HAL
open science

Solutions of subthreshold SRAM in ultra-wide-voltage range in advanced CMOS technologies for biomedical and wireless sensor applications

Anis Feki

To cite this version:

Anis Feki. Solutions of subthreshold SRAM in ultra-wide-voltage range in advanced CMOS technologies for biomedical and wireless sensor applications. Electronics. INSA de Lyon, 2015. English. NNT : 2015ISAL0018 . tel-02003583

HAL Id: tel-02003583

https://theses.hal.science/tel-02003583v1

Submitted on 1 Feb 2019

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.

N_ d'ordre: 2015-ISAL-0018

THÈSE

Présentée

Devant L´ Institut National des Sciences Appliquées de Lyon

Pour obtenir:

LE GRADE DE DOCTEUR

Ecole Doctorale: Electronique Electrotechnique Automatique
Formation Doctorale: Micro et Nano-électronique

Par

Anis FEKI

Ingénieur en Génie Electrique, Ecole National d'Ingénieur de Sfax (ENIS)
Master de Recherche, Université Pierre et Marie Curie (UPMC)

Titre de la thèse:

Conception d'une Mémoire SRAM en tension sous le seuil pour des applications biomédicales et les nœuds de capteurs sans fils en technologies CMOS avancées

Solutions of subthreshold SRAM in Ultra-Wide-Voltage Range in advanced CMOS Technologies for biomedical and wireless sensor applications

Soutenue le 29-05-2015 devant la commission d'examen composée de:

Jean-Michel PORTAL	Professeur, IM2NP, Univ. Provence Polytech'Marseilles	Président
Pascal NOUET	Professeur, LIRMM, Université Montpellier 2	Rapporteur
Luca LARCHER	Assistant Professor, DISMI, Università di Modena e Reggio	Rapporteur
Bruno ALLARD	Professeur, INSA-Lyon, Ampère	Directeur de thèse
David TURGIS	SRAM Manager, STMicroelectro	Encadrant
Olivier THOMAS	Ingénieur, CEA-LETI	Invité

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

INSA Direction de la Recherche - Ecoles Doctorales – Quadriennal 2011-2015

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** http://www.edchimie-lyon.fr Sec : Renée EL MELHEM Bat Blaise Pascal 3e etage 04 72 43 80 46 Insa : R. GOURDON secretariat@edchimie-lyon.fr	M. Jean Marc LANCELIN Université de Lyon – Collège Doctoral Bât ESCPE 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72.43 13 95 directeur@edchimie-lyon.fr
E.E.A.	**ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE** http://edeea.ec-lyon.fr M. Gérard SCORLETTI Secrétariat : M.C. HAVGOUDOUKIAN eea@ec-lyon.fr	M. Gérard SCORLETTI Ecole Centrale de Lyon 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60 97 Fax : 04 78 43 37 17 Gerard.scorletti@ec-lyon.fr
E2M2	**EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION** http://e2m2.universite-lyon.fr Sec : Safia AIT CHALAL Bat Atrium- UCB Lyon 1 04.72.44.83.62 Insa : S. REVERCHON Safia.ait-chalal@univ-lyon1.fr	M. Fabrice CORDEY Laboratoire de Géologie de Lyon Université Claude Bernard Lyon 1 Bât Géode – Bureau 225 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cédex Tél : 04.72.44.83.74 Sylvie.reverchon-pescheux@insa-lyon.fr fabrice.cordey@ univ-lyon1.fr
EDISS	**INTERDISCIPLINAIRE SCIENCESSANTE** http://www.ediss-lyon.fr Sec : Safia AIT CHALAL Bat Atrium – UCB Lyon 1 04 72 44 83 62 Insa : Safia.ait-chalal@univ-lyon1.fr	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 avenue Jean Capelle INSA de Lyon 696621 Villeurbanne Tél : 04.72.11.90.13 Emmanuelle.canet@univ-lyon1.fr
INFOMATHS	**INFORMATIQUE ET MATHEMATIQUES** http://infomaths.univ-lyon1.fr Sec : Renée EL MELHEM Bat Blaise Pascal 3e etage infomaths@univ-lyon1.fr	Mme Sylvie CALABRETTO LIRIS – INSA de Lyon Bat Blaise Pascal 7 avenue Jean Capelle 69622 VILLEURBANNE Cedex Tél : 04.72. 43. 80. 46 Fax 04 72 43 16 87 Sylvie.calabretto@insa-lyon.fr
Matériaux	**MATERIAUX DE LYON** http://ed34.universite-lyon.fr Sec : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry Ed.materiaux@insa-lyon.fr	M. Jean-Yves BUFFIERE INSA de Lyon MATEIS Bâtiment Saint Exupéry 7 avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72.43 71.70 Fax 04 72 43 85 28 Ed.materiaux@insa-lyon.fr
MEGA	**MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE** http://mega.universite-lyon.fr Sec : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry mega@insa-lyon.fr	M. Philippe BOISSE INSA de Lyon Laboratoire LAMCOS Bâtiment Jacquard 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72 .43.71.70 Fax : 04 72 43 72 37 Philippe.boisse@insa-lyon.fr
ScSo	**ScSo*** http://recherche.univ-lyon2.fr/scso/ Sec : Viviane POLSINELLI Brigitte DUBOIS Insa : J.Y. TOUSSAINT viviane.polsinelli@univ-lyon2.fr	Mme Isabelle VON BUELTZINGLOEWEN Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.77.23.86 Fax : 04.37.28.04.48 isavonb@dbmail.com

