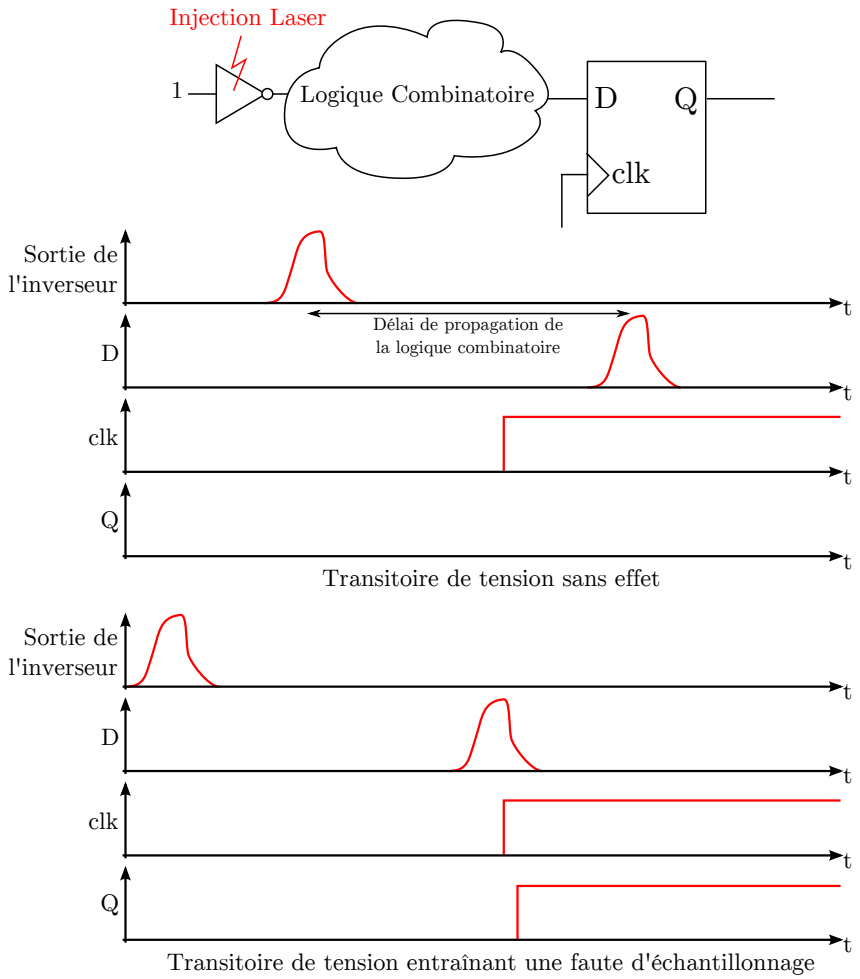


FIGURE 1.15 – Propagation d'un transitoire de tension dans la logique d'un circuit.

1.4.2.2 Injection d'une faute dans un point mémoire

Outre la possibilité d'injecter une faute dans la logique combinatoire d'un circuit comme détaillé dans la section précédente, il est également possible d'induire une faute en ciblant directement un point mémoire. On parle alors de de SEU ou d'aléa logique [2]. Le fonctionnement du passage d'un SET à un SEU peut être montré sur une cellule SRAM.

Le phénomène d'injection d'une faute dans une SRAM est détaillé dans la figure 1.16. Un point mémoire est composé de deux inverseurs rebouclés l'un sur l'autre (a). Ils sont ainsi dans un état stable, un des deux noeuds représente la donnée stockée, et l'autre son complément. Lorsqu'un des deux inverseurs est attaqué comme vu dans

la section 1.4.2.1, sa sortie est inversée pendant la durée du transitoire de tension créé (b). Ce changement d'état peut alors se propager via le second inverseur (c). Une fois la perturbation terminée, la cellule étant dans un état stable, celle-ci ne retourne pas à son état initial (d). Une faute permanente a donc été créée.

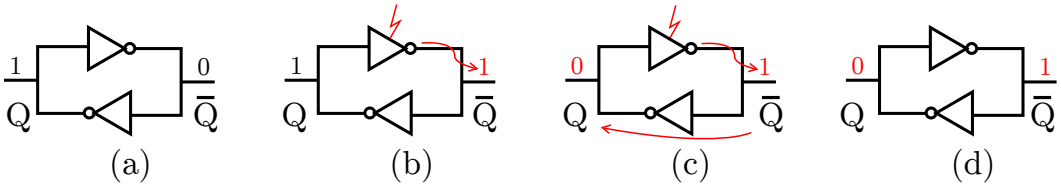


Figure 1.16 – Création d'un *Single Event Upset* sur une cellule SRAM.

1.4.3 Paramètres de l'injection de fautes

De multiples paramètres influent sur le résultat de l'injection laser dont la longueur d'onde, la focalisation du faisceau ou la durée d'impulsion. Leur connaissance et leur maîtrise sont importantes pour obtenir des résultats contrôlés et répétables, ce qui est crucial du point de vue d'un attaquant. En effet, comme vu dans la section 1.3.2, le choix du modèle de faute est primordial pour la réussite d'une attaque. Contrôler au mieux les paramètres de l'attaque et donc la faute injectée, permet de choisir un modèle de faute théorique adapté et d'obtenir des fautes s'en approchant le plus possible. La connaissance de ce modèle de faute mène ensuite au choix de l'attaque mathématique utilisée.

1.4.3.1 La longueur d'onde

La longueur d'onde utilisée influe sur l'énergie des photons du laser et l'absorption du faisceau par le silicium.

La relation entre la longueur d'onde et l'énergie des photons du laser est donnée par l'équation (1.1), où h est la constante de Planck, c la vitesse de la lumière et λ la longueur d'onde du laser [41].

$$E_{ph} = \frac{hc}{\lambda} \tag{1.1}$$

L'énergie des photons doit être supérieure à celle de la bande interdite du silicium pour déclencher l'effet photoélectrique. Ceci impose alors la condition donnée par

l'équation (1.2), où E_{bg} est l'énergie de la bande interdite du silicium (soit 1.1 eV à 300k).

$$\lambda < \frac{hc}{E_{bg}} \tag{1.2}$$

Ce qui donne une longueur d'onde maximale de 1120 nm, on peut donc utiliser des longueurs d'onde allant de l'ultraviolet jusqu'à l'infrarouge.

Le coefficient d'absorption a un impact sur le choix de la longueur d'onde. La figure 1.17 donne le coefficient d'absorption du silicium pour différents niveaux de dopage en fonction de l'énergie des photons. Deux lignes indiquent les énergies correspondant à des longueurs d'onde de 800nm et 1060nm. On constate qu'une longueur d'onde d'environ 1060nm permet d'obtenir un coefficient d'absorption minimal indépendamment du niveau de dopage.

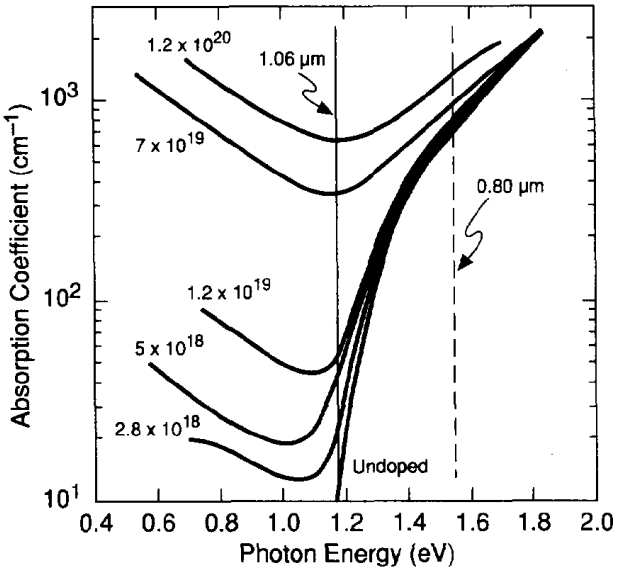


FIGURE 1.17 – Coefficient d'absorbtion du silicium en fonction de l'énergie des photons et du niveau de dopage. [3]

1.4.3.2 Injection via la face avant ou arrière et épaisseur du substrat

L'injection laser peut se faire via la face avant ou la face arrière du circuit, comme illustré sur la figure 1.18.

Bien que l'injection via la face avant permette d'accéder directement à la région active du silicium, celle-ci n'est généralement plus utilisée en raison des nombreuses couches de métallisation. En effet, ces dernières reflètent et bloquent le faisceau

laser, rendant la plupart des zones sensibles non-atteignables. Un autre obstacle majeur à l'injection via la face avant est la présence de boucliers destinés à protéger le circuit contre ces attaques. L'utilisation de boucliers sera détaillée dans la section 1.5 traitant des contre-mesures.

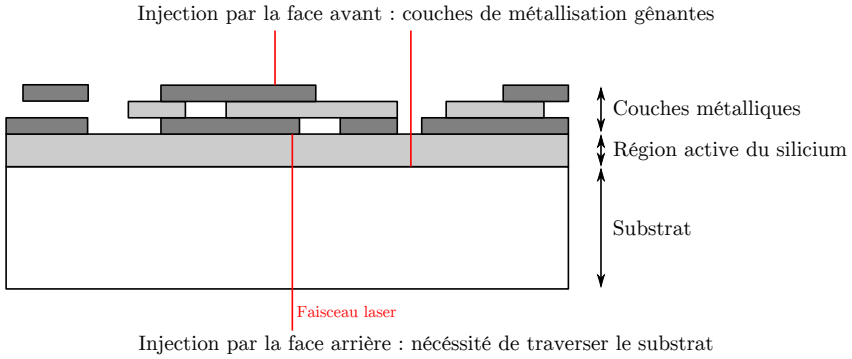


Figure 1.18 – Injection laser via la face avant ou la face arrière.

Pour ces raisons, l'injection via la face arrière est généralement la plus utilisée. Afin de faciliter la traversée du substrat par le faisceau, les circuits attaqués seront généralement amincis. La figure 1.19 [4], montre le taux de générations de porteurs en fonction de la longueur d'onde pour plusieurs épaisseurs de substrat. Pour permettre une attaque laser, les circuits sont généralement affinés à une épaisseur de substrat entre 100 µm et 200 µm (dans le cas des circuits utilisé dans ces travaux, 140 µm à partir d'une épaisseur initiale d'environ 750 µm). Pour ces épaisseurs, la longueur d'onde optimale est située autour de 1000 nm (Infrarouge), tandis que pour une injection via la face avant ne nécessitant pas de traverser le substrat, la longueur d'onde optimale est d'environ 800 nm (vert).

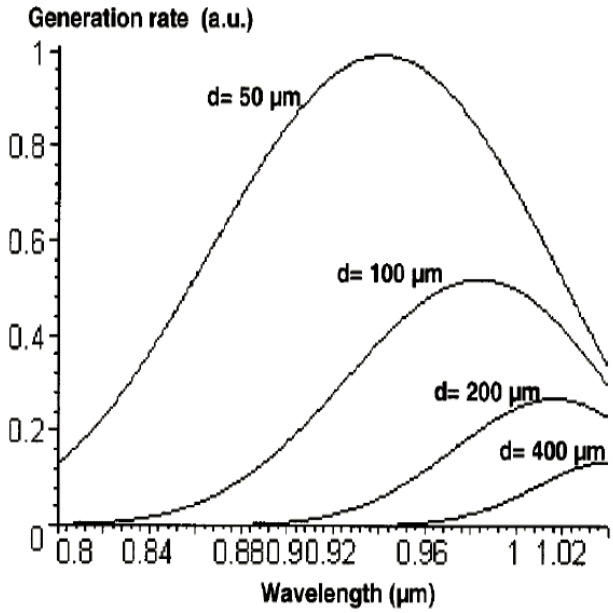


FIGURE 1.19 – Taux de génération de porteurs en fonction de la longueur d'onde pour plusieurs épaisseurs de substrat. [4]

1.4.3.3 La focalisation du faisceau laser

Afin d'avoir une précision spatiale et une densité d'énergie maximale au niveau de la zone ciblée, la focalisation du faisceau est un facteur important.

Lors d'une injection via la face arrière, en raison de l'indice de réfraction élevé du silicium, le faisceau diverge fortement à l'entrée dans le substrat comme montré sur la figure 1.20 [5]. Il faut donc prendre cela en compte lors du réglage de la hauteur d'injection. Sur la figure, z_0 correspond à la hauteur lorsque l'on effectue la mise au point du spot sur la face arrière du circuit. La hauteur pour obtenir le meilleur focus au niveau de la région active du silicium est donnée par l'équation 1.3, où e est l'épaisseur du substrat et n l'indice de réfraction du silicium.

$$z_1 = z_0 + \frac{e}{n} \tag{1.3}$$

Notons que les lois de l'optique fixent le diamètre minimal du spot à la longueur d'onde utilisée, soit environ 1 µm.

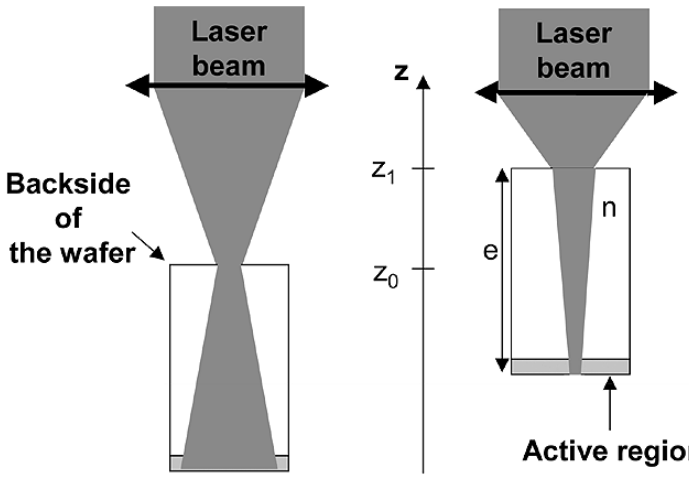


Figure 1.20 – Influence de la hauteur de l'objectif sur la focalisation. [5]

1.4.3.4 Durée d'impulsion

Dans [6] Douin et al. montrent que la durée de l'impulsion a un effet sur la quantité d'énergie nécessaire pour créer un SEU . La figure 1.21 montre les seuils d'énergie et de puissance requises pour injecter une faute dans la cellule SRAM testée en fonction de la durée d'impulsion. Deux domaines peuvent être distingués : les impulsions plus courtes que le temps de basculement de la cellule (noté I sur la figure) et les impulsions plus longues (noté II sur la figure).

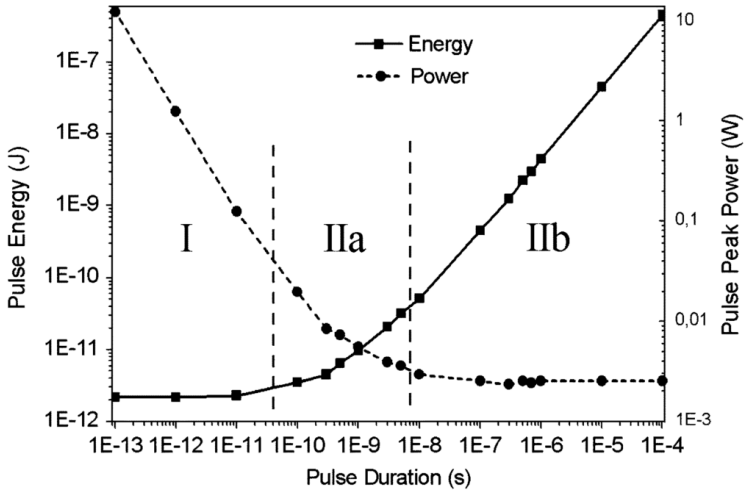


Figure 1.21 – Seuil d'injection d'un SEU dans une cellule SRAM en fonction de la durée d'impulsion. [6]

Dans le cas des impulsions courtes, la totalité du phénomène photoélectrique

a lieu avant que la cellule réagisse, ce qui se traduit par un seuil d'énergie fixe indépendant de la durée d'impulsion. Pour les impulsions plus longues en revanche, on observe un seuil fixe en puissance.

1.4.3.5 Synchronisation du tir

Lorsque la cible visée est un circuit en fonctionnement (par opposition à des tests statiques sur des éléments mémoire), il est également nécessaire de contrôler l'instant du tir laser.

Comme vu dans la section 1.4.2.1 et illustré par la figure 1.15, lors de l'utilisation d'impulsions courtes, la durée du transitoire de tension est plus courte que la période d'horloge de la cible. L'instant d'injection doit alors être contrôlé de manière à ce que le transitoire de tension soit échantillonné pour qu'une faute soit injectée.

A un niveau plus macroscopique, contrôler l'instant d'injection est nécessaire pour impacter l'instant du calcul désiré. Par exemple, comme vu dans la section 1.3.2, certaines attaques requièrent de fauter une ronde précise de l'AES.

Le contrôle de l'instant d'injection est problématique lorsque certains délais comme le temps de réaction de la source laser présentent des gigues importantes (plus d'une période d'horloge du circuit cible). Ainsi, la synchronisation du tir laser est une étape déterminante pour la réussite de l'attaque et peut devenir complexe pour des circuits fonctionnant à des fréquences élevées.

1.4.3.6 Synthèse

Cette section a présenté les différents paramètres physiques d'une attaque laser et leurs effets. En pratique, certains de ces paramètres sont fixés par le banc de test (par exemple, la longueur d'onde) ou le circuit ciblé (injection via la face avant ou arrière, épaisseur du substrat). D'autres paramètres ont une valeur optimale, comme la focalisation du faisceau, pour laquelle on cherche généralement à obtenir un spot le plus petit possible afin d'avoir une densité d'énergie et une précision spatiale optimale. Enfin, les paramètres restants, comme la durée d'impulsion, la puissance du faisceau laser, ou la position ciblée sur le circuit sont ceux sur lesquels la marge de manoeuvre est la plus importante.

Les valeurs de ces paramètre utilisées en pratique seront abordées plus en détail dans le chapitre 2, lors de la présentation du banc de test à l'aide duquel ces travaux out été menés.

1.5 Contre-mesures face aux attaques par injection de fautes

Afin de protéger les circuits face aux attaques décrites précédemment, de nombreuses contre-mesures existent. Celles-ci ne sont pas toutes efficaces contre les mêmes attaques, et peuvent également avoir une efficacité limitée spatialement ou temporellement. Ainsi, différentes contre-mesures sont généralement utilisées en parallèle afin de couvrir les différentes vulnérabilités du circuit à protéger.

Plusieurs approches peuvent être utilisées. Dans un premier temps, il est possible d'empêcher l'attaque en bloquant le mécanisme d'injection. Sinon, des circuits détectant le phénomène d'injection ou la faute elle même pourront être utilisés pour produire une alarme et prendre les mesures nécessaires à la protection du circuit.

1.5.1 Prévention de l'injection de fautes

La première approche pour protéger un circuit des attaques en fautes est d'empêcher l'injection. La méthode employée dépend du type d'attaque que l'on cherche à bloquer.

Une méthode couramment utilisée est l'ajout d'un bouclier sur le dessus du circuit. Cette grille métallique agit comme une barrière, permettant :

- de bloquer l'injection de fautes par laser (et dans une moindre mesure par impulsion electro-magnétique),
- d'empêcher des attaques invasives de type micro-probing ou de modification du circuit,
- de rendre la rétro-conception du circuit plus complexe.

La figure 1.22 montre une photographie d'un bouclier sur la face avant d'un circuit.

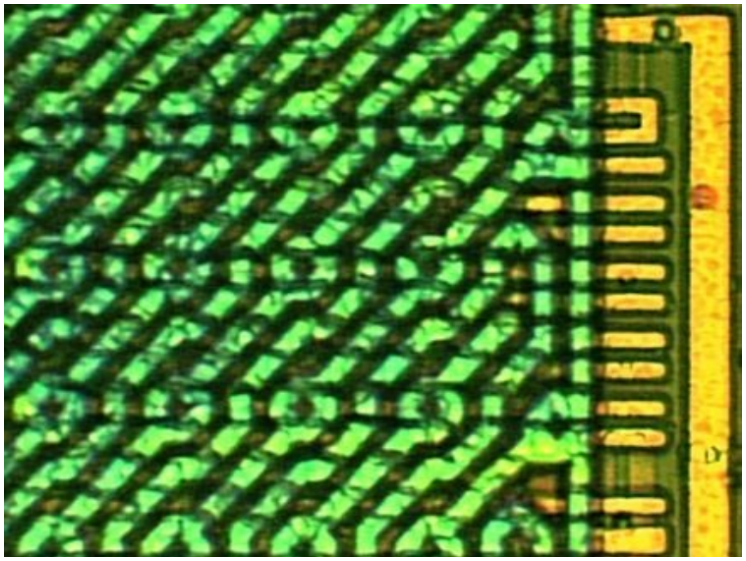


FIGURE 1.22 – Exemple de shield sur la face avant d'un circuit. [7]

Un bouclier peut être passif ou actif. Dans le premier cas, celui ci est une simple couche de métallisation déconnectée du reste du circuit. Dans le cas d'un shield actif, la grille métallique est connectée au circuit, il est alors possible de détecter une atteinte à l'intégrité de celle-ci pour déclencher une alarme [7].

Une seconde approche est l'utilisation de méthodes de conception spéciales pour rendre le circuit moins sensible à certaines perturbations. Par exemple, les cellules SRAM présentées dans [43] utilisent un layout modifié présentant une surface sensible à l'injection laser réduite, les rendant ainsi plus difficiles à attaquer.

1.5.2 Détection du phénomène physique

Une autre approche consiste à détecter l'injection de fautes à l'aide de capteurs, pour permettre ensuite au circuit d'agir en conséquence (reset ou arrêt du circuit, sortie aléatoire, etc.).

Le type de capteur utilisé dépend du type d'injection de faute que l'on cherche à détecter. Ils mesurent les grandeurs physiques impactées par les méthodes d'injection de faute et déclenchent une alarme en cas de valeurs anormales.

- Il est possible de détecter des attaques par perturbation de la tension d'alimentation ou par impulsion électro-magnétiques à l'aide de détecteurs de va-

riations de tension d'alimentation [44].

- Des détecteurs de variations de la fréquence d'horloge l'horloge permettent de détecter les méthodes d'injection qui visent à créer des fautes par violation de contrainte temporelle. (Injection EM, glitchs d'horloge ou d'alimentation) [45]. La figure 1.23 montre le principe de fonctionnement d'un tel détecteur (a). Ces capteurs utilisent une horloge retardée qui est utilisée pour échantilloner l'horloge du circuit (b), en cas de perturbation, la valeur échantillonée change (c), ce qui déclenche une alarme. [1]

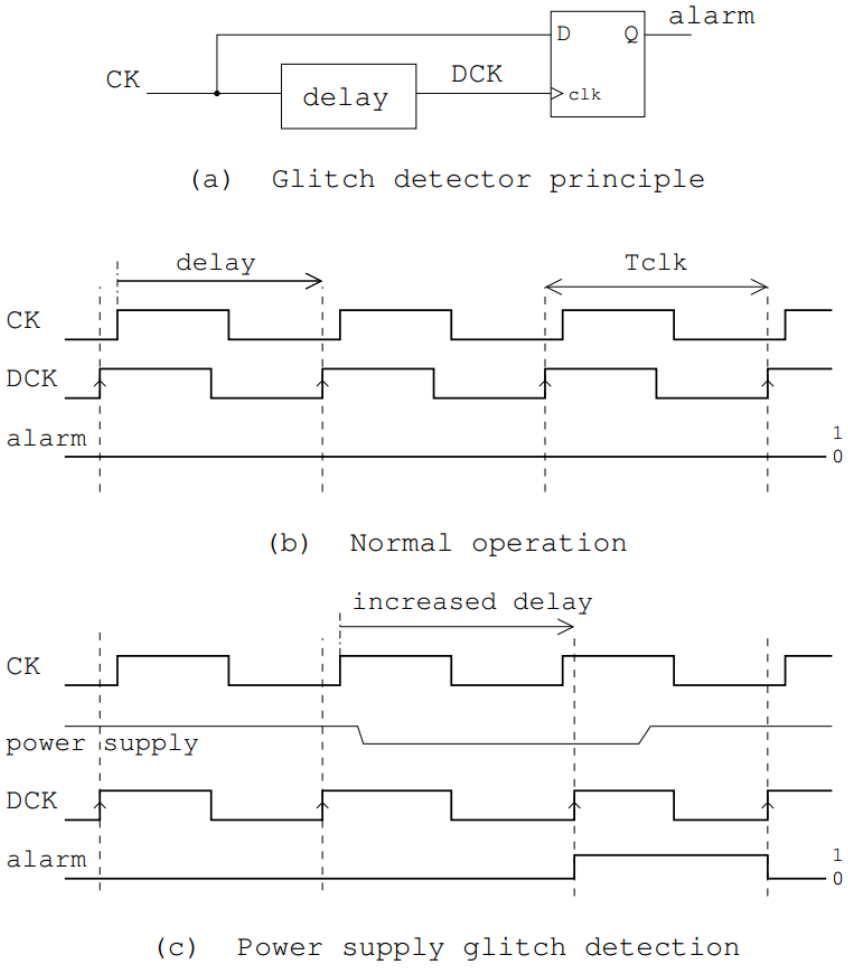


Figure 1.23 – Principe d'un détecteur de perturbation d'horloge [1].

- Pour détecter l'injection de faute laser, des capteurs de luminosité ou de température peuvent être utilisés. Un autre type de capteurs efficaces pour détecter l'injection de faute laser sont les détecteurs de type BBICS [8]. Ces

derniers permettent de détecter les courants transitoires induits par les tirs laser dans le substrat. La figure 1.24 décrit le fonctionnement d'un tel capteur. Les entrées INPWELL et INNWELL sont reliées respectivement aux prises de polarisation du subtrat P et N. Dans le cas d'une attaque sur un transistor NMOS, un courant passera dans le transistor MN1, ce qui rend le transistor MN2 passant, reliant ainsi le noeud OUTA à la masse, déclenchant l'alarme du capteur.

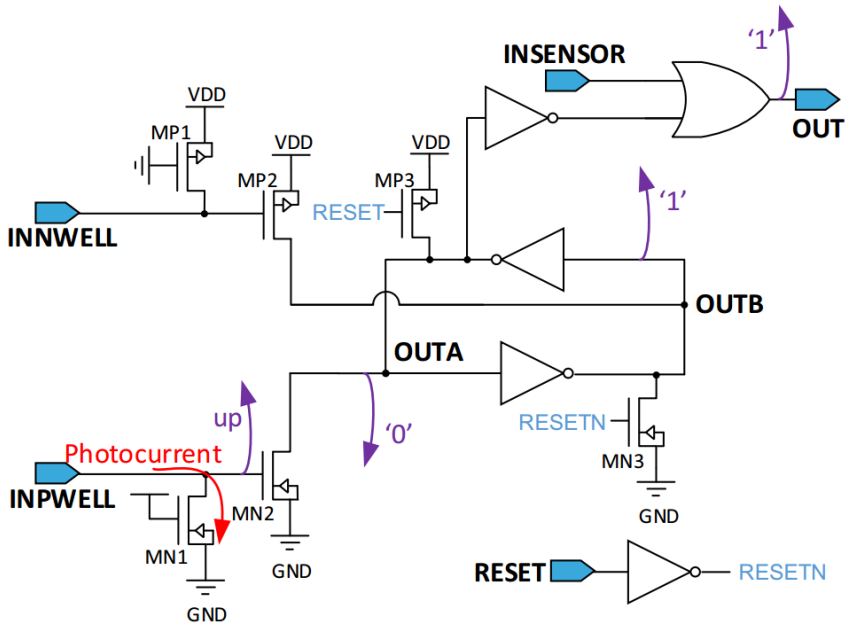


FIGURE 1.24 – Principe de fonctionnement d'un capteur de type BBICS [8].

La plupart des capteurs utilisés pour ces types de contre-mesures ont une couverture spatiale limitée et doivent en conséquences être placés à de multiples endroits sur le circuit afin de couvrir tous les éléments sensibles de celui-ci.

1.5.3 Détection de la faute logique

Enfin, une dernière approche consiste à détecter, atténuer, ou annuler l'impact logique des fautes en cas de succès de l'injection. Pour y parvenir, plusieurs méthodes sont possibles :

- Une première consiste à utiliser des systèmes redondants afin d'effectuer une même opération plusieurs fois. Cela peut se faire soit successivement sur une même partie de circuit dans le cas d'une redondance temporelle (figure 1.25), soit en parallèle via une multiplication d'une même fonction du circuit dans le cas d'une redondance matérielle (figure 1.26) [46]. Dans le cas où une fonction est tripliquée, un vote majoritaire peut aussi être utilisé pour corriger les fautes injectées [47]. La figure 1.27 montre le fonctionnement d'une redondance triple avec votre majoritaire : en cas de différence entre la sortie d'une des fonctions et celle des deux autres, les deux fonctions avec la même sortie ont le résultat correct, qui est alors envoyé en sortie. La redondance temporelle sera abordée de manière plus détaillée dans le chapitre 4. Cependant les méthodes de redondance sont généralement très coûteuses en termes de consommation, de temps d'exécution et de surface [48].

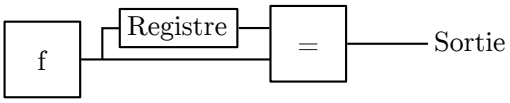


Figure 1.25 – Redondance temporelle

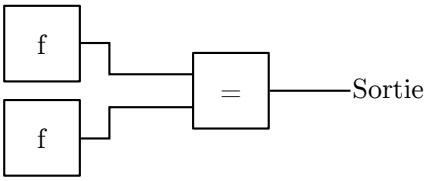


Figure 1.26 – Redondance matérielle

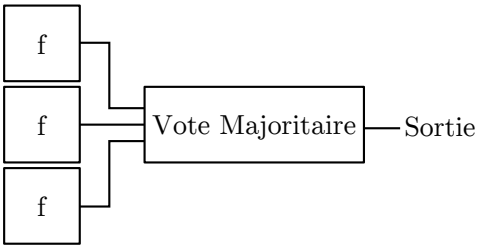


FIGURE 1.27 – Redondance triple

- L'utilisation de code de détection ou de correction d'erreur peut également permettre de parvenir à des résultats similaires avec un coût moindre. Par exemple, dans [49], les auteurs présentent un code de détection d'erreur pour l'AES basé sur la parité permettant de détecter les fautes avec entre 10% et 20% de sur-coût en surface.

1.6 Simulation d'injection laser

L'utilisation de contre-mesures a un coût important en surface et en temps de conception, il est donc nécéssaire de pouvoir prévoir les vulnérabilités d'un circuit pour le protéger au mieux. Cependant, les tests en laboratoires destinés à évaluer la robustesse d'un circuit face à l'injection laser sont très coûteux. Ils sont longs à effectuer, nécessitent un équipement complexe et surtout, s'effectuent post-production, ce qui rend la découverte d'une faille à ce stade extrêmement coûteuse. Ainsi, afin de diminuer ces coûts, l'utilisation de modèles de simulation pour évaluer la robustesse des circuits pendant l'étape de conception et avant l'étape de production est primordiale.

La simulation peut se faire à plusieurs niveaux d'abstraction : au niveau électrique via des simulateurs électriques type SPICE et au niveau physique via TCAD (pour Technology Computer Aided Design). SPICE permet de simuler des circuits plus complexes, tandis que la simulation TCAD permet d'obtenir des résultats plus précis, mais au prix d'une mise en œuvre plus lourde et d'un temps de calcul beaucoup plus important qui limitent son application à des zones plus réduites (souvent un seul transistor).

La simulation par TCAD nécéssite de modéliser le circuit à tester en suivant

les différentes étapes du processus de fabrication avant de simuler ensuite les effets physiques résultant de l'illumination du circuit. Dans [50], il est présenté une cellule SRAM modélisée en trois dimensions et simulée par TCAD. La figure 1.28 montre cette cellule en 3D.

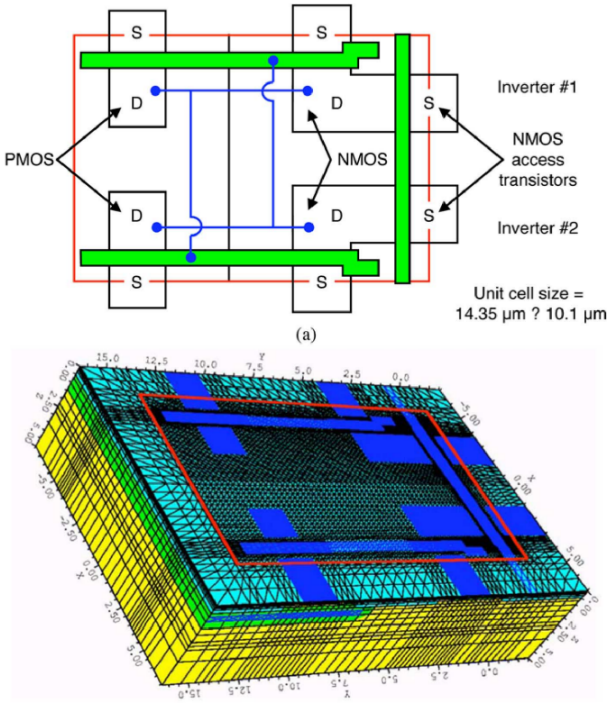


Figure 1.28 – Layout et modélisation TCAD 3D d'une cellule SRAM [9]

La simulation SPICE se fait au niveau électrique et nécessite donc une modélisation préalable de l'effet du laser sur le silicium. Pour cela plusieurs modèles ont été établis [10,51–53]. De manière générale, le photocourant est modélisé par une source de courant au niveau de la jonction PN ciblée, dont l'intensité et le profil dépendent des paramètres du laser (puissance et durée d'impulsion) comme sur la figure 1.29 : la source de courant I_{rad} simule un tir laser sur le drain du transistor.

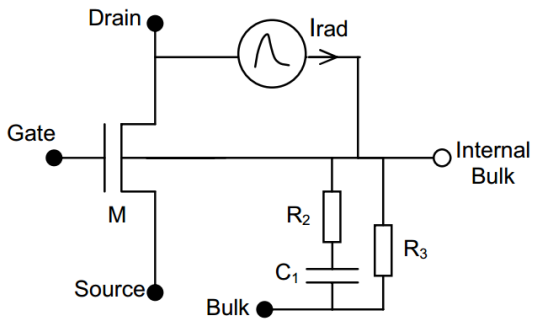


FIGURE 1.29 – Modèlisation d'une impulsion laser sur un transistor [10]

Dans la suite de ces travaux, seule la simulation au niveau électrique sera utisée.

1.6.1 Modèle de simulation électrique utilisé

Le modèle de simulation électrique utilisé dans la suite de ces travaux est celui élaboré par Sarafianos et al. [54–56]. Ce modèle utilise la simulation SPICE, cependant, afin d'améliorer la précision de simulation par rapport aux modèles précédents, celui-ci prend en compte la topographie des circuits et le profil d'intensité du faisceau laser. A partir du layout du circuit et des paramètres du tir laser, le modèle permet de générer une source de courant par jonction PN présente dans l'aire d'effet du laser. Les modèles antérieurs utilisaient uniquement une source pour la jonction ciblée directement. L'intensité du courant de chaque source dépend de la distance entre le centre du faisceau laser et la jonction associée.

