



HAL
open science

Impact des technologies sur les architectures de calcul

Jacques-Olivier Klein

► **To cite this version:**

Jacques-Olivier Klein. Impact des technologies sur les architectures de calcul. Micro et nanotechnologies/Microélectronique. Université Paris Sud - Paris XI, 2009. tel-00408875

HAL Id: tel-00408875

<https://theses.hal.science/tel-00408875>

Submitted on 3 Aug 2009

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

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



Mémoire présenté pour l'obtention de l'
Habilitation à Diriger des Recherches

SPECIALITE : PHYSIQUE

par

Jacques-Olivier Klein

Impact des technologies
sur les architectures de calcul

TOME 1 : Bilan des activités et du parcours scientifique

Soutenue le 30 avril 2009 devant les membres du jury :

Jean-Philippe Bourgoïn,
Claude Chappert,
Patrick Garda
Bertrand Granado, rapporteur,
Ian O'Connor, rapporteur,
Lionel Torres, rapporteur.

Table des matières

A propos de ce document.....	5
------------------------------	---

I. Première partie : CV Détaillé	9
--	---

1. Etat Civil	9
2. Titres Universitaires	9
3. Parcours	9
4. Activités d'enseignements	10
5. Activités liées à l'administration	10
6. Activités liées à la recherche	10
7. Encadrement	12
8. Résumé de mes travaux de recherche.....	14
9. Publications	19

II. Synthèse des travaux de recherche de 1995 à 2004	31
--	----

1. Conclusion de ma thèse	31
2. Traitement d'image stochastique	33
3. Rétine électronique programmable	35
4. Synthèse d'architectures de calcul analogique à capacités commutées	38
5. Conversion thématique	39

II. Troisième partie : Logique à paroi de domaine magnétique	43
--	----

1. Introduction	43
2. Principe de fonctionnement.....	43
3. Composants élémentaires	44
4. Modélisation et simulation comportementales	47
5. Technique de synthèse de circuit à paroi de domaine	52
6. Conclusion	61

III. Quatrième partie : Architectures neuro-inspirées	65
---	----

1. Introduction	65
-----------------------	----

2.	Réseaux de neurones	65
3.	Conception et apprentissage d'un réseau utilisant des nanocomposants	73
4.	Vers des architectures auto-configurable	76
5.	Conclusion	80
IV.	Conclusion et perspectives.....	85
<hr/>		
1.	conclusion	85
2.	Axes de recherche	85
3.	Technologies magnétiques	86
4.	Opportunités architecturales	87
V.	Références bibliographiques	91
<hr/>		

A propos de ce document

Ce document est organisé en 5 parties.

La première partie est un CV détaillé. Il résume mon parcours de chercheur depuis ma thèse.

La deuxième partie décrit mon parcours scientifique au sein du département AXIS de l'IEF, jusqu'en 2005.

La troisième et la quatrième parties constituent des développements scientifiques autour des thèmes traités au sein du département NST de l'IEF : logique magnétique et architectures neuro-inspirées pour les nanocomposants.

La cinquième partie conclue ce travail et présente mes perspectives de recherche.

Un recueil des publications représentatives de mon parcours scientifique constitue le tome 2 de mon mémoire d'HDR.

Publications	Françaises	Internationales
Articles	2	8
Conférences	25	23 (1 invités)
Brevet	1	

Thèses	soutenue	encadrement
A. Nshare	5/09/02	50%
S. Moutault	non soutenue	100% (derog. HDR)
M. He	15/10/08	100% (derog. HDR)
A. Elouardi	25/05/05	20%

Principales Responsabilités
Responsable de l'opération NST Nanoarchi à l'IEF
Animateur du thème «technologies émergentes» du GDR SOC SIP
Coordinateur du projet ANR PANINI
Responsable du semestre 4 (IUT GEii1 CACHAN)

Première partie

CV détaillé

I. Première partie : CV Détaillé¹

1. Etat Civil

Nom	KLEIN
Prénom	Jacques-Olivier
Date de naissance	27 septembre 1967
Situation actuelle	Maître de conférences à l'IUT de Cachan Université Paris-Sud 11
Laboratoire	Institut d'Electronique Fondamentale (IEF) Université Paris-Sud 11, 91405 Orsay cedex
E-mails	Jacques-Olivier.Klein@u-psud.fr
Page web	http://www.ief.u-psud.fr/~jok/ ou http://joklein.free.fr
Adresse postale	4 rue Vincent Van Gogh 91440 Bures-Sur-Yvette

2. Titres Universitaires

Diplômes obtenus depuis le Baccalauréat

- 1987 : DEUG A, Université Paris-Sud 11,
- 1988 : Licence EEA, Université Paris-Sud 11,
- 1989 : Maîtrise EEA, Université Paris-Sud 11,
- 1990 : DEA d'Electronique, Université Paris-Sud 11,
- 1995 : Doctorat, Université Paris-Sud, titre de la thèse : Contribution à l'étude de l'adéquation algorithme-architecture : Machines de Boltzmann et circuits analogiques cellulaires, Directeur : Patrick Garda, soutenue le 5 janvier 1995, à Orsay.

3. Parcours

- 1991 - 1994: Doctorant à l'IEF (Université Paris-Sud) et moniteur CIES,
- 1994 - 1995 : ATER à l'IUT de CACHAN, Université Paris-Sud,
- 01/09/1995 : Maître de conférences stagiaire, IUT de Cachan,

¹ Le titre et le plan de cette partie suivent les consignes fournies par l'université Paris-Sud.

- 01/09/1996 : Titularisation, maître de conférences deuxième classe,
- 1996 – 2000 : Contrat pédagogique,
- 1/9/2005 - 1/2/2006 : 6 mois de CRCT (demi-service). Changement de thématique et de département,
- 1/2/2007 : Passage au sixième échelon des Maîtres de Conférences Classe Normale,
- 2008 - 2009 : Délégation au CNRS.

4. Activités d'enseignements

- Responsable des enseignements d'informatique industrielle de 1999 à 2004 en DUT GEii1 à Cachan (semestres 2 et 3), mise en place deux nouvelles cibles (Coldfire 5204 et MC9S12), développement de cartes à microcontrôleur, mise en place de nouveaux outils (Codewarrior), développement de bibliothèques de pilote de périphérique, rédaction des documentations et des supports de cours, cours magistraux, rédaction des sujets de TD, TP, projets et sujets d'examen.
- Création et responsabilité des projets de robotique du semestre 1 du DUT GEii de Cachan depuis 2004. Organisateur du concours de robotique commun GEii1 - GEii2 en fin de semestre 1 à l'IUT de Cachan (Le Gamel-Trophy).
- Enseignant en informatique industrielle (semestre 2) notamment en VHDL. Rédaction de sujets de TD, TP et mini-projets.
- Création (janvier 2000) et responsabilité du parcours «Robotique et noyaux temps réel» dans le semestre 4 du DUT GEii de Cachan. Portage d'un noyau temps réel et développement d'un générateur de code associé.
- Membre du comité d'organisation du concours national de robotique des IUT GEii. Rédacteur du règlement du concours 2008.

5. Activités liées à l'administration

- **Responsable pédagogique du semestre 4** du DUT GEii à Cachan en 2004 et 2005, puis à nouveau depuis 2007 (gestion des services enseignant, emploi du temps, planning, jury, groupes, parcours, contrôle de connaissance, coordination pédagogique).
- Membre de la CSE 63 d'avril 2002 à mars 2003, puis à nouveau depuis mars 2007.
- Membre élu du conseil de département de GEii-1 de l'IUT de Cachan depuis 2003,
- Membre de la commission consultative de choix des enseignants de l'IUT de CACHAN depuis 2003,

6. Activités liées à la recherche

6.1. Prix

- Prix Jean Demal 2006 pour une communication au congrès de l'AIPU (Association International de Pédagogie Universitaire)

6.2. Participation à des comités, organisation de colloques

- Responsable du pôle Ile de France du réseau doctorale en architecture des systèmes et machines informatiques dès sa création en 1992.
- Participation au comité de programme des journées communes PRC/GDR Architectures Nouvelles de Machines et jeunes chercheurs à Rennes en 1994.
- Participation au comité de lecture pour Neuro-Micro en 1994.
- Relecture pour un numéro spécial CNN de «journal of circuits, systems and computers» en 2002.
- Relecture pour ISCAS 2004 (IEEE International Symposium on Circuits and Systems) en 2004.
- Relecture de trois papiers pour DCIS 2007 (Conference on Design of Circuits and Integrated Systems) en 2007.
- Relecture de quatre papiers pour DCIS 2008 (Conference on Design of Circuits and Integrated Systems) en 2008.
- Relecture d'un papier pour la revue IJRC (International Journal of Reconfigurable Computing) en 2008.
- Relecture de sept papiers pour le colloque nationale du GDR SOC SIP (System on Chip / System in Package) en 2008.
- Organisation de la session orale et de la session poster «Technologies Emergentes» lors du colloque nationale du GDR SOC SIP (System on Chip / System in Package) en 2008.
- • Relecture d'un papier pour la revue IEEE-TRANSMAG (Transactions on magnetics) en 2008.

6.3. Programmes d'échanges, collaborations, réseaux internationaux, projets nationaux et européens,

- Participation au groupe de travail « rétine » (GT 7.3) dans le cadre du GDR ISIS (1998-2001).
- Participation à l'ACI SoC-NT du CNRS (2004),
- Participation à l'AC Nanosciences (Nanosys et CALOMAG) du CNRS jusqu'en 2006.
- Participation au projet européen MAGLOG (Project Exploring the Frontiers of Magnetic Logic FP6-500021), notamment en collaboration avec l'Imperial College de Londres.
- **Animateur du thème «Technologies Emergentes» du GDR SOC SIP depuis fin 2007.**

6.4. Actions de valorisation

- Brevet français : «Synapse nano-électrique et procédé d'apprentissage d'une telle synapse», Klein, J.-O. et Belhaire, E., déposé le 27 juillet 2007, n° 07 05532.

6.5. Administration liée à la recherche

- **Coordinateur du projet ANR-PANINI** (Programme Architectures Nanoélectroniques Intégrées Neuro-Inspirées), sélectionné par l'ANR Architecture du futur en septembre 2007.
- **Responsable de l'opération NanoArchi** (Architectures à base de nanocomposants) dans le département NST (NanoSpintronique) de l'IEF depuis décembre 2007.

7. Encadrement

7.1. Direction de thèse (avec dérogation à l'HDR).

- **Sébastien Moutault**, en thèse de 2001 à 2004. Sujet : Synthèse d'architecture reconfigurable de calcul analogique. Directeur de thèse : J.O.KLEIN. Publications : 4 conférences internationales avec actes et comité de lecture [A4,A6,A8,A12] et 5 communications lors de conférences nationales [N6,N9,N10,N11,N12]. [Note : Sébastien Moutault, a été recruté par l'IUT de Cachan en tant qu'enseignant à temps plein (PRAG) en septembre 2003, alors que sa thèse était en cours (début de la troisième année). Il a tenté, en vain, de mener de front son service d'enseignement à plein-temps (384h) et la rédaction de sa thèse. Malgré des résultats prometteurs, il a renoncé à soutenir fin 2004. Il est aujourd'hui PRAG à l'IUT de Bordeaux.]
- **Michel Hé**, thèse en cours débutée le 01/09/2005, soutenance prévue le 15 octobre 2008. Sujet : Contribution à l'étude de l'impact des nanotechnologies sur les Architectures : Apprentissage d'inspiration neuronale de fonctions logiques pour circuits programmables. Directeur de thèse : J.O.KLEIN. Publications : 1 revue internationale avec comité de lecture [R8], 3 conférences internationales avec actes et comité de lecture [A19,A22, A23] et 3 posters lors de conférences nationales [C9,C10,N26].

7.2. Co-encadrement de thèse

- Co-encadrement de la thèse d'**Abdallah Nshare à 50%** avec Antoine Dupret, soutenue le 5 septembre 2002, Titre : Définition et conception d'une nouvelle génération de rétines programmables. Directeur : Antoine Dupret. Publications : 1 revue internationale à comité de lecture [R2], 3 conférences internationales avec actes et comité de lecture [A2,A3,A6] et 4 communications dans des conférences nationales [N7,N8,N11]. Abdallah Nshare est aujourd'hui à la tête d'une entreprise de conception microélectronique.
- Participation à l'encadrement d'**Abdelhafid Elouardi** (à hauteur de 20%) dans le cadre d'une collaboration avec l'équipe Véhicules Intelligents (IEF/AXIS). Sujet: Évaluation des rétines électroniques pour une définition architecturale d'un système monopuce (SoC) dédié à la vision embarquée. Direction : Roger REYNAUD et Samir BOUAZIZ. Publications communes avec le doctorant : 3 revues internationales à comité de lecture [R3,R5,R6], 9 conférences internationales avec actes et comité de lecture [A7, A9, A10, A11, A13, A14, A15, A17, A21] et 3 communications dans des conférences nationales [C4,N13,N14]. Abdelhafid Elouardi est aujourd'hui maître de conférences à l'université Paris-Sud.

7.3. Encadrement de stage de DEA

- **Stéphane Hafrad**, stage de DEA, co-encadrement à 50% avec H. Pujol. Titre : Evaluation d'un réseau d'interconnexion haut débit pour machines connexionnistes, septembre 1994, direction P. Garda.
- **Abdallah Nshare**, stage de DEA, co-encadrement à 50% avec A. Dupret. Titre : Etude et conception d'un amplificateur à transconductance pour les architectures cellulaires, septembre 1999. Publications : 1 Poster dans un congrès national avec actes et comité de lecture, [N6].
- **Sébastien Moutault**, stage de DEA, co-encadrement à 50% avec A. Dupret. Titre : Méthodologie d'analyse de performances pour les rétines artificielles, septembre 2000.
- **Louis Barthelemy Faber**, stage de Master recherche, co-dirigé à 50% avec Weisheng Zhao. Titre : Compact model of magnetic tunnel junction based on a spin transfer torque mechanism, du 1er mars au 31 août 2008.

7.4. Autres encadrements

- **Belem Brahima**, stage DUT, Titre : Etude de l'effet de la corrélation des échantillons aléatoires dans un algorithmes stochastique, avril-juin 1996. Publication : 1 poster dans une conférence nationale [N5]. Belem Brahima a poursuivi ses études en Angleterre et obtenu un PhD de l'université de Glamorgan en 2004.
- **Sébastien Moutault**, Stage de Magistère EEA, Titre : Contribution à la définition d'une méthode de mise en œuvre de circuits cellulaires de vision : Réalisation de l'interface matériel/Logiciel, juin 1997.
- **Sébastien Moutault**, stage de maîtrise EEA. Titre : implantations d'algorithmes de traitement d'image sur la rétine PARIS, juin-juillet 1998.
- **Michael Helie**, stage DUT. Titre : Etude et développement d'une plate-forme de test pour les capteurs intelligents, avril-juin 1999.
- **Nicolas Martin**, stage DUT. Titre : Mise en œuvre d'algorithmes de traitement d'images sur une architecture programmable de rétine analogique intégrée, juin-2000.
- **Olivier Gavinet**, stage DUT. Titre : Synthèse de microcontrôleurs et microprocesseurs open source en technologie CMOS, avril-juin 2001.
- **David Robillard**, stage Fiupso 2ème année. Titre : Définition d'un protocole de communication avec la rétine électronique PARIS développé à l'IEF, juin-août 2003,
- **Matthieu Simon**, Stage de licence pro de robotique. Titre : Contribution à la mise en œuvre d'un système de vision pour la robotique, juin 2004. Matthieu Simon a obtenu un diplôme d'ingénieur de l'université Paris 6.
- **Florent Ouchet**, stage de Magistère EEA, 15/06/2006 au 31/07/2006 co-encadré à 50% avec Eric Belhaire, Sujet : Modélisation de fonctions logiques à base de domaines magnétiques, Publication : 1 communication à une conférence internationale avec actes et comité de lecture [A20].

8. Résumé de mes travaux de recherche

8.1. Introduction

Les architectures dédiées au calcul sont très largement dominées par le couple constitué d'un côté par le modèle de Von Neumann et de l'autre côté par la technologie CMOS. Cette prédominance est telle qu'elle s'impose aujourd'hui avec une évidence écrasante face aux propositions alternatives. La généricité du modèle de programmation alliée à l'efficacité, notamment d'un point de vue énergétique, de la technologie CMOS donne à ce couple un caractère tellement dominant qu'en sortir serait considéré aujourd'hui comme une rupture de paradigme. C'est pourtant autour des technologies alternatives et des modèles de programmation différents que portent les travaux présentés ici. L'étude de l'impact des premières sur les secondes en constitue le fil directeur.

Au départ, il s'agit de travaux essentiellement fondamentaux, portant sur des applications qui peuvent être qualifiées de niches. Ainsi, dans la deuxième partie de ce manuscrit, je présente les travaux réalisés au sein du département AXIS de l'IEF. Ils sont essentiellement axés sur les architectures bio-inspirées, de réseaux de neurones et de rétines. Ces architectures sont associées à des éléments de calcul analogiques. Dans un premier temps, elles font appel à une organisation relativement classique d'un point de vue des circuits analogiques puisqu'elles sont constituées par un réseau quasi-statique d'opérateurs. Ensuite, elles s'inspireront davantage des architectures, de la programmation et des méthodes de synthèse des systèmes numériques.

Cependant, à mesure que la fin de la loi de Moore se profile, les recherches sur les technologies alternatives concernent de plus en plus l'ensemble du secteur de la micro-nano-électronique et touchent à des marchés énormes. Aujourd'hui, compte tenu du rôle de ce secteur dans l'industrie et dans toute l'économie, cette ambition constitue un enjeu de société majeure, mais toujours un objectif scientifiquement aussi risqué tant il reste difficile de concurrencer le couple en place.

Dans les deux parties suivantes, je présenterai les résultats des travaux menés au sein du département NST de l'IEF à partir de l'année 2005. Dans ce cadre nous avons créé une opération «Architectures pour les nanocomposants», qui vise à concevoir des architectures basées sur des éléments nanométriques, magnétiques ou non magnétiques. La troisième partie de ce manuscrit porte alors sur l'impact en termes d'architecture, du choix de l'aimantation comme grandeur support de l'information dans les circuits à paroi de domaine magnétique. Enfin, je revendrai dans la quatrième partie sur l'approche neuronale, non plus en tant que modèle de programmation alternatif mais principalement comme source d'inspiration pour imaginer les procédures qui permettent d'appliquer un apprentissage à un ensemble de nanocomposants de façon à progresser vers la fonctionnalisation d'une assemblée auto-organisée fabriquée par une approche dites «bottom-up». Outre l'intérêt d'une telle procédure pour compenser les défauts et dispersions de caractéristiques, cet objectif permettrait d'envisager les plus hautes densités d'intégration sans payer le prix pharaonique projeté pour les équipements et les usines de fabrication en CMOS ultime.

8.2. Thèse à l'Institut d'Électronique Fondamentale,

1/10/1991-5/1/1995, Titre : Contribution à l'étude de l'adéquation algorithme-architecture : Machines de Boltzmann et circuits analogiques cellulaires, Directeur P. Garda, soutenue le 5 janvier 1995, Université Paris Sud XI, Orsay, devant le Jury : René Castagné (Président, IEF), Daniel Etiemble (Examinateur, LRI), Patrick Gallinari (Rapporteur, Paris 6), Patrick Garda (Directeur, IEF), Sylvie Le Masson (Examinatrice, Bordeaux 1), Eric Martin (Rapporteur, Univ. Bret. Sud).

J'ai effectué ma thèse à l'Institut d'Electronique Fondamentale sous la direction de Patrick Garda en contribuant au projet d'implantation analogique de réseaux de neurones stochastiques (projet RA) [T-A1]. Ma contribution s'est portée principalement sur l'étude architecturale et algorithmique. J'ai commencé par une évaluation de performance pour la simulation de réseaux de neurones en général et de la machine de Boltzmann en particulier afin d'identifier les réseaux qui nécessitent une architecture dédiée. Contrairement aux idées reçues, nous avons montré que **les stations de travail récentes (de l'époque) suffisaient à simuler bien des réseaux de neurones, même itératifs et stochastiques** [T-N1,T-RN1]. Les architectures spécialisées permettent d'aller encore plus loin en répondant aux exigences extrêmes par la quantité de données manipulées ou par les contraintes temporelles: traitement d'image neuronal ou applications à temps de réponse ultracourts. Dans ces derniers cas, j'ai montré que **le calcul analogique profite de la faculté d'auto-compensation des algorithmes d'apprentissage** neuronaux [T-N4]. Cet effet de **compensation** peut être obtenu par un apprentissage on-chip qui consomme de la surface de silicium sur le circuit, mais, moyennant le développement d'un **algorithme d'apprentissage** adapté [T-R1], il peut également être obtenu en **incluant le circuit analogique dans une boucle d'apprentissage** dont l'essentiel est réalisé off-line. Nous avons aussi démontré, avec Hubert Pujol, qu'un réseau d'interconnexion composé de **bus sécable était adapté aux diffusions multiples** nécessaires pour ce type d'architectures [T-N2,T-N3].

8.3. Traitement d'image stochastique (1996-1998)

L'analyse de l'effet de la corrélation des générateurs aléatoires sur la qualité de l'optimisation réalisée par un algorithme stochastique est le premier thème de recherche que j'ai initié après ma thèse. Nous avons choisi le modèle des verres de spin pour lequel la dynamique stochastique est connue, et rendue nécessaire par la présence évidente de minima locaux du fait de la forme du champ externe. Si la **corrélation spatiale est rapidement néfaste**, une **corrélation temporelle n'affecte** de façon significative la vitesse de convergence vers la solution optimale et **la qualité du résultat obtenu qu'au delà de valeur très importante de la corrélation** [N5]. Nous avons également montré que les réseaux faisant intervenir des cliques d'ordre supérieur à deux et une dynamique synchrone peuvent être utilisés, y compris pour minimiser des énergies quadratiques, dès lors qu'on évite les rares cas où la dynamique synchrone conduit à des oscillations stables. En conséquence, des architectures mixtes, analogiques/numériques particulièrement efficaces peuvent tirer profit des performances des algorithmes stochastiques en traitement d'images. Nous avons notamment démontré qu'**un maillage carré résistif associé à des interrupteurs commandés par un algorithme stochastique permet de fermer des contours, y compris sur des images extrêmement bruitées** [RN2].

8.4. Calcul analogique programmable à capacité commuté (1999-2004)

Travaux de Thèse d'A. Nshare, co-encadré avec A. Dupret.
et collaboration avec l'opération Véhicule intelligent
dans le cadre de la Thèse d'Abdelhafid Elouardi.
Travaux de thèse de S. Moutault.

Persuadé que le caractère figé des traitements analogiques, même s'il peuvent se révéler efficaces sur des images très bruitées grâce à la dynamique stochastique, ne laisse pas assez de degré de liberté, donc pas assez de champ d'innovation, pour in-

téresser les traiteurs d'images et soucieux d'inclure un mécanisme de virtualisation des éléments de calcul analogique, j'ai choisi d'initier un projet de rétine programmable. J'ai proposé une architecture de processeur analogique programmable inspirée des intégrateurs à capacités commutées [N6]. Avec Antoine Dupret, nous avons encadré la thèse d'Abdalha Nshare. Il a montré que l'intégration de minuscules processeurs de calcul analogiques programmables en bordure d'une matrice d'acquisition et de mémorisation d'images permet d'effectuer des algorithmes de vision performants dans des circuits à très faible coût et à faible consommation. Pour cela, nous avons réalisé un premier circuit de validation (PARIS 1) [R2, A2, N7]. Nous avons ensuite optimisé les mémoires analogiques et les circuits de calcul analogiques de façon à minimiser l'effet des capacités parasites [A3,A6,A8]. Notre circuit a servi de support de **collaborations** et d'expérimentations, notamment en collaboration avec l'opération «**véhicules intelligents**» de l'IEF lors de la thèse d'Abdelhafid Elouardi. Il a ainsi **caractérisé notre circuit pour le comparer à l'état de l'art** des systèmes de vision intégrés et a proposé son intégration au sein d'un système temps réel pour l'utiliser dans un contexte d'assistance à la conduite de véhicule [R3,R5,R6].

L'optimisation de l'opérateur de calcul à capacités commutées dans le contexte des architectures de rétines nous a permis de prendre conscience de son universalité, ce qui ouvrait la voie vers la synthèse automatique d'architecture de calcul analogique [A4, N9, N10, N11]. A partir d'une architecture de calcul analogique générique, il est possible d'intégrer des opérateurs de traitement analogique dans une chaîne de synthèse automatique. Nous avons montré qu'un **seul opérateur analogique à capacités commutées, suffisamment universel, permet d'envisager la synthèse automatique** de circuits de calcul analogiques. Nous avons défini l'architecture matérielle adaptée à ce type de traitement et précisé les étapes permettant d'arriver à la synthèse d'algorithmes sur ces architectures, comme le fait l'approche A³ (Adéquation Algorithmes Architecture) pour les circuits numériques programmables.

8.5. Reconversion thématique (2005)

Durant l'année 2005 l'IEF a connu une restructuration touchant particulièrement le département AXIS et conduisant à la disparition de l'opération «rétine» à laquelle j'appartenais jusque là. A cette période, j'ai profité de six mois de CRCT qui m'ont permis d'effectuer une reconversion thématique en **rejoignant le département NST** (Nano-Spin-Tronique), avec Eric Belhaire et Jean-François Pône pour y créer une **nouvelle opération NanoArchi** (Architecture à base de nano-composants). Là, nous avons initié trois thématiques de recherche : l'une sur les architectures à base de MRAM, sujet de thèse de Weisheng Zhao, suivi par Eric Belhaire et Claude Chappert, une seconde sur les architectures de circuits neuro-inspirés utilisant des nanocomposants, sujet de thèse de Michel Hé dont je suis le directeur de thèse et que nous avons suivi ensemble avec Eric Belhaire et une troisième sur la logique à base de parois de domaine magnétique, sujet que j'ai directement pris en charge.

8.6. Logique à parois de domaine magnétique (2005 - ...)

Projet Européen MAGLOG, Collaboration avec le Collège Impérial de Londres.

Dans le cadre du **projet Européen MAGLOG**, nous collaborons avec Russell Cowburn de l'imperial College de Londres. De petits circuits faisant propager des parois de domaines magnétiques en utilisant un champ magnétique externe tournant dans le plan du circuit, ont été démontrés par Russell Cowburn [7]. Ces circuits présentent l'avantage d'une très grande simplicité de réalisation comparée à la complexité et au coût de fabrication des circuits intégrés électroniques ainsi qu'une très faible puissance

dissipée. Dans ce contexte, la problématique, était de construire une méthodologie qui permette de concevoir simplement et sûrement des circuits complexes. Je me suis d'abord attaché à définir les méthodes de simulations adaptées à ce type de circuits de façon à pouvoir vérifier la validité des architectures proposées. Pour cela, j'ai défini une **modélisation comportementale à l'échelle du tronçon** de circuit [R4,A16]. A partir de cette modélisation, j'ai cherché à définir une méthode de conception permettant de traduire une fonction logique en circuit magnétique. J'ai d'abord montré que les méthodes de synthèse inspirées d'une traduction «porte à porte» des architectures CMOS vers leur équivalent magnétique constituait une impasse [A18]. En dépit de ces difficultés, j'ai montré comment resynchroniser les signaux et j'ai présenté une méthode permettant d'effectuer la **synthèse de n'importe quelle machine d'état** [A20,R7].

8.7. Architecture de systèmes à base de nanocomposants (2005 – ...)

Projet Nanosys (AC CNRS Nanosciences)

Travaux de thèse de Michel Hé (direction J.O.Klein)

et de Mathilde Joly (Collaboration Univ. Paris 6, direction P. Garda)

Collaboration avec l'IMS-Bordeaux (anciennement IXL)

Sous l'impulsion du CNRS, 17 équipes intéressées par l'impact des nanotechnologies sur les architectures se sont réunies dans le groupe Nanosys, de l'AC Nanosciences, groupe animé par Eric Belhaire. Ce groupe visait à impulser une recherche pluridisciplinaire, allant de la technologie aux architectures de machines parallèles en passant par la modélisation et la conception. Au cours des réunions, il est apparu une convergence d'intérêt vers les transistors à nanotubes de carbones (CNTFET) leur caractérisation, leur modélisation et leur utilisation dans des circuits numériques ou analogiques. Par ailleurs, la nécessité de trouver des paradigmes de calcul en rupture avec les architectures classiques de processeurs pour prendre en compte les spécificités des nanocomposants ont donné un regain d'intérêt aux réseaux neuronaux. J'ai choisi de viser l'**implantation neuronale de fonctions logiques génériques, à la manière des LUT dans les FPGA**. Par soucis de cohérence avec le reste des participants de Nanosys, nous avons choisi d'utiliser des **CNTFET comme éléments actifs** dans nos circuits et profitons ainsi de l'expertise de nos collègues et des modèles de simulation les plus récents. Par ailleurs, un nanocomposant linéaire présentant un effet mémoire est nécessaire pour réaliser l'apprentissage des synapses qui connectent les neurones. Nous avons choisi des dipôles résistifs présentant des possibilités de configuration continue de leur état de conductance [A22]. Nous avons modélisé la dynamique de la variation de leur conductance de la façon la plus réaliste possible compte tenu du peu de données expérimentales dont nous disposons à ce sujet aujourd'hui. **Nous avons alors proposé une architecture et conçu les circuits permettant de réaliser l'apprentissage de fonctions logiques**, en se limitant pour l'instant aux fonctions linéairement séparables [R8,A23]. La complexité matérielle induite par l'électronique ajoutée pour réaliser l'apprentissage, même si elle peut être multiplexée entre plusieurs synapses, reste relativement importante au regard du gain que peut apporter la grande densité d'intégration des nanocomposants. Nous suivons donc, dès maintenant, d'autres pistes [X1]. Nous avons notamment mis en évidence l'intérêt que peut présenter un composant résistif dont la variation de la conductance serait liée à l'amplitude de la tension ou du courant indépendamment de sa polarité (ou du sens du courant). Le principe d'un tel apprentissage, qui peut alors être réalisé en parallèle sur l'ensemble du réseau pour un coût très réduit en termes de surface a fait l'objet d'un dépôt de brevet [B1]. Il pourrait, par exemple s'appliquer à un nanotube multiparois dont la résistance est augmentée par pelage successif des nanotubes les plus externes, mais nous

chercherons également des solutions réversibles, par exemple en utilisant des CNTFET à grille optique. Ce projet se poursuit maintenant avec le CEA-LIST, l'IXL-Bordeaux, le CEA-LEM et le LE2I-Clermont Ferrand dans le cadre de l'ANR-PANINI dont je suis coordinateur.