***ScSo : Histoire, Geographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie**

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

A la mémoire de mon père

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Acknowledgments

Starting from Tunisia, passing through Paris, then by Brussels and until my landing in Grenoble. When I turn around and I look at the path taken to reach this point, I realize that alone, I never would have succeeded. I would like to thank all people who have accompanied me during this trip. I am thinking here to Mom and dad, my two wonderful brothers (Ghassen and Jihed), all my family and all my friends (Cima, Ramzy, Taieb, Amène, Bilel, Ramona, Camilia, Med amine, Aymen Mili, Hassen, mahmud, Emna, Jung kyu...) without their help and support I would never have done this path.

My research work has been made in the STMicroelectronics Company (Crolles site) and in the Amper laboratory (INSA-Lyon). I'm very gratefull to my advisor Professor Bruno ALLARD for his guidance. Thank you Bruno, for guiding me, helping me and motivating me all the time. All my thanks to David TURGIS my industrial supervisor for including me into his SRAM design team during these three years of thesis, thank you David for your guidance. I want to thank all those who contributed to the progress of this research work: Special thank to Jean-Christophe LAFONT for his precious technical help and guidance. I would like to thank Fady Abouzeid and Sylvain Clerc from Rad-hard team for their help in the ULV aspects. I also thank Céline Le Gloanec and Sebastien Haendller from Tech2Design team for their support for implementation of memory « barrettes » and for silicon test. A great thank to Olivier Thomas from CEA-Leti for his valuable help and guidance at the « bist » level.

I would like to thank Wael Konzali and Med Siala from the I/O design team for helping me to resolve problems related to the « ESD » aspect. Many thanks to Hatem Sakhria and Slim Ellili from SoCs design team for their support.

I will never forget the time spent in the STMicroelectronics Company's site at Grenoble (Crolles). This hives of Abeille where nature and high technology intersect and cohabits harmonically.

I never forget the good times, discussions, jorkyball games and lunches I shared with my friends and collagues (I think here to Guy "le plus heureux", Jean Christophe, Ludovic, Fares, Jean Philippe, kaya, Aymen Mili, mohamed...). My thanks also go to all my colleagues at ST, as well as my colleagues from the Amper laboratory (INSA-Lyon).

My sincere thanks also go to the jury members who accepting to evaluate my thesis. Thank you to Professor Jean-Michel PORTAL to honor me by accepting to chair the jury of this thesis. Many thank to Professors Pascal NOUET and Luca LARCHER for accepting to be «rapporteurs» of this manuscript and giving me the honor of judging this work. Thank you to Dr Olivier Thomas for agreeing to examine my memory and to be part of jury.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Résumé

L'émergence des circuits complexes numériques, ou System-On-Chip (SOC), pose notamment la problématique de la consommation énergétique. Parmi les blocs fonctionnels significatifs à ce titre, apparaissent les mémoires et en particulier les mémoires statiques (SRAM). La maîtrise de la consommation énergétique d'une mémoire SRAM inclue la capacité à rendre la mémoire fonctionnelle sous très faible tension d'alimentation, avec un objectif agressif de 300 mV (inférieur à la tension de seuil des transistors standard CMOS).

Dans ce contexte, les travaux de thèse ont concerné la proposition d'un point mémoire SRAM suffisamment performant sous très faible tension d'alimentation et pour les nœuds technologiques avancés (CMOS bulk 28nm et FDSOI 28nm). Une analyse comparative des architectures proposées dans l'état de l'art a permis d'élaborer deux points mémoire à 10 transistors avec de très faibles impacts de courant de fuite. Outre une segmentation des ports de lecture, les propositions reposent sur l'utilisation de périphéries adaptées synchrones avec notamment une solution nouvelle de réplication, un amplificateur de lecture de données en mode tension et l'utilisation d'une polarisation dynamique arrière du caisson SOI (Body Bias).

Des validations expérimentales s'appuient sur des circuits en technologies avancées. Enfin, une mémoire complète de 32kb (1024x32) a été soumise à fabrication en technologie 28nm FDSOI. Ce circuit embarque une solution de test (BIST) capable de fonctionner sous une tension d'alimentation de 300mV.