Le photocourant de chaque source de courant correspondant à une jonction PN est donné par l'équation 1.4 ou $a(P)$ et $b(E)$ sont des paramètres fonctions de la puissance P du laser déterminés expérimentalement, V_r est la tension de polarisation inverse de la jonction PN cible, A est la surface de la jonction.

$$I_{ph}(t) = [a(P).V_r + b(P)].A.\alpha_{topology}.\Omega_{shape}(w).Z_{focus} \tag{1.4}$$

$\Omega_{shape}(w)$ représente l'influence de la durée de l'impulsion laser. Le profil de l'impulsion a une forme de double exponentielle dépendant de la durée de l'impulsion. Cette forme permet de modéliser le temps de montée rapide de l'impulsion laser, suivie du phénomène de diffusion [57]. Comme montré dans la figure 1.30,

pour les impulsions les plus courtes, le courant n'atteint pas le régime stationnaire. $\Omega_{shape}(w)$ est donné par l'équation 1.5, où w est la durée de l'impulsion laser.

$$\Omega_{shape}(w) = 1 - e^{-\frac{w}{250E-9}} \tag{1.5}$$

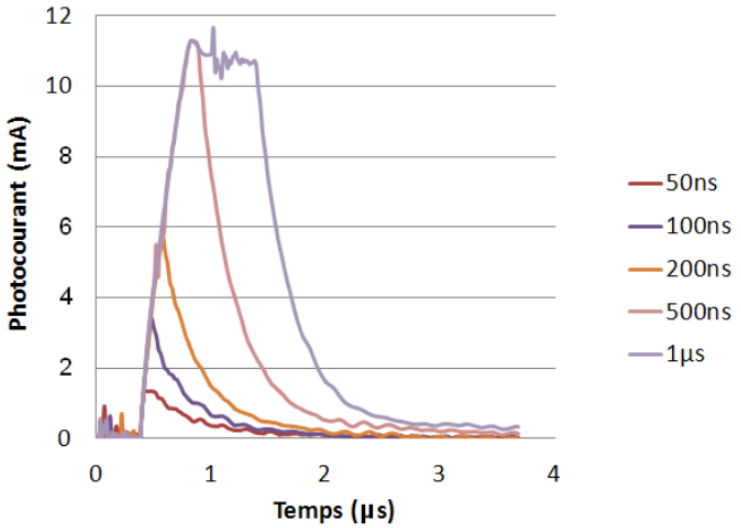


Figure 1.30 – Profil temporel du photocourant dans une jonction PN pour différentes durées d'impulsion. [11]

$\alpha_{topology}$ est un coefficient représentant le profil d'intensité spatiale du faisceau et dépend de la distance entre le centre du faisceau et le centre de la jonction PN. La figure 1.31 montre le coefficient $\alpha_{topology}$ pour une durée d'impulsion de 50 ns et une taille de spot de 1 µm.

De la même manière, le coefficient Z_{focus} représente la focalisation du faisceau et dépend de la hauteur d'injection (comme abordé dans la section 1.4.3.3). Ces paramètres sont mesurés ou déterminés de manière expérimentale.

A partir de l'équation 1.4, il est ainsi possible de configurer les sources de courant correspondant à chaque jonction PN d'un circuit à tester en fonction de la distance entre le centre du faisceau laser et celui de la jonction. On note que les valeurs des coefficients du modèle sont déterminées expérimentalement et varient en fonction de la technologie CMOS utilisée. La figure 1.32 montre la modélisation d'un tir laser sur un transistor NMOS, on note la présence de deux sources de courants : une pour le drain et une pour la source du transistor. Pour un transistor PMOS, une source

de courant est ajoutée à la jonction entre le NWell et le substrat (figure 1.33).

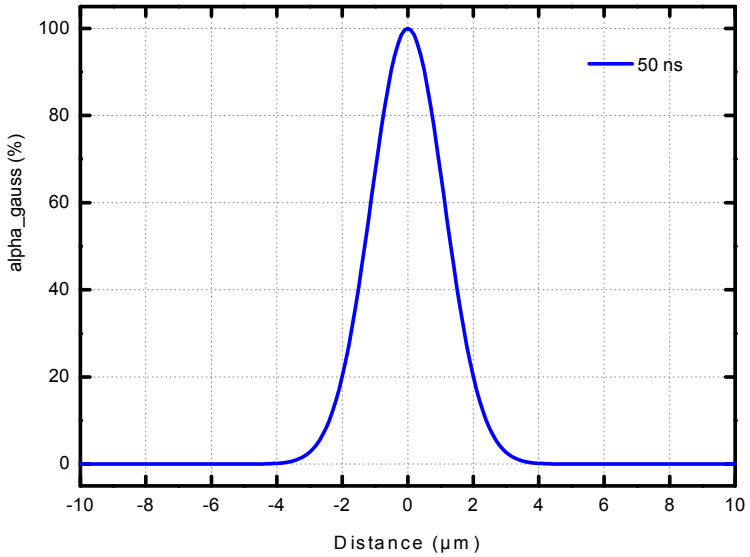


FIGURE 1.31 – Coefficient $\alpha_{topology}$ en fonction de la distance entre les centres de la jonction et du faisceau laser pour une impulsion de 50 ns (technologie CMOS 0.25um) [12]

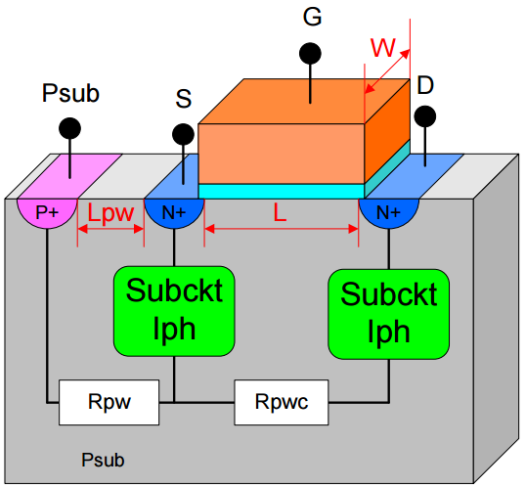


FIGURE 1.32 – Modélisation d'une impulsion laser sur un transistor NMOS [11]

A l'aide de ce modèle il est ainsi possible d'obtenir des simulations précises, tenant compte de la topologie des cellules, ce qui n'est habituellement pas le cas pour les simulations SPICE.

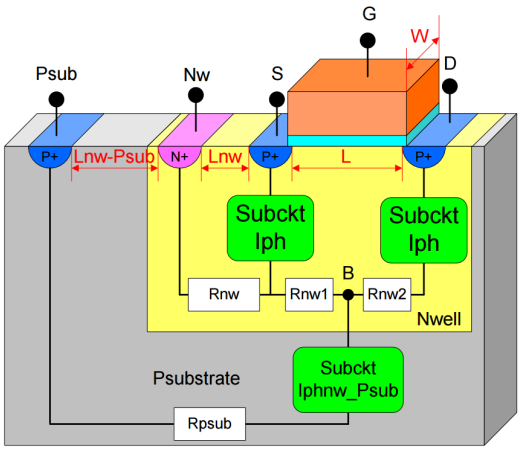


Figure 1.33 – Modélisation d'une impulsion laser sur un transistor PMOS [11]

1.7 Circuits basse consommation

Ce travail de thèse portant sur l'analyse des effets des attaques laser sur des circuits à basse consommation, cette section présente les bases des techniques de conception basse consommation, et plus particulièrement celles étudiées dans la suite de ce manuscrit.

Deux raisons majeures contribuent à la nécessité de concevoir des circuits consommant le moins d'énergie possible. Premièrement, en augmentant la taille et la densité des circuits, la puissance dissipée augmente, ce qui entraîne un phénomène d'échauffement. Cela fait de la consommation d'énergie le facteur limitant pour augmenter le nombre de transistors sur une puce. La seconde raison est la rapide augmentation du nombre d'appareils mobiles : la durée de vie des batteries est directement liée à la consommation du circuit. L'augmentation de la capacité des batteries étant coûteuse en poids et en volume, la réduction de la consommation s'impose pour obtenir une durée de vie suffisante pour les applications modernes.

Il est possible de diviser la consommation d'un circuit en deux parties : la consommation dynamique et la consommation statique. La première résulte de la commutation des portes du circuit tandis que la seconde provient des courants de fuite des transistors. Jusqu'à récemment, la consommation dynamique était la composante la plus importante, toutefois pour les technologies récentes à partir de 90nm, la consommation statique devient plus importante. La puissance dissipée par

le circuit s'exprime donc par l'équation (1.6) [14, 58].

$$P = P_{dynamic} + P_{static} \tag{1.6}$$

1.7.1 La consommation statique

Idéalement, la logique d'un circuit CMOS ne dissipe pas de puissance statique puisque lorsqu'il se trouve dans un état stable il n'existe normalement pas de chemin direct entre V_{dd} et la masse. Cependant en pratique les transistors MOS ne sont pas des interrupteurs parfaits, et des courants de fuite, notés I_{leak}, existent, entraînant une consommation statique. On peut donc exprimer la puissance dissipée statiquement P_{static} selon l'equation (1.7).

$$P_{static} = V_{dd}.I_{leak} \tag{1.7}$$

1.7.2 La consommation dynamique

La puissance dynamique se divise en deux parties. D'une part l'énergie nécessaire pour charger les capacités du circuit lors des commutations des portes et d'autre part les courants de court-circuit.

La figure 1.34 représente les courants entraînant la consommation dynamique dans un inverseur CMOS. I_{dyn} est le courant qui charge la capacité C lorsque la sortie passe de '0' à '1'. I_{sc} est le courant de court-circuit qui apparaît lorsque la porte commute : lors du changement d'état de l'entrée de l'inverseur, les deux transistors seront brièvement dans un état 'ON' en même temps, créant ainsi un chemin direct entre V_{dd} et la masse. I_{sc} est faible devant I_{dyn}. L'énergie nécessaire pour charger la capacité C valant $C.V_{dd}^2$, si on néglige I_{sc} on peut donc exprimer la puissance dynamique moyenne $P_{dynamic}$ comme dans l'équation 1.8.

$$P_{dynamic} = C.V_{dd}^2.\alpha.f \tag{1.8}$$

Dans cette équation, C est donc la capacité totale du circuit, f est sa fréquence de fonctionnement (plus la fréquence est élevée, plus les portes commutent souvent, augmentant ainsi la consommation dynamique), et α est un coefficient compris

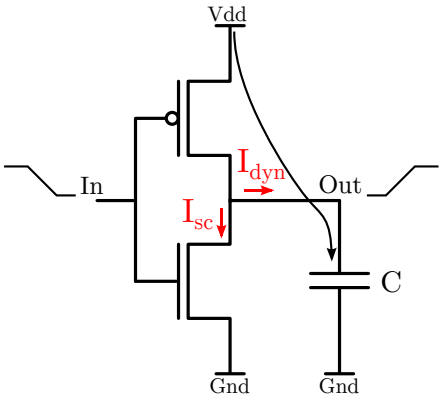


Figure 1.34 – Courants dans un inverseur entraînant la consommation dynamique

entre 0 et 1 qui représente l'activité du circuit (la proportion moyenne de portes qui commutent sur chaque front d'horloge).

Par la suite, nous allons voir les différentes méthodes permettant de réduire la consommation d'un circuit.

1.7.3 Méthodes de réduction de la consommation

De multiples méthodes de réduction de la consommation existent [13,14] à différents niveaux de conception allant de la fabrication des circuits à l'optimisation des algorithmes logiciels. La figure 1.35 liste de manière non-exhaustive des méthodes pour chaque niveau d'abstraction.

La capacité équivalente du circuit C ne peut être réduite qu'en utilisant des méthodes bas niveaux en prêtant une attention particulière à la conception du circuit. En effet, il est possible de jouer sur les dimensions des transistors ou le placement/routage du circuit afin de réduire les capacités de grille et les capacités parasites. Cela se fait en réduisant par exemple les tailles des transistors et la longueur des interconnections [59, 60]

Agir sur l'activité du circuit α revient à jouer sur le nombre de portes qui commutent de '0' vers '1' par front d'horloge. Bien que possible cette méthode n'est pas simple car cela revient à éliminer les transitions de portes non souhaitées qui peuvent survenir lorsque plusieurs signaux en entrée d'une même porte n'ont pas le même temps de propagation sans compter le fait que ces temps de propagations

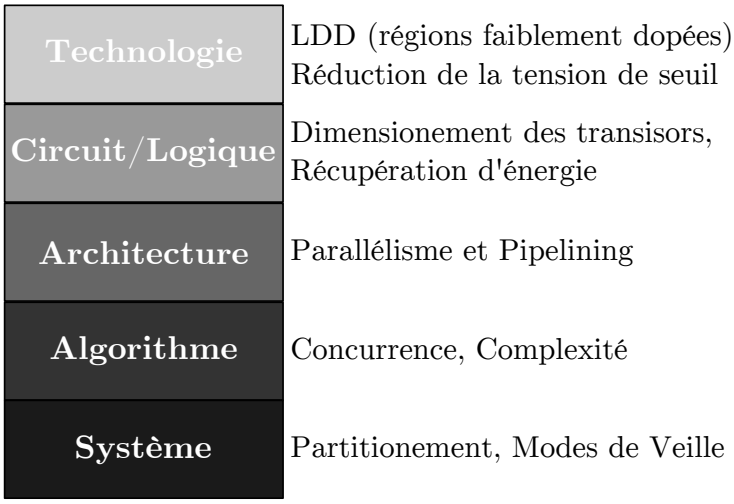


FIGURE 1.35 – Méthodes de réduction de la consommation à différents niveaux d'abstraction [13, 14]

peuvent être sujets à une gigue asynchrone [61].

Il est également possible de réduire la tension d'alimentation V_{dd} pour réduire la consommation dynamique. Cependant, cela se fait au dépend des performances. Une baisse de V_{dd} entraînant une augmentation du temps de commutation (le temps nécessaire à charger la capacité C de 0 à $V_{dd}/2$ ou à décharger C de V_{dd} à $V_{dd}/2$ augmente lorsque V_{dd} diminue), entraînant donc une fréquence maximale de fonctionnement plus faible. Cette méthode est limitée par l'impossibilité de descendre en dessous de $V_{dd} = 2V_{th}$ (avec V_{th} la tension de seuil des transistors) pour assurer un fonctionnement normal du circuit, $V_{dd}/2$ étant la limite entre un '1' et un '0' logique.

A plus haut niveau, il est également possible d'ajuster ces paramètres sur des portions restreintes d'un circuit, afin de réduire la consommation des parties du circuits inactives ou non-critiques du circuit. Une méthode consiste à définir plusieurs zones dans le circuit ayant des tensions d'alimentations différentes, afin de sous alimenter les parties sur circuit dont les performances ne sont pas critiques. On parle alors de "Cluster Voltage Scaling" [58, 62]. Pour aller plus loin, les tensions d'alimentation V_{dd} et les fréquences des différents îlots peuvent être ajustées dynamiquement. On parle alors de "Dynamic Voltage and Frequency Scaling" [58]. En effet, cette méthode permet de sous alimenter des blocs d'un circuit dont l'uti-

lisation est peu fréquente mais dont les performances restent importantes, comme les différents périphériques d'un micro-controlleur.

De la même manière, les méthodes dites de "Clock Gating" et de "Power Gating" permettent de couper complètement l'horloge ou l'alimentation de certains blocs de circuits [58]. Le clock gating a l'avantage d'être peu coûteux et de permettre une granularité très fine. Le power gating en revanche peut nécessiter un temps de remise en route lorsqu'un bloc sort d'hibernation. Plusieurs niveaux de power gating peuvent être définis, afin par exemple de maintenir ou non l'alimentation des éléments mémoire volatils.

Ces méthodes de gestion locale de fréquence et de tension alimentation ont toutefois un coût : la complexification des étages d'alimentation et la nécessité d'ajouter de la logique de contrôle servant à gérer l'alimentation des différents blocs. Ce coût dépend du nombre de zones définies et du nombre de paramètres ajustables. Il est donc important de trouver un compromis afin de s'assurer du gain réel de consommation.

1.7.4 Body biasing et architecture triple-well

La technique de réduction de la consommation étudiée dans le cadre du projet dans lequel ce manuscrit s'inscrit est celle dite du Body-Biasing (ou polarisation des caissons en français). Elle permet d'ajuster l'équilibre entre la consommation et les performances du circuit via la polarisation des caissons des transistors. En modifiant cette dernière, on affecte la tension de seuil V_{th} ainsi que leurs courants de fuite. Une réduction de la tension de seuil des transistors permet une augmentation de la vitesse de commutation des portes logiques et donc une fréquence de fonctionnement plus élevée (pour une tension d'alimentation donnée). En revanche, en abaissant la tension de seuil des transistors, leurs courants de fuite augmente. Cette méthode permet donc un gain en termes de performance du circuit au prix d'une consommation plus importante. Pour compenser cet effet, la méthode de Body Biasing associe une baisse de la tension de seuil des transistors via la polarisation du substrat à une réduction de la tension d'alimentation afin de diminuer la consommation du circuit tout en conservant des performances équivalentes [63].

1.7.4.1 Architecture triple-well

L'utilisation de cette technique nécessite l'emploi d'une architecture triple caissons (*triple-well* en anglais). Dans une architecture CMOS classique, les NMOS sont situés directement dans le substrat dopé P, et les PMOS sont situés dans un caisson dopé N (appelé NWell), comme montré sur la figure 1.36. Le substrat est polarisé à la masse et le NWell à V_{dd}.

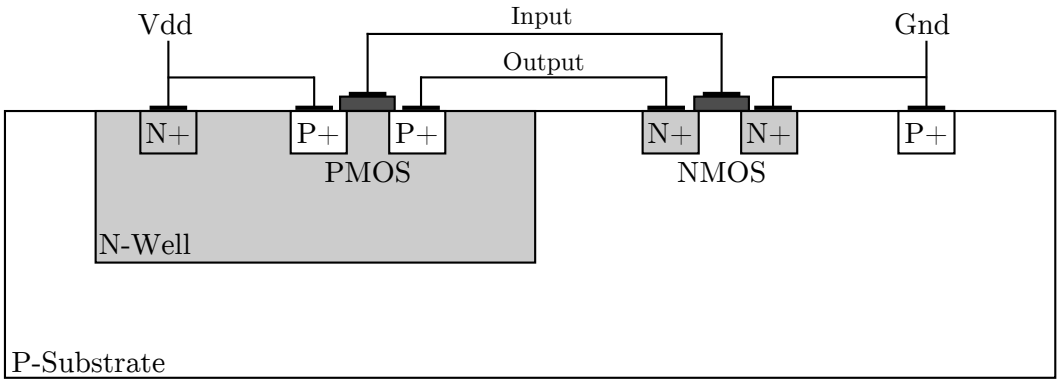


FIGURE 1.36 – Vue en coupe d'un inverseur en technologie CMOS standard.

Afin d'isoler les caissons des NMOS et de permettre de modifier la polarisation de leur caissons, l'architecture *triple-well* ajoute un implant NWell profond (Deep NWell) permettant d'avoir un caisson PWell isolé du substrat du circuit, en dessous par le Deep-NWell et latéralment par le NWell, comme montré sur la figure 1.37.

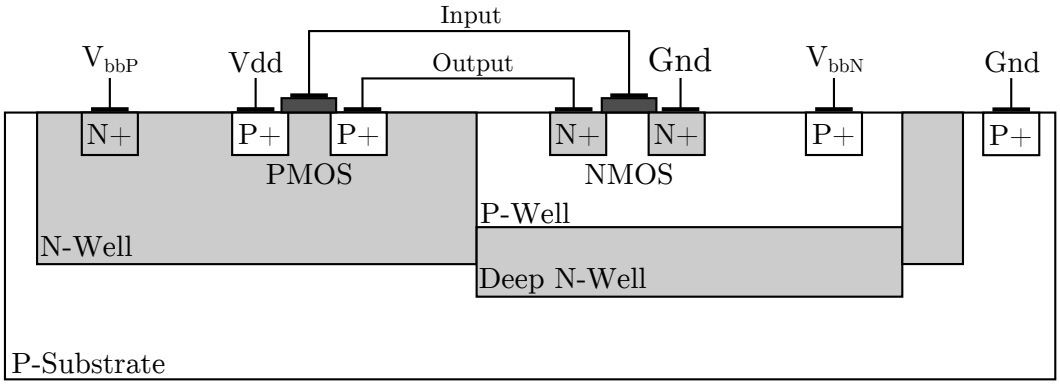


FIGURE 1.37 – Vue en coupe d'un inverseur en technologie CMOS *triple-well*.

Le fait que le PWell soit isolé du substrat permet de le polariser à une tension différente de celle du substrat, permettant ainsi l'emploi de la technique de body-

biasing.

1.7.4.2 Body biasing

Le *body-biasing* consiste à polariser le NWell à une tension différente de Vdd et le PWell à une valeur différente de Gnd. Ce faisant, le V_{th} des transistors ainsi que leurs courants de fuite sont modifiés [63]. Afin d'ajuster la polarisation des caissons, une tension de body-bias V_{bb} est introduite. Ainsi, les caissons PWell des NMOS seront polarisé à $V_{bbP} = V_{dd} - V_{bb}$ et les caissons NWell des PMOS à $V_{bbN} = V_{bb}$.

La tension V_{bb} peut être positive, on parle alors de Forward Body-Biasing (FBB), ou négative, on parle de Reverse Body-Biasing (RBB) comme montré sur la figure 1.38.

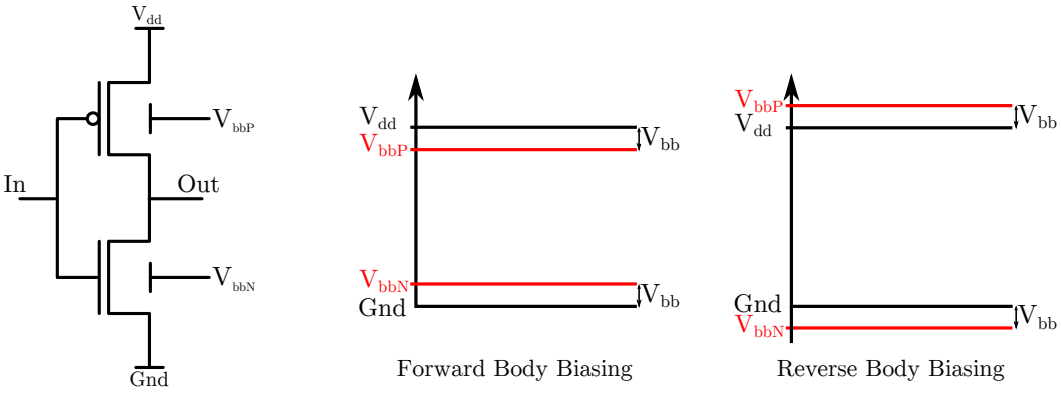


Figure 1.38 – Forward et Reverse Body-Biasing

L'expression de V_{th} pour les NMOS est donnée par l'équation 1.9 où V_{th0} est la valeur de la tension de seuil en l'absence de *body-biasing*, γ le paramètre de body-effect, $2\Psi_B$ le potentiel de surface et V_{BS} la différence de potentiel entre le caisson du transistor et la source (soit V_{bb} dans notre cas) [64].

$$V_{th} = V_{th0} + \gamma(\sqrt{|2\Psi_B - V_{BS}|} - \sqrt{|2\Psi_B|}) \tag{1.9}$$

Ainsi, le RBB a pour effet d'augmenter V_{th} ce qui augmente le temps de commutation des portes, et de réduire les courants de fuites (en augmentant la polarisation inverse de la jonction PN), réduisant ainsi la consommation statique au prix de performances temporelles dégradées, la tension de seuil réduite imposant une réduction de la fréquence de fonctionnement maximale.

Le FBB a l'effet inverse : il diminue V_{th} ce qui diminue le temps de commutation des portes, et augmente les courants de fuites, permettant ainsi d'obtenir de meilleures performances au prix d'une consommation statique augmentée [64].

En couplant l'utilisation du *body-biasing* à des variations de la tension d'alimentation, il est alors possible de contrôler l'équilibre entre les performances et la consommation d'un circuit en temps réel.

1.8 Conclusion

Le but de ces travaux est d'évaluer la robustesse des techniques de réduction de la consommation basées sur le Body-Biasing face à l'injection de faute par laser. Dans ce but, trois points principaux ont été abordés dans ce chapitre : la cryptographie matérielle, l'injection de faute par laser et les modèles de simulation qui y sont associés et enfin les techniques de conception basse-consommation.

La première partie, a présenté en détail l'algorithme AES qui sera utilisé dans la suite de ces travaux comme exemple de test pour la contre-mesure proposée dans le chapitre 4, ainsi que les attaques par injection de fautes et les contre-mesures qui y sont associées.

L'injection de fautes laser a ensuite été approfondie, ainsi que le modèles de simulation qui sera utilisé par la suite pour valider les résultats expérimentaux obtenus.

Enfin les techniques de conception basse-consommations et en particulier celles utilisant le *body-biasing*, qui font l'objet de ces travaux ont été présentées dans la troisième partie de cet état de l'art.

Le prochain chapitre présentera le banc de test utilisé dans la suite des travaux, ainsi qu'une étude préliminaire portant sur l'influence de la durée d'impulsion sur l'injection de fautes sur mémoires SRAM.

CHAPITRE 2

Banc de test et étude de l'injection de faute laser sur mémoires SRAM

2.1 Introduction

A présent que les éléments théoriques utilisés dans la suite de ce manuscrit ont été présentés dans le chapitre précédent, ce chapitre présente le banc de test utilisé dans les parties expérimentales de ces travaux.

Le chapitre se découpe en deux parties, la première présente le banc de test et ses différents composants, ainsi que le travail de développement effectué pour permettre son automatisation.

La seconde partie présente les premiers résultats obtenus avec le banc de test dans le cadre de cette thèse. Des expériences effectuées sur le même banc de test, présentées dans [15], visent à étudier le comportement de cellules SRAM face à des impulsions laser d'une durée de 50 ns. Les travaux présentés ici reproduisent ces tests avec une durée d'impulsion de 30 ps afin d'étudier les effets d'impulsions de très courte durée. Des différences de comportement sont observées, en conséquence, une adaptation du modèle de simulation de Sarafianos et al. (présenté dans le chapitre précédent) est ensuite proposée pour prendre en compte ces différences de

comportement. Enfin, les résultats sont validés sur la RAM d'un microcontrôleur.

2.2 Banc de test

Le banc de test laser utilisé dans le cadre de ces travaux disposes de capacités uniques en Europe au moment de sa mise en place. Il permet l'utilisation de larges plages de puissance (de 0 à 100W) et de durée d'impulsion (de 30 ps à 1 s) avec une excellente précision spatiale (diamètre minimal de spot d'environ 1 µm). Cette section présente son fonctionnement et ses capacités.

La petite taille de la zone d'effet du laser rend nécessaire de tracer des cartes d'une zone donnée d'un circuit pour pouvoir observer les effets de l'injection laser sur cette dernière. La cartographie (le fait de réaliser des cartes) est souvent utilisée dans le domaine de la sécurité matérielle pour évaluer les effets spatiaux des attaques. En effet, les résultats et l'efficacité de certaines méthodes d'injection de fautes (EM, BBI, laser) dépendent de la position du circuit où l'injection est effectuée, et tracer des cartes des résultats d'attaques permet de mieux comprendre les phénomènes en oeuvre. La cartographie peut également être utilisée pour effectuer de la rétro-ingénierie (par exemple, en reliant la position d'injection de faute sur une mémoire, aux bits fautés par l'attaque pour plusieurs emplacements d'injection, il peut être possible de reconstituer le plan d'adressage de la mémoire).

Dans cette optique, le banc est constitué de nombreux éléments permettant la réalisation de cartes détaillées des circuits testés. Les principaux composants sont représentés sur le schéma de la figure 2.1.

- Le circuit testé est placé sur un porte échantillon au format wafer 8 pouces, dont la position est réglable manuellement sur les axes X/Y/Z et en rotation autour de Z.

- L'objectif par lequel le faisceau laser est émis est situé en dessous du porte échantillon, permettant ainsi l'attaque de circuits par la face arrière. Cela permet également d'effectuer des tests sur wafer via la face arrière tout en posant des pointes de mesure sur la face avant de celui-ci. Trois objectifs (5X,

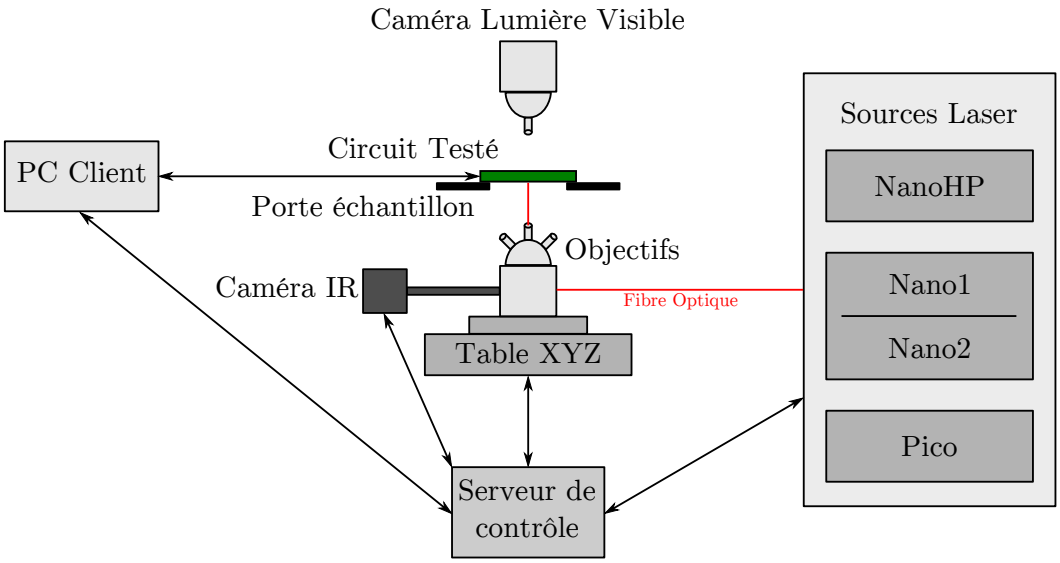


FIGURE 2.1 – Schéma des différents composants du banc de test

20X et 100X) peuvent être utilisés, permettant ainsi d'obtenir trois diamètres de spot laser différents (20 µm, 5 µm et 1 µm respectivement).

- Une caméra infrarouge montée sur le même chemin optique que celui du faisceau laser permet d'observer la face arrière du circuit, ce qui a plusieurs utilités. Premièrement, la caméra infrarouge permet d'observer la face avant du circuit à travers le substrat ce qui permet de se repérer lors de tests effectués via la face arrière d'un composant. Deuxièmement, le plan focal de la caméra est le même que celui du faisceau laser, ce qui permet de visualiser le spot et d'effectuer sa focalisation facilement.

- L'ensemble objectif/caméra est monté sur une table XYZ motorisée. L'axe Z permet d'effectuer la focalisation du faisceau, tandis que les axes X et Y permettent de le déplacer. Le pas minimum de la table est de 0.1 µm. L'avantage de cette méthode est que le circuit testé n'est pas déplacé pendant le processus de cartographie, ce qui est souvent gênant en raison des connectiques ou des pointes de tests.

- Une caméra avec un microscope est montée au dessus du porte échantillon, son rôle principal est d'aider à ajuster le placement des pointes de test lors de travaux sur wafer.

- Quatre sources laser ayant des caractéristiques différentes sont disponibles. Trois sources délivrant des impulsions de durées minimum de l'ordre de la dizaine de nanosecondes et une délivrant des impulsions de 30 ps fixe :

  - Une source Nanoseconde "High Power" (NanoHP) délivre des impulsions d'une durée minimum de 100 ns avec une puissance allant jusqu'à 25W.
  - Deux sources Nanosecondes (Nano1&2) identiques, pouvant délivrer des impulsions d'une durée minimum de 50 ns avec une puissance de 3W maximum.
  - Une source Picoseconde délivrant des impulsion d'une durée fixe de 30 ps avec une énergie de 100 nJ maximum.

  Les quatre sources ont une longueur d'onde identique de 1064 nm, ce qui est optimal pour l'injection via la face arrière (le cas le plus courant), mais reste utilisable pour l'injection via la face avant (moins courant, et une longueur d'onde optimale pour la face avant ne serait pas utilisable en face arrière). Les tirs lasers peuvent être déclenchés de manière continue en réglant une fréquence de répétition, ou en envoyant un signal (électrique ou logiciel). Les sorties des sources sont reliées au bloc optique objectifs/caméra IR via des fibres optiques. Deux sources peuvent être branchées en même temps, permettant ainsi d'utiliser deux configurations laser différentes simultanément.

- Tous ces éléments sont reliés à un serveur de contrôle auquel un PC peut se connecter par TCP/IP pour piloter le banc.

Les photographies des figures 2.2 et 2.3 montrent respectivement le banc de test et l'intérieur de l'enceinte du banc de test avec les différents éléments anotés.

2.2.1 Automatisation du banc

Bien que les différents éléments du banc soient pilotables via le serveur de contrôle, le banc, tel que livré par son fabricant, ne disposait pas de capacités d'automatisation poussées, tout les paramètres de configuration devant être entrés manuellement via

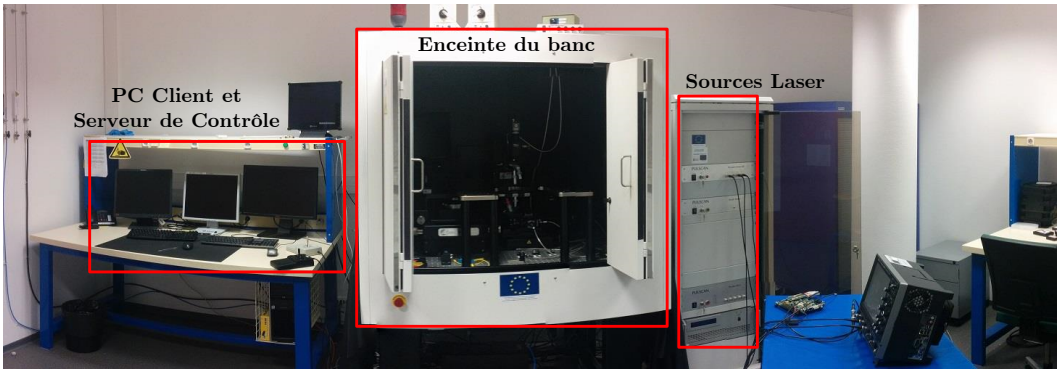


FIGURE 2.2 – Photographie du banc de test

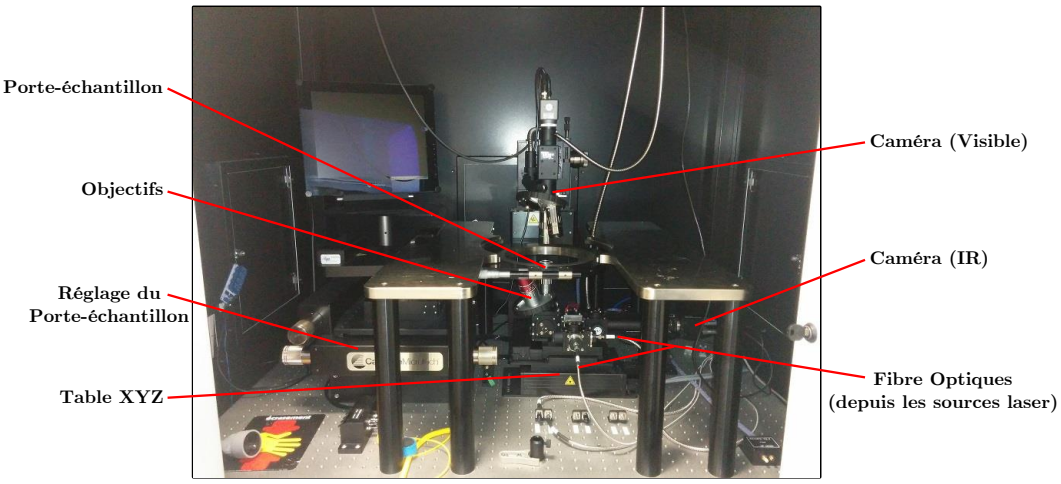


FIGURE 2.3 – Photographie de l'intérieur de l'enceinte du banc de test

la console du serveur, ce qui limitait les conditions d'utilisation du banc à des tests peu complexes.