8.8. Projet de recherche (2008- ...)

Les travaux de recherche auxquels je vais participer dans les prochains mois et les prochaines années sont principalement liés à des projets financés par l'ANR démarrés en 2007 ou qui démarrent cette année 2008.

Je suis coordinateur du projet **ANR-PANINI** (Programme Architectures Nano-électroniques Intrégrées Neuro-Inspirées). Ce projet, démarré en 2008, vise d'abord à développer des procédés d'apprentissage neuronale adaptés aux composants nanoélectroniques puis à étudier les possibilités ainsi offertes pour de nouvelles architectures. Il regroupe des équipes des 5 laboratoires suivants : Le CEA-LEM, pour la fabrication et la caractérisation d'éléments nanoélectroniques présentant des capacités de mémorisation, l'IMS-Bordeaux, pour le développement et l'adaptation des modèles de simulation des nano-composants, l'IEF-Orsay (coordonateur du projet), pour le développement de méthodes et circuits d'apprentissage, le CEA-LIST et LE2I pour l'application de blocs neuronaux au traitement d'image et leur utilisation dans de nouvelles architectures. Naturellement, mon rôle sera de coordonner le projet et de diriger les recherches effectuées à l'IEF. Il s'agit de **développer des procédés d'apprentissage compatibles avec des nanocomposants auto-assemblés** de façon à leur donner une fonction logique (programmation). L'utilisation de procédures de programmation inspirées des algorithmes d'apprentissage neuromimétiques permet de prendre en compte et de corriger les fortes dispersions de caractéristiques attendues dans les composants nanométriques (auto-correction). La redondance des architectures neuronales apporte également une grande tolérance vis-à-vis des défauts (robustesse). Enfin, à plus long terme, le principe même de l'apprentissage des réseaux de neurones, global et distribué, est aujourd'hui le seul qui puisse être envisagé lorsqu'on ne peut pas sélectionner individuellement les composants au sein d'une assemblée autoorganisée, notamment obtenue par une **approche bottom-up** (auto-configuration).

Par ailleurs, dans le cadre de l'**ANR CILOMAG**, je poursuivrai avec Claude Chappert les travaux jusqu'ici dirigés par Eric Belhaire (qui a quitté le laboratoire le 1er janvier 2008). Dans ce cadre, nous travaillerons à la **modélisation et à la conception mixte CMOS-magnétique de composants et de circuits basés sur les jonctions tunnel magnétiques**. Cette conception mènera à développer des cellules adaptées à la logique programmable (FPGA) que ce soit pour mémoriser les configurations des connexions ou pour durcir les FPGA en échantillonnant périodiquement dans une mémoire non-volatile leur état interne. Elle doit également conduire à **développer** avec les autres partenaires du projet CILOMAG **une filière magnétique** susceptible d'être ouverte aux laboratoires par l'intermédiaire du CMP.

Enfin, je poursuivrai dans les mois et les années à venir l'étude de l'architecture des circuits à parois de domaines magnétiques pour la logique et la mémorisation. Le démarrage de ces travaux, effectué en collaboration avec l'Imperial College de Londres au sein du projet Européen MAGLOG, portait sur la synthèse de fonctions logiques à parois de domaine magnétique. Le fonctionnement de ces circuits à aimantation dans le plan requiert des champs magnétiques tournants, elliptiques et asymétriques complexes. Au sein de l'**ANR PNANO ISTRADE**, nous poursuivrons cette étude en montrant l'intérêt d'une **aimantation perpendiculaire** qui simplifie la génération du champ et j'étudierai les nouvelles possibilités architecturales apportées par l'utilisation d'un **courant polarisé en spin** pour le déplacement ou le dépiégeage d'une paroi de domaine

localisée ; cette action locale permettant un adressage individuel des signaux. Enfin, je participerai à l'animation du **GDR SoC SiP** du CNRS, notamment par l'animation du nouveau thème «technologies émergentes».

9. Publications

Note : Les numéros des publications liées à mes travaux de Thèse sont préfixés par la lettre «T-»

9.1. Tableau récapitulatif de toutes les publications

- **10 revues** avec comité de lecture dont **8 internationales**,
- **1 brevet** français déposé à l'INPI, (extension PCT en cours)
- **23 communications** dans des conférences internationales dont 22 avec actes et comité de lecture et **1 communication invitée** sans actes [A5],
- **25 communications** dans des conférences ou séminaires **francophones** dont 15 avec actes et comité de lecture.

9.2. Livres ou chapitres de livres

Néant

9.3. Articles dans des revues internationales avec comité de lecture

- T-R1.** Chip-in-the-loop learning algorithm for Boltzmann machine
Klein, J.-O.; Pujol, H.; Garda, P.
Electronics-Letters. 8 June 1995; 31(12): 986-988
[Impact Factor : 1,15]
- R2.** A DSP-like analogue processing unit for smart image sensors
Dupret, A.; Klein, J.-O.; Nshare, A.
International-Journal-of-Circuit-Theory-and-Applications. Nov.-Dec. 2002 30(6):
595-609, Wiley
[Impact Factor : 0,33]
- R3.** A Smart Sensor Based Vision System: Implementation and Evaluation,
Elouardi A., Bouaziz S., Dupret A., Lacassagne L., Klein J., Reynaud R.
J. Phys. D: Appl. Phys. 39 pp.1694-1705, 2006.
[Impact Factor : 2.2]
- R4.** VHDL simulation of Magnetic Domain Wall Logic
Jacques-Olivier Klein, Eric Belhaire, Claude Chappert, Russel P. Cowburn, Doro-
thée Petit, and Dan Read
IEEE transaction on Magnetics, Vol 42, n°10 , october 2006, pp. 2754-2756.
[impact factor : 1,0]
- R5.** Image Processing Vision Systems: Standard Image Sensors Versus Retinas
A. Elouardi, S. Bouaziz, A. Dupret, L. Lacassagne, J.O. Klein, R. Reynaud
IEEE Transactions on Instrumentation and Measurement, vol. 56, no 5, pp 1675-
1687, october 2007.
[impact factor : 0,70]
- R6.** Time comparison in image processing: APS sensors versus an artificial retina based
vision system,
A Elouardi, S Bouaziz, A Dupret, L Lacassagne, J O Klein and R Reynaud
IOP Measurment Science and technology 18 2817-2826 (20 July 2007).

[Impact Factor 2007 :1,30]

- R7. Synthesis methodology for magnetic domain wall logic,
Jacques-Olivier Klein, Eric Belhaire, Claude Chappert, Russel Cowburn, Dan Read
and Dorothée Petit,
International Journal of Electronics, Vol. 95, No. 3, Mars 2008, 249–263.
[Impact Factor 2006 : 0,46]
- R8. Electrical simulation of on-chip neural learning based on nanocomponents
Michel HE, Jacques-Olivier KLEIN and Eric BELHAIRE
Electronics letters Vol. 44, N°9 p. 575-576, 2008.
[Impact Factor : 1,15]

9.4. Articles dans des revues nationales avec comité de lecture

- T-RN1. Simulation de la machine de Boltzmann en temps réel
Klein, J.-O.; Pujol, H.; Garda, P.
Traitement-du-Signal. 1994; 11(2): 195-204, ISSN : 0765-0019.
- RN2. Algorithmes stochastiques synchrones pour les architectures cellulaires
Klein, J.-O.; Dupret, A.
Traitement-du-Signal. 1999; 16(5): 361-370, ISSN 0765-0019.

9.5. Articles dans des conférences internationales avec comité de lecture

- T-A1. Ra : An analog Neurocomputer for the synchronous Boltzmann machine,
H. Pujol, J.O. Klein, E. Belhaire, P. Garda
MicroNeuro94, 26-28 Septembre 1994, Torino, Italie.
Proceedings of the Fourth International Conference on Microelectronics for Neural Networks and Fuzzy Systems, pp. 449--455, 1994, ISBN 0-8186-6710-9.
- A2. A programmable vision chip for CNN based algorithms
Dupret, A.; Klein, J.-O.; Nshare, A.
Proceedings-of-the-2000-6th-IEEE-International-Workshop-on-Cellular-Neural-Networks-and-their-Applications-CNNA-2000
23-25 mai 2000, Catane (Italie)
ISBN: 0780363442
- A3. Improved ARAM for PARIS, an original programmable vision chip
Nshare, A.; Klein, J.-O.; Dupret, A.
7th IEEE International Workshop on Cellular Neural Networks and their Applications (CNNA 2002),
Frankfurt (Germany), July 22-24, 2002
ISBN: 981238121X
- A4. A universal switched capacitors operator for the automatic synthesis of analog computation circuits
Moutault S., Klein J.-O., Dupret A.
IEEE CAMP 2003, New Orleans (USA), du 12 mai 2003 au 14 mai 2003;
publié dans IEEE Sixth International Workshop on Computer Architecture for Machine Perception, 2003.
ISBN: 0-7803-7971-3
- A5. Decade of Progress in Analog Silicon retinas : the Orsay vision,
Belhaire E., Devos F., Dupret A., Klein J.-O., Vasiliu M.
IEEE CAMP 2003 International Workshop on Computer Architectures for Machine Perception, New Orleans, USA, du 12 mai 2003 au 14 mai 2003 (**invité**).

- A6. A universal switched capacitor computation cell applied to a programmable vision chip
Klein, J.-O.; Dupret, A.; Nshare, A.; Moutault, S.; de-Carne, P.
Proceedings-of-the-16th-European-Conference-on-Circuit-Theory-and-Design, ECCTD'03. 2003: III-225-8 vol.3
ISBN: 8388309951
- A7. Vision system implementation based on smart retina
Elouardi, A.; Bouaziz, S.; Dupret, A.; Klein, J.-O.; Reynaud, R.
Proceedings-of-IEEE-Sensors-2003
Toronto, Canada, October 22-24, 2003
pp. 1130-1135, Vol.2
ISBN: 0780381335
- A8. An improved analog computation cell for Paris II, a programmable vision chip
Moutault, S.; Mathias, H.; Klein, J.-O.; Dupret, A.
2004-IEEE-International-Symposium-on-Circuits-and-Systems (ISCAS2004)
May 23-26 2004, Vancouver (Canada)
2004: V-453-456 Vol.5
ISBN: 078038251X
- A9. Image processing vision system implementing a smart sensor
Elouardi, A.; Bouaziz, S.; Dupret, A.; Klein, J.-O.; Reynaud, R.
Proceedings-of-the-21st-IEEE-Instrumentation-and-Measurement-Technology-Conference
2004: 445-50 Vol.1
Como, Italy, 18-20 may 2004.
ISBN : 078038248X
- A10. On chip vision system architecture using a CMOS retina
Elouardi, A.; Bouaziz, S.; Dupret, A.; Klein, J.O.; Reynaud, R.;
Intelligent Vehicles Symposium, 2004 IEEE
14-17 June 2004 Page(s):206 - 211
Parma, Italy
ISBN: 0780383109
- A11. CMOS Image Sensor versus Retina Experience
Elouardi Abdelhafid, Bouaziz Samir, Dupret Antoine, Lionel Lacassagne, Klein Jacques-Olivier, Reynaud Roger,
IEEE Sensors 2004, Vienna, Austria, 24-27 october 2004.
- A12. Low power image processing: analog versus digital comparison
Klein, J.-O.; Lacassagne, L.; Mathias, H.; Moutault, S.; Dupret, A.
Proceedings of the Seventh International Workshop on Computer Architecture for Machine Perception (CAMP 2005). 2005: 111-15
ISBN : 0769522556
- A13. Modeling and calibrating an on chip vision system with a CMOS retina
Elouardi, A.; Bouaziz, S.; Dupret, A.; Klein, J.-O.; Reynaud, R.
Proceedings.-2004-International-Conference-on-Information-and-Communication-Technologies:-From-Theory-to-Applications
Damascus, Syria , 19-23 april 2004
ISBN : 0780384822
- A14. A Smart Sensor for Automotive Vision Applications
Elouardi A., Bouaziz S., Dupret A., Lacassagne L., Klein J., Reynaud R.
SENSACT2005, ESIEE Noisy-Le-Grand, Paris, du 07 décembre 2005 au 08 janvier 2005;
in SENSACT Sensors and Actuators for Advanced Automotive Applications, vol. 1,
7 pages, 2005.

- A15. A CMOS Retina Based Vision System
Elouardi A., Bouaziz S., Dupret A., Lacassagne L., Klein Jacques O., Reynaud R.
IEEE ICTIS, TETUAN, du 03 juin 2005 au 06 juin 2005;
in IEEE Information and Communication Technologies International Symposium IC-
TIS, p. 128, 7 pages, 2005
- A16. VHDL Simulation of magnetic domain wall logic
Klein, Jacques-Olivier; Belhaire, Eric; Chappert, Claude;
Cowburn, Russell P.; Petit, Dorothée; Read, Dan.
IEEE International Magnetic Conference, San Diego, California, USA, from May 8
to May 12, 2006
- A17. A Smart Sensor for Image Processing: Towards a System on Chip
Abdelhafid Elouardi, Samir Bouaziz, Antoine Dupret, Lionel Lacassagne, Jacques-
Olivier Klein, Roger Reynaud
IEEE-ISIE'2006 International symposium on industrial electronics, 9-13 July 2006,
Montreal, Canada.
- A18. Magnetic Domain wall logic requires new synthesis methodologies
Jacques-Olivier Klein, Eric Belhaire, Claude Chappert, Russell Cowburn, Dan Read
and Dorothée Petit
DTIS 06, Design & test of integrated systems in nanoscale technology, Tunis, 5-7
september 2006 pp. 115-118
ISBN 0-7803-9726-6
- A19. Neural Network for nanoscale architecture
Michel Hé, Jacques-Olivier Klein, Eric Belhaire, Mathilde Joly, Andrea Pinna, Pa-
trick Garda
2006 6th IEEE Conference on Nanotechnology, Cincinnati, Ohio (USA) 17-20 july,
2006.
ISBN: 1-4244-0078-3
- A20. Synthesis of Finite State Machines with Magnetic Domain Wall Logic
Jacques-Olivier Klein, Eric Belhaire, Claude Chappert, Florent Ouchet, Russell
Cowburn, Dan Read and Dorothée Petit
IEEE International Symposium on Circuits and Systems (ISCAS 2007), New Orleans
(USA), 27-30 May 2007, pp 133-136.
ISBN: 1-4244-0921-7
- A21. Active Pixel Sensors and Smart Retinas: Test and Evaluation".
A. Elouardi, S. Bouaziz, A. Dupret, L. Lacassagne, J.O. Klein, R. Reynaud
IEEE Information and Communication Technologies International Symposium IC-
TIS'2007. ISBN: 9954-877-0-2. April 3-5, 2007. Fes, Morocco.
- A22. Architecture of Neural Synaptic Array, Design and Simulation
He M., Klein J.-O., Belhaire E.
in Proceeding of the IEEE-Nano 2007, 7th IEEE International Conference on Nano-
technology, Hong Kong, Chine, du 02 août 2007 au 05 août 2007, vol. 1, p. 601-
604 pages, 2007.
- A23. Mixed analog-digital design of a learning nano-circuit for neuronal architectures
He, Michel; Klein, Jacques-Olivier; Belhaire, Eric;
in Proceeding of the IEEE 3rd International Conference on Design and Technology
of Integrated Systems in Nanoscale Era, DTIS 2008, 25-28 Mars 2008, Tozeur (Tuni-
sie), Pages: 1 - 5, 2008.

9.6. Articles dans des conférences nationales avec comité de lec- ture

- T-N1. La Machine de Boltzmann est-elle vraiment lente ?

Klein J.O., Garda P.
Neuro-Nimes92, Nimes, 2-6 Novembre 1992, pp. 251-262
ISBN: 2906899798

- T-N2.** Le Bus Sécable : réalisation des communications dans un accélérateur de la Machine de Boltzmann,
Pujol H., Klein J.O., Garda P.,
Renpar4, 4ième Rencontres du Parallélisme, Lille, 18-20 mars 1992, pp. 52-55.
- T-N3.** Le bus sécable : un réseau d'interconnexion pour machines parallèles
Pujol H., Klein J.-O., Garda P.
Renpar5, 5ième Rencontres du Parallélisme, Brest, 26-28 mai 1993, pp 11-14.
- T-N4.** Robustesse dans une machine de boltzmann analogique.
Klein J.-O., E. Belhaire, Garda P.
Neural Networks and their Applications , Marseilles, 15-16 Decembre 1994.
- N5.** Optimisation d'un opérateur stochastique par l'étude de l'effet de la corrélation des échantillons aléatoires dans un algorithmes stochastique,
J.-O. Klein, A. Dupret, B.Brahima, E. Belhaire
Journées Adéquation Algorithme Architecture en traitement du signal et images, CEA-LETI, Saclay, France 28-30 janvier 1998, p. 181-186.
- N6.** Vers une nouvelle génération de rétines programmables,
J.O. Klein, A. Dupret, S. Moutault, A. Nshare
GRETSI 99, 13-17 Septembre 1999, Vannes, France, pp.595-598.
- N7.** Une unité de calcul analogique pour architectures cellulaires,
A. Dupret, J.O. Klein et A. Nshare
TAISA'2000, Colloque sur le traitement analogique de l'information, du signal et ses applications, 14-15 Septembre 2000, Toulouse
- N8.** Une RAM analogique originale, Mise en oeuvre pour la rétine PARIS,
A. Dupret, J.O. Klein et A. Nshare
TAISA'2001, Colloque sur le traitement analogique de l'information, du signal et ses applications, Septembre 2001, Bordeaux
- N9.** Un modèle flot de données adapté à la synthèse d'architectures analogiques à temps discret,
Sébastien MOUTAULT, Jacques-Olivier KLEIN, Antoine DUPRET
TAISA'2002, Paris 27-28 septembre 2002.
- N10.** Un opérateur à capacités commutées universel pour la synthèse automatique de circuits de calcul analogique
MOUTAULT S., KLEIN J.-O., DUPRET A.
JFAAA2002, Monastir (Tunisie), du 16 décembre 2002 au 18 décembre 2002, pp.36-40, 2002.
- N11.** Une cellule de calcul universelle à capacités commutées appliquée à une rétine programmable,
Jacques-Olivier Klein, Antoine Dupret, Abdallah Nshare, Sébastien Moutault et Patrice de Carné
TAISA'2003, 4ème colloque sur le traitement analogique de l'information, du signal et ses applications, Louvain (Belgique) 25-26 Septembre 2003, pp.127-130.
- N12.** Une cellule de calcul améliorée pour circuit de vision programmable
Moutault, J. O. Klein, H. Mathias, A. Dupret
TAISA'2004 - 4eme colloque sur le traitement analogique de l'information, du signal et ses applications,
Lausanne, Suisse, du 30 septembre 2004 au 1 octobre 2004
- N13.** Architecture d'un système de vision SoC utilisant une rétine programmable

ELOUARDI ABDELHAFID, BOUAZIZ SAMIR, DUPRET ANTOINE, KLEIN JACQUES-OLIVIER, REYNAUD ROGER
3ème Colloque Interdisciplinaire en Instrumentation C2I'04, Ecole Normale Supérieure, Cachan, France du 29/01/2004 au 30/01/2004.

- N14. Evaluation des performances d'un système de vision à base d'un capteur intelligent
ELOUARDI ABDELHAFID, BOUAZIZ SAMIR, DUPRET ANTOINE, LACASSAGNE LIONEL, KLEIN JACQUES-OLIVIER, REYNAUD ROGER
Rétine électroniques, Asic-FPGA et DSP pour la vision et le traitement d'images en temps réel -READ 2005,
Evry (France), du 1 juin 2005 au 3 juin 2005 pp.69-74, 2005.
- N15. TD, travaux dirigés ? – Non ! Tableau dévissé !
Sébastien MOUTAULT, Joël SENPAU ROCA, Jacques-Olivier KLEIN
Communication au congrès de l'AIPU-Monastir'2006, 15-18 mai 2006
Cette communication a obtenu le prix Jean Demal 2006

9.7. Communications orales sans actes², posters et séminaires

- T-C1. Machine neuronale pour la vision,
Pujol H, Klein J.-O., Garda P., Belhaire E.
Première rencontre du réseau doctoral en architecture des systèmes et des machines informatiques, Orsay, 9 février 93.
- T-C2. Implantation d'un accélérateur matériel pour la machine de Boltzmann synchrone,
Pujol H, Klein J.-O. Belhaire E., Calvet D.
Journées des jeunes chercheurs de Rennes, Rennes, 7-8 décembre 1993.
- T-C3. Implantation d'un réseau d'interconnexion pour la machine de Boltzmann synchrone,
Pujol H, Klein J.-O. Belhaire E., Calvet D
Journées communes PRC/GDR ANM et réseau doctoral en architecture, Orsay, 9-11 mars 1994.
- C4. Modélisation d'un système de vision monopuce à base d'une rétine artificielle
ELOUARDI ABDELHAFID, BOUAZIZ SAMIR, DUPRET ANTOINE, KLEIN JACQUES-OLIVIER, REYNAUD ROGER
VIIème Journées Nationales du Réseau Doctoral de Microélectronique, Marseille, France, du 04/05/2004 au 06/05/2004
- C5. Jacques-Olivier KLEIN,
Simulation VHDL et synthèse de circuits Logiques à Parois de Domaines Magnétiques
Séminaire du groupe Nanosys , 30 janvier 2006, ENST PARIS.
- C6. Belhaire E., Klein J.-O.
Apport des nanotechnologies au Traitement du signal et des images
Séminaire, GDR ISIS : Thème C - Adéquation Algorithme Architecture, 12 mai 2006.
- C7. Belhaire E., Klein J.-O., Zhao W.
Exemples de circuits logiques magnétique et de microarchitectures adaptatives
Séminaire, Ecole de Science de Porquerolles, 01 juillet 2006.
- C8. Belhaire E., Klein J.-O., Zhao W., Chappert C.
Des architectures de circuits pour la nanoélectronique
Journées du GDR G2054 Nanoélectronique, Grenoble, du 19 octobre 2006 au 20

2 ou avec actes à diffusion restreinte.

octobre 2006.

- C9. He M., Klein J., Belhaire E.
Nano-LUT inspirée de réseaux de neurones
Journées Nationales du Réseau Doctoral en Microélectronique JNRDM 2007,
Lille, France, du 14 mai 2007 au 16 mai 2007, Poster sans actes
- C10. He M., Klein J.-O., Belhaire E
Une alternative pour implémenter la LUT : le crossbar synaptique
Colloque National du GDR SOC-SIP, Paris, France, du 13 juin 2007 au 15 juin 2007,
Poster sans actes
- C11. BELHAIRE E., KLEIN J.-O.,
Architectures bio-inspirées,
Séminaire de Réflexion Programmes du LIST, 20 mars 2007.
- C12. BELHAIRE E., KLEIN J.-O.
Architectures électroniques à base de nanodispositifs
Colloque interne IEF, Orsay, France, 23 octobre 2007
- C13. Architectures nanoélectroniques en rupture : La piste neuronale
Jacques-Olivier KLEIN,
Séminaire aux Journées Nanotubes de carbone et Nanotechnologies, Bordeaux,
18-21 mars 2008.
- C14. Nouvelles opportunités de l'approche neuro-inspirée
Jacques-Olivier Klein
Séminaire CEA Chimtronique, 20-21 mai 2008, Paris.

9.8. Brevets

- B1 Synapse nano-électrique et procédé d'apprentissage d'une telle synapse,
Klein et Belhaire, déposé à l'INPI le 27 juillet 2007, sous le n° 07 05532.

9.9. Rapports de recherche

- X.1. Modélisation et simulation fonctionnelles d'un apprentissage neuro-inspiré pour des
nanocomposants.
Jacques-Olivier Klein
Rapport de contrat PANINI N°ANR-07-ARFU-008-1, Délivrable D3.1, 7 juillet 2008.

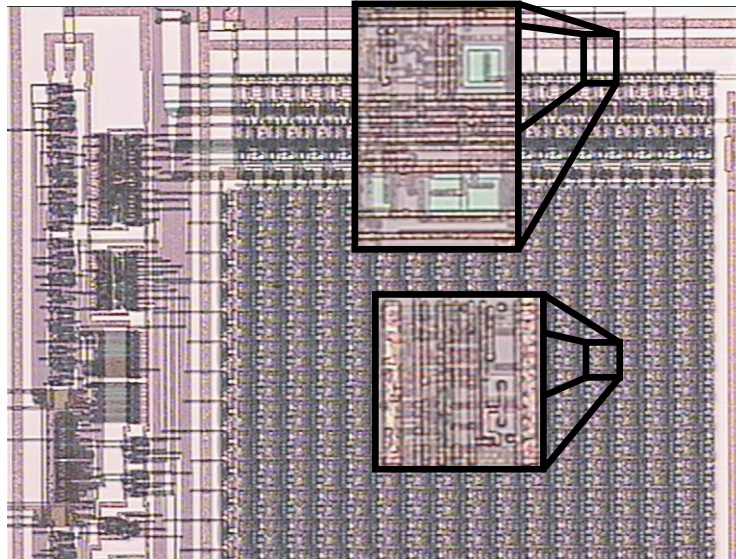
Introduction à la synthèse des travaux de recherche

Les architectures dédiées au calcul sont très largement dominées par le couple constitué d'un côté par le modèle de Von Neumann et de l'autre côté par la technologie CMOS. Cette prédominance est telle qu'elle s'impose aujourd'hui avec une évidence écrasante face aux propositions alternatives. La généricité du modèle de programmation alliée à l'efficacité, notamment d'un point de vue énergétique, de la technologie CMOS donne à ce couple un caractère tellement dominant qu'en sortir serait considéré aujourd'hui comme une rupture de paradigme. C'est pourtant autour des technologies alternatives et des modèles de programmation différents que portent les travaux présentés ici. **L'étude de l'impact des premières sur les secondes en constitue le fil directeur.**

Au départ, il s'agit de travaux essentiellement fondamentaux, portant sur des applications qui peuvent être qualifiées de niches. Ainsi, dans la deuxième partie de ce manuscrit, je présente les travaux réalisés au sein du département AXIS de l'IEF. Ils sont essentiellement axés sur les architectures bio-inspirées, de réseaux de neurones et de rétines. Ces architectures sont associées à des éléments de calcul analogiques. Dans un premier temps, elles font appel à une organisation relativement classique d'un point de vue des circuits analogiques puisqu'elles sont constituées par un réseau quasi-statique d'opérateurs. Ensuite, elles s'inspireront davantage des architectures, de la programmation et des méthodes de synthèse des systèmes numériques.

Cependant, à mesure que la fin de la loi de Moore se profile, les recherches sur les technologies alternatives concernent de plus en plus l'ensemble du secteur de la micro-nano-électronique et touchent à des marchés énormes. Aujourd'hui, compte tenu du rôle de ce secteur dans l'industrie et dans toute l'économie, cette ambition constitue un enjeu de société majeure, mais toujours un objectif scientifiquement aussi risqué tant il reste difficile de concurrencer le couple en place.

Dans les deux parties suivantes, je présenterai les résultats des travaux menés au sein du département NST de l'IEF à partir de l'année 2005. Dans ce cadre nous avons créé une opération «Architectures pour les nanocomposants», qui vise à concevoir des architectures basées sur des éléments nanométriques, magnétiques ou non magnétiques. La troisième partie de ce manuscrit porte alors sur l'impact en termes d'architecture, du choix de l'aimantation comme grandeur support de l'information dans les circuits à paroi de domaine magnétique. Enfin, je revendrai dans la quatrième partie sur l'approche neuronale, non plus en tant que modèle de programmation alternatif mais principalement comme source d'inspiration pour imaginer les procédures qui permettent d'appliquer un apprentissage à un ensemble de nanocomposants de façon à progresser vers la fonctionnalisation d'une assemblée auto-organisée fabriquée par une approche dites «bottom-up». Outre l'intérêt d'une telle procédure pour compenser les défauts et dispersions de caractéristiques, cet objectif permettrait d'envisager les plus hautes densités d'intégration sans payer le prix pharaonique projeté pour les équipements et les usines de fabrication en CMOS ultime.



Deuxième Partie

Synthèse des travaux de recherche de 1995 à 2004

Architectures de calcul analogique

II. Synthèse des travaux de recherche de 1995 à 2004

Architecture de calcul analogique

1. Conclusion de ma thèse

Dans cette partie, je vais présenter une synthèse des activités de recherche que j'ai menées comme maître de conférences à la suite de ma thèse au sein du département AXIS de l'IEF. Cependant, comme il m'a semblé utile d'en montrer la cohérence, j'ai trouvé nécessaire de rappeler brièvement les conclusions de mes travaux de thèse pour comprendre de quelle façon elles ont servi d'impulsion pour choisir une direction de recherche au début de ma carrière.

J'ai effectué ma thèse à l'Institut d'Electronique Fondamentale sous la direction de Patrick Garda en contribuant au projet d'implantation analogique de réseaux de neurones stochastiques (projet RA). Ma contribution ne s'est pas portée sur la conception des cellules élémentaires, déjà conçues par Eric Belhaire, mais sur l'étude architecturale (cf. figure 1.) et algorithmique en relation avec l'usage de circuits analogiques.