Après une introduction générale, le 2ème chapitre du manuscrit décrit l'état de l'art. Le chapitre 3 présente les nouveaux points mémoire. Le 4ème chapitre décrit l'amplificateur de lecture avec la solution de réplication. Le chapitre 5 présente l'architecture d'une mémoire ultra basse tension ainsi que le circuit de test embarqué.

Les travaux ont donné lieu au dépôt de 4 propositions de brevet, deux conférences internationales, un article de journal international est accepté et un autre vient d'être soumis.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Abstract

Emergence of large Systems-On-Chip introduces the challenge of power management. Of the various embedded blocks, static random access memories (SRAM) constitute the angrier contributors to power consumption. Scaling down the power supply is one way to act positively on power consumption. One aggressive target is to enable the operation of SRAMs at Ultra-Low-Voltage, i.e. as low as 300 mV (lower than the threshold voltage of standard CMOS transistors).

The present work concerned the proposal of SRAM bitcells able to operate at ULV and for advanced technology nodes (either CMOS bulk 28 nm or FDSOI 28 nm). The benchmarking of published architectures as state-of-the-art has led to propose two flavors of 10-transitor bitcells, solving the limitations due to leakage current and parasitic power consumption. Segmented read-ports have been used along with the required synchronous peripheral circuitry including original replica assistance, a dedicated unbalanced sense amplifier for ULV operation and dynamic forward back-biasing of SOI boxes.

Experimental test chips are provided in previously mentioned technologies. A complete memory cut of 32 kbits (1024x32) has been designed with an embedded BIST block, able to operate at ULV.

After a general introduction, the manuscript proposes the state-of-the-art in chapter two. The new 10T bitcells are presented in chapter 3. The sense amplifier along with the replica assistance is the core of chapter 4. The memory cut in FDSOI 28 nm is detailed in chapter 5.

Results of the PhD have been disseminated with 4 patent proposals, 2 papers in international conferences, a first paper accepted in an international journal and a second but only submitted paper in an international journal.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Author’s Publications and patents

- **JOURNALS**
  - Anis FEKI, Bruno ALLARD, David TURGIS, Jean-Christophe LAFONT, Faress TISSAFI DRISSI , Fady ABOUZEID and Sebastien HAENDLER «Sub-threshold 10T SRAM Bit cell with Read/Write XY Selection». Accepted for publication to Solid-State Electronics Journal, 2015.
  - Anis FEKI, Bruno ALLARD, David TURGIS, Jean-Christophe LAFONT, Jean-Philippe Noel «Ultra-Wide Voltage Range Sensing Solution for Ultra Low-Power SRAM Memory» submitted to Solid-State Electronics Journal.
  - Abouzeid, F.; Bienfait, A.; Akyel, K.C.; Feki, A.; Clerc, S.; Ciampolini, L.; Giner, F.; Wilson, R.; Roche, P., "Scalable 0.35 V to 1.2 V SRAM Bitcell Design From 65 nm CMOS to 28 nm FDSOI," Solid-State Circuits, IEEE Journal of , vol.PP, no.99, pp.1,7. 2014.
- **International Conferences**
  - Feki, A.; Allard, B.; Turgis, D.; Lafont, J.; Ciampolini, L., "Proposal of a new ultra-low leakage 10T sub threshold SRAM bitcell," SoC Design Conference (ISOCC), 2012 International , vol., no., pp.470,474, 4-7 Nov. 2012.
  - Feki, A.; Turgis, D.; Lafont, J.C.; Allard, B., "280mV sense amplifier designed in 28nm UTBB FD-SOI technology using back-biasing control," SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 2013 IEEE , vol., no., pp.1,2, 7-10 Oct. 2013.
- **Patents**
  - Anis Feki, Jean-Christophe Lafont, David Turgis «Volatile Memory with a Decreased Consumption» US 20130201771 A1
  - Anis Feki «Volatile Memory with a Decreased Consumption and an Improved Storage Capacity» US 20130201766 A1
  - Anis Feki, David Turgis, Jean-Christophe Lafont «A Novel 8T-10T 1R/1W SRAM CELL Invention» (On going) Ref 12-GR1-0188
  - Anis Feki «NEW REPLICA CIRCUIT WITH TOLERANT PVT VARIATION» (On going) Invention Ref 14-GR1CO-0185.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Résumé étendu (français)

Cette partie propose un résumé étendu du contenu du manuscrit.
This section offers an extended summary of the manuscrit content in french.