La première étape des travaux effectués pendant cette thèse a donc consisté à rendre le banc de test automatisable facilement, afin de permettre la réalisation de tests pertinents pouvant durer jusqu'à plusieurs jours, et de développer un environnement logiciel permettant d'utiliser le banc de manière efficace.

Des librairies C/C++ étaient fournies avec le banc, cependant ces langages de programmation nécessitent d'être compilés et sont plus adaptés à l'écriture d'applications complètes que des langages de scripts comme Ruby ou Python, plus flexibles. Le choix a donc été fait de développer des librairies Ruby permettant d'utiliser les fonctions des librairies C/C++ directement.

Le schéma de la figure 2.4 décrit les différents éléments composant l'environe-

ment d'automatisation du banc. Les scripts sont développés directement par les utilisateurs du banc en fonction de leurs besoins, ces scripts peuvent faire appel aux éléments suivants. Une version plus détaillée du schéma décrivant l'architecture logicielle du banc est également disponible en annexe.

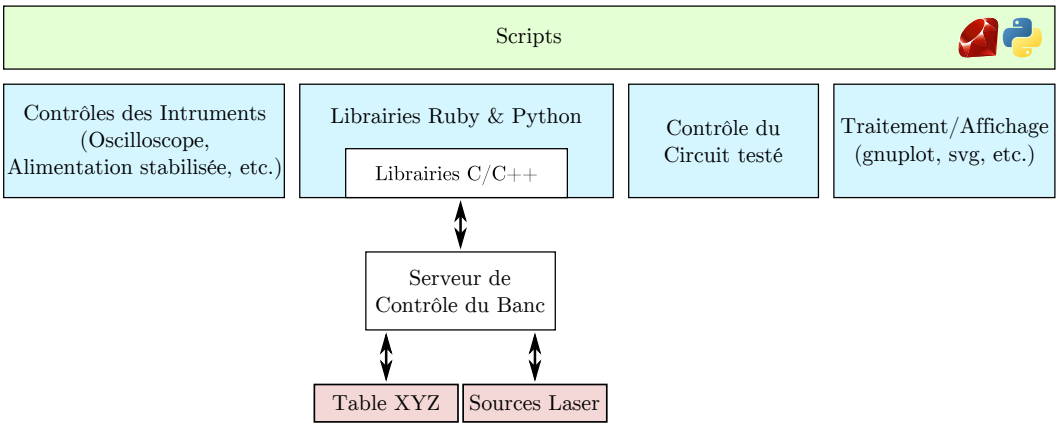


Figure 2.4 – Eléments du système d'automatisation du banc

- Des librairies Ruby permettant d'appeler directement les fonctions des librairies C/C+ permettant de se connecter au serveur de contrôle du banc pour piloter la table XYZ et les sources laser du banc.
- Des librairies permettant de contrôler facilement les instruments supplémentaires du banc ont également été développés (par exemple acquisition via un oscilloscope, pilotage des alimentations stabilisées).
- Ruby permet nativement de contrôler des liaisons séries ou USB, ce qui permet de contrôler les circuits testés
- Enfin, deux outils ont été développés pour faciliter l'affichage des résultats. L'un permet de générer des scripts permettant de tracer directement les cartes des résultats grâce à gnuplot (un logiciel de dessin mathématique). Le second permet de tracer des cartes au format SVG, et peut-être utilisé pour tracer les résultats de cartes en temps réel via une interface web. Une capture d'écran du suivi l'avancement d'une cartographie via cette interface est présentée figure 2.5.

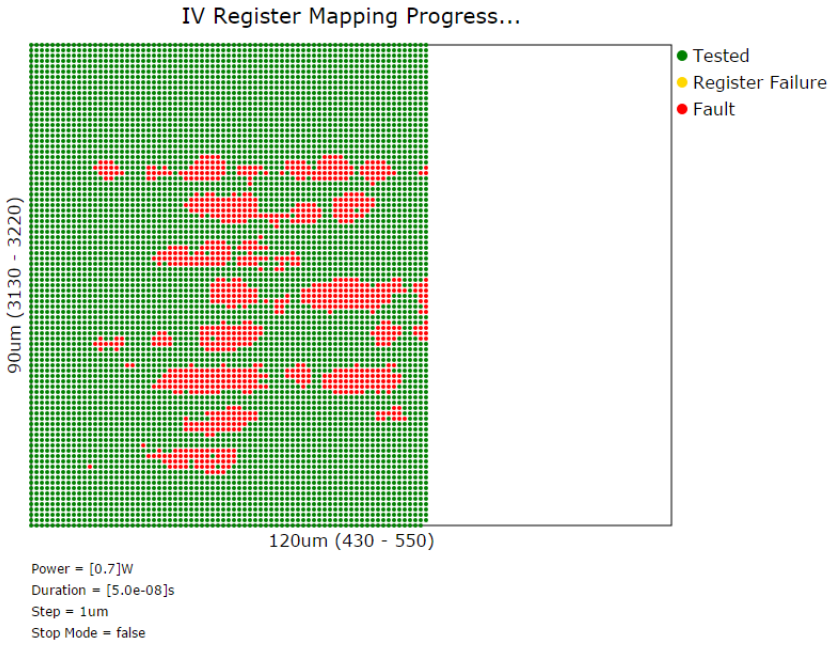


FIGURE 2.5 – Capture d'écran du suivi en temps réel de l'avancement d'une cartographie

2.2.1.1 Fiabilité du banc de test

Le banc de test laser étant un équipement unique et récent, il est hélas sujet à quelques défaillances. N'ayant accès ni à des documents détaillant la conception des sources précisément, ni au code du serveur et de ses différents composants, un travail continu de mise en place de solutions de contournement des bugs, au fur et à mesure de l'apparition des défaillances a ainsi du être effectué, ainsi que la liaison et le suivi des maintenances avec le fournisseur.

Un exemple récent est un cas où les sources Nano cessaient de fonctionner après un temps d'utilisation continu allant d'une trentaine de minutes à quelques heures, ce qui s'avérait bloquant pour effectuer des tests pouvant durer plusieurs jours. Pour une raison inconnue, ouvrir et refermer la porte de l'enceinte semblait corriger le problème ce qui tendait à penser que la défaillance venait de l'interlock du banc (une sécurité qui empêche le laser de déclencher un tir lorsque la porte est ouverte). Un relai piloté par USB a donc été installé en série dans la chaîne de l'interlock permettant ainsi de déclencher une ouverture/fermeture de ce dernier à intervalle réguliers pendant l'exécution des tests, ce qui a résolu le problème en attendant une

solution définitive.

D'autres défaillances sont en revanche restées sans solutions, par exemple, la table XYZ sur laquelle l'objectif du laser est montée, présente des dérives sur les trois axes. Les dérives en X et Y peuvent être minimisées en choisissant des paramètres expérimentaux adaptés, mais les dérives de l'axe Z entraînent une perte de la focalisation du faisceau qui peut fausser les résultats obtenus si les temps de cartographies sont trop longs. Plusieurs causes à ces dérives (variations de température, usure prématurée de la table due à la forme du bloc objectif/caméra IR, etc.) ont été envisagées mais aucune n'a pu être confirmée.

Ces exemples sont les plus gênants rencontrés pendant ces travaux. Cependant le cumul d'une longue liste de bugs mineurs avec des solutions de contournement simples fait que l'utilisation optimale du banc de test demande une bonne connaissance de ce dernier. Du temps à donc également été consacré à la formation et au support des nouveaux utilisateurs.

2.2.1.2 Conclusion

Réunis, ces éléments constituent un système de scripts d'automatisations flexibles et rapide à prendre en main, ce qui permet à un nouvel utilisateur de développer ses propres scripts permettant de mettre en place des tests variés en peu de temps. Les librairies de liaison avec le serveur de contrôle du banc de test ont également été développées pour Python sur demande de certains utilisateurs.

Ce système d'automatisation est utilisé par la totalité des utilisateurs du banc de test, ce qui a contribué à l'heure actuelle, à cinq thèses dont [65] et [66], ainsi qu'à de nombreux autres travaux répartis sur plusieurs projets de recherche.

Certains éléments non-spécifiques au banc laser, comme les modules de traçage de cartes et de pilotage des instruments ont également été réutilisés sur d'autres bancs de tests.

2.3 Injection de fautes laser sur mémoire SRAM

Suite à la mise en oeuvre de l'automatisation du banc, la première étape a été de se familiariser avec son utilisation. Les premiers travaux utilisant les possibilités offertes par l'automatisation du banc, publiés par Roscian et al. dans [15], présentent le modèle de faute obtenu sur des cellules SRAM à l'aide de la source Nanoseconde du banc de test.

Cette section présente ces mêmes expériences reproduites cette fois avec la source Picoseconde, les différences de comportement observées et une adaptation du modèle de simulation pour les impulsions courtes.

2.3.1 Aspects théoriques : zones sensibles et modèle de faute

Comme vu dans le chapitre 1, une cellule SRAM est constituée de deux inverseurs rebouclés l'un sur l'autre. Lorsqu'un tir laser cible un inverseur, la zone sensible pouvant générer un transitoire de tension est le drain du transistor OFF (le NMOS lorsque l'entrée de l'inverseur est à '0' et le PMOS lorsque l'entrée est à 1) comme sur la figure 2.6.

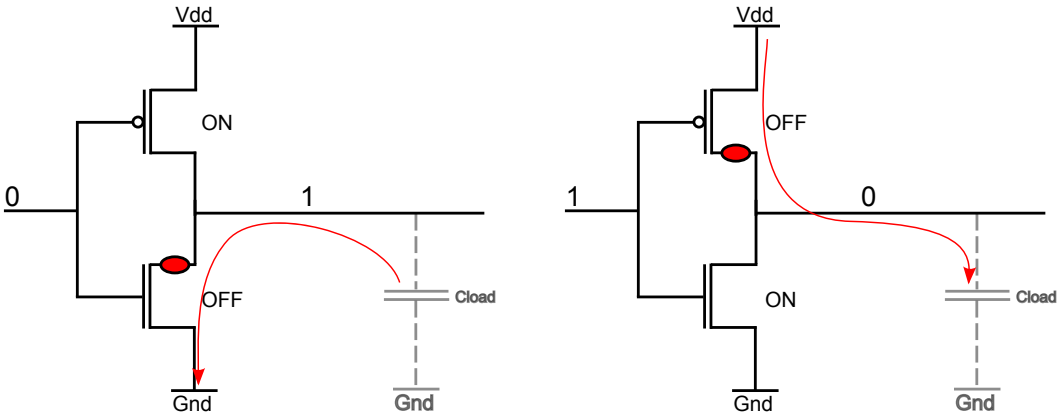


FIGURE 2.6 – Injection Laser dans un inverseur

Le cas d'un tir sur une cellule SRAM est présenté figure 2.7 : au repos, la cellule SRAM possède deux noeuds représentant la donnée stockée et son inverse (a). Lorsqu'un des deux inverseurs est frappé par un tir laser (b), le transitoire de tension généré se propage dans le second inverseur (c). Enfin, quand la perturbation disparaît, la cellule étant dans un état stable, celle-ci ne revient pas à son état initial,

une faute est donc injectée dans le circuit de manière permanente.

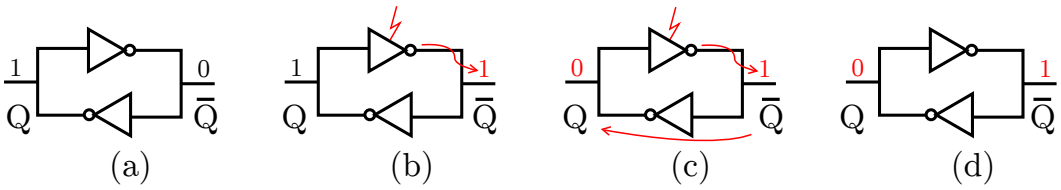


Figure 2.7 – Injection de faute dans une cellule SRAM

La figure 2.8, montre le schéma transistor d'une cellule SRAM 5 transistors (deux par inverseur et un pour la sélection de la cellule), comme celle ciblée par les travaux de [15] et ceux présentés dans la suite de ce chapitre. En fonction de la donnée stockée, les zones sensibles des inverseurs varient. Ainsi lorsque la donnée stockée est '0', les drains des transistors MP1 et MN2 sont sensibles, inversement, lorsque la donnée stockée est '1', les drains des transistors MN1 et MP2 sont sensibles.

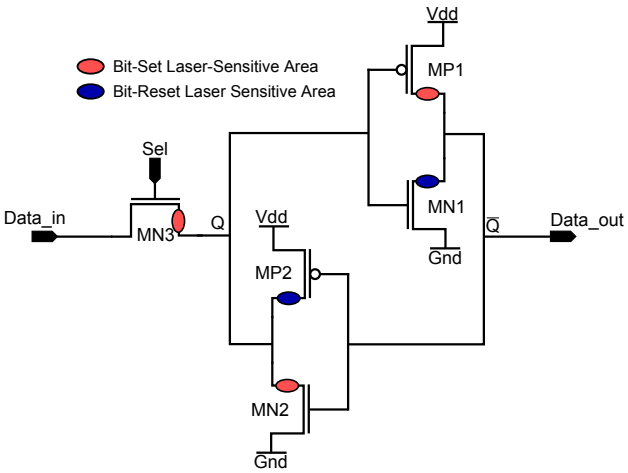


Figure 2.8 – Zones sensibles d'une cellule SRAM 5 Transistors

Il est alors possible de distinguer deux types de zones sensibles sur la cellule. Celles qui peuvent la faire basculer de '0' vers '1' (en rouge sur la figure) et celles qui peuvent la faire basculer de '1' vers '0' (en bleu sur la figure). On parle alors de fautes de type *bit-set* (la cellule est écrite à '1' indépendamment de sa valeur initiale) ou *bit-reset* (la cellule est mise à '0' indépendamment de sa valeur initiale).

Les zones sensibles sont représentées sur le layout de la cellule figure 2.9a (note : le drain des transistors MN2 et MN3 est commun, pour cette raison le drain de MN3 est marqué comme étant une zone sensible au *bit-set* sur la figure 2.8).

L'effet du tir laser touchant une zone plus importante que les surfaces des drains comme montré sur la figure 2.9b, on peut émettre l'hypothèse qu'aux intersections entre deux zones il est possible d'injecter des fautes indépendamment de la valeur initiale stockée dans la cellule. On parle alors de fautes de type *bit-flip*.

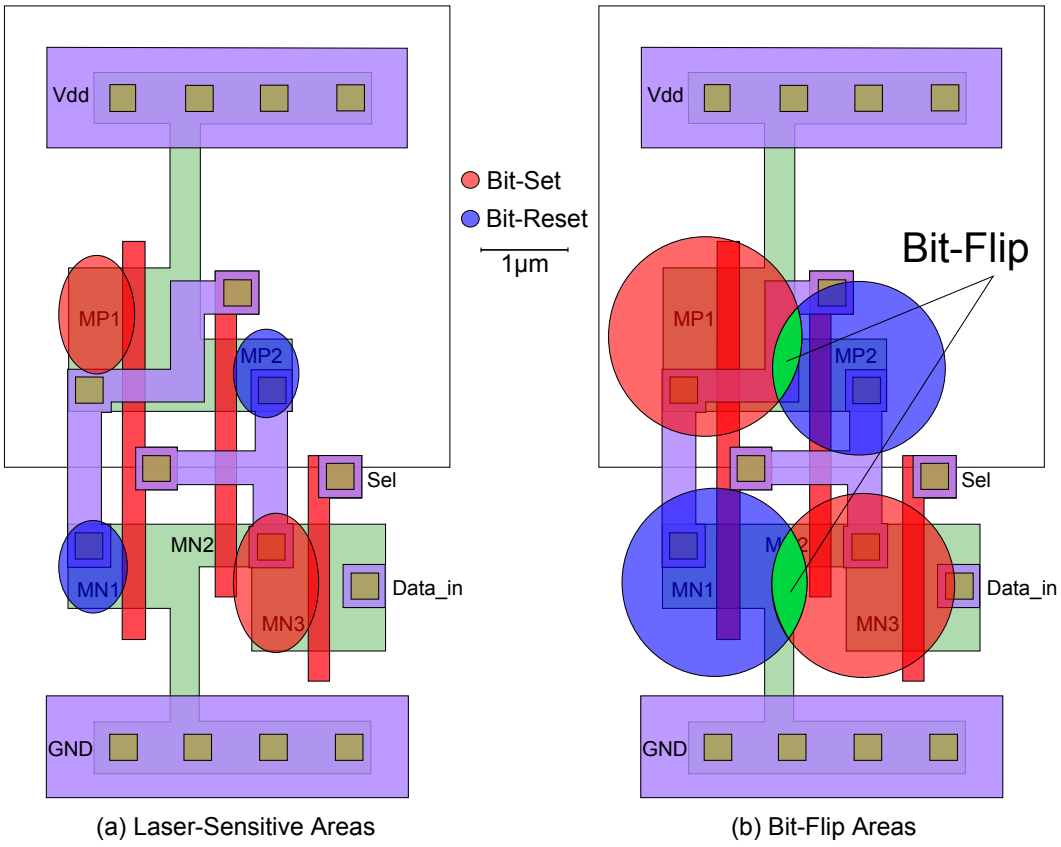


FIGURE 2.9 – Layout de la cellule SRAM testée

Le but des travaux de Roscian et al. ainsi que de ceux présentés par la suite de ce chapitre était de valider ces hypothèses : la présence de quatre zones sensibles différentes et la possibilité ou non d'obtenir des fautes de types *bit-flip*. Une bonne connaissance du modèle de faute est un paramètre important lors du choix de l'attaque à appliquer sur un système sécurisé. Les fautes de type *bit-set*/*bit-reset* permettent par exemple de mettre en place des attaques par "safe error" comme celle présentée dans [67]. Les travaux de ce chapitre ont pour objectif de valider les résultats obtenus par Roscian et al. et d'observer d'éventuelles différences de comportement lors de l'utilisation d'impulsions laser plus courtes.

2.3.2 Injection de fautes dans une cellule SRAM seule

Dans le but d'observer les différentes zones sensibles, les prochaines expériences visent à cartographier une cellule SRAM. La cible est un circuit en technologie CMOS 0.25 µm conçu spécialement pour les tests laser. Il comporte notamment un point mémoire SRAM utilisant le layout présenté dans la section précédente, cette cellule sera la cible des tests. Une description détaillée du circuit est disponible dans [68].

2.3.2.1 Configuration et processus de test

Les travaux de [15] et ceux présentés dans cette section ont été réalisés sur le même circuit avec le même banc de test.

Une carte FPGA connectée au PC de contrôle du banc de test par une liaison série permet de piloter le circuit.

Afin de cartographier la surface ($4\,\mu m \times 9\,\mu m$) occupée par la cellule, celle-ci est quadrillée avec un pas de 0.2 µm. Pour chaque position testée le processus de test est le suivant :

1. La valeur stockée dans la cellule est écrite à '1',
2. La valeur contenue est alors re-lue pour vérifier qu'aucun dommage permanent n'a été infligé à la cellule,
3. Le tir laser est déclenché,
4. La valeur contenue dans la cellule est à nouveau lue pour vérifier si une faute de type *bit-reset* a été injectée.

Le processus est ensuite répété en écrivant '0' dans cellule pour tester l'injection de fautes de type *bit-set*.

A partir des résultats des tests il est ensuite possible de tracer une carte des résultats obtenus. De plus, dans le cas où une position engendrerait à la fois une faute de type *bit-set* et une faute de type *bit-reset*, la possibilité d'injecter des fautes de type *bit-flip* serait alors confirmée.

Le circuit cible ayant été conçu spécialement pour effectuer des tests d'injection laser, il comporte peu de pistes métalliques susceptibles de bloquer le faisceau laser, rendant ainsi possible l'injection par la face avant du circuit. L'objectif 100X du banc est utilisé, donnant ainsi une taille de spot d'environ 1 µm.

Roscian et al., utilisaient la source Nanoseconde du banc de test, la durée d'impulsion est de 50 ns. Dans le cas des travaux présentés ici, la source Picoseconde est utilisée, la durée d'impulsion est alors de 30 ps.

Dans les deux cas la longueur d'onde est de 1064 nm (infrarouge) et la puissance du laser est réglée juste en dessous du seuil à partir duquel des dommages permanents sont infligés au circuit.

2.3.2.2 Impulsions de 50 ns

Les résultats obtenus par Roscian et al. dans [15] sont présentés par la figure 2.10 : pour chaque position de la carte, un point est dessiné lorsqu'une faute est injectée dans la cellule. Les points rouges représentent les fautes de type *bit-set* et les points bleus les fautes de type *bit-reset*.

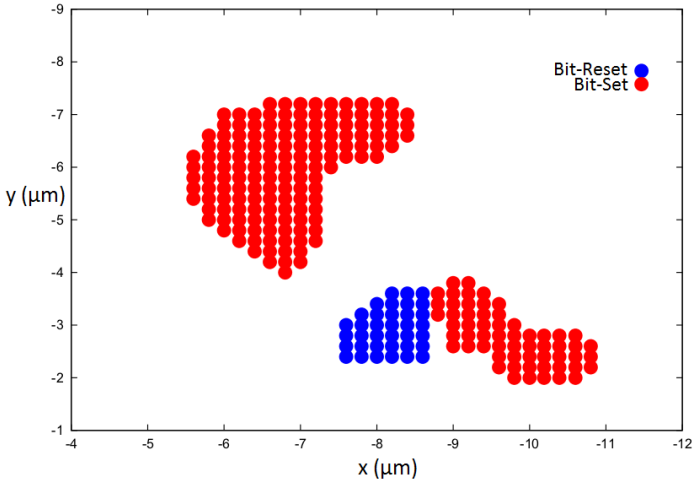


FIGURE 2.10 – Carte de la cellule SRAM obtenue avec une durée d'impulsion de 50 ns [15]

Deux résultats en ressortent : d'une part, la zone sensible censée correspondre au drain de du transistor MP2 n'apparait pas sur la carte. D'autre part, aucune faute de type *bit-flip* n'a été enregistrée.

Via des simulations SPICE utilisant le modèle de Sarafianos et al. décrit dans [54–56] et présenté dans le chapitre 1, les auteurs de [15] montrent que l'absence de sensibilité dans le drain de MP2 s'explique par la grande taille de la zone partagée par le drain de MN2 et MN3 située à proximité. Celle-ci étant sensible à l'état '0', un courant s'y forme, empêchant la cellule de basculer.

De manière similaire, ils montrent que le basculement de la cellule se fait uniquement lorsque le photocourant injecté dans la zone sensible visée est suffisant pour dépasser les photocourants induits dans les zones sensibles adjacentes. Un tir ciblé entre deux zones entraîne des photocourants avec des effets opposés dans les deux zones ciblées, ce qui explique en partie l'absence de *bit-flip* à la limite entre les zones sensibles de MN1 et MN2.

Ces effets ont notamment été utilisés par Sarafianos et al. pour proposer une méthode de durcissement de cellules SRAM. En utilisant un layout spécifique, l'effet de compensation permet de réduire la sensibilité au laser de la cellule [43].

2.3.2.3 Impulsions de 30 ps

Dans cette section, les expériences de la section précédentes sont reproduites avec la source Picoseconde du banc de test, dans le but de valider les résultats de Roscian et al. ou d'observer des différences de comportement liées au changement de durée d'impulsion.

La source Picoseconde délivre des impulsions d'une durée fixe de 30 ps, en dehors de ce changement, les conditions expérimentales restent les mêmes.

La carte obtenue est présentée dans la figure 2.11. La première observations est qu'aucune faute de type *bit-flip* n'a été enregistrée, ce qui va dans le sens des résultats précédents. L'utilisation d'un modèle de faute de type *bit-set*/*bit-reset* semble donc plus approprié qu'un modèle de faute de type *bit-flip* pour l'injection de fautes sur des SRAM.

En revanche, la zone correspondant au drain du transistor MP2 qui n'apparaissait pas avec les impulsions de 50 ns est maintenant visible. Afin de comprendre cet écart de résultat, la prochaine section propose un ajustement du modèle de simulation de Sarafianos et al. présenté dans le chapitre 1 permettant de reproduire ce

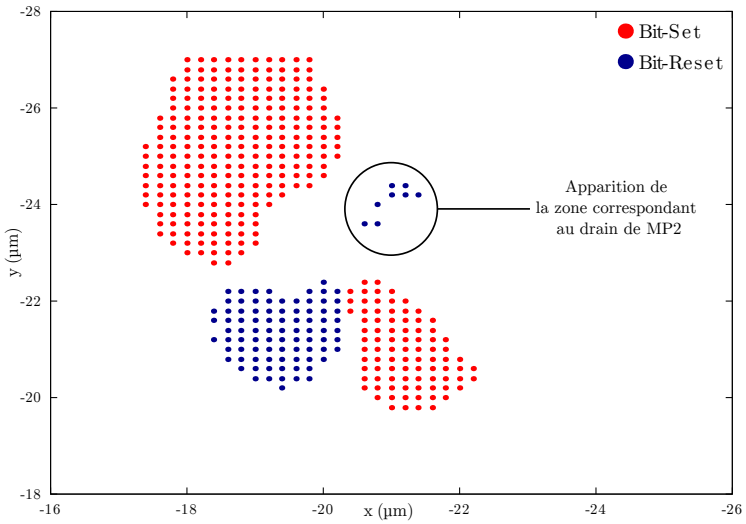


FIGURE 2.11 – Carte de la cellule SRAM 5T obtenue avec une durée d'impulsion de 30 ps

phénomène.

2.3.3 Adaptation du modèle de simulation de Sarafianos et al. aux impulsions courtes

Dans [15] il est montré que l'absence de la zone de sensible correspondant au drain du transistor MP2 lors des tests effectués avec des impulsions de 50 ns est liée à des photocourants créés dans les transistors adjacents à MP2 créant un effet de compensation. L'apparition de la zone sensible dans les résulats obtenus avec des impulsions de 30 ps montre que cet effet est alors moins présent pour des impulsions plus courtes. Une hypothèse possible pour expliquer ce résultat est que les impulsions plus courtes ont un effet spatial moins large. Cette section présente des simulations effectuées afin d'évaluer la validité de cette hypothèse.

On utilise le modèle de simulation présenté dans le chapitre 1 (équation 1.4). La figure 2.12 montre les positions des sources de courants ajoutées pour simuler les photocourants créés dans les drains des transistors. L'intensité délivrée par ces sources suit le profil temporel du photocourant décrit par le modèle, ajusté par un coefficient dépendant des paramètres spatiaux du modèle (distance du spot par rapport au centre de la jonction PN et focalisation du faisceau).

Pour simuler l'hypothèse de l'effet spatial réduit, on ajuste le paramètre $\alpha_{topology}$

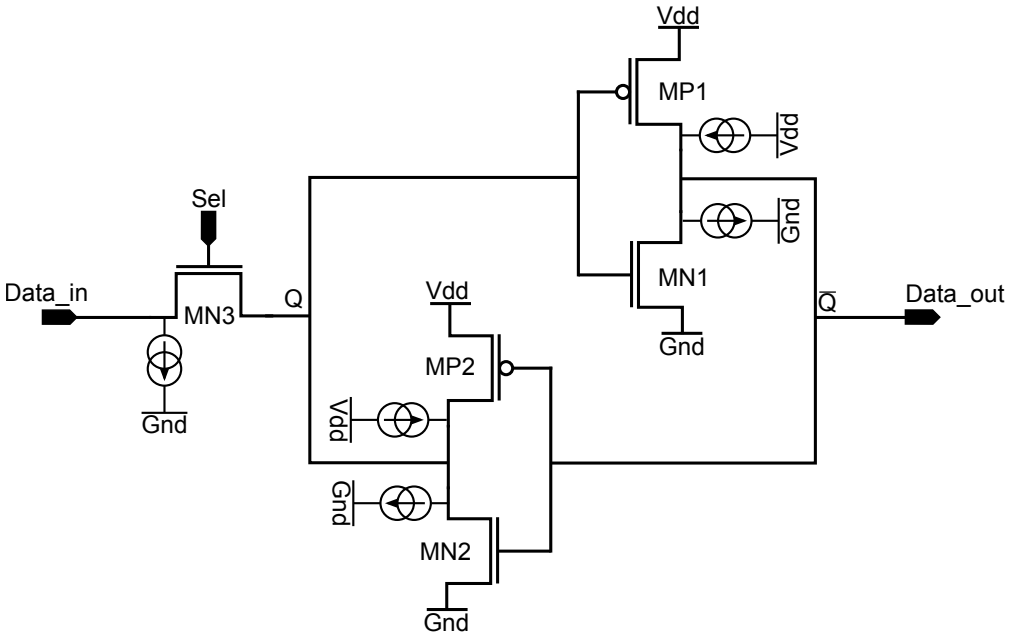


Figure 2.12 – Sources de courants simulant les photocourants injectés sur la cellule SRAM

du modèle correspondant au profil d'intensité du laser. Ce paramètre est un coefficient fonction de la distance entre le centre de la jonction et le centre du spot laser. La figure 2.13 montre la valeur du coefficient pour 50 ns et la valeur ajustée pour 30 ps.

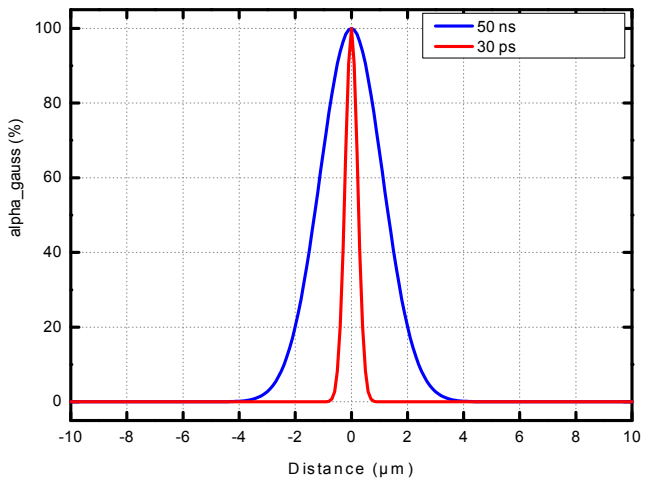


Figure 2.13 – Coefficient $\alpha_{topology}$ pour les impulsions de 50 ns et 30 ps

En utilisant ces nouvelles valeurs, les simulations sont répétées en faisant varier les valeurs du coefficient pour chaque source de courant en fonction de la position

du spot laser pour chaque point de la carte, ce qui donne la figure 2.14. On note que les formes des zones sensibles sont différentes de celles des résultats expérimentaux, ceci est du au fait que en pratique l'injection laser est faite via la face avant et une partie des zones sensibles est ainsi masquée par par les couches de métallisation dans les résultats expérimentaux. Les simulations, elles, en revanche ne tiennent pas compte des couches de métallisation, ce qui donne des formes plus régulières aux zones sensibles.

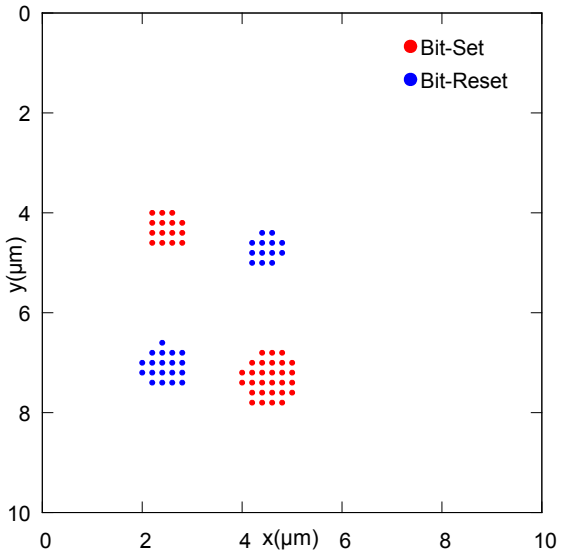


FIGURE 2.14 – Carte obtenue par simulation pour des impulsions courtes

La carte obtenue présente des fautes dans le drain de MP2 ce qui concorde avec les résultats expérimentaux obtenus précédemment.

Afin de mieux comprendre le phénomène, on s'intéresse aux courants créés dans la cellule lorsque le spot laser est situé au dessus du drain de MP2. Les figures 2.15 et 2.16 montrent les résultats des simulations pour des impulsions 50 ns et 30 ps respectivement. Pour les deux figures, le premier graphe représente les tensions des deux noeuds de la cellule, tandis que le second graphe représente les courants créés dans la cellule par le laser. Pour les impulsions de 50 ns un photo-courant non négligeable est créé dans le drain de MN2/MN3, qui s'oppose au courant créé dans le drain de MP2, empêchant ainsi la cellule de basculer. Pour les impulsions de 30 ps en revanche, le photocourant créé dans le drain de MN2/MN3 est quasi nul comparé

au courant créé dans le drain de MP2, et l'état de la cellule bascule.