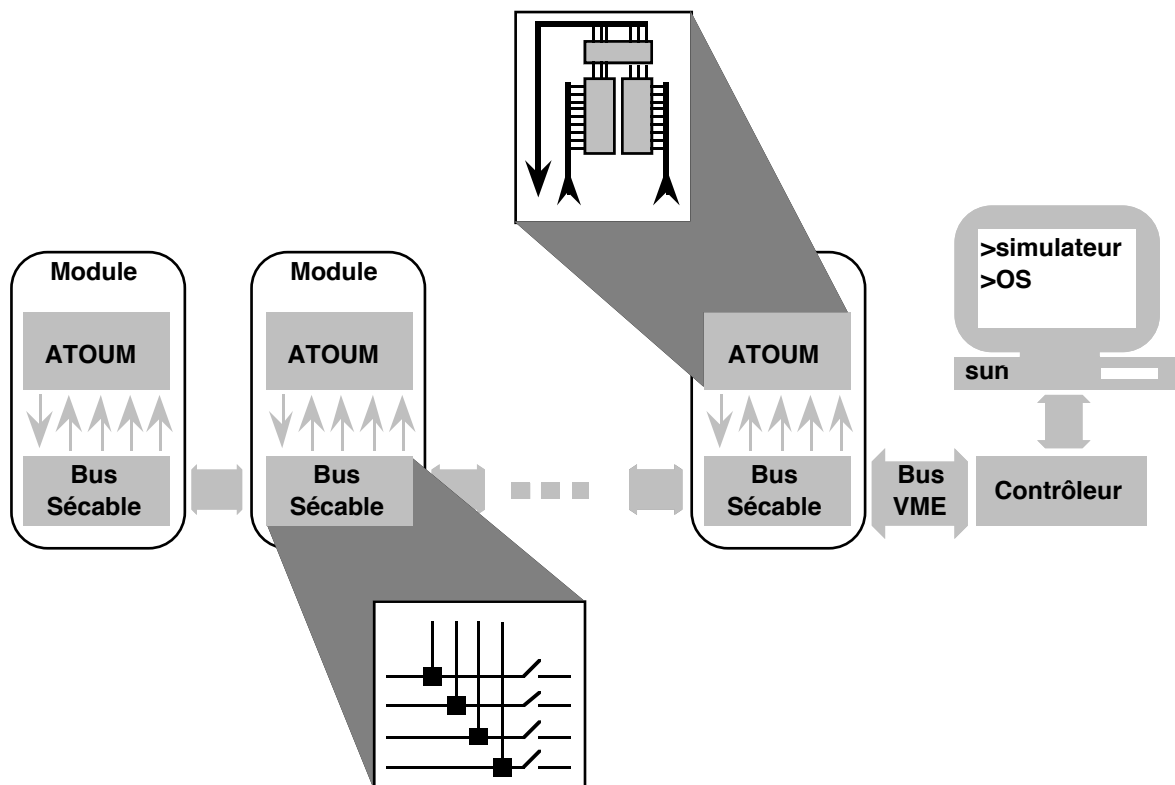


Figure 1 : L'architecture modulaire du projet RA. Chaque module de calcul (carte ATOUM) comprend deux circuits synapses (MBA1) et un circuit neurone (MBA2). Ces modules sont connectés par un bus sécable et relié à un contrôleur communiquant avec la station hôte [T-A1].

Comme ce projet est né d'une collaboration entre Robert Azencott pour la partie algorithmique et l'IEF pour la partie électronique, une telle étude algorithmique n'était pas apparue immédiatement nécessaire à l'IEF. Elle l'est devenu à mesure que les

questions d'architecture, de technologie et d'algorithmie sont devenues de plus en plus imbriquées. J'ai commencé ma thèse par une évaluation de performances pour la simulation de réseaux de neurones en général et de la machine de Boltzmann en particulier, afin d'identifier les réseaux qui nécessitent une architecture dédiées. Contrairement aux idées reçues, les stations de travail récentes (de l'époque) suffisaient à simuler bien des réseaux de neurones, même itératifs et stochastiques [T-N1,T-RN1]. Les architectures spécialisées permettent d'aller encore plus loin en répondant aux exigences extrêmes par la quantité de données manipulées ou par les contraintes temporelles: traitement d'image neuronal ou applications à temps de réponse ultra-courts (cf. figure 2). Dans ces deux derniers cas, le calcul analogique trouve une place de choix, d'autant qu'il profite de la faculté d'auto-compensation des algorithmes d'apprentissage neuronal comme j'ai pu le vérifier.

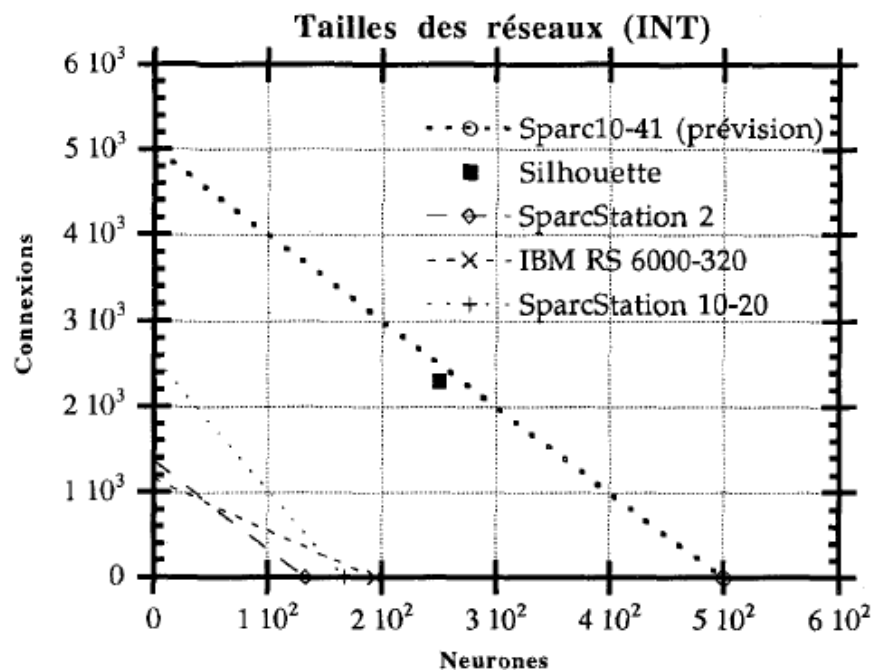


Figure 2 : Tailles des réseaux (en nombre de neurones et synapses) qui peuvent être traités en moins de 40 ms en calcul entier pour différentes stations de travail [T-RN1]. Ces résultats correspondent à 100 itérations de relaxation stochastique pour une machine de Boltzmann synchrone.

De mes travaux de thèse, j'ai tiré trois conclusions qui ont fortement influencé les thèmes de recherche que j'ai choisis au début de ma carrière de maître de conférences.

- D'abord, les évaluations de performances que j'ai menées ont montré que **l'intérêt d'une architecture dédiée est surtout limité aux applications à très forte contrainte temporelle ou celles manipulant d'importantes quantités de données comme le traitement d'image**. C'est vers ces dernières que je me tournerai plus tard, mais l'utilisation de circuits analogiques pour le calcul dans le contexte du traitement d'images soulève un autre problème qui nous amène au deuxième point.
- Les architectures analogiques, jusqu'alors, étaient conçues sous forme de réseaux statiques de cellules de calculs, ce qui implique que les circuits doivent avoir la taille des problèmes traités. On tombe alors dans le paradoxe suivant : le recours à des architectures dédiées analogiques se justifie par la quantité de données mise en jeu en traitement d'image mais **l'absence de multiplexage temporel conduit à des architectures trop rapides et incapables de traiter**

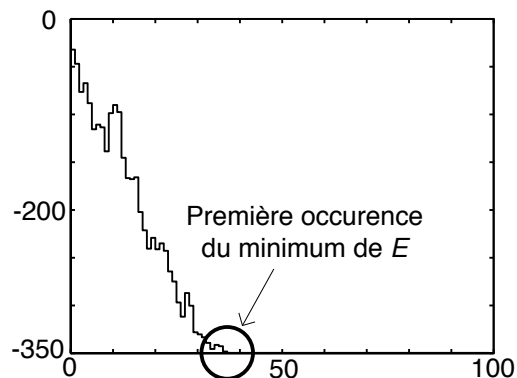
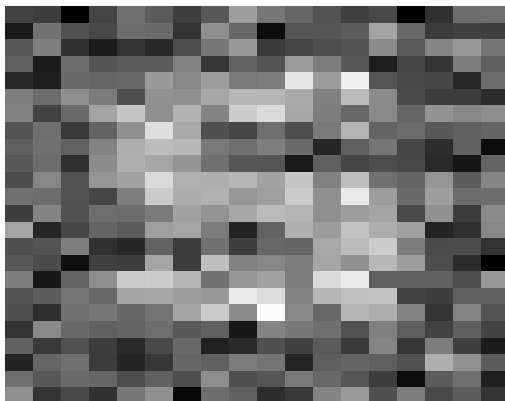
de grosses images par manque de surface. Je m'efforcerais donc par la suite de me tourner vers des architectures analogiques permettant de mettre en place une virtualisation des éléments de traitement de façon à ce que le gain en vitesse de calcul puisse profiter à la taille des problèmes traités.

- Finalement, les études sur la robustesse des implantations analogiques de réseau de neurones stochastiques nous ont montré que **le strict respect des modèles mathématiques** garantissant l'optimalité des solutions apportées par les algorithmes stochastiques, **rarement réalisable** en pratique (temps de simulation infini, décroissance logarithmique de la température, générateurs aléatoires indépendants...) **n'est pas absolument impératif**. J'avais notamment tenté d'évaluer l'impact de la corrélation des générateurs aléatoires, mais les simulations étaient restées difficiles à interpréter : d'une part, l'effet de compensation joué par l'apprentissage nuit à la lisibilité des résultats. D'autre part, la nécessité de réaliser des statistiques sur les résultats m'avait contraint à choisir des problèmes relativement simples de façon à conserver des temps de simulation raisonnables dans lesquels l'effet d'optimisation globale due à la nature stochastique de l'algorithme n'est pas exploité. On peut considérer qu'il s'agissait plutôt d'une version stochastique de réseaux à codage impulsif.

L'analyse plus précise de l'effet de la corrélation des générateurs aléatoires est le premier thème de recherche que j'ai initié après ma thèse avant d'envisager des applications au traitement d'image, mais nous choisissons un modèle de réseau plus simple, pour lequel la dynamique stochastique est connue et rendue nécessaire par la présence évidente de minima locaux.

2. Traitement d'image stochastique

Comme je l'ai évoqué plus haut, le strict respect des modèles mathématiques garantissant l'optimalité des solutions apportées par les algorithmes stochastiques est rarement réalisable en pratique. Nous avons montré, par des simulations, que beaucoup de contraintes peuvent être relâchées sans dégradation notable des résultats. Dans un premier temps, en choisissant le modèle des verres de spin, nous avons montré que **l'impact de la corrélation du générateur aléatoire est très différent selon qu'il s'agit d'une corrélation spatiale ou temporelle** (cf. figure 3). Si la corrélation spatiale est rapidement néfaste, une corrélation temporelle n'affecte de façon significative la vitesse de convergence vers la solution optimale et la qualité du résultat obtenu qu'au delà de valeur très importante de la corrélation [N5].



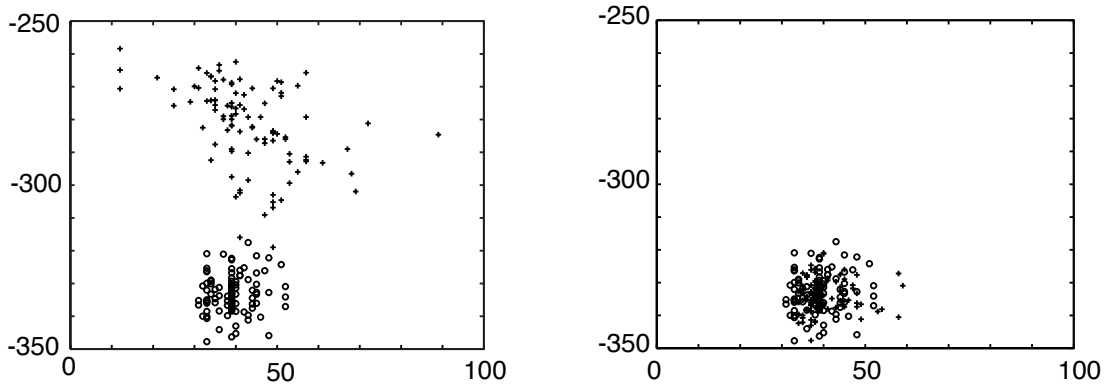


Figure 3 : En haut, à gauche, un exemple d'entrée appliquée sous forme de champ externe dans le modèle des verres de spin. Il s'agit d'une image représentant une forme en «S» particulièrement bruitée. En haut à droite, description du critère considéré : la coordonnée de la première occurrence du minimum d'énergie atteint par un réseau lors du recuit simulé. En bas à gauche, comparaison des résultats suivants que le générateur aléatoire est fortement corrélé **spatialement** (+) ou non (o). La corrélation spatiale nuit clairement à l'efficacité du recuit simulé. En bas à droite, comparaison des résultats suivant que le générateur aléatoire est fortement corrélé **temporellement** (+) ou non (o). Les deux nuages de points sont indiscernables, il n'y a pas d'effet mesurable de la corrélation temporelle [N5].

Nous avons également montré que les réseaux faisant intervenir des cliques d'ordre supérieur à deux et une dynamique synchrone peuvent être utilisés, y compris pour minimiser des énergies quadratiques, dès lors qu'on évite les rares cas où la dynamique synchrone conduit à des oscillations stables (comme les configurations en damiers dans les verres de spin). En conséquence, des architectures mixtes, analogiques/numériques particulièrement efficaces peuvent tirer profit des performances des algorithmes stochastiques en traitement d'images.

Nous avons notamment démontré qu'un **maillage carré résistif associé à des interrupteurs commandés par un algorithme stochastique** (cf. figure 4) **permet de fermer des contours, y compris sur des images extrêmement bruitées** [RN2] (cf. figure 5). Le filtrage est contrôlé par le rapport des résistances mais le filtrage n'opère pas sur les couples de pixels séparés par un contour. La détection des contours est assurée par un algorithme stochastique basée sur la détection du maximum local du gradient et la configuration des contours sur les pixels voisins.

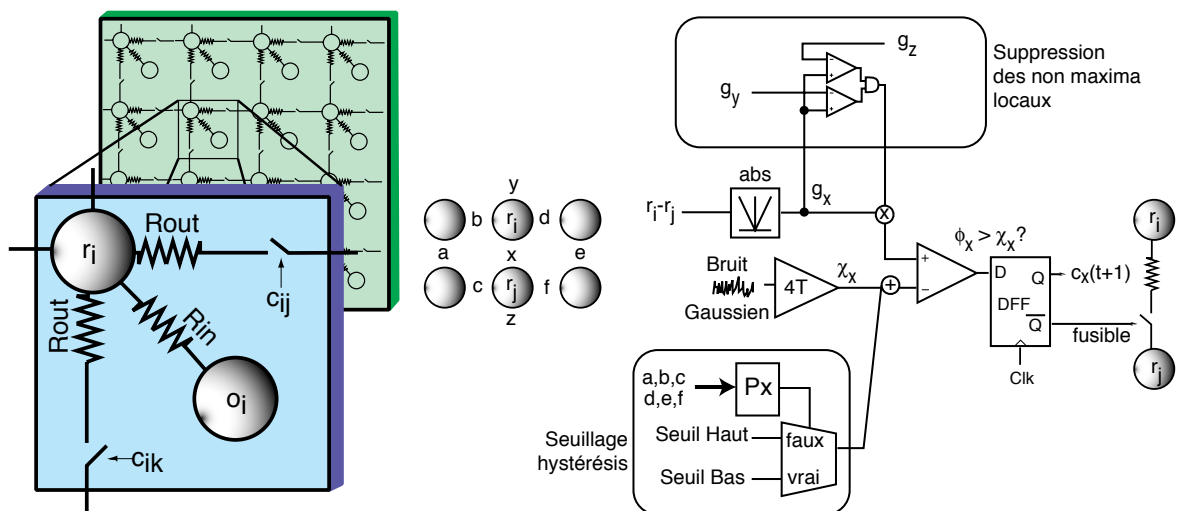


Figure 4 :A gauche : maillage résistif carré associé à des interrupteurs contrôlés par

un algorithme stochastique synchrone. Le rapport R_{in}/R_{out} fixe l'importance relative du filtrage par rapport au terme de rappel aux données. A droite, schéma représentant le principe de l'algorithme contrôlant la détection d'un contour. L'algorithme prend en compte la configuration des contours voisins de façon à privilégier la fermeture des contours et la détection d'un maximum local du gradient de façon à éliminer les détections multiples.



Figure 5 : Images originales (en haut) et images obtenues par un réseau résistif non-linéaire stochastique (en bas). En dépit du niveau de bruit important, l'algorithme permet de combiner filtrage et détection des contours [RN2].

La suite logique de ces travaux aurait pu être de réaliser un circuit analogique spécialisé, dédié à cet algorithme. Pourtant, compte tenu des remarques précédentes sur la nécessité de virtualiser les processeurs, ce n'est pas la direction que j'ai choisie. J'ai également pris conscience de la nécessité de rendre les architectures de calculs analogiques plus attractives pour initier des collaborations qui mèneraient à leur utilisation dans des systèmes temps réel, notamment en robotique ou pour l'automobile. En effet, je suis persuadé que le caractère figé des traitements analogiques, même s'il peuvent se révéler efficaces sur des images très bruitées grâce à la dynamique stochastique, ne laisse pas assez de degré de liberté, donc pas assez de champ d'innovation pour intéresser les traiteurs d'images. Par ailleurs, le caractère stochastique des algorithmes que j'avais envisagés jusqu'alors m'est apparu principalement utile dans les cas où le rapport signal/bruit est extrêmement faible, ce qui n'est pas très courant en pratique. Pour inciter les traiteurs d'images à s'intéresser à ce type d'architecture et favoriser les collaborations visant à la mise en oeuvre des architectures de calcul analogique au sein de systèmes de vision complets, pour inclure la virtualisation, et pour ne pas se restreindre à l'implantation de traitements stochastiques, j'ai choisi d'initier un projet de rétine programmable.

3. Rétine électronique programmable

Travaux de Thèse d'A. Nshare, co-encadré avec A. Dupret, et collaboration avec l'opération Véhicule intelligent dans le cadre de la Thèse d'Abdelhafid Elouardi.

J'ai proposé une architecture de processeur analogique programmable inspirée des intégrateurs à capacités commutées. En utilisant alternativement **les transferts et les répartitions de charge** (cf. figure 6), il est possible de réaliser **une multiplication accumulation à virgule fixe** utilisée dans les opérations de filtrage [N7]. Il faut ici entendre programmable au sens logiciel, c'est-à-dire capable d'exécuter un certain jeu d'instruction généraliste. Comme ce projet n'avait des chances d'aboutir qu'avec l'appui d'une expertise en conception de circuits intégrés analogiques et en capteurs d'images, j'ai proposé à Antoine Dupret de le co-diriger. Ensemble, nous avons encadré la thèse d'Abdalha Nshare.

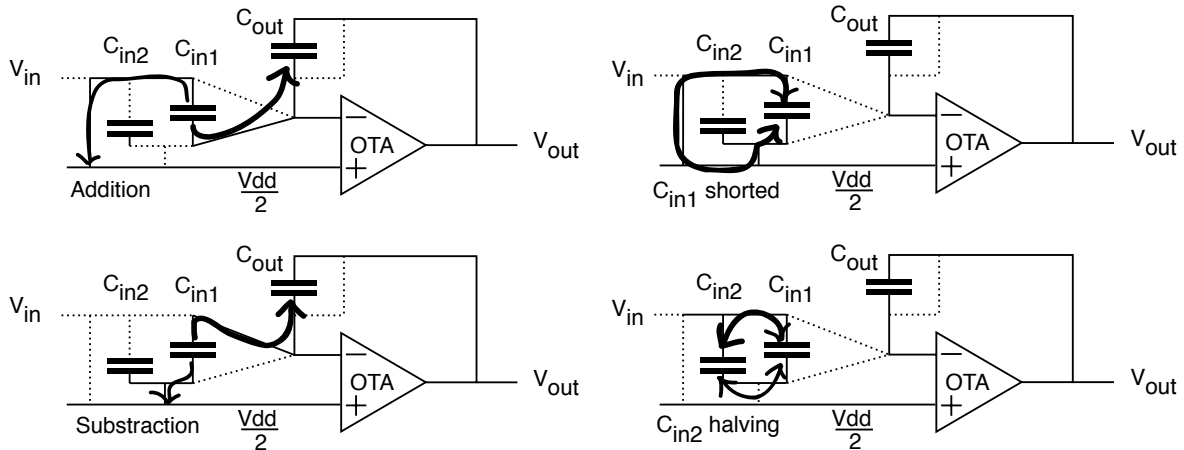


Figure 6 :A gauche, les opérations de transfert de charge réalisent une accumulation (addition) ou une soustraction. A droite, l'alternance entre répartition de charge et mise en court circuit réalise la division par deux. La combinaison des quatres opérations permet de réaliser des multiplications-accumulations par un coefficient à virgule fixe.

Cet unité de calcul analogique étant réalisable sur une surface faible (en technologie CMOS AMS 0,6 chaque processeur occupe $50 \times 300 \mu\text{m}^2$ et dissipe 500 nW pour une fréquence de fonctionnement de 10 MHz.) il est possible de les intégrer en bordure d'une matrice de capteurs d'image associée à des plans mémoires analogiques (cf. figure 7).

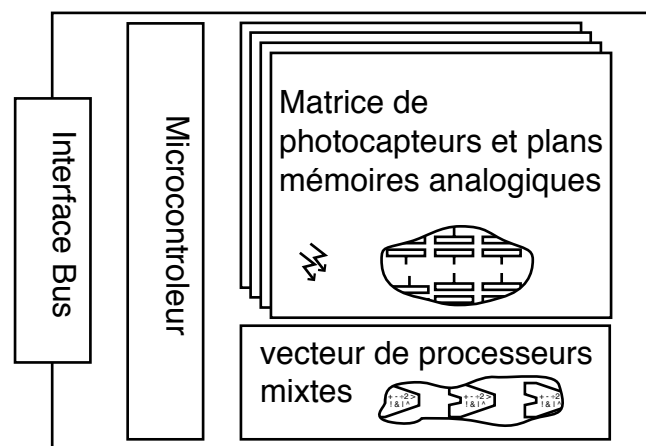


Figure 7 : Architecture de la rétine PARIS.

Abdallah Nshare a montré que l'intégration de minuscules processeurs de calcul analogiques programmables en bordure d'une matrice d'acquisition et de mémorisation d'images permet d'effectuer des algorithmes de vision performants dans des circuits

à très faible coût et à faible consommation. Pour cela, nous avons réalisé un premier circuit de validation (PARIS 1) [R2, A2, N7].

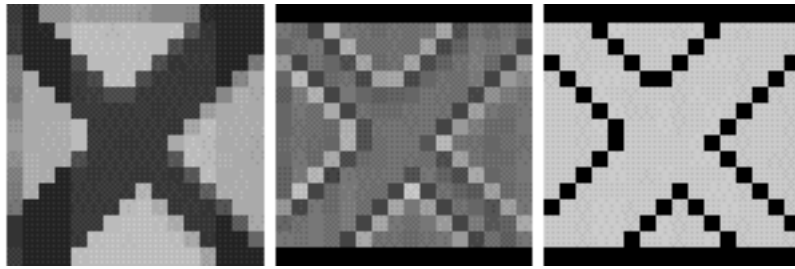


Figure 8 : Image 16 x 16 acquise (à gauche), filtrée (au centre) puis seuillée (à droite) par le prototype de rétine programmable (PARIS1).

L'analyse des expériences de mise en oeuvre de ce premier circuit (cf. figure 8) a permis d'identifier les points clé à améliorer de façon à augmenter la densité d'intégration, la vitesse de calcul et surtout leur précision. Dans ce but, nous avons optimisé les mémoires analogiques et les circuits de calcul analogiques de façon à minimiser l'effet des capacités parasites [A3,A6,A8]. Cette optimisation repose en grande partie sur l'utilisation d'une structure en rétro-action, y compris pour les capacités d'entrée, qui diminue l'impact des capacités parasites d'un facteur proportionnel au gain de l'amplificateur. Cette méthode, déjà utilisée par [47] a été améliorée par l'introduction d'un interrupteur supplémentaire supprimant l'impact de la capacité des longs bus (cf. figure 9).

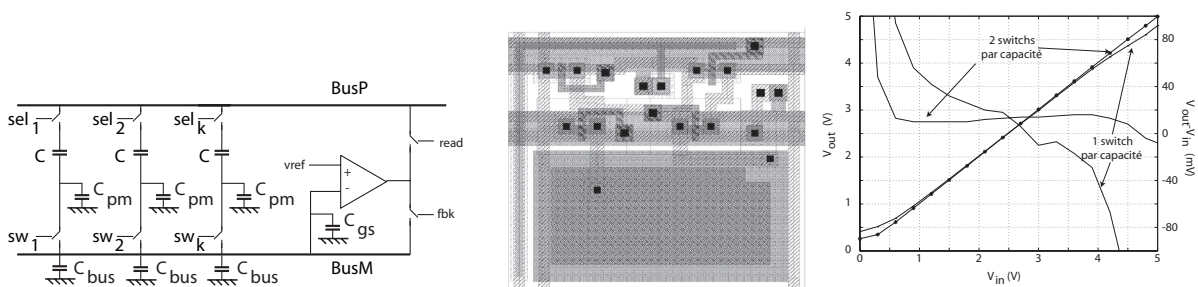


Figure 9 : Schéma (à gauche) et layout (au centre) de la cellule mémoire optimisée. Les résultats de simulation (à droite) comparés à l'état de l'art [47] montre une diminution de l'effet des capacités parasites [A3,A6,A8].

Le concept de programmabilité associé à une architecture de processeurs analogiques au sein d'une rétine était innovant, mais nous n'avons pas été les seuls à explorer cette piste. Il existe maintenant une petite communauté autour de ce thème. A l'université de Manchester, Piotr Dudek a publié en même temps que nous une architecture similaire utilisant des mémoires de courant [1-3] alors qu'à Séville, Rodriguez Vazquez et Servando Espejo (Instituto de Microelectrónica de Sevilla, Centro Nacional de Microelectrónica) faisaient progressivement évoluer l'architecture paramétrable d'un circuit dédié à la simulation de CNN[4] vers une architecture de rétine entièrement programmable [5-6].

Comme je l'avais souhaité, notre circuit a servi de support de collaborations et de nombreuses expérimentations, notamment en collaboration avec l'opération «véhicules intelligents» de l'IEF lors de la thèse d'Abdelhafid Elouardi. Il a ainsi caractérisé notre circuit pour le comparer à l'état de l'art des systèmes de vision intégrés (cf. figure 10) et a proposé son intégration au sein d'un système temps réel pour l'utiliser dans un contexte d'assistance à la conduite de véhicule [R3,R5,R6].

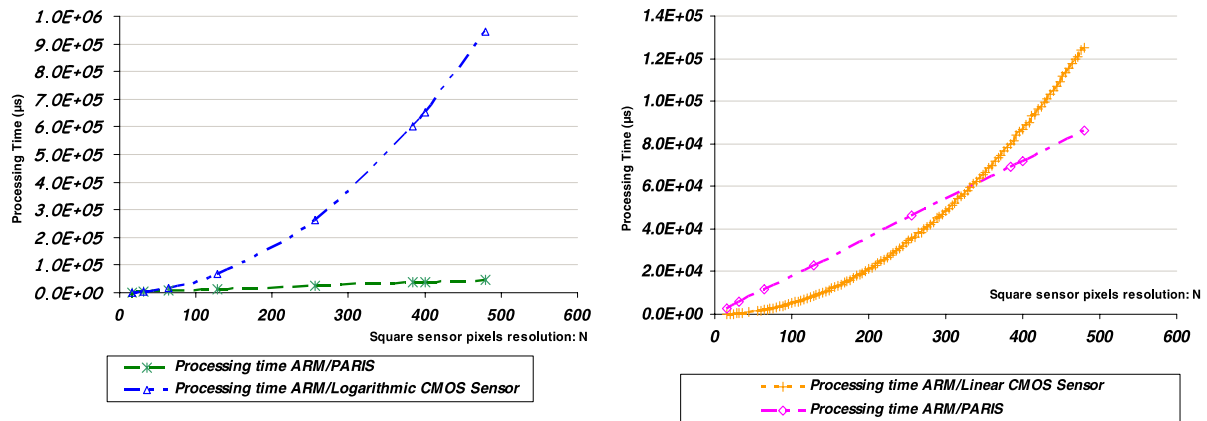


Figure 10 : Comparaison des temps de traitement pour une détection de contours (à gauche) et pour un filtrage de Sobel (à droite) entre la rétine PARIS1 et des capteurs CMOS logarithmique (à gauche) et linéaire (à droite) [R6].

4. Synthèse d'architectures de calcul analogique à capacités commutées

Travaux de thèse de S. Moutault.

L'optimisation de l'opérateur de calcul à capacités commutées dans le contexte des architectures de rétines nous a permis de prendre conscience de son universalité, ce qui ouvrait la voie vers la synthèse automatique d'architecture de calcul analogique.

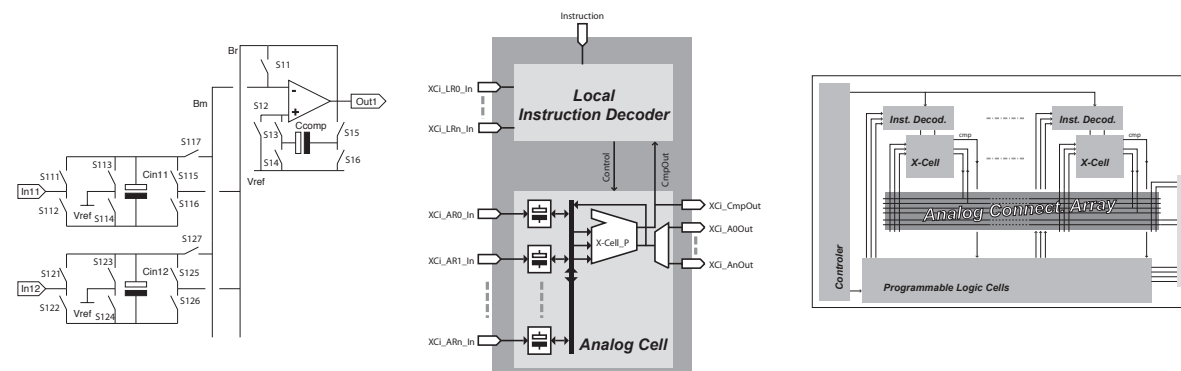


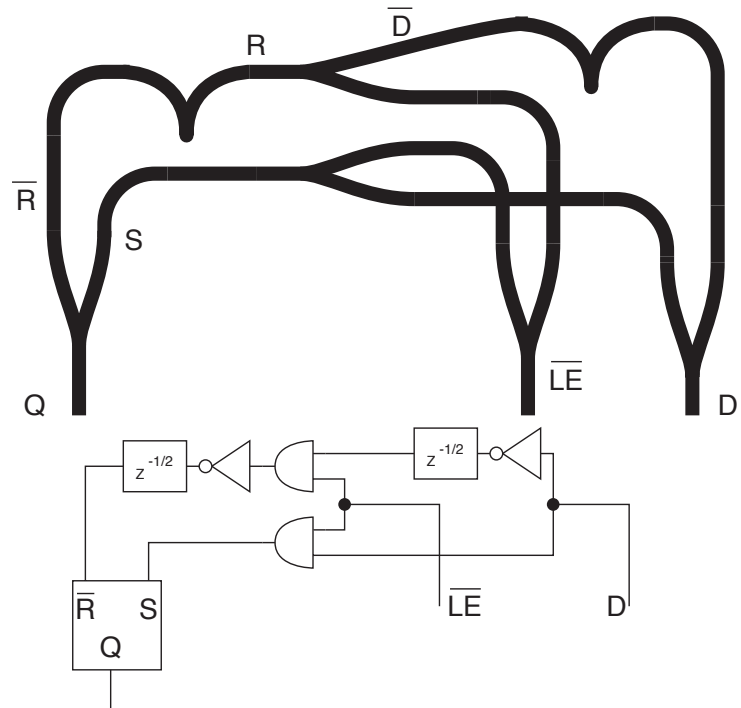
Figure 11 : Cellule de calcul analogique universelle (à gauche), vue fonctionnelle (au centre) et architecture globale (à droite). La synthèse d'un opérateur analogique revient alors à définir sa séquence de contrôle [A4, N9, N10, N11].