La miniaturisation et l'intégration sont les clés de la révolution électronique. Depuis 1960, date de la première apparition du transistor MOSFET en circuit intégré conçue par D.Khang et M.Atalla, [1]. La microélectronique est très largement dominée depuis des années par la technologie des circuits intégrés numériques CMOS, à base de transistors MOSFET silicium. Cette technologie suit la fameuse loi de Moore (1965) [2]. Cette loi prévoit que le nombre de transistors dans un circuit intégrés double tous les deux ans et que la fréquence de fonctionnement du microprocesseur double tous les 18 mois. Après près d'un demi siècle cette loi reste toujours valide et les performances des circuits intégrés ne cessent de s'améliorer. Malheureusement, à chaque passage d'un nœud technologique à l'autre, la tension de seuil et la longueur de grille diminuent ce qui a pour effet l'augmentation de l'énergie statique dû aux courants de fuite. En parallèle, l'évolution des systèmes électronique portable a été très rapide. Ces systèmes deviennent de plus en plus essential pour la vie quotidienne de l'être humain. Ces systèmes varient entre les téléphones portables qui contiennent beaucoup de gadgets, les systèmes biomédicaux (pacemaker, prothèse auditive … etc) et les réseaux de capteurs sans fils. Ces systèmes exigent que le temps mis entre deux chargements de leur batterie soit très large. Cette situation fait naître beaucoup de défis à relever. Ces défis adressent principalement la surface, l'énergie consommée et les performances. La capacité de la batterie s'est largement améliorée durant ces dernières années. Malheureusement, cette amélioration reste insuffisante et ne satisfait pas les besoins des applications basse consommation en terme de budget d'énergie. Les systèmes basés sur la récupération d'énergie ont plus de contraintes en terme d'énergie consommé puisque la quantité d'énergie récupérer de l'environnement est limité et incertaine. Il faut donc conçevoir des circuits à très basse consommation. Les systèmes sur puces (SoC) contiennent des fonctions de plus en plus complexe, ce qui augmente le besoin en terme de budget d'énergie de la batterie. En particulier les mémoires statiques a accès aléatoire (SRAM) sont indispensables dans les SoCs et occupent une grande partie en terme de la surface et de l'énergie totale consommé. La conception d'une mémoire SRAM à très basse consommation devient aujourd'hui vitale, ceci demande beaucoup de défis à relever, d'efforts et de travaux de recherche dans les années a venir.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Les mémoires SRAM font partie des circuits mixtes. Une mémoire contient des cellules SRAM, des amplificateurs de lecture/écritures, des circuits d'assistance en écriture et en lecture et des générateurs d'horloges. Plusieurs topologies de mémoire SRAM existent, l'architecture d'une SRAM peut se diviser en quatre bloques : la matrice, le décodeur de lignes et de colonnes, le bloque de contrôle et le circuit de réplica & les entrées/sorties (Figure 1-1).

La réduction de l'échelle en technologie CMOS s'approche à ces frontières et il est attendu qu'elle atteint ces limites pour le nœud technologique 22nm selon l'ITRS [3]. Les principaux challenges et limitations qui empêchent la continuité de l'utilisation de cette technologie en future sont les limitation physique due à l'augmentation de l'effet tunnel et les courants de fuites ce qui impactent les performances et la fonctionnalité, les limitations technologique due aux techniques de lithographie utilisées qui sont incapable de fournir la résolution souhaités pour la fabrication des composants CMOS en technologies avancées et aussi le challenge économique qui présente un grand obstacle puisque le cout de production, la fabrication et le test des circuits intégrés rendent l'investissement dans une nouvelle technologie inabordable [4]. La diminution de l'épaisseur de grille T_{ox} s'est arrêté du à la limitation atomique, de la même manière la réduction de la tension de seuil s'est arrêté a cause de l'augmentation des courants de fuite [5] ($I_{LEAK} \sim exp(-V_{TH} \times q/nKT)$), L'énergie statique limite la réduction de la tension de seuil à une valeur égale à 200mV [6], la réduction de la tension d'alimentation s'est arrêté aussi puisque on peut plus réduire la tension de seuil (pénalité en terme de l'énergie statique consommé) : tous cela rend la diminution de la consommation et l'augmentation de la fréquence de fonctionnement trop faible (Le rapport « V_{DD}/V_{th} » définie la vitesse d'une porte logique). La Figure 1-4 présente la tendance de la puissance dynamique et statique en fonction de la longueur de canal pour une jonction a une température égale à 25°C dans un transistor standard [5]. Une extrapolation empirique indique que la puissance statique consommé devient égale à la puissance dynamique pour la technologie CMOS 20nm. La consommation statique devient problématique dans les nœuds technologiques avancés. Les courants de fuite présentent une contrainte majeure pour les applications basse consommation. Plusieurs nouvelles techniques de conception de circuit basse consommation ont été proposés durant ces dernières années afin de limiter l'impact de la consommation statique. Des nouvelles solutions technologiques sont attendus afin de rendre la fonctionnalité à très basse tension d'alimentation faisable. Des recherches doivent être mené afin de trouver des composants avec les meilleures performances possibles toute en gardant l'aspect basse consommation [7]. Dans ce contexte, les technologies FDSOI et FINFET apparaissent comme deux alternatives pour dépasser les limitations de la technologie CMOS standard (Figure 1-5). Aujourd'hui, se sont ces deux technologies qui sont capable d'aller en dessous de la technologie 28nm à l'échelle industrielle. Les bénéfices de la technologie FDSOI par rapport à la technologie CMOS sont illustrés