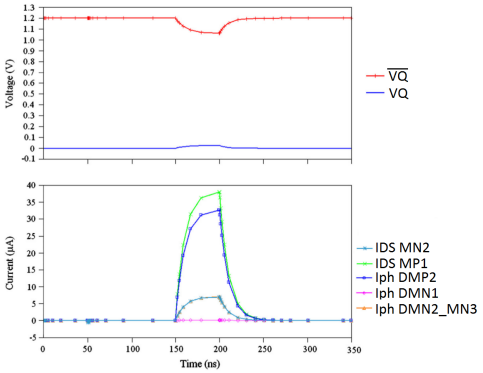


Figure 2.15 – Courants créés dans la cellule pour une durée d'impulsion de 50 ns [15]

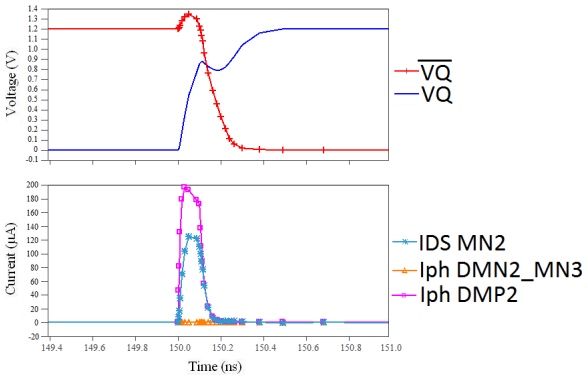


Figure 2.16 – Courants créés dans la cellule pour une durée d'impulsion de 30 ps

Ces résultats tendent à aller dans la même direction que les résultats expérimentaux. Cependant, les valeurs du coefficient utilisé pour les simulation des impulsions courtes ont été choisies de manière arbitraire. Dans le but d'obtenir des simulations de meilleure qualité, il serait intéressant de mesurer des valeurs expérimentales des coefficients. Cela s'avère complexe en raison de la durée extrêmement courte des phénomènes à mesurer lors de l'utilisation d'impulsions de 30 ps et nécessiterait des instruments de mesure que nous n'avions pas à disposition. De plus, cela s'éloigne du but inital de ces travaux, qui est d'étudier l'impact des technologies basse consommation sur l'injection laser.

Afin de valider leurs résultats Roscian et al. avaient reproduit leurs expériences sur la mémoire RAM d'un microcontrôleur, la section suivante présente une comparaison entre leurs résultats et ceux obtenus sur la même cible avec des impulsions de 30 ps.

2.3.4 Validation sur la RAM d'un microcontrôleur

Dans [15], les auteurs ont reproduit leurs résultats sur une cible disponible commercialement afin de les valider. Pour ce faire ils ont cartographié une portion de la RAM d'un microcontrôleur en technologie CMOS 0.35 µm. De même, ces expériences ont été reproduites dans des conditions identiques avec des impulsions de 30 ps.

2.3.4.1 Conditions expérimentales

Les conditions et le protocole de test sont similaires à ceux employés sur la cellule SRAM. Le principal changement est que l'injection se fait cette fois via la face arrière du circuit pour éviter les couches de métallisation. La zone cartographiée est une surface de 40 µm par 40 µm située dans RAM du microcontrôleur (cartographier la surface complète de la RAM avec un pas permettant d'avoir des résultats significatif prendrait un temps déraisonable).

De plus, ces tests nécéssitent d'écrire et relire l'intégralité de la RAM pour chaque tentative d'injection de faute, puisqu'on ne connait pas la position physique correspondant aux différentes adresses mémoires.

2.3.4.2 Résultats

La carte obtenue par Roscian et al. est présentée figure 2.17, sur la figure une cellule est encadrée en noir. Contrairement aux résultats précédents, pour chaque cellule, deux zones sensibles seulement apparaissent. Une raison probable à cela est que les mémoires RAM utilisent généralement des cellules 6 transistors, contrairement à la cellule faisant l'objet des tests précédents. Dans ce cas, un transistor d'accès supplémentaire est ajouté au niveau du noeud $\bar{Q}$ (symmétrique à MN3 sur la cellule précédente), ce qui augmente la surface du drain du NMOS adjacent, et crée ainsi un effet de compensation suffisant pour masquer une zone supplémentaire.

De plus, encore une fois aucun *bit-flip* n'est enregistré, ce qui confirme la validité de l'emploi d'un modèle de faute *bit-set*/*bit-reset* plutôt que celui d'un modèle de faute *bit-flip*.

En reproduisant cette expérience avec les impulsions de 30 ps, la carte présentée figure 2.18 est obtenue. Une cellule est encadrée en noir sur la figure. Les résulats obtenus durant cette cartographie valident les résultats obtenus précémment. Premièrement, cette fois, les quatre zones sensibles de chaque cellule apparaissent distinctement, ce qui confirme également l'hypothèse des cellules 6 transistors entrainant le masquage de deux zones par cellule.

Ensuite, bien que certaines zones de *bit-set* et *bit-reset* semblent se recouvrir sur la carte, en réalité aucun *bit-flip* n'a lieu, ces zones de recouvrement sont en réalité

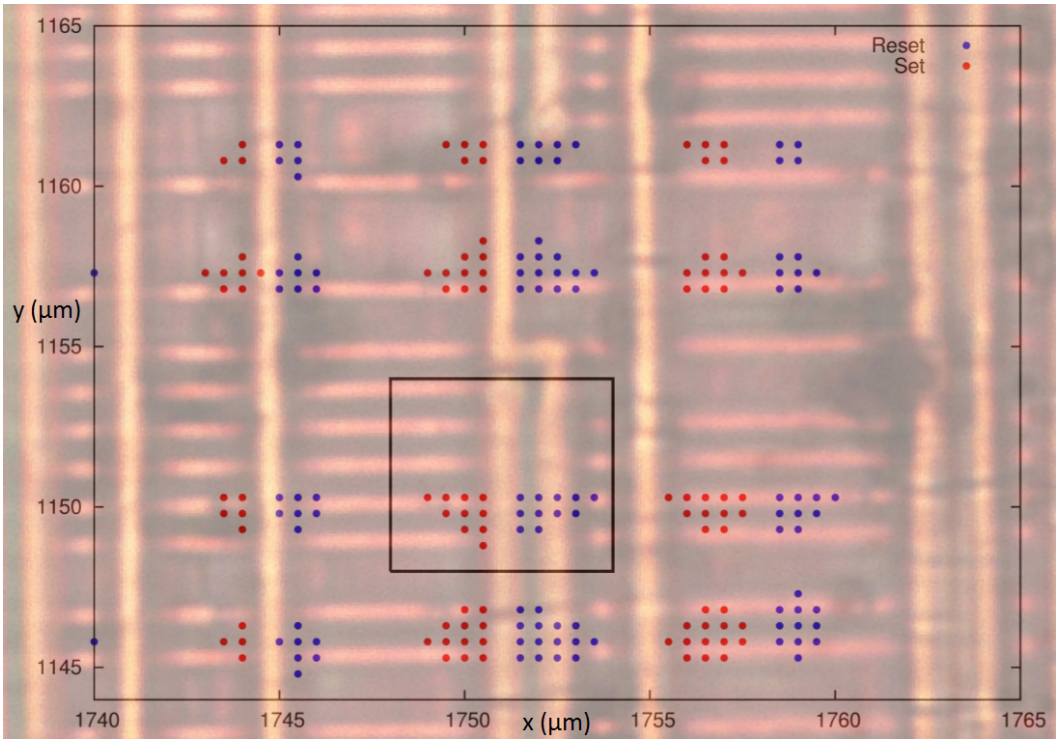


Figure 2.17 – Carte de la RAM d'un microcontrôleur obtenue avec des impulsions de 50 ns [15]

un *bit-set* sur une cellule et un *bit-reset* sur une cellule adjacente. Cela permet de confirmer la validité du modèle *bit-set*/*bit-reset* pour l'injection de faute sur des mémoires SRAM. Notons qu'un article récent de Courbon et al. confirme également ce choix de modèle de faute [69].

2.4 Conclusion

Dans ce chapitre, une première partie a présenté le banc de test utilisé dans ces travaux, ainsi que le système d'automatisation développé pour celui-ci. Ce système est a présent utilisé par tous les utilisateurs du bancs, ce qui contribue directement à de nombreux projets de recherches, et constitue la première contribution de cette thèse, sur laquelle la suite des travaux repose.

La seconde partie présentait les résultats des premières campagnes menées à l'aide du nouveau système d'automatisation. Les expériences présentées dans [15] ont été reproduites avec des durées d'impulsions plus courtes, ce qui entraîne un

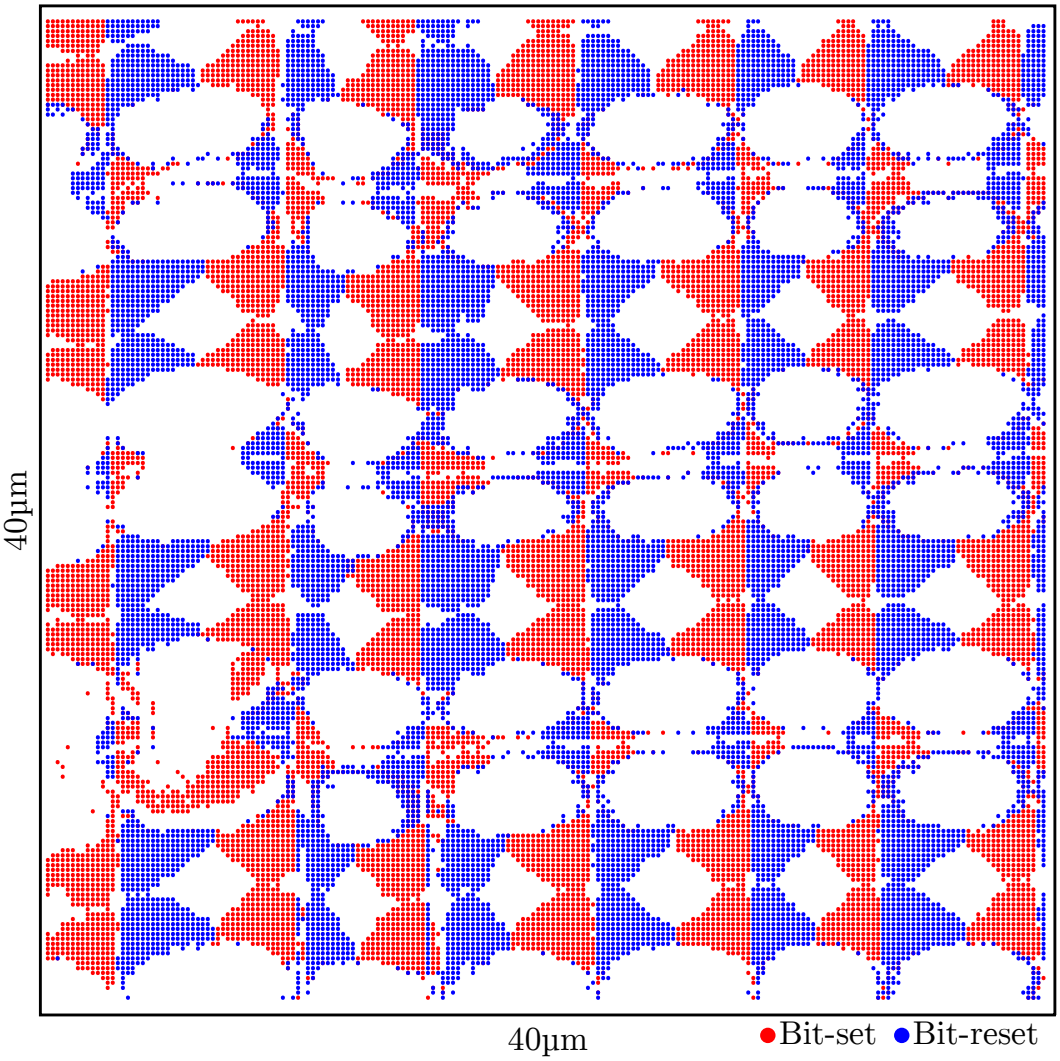


FIGURE 2.18 – Carte de la RAM d'un microcontrôleur obtenue avec des impulsions de 30 ps

comportement différent des cellulles SRAM testées. Les résultats de cette section ont fait l'objet d'une publication à IOLTS 2015 [12]. Les premiers résultats montrent que l'effet de masquage de sensibilité présent avec les impulsions de 50 ns n'avait pas lieu en utilisant des impulsions plus courtes. Puis le modèle de simulation Sarafianos et al. a été adapté aux impulsions courtes, et les résultats obtenus avec le modèle modifié concordent avec ceux obtenus expérimentalement. Enfin, ces résultats ont été validés sur la RAM d'un d'un microcontrôleur disponible dans le commercialement.

En conclusion, la durée d'impulsion doit être choisie avec précaution pour effec-

tuer des tests de sécurité ou fiabilité sur des mémoire SRAM. En effet, se limiter à des tests avec des impulsions de l'ordre de la nanoseconde expose au risque de passer à coté de vulnérabilités potentielles, que ce soit par simulation au moment de la conception ou lors de tests sur des échantillons réels. Par exemple la méthode de durcissement présentée dans [43] qui utilise l'effet de masquage des zones sensibles, est potentiellement inefficace face à des impulsions de 30 ps.

Cependant bien que les résultats de simulation concordent avec les résultas expérimentaux, l'adaptation du modèle de simulation proposée repose sur une extrapolation d'un modèle de simulation concu pour des durées d'impulsions beaucoup plus longues. Pour obtenir des résultats plus précis, l'idéal serait de reproduire les travaux de Sarafianos et al. [54–56] pour obtenir des valeurs expérimentales des coefficients du modèle. Toutefois, cela nécéssite une instrumentation de pointe en raison de la durée extrêmement courte des phénomènes en jeu. De plus, bien que le développement de modèles de simulations soit un point crucial pour la conception de circuits résistants aux attaques laser, cela s'éloigne du cadre de cette thèse qui est l'étude des effets des méthodes de réduction de la consommation sur l'injection laser.

Le prochain chapitre visera donc à étendre les résultats expérimentaux présentés dans ce chapitre en étudiant l'effet des techniques de conception basse consommations présentées dans le chapitre 1 sur des cellules SRAM. Des tests seront effectués sur le micro-contrôleur de la section 2.3.4 lorsque celui ci est sous-alimenté, et un micro-contrôleur capable d'utiliser le FBB sera également étudié.

CHAPITRE 3

Etude de l'injection laser sur des mémoires basse consommation

3.1 Introduction

Les travaux présentés dans le chapitre précédent avaient pour but de préparer ceux présentés dans ce chapitre. Les éléments nécessaires à l'automatisation du banc de test laser ont été développés dans un premier temps, puis une première campagne de test a été effectuée sur des mémoires SRAM. Celle-ci visait à comparer les effets de l'injection laser, avec des impulsions courtes de 30 ps, aux résultats obtenus précédemment par Roscian et al. dans [15] avec des impulsions plus longues de 50 ns. Bien que le modèle de faute obtenu soit identique (les fautes obtenues sont de type *bit-set*/reset et non de type *bit-flip*), le comportement de la cellule était différent. Là où des zones sensibles théoriques n'apparaissaient pas en pratique avec les impulsions de 50 ns, les impulsions de 30 ps permettaient d'y injecter des fautes. Afin de prendre en compte ces différences de comportement, une adaptation pour les impulsions courtes du modèle de simulation de Sarafianos et al. [54–56] a été proposée et validée. L'ensemble des résultats a ensuite été validé sur la RAM d'un microcontrôleur, où les mêmes effets ont été observés : modèle de faute *bit-set*/*bit-*

reset et apparition de zones sensibles absentes lors de l'injection avec des impulsions de 50 ns.

Cependant, l'amélioration de modèles de simulation pour les impulsions courtes n'est pas le but premier de ces travaux. Ce chapitre a donc pour but d'effectuer une première évaluation des effets de l'injection laser sur des éléments similaires à ceux testés dans le chapitre 1 lorsque leur tension d'alimentation est modifiée et lors de l'utilisation de FBB.

Dans cette optique, la première partie cible le microcontrôleur utilisé dans la dernière partie du chapitre 2. De la même manière que dans les tests précédents, la mémoire RAM est ciblée, mais la tension d'alimentation du microcontrôleur est abaissée de manière à déterminer si cela a un impact sur le comportement de ce dernier face à l'injection laser.

Les travaux présentés dans la seconde partie du chapitre ont été réalisés sur un autre microcontrôleur, plus récent, réalisé en technologie CMOS 90 nm. Celui-ci offre la possibilité de faire varier la tension délivrée par le régulateur de tension interne du circuit, ainsi que d'utiliser le FBB. Ils ont pour but d'évaluer plus précisément l'impact des ces techniques permettant d'ajuster le ratio entre la consommation et les performances du circuit sur sa sensibilité au laser, et qui sont celles que le projet MAGE vise à évaluer.

3.1.1 Effets attendus

Il est difficile le prévoir les effets d'une réduction de la tension d'alimentation sur la sensibilité au laser d'un circuit, de nombreux facteurs entrant en jeu et ayant des effets contraires.

Comme vu dans le chapitre 1, le photocourant I_{ph} généré par le tir laser peut être modélisé par l'équation 3.1 [54–56]. Celui-ci est proportionnel à la tension de polarisation inverse de la jonction PN ciblée V_r, or réduire V_{dd} revient à réduire cette tension, et par conséquent l'amplitude du photocourant généré, ce qui devrait donc réduire la sensibilité au laser du circuit.

$$I_{ph}(t) = [a(P).V_r + b(P)].A.\alpha_{topology}.\Omega_{shape}(w).Z_{focus} \tag{3.1}$$

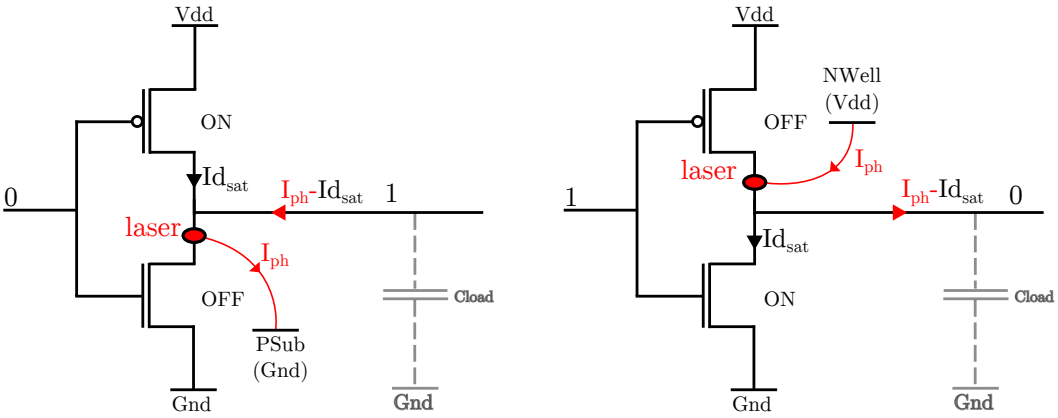


FIGURE 3.1 – Impact du courant de saturation sur l'injection d'une faute

Cependant, réduire V_{dd} réduit également le courant de saturation des transistors Id_{sat}, ce qui réduit le photocourant nécéssaire pour créer une faute. En prenant un inverseur pour exemple comme montré sur la figure 3.1, où I_{ph} est le photocourant créé par le laser entre le drain du transistor OFF et le substrat ou le NWell, dans les deux cas possibles (entrée de l'inverseur à 1 ou 0), le courant entraînant la charge ou la décharge de la capacité vaut $I_{ph} - Id_{sat}$. Une faute est donc injectée si I_{ph} est supérieur à Id_{sat}. Réduire Id_{sat} revient donc à faciliter l'injection d'une faute.

Ces deux phénomènes liés à la réduction de V_{dd} ont donc des effets opposés sur la sensibilité du circuit au laser. Pour cette raison, il est difficile de prévoir dans quelle mesure ces deux effets se compensent ou si l'un prédomine sur l'autre, entraînant une variation de sensibilité.

De la même manière, le FBB réduit la valeur de V_r (le NWell étant polarisé à une valeur inférieure à V_{dd} et le PWell à une valeur supérieure à Gnd), réduisant ainsi le photocourant généré. Cependant il a également pour effet de passer le circuit dans un mode de fonctionnement où l'effet substrat n'est plus négligeable. Cela a des conséquences multiples et difficiles à évaluer, notamment une réduction de la tension de seuil des transistors, ce qui pourrait faciliter les fautes.

Les effets de ces techniques sur la sensibilité au laser étant difficiles à prévoir par la théorie, ce chapitre vise donc à déterminer expérimentalement leur impact.

3.2 Injection de fautes sur la RAM d'un microcontrôleur sous-alimenté

Comme vu dans le chapitre 1, une des méthodes les plus immédiates pour réduire la consommation d'un circuit est de réduire sa tension d'alimentation. Pour évaluer ce paramètre, la première cible de test sera la RAM du microcontrôleur utilisé dans la dernière partie du chapitre précédent. Celle-ci a l'avantage de pouvoir être rapidement mise en oeuvre : les routines de test ont déja été développées durant les tests précédents et les effets du laser sur la RAM dans des conditions de tests normales sont déja connus. Cette section présente les effets obtenus lorsque sa tension d'alimentation est abaissée.

3.2.1 Configuration et processus de test

A la différence des tests du chapitre 2, où l'objectif était simplement d'observer le comportement de la mémoire face à l'injection laser, le but est ici de déterminer l'impact d'une tension d'alimentation réduite sur ce comportement. Dans cette optique, la carte de test a été modifiée pour utiliser non pas l'alimentation 5 V fournie par le lecteur smartcard servant à effectuer la communication, mais une alimentation externe, dont on fera varier la valeur.

Afin de rester dans des conditions de test connues, la configuration de test utilisée est la même que dans le chapitre 2 : une portion de la RAM du microcontrôleur est cartographiée à l'aide de la source Pico du banc de test, la durée d'impulsion est donc de 30 ps. L'énergie est réglée à 10 nJ ce qui se situe juste en dessous du seuil à partir duquel des dommages permanents sont infligés au circuit, et l'objectif 100X du banc de test est utilisé, ce qui donne une taille du spot laser d'environ 1 µm.

Comme précédemment, pour chaque position testée, la RAM est intégralement écrite à '1', puis relue pour détecter d'éventuels bits défectueux. Le tir laser est ensuite déclenché et la RAM est relue pour détecter les fautes de type *bit-set*. Puis ces opérations sont répétées en écrivant la RAM à '0' pour détecter les fautes de type *bit-reset*.

La carte couvre la même zone de 40 µm par 40 µm que dans le chapitre 2. Ce-

pendant, lors de ces tests, la table XYZ du banc a commencé à présenter des dérives sur son axe Z, ce qui entraînait une défocalisation du faisceau lors de campagnes de test de très longue durée. Pour cette raison, le pas est réduit à 1 µm, ce qui permet de réduire le temps nécéssaire pour tracer une carte à environ 2h (contre environ 50 h avec le pas de 0.2 µm précédemment utilisé).

3.2.2 Résultats

Une première carte est tracée dans les conditions de fonctionnement nominales du microcontrôleur (V_{dd}=5 V) afin d'obtenir une référence. La carte ainsi obtenue est présentée figure 3.2. Comme précédemment, les points rouges représentent les fautes monobit de type *bit-set* et les points bleus les fautes monobit de type *bit-reset*. Les points verts eux, représentent des positions pour lesquelles plus de 8 bits sont fautés (à chaque fois le même bit sur plusieurs octets) qui correspondent ainsi vraisemblablement à la logique de contrôle des lignes de sélection de bit utilisées pour lire et écrire dans la mémoire. Ces zones de corruption de la mémoire représentent environ un tiers des fautes obtenues (34%), le reste étant constitué de fautes monobit et d'un nombre extrêmement réduit de fautes sur deux bits (5 fautes sur la totalité de la carte, localisées à la limite entre deux cellules).

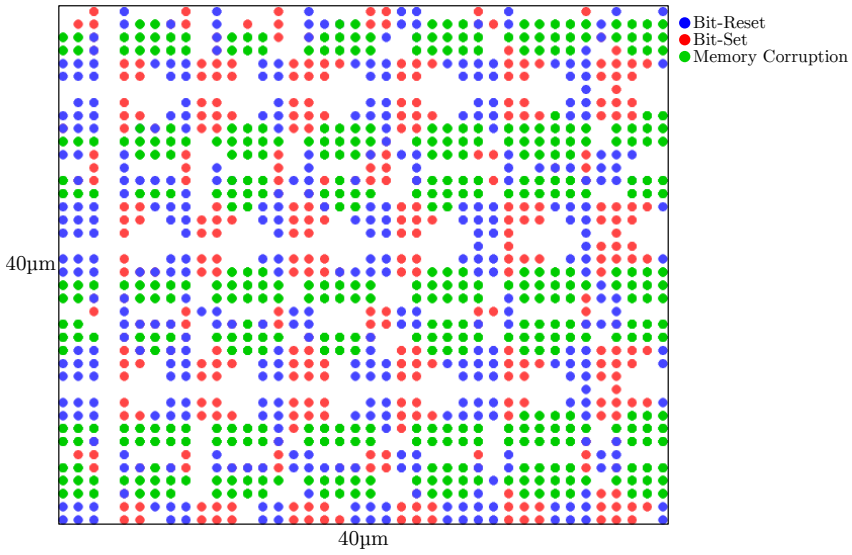


FIGURE 3.2 – Résultats d'injection de fautes sur une partie de la RAM du microcontrôleur avec V_{dd}=5 V

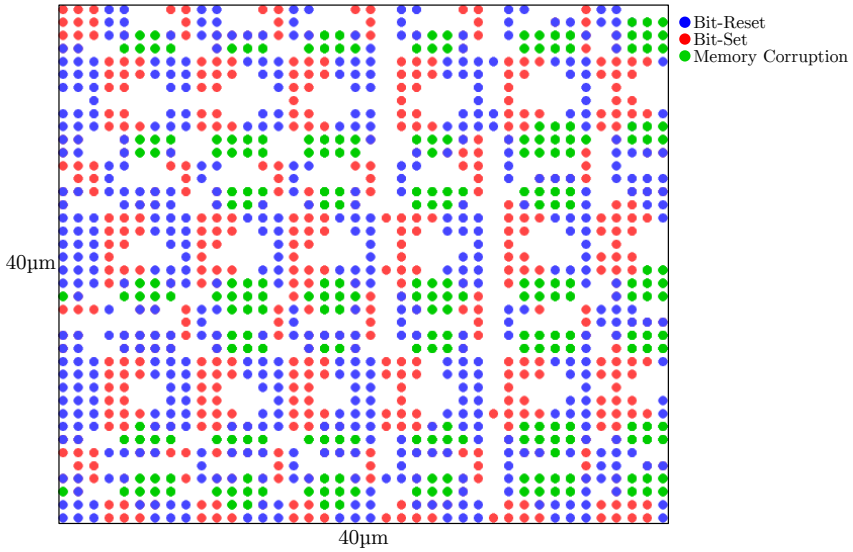


Figure 3.3 – Résultats d'injection de fautes sur une partie de la RAM du microcontrôleur avec V_{dd}=4.5 V

Afin de tester l'impact d'une tension d'alimentation réduite sur la sensibilité au laser du microcontrôleur, la cartographie suivante est effectuée en réduisant la tension d'alimentation du microntrôleur à V_{dd}=4.5 V, ce qui est à la limite de sa plage de fonctionnement nominale. Les résultats obtenus sont présentés figure 3.3. On peut constater que la taille des zones entraînant des corruptions de mémoire diminue (celles-ci ne représentent plus que 19% des fautes obtenues). En revanche la sensibilité des cellules SRAM elles-mêmes augmente : le nombre de fautes mono-bit obtenu sur la surface cartographiée augmente de 14%.

Enfin, une dernière cartographie est effectuée, lors de laquelle la tension est à nouveau abaissée à 4.0V, ce qui cette fois est en dehors de la plage de fonctionnement nominale du circuit, mais semble néanmoins suffisant pour assurer son fonctionnement. Les résultats obtenus sont présentés figure 3.4. Sur cette carte, les zones de corruption de mémoire disparaissent presque complètement (seul 5 points subsistent). Le nombre de fautes monobits injectées lui augmente à nouveau légèrement (8.5% de fautes supplémentaires).

Au vu des résultats obtenus dans les trois cartes précédentes, une baisse de tension d'alimentation permet d'augmenter la probabilité d'injecter une faute mono-bit sur la RAM de ce microcontrôleur lors d'une attaque laser. Cette sous alimentation

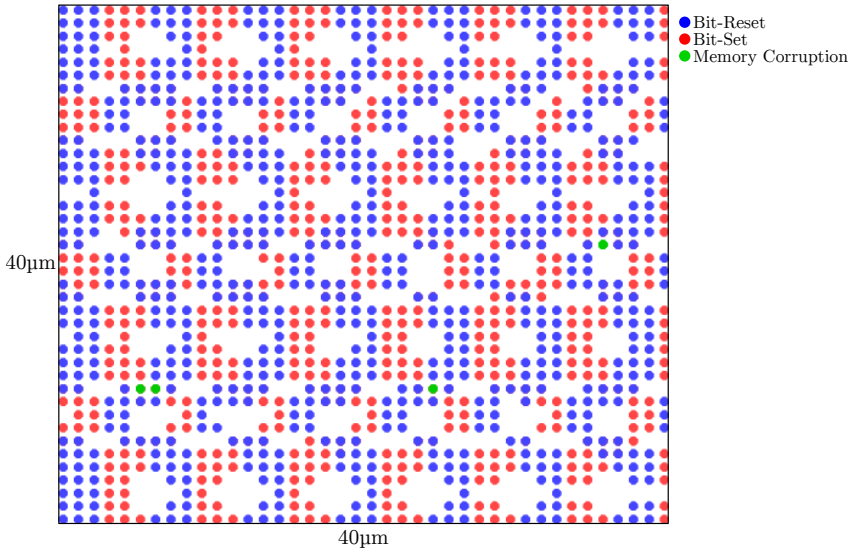


FIGURE 3.4 – Résultats d'injection de fautes sur une partie de la RAM du microcontrôleur avec V_{dd}=4 V

permet également de diminuer la probabilité d'injecter des fautes sur une portion plus importante de la RAM. Ces deux facteurs permettent à un attaquant d'avoir un meilleur contrôle sur les fautes injectées. Une attaque sur ce circuit effectuée en utilisant une tension d'alimentation de 4 V sera donc plus efficace qu'en utilisant une tension d'alimentation de 5 V.

Bien que ces résultats montrent que la tension d'alimentation du circuit a un impact sur sa sensibilité au laser, il est cependant difficile de déterminer avec certitude le phénomène à l'origine des fautes entraînant les corruptions de la mémoire sans connaître les détails de conception du microcontrôleur. En conséquence de quoi il est également difficile de savoir si l'augmentation de la surface sensible permettant de créer des fautes monobit est une conséquence d'une réduction de sensibilité des zones entraînant des corruptions mémoire ou si la sensibilité des cellules constituant la mémoire RAM augmente réellement lorsque la tension d'alimentation est abaissée.

Ces cartographies permettent donc de donner une première idée de ce à quoi il est possible de s'attendre, mais ne sont pas suffisantes. De plus, le microcontrôleur n'est pas une cible idéale : il est réalisé en technologie CMOS 350 nm, peu récente. Seule sa tension d'alimentation externe peut être ajustée : impossible de connaitre

l'impact réel sur la sortie du régulateurs de tension interne. Enfin il ne permet pas d'évaluer les effets des techniques de *body-biasing*.

En conséquence, afin d'obtenir des résultats plus précis et plus pertinents, la suite de ce chapitre se concentrera sur un microcontrôleur plus récent et conforme aux objectifs du projet, sur lequel il est possible d'ajuster la tension de sortie du régulateur interne et d'utiliser du FBB.

3.3 Injection de fautes sur un registre d'un accélérateur cryptographique matériel

Les tests effectués jusqu'à présent l'étaient sur des circuits réalisés dans des technologies CMOS proposant de larges dimensions (0.25 µm pour les points SRAM seuls, et 0.35 µm pour le microcontôleur.

La section précédente a permis de montrer que la tension d'alimentation a un impact sur la sensibilité au laser du circuit, mais le circuit utilisé ne permet pas d'obtenir des résultats suffisamment précis, ni d'évaluer l'impact du *body-biasing*. Pour pallier à cela, la suite des tests s'effectuera sur un micro contrôleur plus récent. Celui-ci est capable d'ajuster à la volée la tension du régulateur interne et d'utiliser du *body-biasing*, ce qui permet de modifier le ratio entre la consommation et les performances en fonction de son activité et ainsi de réduire sa consommation globale.

Le but de cette section sera ainsi d'observer l'impact de ces paramètres sur la sensibilité du microcontrôleur à l'injection laser.

3.3.1 Microcontrôleur cible

La cible des prochains tests est un microcontrôleur 32 bits fabriqué en technologie CMOS 90 nm. La majeure partie de sa logique utilise une architecture *triple-well* (détaillée dans le chapitre 1), ce qui permet d'utiliser du FBB avec V_{bb} allant de 0 à 400 mV. Il est également possible de choisir entre trois valeurs de tension d'alimentation du coeur (V_{dd}=1.13 V, V_{dd}=1.26 V et V_{dd}=1.32 V).

La figure 3.5 montre une vue reconstituée de la face avant du circuit obtenue avec la caméra infrarouge du banc de test laser (à travers la face arrière du substrat

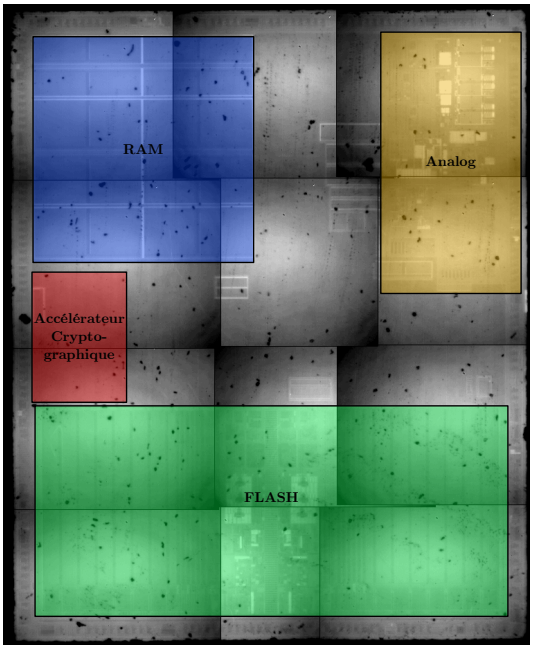


FIGURE 3.5 – Vue reconstituée du microcontrôleur cible via la caméra infrarouge du banc de test

donc). Les blocs mémoires sont encadrés, ainsi que la partie analogique (régulateurs de tension etc.) et l'accélérateur cryptographique qui sera la cible des tests à venir.

L'accélérateur cryptographique est inclus dans la partie *triple-well* de la logique. Il permet d'effectuer des opérations de chiffrement en utilisant AES, DES ou 3DES ainsi que le calcul des fonctions de hachage SHA1, SHA2, MD5 et HMAC.

3.3.1.1 Registres d'initialisation des modes de chiffrement chainés

La première cible au sein de l'accélérateur cryptographique sera un registre 128 bits utilisé pour stocker les vecteurs d'initialisations lors de chiffrement AES en mode chaîné.

Les modes de chiffrement chaînés permettent d'éviter la répétition de blocs de 128 bits lors du chiffrement de messages plus longs que la taille de bloc. Si l'on chiffre un message de plusieurs blocs "naïvement" (appellé mode ECB pour Electronic Code Book [70]) comme sur la figure 3.6 (chaque bloc est chiffré individuellement), des blocs identiques dans le message clair seront également identiques dans le message chiffré. Ainsi, ce mode de chiffrement sera déconseillé [70] car il peut permettre à un attaquant d'identifier des motifs dans les communications.