Il s'agit cette fois d'une architecture programmable au sens matériel, comme l'est un FPGA dans le monde numérique. Alors que la conception des systèmes numériques s'est largement automatisée, la conception analogique reste principalement artisanale. Si les langages de description matérielle, qui servent d'entrée aux outils de synthèse numérique se sont vus augmentés d'extensions analogiques ou mixtes, seule la modélisation a été envisagée et non la synthèse. Pourtant, à partir d'une architecture de calcul analogique générique, il est possible d'intégrer des opérateurs de traitement analogique dans une chaîne de synthèse automatique. Nous avons montré

qu'un seul opérateur analogique à capacités commutées, suffisamment universel, permet d'envisager la synthèse automatique de circuits de calculs analogiques. Nous avons défini l'architecture matérielle adaptée à ce type de traitement et précisé les étapes permettant d'arriver à la synthèse d'algorithmes sur ces architectures, comme le fait l'approche «adéquation algorithme-architecture» pour les circuits numériques programmables.

5. Conversion thématique

La poursuite des travaux sur les architectures de calcul analogiques à capacités commutées, notamment les rétines programmables, aurait nécessité de rencontrer un soutien académique ou industriel pour trouver un second souffle. Nos tentatives dans ce sens, auprès des industriels de la micro-électronique ou dans le cadre d'ACI jeunes chercheurs sont restées sans effet. Il semble que le thème des rétines programmables paraissait encore trop lointain aux uns pour envisager des applications immédiates et trop ancien aux autres puisqu'il s'agit d'un des thèmes fondateurs du département AXIS de l'IEF. Par ailleurs, cette époque correspond à une restructuration de l'IEF, plus particulièrement de mon département et dans celui-ci de l'opération «rétine» qui a finalement disparu. A cette période, j'ai profité de six mois de CRCT qui m'ont permis d'effectuer une reconversion thématique en rejoignant le département NST (Nano-Spin-Tronique), avec Eric Belhaire et Jean-François Pône pour y créer une nouvelle opération NanoArchi (Architecture à base de nano-composants). Ma contribution scientifique dans ce cadre est détaillée dans les deux chapitres suivants.



Troisième partie

Logique à paroi de domaine magnétique

II. Troisième partie : Logique à paroi de domaine magnétique

1. Introduction

L'accroissement exponentiel de la densité d'intégration dans les technologies microélectroniques au cours des 4 dernières décennies a fortement limité la motivation pour la recherche de solutions concurrentes à la technologie CMOS. Nous savons pourtant depuis toujours que la loi de Moore connaîtra nécessairement une fin. Aujourd'hui, alors que cette date approche, les technologies alternatives suscitent un regain d'intérêt. Il nous faut notamment trouver des solutions permettant de limiter l'explosion des coûts de fabrication, augmenter l'autonomie des systèmes portables trouver le moyen d'évacuer la chaleur, mobiliser la troisième dimension notamment pour les interconnexions ou pour les mémoires et mettre au point une technologie non volatile permettant de suppléer les mémoires flash alors qu'elles atteignent déjà leurs limites. Dans ce contexte, les technologies utilisant le magnétisme comme support de l'information présentent beaucoup de qualités. Les technologies de fabrications de pistes magnétiques sont bien moins complexes et moins coûteuses que les technologies CMOS. L'apport d'énergie peut provenir d'un champ magnétique externe, ce qui découple la question de l'autonomie et celle de la dissipation. Le support magnétique étant intrinsèquement non-volatile, il offre à la fois une alternative aux mémoires flash et une solution pour limiter la consommation statique puisque les blocs inutilisés n'ont plus besoin d'énergie du tout. Enfin les technologies de mémoires magnétiques (MRAM) paraissent facilement empilables, manière d'exploiter intelligemment la troisième dimension. En dehors des mémoires magnétiques (MRAM), objet d'intenses recherches aujourd'hui, ils existent des composants basés sur la propagation de paroi de domaine magnétique le long de nanofils [12]. Nous allons nous intéresser ici aux perspectives offertes par la logique à paroi de domaine magnétique. Des circuits ont déjà démontré le bon fonctionnement des éléments fondamentaux, mais il restait à vérifier que ces éléments sont suffisants pour envisager la synthèse d'architectures plus complexes. Je me suis attaché à démontrer la faisabilité de n'importe quelle fonction logique synchrone en décrivant une méthode de synthèse de machine d'états (automate). Ces travaux sont décrits dans le présent chapitre.

Dans un premier temps, je rappellerai le principe de fonctionnement des circuits logiques à parois de domaines magnétiques utilisant une aimantation dans le plan, puis je décrirai les différents composants élémentaires. Ensuite je présenterai une méthode de simulation fonctionnelle adaptée aux circuits magnétiques, puis j'expliquerai les spécificités des circuits magnétiques que les simulations nous ont permis de mettre en lumière. Enfin, je présenterai une méthode de synthèse d'automate adaptée à cette technologie.

2. Principe de fonctionnement

La largeur de ces nanofils magnétiques (de l'ordre de 200 nm dans [7], (cf. figure 12) induit un comportement anisotropique de sorte que l'état de l'aimantation est principalement binaire, dans un sens ou dans l'autre le long de l'axe des nanofils.

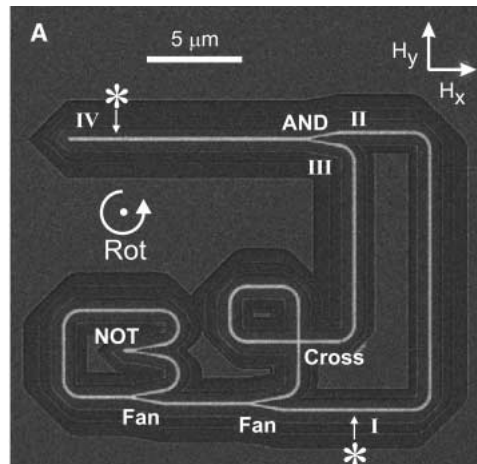


Figure 12 : Circuit à paroi de domaine magnétique, d'après [12].

De plus, la tendance du matériau à favoriser une aimantation uniforme va conduire à l'apparition de domaines magnétiques à l'intérieur desquels l'aimantation est orientée dans le même sens. **Entre deux domaines magnétiques d'aimantations différentes se trouve une paroi de domaine susceptible de se propager lorsqu'elle y est aidée par un champ magnétique extérieur, orienté dans l'axe du nanofil.** Il a été montré que les parois de domaines pouvaient se propager à travers une intersection de nanofils ou se scinder en deux lorsqu'elles rencontrent une bifurcation. Par ailleurs, si le champ extérieur est tournant (circulaire), les parois de domaines peuvent se propager dans des virages à 90°, sous réserve que le sens de la courbure des nanofils suive le sens de rotation du champ externe et au prix d'un retard de phase de 90°. L'exploitation d'un champ externe rotatif dans les circuits magnétiques entraîne trois conséquences. D'abord, le tracé des circuits sera particulier puisque les pistes devront toujours tourner dans le même sens. Ensuite, le sens de rotation du champ va décider du sens de propagation du signal dans les virages, assurant au passage le caractère unidirectionnel de cette propagation. Ainsi, [12], définit la convention suivante : l'état logique '1' (resp. '0') correspond à une aimantation dans le sens de propagation du signal magnétique (resp. le sens opposé à la propagation du signal). Enfin, pour des tronçons de nanofils de même orientation, **le sens du champ externe tournant favorise tantôt la propagation des '0' et tantôt la propagation des '1', de sorte que les fronts montants et descendants sont toujours décalés d'une demi-période** du champ externe. Nous verrons plus loin que ce déphasage n'est pas un détail anodin.

3. Composants élémentaires

Deux composants principaux peuvent être insérés sur le circuit d'un nanofil : la porte inverseuse [13] et l'élément en forme de 'Y' [14] (cf. figure 13). Ces composants sont reliés entre eux par des pistes formant des sections droites, des courbes et des croisements à 90°. Nous allons préciser le fonctionnement de ces différents éléments avant d'envisager leur utilisation.

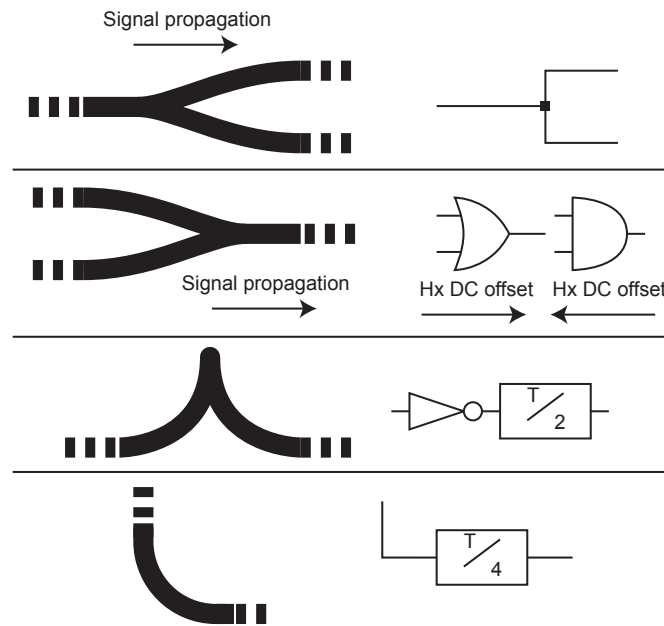


Figure 13 : Composants magnétiques et leur équivalent logique.

3.1. Pistes rectilignes et croisements

Lorsqu'une paroi de domaine est présente sur une piste magnétique, elle peut se propager en étant poussée par un champ magnétique externe d'intensité suffisante et orienté dans la direction de la piste. Le champ magnétique fait croître le domaine dont l'aimantation est orientée suivant la direction du champ externe, au détriment de celui dont la direction d'aimantation s'oppose au champ externe. La croissance du premier domaine se traduit par une propagation de la paroi séparant les deux domaines. On peut noter que le champ externe doit être suffisamment intense pour pousser la paroi tout en restant inférieur au champ de nucléation qui provoquerait le retournement de l'aimantation même en l'absence d'une paroi préexistante. Par ailleurs, la propagation de la paroi résulte principalement de la composante du champ parallèle au nanofil. On peut donc créer des croisements à 90° qui simplifieront le routage du circuit sans que la propagation dans l'une des directions n'interfère avec celle dans la direction orthogonale. Si le champ externe est tournant, circulaire, la propagation aura lieu à des instants différents, en quadrature.

3.2. Virage et direction de propagation

Lorsqu'une paroi de domaine atteint un virage, l'orientation du nanofil cesse d'être alignée sur le champ externe. L'anisotropie de forme due à l'étrécissement de la paroi interdit que l'aimantation puisse s'aligner sur le champ externe. La paroi reste donc bloquée dans le virage jusqu'à ce que le champ externe change de direction. Si le champ externe est rotatif, deux situations peuvent se présenter. Soit le sens de rotation du champ magnétique suit la courbure du virage, soit le virage et le champ magnétique ne tourne pas dans le même sens. Dans le premier cas, le champ externe va s'aligner sur la direction du nanofil provoquant la propagation de la paroi dans le virage et au-delà. Dans le second cas, le champ externe se trouvera aligné avec la direction du nanofil après un demi tour, mais sa direction sera opposée à celle qu'il avait précédemment, ce qui provoquera le rebroussement de la paroi. Ainsi, **la courbure des virages doit suivre le sens de rotation du champ magnétique** externe, sens **qui impose la direction de propagation** du signal dans l'ensemble du nano-circuit.

3.3. La porte inverseuse

La porte inverseuse prend la forme de deux virages courbes symétriques se joignant en un sommet pointu (cf. figure 13). Conformément aux remarques du paragraphe précédent, les parois de domaine vont se propager dans les virages qui composent la porte inverseuse en suivant le sens de rotation du champ magnétique externe. Lorsqu'un domaine magnétique s'étend à partir d'une des deux branches de la porte inverseuse jusqu'à la partie interne, il reste une paroi de domaine entre la partie interne et la deuxième branche. Cette paroi est prête à se propager une demi-période plus tard, c'est à dire lorsque le champ magnétique externe sera inversé. Compte tenu du rebroussement opéré par la piste magnétique, la direction de l'aimantation relative à la direction de propagation est désormais inversée. Ainsi, une porte inverseuse fournit au prochain retournement du champ externe le complément de l'état de son entrée. Si le champ externe est considéré comme une horloge, **elle conduit aussi à un déphasage d'une demi-période d'horloge (T/2) en plus de sa fonction de complément logique.**

3.4. La bifurcation

La bifurcation est réalisée en faisant se rejoindre deux nanofils. A l'approche du point de rencontre, la largeur des deux nanofils symétriques diminue, pour passer de 200 nm à environ 125 nm, avant de rejoindre le troisième nanofils. Cette structure peut être utilisée soit comme une porte logique à deux entrées si les deux pistes symétriques sont utilisées comme des entrées, soit comme un élément de distribution du signal (fan-out) si ce sont des sorties. Dans tous les cas, le point de rencontre entre les trois pistes se comporte comme un obstacle à la propagation des parois de domaines, il faudra donc un champ externe suffisant pour franchir cet obstacle, mais le seuil de propagation dépend des configurations rencontrées.

- Si une paroi de domaine apparaît sur **la piste large ou simultanément sur les deux pistes étroites**, le seuil de propagation est proche de celui des autres éléments composant les circuits magnétiques, appelons le «seuil **normal**».
- A contrario, lorsque la paroi provient d'**une seule des deux branches étroites**, la propagation requiert un champ magnétique externe légèrement plus important, tout en restant inférieur au seuil de nucléation, appelons le «seuil **haut**».

Pour utiliser l'élément de bifurcation en tant que porte logique, **il faut un champ asymétrique**. Dans ce cas, le champ tournant est elliptique, avec un décalage. Ce dernier permet au champ externe de ne dépasser le seuil haut que dans une direction. La propagation des domaines correspondant à cette direction est facilitée par le décalage qui rend possible le basculement des sorties des portes à deux entrées, même si les deux entrées ont des états différents. On peut noter dans ce cas qu'il reste une paroi de domaine sur l'autre entrée, or, la structure matérielle d'une porte logique étant identique à celle d'un élément de distribution, **cette nouvelle paroi repart en arrière**, au moins temporairement, jusqu'à rencontrer un virage qu'elle ne pourra pas franchir du fait de la direction de sa courbure. Par ailleurs, l'état d'aimantation facilité par le décalage du champ magnétique externe, puisqu'il permet de faire basculer la sortie quel que soit l'état de l'autre entrée, correspond d'un point de vue logique, à un état absorbant pour la porte logique considérée. Bien entendu, si l'état absorbant correspond au '0' logique, nous avons affaire à une porte ET alors que s'il s'agit d'un état '1', nous avons affaire à une porte OU. Par ailleurs, comme le champ magnétique et son décalage sont globaux sur l'ensemble du circuit, toutes **les portes logiques orientées dans la direction du décalage du champ sont des portes OU alors que celles qui s'y opposent sont des portes ET** (cf. figure 14). En conséquence, il ne sera possible

de cascader les deux types de portes qu'après avoir opéré un retournement de la direction de propagation, c'est-à-dire après deux virages à 90° qui induisent un retard de phase de 180°.

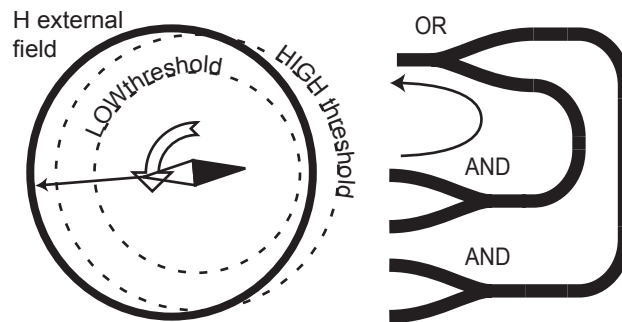


Figure 14 : Le décalage du champ elliptique externe détermine les fonctions des portes logiques suivant leur orientation.

En conclusion, **nous disposons effectivement des opérateurs fondamentaux de la logique booléenne : portes NON, ET et OU**. A priori, cela est suffisant pour réaliser n'importe quelle fonction combinatoire. Cependant les déphasages qui viennent s'intercaler pourraient compliquer cette démarche, ce que nous allons vérifier.

4. Modélisation et simulation comportementales

La synthèse de circuits logiques basés sur la propagation de paroi de domaine magnétique requiert d'une part la validation d'une bibliothèque de fonctions élémentaires et d'autre part une méthode de vérification des circuits complexes avant fabrication. Cette validation passe principalement par la simulation. Je vais maintenant décrire une méthode de modélisation et de simulation comportementale adaptée à la vérification des circuits à paroi de domaine magnétique.

4.1. Choix d'un simulateur événementiel

La simulation de circuits logiques à paroi de domaine nécessite de manipuler des grandeurs continues voire vectorielles (champs magnétiques, aimantation), ce qui pourrait conduire à utiliser un simulateur mixte pour prendre en compte les grandeurs magnétiques. Pourtant, la simulation que nous envisageons ne requiert pas de résolution d'équation différentielle non-linéaire. Dans ces conditions, un simulateur logique purement événementiel est plus efficace et beaucoup plus rapide. Il permet notamment d'éviter les problèmes de convergence rencontrés aux discontinuités lorsque la simulation passe par la résolution d'équation différentielle. Pour ces raisons, j'ai choisi VHDL comme langage de description. Les simulations ont été effectuées d'abord avec le simulateur Model Sim XE III starter 6.0a de ModelTechnology (outil disponible en téléchargement sur le site de Xilinx) puis sous Cadence, avec le simulateur simVision.

4.2. Codage des grandeurs magnétiques

Les informations manipulées dans une simulation de circuit magnétique sont représentées par le champ magnétique externe et par l'aimantation des différents tronçons de piste du circuit. Un composant VHDL est chargé de générer deux signaux représentant le champ magnétique externe. Un signal représentera l'orientation du champ et l'autre son intensité. L'orientation du champ est codée sur 8 valeurs par pas de 45° en tournant dans le sens horaire, le code 0 correspond à un champ horizontal dirigé

vers la droite. L'intensité en Oestred est représentée par un nombre réel. Ces deux signaux sont distribués à tous les modèles de composants magnétiques. La génération de l'orientation du champ est faite par un compteur (resp. un décompteur) modulo 8 lorsque le champ tourne dans le sens horaire (resp. anti-horaire). La génération de l'amplitude est tabulée (cf. figure 15).

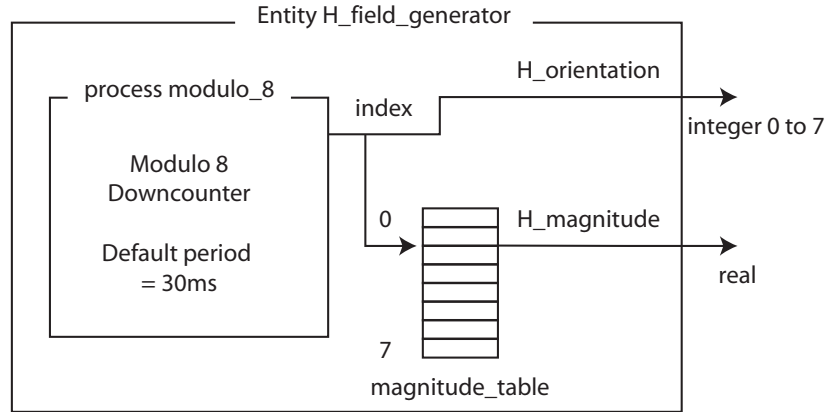


Figure 15 :Génération des signaux correspondant au champ magnétique externe

Chaque élément magnétique est composé de tronçons ayant chacun une orientation codée de la même manière que la direction du champs externe. A chacun de ces tronçons sont associés deux signaux représentant la direction de l'aimantation. Le premier est absolu (codé de la même manière que le champ externe) et le deuxième, binaire, représente la direction relative à la direction de propagation habituelle des parois de domaine : un niveau haut lorsque la direction de l'aimantation suit le sens de propagation et un niveau bas lorsqu'elle s'y oppose. La détermination du sens «normal» de propagation est précisée en associant une étiquette d'entrée magnétique ou de sortie magnétique à chaque extrémité du tronçon. Cela n'empêche pas la modélisation de prendre en compte toutes les directions de propagation possibles, des entrées vers les sorties mais également l'inverse. Pour ce faire, la connexion des différents composants magnétiques est réalisée à chaque extrémité via deux ports VHDL, l'un en entrée et l'autre en sortie, réalisant un double chaînage. Ceux-ci transmettent les informations d'aimantation relatives aux composants voisins, de façon à détecter la présence d'éventuelles parois de domaine aux extrémités des composants (cf. figure 16).

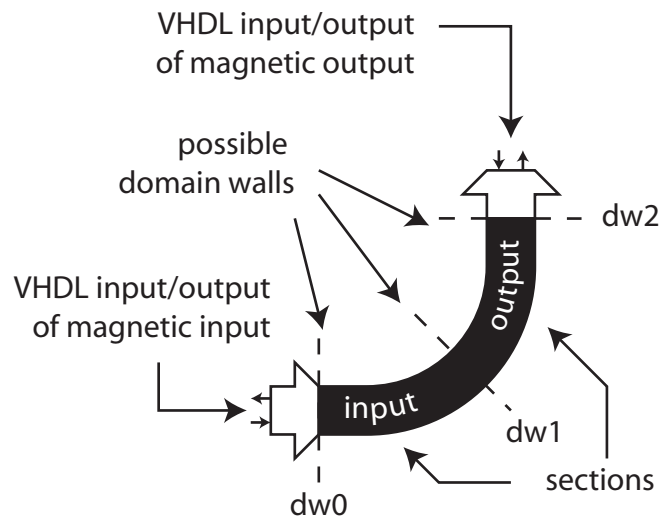


Figure 16 : Organisation d'un modèle de composant magnétique.

Lorsque la présence de directions d'aimantation relative différentes à l'interface entre deux tronçons est détectée par les deux éléments adjacents, ils en déduisent la présence d'une paroi de domaine susceptible de se propager si les conditions requises sont vérifiées dans le tronçon considéré. Ces conditions définissent les modèles comportementaux des composants magnétiques. Nous allons les préciser maintenant.

4.3. Modèle comportemental de propagation des parois

Dans presque tous les cas, la propagation d'une paroi détectée à l'interface entre deux tronçons adjacents a lieu si les deux conditions suivantes sont vérifiées :

- la direction du champ externe est opposée à celle de l'aimantation,
- le champ externe est supérieur au seuil de propagation seuil «normal» (typiquement 65 Oe).

Cependant, les conditions de propagations d'une paroi de domaine diffèrent lorsqu'on s'intéresse au point de rencontre de trois pistes magnétiques dans un élément en 'Y'. Dans ce cas, la propagation d'une paroi survient lorsqu'une des trois conditions suivantes est vérifiée :

- la direction du champ externe est opposée à celle de l'aimantation et il existe au moins **une** paroi de domaine magnétique **large** adjacente au tronçon considéré et le champ externe est supérieur à un seuil **normal** (65 Oe).

ou

- la direction du champ externe est opposée à celle de l'aimantation et il existe au moins **une** paroi de domaine magnétique **étroite** adjacente au tronçon considéré et le champ externe est supérieur à un seuil **haut** (75 Oe).

ou

- la direction du champ externe est opposée à celle de l'aimantation et il existe au moins **deux** parois de domaine magnétique **étroites** adjacente au tronçon considéré et le champ externe est supérieur à un seuil **normal** (65 Oe).

On peut noter que ce modèle comportemental est commun aux deux éléments en forme de 'Y', qu'il s'agisse des éléments de distribution du signal (fan-out) ou des portes logiques à deux entrées. Seule la convention entrée/sortie magnétique qui définit la direction d'aimantation relative et les niveaux logiques associés, est inversée (cf. figure 17).

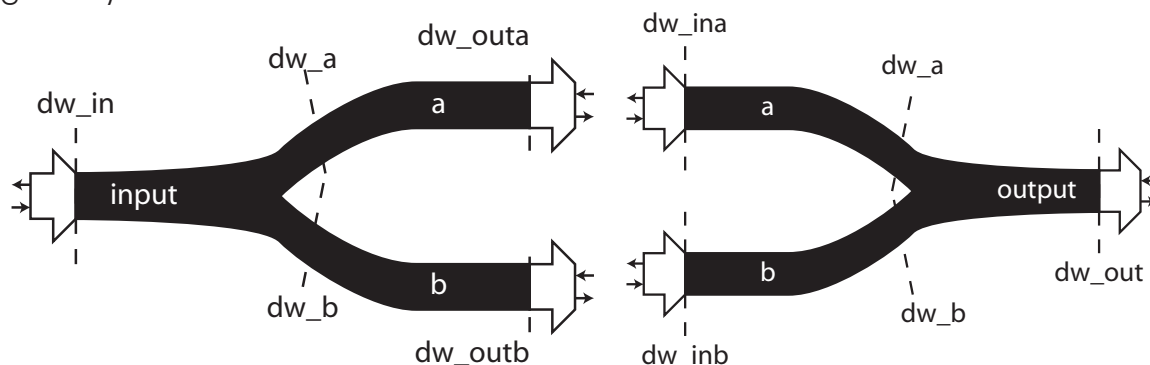


Figure 17 :Modélisation des éléments de distribution (à gauche) et des portes logiques (à droite). Seul le sens conventionnel de propagation du signal change, le modèle comportemental est identique.

Enfin, la porte inverseuse nécessite également un traitement particulier. Du fait du

rebroussement du sens de propagation dans la partie centrale (cf. figure 18), il y a nécessairement une rupture de continuité dans la convention du sens de propagation des parois. De plus, sa partie centrale étant orthogonale au sens habituel de propagation des parois, on ne peut pas y définir de direction privilégiée de propagation évidente.

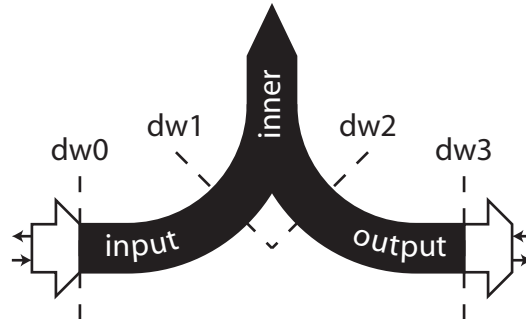


Figure 18 : Organisation du modèle de porte inverseuse. Par convention, nous considérons que le tronçon central est dans la continuité de l'entrée. En conséquence la paroi dw2 est présente si les orientations relatives sont **identiques** entre le tronçon central et le tronçon de sortie.

Nous avons considéré arbitrairement que le sens normal pour ce tronçon suivait celui de l'entrée, s'opposant donc à celui de la sortie. De ce fait, contrairement aux autres situations, une paroi est détectée à l'interface entre le tronçon interne et la sortie si la direction relative de l'aimantation est identique dans les deux tronçons.

4.4. Construction hiérarchique et orientation des composants

Lors de la modélisation hiérarchique d'un ensemble de composants connectés, l'orientation de chaque composant relativement au niveau hiérarchique courant doit être précisée au moment de son instanciation. L'orientation de chaque composant est codée par pas de 45° tournant dans le sens horaire, par un entier compris entre 0 à 7. Le code 0 correspond à un composant dont les entrées sont horizontales et dirigées vers la droite. Lorsque la hiérarchie comporte plusieurs niveaux imbriqués, les orientations des différents niveaux hiérarchiques s'ajoutent (modulo 8) à chaque niveau d'imbrication de façon à obtenir l'orientation absolue au niveau des composants primitifs (tronçons).

4.5. Résultats de simulations

Un nano-circuit à trois composants a été modélisé sous la forme du schéma de la figure 19. Il est fonctionnellement proche de celui de la figure 12.

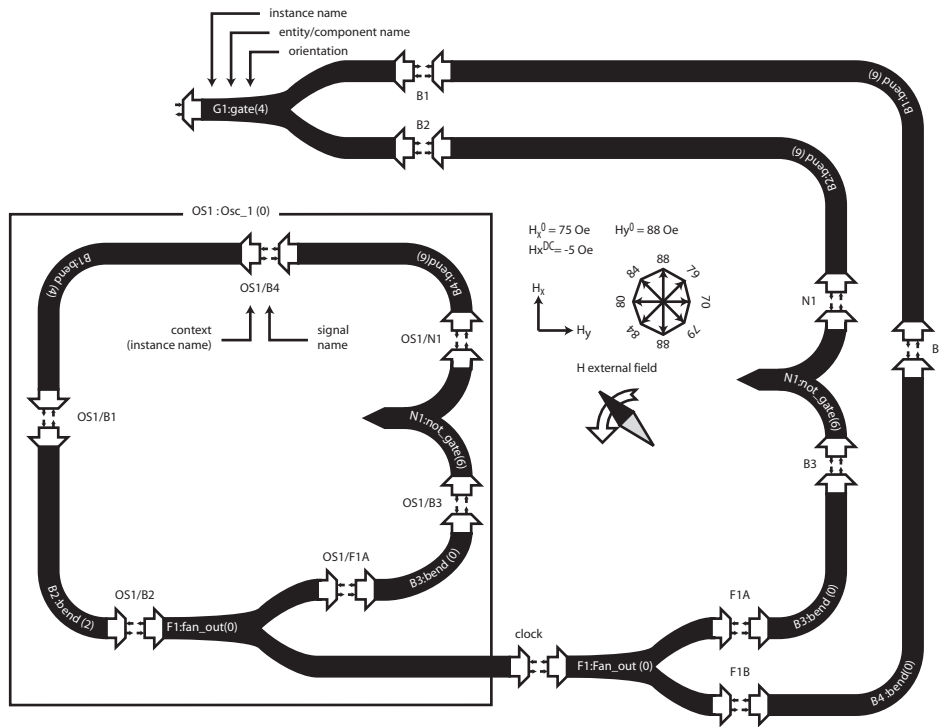


Figure 19 : Schéma du circuit simulé sur la figure 20.

Le résultat de la simulation est donné sur la figure 20. On peut comparer ce résultat avec celui fournie dans [12] et vérifier notamment que les sorties correspondent.