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

dans [8], [9], [10]. La technologie *Planar Fully-Depleted SOI* (FDSOI) présente des performances meilleures avec un rapport de 30% par rapport à la technologie CMOS standard pour le nœud 28nm [11]. De même, le FINFET a été récemment promu pour avoir du gain en coût et en compatibilité avec la technologie CMOS. La technologie FINFET est la première technologie qui a été commercialisé pour le nœud 22nm, qui a ouvert une nouvelle époque, celle des circuits intégrés 3D pour les applications basse consommation. Cette technologie a ouvert un nouveau chemin qui assure la continuité de la loi de Moore au-delà du nœud 20nm puisqu'elle présente des performances meilleures pour une consommation donnée par rapport a la technologie CMOS. Dans [12] ils prévoient que les nœuds technologiques 16nm/14nm en FINFET vont offrir entre 40-50% de plus en terme de performance et au alentour de 50% de réduction en terme de consommation comparé a la technologie CMOS 28nm. Une comparaison entre les deux technologies FDSOI et FINFET est illustrée dans [13]. Dans ce contexte la technologie 28nm FDSOI a été utilisé dans le travail présenté dans ce manuscrit.

Dans le but de réduire l'impacte des courants de fuite dans les nœuds technologiques avancés, plusieurs techniques ont été développées. Les recherches s'articulent autour de deux axes : la modulation de la tension de seuil dans le transistor et la gestion de la consommation dans les circuits intégrés. Deux principales méthodes peuvent être utilisées pour moduler la tension de seuil d'un transistor en technologie CMOS. La première consiste à l'ajustement de V_{th} à travers la modulation de l'épaisseur de la grille ou le dopage du canal [14]. Dans ce contexte plusieurs compagnies de semi-conducteur utilisent trois types de transistors (Low-V_T, Regular-V_T and High-V_T) dans leur plateforme technologique. Ces transistors sont caractérisés par leur tension de seuil : dans les circuit numérique les transistors de type LVT son utilisés puisque ils permettent d'avoir des fréquences élevées, cependant les transistors de type RVT et HVT sont utilisés dans le but de réduire les courants de fuite et pour minimiser la consommation statique. En résumé, il faut utilisé les transistor de type LVT pour avoir des gains en performance et les transistor de type HVT pour couper le chemin face aux courants de fuite [15] [16]. La deuxième méthode pour la modulation de V_{th} consiste à l'utilisation de la technique de « back biasing » [17]. Figure 1-6 présente les techniques de réduction de la consommation statique. L'empilement de transistors en série permet de réduire le courant de fuite avec un facteur entre 2-10x. L'ajustement de la tension de seuil V_{th} en utilisant la technique de « back biasing » permet une réduction entre 2-1000x et finalement, l'utilisation des transistors en mode « SLEEP » permet d'avoir une réduction entre 5-10x.

La consommation dynamique reste la principale source de perte d'énergie. La réduction de la tension d'alimentation est la meilleure méthode qui permet de réduire la consommation énergétique.

La réduction de la tension d'alimentation provoque la diminution quadratique de l'énergie dynamique et la diminution linéaire de l'énergie statique.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

La tension d'alimentation minimale (V_{MIN}) d'un syetème sur puce (SoC) est limitée par La tension d'alimentation minimale de la mémoire SRAM pour deux raisons : En premier lieu, avec la réduction de V_{DD} le délai dans la mémoire SRAM augmente d'une manière plus significative par rapport au délai dans la partie logique. Deuxièmement, la réduction de la tension d'alimentation provoque la dégradation de la stabilité de la cellule SRAM ce qui aura comme résultat l'apparition des erreurs de fonctionnement. D'ou la nécessité de développer des mémoires SRAMs qui travaillent a très basse tension.