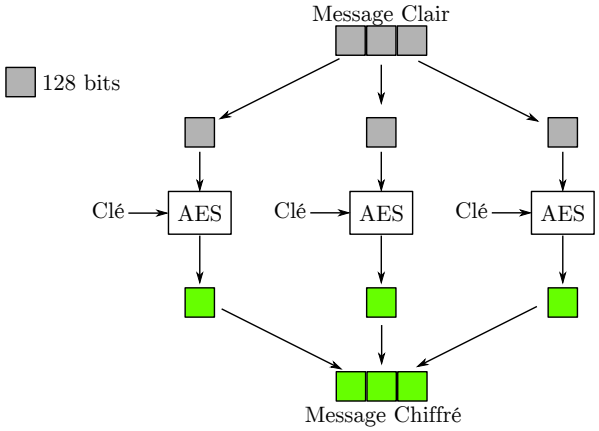


Figure 3.6 – Chiffrement AES en mode ECB

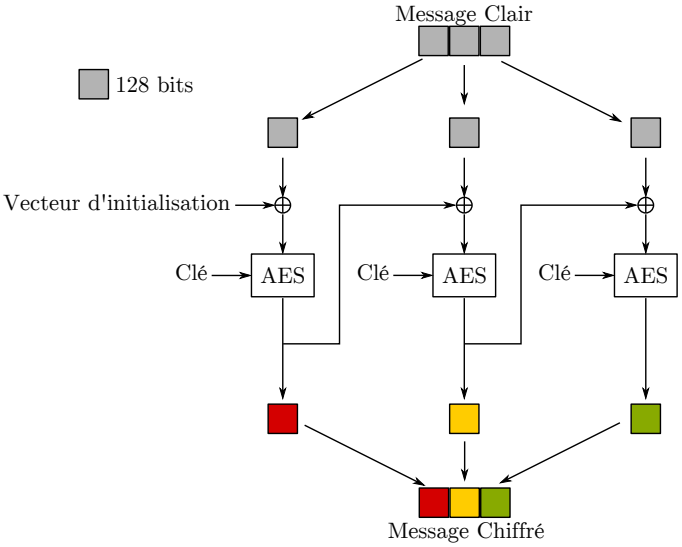


Figure 3.7 – Chiffrement AES en mode CBC

Afin d'éviter cela, des modes de chiffrement "chaînés" sont généralement utilisés. Dans ce cas, chaque bloc est mélangé avec le résultat du chiffrement du bloc précédent avant d'être chiffré à son tour, évitant ainsi l'apparition de motifs, puisque chaque chiffrement dépend des résultats de tous les blocs chiffrés avant lui. Le mode de chiffrement chaîné le plus simple appelé CBC pour Chipher Block Chaining [70], est montré en exemple figure 3.7 : avant chaque chiffrement, on effectue un XOR bit à bit entre le bloc à chiffrer et le chiffré du bloc précédent. L'utilisation de modes chainés requiert alors l'utilisation d'un vecteur d'initialisation pour le chiffrement du premier bloc [70].

Le registre utilisé pour stocker ce vecteur a l'avantage de faire partie de l'accélérateur

cryptographique, d'être situé dans la logique *triple-well* et d'être accessible en lecture et en écriture (contrairement aux autres registres de l'accélérateur cryptographique), ce qui en fait une cible de test idéale.

3.3.2 Injection sur registre avec V_{dd} variable

Les résultats des tests précédents nous donnent une indication sur la tendance que l'on peut attendre lorsqu'on abaisse la tension d'alimentation d'un circuit (une baisse de V_{dd} entraîne une augmentation de la sensibilité au laser). Cependant, la technologie de fabrication du microcontrôleur précédent est ancienne et la variation de tension d'alimentation se faisait sur l'alimentation externe du circuit et non sur le régulateur de tension interne.

Le but des prochains tests est d'effectuer une expérience similaire sur le microcontrôleur présenté précédemment en agissant directement sur la tension délivrée par le régulateur afin de préciser la tendance observée dans la section 3.2.

3.3.2.1 Configuration et processus de test

La configuration et le processus de test utilisés sont relativement similaires aux conditions précédentes.

Le boîtier du microcontrôleur est ouvert face arrière et le substrat aminci à 140 µm afin de permettre une attaque laser (comme vu dans le chapitre 1). En raison d'une panne de la source Picoseconde du banc, la source Nano est utilisée, ce qui donne les paramètres de test suivants :

- La durée d'impulsion est réglée à 50 ns, ce qui est le minimum possible avec cette source.
- La longueur d'onde est de 1064 nm.
- La puissance est réglée à 0.7W, ce qui est situé juste en dessous du seuil à partir duquel des dommages permanents sont infligés au circuit.
- L'objectif 100X est utilisé afin d'avoir le plus petit spot possible, soit environ 1 µm.

Le processus de test est identique à celui employé précédemment. A savoir, pour chaque position à tester, tous les bits du registre sont mis à 1 puis lu afin de vérifier qu'il n'a pas été endommagé de manière définitive. Le tir laser est ensuite déclenché et la valeur du registre est relue. Ceci est ensuite répété en mettant le registre à 0.

La cartographie est effectuée avec un pas de 1 µm sur une zone centrée autour du registre. Enfin, comme le microcontrôleur est capable de basculer d'une valeur de V_{dd} à la volée, il est possible d'effectuer le test avec les trois valeurs de V_{dd} possibles (1.32 V, 1.26 V et 1.13 V) en une seule campagne de test. Ce qui permet d'atténuer les variations liées aux conditions extérieures (dérives de la table XYZ, température etc.) dans les résultats. En effet, à présent, les trois valeurs de tension sont traitées successivement pour chaque position, plutôt que de réaliser une cartographie avec chaque valeur comme dans la section précédente, ce qui avait pour effet d'espacer de plusieurs heures les tests sur une même position.

3.3.2.2 Résultats

Les résultats obtenus sont présentés dans les cartes des figures 3.8 (pour V_{dd}=1.32 V), 3.9 (pour V_{dd}=1.26 V) et 3.10 (pour V_{dd}=1.13 V). Pour chaque carte, la couleur des points représente le nombre de bits maximum fautés pour chaque position (en *bit-set* ou *bit-reset*). Les points verts sont ainsi les positions où des fautes monobits sont injectées et les autres couleurs représentent les fautes multi-bits.

Comme le nombre points par carte est fixe, il est possible de mesurer la surface sensible totale en nombre de fautes. La table 3.1 récapitule la surface sensible en nombre de fautes pour chaque valeur de V_{dd}. Il est ainsi possible de voir que l'augmentation de la surface sensible, lorsque V_{dd} est abaissé de 1.32 V à 1.13 V, est environ de 30%.

V_{dd}	1.13 V	1.26 V	1.32 V
Fautes	1227	998	902

Table 3.1 – Fautes par carte pour chaque valeur de V_{dd}

Ces cartes permettent donc de confirmer la tendance qui se dégageait dans les tests de la section précédente : une baisse de tension d'alimentation entraîne une augmentation de la surface sensible au laser du registre. Cette augmentation de

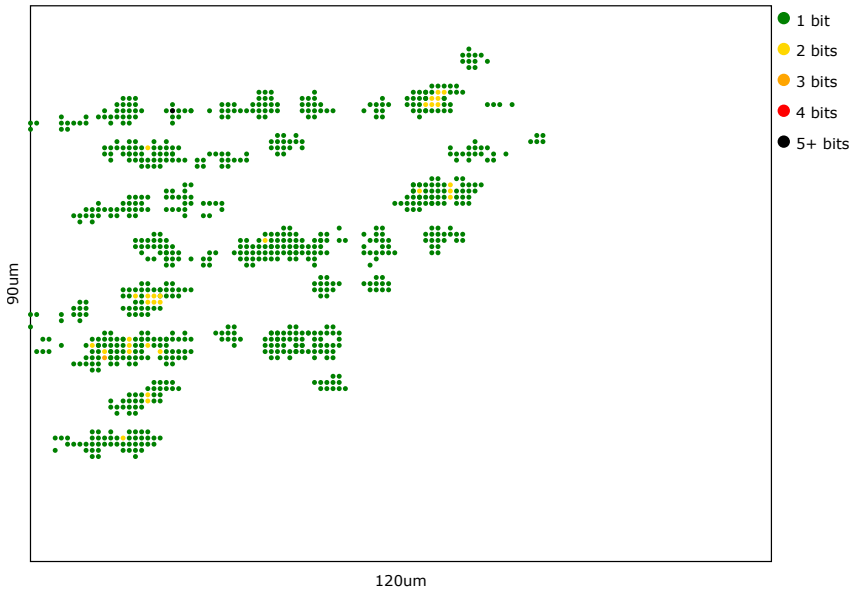


FIGURE 3.8 – Résultats d'injection de faute sur le registre pour V_{dd}=1.32 V

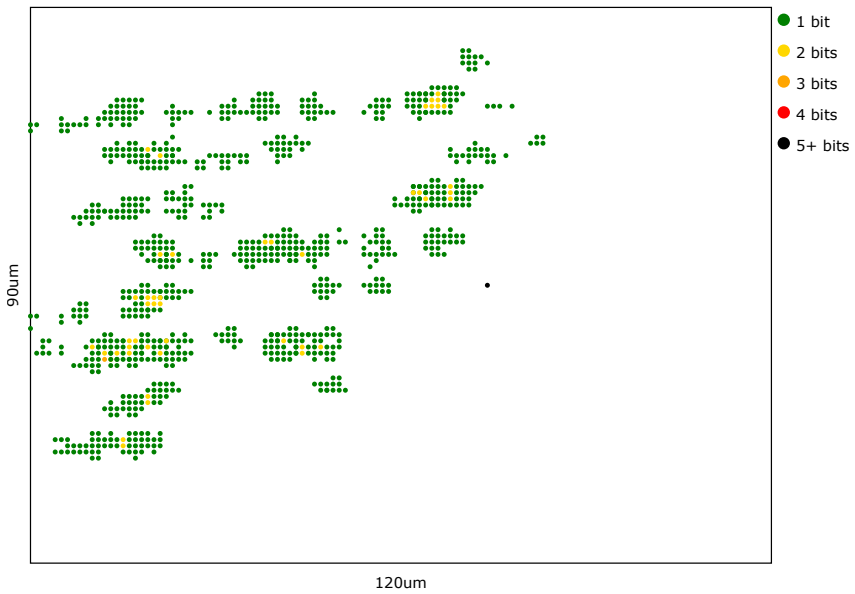


FIGURE 3.9 – Résultats d'injection de faute sur le registre pour V_{dd}=1.26 V

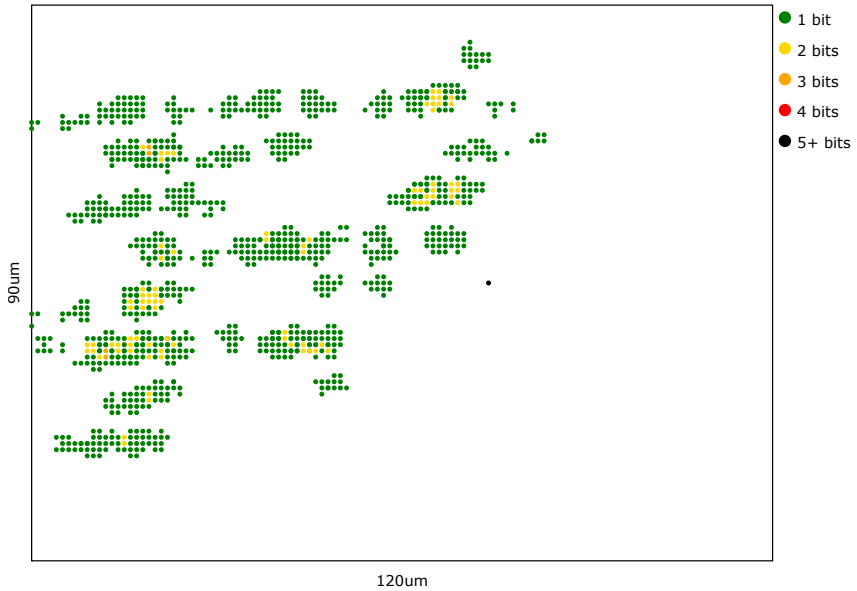


Figure 3.10 – Résultats d'injection de faute sur le registre pour V_{dd}=1.13 V

sensibilité est non-négligeable, cependant, il est possible d'obtenir des variations de surface sensible identiques simplement en ajustant la puissance du laser. Comme vu dans le chapitre 1, l'intensité du photocourant généré diminue avec la distance entre le faisceau et la jonction PN ciblée et augmente avec la puissance du laser. Ainsi, en augmentant la puissance, le seuil de distance pour obtenir un photocourant suffisant pour injecter une faute augmente également, ce qui a pour effet d'augmenter la surface sensible. A première vue, ces variations de sensibilité ne devrait donc pas entraîner de risques sécuritaires nouveaux.

Dans la prochaine section, le but est donc d'observer les effets de l'injection laser lorsque le FBB est ajouté aux variations de V_{dd}.

3.3.3 Injection sur registre avec V_{dd} et V_{bb} variables

Une fois l'effet d'une variation de V_{dd} est connu, la prochaine étape est de caractériser les variations de sensibilité liées à l'activation du FBB . Le microcontrôleur utilisé est capable d'activer un mode "overdrive" qui correspond à du FBB avec V_{bb}=400 V sur la quasi-totalité de la logique (dont l'accélérateur cryptographique). Les prochains tests viseront à coupler ce mode aux variations de V_{dd} des tests précédents, ce qui correspond à la technique de réduction de la consommation que l'on cherche à

étudier : en jouant sur ces deux paramètres, il possible d'ajuster dynamiquement le ratio performances/consommation en fonction de l'activité du circuit et ainsi de réduire sa consommation globale.

3.3.3.1 Configuration et processus de test

La configuration de test est la même que pour les tests précédents à la taille de la zone cartographiée près. L'ajout d'une variable (l'activation ou non du FBB) augmente le temps de cartographie (une cartographie dans les mêmes conditions que les tests précédents prendrait environ 10 heures), ce qui en soit ne serait pas problématique. Cependant la table XYZ du banc test présente une dérive lorsque les cartographies durent trop longtemps, ce qui force à limiter le temps des cartographies aux alentours de 5 heures. La zone de test est donc réduite à une zone de 25 µm par 25 µm avec un pas de 0.5 µm située vers le centre de la surface du registre.

Pour chaque position, le registre est testé avec chaque paire de (V_{dd},V_{bb}) possible, V_{dd} ayant les mêmes valeurs que précédemment, et V_{bb} valant soit 400 mV, soit 0 mV.

Enfin, pour minimiser les effets de la dérive de la table XYZ du banc de test, toutes les configurations sont testées sur une position avant de passer à la position suivante.

3.3.3.2 Résultats

Les cartes obtenues sont présentées dans les figures 3.11 à 3.16. Encore une fois, les points verts représentent les positions sur lesquelles des fautes mono-bit sont injectées, et les autres couleurs des fautes multi-bits.

En comparant deux à deux les cartes correspondant à des valeurs de V_{dd} identiques, il est possible de constater une augmentation de la sensibilité au laser du registre lorsque le FBB est activé.

Par exemple sur la figure 3.12 (V_{dd}=1.32 V et V_{bb}=400 mV) une importante zone de fautes sur deux bits, en bas à droite de la carte, (qui correspond à l'augmentation de la surface de recouvrement entre deux zones sensibles) est plus réduite sur la figure 3.11 (V_{dd}=1.32 V et V_{bb}=0 mV).

Sur les figures 3.14 (V_{dd}=1.26 V et V_{bb}=400 mV) et 3.13 (V_{dd}=1.26 V et V_{bb}=0 mV)

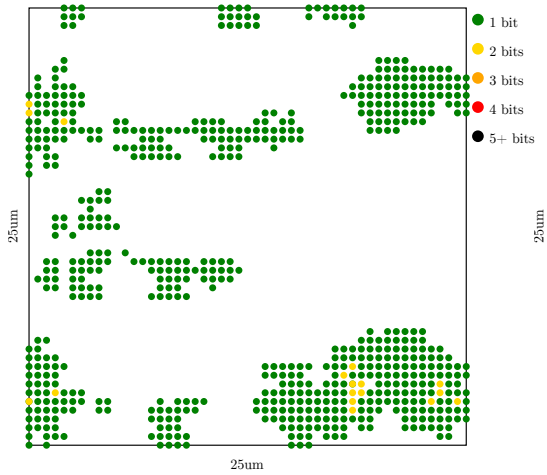


Figure 3.11 – Résultats pour V_{dd}=1.32 V et V_{bb}=0 mV

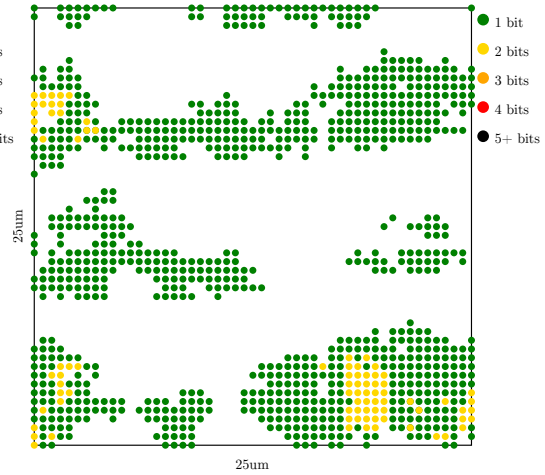


Figure 3.12 – Résultats pour V_{dd}=1.32 V et V_{bb}=400 mV

Figure 3.13 – Résultats pour V_{dd}=1.26 V et V_{bb}=0 mV

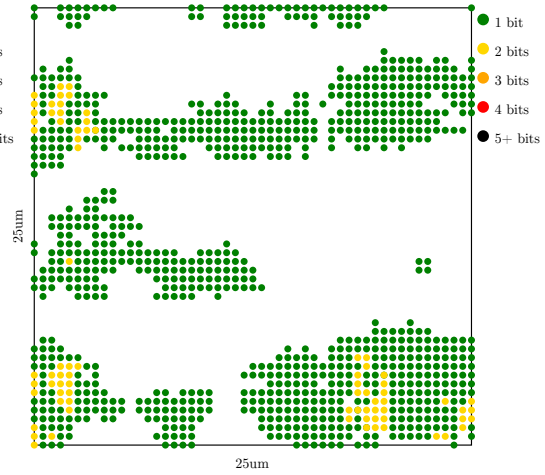


Figure 3.14 – Résultats pour V_{dd}=1.26 V et V_{bb}=400 mV

les zones sensibles situées vers le centre de la carte sont plus denses lorsque le FBB est activé. Idem sur les figures 3.16 (V_{dd}=1.13 V et V_{bb}=400 mV) et 3.15 (V_{dd}=1.13 V et V_{bb}=0 mV).

En plus de ces variations de sensibilité au laser liées à l'activation de FBB, les variations liées aux changements de tension d'alimentation, observées dans la section 3.3.2, sont toujours présentes : les zones sensibles sont plus grandes lorsque V_{dd} est diminué.

Ces variations sont récapitulées dans l'histogramme de la figure 3.17 : comme pour les cartes de la section 3.3.2, il est possible de représenter la somme des surfaces

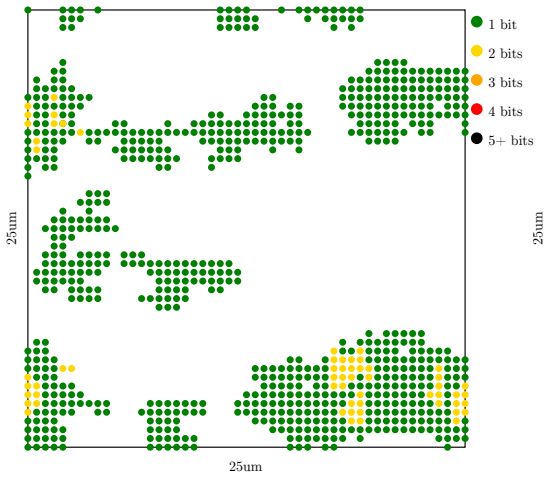


FIGURE 3.15 – Résultats pour V_{dd}=1.13 V et V_{bb}=0 mV

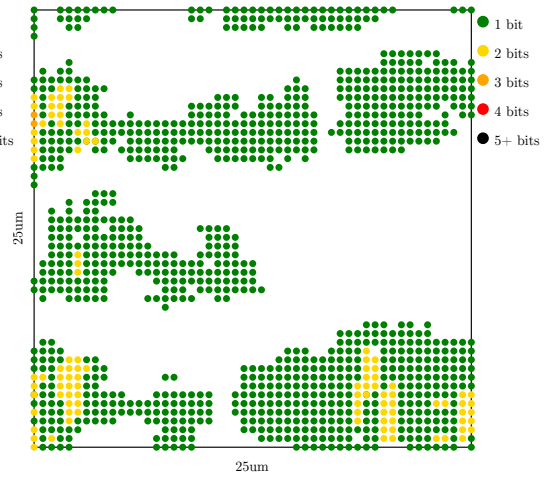


FIGURE 3.16 – Résultats pour V_{dd}=1.13 V et V_{bb}=400 mV

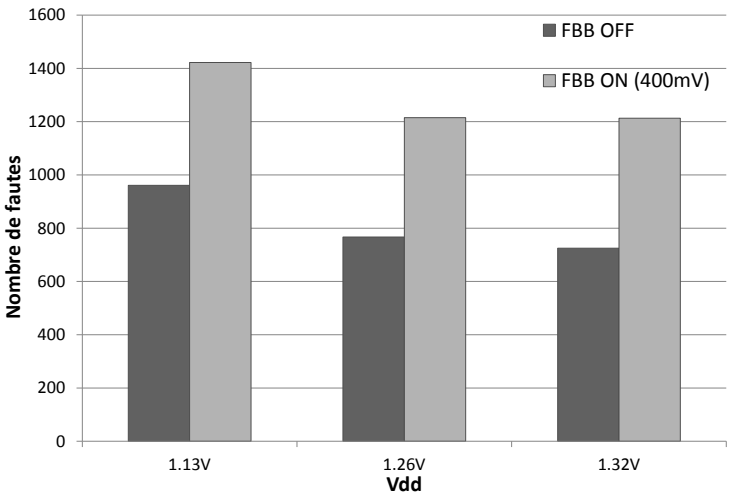


FIGURE 3.17 – Récapitulatif de la sensibilité pour chaque configuration

sensibles de la zone testée par le nombre de fautes obtenues pour chaque carte. Le graphique donne ce nombre de fautes pour chaque valeur de V_{dd} avec et sans FBB.

Comme observé sur les cartes, l'activation du FBB se traduit par une augmentation de la sensibilité au laser du registre. Couplé avec les variations de sensibilité liées à V_{dd} observées dans la section 3.3.2, cela augmente la plage de sensibilité au laser du registre. En faisant varier uniquement V_{dd}, l'écart de surface sensible entre la configuration la moins sensible et la configuration la plus sensible était de 30%, en utilisant le FBB, l'écart de surface sensible varie quasiment du simple au double. La figure 3.18 permet de visualiser l'augmentation de sensibilité. Elle montre les po-

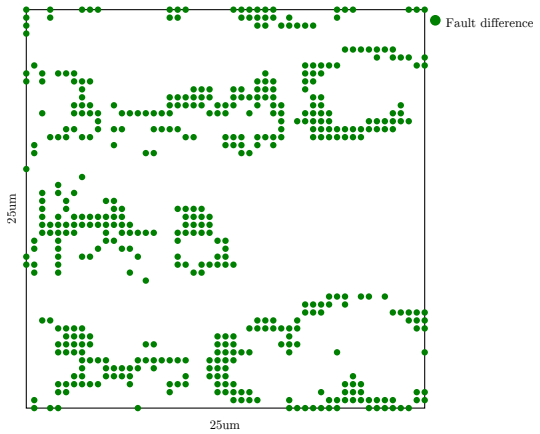


Figure 3.18 – Augmentation de sensibilité entre (V_{dd}=1.32 V, V_{bb}=0 mV) et (V_{dd}=1.13 V,V_{bb}=400 mV).

sitions pour lesquelles des fautes apparaissent dans la configuration la plus sensible et pas dans la configuration la moins sensible.

L'augmentation de la sensibilité au laser du registre est donc plus importante en activant le FBB que lorsque la seule variable était V_{dd}. Cependant, il est toujours possible d'obtenir des variations de surface sensible similaires en faisant simplement varier la puissance du laser. L'utilisation de FBB ne devrait donc pas non plus ouvrir de nouvelles failles de sécurité à première vue.

3.3.4 Seuil d'apparition des fautes

Bien que la surface sensible soit une métrique facilement mesurable pour évaluer les variations de sensibilité au laser, une autre métrique intéressante d'un point de vue sécuritaire est le seuil de puissance laser requise pour injecter une faute. En effet, réussir à injecter des fautes en utilisant une puissance réduite peut permettre de contourner des contre-mesures de détection de l'injection de faute.

3.3.4.1 Configuration de test

Quelques cartographies rapides (et donc peu précises) sont d'abord effectuées avec des puissances laser entre 0 W et 0.7 W (la valeur utilisée pour les tests précédents) afin d'obtenir une estimation grossière du seuil de puissance laser nécéssaire pour injecter une faute. Ces tests préliminaires permettent de définir une plage de puis-

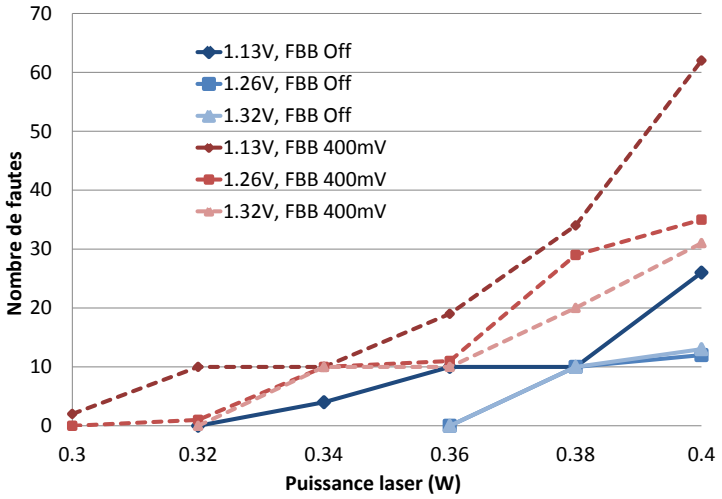


FIGURE 3.19 – Sensibilité pour chaque configuration en fonction de la puissance laser

sance allant de 0.20 W à 0.40 W dans laquelle le seuil recherché est compris.

Des cartographies plus précises peuvent alors être effectuées pour déterminer la valeur du seuil. Pour chaque position de la carte, chaque paire de (V_{dd},V_{bb}) est testée en balayant la plage de puissance de 0.20 W à 0.40 W avec un pas de 0.02 W.

Encore une fois, le nombre de configurations à tester augmente la durée totale des tests, ce qui force à réduire la résolution de la carte. La zone de test est à présent un carré de 10 µm par 10 µm quadrillé avec un pas de 1 µm. Enfin, pour chaque position et configuration, 10 tirs laser sont effectués au cas où les chances d'injecter une faute lorsque la puissance laser est proche du seuil soient inférieures à 100%.

3.3.4.2 Résultats

Pour chaque configuration testée, le graphique 3.19 donne le nombre de fautes obtenues en fonction de la puissance laser. Ces résultats présentent les mêmes tendances que celles des résultats des sections 3.3.3 et 3.3.2 : une augmentation de la sensibilité au laser liée à une baisse de tension d'alimentation, couplée à une nette augmentation de sensibilité lorsque le FBB est activé.

La table 3.2 quant à elle, donne les valeurs de puissance laser à partir desquelles les fautes commencent à apparaître pour chaque configuration. Les écarts de puis-

sance sont faibles, ce qui rend les résultats peu précis, mais les mêmes tendances que précédemment continuent à apparaître.

V_{dd}	FBB OFF	FBB ON (400mV)
1.13 V	0.34 W	0.30 W
1.26 V	0.38 W	0.32 W
1.32 V	0.38 W	0.34 W

Table 3.2 – Seuil de puissance laser requise pour injecter une faute dans le registre

Les variations obtenues sont encore une fois peu importantes par rapport aux capacités de la source laser (la source Nano peut déliver une puissance allant de 0 W à 3 W). Cependant, elles pourraient tout de même représenter un risque de sécurité dans le cas ou elles permettraient de passer sous le seuil de détection d'un capteur.

3.3.5 Mode veille

Outre les possibilités de variations de tension d'alimentation et de FBB, le microcontrôleur permet également d'effectuer du clock-gating. Il est ainsi possible de désactiver l'arbre d'horloge de l'accélérateur cryptographique lorsque celui ci n'est pas utilisé. De plus, un mode de veille où l'intégralité des horloges du circuit sont arrêtées est également disponible.

Une cartographie du registre a été effecutée dans ce mode de veille, en suivant le même protocole que pour les tests de la section 3.3.3, en passant le microcontrôleur en mode veille avant de tirer. Aucune variation de sensibilité notable par rapport au fonctionnement normal du circuit n'a été notée observée.

Cependant, d'un point de vue sécuritaire, la possibilité d'injecter des fautes lorsque le circuit est en mode veille est préoccupant. En effet, le facteur limitant de l'injection laser sur les mémoires est le temps nécessaire pour déplacer le faisceau : lorsque le circuit fonctionne, le temps nécéssaire pour déplacer le faisceau entre deux positions est trop important pour viser plusieurs bits sur la durée d'un calcul. Un attaquant capable de passer le circuit en mode veille s'affranchit de cette restriction et peut alors prendre le temps d'injecter des fautes sur plusieurs positions différentes avant de réactiver le circuit, ce qui ouvre la porte à de nombreuses attaques.

Par exemple, en effectuant des cartographies des différents éléments mémoires (RAM ou registres) du microcontrôleur, et en notant l'adresse des bits fautés pour chaque position, il est possible de construire (au moins partiellement pour des contraintes de temps) une table liant les adresses mémoire aux positions à cibler pour y injecter des fautes. Ainsi, en mettant en pause le fonctionnement du circuit via le mode veille, il est possible pour un attaquant d'écrire les valeurs qu'il souhaite dans des éléments mémoires précis avant de réactiver le circuit.

3.4 Conclusion

Dans ce chapitre, les techniques permettant d'ajuster le ratio performances/consommation d'un circuit en jouant sur la tension d'alimentation et la polarisation des caissons sont évaluées face à l'injection laser.

La première partie du chapitre présente des résultats obtenus sur le microcontrôleur en technologie CMOS 350 nm utilisé dans le chapitre 2. Afin de déterminer l'impact de la tension d'alimentation sur la sensibilité au laser du circuit, des cartes de la RAM de ce dernier on été tracées alors que celui ci est alimenté avec différentes valeurs de tension (5V, 4.5V et 4V). La tendance observée est que la sensibilité au laser des cellules SRAM augmente un peu lorsque V_{dd} diminue. En revanche, d'autres zones sensibles induisant des fautes 8 bits ou plus, elles, diminuent en sensibilité.

Afin d'obtenir des résultats plus précis et de pouvoir évaluer l'effet du *body-biasing* sur l'injection laser, la deuxième partie présente des résultats obtenus sur une nouvelle cible de test : un microcontrôleur en technologie CMOS 90 nm ayant la capacité de basculer dynamiquement entre plusieurs valeurs de V_{dd} et d'activer ou non le FBB. Des cartographies sont effectuées sur un registre situé dans l'accélerateur cryptographique du microcontrôleur afin d'évaluer l'impact de ces paramètres. Les résultats permettent de montrer qu'une diminution de V_{dd} entraîne un augmentation de la sensibilité du registre, et que l'activation du FBB a le même effet. L'augmentation de sensibilité se traduit par une baisse du seuil de puissance laser requise pour injecter, et par conséquent une augmentation de la surface sensible du registre.

Dans tout les cas, il est possible d'obtenir des effets similaires à ceux observés

en modifiant simplement la puissance du laser (une augmentation de la puissance laser augmente également la surface sensible). Ainsi, ces variations de sensibilité ne devraient pas ouvrir la porte à des vulnérabilités majeures en termes de sécurité.

En revanche, ces variations pourraient amplifier des faiblesses déja existantes en impactant les contre-mesures destinées à détecter l'injection laser. La sensibilité accrue pouvant peut être permettre d'injecter des fautes en dehors des zones de détection des capteurs ou d'utiliser une puissance en dessous de leur seuil de détection.

Le chapitre suivant se focalisera sur la manière dont ces techniques de réduction de la consommation peuvent être détournées de leur usage initialement prévu pour améliorer la sécurité du circuit en utilisant les variations de sensibilité qu'elles entraînent.

CHAPITRE 4

Etude d'un AES matériel

4.1 Introduction

Jusqu'à présent, les tests effectués dans les deux chapitres précédents l'ont été sur des éléments mémoire statiques (SRAM et registres). Cependant, ces cibles sont limitées dans l'impact que les techniques de réduction de la consommations peuvent avoir sur le plan sécuritaire. Comme vu dans le chapitre 3, les effets observés sont des variations de sensibilité d'un ordre largement inférieur à ce que les sources laser sont capables de délivrer. Ainsi, les seules vulnérabilités potentiellement introduites seraient les cas où le seuil de sensibilité réduit permettrait de passer sous le seuil de détection d'un capteur. Le risque sécuritaire introduit par les méthodes de réduction de la consommation basées sur le FBB sur des éléments mémoire est donc à première vue relativement faible.

Les conclusions du chapitre précédent permettent d'effectuer des premières observations sur l'impact de ces technologies basse-consommation lorsqu'elles sont utilisées en suivant leur fonctionnement normal. Dans ce chapitre, les prochains tests se focaliseront sur un algorithme AES implanté sur le microcontrôleur 90 nm utilisé dans le chapitre 3, afin de voir si les capacités du microcontrôleur, initialement prévues pour réduire sa consommation, peuvent être détournées de leur usage initial pour améliorer la résistance du circuit face aux attaques laser. L'objectif sera de déterminer si les variations de sensibilité au laser observées sur le registre qui

sont des conséquences des changements de configuration d'alimention peuvent être utilisées pour durcir le circuit.

Ce chapitre se découpe en deux parties : la première présente une étude sur l'AES matériel. Cette étude a pour but de repérer l'AES sur le circuit, d'ajuster les paramètres de l'injection laser à utiliser et enfin d'identifier des vulnérabilités potentielles du circuit.

La seconde partie présente une contre-mesure utilisant les variations de sensibilité observées dans le chapitre 3 pour rendre les fautes injectées dans le circuit moins prévisibles en basculant d'une configuration à une autre entre deux chiffrements. La contre-mesure est ensuite appliquée à l'AES couplée à une redondance temporelle en utilisant deux valeurs différentes d'alimentation pour les deux calculs de la redondance. Des tests d'injection laser sont alors effectués sur l'AES durci de cette manière et les résultats obtenus sont présentés.

4.2 Campagne de tests préliminaires

L'AES est implémenté de manière matérielle sur le microcontrôleur (directement sous forme de portes logiques, par opposition à une implémentation logicielle sous la forme d'un programme exécuté par le coeur du microcontrôleur). Pour pouvoir effectuer des tests ciblés sur l'AES, il est donc nécéssaire de connaître la position physique de son implémentation sur la puce.