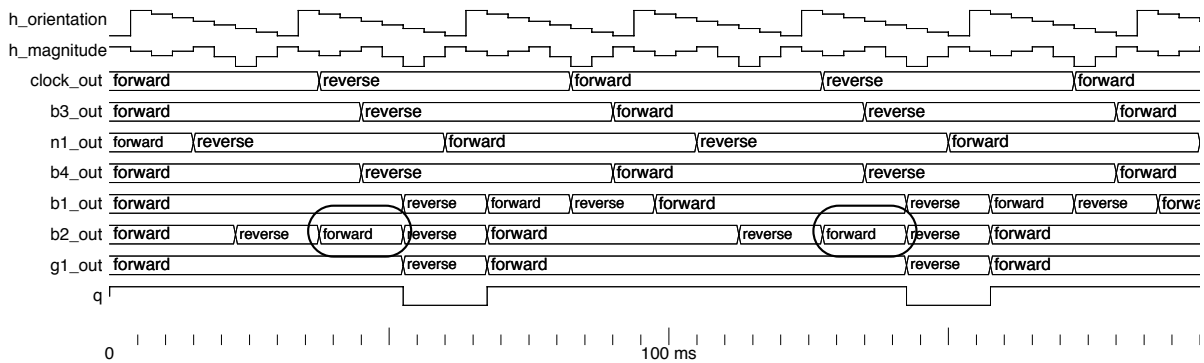


Figure 20 : Résultats de simulation du circuit magnétique de la figure 19.

Seule différence, la rétropropagation des parois de domaines correspondant à des niveaux logiques absorbants pour cette porte logique n'apparaît pas sur la forme d'onde dans la référence [12] tableau 1 page 1689¹.

4.6. Conclusion sur la simulation

Nous avons défini une méthode de simulation comportementale des circuits à propagation de paroi de domaine magnétique. Cette méthode conduit à décrire structurellement un circuit par instanciation d'un des quatre composants primitifs décrits en VHDL : porte inverseuse, virage, bifurcation et porte logique. Ainsi, la simulation événementielle d'une architecture de circuits complexes devient ultra rapide. Ceci ouvre la

¹ La légende de la figure 2A de la référence [12] précise qu'il ne s'agit pas d'une mesure de l'état d'aimantation mais d'une trace **inférée** à partir de la sortie de la porte logique.

voie à la recherche des fonctions de haut niveau réalisables avec une technologie à propagation de paroi magnétique.

5. Technique de synthèse de circuit à paroi de domaine

La démonstration de la faisabilité des fonctions logiques fondamentales (porte inverseuse [13], porte ET [14]) fait espérer que les puissantes techniques de synthèse automatique utilisées pour les systèmes sur puce pourront être directement transposées aux architectures magnétiques. La suite de ce chapitre va montrer pourquoi ces espoirs de transpositions directes rencontrent rapidement leurs limites. Nous proposons alors une technique de synthèse adaptée à ces nouvelles architectures.

5.1. Fonctions à implanter

La synthèse d'un système sur puce (SoC) va requérir principalement l'ensemble des éléments qui entre dans la composition d'une architecture de système à microprocesseur : périphériques d'entrées-sorties, mémoires, bus, unités de calcul, registres et automate de contrôle. Dans cette liste, les MRAMs constituent sans aucun doute l'élément magnétique le plus avancé aujourd'hui. Nous ne nous y attarderons pas plus ici. Nous n'apporterons aucun élément nouveau du côté des entrées-sorties non plus. Les bus requièrent un élément de multiplexage de façon à être partagés par plusieurs émetteurs. Le multiplexeur sera notre premier objectif de synthèse. Le calcul requiert de la logique booléenne pour les fonctions de base et un automate pour les fonctions plus évoluées. Les registres, notamment synchrones, seront utiles en tant qu'éléments temporaires de mémorisation comme pour les automates ou les éventuels calculs pipelinés. Il apparaît donc que la fonction bascule synchrone et la fonction multiplexeur seraient bien utiles. Pour aller plus loin, une méthode de synthèse d'automates permettrait d'envisager la plupart des applications. Pour commencer, intéressons nous déjà à l'élément le plus simple : le multiplexeur.

5.2. Synthèse d'un multiplexeur magnétique

Un multiplexeur recopie sur son unique sortie Y une de ses entrées de donnée I_n parmi 2^N désignée par une entrée d'adresse SEL codée sur N bits. Dans sa forme la plus simple ($N=1$), l'entrée I_0 est recopiée sur la sortie Y quand l'adresse de sélection SEL = 0 ou l'entrée d'indice I_1 est recopiée sur la sortie Y quand SEL = 1. Cette dernière proposition se traduit trivialement par l'expression booléenne suivante :

$$Y = (I_0 \text{ ET NON (SEL)}) \text{ OU } (I_1 \text{ ET SEL}) \quad (1)$$

ou encore, par application du théorème de De Morgan

$$Y = \text{NON (NON (I_0 \text{ ET NON (SEL)}) \text{ ET NON (I_1 \text{ ET SEL}))} \quad (2),$$

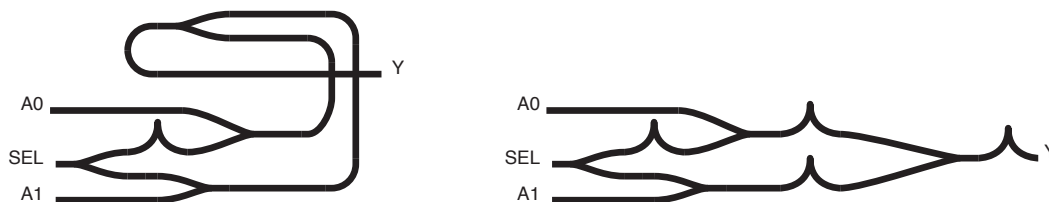


Figure 21 : Deux architectures possibles de multiplexeurs magnétiques.

Considérons maintenant la transposition directe de ces deux expressions logiques en éléments magnétiques. La première expression requiert des portes ET et des portes OU, ce qui ne peut pas se faire avec un architecture rectiligne, contrairement à la seconde solution (cf. figure 21). Pourtant, si on souhaite que le signal de sortie ait une

direction de propagation identique à celle du signal d'entrée, la sortie sera identique dans les deux cas : globalement retardée d'une période mais retardée d'une demi-période supplémentaire pour le signal NOT (SEL) :

$$Y(t) = (I0(t-1) \text{ ET NON (SEL (t-3/2))}) \text{ OU } (I1(t-1) \text{ ET SEL (t-1)}) \quad (3).$$

Le retard dû à la porte NON fait apparaître une demi-période de transition au niveau de chaque front du signal de sélection : l'une où aucun signal n'est dirigé vers la sortie, qui prendra donc la **valeur logique '0'** et l'autre où les deux entrées entrent en concurrence pour imposer l'état de la sortie qui sera alors **le OU logique des entrées**. La fonction multiplexeur paraît réalisable tant que les états parasites, qui vont apparaître avec les fronts du signal de sélection, n'obèrent pas le fonctionnement du reste du système. Voyons par exemple ce qui se passe si on tente de reboucler ce multiplexeur pour en faire une bascule D-latch.

5.3. Bascules D-latch et D-edge

Une bascule D mémorise son entrée de donnée 1 bit D. On distingue les bascules D-latch (verrou), sensibles au niveau du signal d'écriture et les bascule D-edge, sensibles aux fronts du signal d'horloge. Une bascule D-latch est transparente lorsque le signal de verrouillage LE (Latch Enable) est inactif, ce qui signifie que toutes variations de l'entrée se répercutent aussitôt sur sa sortie. À l'instant précis où le signal LE devient actif, la donnée présente sur son entrée est échantillonnée et restera verrouillée sur la sortie tant que le signal LE restera actif. Du fait de l'existence d'un état transparent, la bascule D-Latch ne peut pas être utilisée dans un système bouclé synchrone. Au contraire, une bascule D-edge n'est jamais transparente, ce qui permet de l'utiliser comme registre d'état d'une machine de Moore [15-16] pour séparer l'état futur de l'état courant (cf. figure 22).

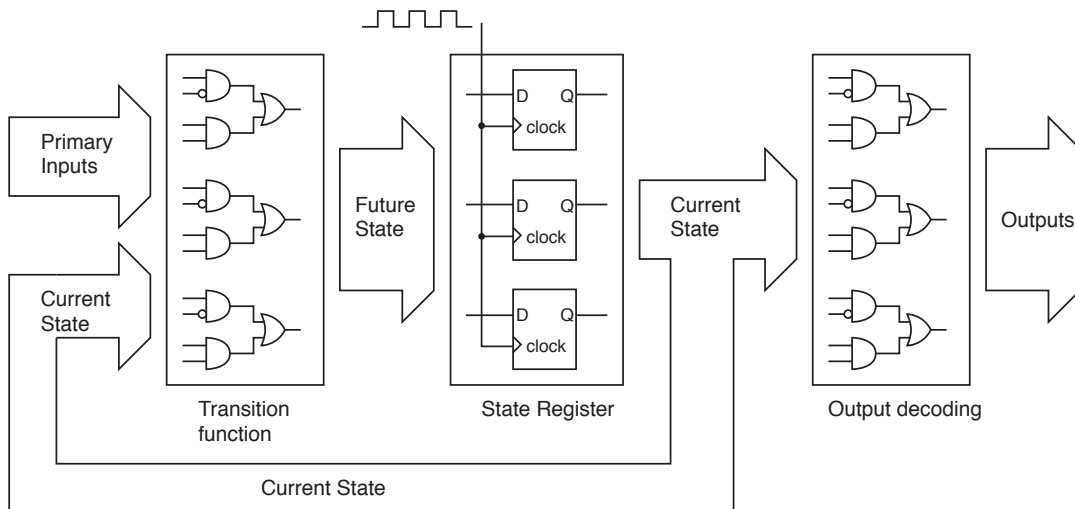


Figure 22 : Machine de Moore

Deux architectures de bascules D-latch peuvent être envisagées suivant que le signal d'écriture sera actif à l'état haut ou à l'état bas.

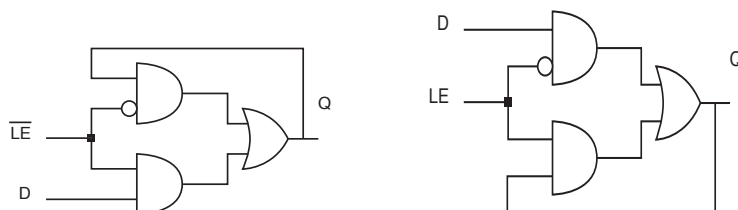


Figure 23 : Bascule D-Latch à signal de verrouillage actif sur niveau bas (à gauche) ou sur niveau haut (à droite).

On peut ainsi espérer cascader les deux types de bascules de façon à réaliser une bascule D-edge de type maître esclave (cf. figure 23). Malheureusement, l'analyse de l'effet d'une écriture sur l'une ou l'autre de ces bascules D-latch révèle quelques surprises. Une bascule D-latch peut être réalisé à partir d'un multiplexeur dont la sortie est rebouclée sur l'une de ses deux entrées de donnée. Le fonctionnement de la bascule D-latch se déduit donc de l'analyse du fonctionnement du multiplexeur. Les figures 24 et 25 montrent respectivement les formes d'onde obtenues lors d'une écriture dans une D-latch à signal de verrouillage LE actif sur état bas ou actif sur état haut.

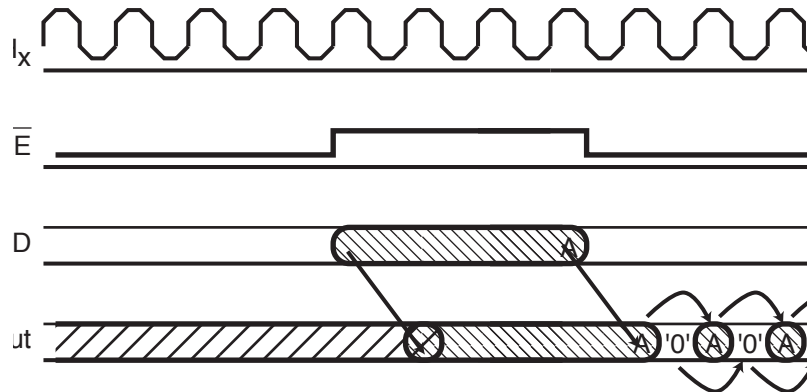


Figure 24 :Formes d'ondes obtenues lors d'une écriture dans une bascule D-latch à verrouillage actif sur état bas.

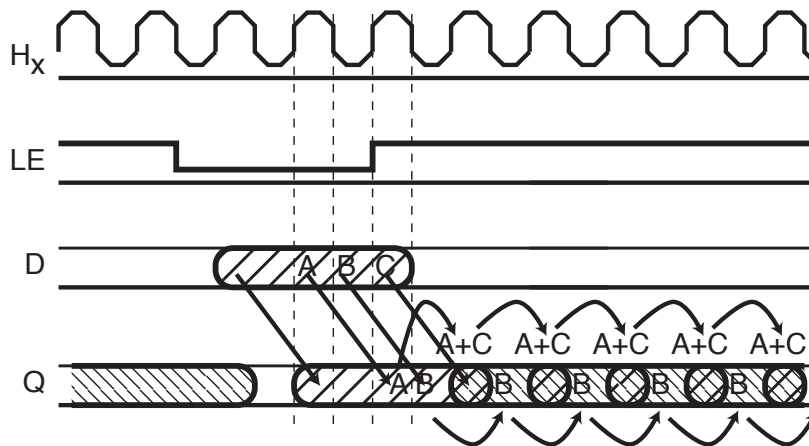


Figure 25 :Formes d'ondes obtenues lors d'une écriture dans une bascule D-latch à verrouillage actif sur état haut.

Dans le cas d'un signal de verrouillage actif sur niveau bas (cf. figure 24), **la sortie fait apparaître un '0'** en fin de cycle d'écriture, inséré par le front descendant du signal / LE. **Ce '0' est mémorisé et se répète en sortie de façon périodique** toutes les demi-périodes du champ externe, ou plus précisément lors de l'état bas du champ externe. De façon symétrique, dans le cas d'un signal de verrouillage actif sur niveau haut (cf. figure 25), la sortie montre que **le «conflit» en fin de cycle** d'écriture inséré par le front montant du signal LE est mémorisé et **se répète en sortie de façon périodique** à chaque état haut du champ externe. La sortie dans ce cas représente le OU logique de l'état en mémoire et de la nouvelle entrée. Ainsi, il sera possible d'écrire un état haut, mais jamais de revenir à un état bas.

En conclusion, utiliser un multiplexeur pour construire une bascule D-latch puis une bascule D-edge de type maître-esclave mène à une impasse. La désynchronisation des signaux d'entrée par le multiplexeur introduit des états parasites périodiques en sortie des bascules D-latch, ce qui les rend inutilisables. Pour contourner cette difficulté, et sachant qu'il s'agit de composants magnétiques possédant des propriétés intrinsèques de non volatilité, voyons s'il est possible de construire un automate sans utiliser d'élément de mémorisation explicite. Notre première tentative concerne un compteur, simple à tester matériellement par lecture magnéto-optique du fait de la périodicité des signaux, si cela s'avérait utile.

5.4. Compteur binaire magnétique

Un compteur binaire N bits fournit à chaque cycle d'horloge la configuration de bit suivante, modulo 2^N , dans l'ordre du codage binaire naturel. Pris séparément, chaque bit de rang n fournit un signal dont la fréquence correspond à celle de l'horloge divisée par 2^n . Deux approches peuvent être suivies pour synthétiser un compteur : l'une asynchrone et l'autre synchrone. Le compteur asynchrone est réalisé en cascade de des éléments de division de fréquence par deux.

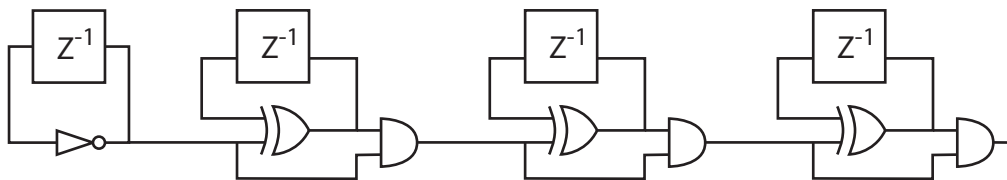


Figure 26 : Tentative de construction d'un compteur par mise en cascade de bascules T.

Comme tout élément asynchrone, il est très sensible aux impulsions parasites, c'est pourquoi on lui préfère souvent le compteur synchrone. Dans ce dernier, l'horloge est globale, d'où son nom. Il est réalisé grâce à un mécanisme de chaînage qui propage de façon combinatoire l'information de retenu pour l'étage suivant. Cette opération consiste simplement à agréger les sorties des bascules par un opérateur ET au niveau de chaque bit (cf. figure 26).

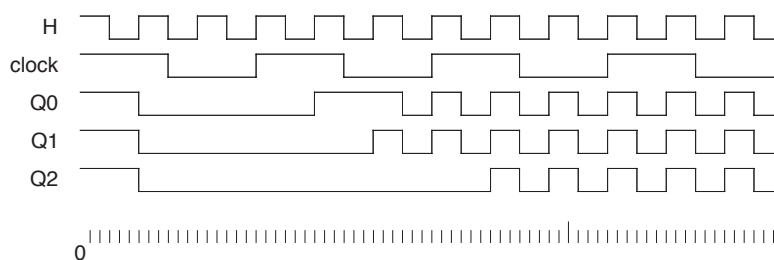


Figure 27 : Chronogramme des sorties d'un compteur magnétique 3 bits formé par la mise en cascade de bascules T.

Comme le ET magnétique est effectivement combinatoire, cela ne pose pas de problème particulier. Reste à vérifier qu'il est possible de cascader des opérateurs de division de fréquence par deux. Pour cela, l'opération à réaliser est le ou-exclusif de l'entrée et de la donnée mémorisée. On appelle cet élément une bascule T (Toggle). La simulation de ce compteur a été effectuée par le modèle VHDL présenté sur la figure 26. On constate sur la figure 27 que le compteur rentre rapidement dans un cycle d'oscillation où toutes ses sorties se mettent à osciller en suivant le champ externe. Pour comprendre ce qui se produit, nous pouvons détailler l'analyse de l'opérateur

OU-EXCLUSIF et étudier son impact sur le fonctionnement de la bascule T.

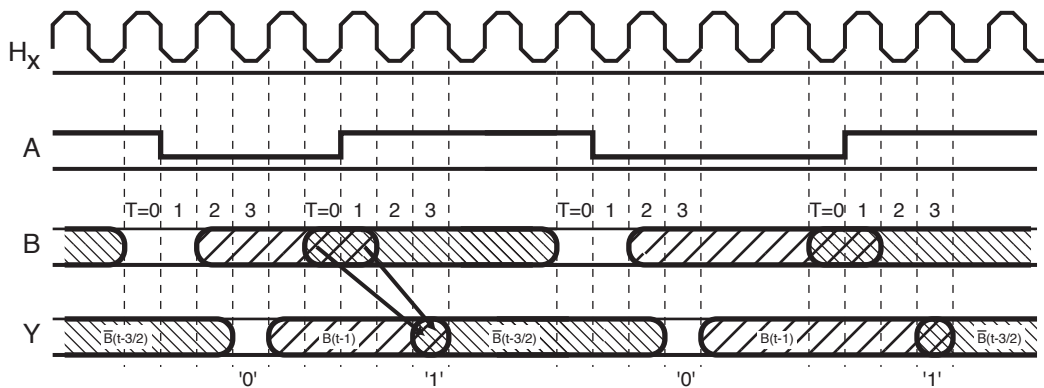


Figure 28 : Chronogramme du fonctionnement d'un OU-EXCLUSIF

Rappelons que la sortie d'une porte OU-EXCLUSIF à deux entrées recopie l'une des entrées lorsque l'autre est au niveau bas ou recopie le complément de l'une des entrées si l'autre est au niveau haut. On s'attend donc à ce que la sortie Y soit la copie de A lorsque B est à l'état bas et le complément de A lorsque B est à l'état haut. C'est le cas, sauf lors des transitions de B : trois demi-périodes après le front montant de B, on trouve systématiquement un état '1' et trois demi-périodes après un front descendant de B on trouve un état '0' (cf. figure 28). Il s'agit du même phénomène que celui observé pour le multiplexeur, ce qui est logique au demeurant puisqu'un OU-EXCLUSIF multiplexe les entrées A et /A. Ici, la collision OU se fait entre /A avant le front de B et A après le front de B. Comme il s'agit du front montant de B et que les fronts montants et descendants ne peuvent pas être synchrones, A ne peut pas présenter de front descendant à cet instant, si bien que la configuration /A (0)=0 (c'est-à-dire A (0)=1) et A (1)=0 n'est pas possible. Or c'est la seule qui conduirait à un état '0' en sortie de la porte OU, ce qui explique l'état '1' systématiquement introduit. De même, comme pour le multiplexeur, on trouve un état '0' trois périodes après un front descendant du signal B.

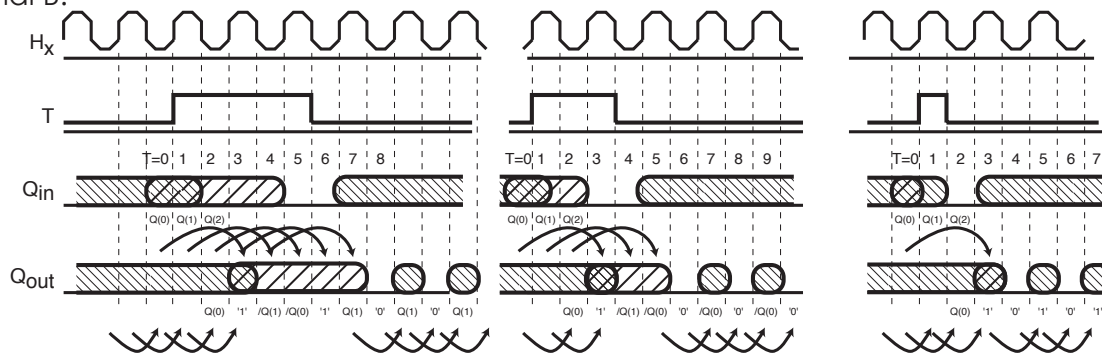


Figure 29 : Chronogramme du fonctionnement d'une bascule T.

L'analyse du fonctionnement de la bascule T montre de même que l'introduction d'états parasites perturbe son fonctionnement (cf. figure 29).

- Lors d'une impulsion de largeur cinq demi-périodes sur le signal T on trouve en sortie un état '0' aux instants $8+2k$ après le front et $Q(1)$ aux instants $7+2k$. Ceci est dû à une double inversion du signal $Q(1)$.
- Lors d'une impulsion de largeur trois demi-période sur le signal T on trouve en sortie un état '0' aux instants $6+2k$ après le front et $/Q(0)$ aux instants $5+2k$. Il y a donc bien une seule inversion du signal $Q(0)$.

- Lors d'une impulsion de largeur demi-période sur le signal T on trouve en sortie un état '0' aux instants $4+2k$ après le front et $Q(1)$ aux instants $1+2k$.

Ainsi, avec une durée d'impulsion adaptée, l'inversion de l'état mémorisé est possible mais dans tous les cas la sortie est polluée par l'introduction de '0' parasites.

De cette tentative, nous pouvons conclure que l'introduction de retard associé à chaque inversion logique gêne considérablement l'usage des méthodes de synthèse logique traditionnelles, adaptées aux technologies électroniques. Il faudra donc développer des méthodes de synthèse adaptées à la logique à paroi de domaine. Par ailleurs, nous allons voir que la direction de propagation des signaux nécessite aussi une attention particulière.

5.5. Retro-propagation des parois de domaines

Afin d'illustrer l'impact de la rétro-propagation du signal considérons le circuit de la figure 30 où le champ magnétique externe est décalé vers la gauche. Lors de l'étape 1, une paroi de domaine entre dans la porte G1 par le point A. Grâce au décalage du champ qui lui est favorable, la paroi de domaine se propage à travers la porte G1 et fait basculer sa sortie Q1.

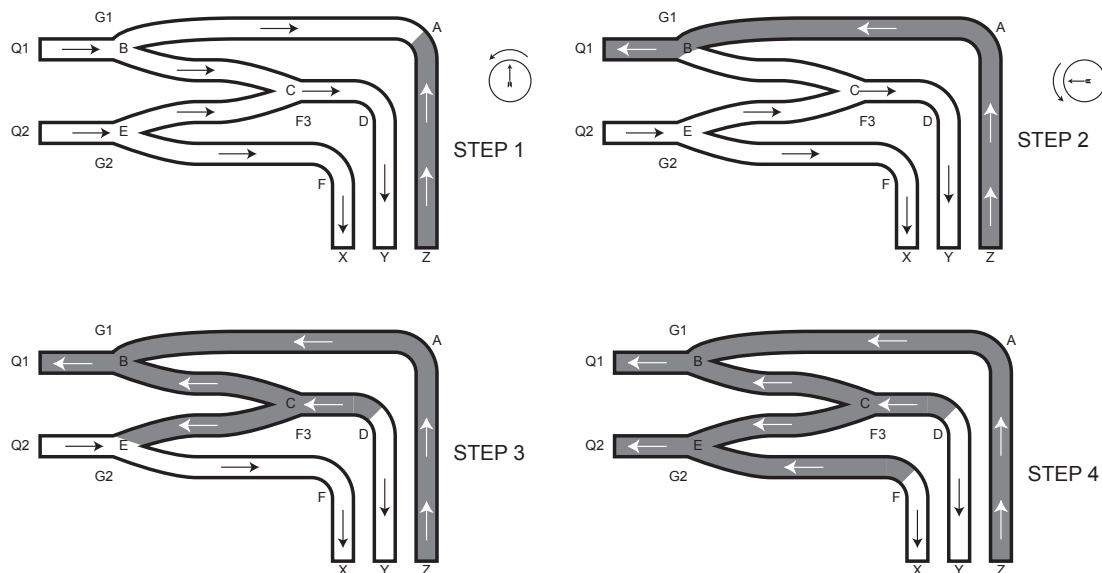


Figure 30 : Retropropagation des signaux dans les circuits magnétiques.

Comme nous l'avons déjà dit, une nouvelle paroi de domaine apparaît (étape 2, point B) qui va se propager à rebours vers l'entrée. Elle atteint alors la bifurcation F3 (étape 3 point C) qui est également réversible et se comporte donc comme une porte logique. En conséquence, ce nouvel état se propage à travers F3 (étape 3) puis G2 (étape 4) jusqu'aux points D, F et Q2. La propagation jusqu'au point B et C ne pose pas de problème puisqu'elle est arrêtée par la courbure de la piste magnétique en ces points. A contrario, le basculement de la sortie Q2 constitue un dysfonctionnement intolérable. Dans cet exemple, les difficultés proviennent de la connexion directe d'un élément de distribution (fanout F3) vers des portes logiques. Comme le montre la figure 31, l'insertion systématique d'un virage entre ces deux éléments permet d'éviter ce problème.

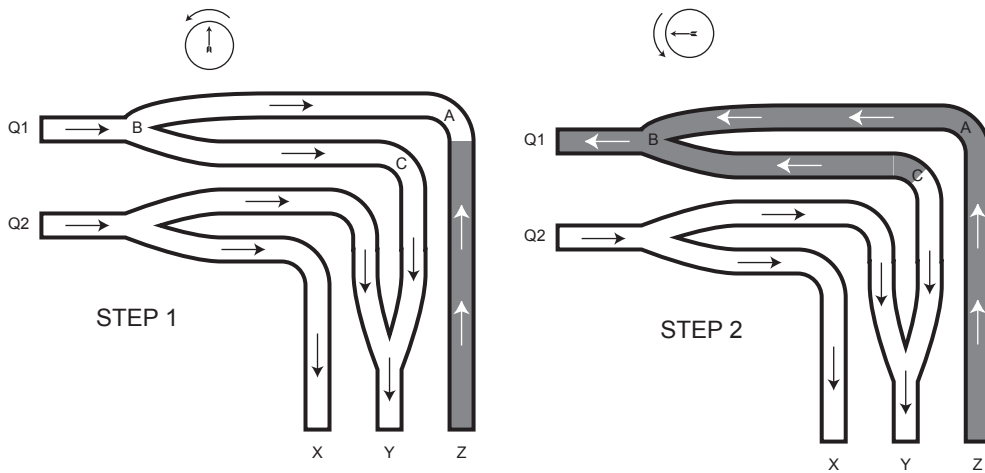


Figure 31 : L'insertion systématique d'un virage avant chaque porte logique supprime les effets néfastes de la rétropropagation des signaux.

En conclusion, l'effet de rétro-propagation du signal dans les portes logiques ne limite pas les possibilités d'utilisation de la logique à parois de domaines. Inversement, l'utilisation du seul champ externe pour cadencer le fonctionnement d'une machine de Moore magnétique n'est pas possible. Les portes inverseuses introduisent des retards qui se traduisent par des états parasites transitoires. Ceux-ci déclenchent des erreurs périodiques qui s'avèrent rédhibitoires dans un système bouclé. Nous allons donc considérer une autre piste. Puisque l'horloge ne peut pas être le champ externe, nous utiliserons un signal magnétique. Pour cela, nous allons construire une bascule D à partir d'une nouvelle bascule /RS verticale.

5.6. Bascule /RS verticale

En supposant un champ déséquilibré dans le sens horizontal, nous avons vu qu'il était possible de réaliser des portes logiques OU et ET. Si maintenant, le champ magnétique externe est équilibré et inférieur au seuil haut dans la direction verticale, on obtient le fonctionnement suivant :

- La sortie suivra les deux entrées si celles-ci sont identiques (fonction SET ou RESET).
- La sortie ne sera pas modifiée si les deux entrées sont différentes (fonction mémorisation).

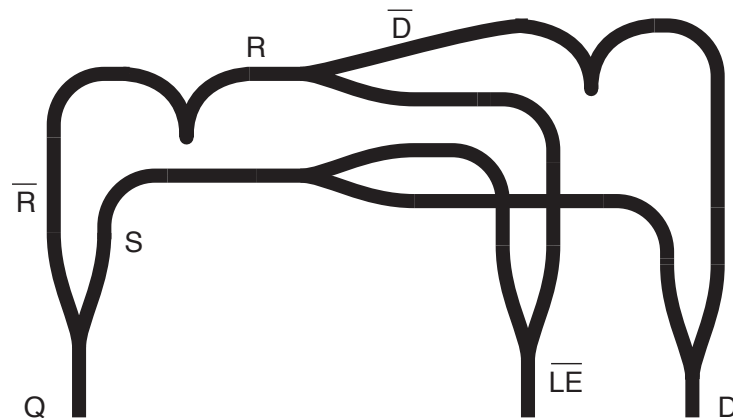


Figure 32 :Bascule D-Latch magnétique

Nous avons là un effet de mémorisation contrôlable. Si on appelle l'une des entrées /RESET et l'autre SET, nous obtiendrons une bascule /RS avec un RESET actif à l'état BAS et un SET actif à l'état HAUT. Bien sûr, il s'agit là d'une pure convention puisque les deux entrées sont symétriques. À partir de cette bascule, il devient possible de construire une bascule D-latch (cf. figure 32).