L'étude du fonctionnement des circuits digitaux en tension sous le seuil et pré de la tension de seuil a été abordée en [18]. Récemment, plusieurs travaux sur la conception des microprocesseurs fonctionnant en sous le seuil ont été apparu dans le but de répondre aux exigence des applications « Internet of Things », les nœuds de capteur sans fils et les applications biomédicales. Dans [19], un processeur fonctionnant avec une tension d'alimentation en sous le seuil a été conçu en technologie 130nm. Ce processeur consomme 11nW pour une tension d'alimentation égale à 160mV et 3.5pJ/instruction pour une tension d'alimentation égale à 350mV. Egalement, un processeur FFT fonctionnant avec une tension d'alimentation minimal égale à 180mV a été présenté dans [20]. Cependant, la mémoire SRAM présente une partie critique dans un SoC car c'est elle qui limite l'efficacité énergétique puisque $VDD_{MIN,\ SRAM} > VDD_{MIN,\ LOGIC}$. Le travail qui a été fait dans [21] indique que le point d'énergie optimale de la mémoire SRAM est localisé dans la plage de tension sous le seuil. Donc, la conception d'une mémoire SRAM fonctionnant à très basse tension d'alimentation présente un axe de recherche promouvant pour concevoir des SoCs à très basse consommation. Plusieurs travaux sur la conception de mémoires SRAM à très basse tension d'alimentation ont été achevés durant les dernières années. Cependant, la plupart de ces travaux de recherche peuvent être classés dans deux catégories : la première concerne les mémoires SRAMs fonctionnant au dessus de la tension de seuil du transistor ($V_{DD} > V_{th}$), ce qui est le cas dans [22], [23]. Alors que La deuxième catégorie concerne les mémoire SRAM travaillant en dessous de la tension sous le seuil ($V_{DD} < V_{TH}$) [24], [25].

Cependant, il n'y a pas assez de travaux qui ciblent la conception de mémoires SRAM fonctionnant à la fois au dessus et au dessous de la tension de seuil comme c'est le cas dans [26], [27] et [28]. Afin de mettre en place ce genre de mémoire, des nouvelles techniques au niveau de la conception de circuits ainsi que du processus de fabrication, doivent être explorées pour assurer un bon fonctionnement dans cette plage de tension. Cette thèse s'adresse à cet axe de recherche et propose des nouvelles techniques pour permettre le fonctionnement dans une très large plage de tension d'alimentation (UWVR).

L'avancement de la technologie CMOS est limité par l'augementation de la variabilité locale qui affecte directement le rendement des circuits intégrés. Cette limitation devient plus significative pour

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

les nœuds technologiques avancés et provoque de lourdes pertes financière pour les compagnies de semi-conducteur. Les variations des caractéristiques électriques dues au désappariement entre deux dispositifs limitent les performances des circuits. Le désappariement est l'un des phénomènes les plus critique dans les circuits intégrés mixtes et spécialement dans les mémoires SRAM. En 1989, Marcel Pelgrom [30] a proposé la fameuse loi qui modélise le désappariement « mismatch » (voir équation 2-1). La principale source de variation en technologie CMOS est la fluctuation aléatoire de dopage (RDF) ou la variation de nombre de dopants dans la zone de déplétion (voir Figure 2-1). En général, le nombre de dopants dans un volume donné est contrôlé par la dose moyenne, Na. Ce nombre est variable d'un volume à l'autre dû au processus de fabrication aléatoire. La Figure 2-2 présente une simulation du comportement du courant drain-source pour différents rapports W/L pour les transistors NMOS et PMOS en technologie 28nm CMOS. Les résultats confirment la dépendance du désappariement de la taille du transistor MOS choisit, En augmentant la surface du composant on réduit l'impact du désappariement. Ce paramètre est très important puisque il impacte directement le rendement et les performances des mémoires SRAM. C'est pour cette raison que les concepteurs de mémoires SRAM doivent prendre en considération l'impact du désappariement afin d'avoir des mémoires SRAM fiables. Il existe deux sources de consommation d'énergie qui composent la puissance consommée dans une mémoire: la puissance dynamique qui est le résultat du phénomène de charge et de décharge des capacités durant les opérations de lecture et d'écriture, et la puissance statique qui est due aux courants de fuite (voir Figure 2-3). Les équations de la puissance dynamique et statique sont présentées dans (2-2) et (2-3). L'énergie par opération est une métrique qui est largement utilisée afin d'évaluer l'efficacité énergétique dans les mémoires SRAM. L'une des méthodes les plus utilisées pour diminuer l'énergie consommée est la réduction de la tension d'alimentation, cette méthode permet de réduire quadratiquement l'énergie dynamique et linéairement l'énergie statique comme c'est indiqué dans l'équation (2-4). La réduction de la tension d'alimentation V_{DD} a pour résultat l'augmentation exponentielle du délai, ce qui provoque l'augmentation de l'énergie statique qui domine l'énergie totale à très basse tension (voir Figure 2-5 et 2-6).

La surface et la stabilité sont les deux principaux paramètres qui caractérise une cellule SRAM : la surface occupée par une cellule SRAM représente environ 2/3 de la surface totale occupée par une mémoire SRAM, de plus la mémoire « L1 CACHE » occupe une grande partie de la surface de la plupart des SoCs « System on Chip ». La stabilité d'une cellule mémoire définie sa sensibilité face aux variations due au processus de fabrication, de la tension d'alimentation et de la température (PVT). Selon la loi de Pelgrom la stabilité et la surface sont inversement proportionnelles: on augmentant la surface de la cellule on diminue la sensibilité face aux variations. Dans les technologies CMOS avancées la cellule SRAM devient de plus en plus sensible aux différentes sources de bruit dû à

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

l'augmentation de la variabilité. Le défi qu'on doit relever pour la conception de mémoire fonctionnant à très basse tension est l'augmentation de la fiabilité durant les opérations de lecture et d'écriture. Dans ce contexte, plusieurs métriques ont été définies dans l'état de l'art pour caractériser la stabilité d'une cellule SRAM. Dans ce qui suit on va illustrer les différents métriques utilisés pour la caractérisation de la cellule SRAM 6T.