De plus, contrairement aux tests précédents, effectués sur des éléments mémoire statiques, l'aspect temporel entre à présent en jeu : il est nécéssaire de cibler précisément l'instant du calcul, voire une ronde précise, pour obtenir des fautes exploitables par un attaquant.

La première partie des travaux à effectuer consiste donc en une campagne de tests préliminaires pour repérer l'AES spatialement et temporellement afin de pouvoir mener des attaques plus précises par la suite.

4.2.1 Repérage de l'AES

4.2.1.1 Repérage spatial

La position approximative de l'accélérateur cryptographique incluant l'AES ainsi que le registre ciblé dans le chapitre 3 est connue. Cependant, l'accélérateur cryptographie inclut de nombreuses fonctionnalités, et la position exacte de l'AES n'est pas connue. En conséquence, la première étape consiste à effectuer une cartographie large de la zone supposée de l'accélérateur cryptographique afin de repérer l'emplacement des portes logiques constituant l'AES.

Pour synchroniser le tir avec le calcul, le microcontrôleur envoie directement le signal de tir à la source laser. Précédemment, lors des tests sur mémoires et registres, l'ordre de tir était envoyé directement par le PC de contrôle, ce qui impliquait un délai variable de quelques millisecondes, ce qui est inutilisable dans le cas d'attaques sur l'AES, le chiffrement ne durant que quelques centaines de nanosecondes. Le signal envoyé directement par le circuit permet de réduire ce délai. Le signal de déclenchement est réglable à quelques dizaines de nanosecondes avec une variation aléatoire (jitter) de quelques nanosecondes.

Le schéma figure 4.1 représente les différents instruments utilisés et leurs liaisons : un PC contrôle la table XYZ et la configuration de la source laser. Pour chaque position de la carte le protocole est le suivant :

- Le PC de contrôle envoie un ordre de chiffrement au microcontrôleur.
- Le microcontrôleur délivre alors un signal de synchronisation directement à la source laser.
- Le signal de synchronisation passe à l'état haut avant le début du calcul, et passe à l'état bas une fois celui-ci terminé.
- Une fois le calcul terminé, le microcontrôlleur renvoie le chiffré au PC de contrôle via la liaison série.

La clé et le message clair utilisés pour le chiffrement sont fixes afin de simplifier le traitement des résultats et la programmation du microcontrôleur.

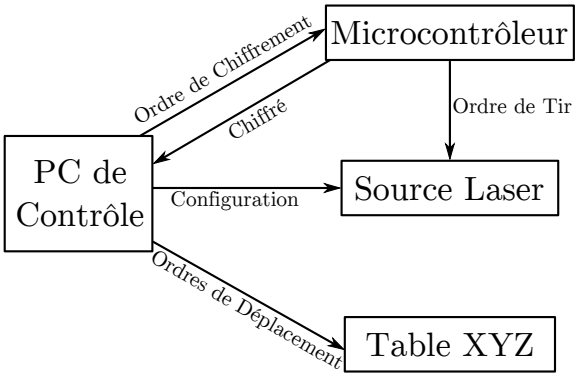


Figure 4.1 – Liaisons entre les différents appareils.

Sur le plan temporel, le calcul lui même dure environ 200 ns (la fréquence d'horloge de l'accélérateur cryptographique est de 50 MHz, soit une période de 20 ns et une ronde d'AES est calculée par cycle d'horloge). Cependant, le signal de synchronisation est plus long que la durée du calcul. On utilise par conséquent, dans un premier temps, une durée d'impulsion longue pour être certain d'affecter le circuit pendant au moins une partie de la durée du chiffrement. La source Nanoseconde du banc de test est utilisée.

Un repère visuel est choisi via la caméra infrarouge du banc pour s'assurer de la répétabilité des cartes entre plusieurs puces dans l'éventualité où l'une serait endommagée lors des tests (par exemple par une puissance laser trop importante). Le repère est montré en rouge sur la figure 4.2.

Une cartographie de la zone supposée de l'accélérateur cryptographique peut alors être effectuée. La durée d'impulsion choisie est de 300 ns. La zone à couvrir étant grande (1 mm par 1 mm) le pas entre deux positions testées est de 10 µm afin de garder un temps de cartographie raisonnable. En conséquence de ce pas de cartographie assez large, l'objectif 5X du banc de test est utilisé afin d'obtenir un spot de 20 µm de diamètre et ainsi avoir une couverture complète de la surface cartographiée.

Le résultat obtenu est présenté figure 4.3. Chaque point représente une position pour laquelle une faute a été injectée dans le calcul, et la couleur du point représente le nombre de bits fautés sur le chiffré obtenu en sortie. On peut donc raisonnablement supposer que l'AES se trouve dans la zone fautée.

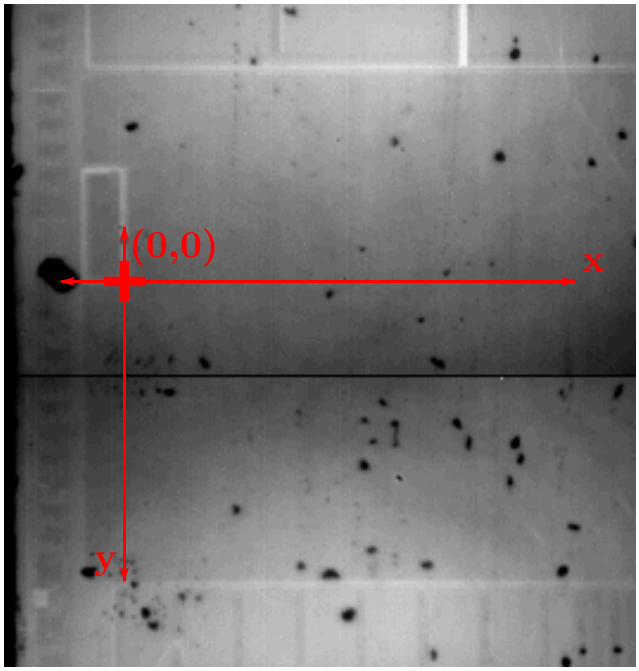


FIGURE 4.2 – Origine du repère des cartes (gros plan sur la zone de l'accélérateur cryptographique, via la caméra infrarouge du banc de test)

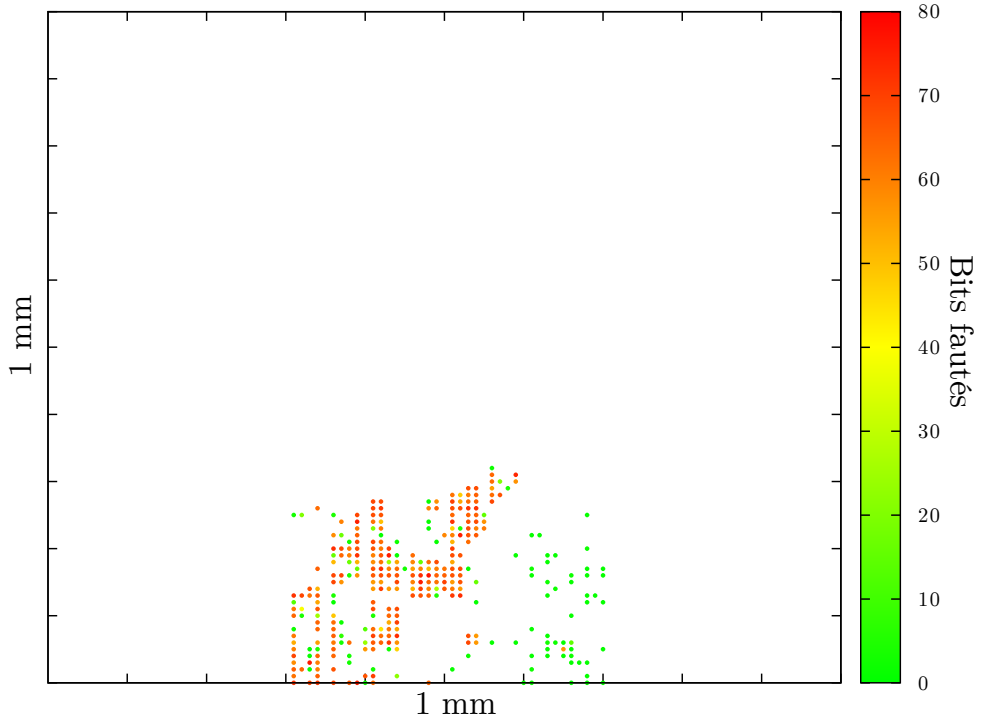


FIGURE 4.3 – Première carte de repérage de l'AES

La zone de test est alors resserrée sur la partie de la carte précédente où des fautes ont été enregistrées. On réduit également le pas de cartographie à 5 µm et la taille du spot à 5 µm de diamètre. La carte présentée par la figure 4.4 est alors obtenue.

La couleur des points dépend toujours du nombre de bits fautés sur le chiffré de sortie. La présence d'une zone située à droite de la carte, dans laquelle la quasi-totalité des fautes obtenues sont des fautes mono-bit, peut être notée. Le fait que les fautes soient des fautes d'un seul bit sur le chiffré de sortie indique que celles-ci ont lieu au plus tard juste après l'opération SubBytes de la dernière ronde du calcul. Dans le cas contraire, celles-ci seraient propagées sur plusieurs bits par l'opération en question. L'hypothèse la plus probable, étant donné la longueur de l'impulsion utilisée comparée à la durée du calcul, est que cette zone corresponde aux registres contenant le résultat final du calcul et que les fautes soient injectées après ce dernier. Cela n'a donc que peu d'intérêt du point de vue d'un attaquant, ces fautes n'étant pas exploitables pour attaquer l'algorithme.

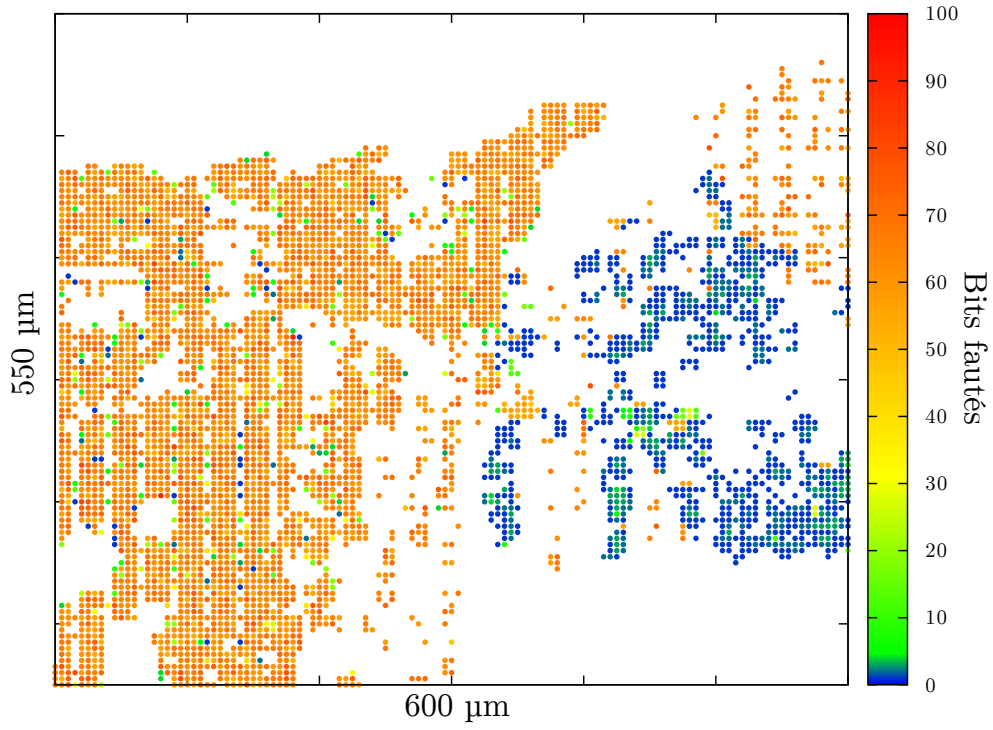


Figure 4.4 – Seconde carte de repérage de l'AES

Ces cartes permettent de connaître précisément la position de l'AES sur le circuit. Pour pouvoir mettre en place une attaque, il reste donc à repérer le calcul dans le temps afin de pouvoir ajuster la durée d'impulsion et l'instant visé pour obtenir des fautes exploitables.

4.2.1.2 Repérage temporel

A présent que la position spatiale de l'AES est connue, l'étape suivante est d'affiner la précision temporelle de l'attaque. Le but est d'utiliser la durée d'impulsion minimum de la source laser (50 ns dans le cas de la source Nano), et de viser la ronde 9 du calcul pour se placer, par exemple, dans le cas de l'attaque de Piret et al. [34] décrite dans le chapitre 1.

Pour déterminer l'instant où la faute est injectée dans le calcul, la méthode utilisée est la suivante : à partir d'un chiffré fauté, on effectue une opération de déchiffrement de l'AES en utilisant la clé correcte. Comme vu dans le chapitre 1, l'opération MixColumns de l'AES propage les fautes sur un ou plusieurs octets d'une même colonne sur toute la colonne du *state*, tandis que l'opération ShiftRows déplace les octets fautés. Les autres opérations en revanche ne propagent pas les fautes entre plusieurs octets. Comme montré dans la figure 4.5, cela a pour effet de propager une faute sur un octet à l'intégralité du *state* en deux rondes. De la même manière, lors du déchiffrement, le nombre d'octets fautés diminue jusqu'au point où la faute a été injectée. En revanche, à partir de ce point, comme la faute injectée ne disparaît pas au moment où elle était apparue durant l'opération de chiffrement, celle-ci se propage à nouveau dans la direction opposée jusqu'à contaminer l'intégralité du *state* deux rondes plus tard. Ainsi en comparant à chaque étape intermédiaire, la valeur du *state* avec celle qui serait obtenue dans le cas d'un déchiffrement à partir du message chiffré non fauté, on peut raisonnablement supposer que la faute a été injectée à l'instant du calcul où le *state* présente le moins d'octets faux. Cela permet également de connaître la position du ou des octets fautés.

Dans le cas de la figure 4.5, une faute est injectée à la ronde n. Lors du déchiffrement, un seul octet est fauté sur le *state* de la ronde n, tandis que 4 octets

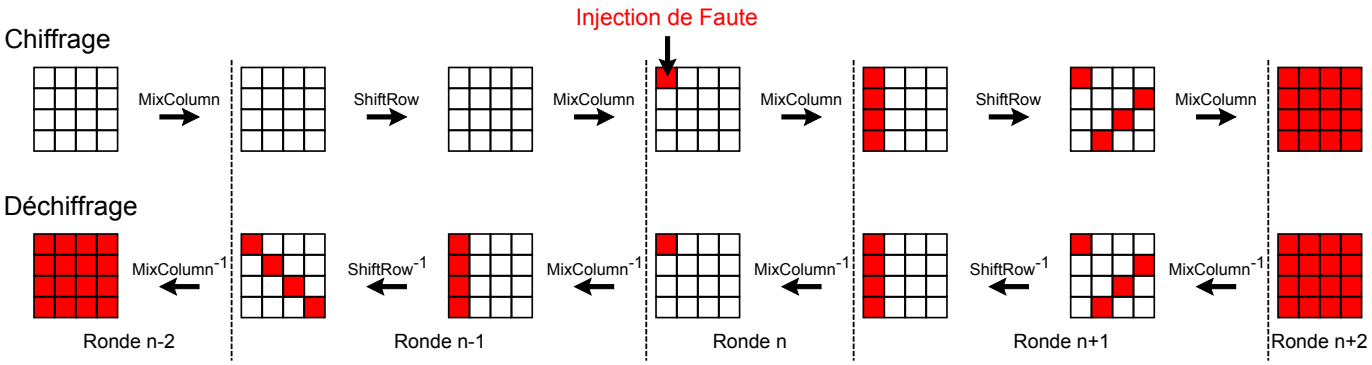


Figure 4.5 – Propagation des fautes dans l'AES, seules les opérations ayant un effet sur la propagation des fautes sont représentées

sont fautés sur les *state*s des rondes $n+1$ et $n-1$, et 16 sur ceux des rondes $n+2$ et $n-2$, ce qui permet de déterminer que la faute a été injectée à un instant situé entre les MixColumns des rondes n et $n-1$.

Afin de déterminer le délai optimal à utiliser entre la réception du signal de tir par la source laser et le tir lui même, des cartographies successives sont effectuées en utilisant un délai variable par pas de 10 ns et une durée d'impulsion de 50 ns (durée minimale possible avec la source Nano). Comme l'aspect spatial n'est pas important dans ce cas, et en raison du grand nombre de cartes à tracer, pour accélérer le processus, la zone de test est réduite à un carré de 50 µm par 50 µm, situé au centre de la zone de l'AES précédemment déterminée.

La carte obtenue avec le plus grand taux de fautes sur la ronde 9 est celle effectuée en utilisant un délai ajustable de 80 ns.

4.2.2 Observations sur l'implémentation de l'AES

Grâce aux résultats des tests précédents, la position spatiale de l'AES et l'instant où le calcul a lieu peuvent à présent être ciblés précisément. Ces informations permettent maintenant de tracer une carte plus détaillée afin d'en extraire des informations sur la conception de l'AES utiles pour mieux cibler la suite des tests ou pour identifier des vulnérabilités potentielles. La durée d'impulsion utilisée est de 50 ns avec un délai de 80 ns, le pas de cartographie est de 1 µm et la taille du spot 1 µm également.

4.2.2.1 Génération des sous-clés de ronde

La carte obtenue est présentée figure 4.6. Sur celle-ci, les fautes identifiées sont cette fois colorées par ronde fautée. En excluant les registres de sortie situés à droite de la figure et identifiés précédemment, il est possible de distinguer deux zones : une où la ronde 9 est touchée (en rouge sur la gauche et en haut de la carte), et une où la ronde 10 est touchée (en noir, en bas et au centre).

Une explication probable à ce phénomène serait la présence de l'implémentation du KeyScheduler de l'AES dans la zone où la ronde 10 est touchée. Le KeyScheduler permet de calculer les sous clés de ronde à partir de la clé de chiffrement initiale. Cette opération peut être effectuée en parallèle du reste du calcul. La sous clé k_n n'étant pas utilisée avant la ronde n, k_n peut être calculée pendant la ronde $n-1$. De cette manière, de nombreuses implémentations de l'algorithme calculent en parallèle une ronde et la sous-clé de la ronde suivante. Une faute injectée sur l'implémentation du KeyScheduler sera alors propagée de manière effective dans le calcul lors de l'opération AddRoundKey de la ronde suivante, ce qui explique l'apparition de ces deux zones distinctes pour un tir ayant lieu à un moment identique du calcul.

La connaissance de la position du KeyScheduler peut être exploitée dans le cadre de certaines attaques visant spécifiquement celui-ci comme par exemple l'attaque présentée dans [39].

4.2.2.2 Colonne fautée

En remontant à la faute injectée comme montré sur la figure 4.5 plus haut, il est possible d'identifier les octets fautés et ainsi de regrouper les fautes injectées en fonction de la colonne du *state* touchée. Celles-ci sont représentées sur la figure 4.7. La couleur des points représente l'indice de la colonne fautée. Sur la figure, il est possible d'observer des zones bien définies correspondant à chaque colonne. La connaissance de la position de ces zones permet donc de choisir sur quelle colonne une faute est injectée ce qui est particulièrement utile dans le cadre de l'attaque de Piret et al, celle-ci nécessitant deux fautes sur chaque colonne pour retrouver la clé de chiffrement. Un attaquant capable de choisir la colonne sur laquelle les fautes sont injectées peut ainsi retrouver la clé en un nombre réduit de tirs laser.

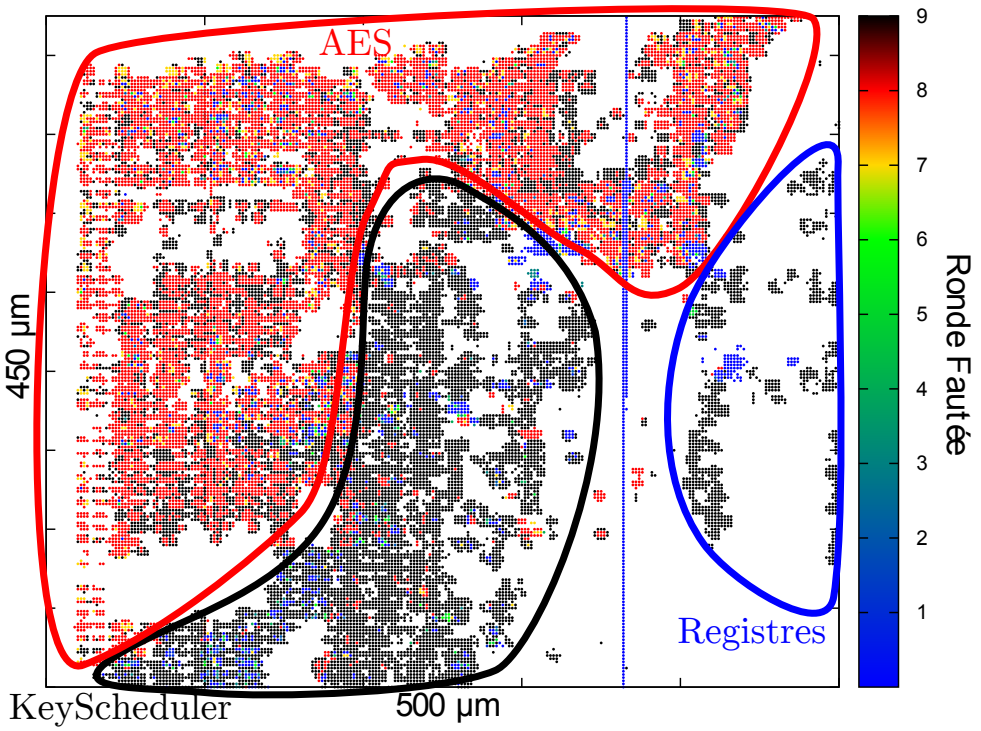


Figure 4.6 – Fautes colorées par ronde fautée

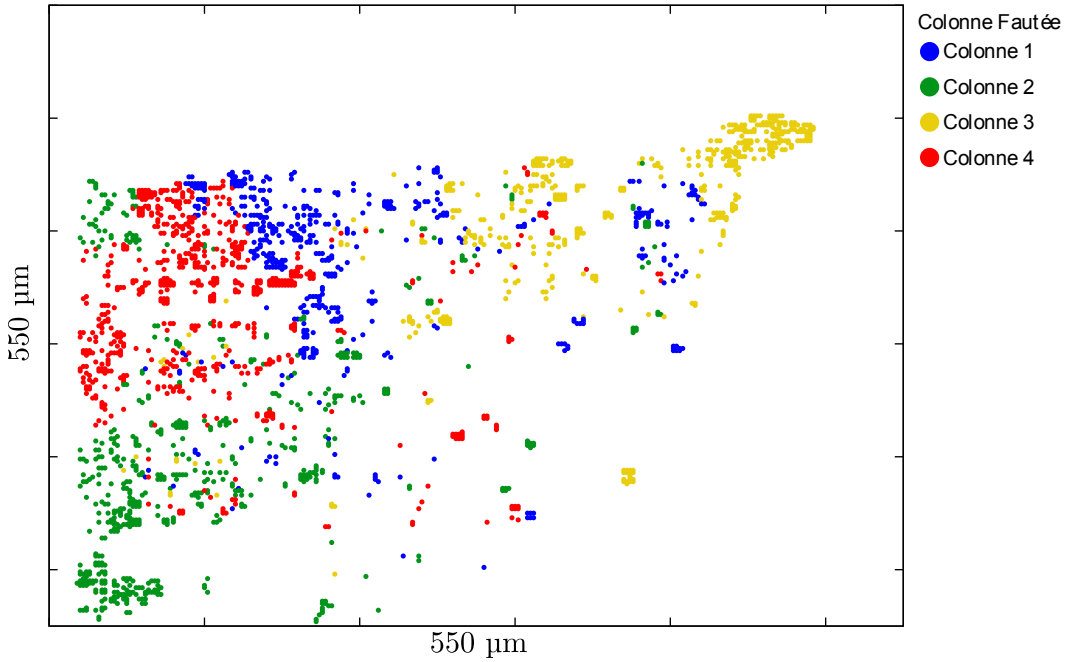


Figure 4.7 – Fautes sur la ronde 9, classées par colonne fautée.

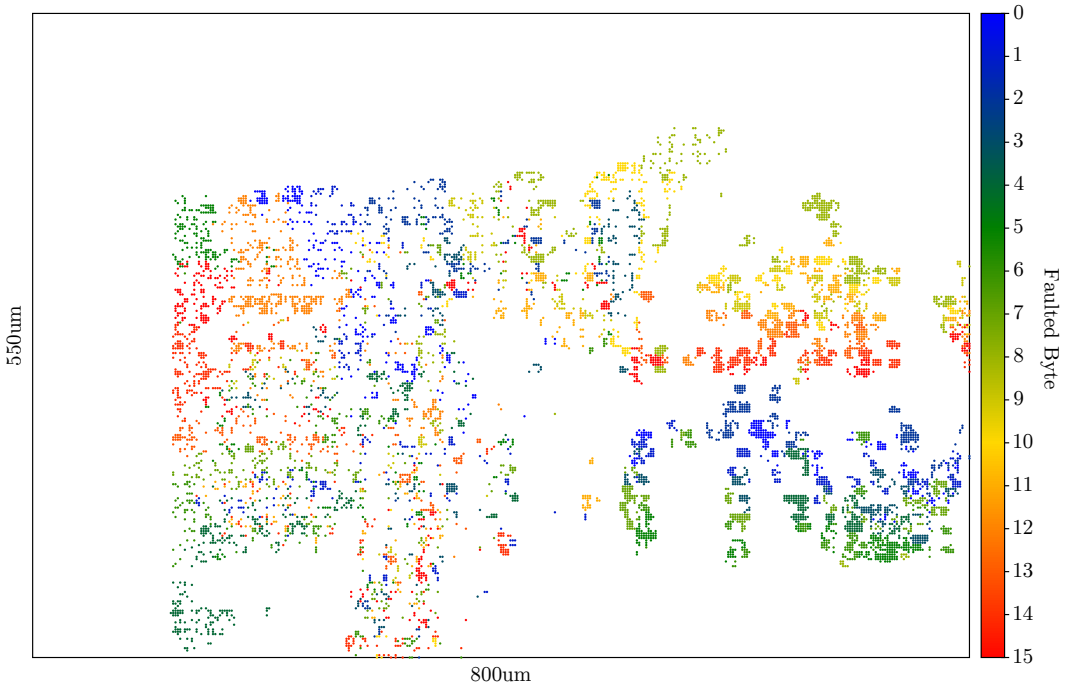


FIGURE 4.8 – Fautes mono-octet, classées par octet.

4.2.2.3 Octet fauté

De la même manière, sur la figure 4.8, les fautes mono-octet obtenues sont classées par octet touché. Similairement à ce qui était observé pour les colonnes du *state*, il est possible d'observer des zones correspondant aux différents octets du *state*. Ces zones sont toutefois moins bien définies que celles correspondant aux colonnes du *state* mais peuvent tout de même être utilisées par un attaquant pour cibler l'octet souhaité. Encore une fois, cette capacité peut être exploité par exemple dans l'attaque présentée dans [71].

4.2.3 Conclusion

Ces tests préliminaires ont permis de repérer l'AES spatialement sur le circuit et de déterminer l'instant d'injection optimal pour viser la ronde souhaitée (le délai de 80 ns permet de viser la ronde 9, mais comme la période d'horloge de l'AES est connue, il est possible de retrouver facilement les autres rondes).

De plus, en effectuant une cartographie plus détaillée du circuit, il a été montré qu'il est possible de contrôler la faute injectée sur le circuit de manière précise,

permettant ainsi d'adapter le type de faute injectée au modèle mathématique de l'attaque choisie (attaques sur la génération des sous-clés, attaques sur une colonne ciblée ou sur un octet ciblé). Un attaquant capable de contrôler ces paramètres peut ainsi obtenir les fautes nécessaires à la conduite d'une attaque en un nombre d'essais réduit, ce qui diminue ses chances d'endommager la cible ou de déclencher une contre-mesure, augmentant ainsi ses chances de succès. Ces résultats sont inquiétants sur le plan sécuritaire, d'autant plus que l'implantation de l'AES est matérielle, il n'est donc pas possible de la modifier pour en combler les vulnérabilités.

La suite de ce chapitre se focalisera sur la protection de l'AES en utilisant les variations de sensibilité liées au FBB observées dans le chapitre 3, ce qui a justement l'avantage de ne pas nécessiter d'apporter de modifications à l'implémentation matérielle.

4.3 Utilisation des paramètres d'alimentation pour durcir l'AES

Dans le chapitre 3, il a été montré que le FBB et les changements de tension d'alimentation ont un effet sur la sensibilité au laser des registres du microcontrôleur ciblé. Un baisse de V_{dd} rend le registre testé plus sensible au laser et l'activation du FBB se traduit par une autre nette augmentation de sensibilité. Cette variation de sensibilité se traduit par une augmentation de la taille des zones sensibles du registre (ce qui accroît les recouvrements entre plusieurs zones et donc les cas où plusieurs bits du registre sont fautés) et une baisse du seuil de puissance laser requis pour injecter une faute dans le registre.

La section suivante présente comment, à partir de ces résultats, il est possible de protéger le circuit en utilisant ses capacités de FBB, face aux attaques laser, sans nécessiter de modifications invasives du circuit.

4.3.1 Principe de base

Comme vu dans le chapitre 1, pour réaliser une attaque en faute, la répétabilité et la prédictabilité des fautes sont des paramètres qui influent grandement sur la réussite

de l'attaque. L'attaquant a besoin de pouvoir prédire au mieux les fautes qu'il est capables d'injecter sur le circuit pour pouvoir adapter le modèle mathématique utilisé. Un meilleur contrôle sur les fautes injectées réduit également le nombre d'injections nécessaires pour obtenir suffisamment de chiffrés fautés afin de mener l'attaque à terme, ce qui réduit les chances d'activer des contre-mesures ou d'endommager le circuit. Ainsi il est dans le meilleur intérêt du concepteur de circuit de chercher à minimiser la répétabilité et la prédictabilité des fautes.

En supposant que l'AES est sujet à des variations similaires à celles observées sur le registre lors de l'utilisation de FBB, il devrait alors être possible d'utiliser ces effets pour modifier les fautes injectées en fonction de la configuration d'alimentation utilisée.

La méthode présentée dans la suite de cette section utilise ces variations de sensibilité au laser liées aux paramètres d'alimentation et de polarisation, observées dans le chapitre 3, pour agir sur la répétabilité des fautes. La figure 4.9 décrit le principe de base de cette méthode : lorsqu'on injecte des fautes dans un circuit à l'aide d'un laser, le type de faute injecté varie en fonction de la puissance laser.

- A très basse puissance aucune faute n'est injectée.
- En augmentant petit à petit la puissance, un premier seuil est atteint, à partir duquel une faute de faible ampleur est injectée (mono-bit ou mono-octet).
- En augmentant encore la puissance du laser, un autre seuil apparaît, à partir duquel plusieurs bits sont fautés.
- Enfin, un dernier seuil de puissance existe, à partir duquel des dommages permanents sont infligés au circuit (collage de bits à 0 ou à 1 de manière définitive, destruction, etc.)

Ces seuils définissent ainsi plusieurs domaines de puissance laser au sein desquels les effets obtenus sur le circuit sont différents. Les résultats du chapitre 3 montrent que les seuils entre ces domaines varient en fonction de la tension d'alimentation et de la polarisation des caissons. En activant FBB, les seuils sont abaissés et, de la

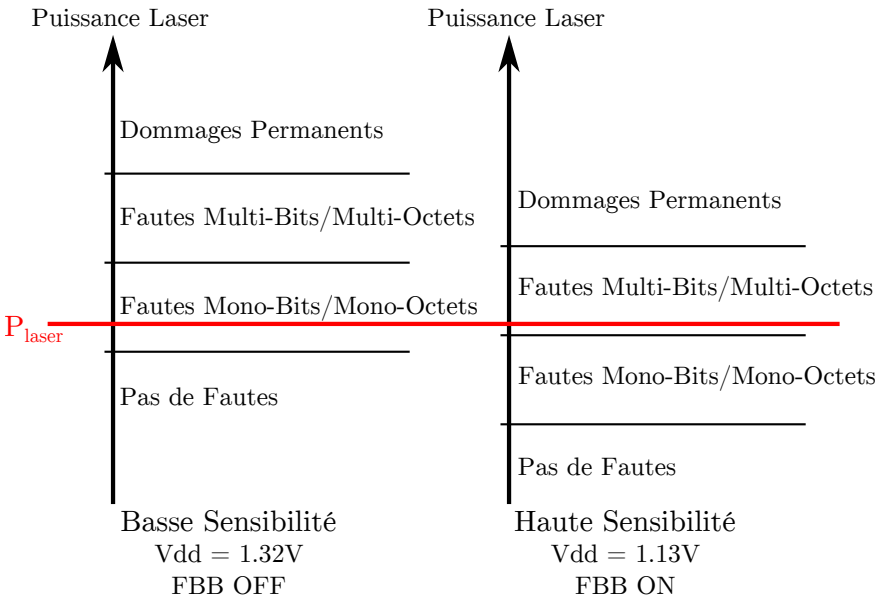


Figure 4.9 – Le circuit réagit différement pour différentes configurations d'alimentation à puissance laser fixe.

même manière, les seuils suivent les variations de V_{dd} (une baisse de V_{dd} entraîne une diminution des valeurs des seuils, et vice-versa).

En définissant plusieurs configurations d'alimentation, différents modes de fonctionnement du circuit, dans lesquels celui-ci est plus ou moins sensible au laser, peuvent être obtenus. De plus, les fautes injectées devraient alors être différentes d'une configuration à l'autre pour une puissance laser donnée. Par exemple, sur la figure 4.9, deux modes sont définis, un mode basse sensibilité (FBB désactivé, V_{dd} élevé) et un mode haute sensibilité (FBB activé et V_{dd} bas) et le circuit est attaqué avec une puissance laser fixe P_{laser}. Lorsque le circuit est attaqué pendant qu'il fonctionne en mode basse sensibilité, des fautes mono-bit ou mono-octets sont obtenues. En passant alors en mode haute sensibilité, les seuils entre les domaines de fautes sont abaissés et les fautes, à présent obtenues, sont des fautes sur plusieurs bits ou octets.

En faisant varier les conditions de polarisation et de tension d'alimentation d'un chiffrement à un autre, il est ainsi possible de réduire la prédictabilité et la répétabilité des fautes injectées sur l'AES

4.3.2 Preuve de concept

Afin de tester la validité du concept de cette contre-mesure, elle est appliquée à un AES temporellement redondant. Cela consiste à effectuer plusieurs fois le même calcul consécutivement sur le même équipement, puis à comparer les résultats. Si les résultats sont identiques, le résultat est retourné. Le cas échéant, le circuit peut effectuer un action destinée à empêcher l'attaque (bloquage de la sortie du calcul, sortie aléatoire, etc.) ou même des attaques ultérieures (effacement de la clé de chiffrement, remise à zéro du circuit, etc.). Ainsi, pour effectuer une attaque en faute avec succès, un attaquant doit être capable d'injecter une faute identique sur plusieurs calculs consécutifs. Cela implique, généralement, dans le cas d'une attaque laser, de pouvoir tirer plusieurs fois consécutives sur une même position avec une puissance fixe à un même instant exact du calcul.