La figure 33 montre un exemple de fonctionnement d'une telle bascule.

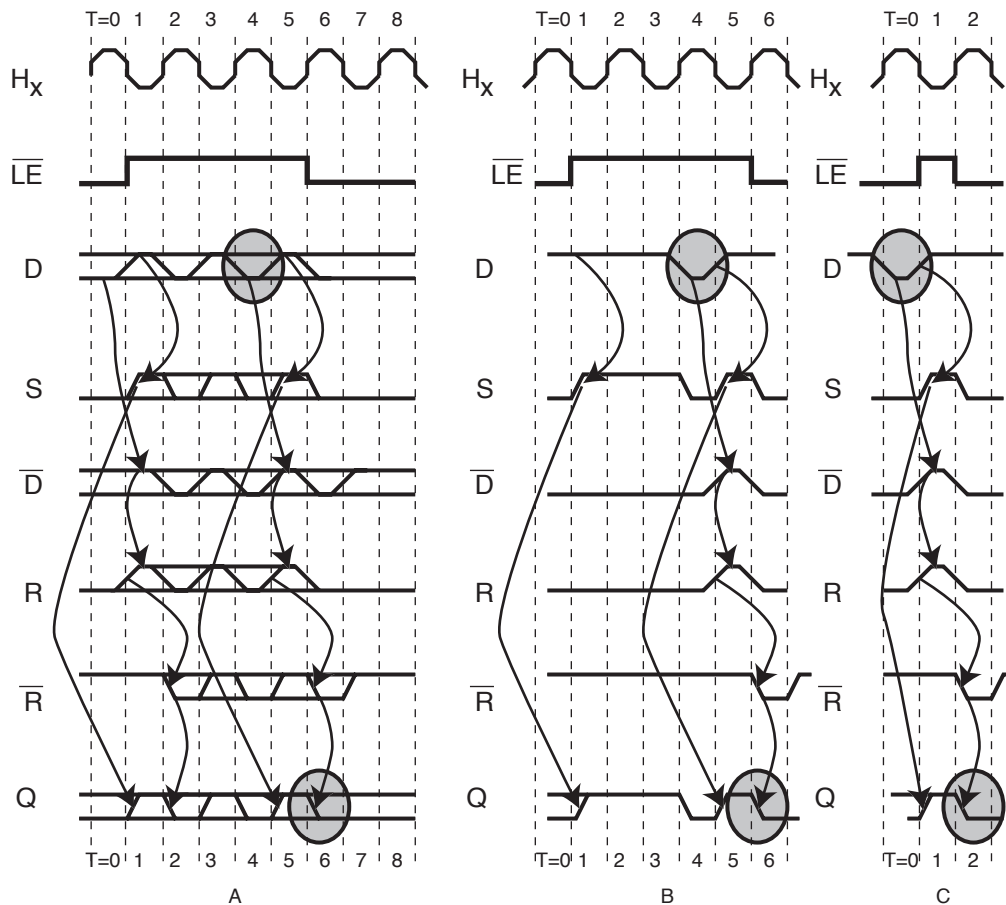


Figure 33 :Chronogramme de fonctionnement d'une bascule D-Latch magnétique.

Si l'entrée D est au niveau haut, la sortie réagira $1/4$ de période après le front montant du signal \overline{LE} . Si l'entrée est au niveau bas, la sortie sera actualisée $3/4$ de période après le front montant du signal \overline{LE} . Contrairement à la précédente version de bascule D-latch, la fonction d'écriture et de mémorisation n'est pas perturbée par ces retards. En particulier, lors du front descendant du signal \overline{LE} , aucune configuration parasite destructrice n'apparaît, ce qui garantit un échantillonnage correct de la donnée écrite. Notons au passage que l'état transparent, contrairement aux circuits électriques, ne fait pas courir le risque d'un retour de l'état sur lui-même. Le signal de sortie n'aura un impact sur le signal d'entrée qu'après au moins $3/4$ de période du champ externe. Partant de ces constats, on peut utiliser les bascule D-latch pour mémoriser l'état d'une machine de Moore. C'est ce que nous allons montrer dans le paragraphe suivant.

5.7. Architecture d'une Machine de Moore magnétique

Une machine de Moore implémente un automate synchrone. L'automate est un gra-

pe dont les sommets sont les états et les arcs des transitions.

- À chaque état est associée une sortie. Le décodage des sorties est une fonction combinatoire de l'état courant.

- À chaque transition est associée une condition sur les entrées. Les transitions sont représentées par une fonction combinatoire des entrées primaires et de l'état courant pour déterminer l'état futur.

Ainsi, la synthèse d'un automate revient à synthétiser deux fonctions combinatoires : l'une pour les transitions et l'autre pour le décodage des sorties.

Les bascules D-latch précédemment décrites permettent de verrouiller l'état courant d'un automate. Leur structure en angle nous conduit à placer ces bascules dans un virage du circuit magnétique. Nous proposons alors l'architecture de la figure 34.

La fonction de transition est réalisée par des produits de sommes logiques (et non par des sommes de produits). Cela permet d'enchaîner la fin les produits logiques directement avec les entrées des bascules, les deux fonctions nécessitant des portes ET. Le paragraphe suivant, présente une instance de cette architecture pour un compteur.

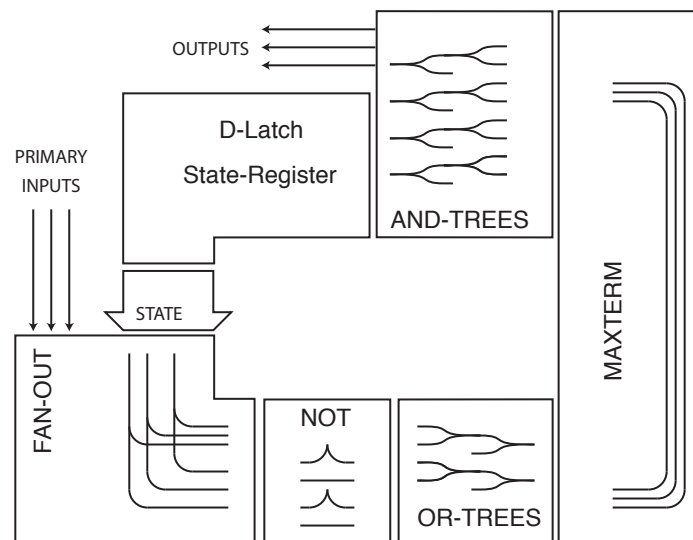


Figure 34 :architecture d'une machine de Moore magnétique

5.8. Compteur modulo 8

A titre d'illustration du paragraphe précédent, nous avons choisi de concevoir et de simuler un compteur 3 bits. Son implantation sous forme de produit de somme correspond aux expressions logiques suivantes :

$$D0 = \text{NOT } Q0$$

$$D1 = (\text{NOT } Q0 \text{ OR } \text{NOT } Q1) \text{ AND } (Q0 \text{ OR } Q1)$$

$$D2 = (\text{NOT } Q0 \text{ OR } \text{NOT } Q1 \text{ OR } \text{NOT } Q2) \text{ AND } (Q2 \text{ OR } Q1) \text{ AND } (Q2 \text{ OR } Q0)$$

Le schéma de la figure 35 correspond à son implantation magnétique.

Nous avons d'abord modélisé le comportement de ce circuit sous forme de portes logiques et de retards afin de vérifier que la désynchronisation opérée par les éléments inverseurs est effectivement contrecarrée par l'introduction de bascules D-Latch. Nous avons ensuite modélisé l'ensemble du compteur de façon comportementale. Cette deuxième simulation prend donc en compte les éventuels effets de retropropagation. La simulation de la figure 36 nous permet de vérifier qu'il s'agit bien d'un compteur modulo 8 parfaitement fonctionnel.

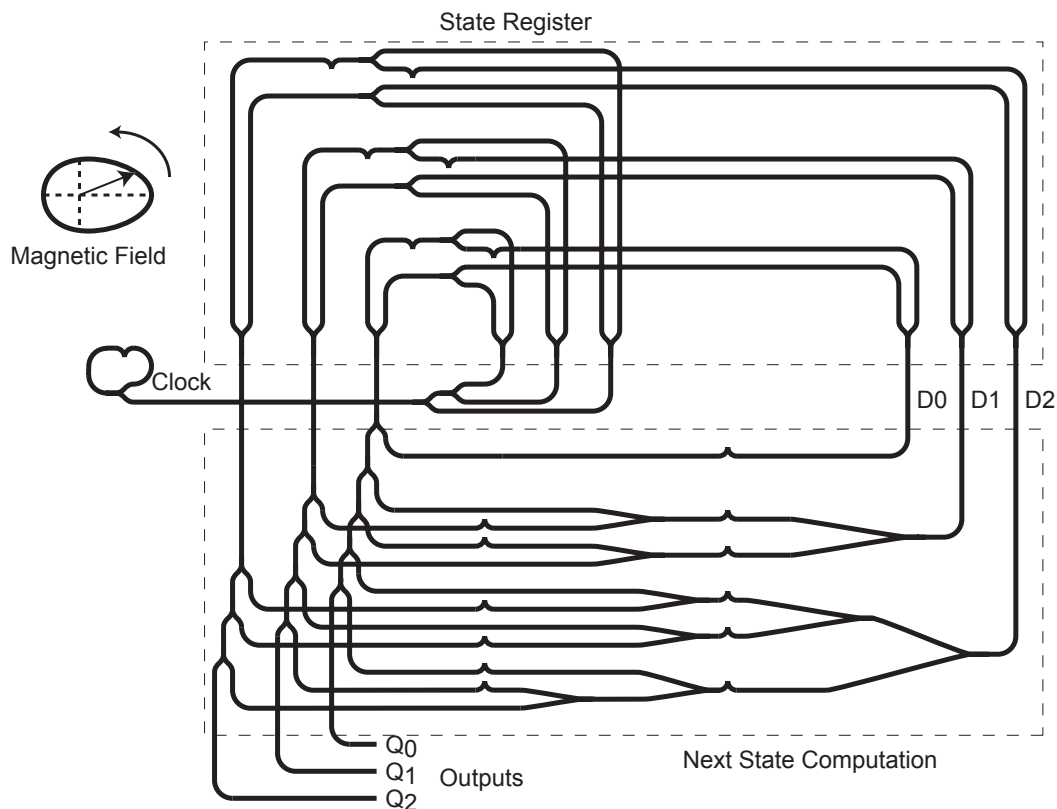


Figure 35 :Compteur magnétique

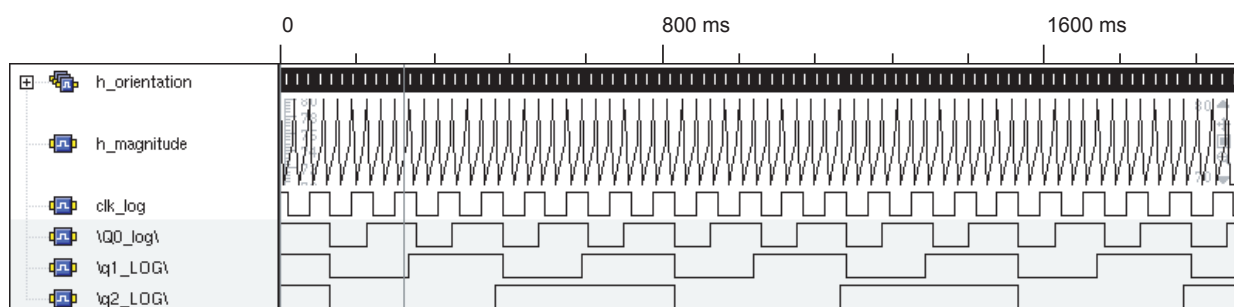


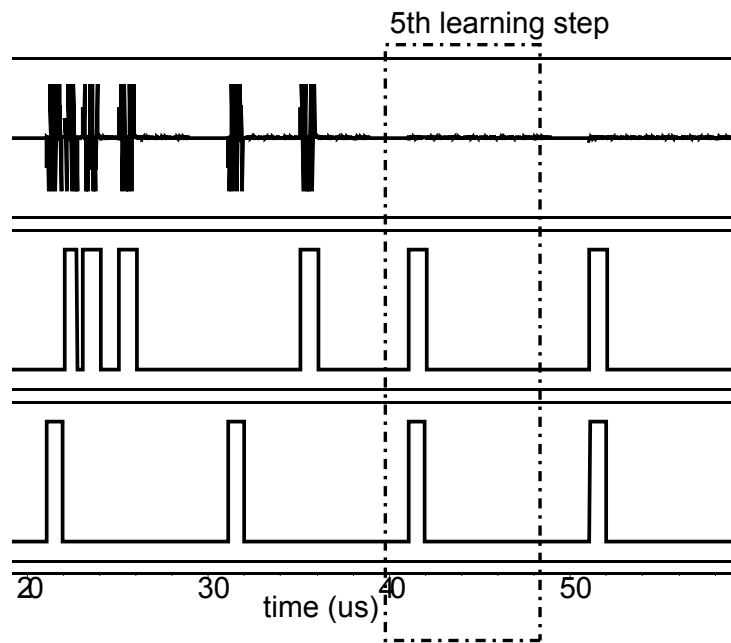
Figure 36 :Simulation VHDL comporementale d'un compteur magnétique

6. Conclusion

Nous avons étudié la transposition des techniques de synthèse d'architecture numérique CMOS pour des circuits à paroi de domaine magnétique. Cette étude montre que la transposition n'est pas directe. En particulier, nous avons découvert l'importance du déphasage introduit par les portes NON et montré que le champ externe ne peut pas directement faire office de signal d'horloge pour un automate. Pour contourner ces difficultés, nous avons proposé une nouvelle bascule /R S verticale à partir de laquelle nous avons construit une D-latch qui permet de réaliser n'importe quel automate. La très grande généralité de ce modèle permet d'envisager la synthèse de la plupart des circuits numériques.

La présente étude concernait les circuits magnétiques utilisant une aimantation dans le plan du circuit. Parmi les perspectives, le recours à des pistes magnétiques présen-

tant une anisotropie perpendiculaire ouvre de nouvelles perspectives. D'une part, ce type de circuits autorise une densité d'intégration supérieure. D'autre part, la génération du champs externe nécessaire pour la propagation est bien plus simple puisqu'il est simplement alternatif et non elliptique. Enfin, la topologie des circuits n'est pas contrainte par le rôle des virages. Par ailleurs, les parois de domaines peuvent également être poussées par des courants polarisés en spin, ce qui permet d'avantage de sélectivité qu'un champs globale à l'ensemble du circuit.



Quatrième partie
Architectures neuro-inspirées

III. Quatrième partie : Architectures neuro-inspirées

1. Introduction

L'industrie microélectronique a besoin de faire croître constamment la densité d'intégration des composants de façon à proposer de nouvelles fonctions et créer ainsi de nouveaux marchés. Comme le souligne Claude Saunier dans ses rapports parlementaires, [17] cette progression permet une croissance importante de l'économie numérique à travers les nouveaux services proposés (internet, téléphonie mobile, ...) mais elle impacte également l'ensemble de l'économie mondiale par l'accroissement de productivité que ces technologies engendrent. De ce point de vue, l'imminence annoncée de la fin de la loi de Moore pour les circuits CMOS vers les années 2015-2020 apparaît comme catastrophique pour l'économie ; c'est pourquoi des solutions alternatives ou complémentaires aux technologies CMOS doivent être explorées. L'un des moyens pour continuer à faire progresser la densité d'intégration est de recourir à des objets dont les dimensions intrinsèques sont nettement inférieures à ce qu'on peut espérer faire de mieux avec des transistors CMOS. C'est le cas des transistors à nanotube de carbone ou à nanofil par exemple. Malheureusement, les caractéristiques électriques fortement dispersées de ces composants comme le grand nombre de défauts attendus ne laissent que peu d'espoir d'obtenir des circuits 100% fonctionnels en utilisant ces composants sans repenser fondamentalement les architectures. Dans ce contexte, les réseaux de neurones connaissent un regain d'intérêt. En effet, ils sont considérés naturellement robustes, notamment grâce à la nature distribuée de l'information, et susceptible de corriger une grande quantité de défauts grâce à l'apprentissage. Notre approche consiste à proposer une architecture et à confronter l'idée d'une telle architecture neuronale aux contraintes d'une conception détaillée. Pour autant, de nombreux problèmes restent à résoudre pour passer de cette idée à sa réalisation. D'abord, nous ignorons encore quels nanocomposants vont finalement se révéler être les mieux adaptés à remplacer ou compléter la technologie CMOS. De plus, les technologies ne sont évidemment pas encore disponibles, la physique sous-jacente n'est pas toujours bien comprise et les modèles de composants pour la conception sont en cours de développement. Toutes les méthodes de conception doivent donc être revisitées et tous ces problèmes ne seront pas résolus facilement. Dans ce chapitre, on se focalise sur l'utilisation de matrice de connexions synaptiques réalisables à partir d'éléments dont la conductance peut être ajustée de façon à implémenter une forme de configuration de fonction logique inspirée des algorithmes d'apprentissage des réseaux de neurones. Lorsque des éléments actifs sont nécessaires, nous pourrions considérer l'utilisation de transistors CMOS (approche hybride) ou des transistors à nanotube de carbone. Cette dernière piste a notamment fait l'objet de la thèse de Michel Hé. La mémorisation de la configuration y est effectuée par un dipôle résistif présentant un effet de mémorisation. Nous avons choisi d'étudier plus particulièrement des dispositifs à électrolyte solide (CBRAM). L'étude que nous présentons ici décrit la conception d'une architecture et des cellules à base de nanotubes de carbone permettant d'envisager un apprentissage.

2. Réseaux de neurones

Un réseau de neurone formel est constitué de cellules identiques interconnectées. Son fonctionnement, comme les méthodes d'apprentissage permettant de définir le rôle

de chaque neurone, s'inspire de la biologie.

Nous allons donc commencer par rappeler brièvement les caractéristiques des neurones biologiques avant de n'en retenir qu'un formalisme très simplifié.

2.1. Inspiration biologique

Le neurone est d'abord une cellule vivante présente dans le système nerveux central des animaux. Cette cellule comporte un noyau dans sa partie principale (corps cellulaire), entouré d'une membrane et renfermant le noyau. Le corps cellulaire mesure quelques dizaines de micromètres et fait apparaître des ramifications appelées des dendrites. De plus, le neurone possède un long tubule appelé **axone** pouvant atteindre 1m. L'axone se termine généralement par des ramifications. A la différence de la plupart des cellules somatiques, les neurones cessent très rapidement de se multiplier, mais le système nerveux central n'a pas pour autant terminé son développement. En contrepartie, les connexions entre les cellules se construisent tout au long de la vie. La communication entre neurones se fait par l'émission de neurotransmetteurs libérés par des éléments appelés synapses, souvent «connectés» à l'extrémité des ramifications dendritiques et détectés par des neuro-récepteurs au niveau du corps cellulaire ou des dendrites. L'ensemble des synapses contribue à modifier le potentiel électrique du milieu dans lequel baigne le corps cellulaire que nous appellerons le potentiel post-synaptique. Lorsque le potentiel post-synaptique atteint un seuil (typiquement -30mV) le neurone est le siège d'une suite de réactions menant au déclenchement d'une onde de dépolarisation de la paroi de l'axone appelé potentiel d'action. Cette activité est mesurable électriquement sur l'axone et permet d'agir sur les autres neurones par l'intermédiaire des synapses situées à l'extrémité de l'axone. Certaines synapses vont avoir pour effet d'inciter au déclenchement d'un potentiel d'action dans le neurone destination alors que d'autres vont avoir tendance au contraire à inhiber le déclenchement de potentiel d'action. Par ailleurs, **les synapses ont des efficacités différentes, elles peuvent varier environ d'un facteur 10 d'une synapse à l'autre.** L'un des éléments clés de l'apprentissage réside dans la manière dont les synapses évoluent, notamment leur efficacité, en fonction de l'activité des neurones environnants. La logique comme l'expérience conduit à considérer que l'efficacité d'une synapse doit pouvoir être modifiée par l'histoire des potentiels d'action des deux neurones concernés. Hebb avait postulé en 1949 que le déclenchement d'un potentiel d'action renforce l'efficacité des synapses qui y ont contribué. Ce postulat revient à supposer que le chemin suivi par l'influx nerveux à travers le système nerveux central laisse une trace, support d'une forme de mémorisation. On retrouve ce principe dans la plupart des algorithmes d'apprentissage de réseaux de neurone qui prennent en compte **les cooccurrences d'activité neuronale pour définir l'évolution des efficacités synaptiques.** Il apparaît également que le déphasage des potentiels d'action peut être également support d'information tout comme il peut jouer un rôle dans l'apprentissage [45].

2.2. Neurone formel

La définition de modèles de réseaux de neurones peut viser un biomimétisme aussi fidèle que possible de façon à comparer et valider des modèles biologiquement vraisemblables. Cependant, à cette recherche de biomimétisme, on peut opposer une autre approche que nous pouvons qualifier de bio-inspirée, dans laquelle la fidélité avec la biologie importe moins que les nouvelles applications qui peuvent être envisagées. Pour différencier les modèles visant la fidélité du biomimétisme de ceux qui ne cherchent qu'à y puiser une source d'inspiration, nous appellerons les seconds des modèles de réseaux de neurones «formels» (RNF).

Le fonctionnement de la cellule est modélisé de façon très simpliste dans la plupart des modèles de réseaux de neurones formels. La fidélité avec le fonctionnement des

neurones vivants n'est généralement pas l'objectif principal, **un RNF est d'abord un algorithme de traitement de signal non-linéaire permettant de réaliser une fonction sans l'avoir explicitement décrite** par une procédure algorithmique. On y retrouve cependant beaucoup de vocabulaire tiré des neurones biologiques. Un réseau de neurone formel est constitué d'éléments appelés neurones n_i dont l'activité x_i peut être binaire (+1/-1), correspondant à la présence ou à l'absence d'un potentiel d'action, ou continue, correspondant à la fréquence d'un potentiel d'action, mais qu'elle soit continue ou binaire, l'activité reste bornée. Les neurones sont connectés les uns aux autres, formant un graphe dont les neurones sont les sommets et les synapses les arcs. Chaque synapse reliant le neurone n_i au neurone n_j porte un poids synaptique w_{ij} représentant son efficacité. Ce poids est signé, positif pour une synapse excitatrice et négatif pour une synapse inhibitrice. Le traitement effectué par chaque neurone est décomposé en deux phases. La première est une opération purement linéaire. Elle consiste à calculer l'influence du voisinage d'un neurone en faisant la somme pondérée des états d'activité des neurones qui lui sont connectés. Les pondérations correspondent aux poids (ou efficacités) synaptiques et le résultat est l'image du potentiel post-synaptique.

$$V_j = \sum_i w_{ij} \times x_i \quad (1)$$

La deuxième opération, appelée fonction d'activation, est un opérateur non linéaire qui réalise une décision. La nature précise de cette fonction dépend du modèle choisi et du codage des états d'activité. Par exemple, dans les cas les plus simples, avec des états d'activité binaires (-1,+1), il peut s'agir d'un seuillage par une fonction signe :

$$x_j = \text{signe}(V_j) \quad (2)$$

Lorsque les états d'activité des neurones ne sont plus binaires, la fonction d'activation est souvent une sigmoïde

$$x_j = \frac{1}{1+\exp(-V_j)} \quad (3)$$

Ce fonctionnement basique peut se décliner en de très nombreuses variantes : en prenant en compte une dynamique stochastique, des états continus ou en faisant intervenir des éléments de délais par exemple. La manière d'interconnecter les neurones (tous interconnectés dans un réseau récurrent ou organisés en couches successives, l'organisation des connexions entre les couches) conduit à des architectures différentes correspondant à des classes d'application différentes (classement, identification, mémoire associative, traitement d'image) portant sur des structures de données différentes (1D, 2D). La combinaison de ces différentes possibilités mènent à définir des modèles de réseaux de neurones formels caractérisés par la dynamique des neurones, les architectures de réseaux et une procédure d'apprentissage adaptée.

L'apprentissage consiste à donner une fonction au réseau à partir d'un sous-ensemble représentatif de stimuli et de réponses attendues. On appelle la base d'apprentissage cet ensemble et patterns les couples entrée/sortie qui la constituent. L'apprentissage permet d'ajuster les poids synaptiques de façon à obtenir la réponse attendue pour les entrées de la base d'exemple. Nous préciserons plus loin quelques exemples d'algorithmes d'apprentissage. Si son architecture est bien choisie, le réseau peut présenter une faculté de généralisation : lorsqu'on lui présente des entrées qui ne font pas partie de la base d'apprentissage, il répond correctement. Cette faculté de généralisation, associé au principe de programmation à partir d'exemple constitue la caractéristique la plus recherchée dans le réseaux de neurones quand on les considère uniquement comme une méthode du traitement de signal non-linéaire, indépendamment du support matériel de l'exécution, ce qui est le cas le plus fréquent. Pourtant, les réseaux de neurones présentent également des qualités de robustesse qui en font un modèle potentiellement intéressant dans le cadre d'une implémentation matérielle dédiée. Cette robustesse provient d'une part de la nature distribuée de l'information, répartie

dans tous les poids synaptiques du réseau, et d'autre part de la faculté de compenser des défauts grâce à l'apprentissage. Elle paraît particulièrement bien adaptée lorsque les caractéristiques électriques des composants sont dispersées et qu'ils présentent un fort taux de défaut, ce qui semble bien être le cas aujourd'hui pour la plupart des nanocomposants. Dans ce contexte, les approches neuronales connaissent un regain d'intérêt. Ayant une expérience dans le domaine [T-R1,T-A1,T-N4], nous nous sommes intéressés à cette piste. Cependant, à la différence des équipes qui se lancent dans ce domaine pour la première fois, nous savons qu'il devient rapidement difficile d'effectuer un apprentissage dès lors qu'on considère des fonctions complexes. **Nous nous sommes donc concentrés sur une approche de granularité très fine** visant à effectuer des fonctions logiques dont la complexité est comparable à celle des fonctions tabulées (Look Up Table) dans les FPGA. A l'opposé, on trouvera des recherches visant à réaliser des neuro-processeurs complets et généralistes, prenant en charge globalement toutes les étapes d'un traitement [8]. Bien que des exemples biologiques, à commencer par nous-mêmes, nous montrent que de telles solutions doivent nécessairement exister, l'expertise que nous avons pu acquérir dans le domaine de l'apprentissage pour différents modèles de réseaux de neurones me conduisent à douter très fortement que cet objectif puisse être accessible à moyen terme. En comparaison, notre approche peut paraître très modeste, mais elle entre directement en concurrence avec les fonctions équivalentes qui seront réalisées en CMOS ultime, ce qui constitue un enjeu majeur pour l'industrie microélectronique.

2.3. Neurone formel et architecture nanoélectronique

D'abord, chaque neurone doit diffuser son état d'activité x_i à ses voisins puis il est nécessaire de réaliser la somme des contributions ($w_{ij} \times x_i$) de chacun de façon à évaluer le potentiel post-synaptique V_j qui sera seuillé pour déterminer l'état d'activité du neurone x_j . L'architecture matérielle qui permet d'implémenter efficacement un réseau de neurone découle de ces deux nécessités. La diffusion d'un état peut se faire sans dégradation sous forme d'une tension, que le codage soit analogique, binaire, ou impulsionnel, mais la somme des influences de chaque neurone qui permet de calculer le potentiel post-synaptique s'effectue plus efficacement par un codage en courant. Les synapses doivent alors être caractérisées par des conductances à l'intersection entre les conducteurs porteurs des états d'activité des neurones d'entrée et les conducteurs représentant les potentiels post-synaptiques. Le seuillage peut opérer sur le signe du courant, bien sûr, mais également sur le potentiel électrique du conducteur post-synaptique. En effet, ce potentiel représente aussi, à une normalisation près, le potentiel post-synaptique. La normalisation n'affecte ni l'influence relative de chacune des entrées ni le signe du résultat, ce qui conduit au même état. Le comparateur peut alors être réalisé de façon compacte et peu gourmande en énergie, par exemple à partir d'un simple inverseur (CMOS ou CNTFET).

Bien évidemment, **les conductances présentes à l'intersection des conducteurs d'entrée et des conducteurs post-synaptiques resteront positives alors que les poids synaptiques doivent impérativement être signés** de façon à autoriser des synapses inhibitrices ou excitatrices. Un codage différentiel des entrées permet de contourner cette difficulté. Dans une telle architecture, chaque entrée est injectée dans la matrice synaptique sous forme complémentaire : $(x_{i+}, x_{i-}) = (+1, -1)$ ou $(-1, +1)$. Chaque poids synaptique w_{ij} est alors codé par le couple de conductances (g_{ij+}, g_{ij-}) reliant les deux conducteurs d'entrée (x_{i+}, x_{i-}) au conducteur post-synaptique. Le poids synaptique résultant est donné par la différence des deux conductances :

$$w_{ij} = g_{ij+} - g_{ij-}. \quad (4)$$

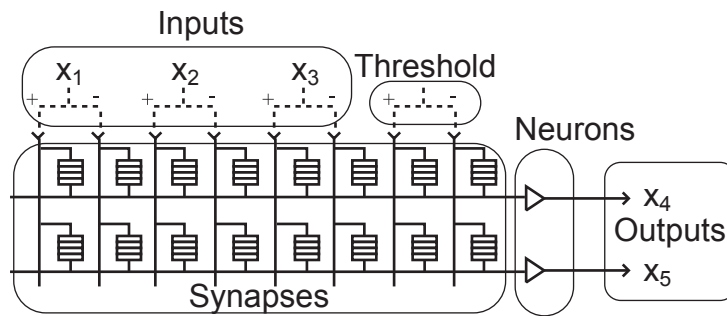


Figure 37 : Architecture d'un réseau à 3 entrées complémentaires et deux sorties.

Conséquence bénéfique de ce codage différentiel, l'utilisation de composants présentant une dynamique de conductance limitée peut mener à une importante dynamique des poids synaptiques.