La cellule SRAM 6-Transistors est la plus utilisée aujourd'hui dans le milieu industriel. Cette cellule est composée de deux inverseurs montés en tête-bêche avec deux transistors (pass-gate) assurant la connexion avec les deux bitlines. Les deux transistors « pass-gate » (PG0 and PG1) sont contrôlés par le signal « WL » (wordline) pour assurer l'opération de lecture et d'écriture selon l'état des deux bitlines (BLT) et (BLF) (voir Figure 2-7). Ces deux bitlines agissent comme étant des nœuds d'entrée/sortie assurant le cheminement de l'information à partir des nœuds internes de la cellule vers l'amplificateur de lecture durant l'opération de lecture et depuis le circuit d'écriture vers la cellule durant l'opération d'écriture. La marge de bruit statique (SNM) est une approche qui a été étudié dans un premier temps sur les chaines de délais logique pour déterminer le minimum de tension (bruit) qu'on doit appliquer à l'entrée pour provoquer le basculement du niveau logique à la sortie. Cette approche a été adaptée par « Seevinck » pour établir la marge de bruit statique de la cellule [34] [35]. La méthode de « Seevinck » [36] est devenu alors l'approche la plus utilisée dans le flot de caractérisation des mémoires SRAM en industrie pour évaluer la « SNM » (Figure 2-8). Cette méthode peut être utilisée pour déterminer la SNM durant l'opération de lecture (SNM_{READ}), de l'écriture (WSNM) et le maintien (SNM_{HOLD}), cette métrique correspond au coté du plus grand carré qui peut être insérer dans la courbe « butterfly » (voir Figure 2-9).

La « write margin » (WM) caractérise la stabilité d'une cellule SRAM durant l'opération d'écriture. Plusieurs approches existent pour la détermination de la marge de stabilité en écriture : par la méthode de Seevink (voir Figure 2-12), par le balayage de la bitline (voir Figure 2-13) et par le balayage de la wordline (voir Figure 2-14). La tension de rétention des données « DRV » dans une cellule SRAM est la tension minimale pour maintenir les données dans les nœuds internes dans une cellule mémoire en mode « standby » (voir Figure 2-15). La consommation statique a largement augmenté avec les technologies CMOS avancées. Durant le mode de rétention, la consommation statique est la principale contribuable à la consommation totale de la mémoire SRAM (Equation 2-2). Durant le mode de rétention de données, les transistors d'accès sont ouverts, et la tension d'alimentation est ramenée jusqu'à la tension de rétention (DRV) afin de réduire la consommation statique [37]. L'influence de la diminution de V_{DD} sur les courants de fuite est illustrée dans la Figure 2-16. Cette Figure montre qu'une réduction de V_{DD} avec 40% provoque une diminution d'environ 70 % en terme de consommation statique. La diminution de V_{DD} durant le mode de rétention est une méthode très

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

bénéfique permettant réduire l'énergie statique consommée. Cependant, cette option peut provoquer la destruction des données dans les nœuds internes de la cellule (hold failures). Donc, il faut faire un compromis pour diminuer la consommation tout en gardant la fiabilité. La stabilité en lecture et en écriture sont les principaux métriques qui définie la stabilité de la cellule SRAM. Ces deux marges dépondent de la robustesse des facteurs α, β and γ: $\boldsymbol{\alpha} = \frac{(W/L)_{PG}}{(W/L)_{PU}}$, $\boldsymbol{\beta} = \frac{(W/L)_{PD}}{(W/L)_{PG}}$ *and* $\boldsymbol{\gamma} = \frac{(W/L)_{PD}}{(W/L)_{PU}}$

Afin de concevoir des cellules SRAM avec une bonne marge de stabilité en écriture, le transistor PU (Figure 2-7) doit être plus petit que le transistor PG (augmenter le facteur α), de même le transistor PD doit être plus grand que le transistor PG pour garantir une marge de stabilité en lecture acceptable (augmenter le facteur β). La relation (2-6) montre que l'amélioration du « WM » est limitée par la dégradation du « SNM_{READ} ». Similairement, l'amélioration de la stabilité en lecture est limitée par la taille du transistor PD, qui définie indirectement la surface occupée par la cellule qui doit être la plus petite que possible vu sont impact sur la surface totale de la mémoire. L'optimisation de la cellule SRAM doit être faite toute en minimisant le nombre d'échec durant l'opération de lecture ou d'écriture.