Cela fait de la redondance temporelle une application directe de la méthode de durcissement proposée. Si les hypothèses émises dans la sous-section précédente sont vraies, un changement de mode d'alimentation entre deux calculs d'une redondance devrait compliquer considérablement l'injection de deux fautes identiques consécutives. La sensibilité du circuit variant d'un chiffrement à l'autre, les fautes obtenues sur les deux calculs consécutifs devraient être différentes même si l'attaquant est capable d'injecter une faute deux fois de suite au même instant.

La figure 4.10 illustre ces cas. Lorsqu'une redondance temporelle double est utilisée, dans le cas normal, le chiffrement est effectué deux fois, les résultats obtenus sont identiques et le résultat est renvoyé (a). Si le circuit est attaqué pendant le calcul à deux instants différents, deux résultats distincts sont obtenus et l'attaque est ainsi détectée (b). En revanche, si l'attaquant parvient à injecter deux fautes identiques consécutives en tirant deux fois au même instant du calcul, les deux chiffrés fautés sont identiques, l'attaque n'est donc pas détectée et le résultat est renvoyé (c). En effectuant un calcul dans un mode haute sensibilité et un calcul dans un mode basse sensibilité, même si l'attaquant parvient à tirer deux fois consécutivement au même instant du calcul, les deux chiffrés ont alors une chance d'être différents et ainsi de permettre une détection de l'attaque (d).

La redondance temporelle est une contre-mesure coûteuse en temps de calcul (et

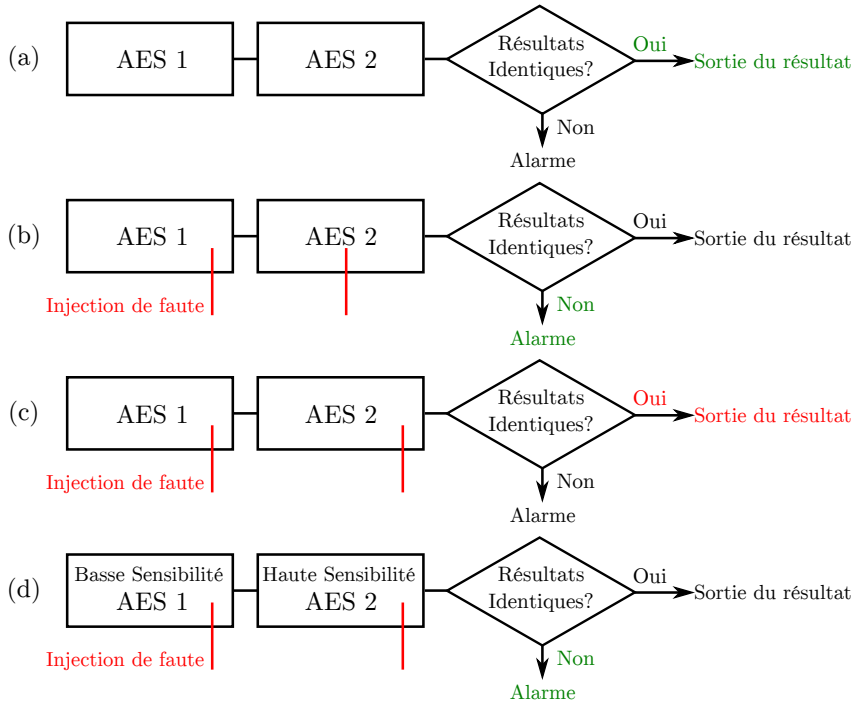


Figure 4.10 – Exemple d'utilisation des variations de sensibilité pour durcir une redondance temporelle.

donc en consommation), cependant dans le cas présent, elle présente deux avantages importants. Elle ne nécessite pas d'apporter de modifications au circuit, et peut être implémentée facilement de manière logicielle sur le microcontrôleur. De plus, c'est une application directe de la méthode proposée dans cette section : le calcul étant répété plusieurs fois, la répétabilité des fautes devient d'autant plus importante, et une méthode permettant de réduire cette dernière est donc particulièrement adaptée à la redondance.

Cependant, la redondance reste uniquement une application de la méthode de durcissement via le FBB proposée et est utilisée comme exemple dans cette section. D'autres applications, comme l'utilisation de valeurs aléatoires de (V_{dd}, V_{bb}) à chaque chiffrement, sont également envisageables.

4.3.3 Conclusion

Dans cette section, une méthode permettant d'améliorer la résistance du microcontrôleur aux attaques en utilisant les capacités de FBB et de variation de la

tension d'alimentation dont il dispose a été décrite. Cette méthode s'appuie sur les résultats du chapitre 3 qui montrent des variations de sensibilité au laser du microcontrôleur. L'activation du FBB et la baisse de V_{dd} entraînent une baisse des seuils de puissance laser requis pour injecter une faute. Cet effet peut alors être utilisé pour définir plusieurs niveaux de sensibilité entre lesquels le circuit est capable de basculer. Pour chaque niveau, les fautes obtenues sont alors différentes, ce qui permet de réduire la répétabilité et la prédictaibilité des fautes en changeant de niveau de sensibilité avant chaque calcul. L'utilisation de la redondance temporelle est ensuite proposée pour tester la validité de ces hypothèses. La prochaine section présentera l'implémentation de ces tests ainsi que leurs résultats.

4.4 Implémentation et test de la méthode de durcissement

Dans cette section, l'utilisation de la méthode de durcissement basée sur l'utilisation du FBB présentée précédemment sera implémentée sur le microcontrôleur, couplée à une redondance temporelle comme décrite dans la section 4.3.2. Afin de juger de l'efficacité de la méthode proposée, les résultats de ces tests seront présentés et comparés aux résultats des mêmes tests effectués sur une redondance temporelle n'utilisant pas les variations de sensibilité liées au FBB et aux variations de V_{dd}.

4.4.1 Procédure de test

Deux niveaux de sensibilité sont définis, basés sur les deux configurations correspondant à la plus haute et la plus basse sensibilité observées lors des travaux effectués sur le registre dans le chapitre 3.

- Mode haute sensibilité : Vdd=1.13V et Vbb=400mV.
- Mode basse sensibilité : Vdd=1.32V et Vbb=0mV.

Le but de ces tests est de comparer une redondance temporelle où les deux calculs sont effectués chacun dans un mode de sensibilité différent, à une redondance temporelle classique où les deux calculs sont effectués dans des conditions identiques.

Pour cela, une cartographie de l'AES est effectuée en utilisant la configuration suivante :

- La zone cartographiée mesure 200 µm par 300 µm. Elle est située vers la gauche de la zone couverte par l'AES (pour viser principalement le chemin de données de l'AES (et non le KeyScheduler) et les positions où uniquement des fautes mono-bit étaient enregistrées dans la section 4.2.1.1). Le pas de la cartographie est de 5 µm.

- Pour absorber les variations des conditions expérimentales, non contrôlables (gigue sur l'instant d'injection etc.), il est nécessaire d'effectuer de nombreux chiffrements sur chaque position. Cependant, le temps de cartographie s'en retrouve augmenté d'autant. Afin d'accélérer la procédure, une tentative d'injection de faute avec le circuit en mode basse sensibilité est effectuée pour chaque position de la zone de test. Lorsqu'une faute est injectée avec succès, le processus de test est alors arrêté sur cette position et 2000 tentatives d'injection sont effectuées en alternant une injection en mode basse sensibilité avec une injection en mode haute sensibilité.

- Les autres paramètres sont identiques à ceux déterminés dans la section 4.2.1. La durée d'impulsion est de 50 ns, le diamètre du spot est de 1 µm et le délai réglable est de 80 ns.

Pour chaque position de la carte, les résultats obtenus sont ainsi une suite de 2000 messages chiffrés, alternant un chiffrement effectué en mode basse sensibilité et un chiffrement effectué en mode haute sensibilité. Afin de trier ces résultats et évaluer l'impact de la méthode proposée, ceux-ci sont classés de la manière décrite sur la figure 4.11. Les résultats consécutifs sont comparés deux à deux pour déterminer la sortie de la redondance qu'ils forment. De cette manière, il est possible d'obtenir en une seule cartographie les résultats qui auraient été obtenus via trois types de redondances différents :

- Deux calculs consécutifs forment une redondance "durcie" où le premier calcul est effectué en mode haute sensibilité et le second en mode basse sensibilité.

- Deux calculs en mode basse sensibilité consécutifs forment une redondance basse sensibilité.

- Deux calculs en mode haute sensibilité consécutifs forment une redondance haute sensibilité.

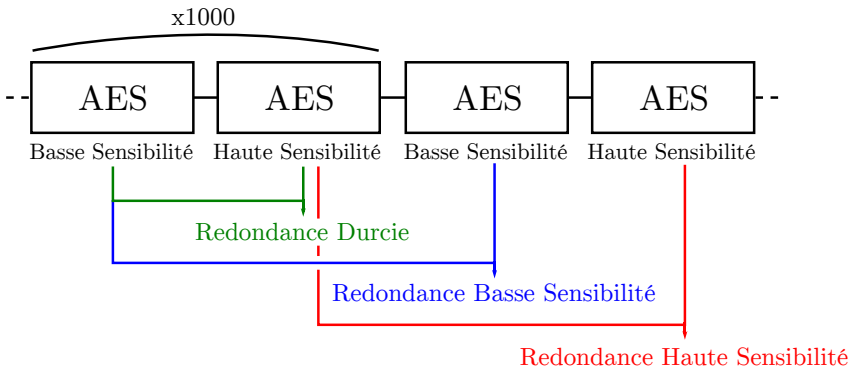


FIGURE 4.11 – Procédure de Test

En traitant les résultats de cette manière, il est ainsi possible de comparer la méthode de durcissement à une redondance temporelle simple dans les cas où le circuit fonctionne en mode haute et basse sensibilité. Ceci afin de s'assurer que les gains de sécurité sont indépendants du mode de fonctionnement de la redondance simple.

4.4.2 Premiers résultats

Pour chaque paire de chiffrés équivalente à une des redondances définies précédemment, il est possible de distinguer trois résultats :

- Double faute : Les deux chiffrés sont fautés de manière identique. Dans ce cas, l'injection de faute a été réalisée avec succès et le circuit est compromis comme sur la figure 4.10c.

- Détection : Les résultats des deux chiffrements sont différents (soit deux fautes différentes ont été injectées, soit un seul des calculs a été fauté). Dans ce cas, le circuit peut alors détecter l'attaque ce qui lui permet alors de réagir (alarme, sortie aléatoire, remise à zéro etc.) comme sur la figure 4.10b.

Résultat	Redondance Durcie	Redondance Haute Sensibilité	Redondance Basse Sensibilité
Double faute	35.6%	74.3%	51.7%
Détection	58.6%	13.6%	18.9%
Aucune faute	5.8%	12.1%	29.4%

Table 4.1 – Chances d'obtenir chaque résultat pour chaque type de redondance sur l'ensemble des positions de la carte.

- Aucune faute : Aucun des deux chiffrés n'a été fauté. Dans ce cas, la sécurité du circuit n'est pas compromise, mais l'attaque n'est pas détectée comme sur la figure 4.10a.

Parmi ces trois cas, le but de la contre-mesure est de minimiser la quantité de doubles-fautes obtenues, celles-ci constituant les résultats exploitables par un attaquant. En revanche, la quantité de détections est le paramètre que l'ont cherche à maximiser. Enfin, le cas où aucune faute n'a été injectée est neutre. Il ne permet pas d'attaquer le circuit, mais ne permet pas non plus de détecter une attaque manquée, ce qui permet à l'attaquant de tenter une nouvelle injection.

La table 4.1 rapporte les chances d'obtenir chaque résultat pour chaque type de redondance sur l'ensemble des positions testées durant le processus de cartographie.

La première observation, à partir de ces résultats, est que la redondance durcie présente un taux de double faute plus faible que les deux autres configurations et un taux de détection plus élevé. Ce résultat est encourageant, puisqu'il indique que la redondance durcie est plus à même d'empêcher l'injection de fautes et permet d'augmenter les chances de détecter une attaque. Cela signifie que la contre-mesure semble fonctionner, d'autant plus que l'écart de performance avec les autres types de redondance est important.

Les performances de la redondance haute sensibilité sont particulièrement mauvaises. Le taux de double faute est plus de deux fois plus grand et le taux de détection plus de quatre fois plus faible que celui de la redondance durcie. Les résultats de la redondance basse sensibilité sont un peu meilleurs, mais restent inférieurs à ceux de la redondance durcie.

En revanche, le cas où aucune faute n'est injectée présente des résultats plus inattendus. En effet, celui-ci est plus élevé pour la redondance haute sensibilité que pour la redondance durcie. Cela signifie que, pour un nombre non négligeable de cas, des fautes sont injectées en mode basse sensibilité et pas en mode haute sensibilité. Cela va à l'encontre du modèle de variations de sensibilité établi à partir des résultats sur le registre. Une hypothèse permettant d'expliquer ces résultats sera présentée par la suite mais reste à valider.

En conclusion, ces premiers résultats indiquent que la contre-mesure proposée est plus efficace qu'une redondance temporelle classique. La redondance durcie permettant de détecter les tentatives d'injection plus fréquemment et d'avoir un meilleur taux de détection d'injection de fautes. Cependant, ces résultats sont moyennés sur l'ensemble du circuit. La suite présente ces même résultats en prenant en compte l'aspect spatial afin de vérifier qu'il n'y a pas de positions pour lesquelles les tendances s'inversent.

4.4.3 Analyse spatiale des résultats

La table 4.1 récapitule les résultats obtenus sur l'ensemble de la surface cartographiée et ne prennent donc pas en compte d'éventuels effets spatiaux, où par exemple la contre-mesure serait moins efficace sur certaines portions de l'AES. Afin de vérifier cela, cette partie des résultats montre les cartes obtenues.

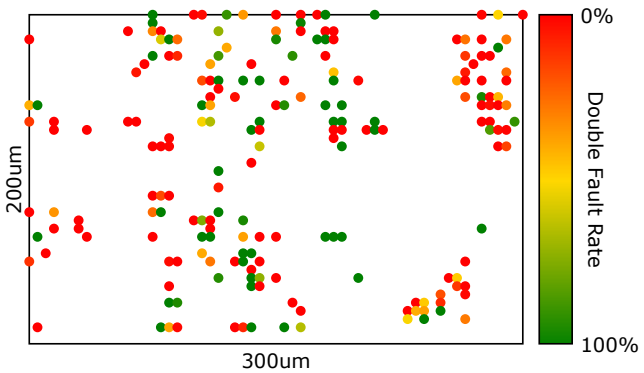


FIGURE 4.12 – Carte du taux de double faute pour la redondance durcie.

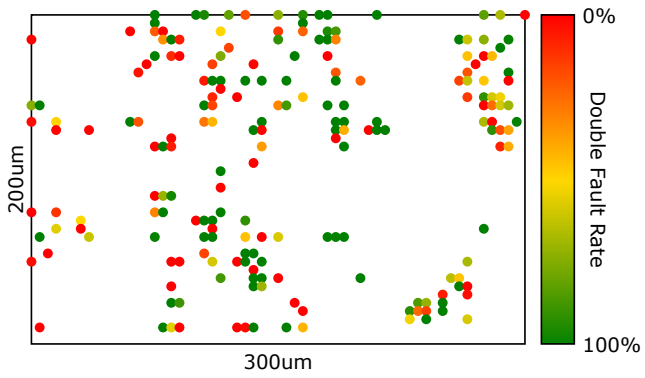


Figure 4.13 – Carte du taux de double faute pour la redondance simple en mode basse sensibilité.

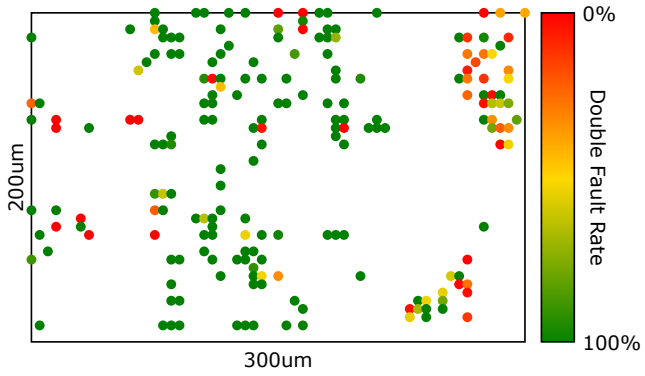


Figure 4.14 – Carte du taux de double faute pour la redondance simple en mode haute sensibilité.

Les figures 4.12, 4.13 et 4.14 présentent les cartes des taux de double faute pour les redondances durcie, basse sensibilité et haute sensibilité respectivement. Chaque point représente une position testée (les positions sur lesquelles le processus de cartographie s'est arrêté pour effectuer 2000 tests) et la couleur du point donne le taux de double faute pour cette position.

Ces cartes confirment les résultats de la table 4.1 : le taux de double faute observé est plus faible pour la redondance durcie que pour les deux autres. En revanche, elles ne révèlent pas de tendance particulière vis à vis de la répartition spatiale des fautes. L'efficacité des trois types de redondances semble uniforme sur toute la surface de l'AES La même tendance peut être observée sur les cartes correspondant au taux de détection, celles-ci confirment les résultats globaux, mais n'affichent pas de tendance quant à la répartition des fautes.

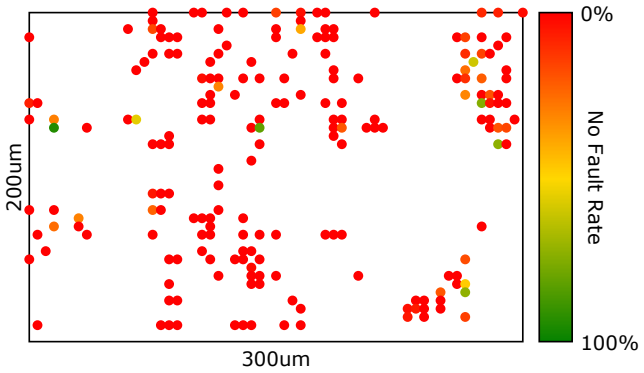


FIGURE 4.15 – Carte des chances d'obtenir deux chiffrés non fautés pour la redondance durcie.

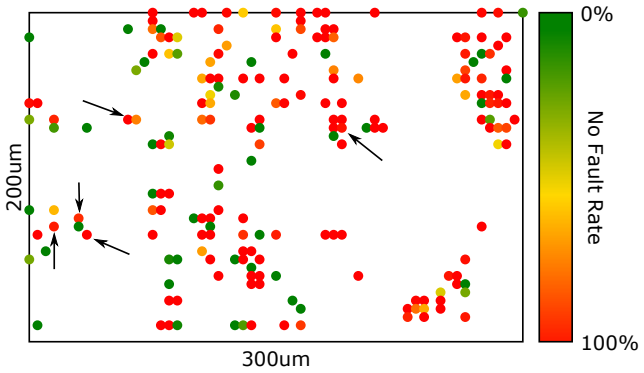


FIGURE 4.16 – Carte des chances d'obtenir deux chiffrés non fautés pour la redondance simple en mode basse sensibilité.

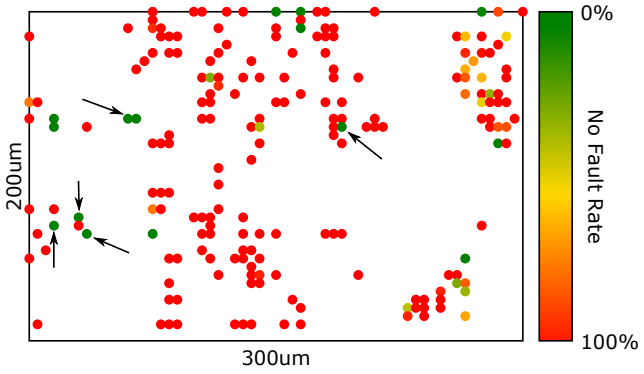


FIGURE 4.17 – Carte des chances d'obtenir deux chiffrés non fautés pour la redondance simple en mode haute sensibilité.

De la même manière, les figures 4.15, 4.16 et 4.17 représentent les chances d'obtenir deux chiffrés non fautés pour les redondances durcie, basse sensibilité et haute sensibilité respectivement. Il n'y a pas non plus de tendance spatiale qui se dégage de ces cartes. En revanche, celles-ci confirment les résultats étranges observés dans la table 4.1. Il est effectivement possible d'observer des positions (indiquées par des flèches sur les cartes) pour lesquelles des fautes sont injectées en mode basse sensibilité et non en mode haute sensibilité, ce qui confirme qu'un effet supplémentaire s'ajoute aux variations de sensibilité observées sur le registre.

Les résultats observés sur les cartes confirment donc les résultats globaux donnés par la table 4.1, y compris les résultats étranges sur le taux de non-faute. Cependant, elles ne permettent pas d'observer de comportement spatial particulier ou d'apporter des éléments supplémentaires permettant d'expliquer les écarts inattendus du taux de non-faute.

4.4.4 Comparaisons position par position

Les cartes présentées précedemment confirment les tendances globales, mais ne permettent pas d'observer d'effets spatiaux notables comme des zones où l'efficacité des différentes redondances varie. Cependant, l'efficacité de la contre-mesure n'est pas pour autant la même sur toutes les positions testées. Pour cette raison, cette sous-section présente des graphes permettant de comparer, position par position, les performances des trois types de redondances.

Le graphe présenté figure 4.18 montre la comparaison du taux de double faute, position par position, entre la redondance durcie et celle en mode haute sensibilité. Chaque point de la figure représente une des 227 position testées. L'abscisse du point donne le taux de double faute pour la redondance durcie et l'ordonnée le taux de double faute pour la redondance haute sensibilité. Par exemple, un point situé dans le coin supérieur droit du graphe représente une position pour laquelle le taux de double faute est de 100% en mode haute sensibilité et de 0% en utilisant la redondance durcie. En conséquence, les points au-dessus de la diagonale sont ceux pour lesquel le taux de double faute est plus élevé pour la redondance haute sensibilité et ceux au-dessous, ceux pour lesquel la performance de la redondance

durcie est moins bonne. Enfin, les points situés sur la diagonale représentent les positions pour lesquelles les performances des deux redondances comparées sont identiques.

Ce graphe montre que pour certaines positions, peu nombreuses, la redondance durcie est moins efficace que la redondance haute sensibilité, ce qui n'était pas évident à la vue des cartes et des résultats globaux. Pour 18 positions sur les 227 testées, le taux de double faute est inférieur en utilisant la redondance haute sensibilité. Cependant, l'écart de performance est supérieur à 10% de double fautes supplémentaires pour seulement 8 d'entre elles (avec un maximum de 19% d'écart en faveur de la redondance haute sensibilité). En comparaison, pour 52 positions la performance des deux redondances est égale et pour les 157 positions restantes, la redondance durcie présente un taux de double faute plus faible, dont 55 pour lesquelles le taux de double faute est inférieur de 90% ou plus à celui de la redondance haute sensibilité. La redondance durcie reste donc largement plus performante.

De la même manière, le graphe 4.19 compare le taux de détection, position par position entre ces deux mêmes redondances : les points situés au-dessous de la diagonale sont les positions pour lesquelles le taux de détection est plus élevé en utilisant la redondance durcie. Similairement à ce qui pouvait être observé pour le taux de double faute, pour certaines positions le taux de détection est plus élevé pour la redondance haute sensibilité. Mais la tendance générale est largement en faveur de la redondance durcie. Sur les 227 positions testées, 14 présentent un taux de détection qui favorise la redondance haute sensibilité avec un maximum de 16% d'écart. Pour 41 points, les deux redondances sont équivalentes. Pour les 156 points restants la redondance durcie a plus de chances de détecter l'attaque, dont 58 points pour lesquels le taux de détection est plus de 90% plus élevé en utilisant la redondance durcie.

Le graphe figure 4.20 montre la comparaison entre la redondance durcie et la redondance basse sensibilité. De manière identique aux résultats pour la redondance haute sensibilité, pour certaines positions, les performances de la redondance basse sensibilité sont en réalité meilleures que celles de la redondances durcie. Comme les résultats de la table 4.1 pouvaient le laisser présager, l'écart de performance est

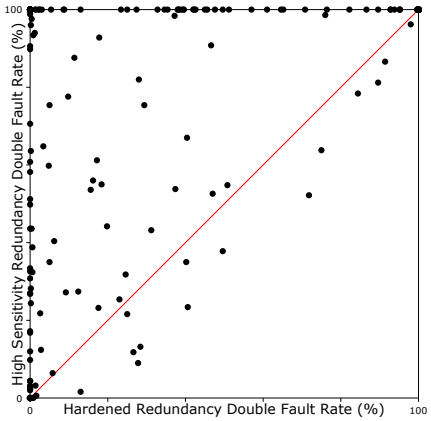


Figure 4.18 – Taux de double faute pour les redondances haute sensibilité et durcie.

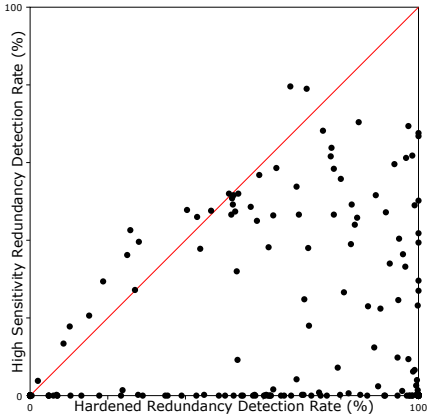


Figure 4.19 – Taux de détection pour les redondances haute sensibilité et durcie.

moins important qu'entre la redondance durcie et la redondance haute sensibilité, mais reste en faveur de la redondance durcie. Sur les 227 positions testées, la redondance basse sensibilité présente un taux de double faute plus faible sur 59 positions (avec un écart maximum de 25%). Pour 75 positions, les performances des deux redondances sont équivalentes. Enfin, pour les 93 positions restantes, la redondance durcie présente un taux de double faute plus faible, dont 20 positions pour lesquelles la différence de taux de double faute est de 90% ou plus.

Similairement, le graphe figure 4.21 compare le taux de détection pour ces deux redondances. Encore une fois, pour certaines positions la redondance durcie est en réalité légèrement moins performante mais reste globalement la plus sûre : sur les 227 positions testées, la redondance durcie est inférieure sur 23 d'entre elles

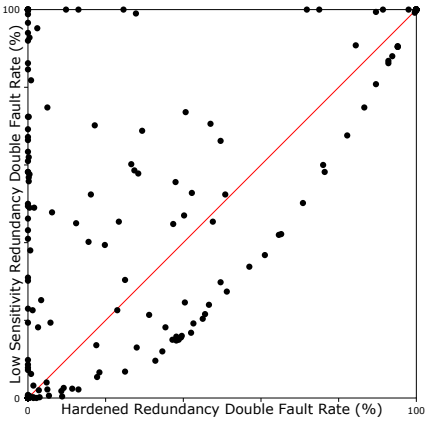


FIGURE 4.20 – Taux de double faute pour les redondances basse sensibilité et durcie.

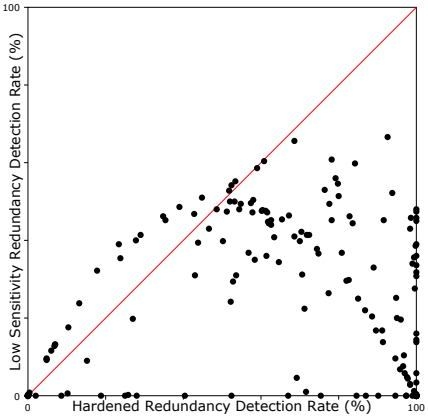


FIGURE 4.21 – Taux de détection pour les redondances basse sensibilité et durcie.

(l'écart maximal entre les deux taux étant de 15.5%). Les deux redondances ont des performances égales pour 43 positions. Sur les 161 positions restantes, la redondance durcie est supérieure, dont 51 positions pour lesquelles la différence entre les deux taux est de 90% ou plus.

Les résultats obtenus à partir de ces graphes montrent que l'efficacité de la contre-mesure est dépendante de la position de l'injection laser, et confirment son efficacité générale. De plus, l'existence de positions où la performance de la contre-mesure est inférieure à une redondance classique, tend à confirmer la présence d'effets supplémentaires non observés dans les résultats du chapitre 3. Des hypothèses sur l'origine de ces effets sont présentés dans la section suivante.

4.4.5 Sensibilité et aspect temporel

Comme observé dans les résultats globaux (table 4.1) puis dans les cartes, pour certaines positions (41 sur les 227 testées), les chances d'obtenir deux chiffrés non fautés sont plus élevées en mode haute sensibilité qu'en mode basse sensibilité. Autrement dit, pour ces positions, les chances d'injecter une faute sont plus élevées en mode basse sensiblité qu'en mode haute sensibilité. Cela va à l'encontre des résultats obtenus dans le chapitre 3 et sur lesquels se base la contre-mesure présentée ici.

Une explication potentielle à cet écart de résultats pourrait être l'aspect temporel des tests sur l'AES Les tests sur registre par contraste étaient effectués de manière statique : le registre était écrit, puis le tir était déclenché et enfin la valeur du registre était relue. Les fautes étaient injectées directement sur les éléments mémoire du registres et non sur la valeur échantillonée lors de l'écriture ou de la lecture.

L'AES en revanche est composé en partie de logique combinatoire. Cela introduit un autre mécanisme d'injection de faute, où comme vu dans le chapitre 1, lorsque qu'une faute est provoquée dans la logique combinatoire, elle se propage en aval. La faute peut ensuite être échantillonnée par un registre si le signal est perturbé lorsque celui-ci échantillonne son entrée.

Bien que la fréquence d'horloge de l'AES soit la même pour les deux modes de sensibilité, le FBB a quand même pour effet de réduire la tension de seuil des transistors, ce qui réduit les temps de propagation des signaux entre deux registres. Cela pourrait ainsi avoir pour effet de réduire la fenêtre temporelle pendant laquelle il est possible d'injecter une faute, comme montré sur la figure 4.22. Si le temps de propagation est plus court, le signal devrait mettre moins de temps à retrouver sa valeur correcte après une perturbation liée à une injection de faute. Ainsi, la durée de la fenêtre temporelle pendant laquelle une faute injectée sera échantillonnée lors du front d'horloge suivant est raccourcie.

En réduisant la durée de la fenêtre d'injection, la probabilité d'injecter une faute est réduite, ce qui pourrait expliquer les effets observés.

Cependant ces hypothèses, sont difficiles à vérifier en l'absence d'informations

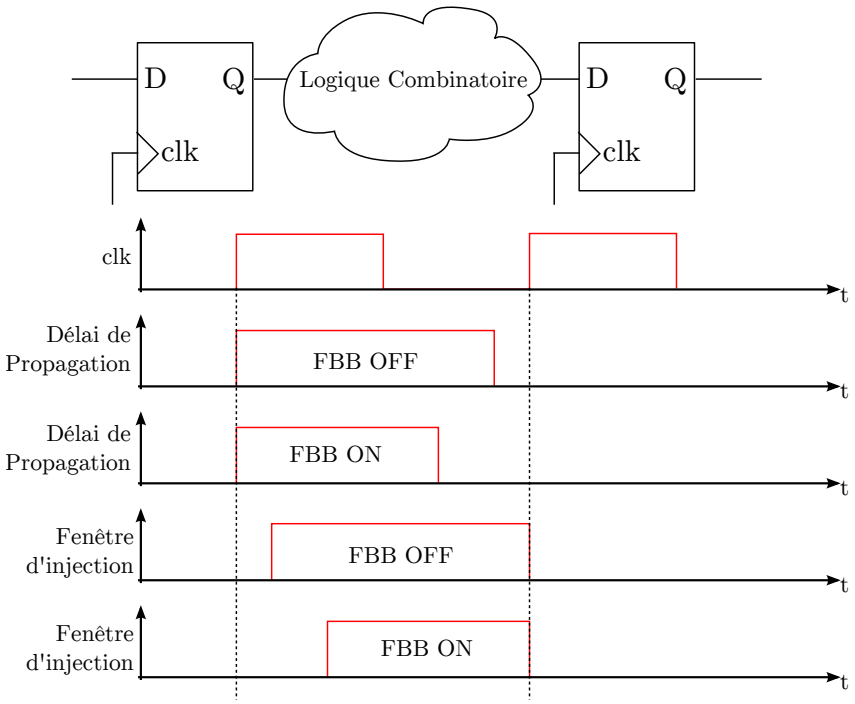


FIGURE 4.22 – Temps de propagation dans la logique combinatoire

exactes sur la conception du circuit. Une méthode pour en tester la validité serait d'utiliser la source Picoseconde du laser et d'injecter des fautes sur l'AES en faisant varier l'instant d'injection de manière à balayer la durée du calcul par incréments de quelques picosecondes. En répétant l'opération avec et sans FBB, il serait ainsi possible d'observer les plages temporelles sur lesquelles chaque ronde est fautée. Puis, en comparant la durée des plages entre le calcul avec et sans FBB, il devrait être possible d'avoir plus d'informations sur ces phénomènes.

4.5 Conclusion

Dans le chapitre 3, il a été montré que les paramètres d'alimentation avaient un impact sur la sensibilité d'un registre à l'injection de fautes laser. En effet, lorsque la tension d'alimentation du circuit est abaissée, il devient plus facile d'injecter des fautes. De même, il a été présenté que l'activation du FBB augmente également la sensibilité du registre testé. Cette augmentation se traduit par une surface sensible plus grande et un seuil de puissance laser plus bas. Ainsi, ce dernier chapitre montre comment utiliser ces effets pour améliorer la résistance d'un circuit face aux attaques

laser.

Dans la première partie de ce chapitre, une étude de l'AES matériel implémenté sur le microcontrôleur utilisé dans le cadre de ces travaux a été présentée. Celle-ci avait pour but de définir les paramètres de l'injection de faute sur ce circuit mais aussi d'en identifier les vulnérabilités afin de définir par la suite une contremesure permettant de le protéger. Les résultats de cette analyse ont mis en avant différentes vulnérabilités du circuit. En effet, nous avons vu qu'il est par exemple possible, en contrôlant les paramètres d'injection de faute, de cibler précisément un octet, ou une étape spécifique du calcul. Afin de palier à ces vulnérabilités, une contremesure utilisant les capacités de variations d'alimentation et de polarisation des caissons a été présentée. Cette méthode propose de réduire la répétabilité et la prédictabilité des fautes en modifiant de manière dynamique les conditions de polarisation du circuit puisque celles-ci ont un impact sur sa sensibilité aux attaques laser. En utilisant des conditions d'alimentation variables d'un chiffrement à un autre, la sensibilité du circuit est modifiée et par conséquence les fautes obtenues sont moins prévisibles.