2.4. Algorithmes d'apprentissage

Pour qu'un réseau de neurone puisse tirer parti de la phase d'apprentissage afin de compenser ses défauts et ses dispersions de caractéristiques, il est nécessaire que le circuit participe physiquement à la phase d'apprentissage. Il est possible de réaliser l'apprentissage totalement sur le circuit, on parle alors d'un apprentissage «on-chip» ou bien d'intégrer le circuit dans la boucle de l'apprentissage tout en conservant une approche logicielle extérieure au circuit pour le calcul des variations des poids synaptiques, on parle dans ce dernier cas d'un apprentissage «chip-in-the loop». L'apprentissage on-chip est bien plus intéressant dans le contexte de nanocomposants puisqu'il réduit le temps de configuration des circuits par la parallélisation des traitements liés à l'apprentissage. A contrario, il impose des contraintes de surface ajoutée sur le circuit qui doivent restée modeste au regard du gain en termes de densité d'intégration apporté par l'utilisation de nanocomposants. Cette contrainte se répercute sur le choix des algorithmes d'apprentissage qui peuvent être mis en oeuvre matériellement. Nous allons donc visiter quelques-uns des algorithmes d'apprentissage les plus classiques de façon à identifier ceux qui paraissent les plus adaptés.

2.5. Règle de HEBB

Le premier algorithme d'apprentissage connu consiste à optimiser la corrélation entre les sorties attendues y_j et les sorties obtenus avant seuillage (potentiel post-synaptique) V_j dans un un réseau ne possédant qu'une seule couche de neurones actifs. Cet algorithme conduit à modifier, pour chaque pattern de la base d'apprentissage, les poids synaptiques dans le sens du produit $x_i \times y_j$ où x_i représente l'état d'activité de l'entrée pour le pattern considéré.

$$\Delta w_{ij} = x_i \times y_j \quad (5).$$

Cet algorithme peut être utilisé pour des réseaux ne possédant qu'une seule couche de neurone actif, que nous appelons des classifieurs linéaires ou bien pour des réseaux récurrents entièrement connectés appelés réseaux de Hopfield [18]. Le principal avantage de cet algorithme est sa grande simplicité tant conceptuelle que de mise en oeuvre. Cependant, il présente pour nous l'inconvénient majeur de ne pas prendre en compte la réponse du réseau (x_j) pour modifier les poids synaptiques. Ne considérant que les éléments de la base d'exemple (entrées x_i et sorties attendues y_j) il ne permet ni de corriger les défauts du circuit ni d'importantes dispersions de caractéristiques. Nous pouvons le qualifier de ce point de vue de méthode de configuration «en boucle ouverte». Reste que le

principe de multiplication entre les sorties attendues et les entrées se retrouve dans la plupart des algorithmes d'apprentissage, comme nous allons le voir.

2.6. Règle DELTA (ADALINE)

L'algorithme d'apprentissage suivant est une variation de la règle de Hebb et consiste à minimiser l'erreur quadratique entre la sortie obtenue et la sortie désirée dans le cas d'un codage d'état de neurone continu. Elle conduit par application d'une méthode de gradient à modifier les poids synaptiques w_{ij} à chaque pattern suivant le produit entre l'entrée x_i et l'erreur sur la sortie $y_j - x_j$.

$$\Delta w_{ij} = \alpha \times x_i \times (y_j - x_j) \quad (6).$$

Le paramètre α représente le pas de l'apprentissage. Dans le cas où les états d'activité de neurones sont binaires (+1 / -1), la règle de DELTA correspond à appliquer la règle de HEBB pour les synapses associées à une sortie en erreur et à ne rien faire pour les autres. Cette fois il s'agit bien d'un algorithme en boucle fermée dans le sens où il exploite l'erreur sur les sorties. Il est donc susceptible d'opérer une compensation de dispersion ou une correction de certains défauts.

2.7. BACKPROP (Perceptron)

L'algorithme de la rétropropagation du gradient (BACKPROP) est une extension de la règle DELTA pour des réseaux multicouches. Dans ce type de réseaux, il existe des éléments appelés neurones cachés qui ne sont ni des entrées, ni des sorties. La difficulté pour l'apprentissage d'un tel réseau consiste à évaluer l'erreur commise par les neurones cachés. Par définition, nous ne connaissons pas l'état attendu pour ces éléments, ce qui exclu de le comparer à l'état obtenu. L'erreur est donc calculée à partir de l'erreur effectuée par les sorties en attribuant aux neurones cachés la responsabilité d'une erreur d'autant plus fortement que celui-ci est lié par une synapse des poids important à la sortie erronée. L'évaluation de l'erreur se propage donc des sorties vers les entrées, en sens inverse du sens de propagation normal dans un tel réseau en couches, d'où son nom.

$$\Delta w_{ij} = \alpha \times x_i \times (y_j - x_j) f'(V_j) \quad (7).$$

Cette règle s'applique à des réseaux en couches, dont les états de neurones sont continus. La fonction d'activation $f(x)$ qui ne peut plus être un simple seuil est alors le plus souvent une sigmoïde (équation 3). Elle présente notamment l'avantage d'être particulièrement simple à dériver.

2.8. Machine de Boltzmann

La machine de Boltzmann est un modèle de réseau récurrent caractérisé par la dynamique stochastique de ses neurones binaires. La fonction d'activation est remplacée par un tirage au sort suivant un probabilité

$$P(x_j = 1 | V_j) = \frac{1}{1 + \exp(\frac{-V_j}{T})} \quad (8).$$

où T est un paramètre appelé température par analogie avec un recuit simulé. La décroissance de la température tend à rendre le réseau déterministe à la fin de la relaxation. Son architecture est organisée en couches successives, comprenant des neurones d'entrée, cachés et de sortie, mais les connexions sont symétriques ($w_{ij} = w_{ji}$), ce qui implique que le processus qui mène à la détermination de la réponse du

réseau est d'avantage une relaxation qu'une simple propagation des entrées vers les sorties. Dans sa version initiale, [19] le tirage au sort ne doit concerner qu'un seul neurone à la fois, mais R. Azencott a proposé une version synchrone [20] dans laquelle les états de tous les neurones sont tirés au sort simultanément. Cette version a servi de base aux travaux du groupe dirigé par Patrick Garda à l'IEF dans le cadre du projet «Réseau Analogique» auxquels j'ai contribué durant ma thèse.

L'apprentissage est basé sur la comparaison de la mesure de co-activité

$$P_{ij} = \frac{1}{N} \sum_{n=1}^N x_i \times x_j \quad (9).$$

des neurones n_i et n_j situés aux extrémités d'une synapse w_{ij} , pour un réseau dont les sorties sont imposées (phase forcée) noté P_{ij}^+ avec un réseau dont les sorties évoluent librement (phase libre) noté P_{ij}^- . La modification des poids synaptiques est proportionnelle à la différence de cette mesure

$$\Delta w_{ij} = \alpha(P_{ij}^+ - P_{ij}^-) \quad (10).$$

Dans le cas d'un réseau à dynamique synchrone, la mesure des co-activités simultanées doit être remplacée par la mesure des co-activités successives de façon à prendre en compte la causalité :

$$P_{ij} = \frac{1}{N} \sum_{t=0}^N x_i^t \times x_j^{t+1} + \frac{1}{N} \sum_{t=0}^N x_i^{t+1} \times x_j^t \quad (11).$$

Ce deuxième algorithme paraît plus difficile à mettre en oeuvre de façon efficace en utilisant des nanocomposants, mais l'expérience montre qu'il est très souvent possible en pratique d'utiliser le premier algorithme d'apprentissage avec un réseau synchrone.

2.9. Discussion

La règle de Hebb est séduisante par sa simplicité puisqu'il suffit d'incrémenter ou décrémente les conductances des synapses suivant que les états d'activité des entrées et des sorties attendues sont de même signes ou de signes opposés. Nous y reviendrons dans la section 4 au sujet des perspectives d'apprentissage par auto-configuration, mais il présente un défaut majeur. Comme il ne prend pas en compte l'état des neurones obtenus en sortie, il fonctionne en boucle ouverte et ne peut compenser ni des défauts ni des dispersions de caractéristiques. Il s'agit donc d'avantage d'un algorithme de programmation que d'apprentissage.

La règle Delta pour un classifieur linéaire corrige ce défaut tout en restant très simple à mettre en oeuvre. Il suffit de n'appliquer la règle de Hebb qu'aux synapses associées aux sorties en erreur. C'est donc cet algorithme que nous avons d'abord choisi d'implanter.

La rétropropagation paraît bien plus complexe à mettre en oeuvre. De plus, elle s'applique à des neurones dont les états sont continus, ce qui ne correspond pas directement à notre objectif de réalisation de fonctions logiques. En outre, le nombre d'itérations permettant d'obtenir la convergence d'un apprentissage augmente de plusieurs ordres de grandeur lorsqu'on passe d'un réseau monocouche à un Perceptron Multicouche. Cette augmentation induit des contraintes sur la dynamique des poids synaptiques et sur l'endurance des éléments non-volatiles utilisés pour les mémoriser. Ces contraintes vont diminuer les chances de découvrir un jour un composant présentant toutes les qualités requises pour jouer le rôle de synapse. Dans ce contexte,

le fait que la rétropropagation du gradient donne accès à toutes les fonctions logiques, y compris celles qui ne sont pas séparables linéairement, ne nous semble pas suffisant pour justifier la complexité ajoutée sachant que toutes les fonctions logiques non-linéairement séparables peuvent être facilement décomposées en fonctions linéairement séparables. Au vue de la complexité de son implantation, nous avons donc provisoirement écarté cet algorithme, mais le point évoqué juste après nous semble suffisamment important pour chercher à dépasser les difficultés d'implantation et quelques pistes permettant d'implanter des réseaux multicouches sans recourir à la rétropropagation on-chip sont aujourd'hui à l'étude.

En effet, la présence de neurones cachés permet également d'assurer une forme de redondance qui confère au réseau une bonne robustesse face à certains défauts. Au début de la thèse de Michel HE, par des simulations numériques fonctionnelles, nous avons étudié les possibilités de correction des défauts par apprentissage [A19]. Ces simulations montrent que, pour un problème donné, il existe une taille minimale de réseau en dessous de laquelle il ne faut pas descendre pour que l'apprentissage puisse converger. Lorsqu'on injecte des défauts dans ce réseau, l'apprentissage converge tant que le nombre de neurones fonctionnels reste suffisant pour reconstituer un tel réseau minimal. Dans ce cas, l'apprentissage parvient toujours à corriger les défauts. C'est évidemment un point clé dans la mesure où la redondance et la robustesse constituent l'une des principales motivations des approches neuronales dans les architectures à base de nanocomposants.

De ce point de vue, la machine de Boltzmann semble de prime abord présenter l'intérêt de combiner une approche en boucle fermée (grâce à la phase «libre») avec un algorithme d'apprentissage locale, ne prenant en compte que les états des deux neurones présents aux extrémités de chaque connexion synaptique. A priori la localité de l'apprentissage garantit à une réalisation matérielle plus simple que la rétropropagation du gradient. Malheureusement, l'intégration on-chip de cet algorithme d'apprentissage nécessite d'implémenter autant de compteurs de cooccurrence que de synapses. Dans ce contexte, il ne paraît pas possible de justifier l'usage d'architectures neuronales par l'augmentation de la densité d'intégration rendue accessible grâce au recours à des nano-composants si dans le même temps un compteur doit être associé à chaque synapse nanoscopique. La seule solution acceptable consiste à considérer que la synapse doit elle-même réaliser l'opération de comptage. On est ramené à réaliser localement une incrémentation ou une décrémentation immédiate de la conductance des synapses en fonction des états des deux neurones présents à ses extrémités, cette opération devant agir de façon contraire suivant qu'on se trouve dans la phase libre ou dans la phase forcée. L'absence de compteur auxiliaire implique d'effectuer immédiatement les modifications de conductance, ce qui risque de poser des problèmes quand à la convergence de l'apprentissage. Nous sommes alors conduit à alterner aussi fréquemment que possible les phases libres et forcées de façon à éviter que l'effet l'une des phases ne modifie trop le comportement du réseau. A l'extrême, si le réseau est déterministe et ne contient qu'une seule couche de sortie, on peut alterner entre phases libres et forcées à chaque pattern. Dans ce cas, les états des neurones de sorties (x_i) (eq. 9) en phases libre et forcée correspondent exactement aux états des neurones de sortie obtenus (x_i) et attendus (y_i) de l'équation 6. La comparaison des cooccurrences (eq. 10) se ramène alors simplement à la règle DELTA (eq. 6). Ainsi, sans nier l'intérêt que pourrait présenter l'approche stochastique de la machine de Boltzmann, il apparaît que **la règle DELTA constitue un algorithme de référence que nous retrouvons dans pratiquement tous les autres algorithmes d'apprentissage**. Il m'a donc semblé important, dans un premier temps, de vérifier que des nanocomposants étaient susceptibles d'être utilisés efficacement pour implanter ce type d'apprentissage.

3. Conception et apprentissage d'un réseau utilisant des nanocomposants

3.1. Architecture de base en reconnaissance

Qu'elle soit ou non basée sur des nanocomposants, on retrouve dans la conception d'une architecture neuronale les deux étapes de détermination de l'état d'un neurone, d'abord de combinaison linéaire, puis de décision non linéaire. Ces deux étapes conduisent à choisir des éléments correspondants :

- linéaires (et potentiellement passifs) pour la première étape, dont la conductance est le paramètre principal qui doit pouvoir être ajusté,
- de seuillage non-linéaire (et nécessairement actifs) pour la deuxième étape correspondant à une restauration de niveau logique.

Il existe aujourd'hui une gamme relativement importante de composants dont la conductance peut être modifiée et présentant un effet mémoire non-volatile de cet état de conductance [21-40]. Si de plus, plusieurs niveaux de conductances sont accessibles, ces composants constituent une opportunité pour la réalisation nano-électronique de réseaux de neurones de très haute densité. C'est clairement le cas pour certains d'entre eux [21,22,27,30,33,37,39] et ce n'est pas exclu pour les autres.

Concernant la réalisation des neurones, dans la mesure où la surface d'un réseau de neurones est habituellement dominée par les connexions synaptiques, il est possible de recourir à des architectures mixtes dans lesquelles les connexions sont réalisées par des nano-composants alors que les neurones sont conçus en utilisant une technologie CMOS plus classique. C'est notamment la voie suivie par Likharev [8-10] qui se trouve être le porteur de l'une des propositions architecturales basée sur des nanocomposants les plus connues aujourd'hui. Sans nier l'intérêt d'une telle piste, je souhaite également envisager l'intégration des éléments de décision non-linéaire en nanocomposants. D'une part, il se trouve que les mêmes nano-composants présentent des caractéristiques qui les font aussi bon candidats à la l'intégration des synapses que des neurones [41]. D'autre part, l'intérêt du projet NANOSYS (ACI NANOSCIENCE du CNRS) s'est rapidement focalisé sur les transistors à nanotubes de carbones, leur caractérisation, leur modélisation et l'exploitation de leurs caractéristiques dans des architectures. Il m'est donc apparu opportun dans le cadre de la thèse de Michel HE d'envisager d'utiliser et de valider les nouveaux modèles disponibles dans nos circuits. Ceci permet notamment d'évaluer l'impact qu'ils pourraient avoir sur les techniques de conception à base de transistor, qu'elles soient analogiques ou numériques. Partant de ce principe, Michel HE a choisi de s'intéresser plus particulièrement aux composants de type CBRAM pour réaliser les synapses, choix basé principalement la dynamique importante des valeurs de conductance accessibles et sur la valeur limitée de son courant de programmation qui le rend compatible avec l'utilisation de CNTFET. L'architecture qui en découle est illustrée sur la figure 38.

3.2. Technique d'apprentissage

Comme nous l'avons évoqué, différents modèles de réseaux de neurones existent et chacun utilise une procédure d'apprentissage qui lui est propre. Pourtant, lorsqu'on examine attentivement ces différents algorithmes, on peut y retrouver un principe commun. Dans la plupart des cas, lorsqu'on considère des architectures simples, on retrouve un algorithmes proche de la règle Delta. C'est bien sûr le cas pour le Perceptron lorsqu'il n'a qu'une seule couche, mais c'est également le cas pour la machine de Boltzmann en l'absence de neurones cachés lorsque la température tend vers

zéro. A ce stade, le problème de l'apprentissage revient alors à trouver une procédure permettant de faire varier les poids synaptiques afin de réaliser la règle Delta. **La complexité de cette procédure et du matériel associé doit rester modeste de façon à ne pas faire perdre tout l'intérêt que l'utilisation de nanocomposants pourrait apporter en termes de densité d'intégration.** Dans la thèse de Michel HE, l'architecture du réseau utilisé en reconnaissance a été modifiée pour introduire des éléments de sélection au niveau de chaque synapse et des éléments de programmation associés à chaque neurone : l'unité d'apprentissage. Cet élément est composé de deux parties : une partie numérique, conçue à partir de schéma de porte logique tout à fait classique à ceci près que le modèle de transistor utilisé dans les simulations est un CNTFET [42-43], et d'une partie analogique conçue pour contrôler l'augmentation ou la diminution de la conductance du composant CBRAM utilisé en synapse et modélisé pour l'occasion. La partie numérique de la cellule d'apprentissage génère deux signaux d'incrémentement et de décrémentation de la conductance en suivant une version binaire de la règle de Hebb, à partir des valeurs de sortie obtenues x_j et souhaitées y_j , **mais également à partir de l'état d'une entrée sélectionnée (x_i).** Le fait que l'entrée soit utilisé dans le calcul du sens de variation de la conductance **impose d'appliquer une programmation séquentielle, colonne par colonne.** Cette démarche conduit à insérer dans la matrice des éléments de sélection permettant de déconnecter de son entrée la colonne considérée pour la connecter à sa cellule d'apprentissage (cf. figure 38). Cette approche **limite la densité d'intégration et constitue une première étape**, mais nous verrons plus loin qu'une autre solution est envisageable où les éléments de sélection sont éliminés.

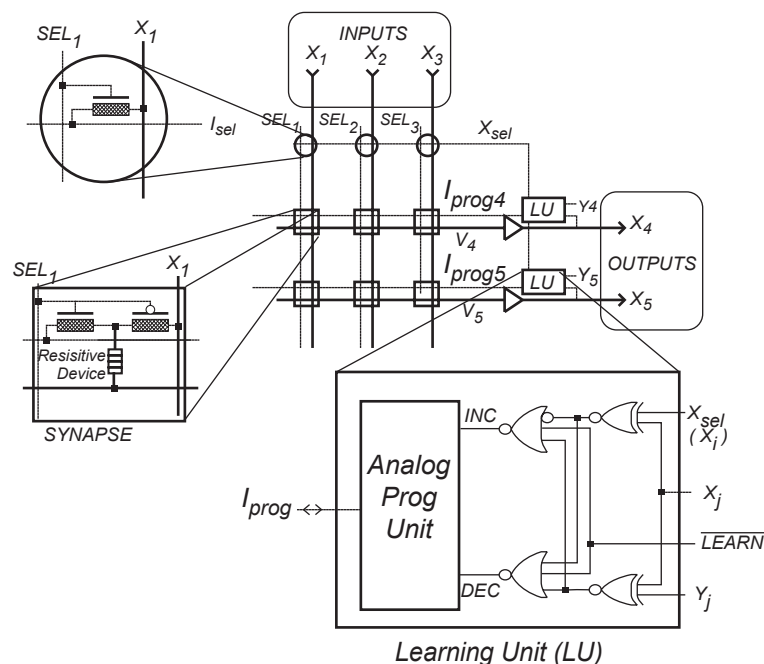


Figure 38 : Architecture d'un réseau de neurone comportant un apprentissage on-chip utilisé dans la thèse de Michel He. Comme le sens dans lequel le courant de programmation, définissant le sens de variation de la conductance, dépend de l'état de l'entrée (suivant la règle de Hebb ou Delta), il est nécessaire d'isoler une colonne de synapses, de la déconnecter de son entrée pour la connecter aux unités d'apprentissage. En appliquant une version binaire de la règle Delta, il est ainsi possible de réaliser un apprentissage, séquentiellement, colonne par colonne.

Pour la partie analogique, un injecteur de charge similaire à celui utilisé dans la thèse de Yimin Zhu [44] a été conçu et simulé (cf. figure 39). Il utilise un modèle de CNT-FET[43].

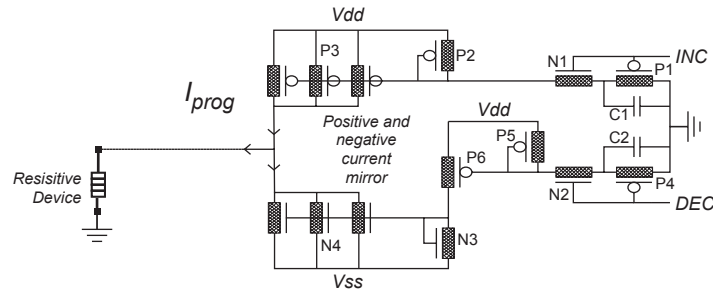


Figure 39 : Partie analogique de l'injecteur de charge inspiré de [44] et utilisé pour modifier les conductances des CBRAM.

Le fonctionnement en charge permet un contrôle automatique de la durée des impulsions de programmation, particulièrement important pour assurer que la modification de la conductance sera limitée et n'atteindra pas immédiatement une de ses valeurs extrêmes.

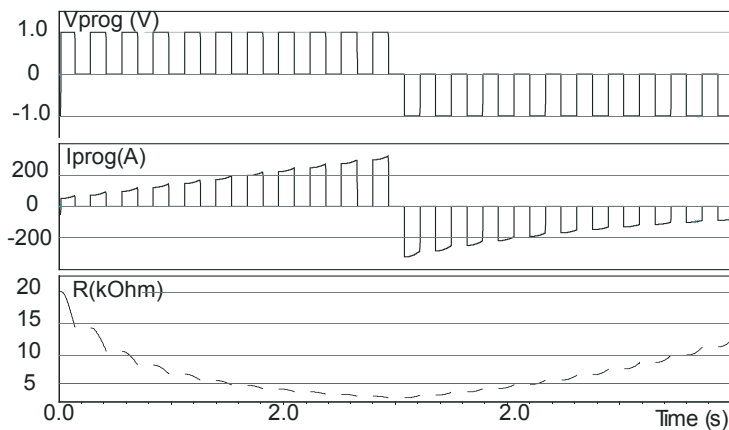


Figure 40 : Evolution de la valeur de la résistance, (dans un sens puis dans l'autre) utilisant l'injecteur de charge de la figure 39. Le modèle de conductance ajustable considère une variation de la conductance proportionnelle à l'amplitude des impulsions et à la durée (et au nombre) des impulsions. Ce modèle est, par exemple, cohérent avec les résultats de [22].

A contrario, pour un composant dont la variation de conductance est supposée varier sous le contrôle d'une tension au-delà d'un seuil de programmation, le contrôle direct de la tension de programmation permet d'envisager beaucoup plus facilement de faire converger un apprentissage.

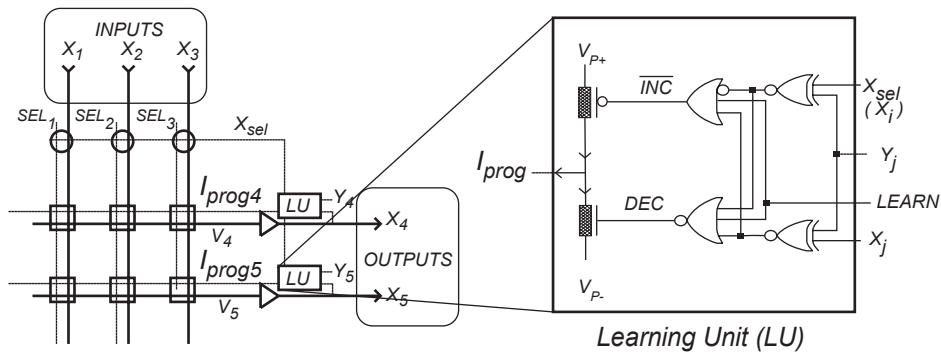


Figure 41 : Architecture de réseau utilisé pour la simulation électrique d'un apprentissage de fonction logique (cf. figure 42). Pour alléger la figure, les seuils et les entrées complémentaires ne sont pas représentés.

Dans ces conditions, des simulations électriques montrent qu'un tel apprentissage est envisageable et qu'il peut converger très rapidement (cf. figure 42) [R8,A23].

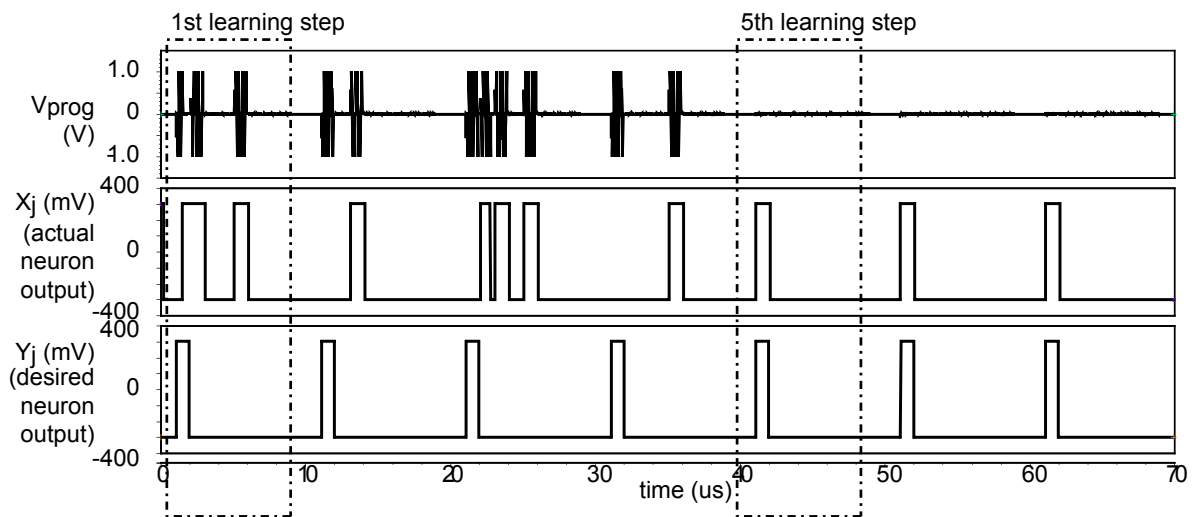


Figure 42 : Résultats de simulation électrique d'un apprentissage neuronal appliqué au réseau de neurone de la figure 41. La première ligne correspond aux impulsions de programmation. La trace correspond à l'état de la sortie du neurone considéré. La troisième ligne est l'état attendu. Il correspond ici à la fonction logique NOR. Les entrées ne sont pas représentées. Elles balaient pour chaque itération toutes les configurations binaires dans l'ordre du code binaire naturel. A la cinquième itération, l'état obtenu et attendu sont identiques et les impulsions de programmation ont cessé.

4. Vers des architectures auto-configurable

Dans l'étude précédente, Michel HE a introduit des **éléments de sélection à l'intérieur de la matrice de connexions synaptiques**. Cette solution permet d'isoler une colonne pour associer un cellule d'apprentissage à chaque synapse de façon à y opérer la modification de conductance dans le bon sens, en suivant la règle de Hebb ou la règle Delta. Cependant, une telle solution **n'est pas compatible avec une grande densité d'intégration**. A l'inverse, **il serait souhaitable de définir une méthode** de programmation permettant de réaliser l'apprentissage suivant la règle de Hebb en **ne jouant que sur les conducteurs accessibles** à partir du pourtour de la matrice de

connexion synaptique (cf. figure 43). Dans le cadre du projet ANR-PANINI , j'ai commencé à explorer les possibilités d'une telle programmation. J'ai considéré différents cas, correspondant à différents comportements de composants utilisables en tant que synapses. Certains comportements correspondent à des composants identifiés [21-40], d'autres pas encore. Tous ces modèles fonctionnels de composants ont en commun un contrôle de la variation de la conductance par la tension et présentent une plage de tension où la variation est nulle garantissant la non-volatilité. Ils diffèrent par leur sensibilité à la polarité des tensions appliquées où seulement à leur amplitude, par leur caractère monopolaire ou dipolaire, sont programmable de façon réversible ou irréversible. **Dans tous les cas, j'ai défini une méthode d'apprentissage adaptée** permettant de converger rapidement vers une configuration des poids synaptiques correspondant à la fonction logique linéairement séparable apprise [X1]. Parmi ces méthodes, certaines ont déjà fait ou font l'objet d'un dépôt de brevet [B1]. A ce jour, le processus d'extension du brevet étant en cours et il ne m'est pas encore possible de détailler toutes les méthodes envisagées. Je ne peux donc en faire qu'une description partielle. Toutes ces méthodes ont en commun d'appliquer un potentiel sur les conducteurs post-synaptiques qui ne dépend que des états obtenus (x_j) et attendus (y_j) des différents neurones.

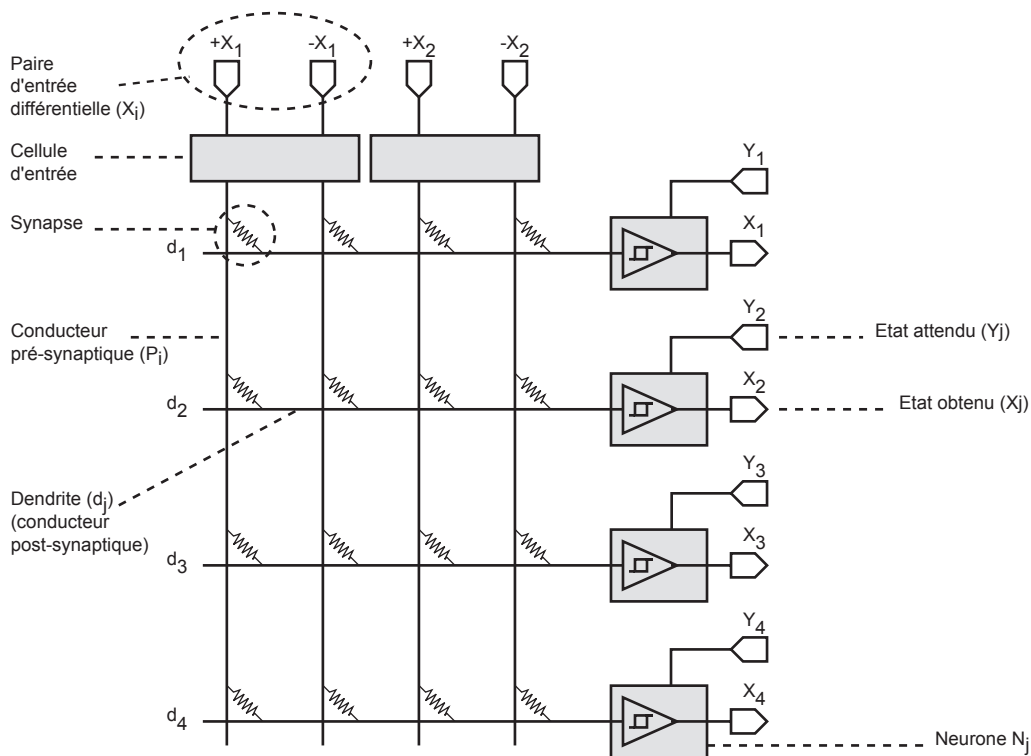


Figure 43 : Architecture d'un réseau avec apprentissage, sans éléments de sélection des synapses. La cellule neurone réalise l'apprentissage en imposant un potentiel sur les conducteurs post-synaptiques (dendrites). Les synapses verront leur conductance modifiée en fonction de la différence entre ce potentiel et celui imposé par les entrées.

La difficulté d'une telle entreprise peut être illustrée de la manière suivante. Nous devons considérer des nanocomposants dont la conductance est non-volatile mais peut être modifiée. Nous pouvons considérer que le contrôle de la conductance s'opère à partir de la tension aux bornes d'un tel dipôle. La non-volatilité implique qu'il existe une plage de tension neutre pour laquelle les conductances ne sont pas affectées et qui permet d'exploiter le réseau sans modifier sa programmation, en utilisant des ni-

veaux de tension V_H et V_L pour les niveaux logiques hauts et bas. Comme l'algorithme d'apprentissage nécessite d'augmenter ou de diminuer cette conductance on peut considérer qu'au delà d'un seuil de tension (arbitrairement positif) V_{T+} la conductance augmente alors qu'en deçà d'un seuil négatif V_{T-} , la conductance diminue. On se trouve typiquement avec un composant dont le comportement est similaire à celui d'une CBRAM. Dans ces conditions, effectuer un apprentissage pour un tel réseau revient à appliquer les potentiels sur les extrémités accessibles (entrées et sorties) qui conduiront à appliquer sur les synapses dont la conductance doit être modifiée des tensions supérieures au seuil de programmation de façon que leur conductance varie dans le sens indiquée par la règle DELTA. **Supposons qu'un tel jeu de potentiel existe.** Considérons le cas de deux conducteurs pré-synaptiques $P1+$ et $P1-$ correspondant à des états différents (V_H et V_L) les deux étant connectés à deux dendrites (d_2 et d_3) devant mener à des états de sorties (Y_j) également différents. La règle DELTA (et de HEBB) imposent le signe de tous les potentiels $V_d - V_p$ aux bornes des synapses comme indiqué sur la figure 44. En partant du conducteur $P1+$ et en suivant le parcours désigné par les flèches A,B,C et D, **on ne suit que des tensions positives (supérieures à V_{T+} ou $-V_{T-}$), mais le circuit est fermé.** Ceci est clairement en contradiction avec la loi des mailles, et donc **impossible**. Nous pouvons en déduire qu'**aucun jeu de potentiel appliqué au réseau ne permet d'effectuer simultanément toutes les modifications réclamées par la règle DELTA.**

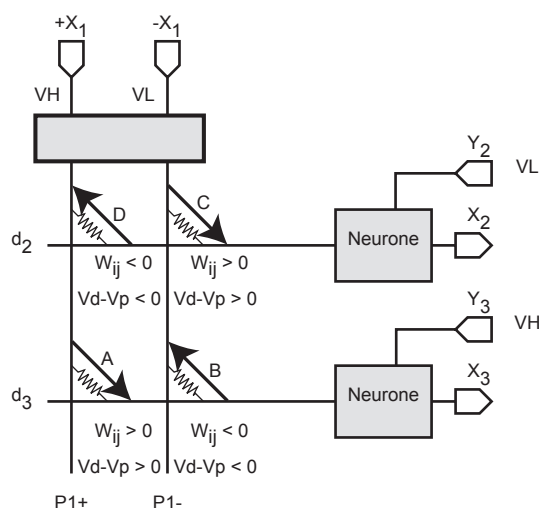


Figure 44 : En considérant la configuration d'entrées et sorties de cette figure, on est conduit en supposant qu'une configuration de potentiel permet de réaliser la règle DELTA pour toutes les synapses à admettre qu'il existe un parcours fermé que l'on peut suivre en ne rencontrant que des tensions positives. Ceci démontre, par l'absurde, qu'une telle configuration n'existe pas.

Cependant, nous avons trouvé différentes solutions pour contourner cette difficulté. Celles qui s'appliquent exactement au cas précédent, c'est-à-dire à un composant dont le contrôle de la variation de conductance est similaire à celui d'une CBRAM ne peut pas encore être dévoilées compte tenu des procédures de dépôts de brevet en cours. A contrario, nous pouvons maintenant révéler celles qui s'appliquent à un composant dont la conductance ne dépend que de l'amplitude de la tension à ses bornes. Il peut s'agir d'un composant dont la programmation est irréversible (par exemple par claquage successif des nanotubes les plus externes dans un nanotube multiparoï [46]) ou d'un composant dans lequel la réversibilité est assurée par une autre grandeur (par exemple la lumière [30]) et/ou par une troisième électrode. Dans ces conditions, **l'apprentissage peut être effectué en appliquant sur les conduc-**

teurs post-synaptiques un potentiel qui dépend uniquement de l'erreur effectuée par le neurone de sortie, mais pas des entrées. Celles-ci interviendront par le jeu d'un effet de seuil. Par exemple, si l'application d'un potentiel suffisant fait décroître la conductance (ce qui serait le cas par claquages successifs dans un nanotube multi-paroi), il suffit d'appliquer sur les conducteurs post-synaptiques (dendrites) correspondant à une sortie en erreur un potentiel correspondant au seuil de programmation (V_{T+} ou V_{T-}) de même signe que la sortie **attendue** (Y_j). La figure 45 illustre une réalisation possible de la cellule neurone qui réalise cette opération.

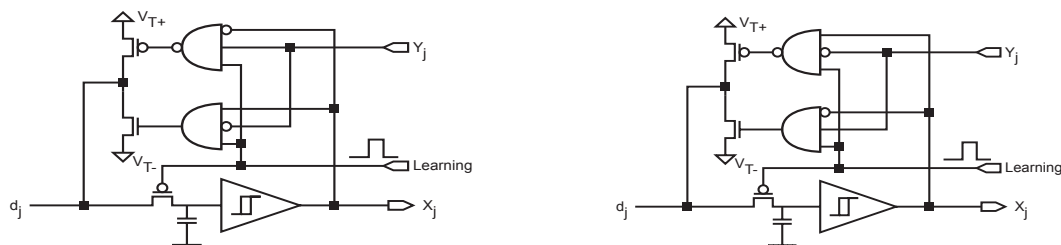


Figure 45 : Une réalisation possible d'un neurone avec apprentissage susceptible de réaliser la règle DELTA pour des synapses dont la conductance décroît (à gauche) ou croît (à droite) lorsqu'on applique une impulsion de tension d'amplitude supérieur aux seuils V_{T+} ou V_{T-} .

Seuls les synapses dont l'entrée est de signe opposée à ce potentiel verront une tension suffisante pour que leur conductance diminue, les autres étant soumises à une tension inférieure au seuil de programmation ne seront pas modifiées. Il est facile de vérifier que les synapses dont la conductance aura diminuée sont celles pour lesquelles la règle DELTA indiquent que le poids synaptique doit diminuer. A contrario, les synapses qui devraient voir leur conductance augmenter sont inchangées. Il reste à vérifier qu'un apprentissage n'exploitant que la diminution des conductances peut malgré tout converger. J'ai réaliser des simulations fonctionnelles qui permettent de le vérifier. La convergence est même étonnamment rapide comme le montre la figure 46.

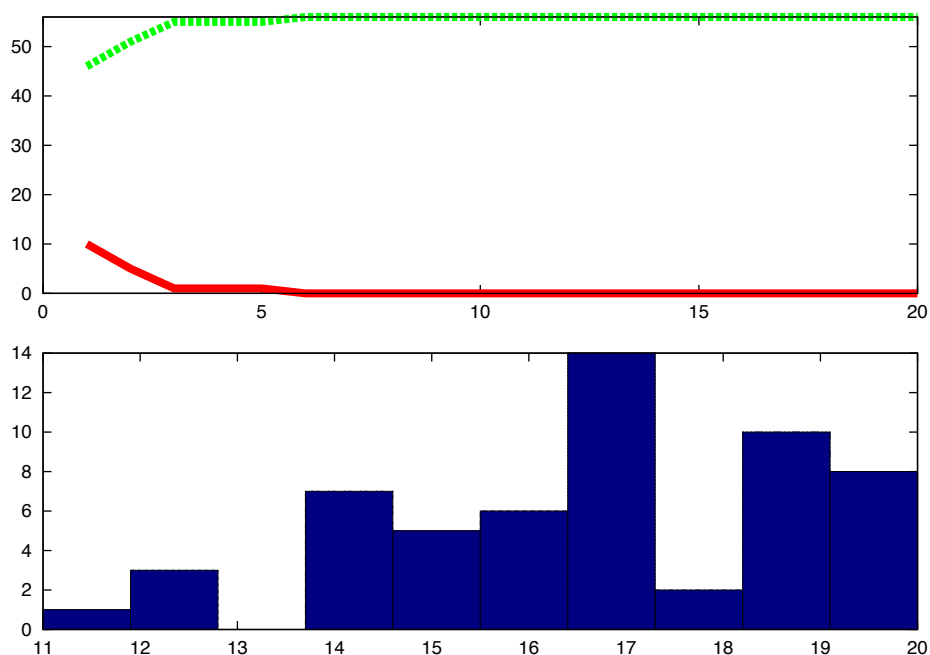


Figure 46 : En haut, convergence de l'algorithme d'apprentissage dans lequel les conductances ne peuvent que **décroître**. L'apprentissage porte sur un ensemble

de 7 fonctions logiques à trois entrées représentatives des fonctions linéairement séparables. La courbe verte en pointillée représente le nombre de bonnes réponses, la courbe rouge pleine le nombre d'erreurs. En bas, distribution des valeurs de conductance obtenues à la fin de l'apprentissage pour une initialisation de l'ensemble des poids à la valeur 20. Toutes les modifications des conductances sont unitaires.

Si nous considérons maintenant un composant dont le fonctionnement est inversé, c'est à dire dont la conductance **croît** lorsque l'amplitude de la tension appliquée sort de la plage neutre définie par l'intervalle $[VT+, VT-]$, la règle DELTA sera vérifiée en appliquant sur les conducteurs post-synaptiques correspondant à une sortie en erreur un potentiel correspondant au seuil de programmation ($VT+$ ou $VT-$) de même signe que la sortie **obtenue (X_j)**. Là encore, la convergence d'un tel apprentissage est particulièrement rapide comme le montre la figure 47.

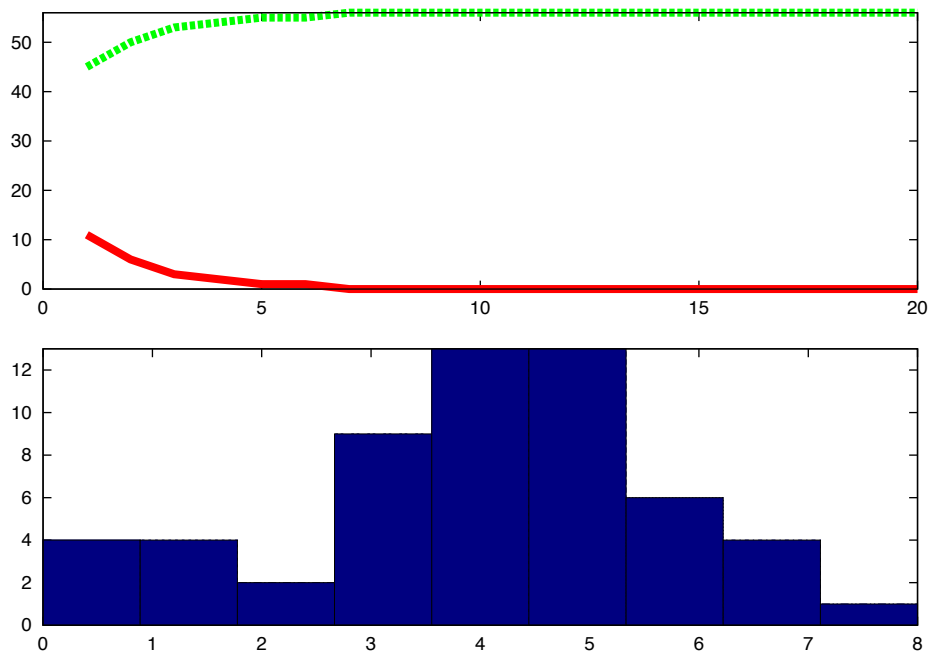


Figure 47 : En haut, convergence de l'algorithme d'apprentissage dans lequel les conductances ne peuvent que **croître**. L'apprentissage porte sur un ensemble de 7 fonctions logiques à trois entrées représentatives des fonctions linéairement séparables. La courbe verte en pointillée représente le nombre de bonnes réponses, la courbe rouge pleine le nombre d'erreurs. En bas, distribution des valeurs de conductance obtenues à la fin de l'apprentissage pour une initialisation nulle de l'ensemble des poids. Toutes les modifications des conductances sont unitaires.

Ainsi, sous réserve que des composants se comportent comme nous les avons modélisés, (c'est à dire qu'une variation incrémentale des conductances soit réalisable) il devient même possible d'appliquer ce type de procédure d'apprentissage à une assemblée de composants dont les connexions sont construites par une méthode bottom-up. La fonctionnalisation d'un tel réseau est un défi majeur tant il permet d'envisager les plus hautes densités d'intégration tout en conservant un coût de fabrication très réduit.

5. Conclusion

A partir des traits communs des algorithmes d'apprentissage neuronal, j'ai défini des

méthodes d'apprentissage adaptées à des nanocomposants présentant plusieurs états de conductance de façon à réaliser des fonctions logiques. Ces méthodes ouvrent la voie vers la fonctionnalisation d'assemblée auto-organisée d'éléments nanoscopiques obtenue par une approche bottom-up. Il s'agit d'une étape importante pour poursuivre l'accroissement des densités d'intégration tout en contrôlant les coûts de fabrication, enjeu économique majeur pour l'industrie microélectronique et pour tous les secteurs économiques qui dépendent de sa croissance. Cette piste sera donc poursuivie, notamment dans le cadre du projet ANR PANINI.



Coucher de soleil vers Königswalde, illustrant le terme «horizon» sur wikipedia. André Karwath. Licence : Creative Commons Attribution-ShareAlike 2.5

Conclusion et perspectives

IV. Conclusion et perspectives

1. conclusion

J'ai étudié l'impact des technologies sur les architectures de circuits de calcul. Cette étude m'a amené à considérer des approches que nous pourrions qualifier de «non-standards» par les modèles de programmation comme par les technologies mises en jeu. D'abord, l'étude des réseaux de neurones a mis en évidence l'influence que peut avoir la technologie de calcul analogique sur les algorithmes à mettre en oeuvre, l'importance de leur robustesse intrinsèque, les différentes contraintes de précision et de résolution et l'importance des évaluations de performance, en particulier en comparaison des architectures de processeurs numériques. L'ensemble de ces critères est à considérer en regard des familles d'applications envisagées. En particulier, en fonction des couples de technologie et des modèles de programmation choisis, nous pourrions plutôt envisager des applications dites «de niche» ou de masse. Clairement les architectures de réseaux de neurones que nous avons étudiés, très rapides mais dénuées de capacité de virtualisation, entrent dans le premier cas. Par la suite, ce dernier point m'a conduit à imaginer une nouvelle voie pour le calcul où l'originalité tient dans le choix des techniques de calcul analogique à temps discret alors que l'architecture et le modèle de programmation restent classiques pour une architecture de calcul massivement parallèle. En dépit du caractère totalement original des structures à capacités commutées envisagées comme partie opérative d'une architecture de calcul, l'approche à temps discret, similaire aux architectures numériques parallèles, permet de s'inspirer des méthodes de programmation et de synthèse automatique d'architecture numériques dont l'efficacité est prouvée. Enfin, toujours dans l'esprit d'appliquer des techniques standards de synthèse à des architectures qui ne le sont pas, j'ai étudié les architectures de calcul complètement magnétiques, utilisant les parois de domaines magnétiques poussées par des champs magnétiques comme support de l'information. Cette première étude portant sur les technologies à aimantation dans le plan a permis d'identifier les spécificités de telles architectures, en particulier celles liées aux difficultés de synchronisation et à l'absence de directionnalité des opérateurs, mais également de définir les outils et les méthodes de synthèse adaptées, se rapprochant des modèles les plus utilisés dans les architectures de circuits numériques dédiés au contrôle, tels que les automates.

2. Axes de recherche

L'orientation que je donne maintenant à ma recherche est résolument tournée vers les composants émergents et les architectures permettant d'en tirer parti. Compte tenu de l'impact qu'aurait la fin de la loi de Moore, il s'agit d'un enjeu majeur tant sur le plan scientifique qu'économique [17]. Nous retrouverons dans ce contexte les objectifs classiques pour les architectures de calcul : stocker, traiter et communiquer qui dessinent les défis scientifiques à relever : technologies de mémoires émergentes, nouveaux dispositifs de calcul et interconnexions. Au sein du département NST et de l'opération Nanoarchi que je dirige, nous sommes d'ors et déjà présents sur les deux premiers axes à travers la poursuite ou le démarrage des travaux sur les mémoires magnétiques (ANR-CIOMAG coordination IEF, Claude Chappert), sur les méthodes d'apprentissage d'inspiration neuronale dans le cadre du projet ANR que je coordonne (ANR-PANINI), mais également dans les autres opérations du département NST, notamment sur la logique à parois de domaines magnétiques dans le cadre de

l'ANR-ISTRADÉ coordonnée par Dafiné Ravelosona. Il sera cependant indispensable de prendre en compte le troisième axe tant les interconnexions et l'utilisation de la troisième dimension apparaissent aujourd'hui incontournables dans les architectures de Systèmes sur Puces et de «System In Package».

3. Technologies magnétiques

3.1. Projets à court terme

A court terme, nous poursuivrons bien entendu les travaux initiés dans le cadre des différents projets (CALOMAG, STREP-MAGLOG, CILOMAG) pour caractériser, modéliser et intégrer dans une chaîne de conception les mémoires magnétiques, en évoluant progressivement vers les technologies qui offrent le plus de perspectives en termes de scalabilité et de consommation. Ainsi, nous assurerons une transition des méthodes d'écriture par champ magnétique induit (qui nécessitent des courants relativement importants et rendent ces mémoires difficiles à faire évoluer vers les noeuds technologiques ultimes) vers les méthodes d'écritures plus économes en énergie lorsque la taille diminue : écriture thermiquement assistée, écriture par courant polarisé en spin. Dans ces derniers cas, les courants critiques d'écriture sont déterminés par les densités de courants traversant les composants, ainsi, toute diminution de la taille de ces composants se traduit par une diminution des courants requis et donc de la consommation.

3.2. Projets à moyen terme

Nous étudierons également les architectures de mémoires exploitant le déplacement des parois de domaines comme principe d'écriture ou comme principe de calcul. J'ai déjà étudié les architectures exploitant un champ magnétique externe rotatif, mais nous devons également étudier l'impact des différentes technologies envisagées aujourd'hui. En particulier, l'utilisation de dispositif à anisotropie perpendiculaire simplifie la génération d'un champ magnétique externe puisque le champ magnétique généré n'a plus besoin d'être tournant. Nous pouvons cependant noter que cela change profondément le mode de propagation des domaines magnétiques, les courbures ne permettant plus d'imposer un sens de propagation. Enfin, l'utilisation de courants polarisés en spin pour déplacer les parois de domaines magnétiques a aussi un impact fort sur les architectures puisqu'il devient alors possible d'agir sélectivement sur les parois de domaines alors que les champs magnétiques externes ne permettent que des actions globales.

3.3. Projets à long terme

A plus long terme, nous comptons explorer les méthodes d'écriture qui paraissent les plus intéressantes en termes de rendement énergétique et de scalabilité : par champs électriques et utilisant des courants de spin, sans déplacement de charge électrique. Enfin, la démarche entreprise au sein du département NST et de l'opération que je dirige de caractérisation, modélisation, intégration dans une chaîne de conception, et conception d'architecture peut s'appliquer à toutes les technologies émergentes de mémoires. A travers l'ANR PANINI comme à travers le thématique du GDR SOC SIP que j'anime «technologies émergentes» je suis en contact avec des partenaires qui pourront apporter leurs compétences en technologies, en caractérisation comme en modélisation compacte de façon à créer une dynamique équivalente à celle initiée sur les MRAM autour des différentes technologies de mémoires émergentes (OXRAM, CBRAM, PCRAM,...).

4. Opportunités architecturales

4.1. Exploitation de la non-volatilité

Par leur non-volatilité ces différentes technologies ouvrent de nouvelles possibilités en termes d'architecture afin de répondre aux défis qui se profilent pour l'électronique dans les années à venir. J'étudierai particulièrement, deux pistes : l'augmentation des performances par la reconfiguration dynamique et la diminution de la consommation statique. D'une part, nous pouvons exploiter la non-volatilité dans le cadre des architectures multicontexte reconfigurables dynamiquement. Le fait que les différents contextes de configuration soient mémorisés dans les éléments magnétiques placés au-dessus des éléments actifs, éventuellement organisés en plusieurs couches, permet de mémoriser plusieurs contextes sans affecter notablement la densité d'intégration. Par exemple, on imagine tout de suite l'intérêt que pourrait avoir une telle architecture pour implémenter des codecs multistandards dans le cadre d'applications multimédias nomades. Les schémas d'utilisation d'une telle reconfiguration, leur granularité et leur efficacité en termes de performances comme leur efficacité énergétique devront être évalués. D'autre part, la non-volatilité du circuit de calcul lui-même (et non de sa seule configuration) permet d'envisager des conditions d'utilisation radicalement différentes de celles en cours aujourd'hui. Par exemple la mise hors circuit des unités fonctionnelles inutilisées de façon temporaire, même à une échelle de temps très court, permet de réduire de façon drastique la consommation statique. Là encore, on peut imaginer tout l'intérêt qu'une telle approche peut apporter par exemple dans le cadre d'une application de réseaux de capteurs abandonnés auto-alimentés.

4.2. Architectures tolérantes aux défauts

La tolérance aux défauts et aux dispersions technologiques constitue un enjeu majeur pour être en mesure d'exploiter des composants produits à grande densité à un coût raisonnable, notamment par une approche bottom-up. Trois pistes sont envisageables aujourd'hui pour faire face aux taux de défauts important attendu dans les nanotechnologies : la reconfiguration, le codage redondant et l'approche neuronale. La première méthode qui nécessite l'identification des défauts et le placement routage des opérateurs spécifiques à chaque circuit fabriqué en fonction des défauts cartographiés ne me semble pas réaliste. J'étudierai donc les approches redondantes et neuronales, que ce soit pour des composants magnétiques ou pour d'autres type de mémoires émergentes. L'objectif ultime de la piste neuronale sera de définir des procédures d'apprentissage permettant de fonctionnaliser des réseaux aléatoires de composants obtenus par une approche bottom-up. Il sera alors possible d'exploiter des éléments nanoscopiques de très hautes densités, qui ne nécessitent pas de recourir à des méthodes de fabrication coûteuses, pour des traitements combinatoires classiques comme pour des tâches de classification typiquement neuronale.

Références Bibliographiques

V. Références bibliographiques

- [1] Dudek et al. A CMOS general-purpose sampled-data analog processing element. *Ieee T Circuits-II* (2000) vol. 47 pp. 467-473
- [2] Foldesy et al. A behavioural modelling technique for visual microprocessor mixed-signal VLSI chips. *International Journal of Circuit Theory and Applications* (2002) vol. 30 pp. 139-163
- [3] Dudek et al. A general-purpose processor-per-pixel analog SIMD vision chip. *Ieee T Circuits-I* (2005) vol. 52 pp. 13-20
- [4] Roca et al. A programmable imager for very high speed cellular signal processing. *J Vlsi Sig Process S* (1999) vol. 23 pp. 305-318
- [5] Roska et al. Toward visual microprocessors. *P Ieee* (2002) vol. 90 pp. 1244-1257
- [6] Galan et al. A bio-inspired two-layer mixed-signal flexible programmable chip for early vision. *Ieee T Neural Networ* (2003) vol. 14 pp. 1313-1336
- [7] Allwood et al. Magnetic domain-wall logic. *Science* (2005) vol. 309 (5741) pp. 1688-1692
- [8] Likharev et al. CrossNets - High-performance neuromorphic architectures for CMOL circuits. *Ann Ny Acad Sci* (2003) vol. 1006 pp. 146-163
- [9] Türel et al. Neuromorphic architectures for nanoelectronic circuits. *Int J Circ Theor App* (2004) vol. 32 pp. 277-302
- [10] Türel et al. Architectures for nanoelectronic implementation of artificial neural networks: new results. *Neurocomputing* (2005) vol. 64 pp. 271-283
- [11] Snider. Self-organized computation with unreliable, memristive nanodevices. *Nanotechnology* (2007) vol. 18
- [12] Allwood et al. Magnetic domain-wall logic. *Science* (2005) vol. 309 (5741) pp. 1688-1692
- [13] Allwood et al. Submicrometer ferromagnetic NOT gate and shift register. *Science* (2002) vol. 296 pp. 2003-2006
- [14] Faulkner et al. Controlled switching of ferromagnetic wire junctions by domain wall injection. *Ieee T Magn* (2003) vol. 39 pp. 2860-2862
- [15] Moore, E. F., «Gedanken Experiments on Sequential Machines,» *Automata Studies*, Princeton Univ. Press; 1956.
- [16] E. F. Moore, «Sequential Machines: Selected Papers,» Addison Wesley Publishing Company, Inc, Reading, Mass. (1964).
- [17] C. Saunier. L'évolution du secteur des semi-conducteurs et ses liens avec les micro et nanotechnologies, Rapport de l'OPECST n° 244 (2002-2003), déposé le 8 avril 2003.
- [18] Hopfield. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences of the USA*, 79(8), pp. 2554-2558 (1982)
- [19] Hinton, G. E., Sejnowski, T. J., «Learning and Relearning in Boltzmann Machines». In D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, pp 282-317. Cambridge: MIT Press, (1986).

- [20] AZENCOTT et al. SYNCHRONOUS BOLTZMANN MACHINES AND CURVE IDENTIFICATION TASKS. *Network-Comp Neural* (1993) vol. 4 (4) pp. 461-480
- [21] Jo et al. CMOS compatible nanoscale nonvolatile resistance, switching memory. *Nano Lett* (2008) vol. 8 pp. 392-397
- [22] Lai et al. An organic/Si nanowire hybrid field configurable transistor. *Nano Lett* (2008) vol. 8 pp. 876-880
- [23] Inoue et al. Nonpolar resistance switching of metal/binary-transition-metal oxides/metal sandwiches: Homogeneous/inhomogeneous transition of current distribution. *Physical Review B* (2008) vol. 77 pp. 035105
- [24] Meijer. Materials science - Who wins the nonvolatile memory race?. *Science* (2008) vol. 319 pp. 1625-1626
- [25] Waser et al. Nanoionics-based resistive switching memories. *Nat Mater* (2007) vol. 6 pp. 833-840
- [26] Lee et al. Two series oxide resistors applicable to high speed and high density non-volatile memory. *Adv Mater* (2007) vol. 19 pp. 3919
- [27] Gilbert et al. An embeddable multilevel-cell solid electrolyte memory array. *Ieee J Solid-St Circ* (2007) vol. 42 pp. 1383-1391
- [28] Schindler et al. Bipolar and unipolar resistive switching in Cu-doped SiO₂. *Ieee T Electron Dev* (2007) vol. 54 pp. 2762-2768
- [29] Park et al. Nonvolatile programmable metallization cell memory switching element based on Ag-doped SbTe solid electrolyte. *Applied Physics Letters* (2007) vol. 91 pp. 162107
- [30] Borghetti et al. Optoelectronic switch and memory devices based on polymer functionalized carbon nanotube transistors . *TNT2006*, 04-08 September, 2006, Grenoble-France (2006) pp. 2
- [31] Kaeriyama et al. A nonvolatile programmable solid-electrolyte nanometer switch. *Solid-State Circuits* (2005)
- [32] Baek et al. Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application. *Electron Devices Meeting* (2005)
- [33] Kund et al. Conductive bridging RAM (CBRAM): an emerging non-volatile memory technology scalable to sub 20nm. *Electron Devices Meeting* (2005)
- [34] Kozicki et al. Programmable metallization cell memory based on Ag-Ge-S and Cu-Ge-S solid electrolytes. *Non-Volatile Memory Technology Symposium* (2005)
- [35] Kaeriyama et al. A nonvolatile programmable solid-electrolyte nanometer switch. *Ieee J Solid-St Circ* (2005) vol. 40 pp. 168-176
- [36] Gilbert et al. A macro model of programmable metallization cell devices. *Solid State Electron* (2005) vol. 49 pp. 1813-1819
- [37] Stewart et al. Molecule-independent electrical switching in Pt/organic monolayer/Ti devices. *Nano Lett* (2004)
- [38] Specht et al. 20 nm tri-gate SONOS memory cells with multi-level operation. *Electron Devices Meeting* (2004)
- [39] Li et al. Multilevel memory based on molecular devices. *Applied Physics Letters* (2004)
- [40] Balakrishnan et al. Electrical characterization of solid state ionic memory ele-

- ments. Proceedings of the Non-Volatile Memory Technology Symposium (2003)
- [41] Kuekes et al. The crossbar latch: Logic value storage, restoration, and inversion in crossbar circuits. *Journal of Applied Physics* (2005)
 - [42] O'Connor et al. CNTFET modeling and reconfigurable logic-circuit design. *Ieee T Circuits-I* (2007) vol. 54 pp. 2365-2379`
 - [43] Fregonese et al. Computationally efficient physics-based compact CNTFET model for circuit design. *Ieee T Electron Dev* (2008) vol. 55 (6) pp. 1317-1327
 - [44] ZHU et al. Autocompensated capacitive circuit for stochastic neural networks. *Electron Lett* (1994) vol. 30 (4) pp. 330-331
 - [45] Markram et al. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* (1997) vol. 275 (5297) pp. 213-215
 - [46] Lee et al. Characteristics of multiwalled carbon nanotube nanobridges fabricated by poly(methylmethacrylate) suspended dispersion. *J Vac Sci Technol B* (2002) vol. 20 pp. 2773-2776
 - [47] Carmona et al. A 0.5 μm CMOS CNN analog random access memory chip for massive image processing. *Cellular Neural Networks and Their Applications Proceedings* (1998)