Cette méthode est ensuite utilisée pour améliorer une redondance temporelle sur l'AES testé, et les résultats obtenus sur cette implémentation sont présentés. Ceux-ci montrent une amélioration de la sécurité du circuit lorsque les deux calculs de la redondance sont effectués dans deux conditions d'alimentation différentes : les chances d'injecter deux fois de suite la même faute sur les deux calculs, et ainsi casser la redondance, sont réduites et les chances de détecter l'attaque augmentées. Ces résultats, couplés à l'étude de sensibilité du registre présentée dans le chapitre 3, ont fait l'objet d'une publication à DATE 2016 [72]. Bien que des techniques similaires utilisant des variations de fréquence et de tension d'alimentation aient déjà été proposées pour protéger les circuits face aux attaques par analyse de la consommation [73–75], ceci constitue à notre connaissance la première évaluation face à des attaques par injection de fautes et en utilisant le FBB.

Les résultats de ce chapitre montrent que les capacités d'un circuit destinées à réduire sa consommation peuvent être détournées de leur usage initial pour améliorer sa sécurité. En effet, bien que la contre-mesure proposée ici exploite la capacité du

circuit à utiliser un V_{dd} réglable et le FBB, elle le fait d'une manière totalement différente de leur utilisation prévue. Elle n'est donc pas une contre-mesure basse consommation.

Bien qu'efficace, cette contre-mesure présente quelques limitations. Elle présente un coût additionnel non négligeable en temps nécéssaire pour basculer d'un mode d'alimentation à un autre entre deux chiffrements. Elle n'est pas suffisante pour assurer la sécurité du circuit à elle seule (même en utilisant la redondance durcie, il reste 35% de chances d'injecter une double faute en moyenne). Son impact face à d'autre moyens d'injection (impulsions électro-magnétiques, etc.) ou aux attaques par analyse de canaux cachés n'a pas été étudié.

En revanche, son avantage principal est de pouvoir s'utiliser sur le circuit sans nécessiter d'y apporter des modifications matérielles et de pouvoir être activée ou désactivée à tout moment de manière logicielle. Elle représente ainsi un moyen d'ajuster dynamiquement le compromis entre la sécurité, la consommation et les performances du circuit, un peu de la même manière que le FBB et un V_{dd} réglable permettent d'ajuster le ratio consommation/performances.

Le principal axe d'amélioration de la méthode présentée est le temps nécéssaire pour basculer d'une configuration d'alimentation à une autre. Celui ci peut être amélioré au niveau de la conception du circuit mais également via la manière dont la contre-mesure est implémentée. Par exemple, dans le cas de l'application à la redondance temporelle, il est possible d'atténuer ce coût lorsque les messages à chiffrer sont longs. En effet, lors d'un chiffrement en mode chaîné, par exemple, il n'est pas nécessaire de changer les paramètres d'alimentation entre chaque bloc à chiffrer mais seulement entre les deux chiffrements d'un même message.

D'autres améliorations ou applications sont également imaginables. Par exemple, l'utilisation du RBB pourrait permettre d'étendre la plage sur laquelle il est possible de faire varier la sensibilité du circuit. L'utilisation d'un oscillateur chaotique, pour générer les valeurs de polarisation des caissons et ainsi obtenir une sensibilité aléatoire, est également envisageable

Enfin, une étude plus poussée visant à valider les hypothèses présentées dans la section 4.4.5 pourrait ouvrir de nouvelles voies d'amélioration de cette méthode et

apporter une meilleure compréhension des phénomènes physiques en jeu.

Conclusion générale

La généralisation de de l'utilisation de communications sécurisés et le nombre grandissant d'appareils mobiles connectés, font de la réduction de la consommation et la sécurité des circuits deux problématiques majeurs auxquels les concepteurs de circuits font face aujourd'hui. Dans ce contexte, le projet MAGE, dont cette thèse fait partie, vise à développer des microcontrôleurs sécurisés basse consommation. L'objectif des travaux présentés dans ce manuscrit était de déterminer l'impact des méthodes de réduction de la consommation ciblées par le projet (*body-biasing* et ajustement de la tension d'alimentation) sur la vulnérabilité des circuits aux attaques par injection de faute laser.

Après un état de l'art ayant introduit les différents concepts nécessaires à la suite des travaux, le banc de test utilisé pendant cette thèse ainsi que le travail effectué pour permettre l'automatisation des campagnes de test ont d'abord été présenté. Pour en valider le fonctionnement, une première étude sur l'influence de la durée d'impulsion sur l'injection de fautes laser sur des mémoires de type SRAM a ensuite été rapportée. Des tests d'injection laser ont été effectués sur des cellules SRAM avec des impulsions d'une durée de 30 ps et les résultats ont été comparés avec des résultats obtenus précédemment avec des impulsions de 50 ns. Cela a permis de confirmer la validité du modèle de faute bit-set/bit-reset établi par la première étude. De plus, certaines zones sensibles théoriques qui n'apparaissaient pas avec les impulsions de 50 ns ont été observées avec les impulsions de 30 ps. Une adaptation du modèle de simulation de Sarafianos et al. pour les impulsions courtes a alors été proposée, et a permis de valider les résultats par la simulation. Enfin, les résultats de ces expériences ont été reproduits sur la RAM d'un microcontrôleur commercial en technologie CMOS 350 nm. Ces travaux constituaient également la première étude

effectuée a l'aide de la source picoseconde du banc de test et ont permis de valider son fonctionnement. Ils ont fait l'objet d'une publication à IOLTS 2015 [12].

Le troisième chapitre a ensuite permis d'apporter une première évaluation des effets du *body-biasing* et de la tension d'alimentation sur la sensibilité au laser d'un circuit CMOS. Une première étude a été effectuée sur le microcontrôleur utilisé dans le chapitre 2. Celle-ci a montré une augmentation du nombre de fautes mono-bit obtenues lorsque la tension d'alimentation du microcontrôleur est abaissée. Une seconde étude a ensuite été réalisée sur un deuxième microcontrôleur plus récent en technologie CMOS 90 nm - celle visée dans le cadre du projet dont cette thèse fait partie - et capable d'utiliser le forward body biasing ainsi que plusieurs valeurs de V_{dd} interne. Cette étude se focalisait sur un registre de l'accélérateur cryptographique inclus dans le microcontrôleur. La sensibilité au laser du registre a été évaluée lorsque celui-ci était soumis à différentes valeurs de tension d'alimentation avec et sans forward body biasing. Les résultats montrent une augmentation de la sensibilité du registre au laser lorsque sa tension d'alimentation est abaissée, et un autre acroissement de sensibilité lorsque le forward body biasing est activé. Cette sensibilité accrue se traduit par un seuil de puissance laser nécessaire pour injecter une faute plus bas, ce qui a également pour effet d'augementer la surface sensible au laser du registre. Cependant l'amplitude de ces variations est telle qu'elles ne devraient pas introduire de nouvelles vulnérabilités d'un point de vue sécuritaire.

Enfin, le dernier chapitre propose une méthode pour améliorer la robustesse d'un circuit basée sur les variations de sensibilité observées dans le chapitre 3. Une étude préliminaire de l'AES matériel inclus sur le microcontrôleur CMOS 90 nm du chapitre 3 est d'abord effectuée. Celle-ci permet de repérer la position de l'AES sur le circuit pour préparer les tests suivants, et met également en lumière des vulnérabilités dans l'implémentation de l'algorithme. La seconde partie du chapitre commence par présenter la méthode de durcissement basée sur les résultats du chapitre 3 : en utilisant les variations de sensibilité introduites par l'utilisation du forward body biasing et d'une tension d'alimentation variable, il est possible de réduire la répétabilité et la prédictabilité des fautes. Ce principe est ensuite appliqué à une redondance temporelle sur l'AES effectuant un premier calcul dans

un mode de sensibilité bas (V_{dd} élevé et forward body biasing désactivé), et un second calcul dans un mode de sensibilité élevé (V_{dd} bas et foward body biasing activé). Cette redondance est alors soumise à l'injection de faute laser et comparée à deux autres redondances, une effectuant deux calculs en mode basse sensibilité et l'autre effectuant les deux calculs en mode haute sensibilité. Les résulats des tests montrent que la redondance améliorée permet de réduire les probabilité qu'un attaquant parvienne à injecter deux fautes identiques successives et augmente les chances d'obtenir deux résultats différents successifs. Ainsi, la méthode proposée dans ce chapitre améliore la capacité de la redondance à bloquer et détecter les attaques laser. Elle a fait l'objet d'une publication à DATE 2016 [72] et constitue à notre connaissance la première évaluation d'une contre-mesure de ce type face à des attaques attaques par injection de fautes et en utilisant le FBB.

Plus globalement, ces travaux constituent une première évaluation de l'effet des techniques basse consommation basées sur la tension d'alimentation et le body biasing sur la vulnérabilité au laser d'un circuit. Les résultats montrent que ces techniques ont un impact négatif limité sur la sécurité d'un circuit. Cependant il est également montré qu'une utilisation détournée des capacités offertes par ces techniques (elles n'ont alors pas d'impact positif sur la consommation du circuit) peut améliorer sa résistance face aux attaques en fautes. Il est ainsi possible d'utiliser ces techniques, initialement conçues pour ajuster dynamiquement l'équilibre entre consommation et perfomances d'un circuit, pour ajuster le compromis entre consommation et sécurité.

Plusieurs pistes sont possibles pour approfondir ces résultats. Dans le chapitre 4, certaines anomalies dans les résultats obtenus sur la redondance laissent à penser que le body biasing a également un impact sur les fautes se propageant dans la logique et il serait être intéressant d'étudier ces effets plus en détail. Par exemple une méthode pourrait consister à effectuer des calculs d'AES, et en ciblant une position fixe avec la source Picoseconde du laser injecter des fautes à différents instants. En balayant la durée de l'algorithme avec un pas très fin (quelques picosecondes), il devrait être possible de reconstruire les plages temporelles durant lesquelles chaque ronde du calcul est fauté. En répétant l'opération avec et sans FBB, les variations des durées

des plages temporelles devraient alors fournir des informations sur l'impact du FBB sur la propagation des signaux.

La contre-mesure présentée dans le chapitre 4 pourrait également être la cible d'améliorations. L'utilisation du Reverse Body Biasing pourrait notamment permettre d'étendre la plage de sensibilité du circuit et ainsi d'augmenter l'efficacité de la contre-mesure. D'autres applications que la redondance temporelle pourraient également être évaluées, comme par exemple l'utilisation d'un générateur chaotique pour polariser les caissons, et obtenir ainsi une sensibilité aléatoire du circuit.

De plus, les résultats observés ont été obtenus principalement sur un circuit en technologie CMOS 90 nm. Il a été montré dans [76] que l'injection laser permet d'obtenir des fautes mono-bit au moins jusqu'au 28 nm. Comme expliqué dans le chapitre 3, les effets du FBB sont difficiles à prévoir par la théorie, ainsi il serait intéressant d'étendre l'étude à d'autres noeuds technologiques, afin d'observer si les effets restent les mêmes.

Enfin, l'impact du *body-biasing* sur l'injection électro-magnétique et les attaques par canaux cachés a déja été étudié dans le cadre d'autre travaux liés au projet MAGE. En revanche, l'injection par perturbation de la tension de polarisation du substrat (*Body-Bias Injection*) n'a pas été évaluée, il serait donc également intéressant d'approfondir cet aspect. Des démarches ont été entamées dans ce sens et la mise en place d'un banc de test BBI a été présentée dans [32].

Glossaire & Acronymes

Bit-flip

Inversion de la valeur d'un bit indépendamment de sa valeur initiale.

Bit-reset

Passage d'un bit de 1 à 0.

Bit-set

Passage d'un bit de 0 à 1.

Body-biasing

Polarisation du substrat et/ou des caissons.

State

Matrice de 4x4 octets utilisée pour représenter les états intermédiaires de l'AES.

Triple-well

Architecture Triple Caissons, une architecture CMOS où les transistors NMOS sont placés dans un caisson dopé P isolé du substrat par un implant profond dopé N.

AES

Advanced Encryption Standard.

BBICS

Bulk Built-In Current Sensor.

CBC

Chipher Block Chaining, mode de fonctionnement d'un algorithme de chiffrement par bloc où les chiffrements des différents blocs sont chainés, afin d'éviter l'apparition de motifs dans les données chiffrées.

DES

Data Encryption Standard.

ECB

Electronic Code Book, mode de fonctionnement d'un algorithme de chiffrement par bloc où chaque bloc est chiffré/déchiffré individuellement.

FBB

Forward Body-Biasing.

NIST

National Institute of Standards and Technology, l'organisme américan responsable de la définition des standards technologiques.

RBB

Reverse Body-Biasing.

SET

Single Event Transient.

SEU

Single Event Upset.

SPICE

Simulation Program with Integrated Circuit Emphasis.

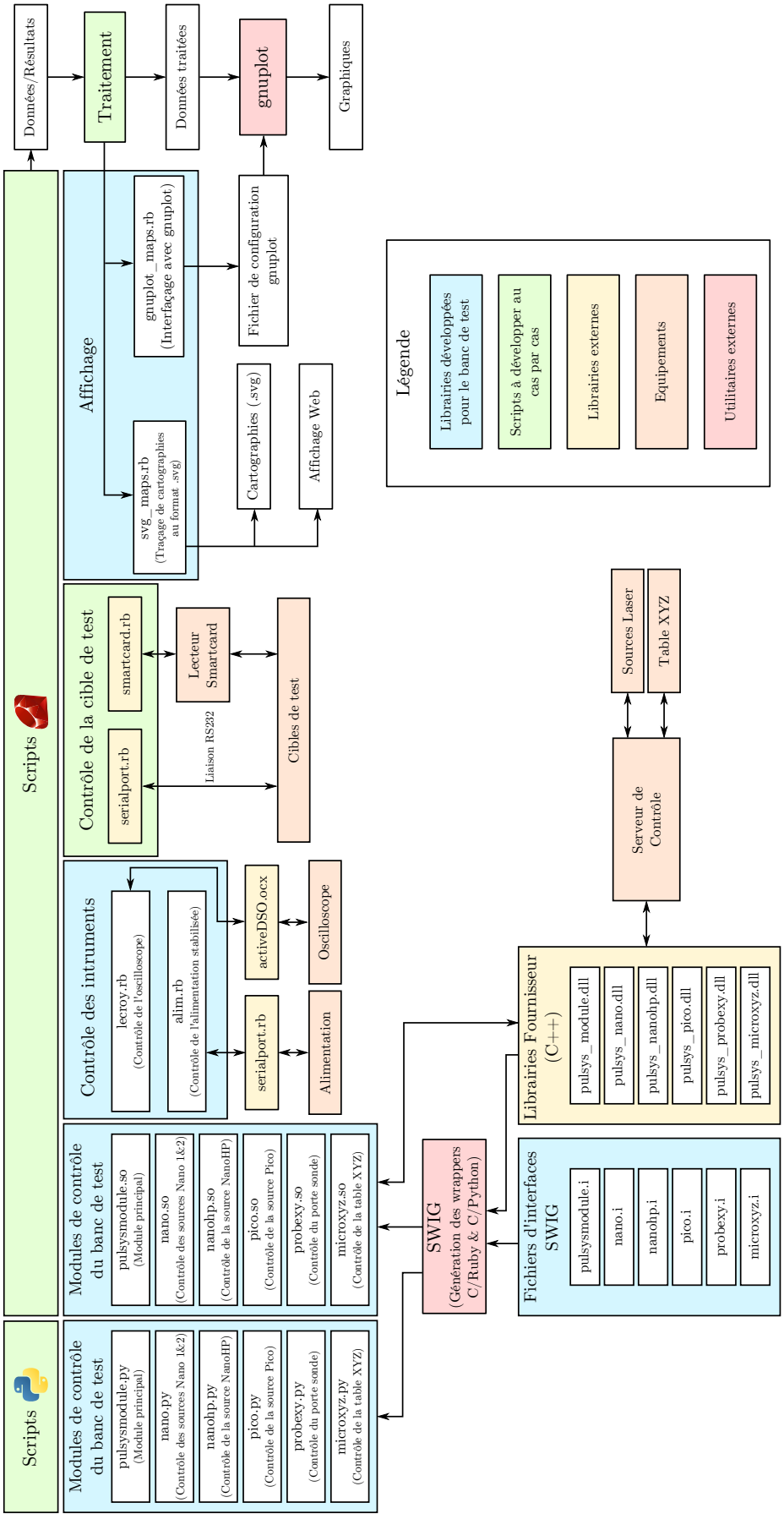
TCAD

Technology Computer Aided Design.

TCP/IP

Transmission Control Protocol/Internet Protocol, un protocole de communication réseau.

Annexe I : Architecture logicielle du banc de test



Liste des Communications

Publications

- M. Lacruche, N. Borrel, C. Champeix, C. Roscian, A. Sarafianos, J.-B. Rigaud, J.-M. Dutertre, and E. Kussener, "Laser fault injection into SRAM cells: Picosecond versus nanosecond pulses," in *On-Line Testing Symposium (IOLTS), 2015 IEEE 21st International*, pp. 13–18, 2015

- M. Lacruche, N. Beringuier-Boher, J.-M. Dutertre, J.-B. Rigaud, and E. Kussener, "On the use of forward body biasing to decrease the repeatability of laser-induced faults," in *Design, Automation and Test in Europe (DATE)*, vol. to be published, 2016

- N. Beringuier-Boher, M. Lacruche, D. El-Baze, J.-M. Dutertre, J.-B. Rigaud, and P. Maurine, "Body biasing injection attacks in practice," in *Proceedings of the Third Workshop on Cryptography and Security in Computing Systems*, pp. 49–54, ACM, 2016

- J.-M. Dutertre, P. Candelier, C. Champeix, S. DeCastro, G. DiNatale, M.-L. Flottes, M. Lacruche, M. Lisart, J.-B. Rigaud, C. Roscian, B. Rouzeyre, and A. Sarafianos, "Assessment of the laser-induced fault model towards continuous cmos technology shrinkage," in *Trustworthy Manufacturing and Utilization of Secure Devices (TRUDEVICE)*, vol. to be published, 2016

Posters

- Laser Fault Injection on SRAM Cells : Pulse duration impact (50ns vs 30ps), Marc Lacruche, Nicolas Borrel, Clément Champeix, Alexandre Sarafianos, *PHISIC Workshop, 2015*

- Influence of Low-Power Design Techniques on Fault Injection Vulnerability, Marc Lacruche, Jean-Baptiste Rigaud, Edith Kussener, *Journée de la Recherche Mines Saint-Etienne, 2013*

Bibliographie

[1] L. Zussa, A. Dehbaoui, K. Tobich, J.-M. Dutertre, P. Maurine, L. Guillaume-Sage, J. Clediere, and A. Tria, "Efficiency of a glitch detector against electromagnetic fault injection," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), 2014*, pp. 1–6, March 2014.

[2] F. Wang and V. D. Agrawal, "Single event upset : An embedded tutorial," in *VLSI Design, 2008. VLSID 2008. 21st International Conference on*, pp. 429–434, IEEE, 2008.

[3] J. S. Melinger, S. Buchner, D. McMorrow, W. J. Stapor, T. R. Weatherford, A. B. Campbell, and H. Eisen, "Critical evaluation of the pulsed laser method for single event effects testing and fundamental studies," *Nuclear Science, IEEE Transactions on*, vol. 41, no. 6, pp. 2574–2584, 1994.

[4] D. Lewis, V. Pouget, F. Beaudoin, P. Perdu, H. Lapuyade, P. Fouillat, and A. Touboul, "Backside laser testing of ICs for SET sensitivity evaluation," *Nuclear Science, IEEE Transactions on*, vol. 48, no. 6, pp. 2193–2201, 2001.

[5] F. Darracq, H. Lapuyade, N. Buard, F. Mounsi, B. Foucher, P. Fouillat, M.-C. Calvet, and R. Dufayel, "Backside SEU laser testing for commercial off-the-shelf SRAMs," *IEEE Transactions on Nuclear Science*, vol. 49, no. 6, pp. 2977–2983, 2002.

[6] A. Douin, V. Pouget, F. Darracq, D. Lewis, P. Fouillat, and P. Perdu, "Influence of Laser Pulse Duration in Single Event Upset Testing," *Nuclear Science, IEEE Transactions on*, vol. 53, no. 4, pp. 1799–1805, 2006.

[7] O. Kömmerling and M. G. Kuhn, "Design principles for tamper-resistant smartcard processors," in *USENIX workshop on Smartcard Technology*, vol. 12, pp. 9–20, 1999.

[8] 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 *On-Line Testing Symposium (IOLTS), 2015 IEEE 21st International*, pp. 150–155, July 2015.

[9] P. Dodd, "Physics-based simulation of single-event effects," *Device and Materials Reliability, IEEE Transactions on*, vol. 5, pp. 343–357, Sept 2005.

[10] a. Douin, V. Pouget, D. Lewis, P. Fouillat, and P. Perdu, "Electrical modeling for laser testing with different pulse durations," *11th IEEE International On-Line Testing Symposium*, pp. 1–5, 2005.

[11] A. S. Sarafianos, *Fault injections by laser impulsion in secured microcontrollers*. Theses, Ecole Nationale Supérieure des Mines de Saint-Etienne, Sept. 2013.

[12] M. Lacruche, N. Borrel, C. Champeix, C. Roscian, A. Sarafianos, J.-B. Rigaud, J.-M. Dutertre, and E. Kussener, "Laser fault injection into SRAM cells : Picosecond versus nanosecond pulses," in *On-Line Testing Symposium (IOLTS), 2015 IEEE 21st International*, pp. 13–18, 2015.

[13] A. P. Chandrakasan and R. W. Brodersen, *Low power digital CMOS design*. Springer Science & Business Media, 2012.

[14] J. M. Rabaey and M. Pedram, *Low power design methodologies*, vol. 336. Springer Science & Business Media, 2012.

[15] C. Roscian, A. Sarafianos, J.-M. Dutertre, and A. Tria, "Fault Model Analysis of Laser-Induced Faults in SRAM Memory Cells," in *Fault Diagnosis and Tolerance in Cryptography (FDTC), 2013 Workshop on*, pp. 89–98, IEEE, 2013.

[16] A. Kerckhoffs, "La cryptographie militaire," in *Journal des sciences militaires*, vol. 9, pp. 5–83, IEEE, 1883.

[17] J. Daemen and V. Rijmen, *The design of Rijndael : AES-the advanced encryption standard*. Springer Science & Business Media, 2013.

[18] NIST, "Fips 46 data encryption standard," January 1977.

[19] B. Schneier, "Description of a new variable-length key, 64-bit block cipher (blowfish)," in *Fast Software Encryption*, pp. 191–204, Springer, 1994.

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

[21] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[22] V. Miller, "Use of elliptic curves in cryptography," in *Advances in Cryptology CRYPTO85 Proceedings*, pp. 417–426, Springer, 1986.

[23] "Announcing the advanced encryption standard (aes)," in *Federal Information Processing Standards Publication 197*, National Institute of Standards and Technology, 2001.

[24] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Advances in Cryptology CRYPTO99*, pp. 388–397, Springer, 1999.

[25] J.-J. Quisquater and D. Samyde, "Eddy current for magnetic analysis with active sensor," in *Proceedings of Esmart*, vol. 2002, 2002.

[26] P. C. Kocher, "Timing attacks on implementations of diffie-hellman, rsa, dss, and other systems," in *Advances in Cryptology CRYPTO96*, pp. 104–113, Springer, 1996.

[27] L. Zussa, J.-M. Dutertre, J. Clédière, and A. Tria, "Power supply glitch induced faults on FPGA : an in-depth analysis of the injection mechanism," in *On-Line Testing Symposium (IOLTS), 2013 IEEE 19th International*, (Chania, Greece), July 2013.

[28] M. Agoyan, J.-M. Dutertre, D. Naccache, B. Robisson, and A. Tria, "When clocks fail : On critical paths and clock faults," in *Smart Card Research and Advanced Application*, pp. 182–193, Springer, 2010.

[29] A. Dehbaoui, J.-M. Dutertre, B. Robisson, P. Orsatelli, P. Maurine, and A. Tria, "Injection of transient faults using electromagnetic pulses-Practical results on a cryptographic system-.," *IACR Cryptology ePrint Archive*, vol. 2012, p. 123, 2012.

[30] S. P. Skorobogatov and R. J. Anderson, "Optical fault induction attacks," in *Cryptographic Hardware and Embedded Systems-CHES 2002*, pp. 2–12, Springer, 2003.

[31] P. Maurine, K. Tobich, T. Ordas, and P. Y. Liardet, "Yet Another Fault Injection Technique : by Forward Body Biasing Injection," in *YACC'2012 : Yet Another Conference on Cryptography*, (Porquerolles Island, France), 2012.

[32] N. Beringuier-Boher, M. Lacruche, D. El-Baze, J.-M. Dutertre, J.-B. Rigaud, and P. Maurine, "Body biasing injection attacks in practice," in *Proceedings of the Third Workshop on Cryptography and Security in Computing Systems*, pp. 49–54, ACM, 2016.

[33] D. Boneh, R. A. DeMillo, and R. J. Lipton, "On the importance of checking cryptographic protocols for faults," in *Advances in Cryptology EUROCRYPT97*, pp. 37–51, Springer, 1997.

[34] G. Piret and J.-J. Quisquater, "A differential fault attack technique against SPN structures, with application to the AES and KHAZAD," in *Cryptographic Hardware and Embedded Systems-CHES 2003*, pp. 77–88, Springer, 2003.

[35] T. Fuhr, E. Jaulmes, V. Lomné, and A. Thillard, "Fault attacks on AES with faulty ciphertexts only," in *Fault Diagnosis and Tolerance in Cryptography (FDTC), 2013 Workshop on*, pp. 108–118, IEEE, 2013.

[36] R. Lashermes, G. Reymond, J.-M. Dutertre, J. Fournier, B. Robisson, and A. Tria, "A dfa on aes based on the entropy of error distributions," in *Fault*

Diagnosis and Tolerance in Cryptography (FDTC), 2012 Workshop on, pp. 34–43, IEEE, 2012.

[37] C. Giraud, "Dfa on aes," in *Advanced Encryption Standard–AES*, pp. 27–41, Springer, 2005.

[38] C. H. Kim and J.-J. Quisquater, "New differential fault analysis on aes key schedule : two faults are enough," in *Smart Card Research and Advanced Applications*, pp. 48–60, Springer, 2008.

[39] C.-N. Chen and S.-M. Yen, "Differential fault analysis on aes key schedule and some countermeasures," in *Information Security and Privacy*, pp. 118–129, Springer, 2003.

[40] D. H. Habing, "The use of lasers to simulate radiation-induced transients in semiconductor devices and circuits," *Nuclear Science, IEEE Transactions on*, vol. 12, no. 5, pp. 91–100, 1965.

[41] S. P. Buchner, F. Miller, V. Pouget, and D. P. McMorrow, "Pulsed-Laser Testing for Single-Event Effects Investigations," *Nuclear Science, IEEE Transactions on*, vol. 60, no. 3, pp. 1852–1875, 2013.

[42] M. J. Gadlage, R. D. Schrimpf, J. M. Benedetto, P. H. Eaton, D. G. Mavis, M. Sibley, K. Avery, and T. L. Turflinger, "Single event transient pulse widths in digital microcircuits," *Nuclear Science, IEEE Transactions on*, vol. 51, no. 6, pp. 3285–3290, 2004.

[43] A. Sarafianos, M. Lisart, O. Gagliano, V. Serradeil, C. Roscian, J.-M. Dutertre, and A. Tria, "Robustness improvement of an SRAM cell against laser-induced fault injection," in *Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), 2013 IEEE International Symposium on*, pp. 149–154, IEEE, 2013.

[44] A. G. Yanci, S. Pickles, and T. Arslan, "Detecting voltage glitch attacks on secure devices," in *Bio-inspired Learning and Intelligent Systems for Security, 2008. BLISS'08. ECSIS Symposium on*, pp. 75–80, IEEE, 2008.

[45] D. El-Baze, J.-B. Rigaud, and P. Maurine, "A fully-digital em pulse detector," in *Design, Automation and Test in Europe (DATE)*, vol. to be published, 2016.

[46] J. Patel and L. Fung, "Concurrent error detection in alu's by recomputing with shifted operands," *Computers, IEEE Transactions on*, vol. C-31, pp. 589–595, July 1982.

[47] R. Lyons and W. Vanderkulk, "The use of triple-modular redundancy to improve computer reliability," *IBM Journal of Research and Development*, vol. 6, pp. 200–209, April 1962.

[48] M. Joye, P. Manet, and J.-B. Rigaud, "Strengthening hardware aes implementations against fault attacks," *IET Information Security*, vol. 1, no. 3, p. 106, 2007.

[49] G. Bertoni, L. Breveglieri, I. Koren, P. Maistri, and V. Piuri, "A parity code based fault detection for an implementation of the advanced encryption standard," in *Defect and Fault Tolerance in VLSI Systems, 2002. DFT 2002. Proceedings. 17th IEEE International Symposium on*, pp. 51–59, IEEE, 2002.

[50] P. Dodd, "Physics-based simulation of single-event effects," *Device and Materials Reliability, IEEE Transactions on*, vol. 5, pp. 343–357, Sept 2005.

[51] J. Rockett, L.R., "Simulated seu hardened scaled cmos sram cell design using gated resistors," *Nuclear Science, IEEE Transactions on*, vol. 39, pp. 1532–1541, Oct 1992.

[52] T. Calin, M. Nicolaidis, and R. Velazco, "Upset hardened memory design for submicron cmos technology," *Nuclear Science, IEEE Transactions on*, vol. 43, pp. 2874–2878, Dec 1996.

[53] T. Monnier, F. Roche, J. Cosculluela, and R. Velazco, "Seu testing of a novel hardened register implemented using standard cmos technology," *Nuclear Science, IEEE Transactions on*, vol. 46, pp. 1440–1444, Dec 1999.

[54] A. Sarafianos, R. Llido, J.-M. Dutertre, O. Gagliano, V. Serradeil, M. Lisart, V. Goubier, A. Tria, V. Pouget, and D. Lewis, "Building the electrical model of

the Photoelectric Laser Stimulation of a PMOS transistor in 90nm technology," *Microelectronics Reliability*, vol. 52, no. 9, pp. 2035–2038, 2012.

[55] A. Sarafianos, O. Gagliano, V. Serradeil, M. Lisart, J.-M. Dutertre, and A. Tria, "Building the electrical model of the pulsed photoelectric laser stimulation of an NMOS transistor in 90nm technology," in *Reliability Physics Symposium (IRPS), 2013 IEEE International*, pp. 5B—-5, IEEE, 2013.

[56] A. Sarafianos, C. Roscian, J.-M. Dutertre, M. Lisart, and A. Tria, "Electrical modeling of the photoelectric effect induced by a pulsed laser applied to an SRAM cell," *Microelectronics Reliability*, vol. 53, no. 9, pp. 1300–1305, 2013.

[57] M. Bajura, Y. Boulghassoul, R. Naseer, S. DasGupta, A. Witulski, J. Sondeen, S. Stansberry, J. Draper, L. Massengill, and J. Damoulakis, "Models and algorithmic limits for an ecc-based approach to hardening sub-100-nm srams," *Nuclear Science, IEEE Transactions on*, vol. 54, pp. 935–945, Aug 2007.

[58] A. Gibbons, A. Robert, K. Shi, M. Keating, and D. Flynn, "Low power methodology manual," 2007.

[59] H. B. Bakoglu, "Circuits, interconnections, and packaging for vlsi.," 1990.

[60] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital integrated circuits*, vol. 2. Prentice hall Englewood Cliffs, 2002.

[61] C.-Y. Tsui, M. Pedram, and A. M. Despain, "Efficient estimation of dynamic power consumption under a real delay model," in *Proceedings of the 1993 IEEE/ACM international conference on Computer-aided design*, pp. 224–228, IEEE Computer Society Press, 1993.

[62] D. Chinnery and K. Keutzer, *Closing the Power Gap Between ASIC and Custom*. Springer, 2007.

[63] J. T. Kao, M. Miyazaki, and A. P. Chandrakasan, "A 175-mv multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *Solid-State Circuits, IEEE Journal of*, vol. 37, no. 11, pp. 1545–1554, 2002.

[64] K. von Arnim, E. Borinski, P. Seegebrecht, H. Fiedler, R. Brederlow, R. Thewes, J. Berthold, and C. Pacha, "Efficiency of body biasing in 90-nm cmos for low-power digital circuits," *Solid-State Circuits, IEEE Journal of*, vol. 40, pp. 1549–1556, July 2005.

[65] C. Roscian, *Cryptanalyse physique de circuits cryptographiques Ã l'aide de sources LASER*. PhD thesis, 2013. ThÃ¨se de doctorat dirigÃ©e par Tria, Assia MicroÃ©lectronique Saint-Etienne, EMSE 2013.

[66] N. Borrel, *Evaluation d'injection de fautes Laser et conception de contre-mesure sur une architecture Ã faible consommation*. PhD thesis, 2015.

[67] E. Biham and A. Shamir, "Differential fault analysis of secret key cryptosystems," in *Advances in Cryptology-CRYPTO'97*, pp. 513–525, Springer, 1997.

[68] J.-M. DUTERTRE, *Robust Reconfigurable Circuits*. Theses, Université Montpellier II - Sciences et Techniques du Languedoc, Oct. 2002.

[69] F. Courbon, P. Loubet-Moundi, J. Fournier, and A. Tria, "Adjusting laser injections for fully controlled faults," in *Cosade 2014 Fifth International Workshop on Constructive Side-Channel Analysis and Secure Design*, 2014.

[70] M. Dworkin, "Recommendation for block cipher modes of operation. methods and techniques," tech. rep., DTIC Document, 2001.

[71] C. H. Kim, "Differential fault analysis against aes-192 and aes-256 with minimal faults," in *Fault Diagnosis and Tolerance in Cryptography (FDTC), 2010 Workshop on*, pp. 3–9, IEEE, 2010.

[72] M. Lacruche, N. Beringuier-Boher, J.-M. Dutertre, J.-B. Rigaud, and E. Kussener, "On the use of forward body biasing to decrease the repeatability of laser-induced faults," in *Design, Automation and Test in Europe (DATE)*, vol. to be published, 2016.

[73] K. Baddam and M. Zwolinski, "Evaluation of dynamic voltage and frequency scaling as a differential power analysis countermeasure," in *VLSI Design, 2007.*

Held jointly with 6th International Conference on Embedded Systems., 20th International Conference on, pp. 854–862, IEEE, 2007.

[74] S. Yang, W. Wolf, N. Vijaykrishnan, D. N. Serpanos, and Y. Xie, "Power attack resistant cryptosystem design : a dynamic voltage and frequency switching approach," in *Proceedings of the conference on Design, Automation and Test in Europe-Volume 3*, pp. 64–69, IEEE Computer Society, 2005.

[75] S. Ordas, M. Carbone, G. Ducharme, S. Tiran, and P. Maurine, "Efficiency of the rdvfs countermeasure," in *Faible Tension Faible Consommation (FTFC), 2014 IEEE*, pp. 1–4, IEEE, 2014.

[76] J.-M. Dutertre, P. Candelier, C. Champeix, S. DeCastro, G. DiNatale, M.-L. Flottes, M. Lacruche, M. Lisart, J.-B. Rigaud, C. Roscian, B. Rouzeyre, and A. Sarafianos, "Assessment of the laser-induced fault model towards continuous cmos technology shrinkage," in *Trustworthy Manufacturing and Utilization of Secure Devices (TRUDEVICE)*, vol. to be published, 2016.