Vu la surface occupée par la cellule dans un système sur puce « SoC » beaucoup de compromis doit être tenue en compte, ceci rend la conception d'une cellule SRAM bien optimisée très délicate. La cellule SRAM devient plus sensible et la probabilité aux échecs « failures » augmente si le dimensionnement de la cellule ne tient pas compte de ces compromis. Même si la cellule est optimisée elle peut etre affectée par d'autre sources de « failures » comme la variation du « Process-Voltage-Temperature », le vieillissement, « the random telegraph noise » … etc. ces perturbations impactent directement la tension d'offset de l'amplificateur de lecture ainsi que le délai ce qui provoque des échecs au niveau du fonctionnement la cellule mémoire. Afin de concevoir une mémoire SRAM ayant un rendement acceptable, il faut prendre de la marge afin d'anticiper ces effets indésirables.

Comme il l'a déjà été expliqué précédemment, la diminution de la tension d'alimentation est la méthode la plus utilisée pour réduire la consommation totale dans une mémoire SRAM. La tension d'alimentation minimale (V_{MIN}) dans un SoC est limitée par la tension V_{MIN} de la mémoire. Dans une SRAM, la tension V_{MIN} est limitée principalement par la stabilité en lecture et en écriture de la cellule, puisque la stabilité en lecture SNM et celle en écriture WM dans une cellule 6T standard sont deux métriques inversement proportionnel. Ceci crée une limitation en terme de V_{MIN}. La tension minimale pour une cellule SRAM est le maximum entre $VDD_{MIN,\ SNM}$ et $VDD_{MIN,\ WM}$. Avec $VDD_{MIN,\ SNM}$ et $VDD_{MIN,\ WM}$ présentent la tension d'alimentation minimale correspondant à SNM ≥ 0 mV et WM ≥ 0 mV respectivement. Le fonctionnement de la cellule standard 6T à très basse tension avec un rendement acceptable est un défi, puisque ces deux métriques de stabilité en lecture et en écriture

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

seront dégradés et ils ne peuvent pas être optimisé en même temps. Les recherches et les prédictions ont montré que VDD_{MIN} de la cellule 6T est au alentours de 600mV [38].

Durant les dernières années, deux axes de recherches sont prédominants dans le but de concevoir des mémoires SRAMs fonctionnants à très basse tension d'alimentation. Le premier axe consiste à l'utilisation de la cellule 6T avec des techniques d'assistances pour l'opération de lecture et d'écriture permettant de fonctionner à des tensions plus basse. Cependant, l'utilisation de ces techniques aura comme impact l'augmentation de la complexité au niveau de la périphérie et le rajout de plus de source de consommation. Le deuxième axe consiste à chercher d'autres architectures de cellules SRAM qui seront capables de fonctionner en tension de sous le seuil sans avoir besoin d'être aidées par des circuits d'assistance.

Afin de résoudre les limitations en terme de WM et SNM à très basse tension dû au conflit au niveau du dimensionnement des transistors, les cellules dites favorisées en lecture ou en écriture sont apparus dans l'état de l'art, basé sur le principe de l'amélioration de la stabilité de lecture au détriment de la stabilité d'écriture et vice-versa, ceci par l'ajustement de la taille des transistors (W et L) et par l'ajustement de la tension de seuil V_{th} de la cellule mémoire [39]. En même temps, afin de compenser la dégradation de la stabilité d'écriture WM ou la stabilité de lecture SNM_{READ} des techniques d'assistance d'écriture ou de lecture doivent être utilisées. Pour concevoir des cellules SRAM favorisées en écriture, le facteur α doit avoir la valeur la plus grande (voir la relation 2-6). Ceci en rendant les transistors PG plus forts que possible ou en rendant les transistors PU plus faibles. Plusieurs techniques existent pour compenser la dégradation de la stabilité en lecture. Parmi les techniques les plus utilisées: la diminution de la tension V_{SS} [40] (Figure 2-22) et la diminution de la tension du signal « WL » (Figure 2-23). Similairement, pour obtenir des cellules SRAM favorisées en lecture, on doit maximiser les facteurs β et γ. Ceci s'obtient en rendant les transistors PD plus forts ou en rendant les transistors PG et PU faibles. Pour compenser la faiblesse de la stabilité d'écriture dans ces cellules, plusieurs circuits d'assistance sont utilisés : l'application d'une tension négative au niveau de la bitline (l'écriture d'un « 0 ») (voir Figure 2-24), l'augmentation du niveau de tension dans le signal wordline [45] (voir Figure 2-25). La réduction de la tension d'alimentation ou l'augmentation de V_{SS} (Figure 2-26). Durant ces dernières années, des nouvelles architectures de cellules SRAM ont été proposées pour résoudre la limitation de la tension V_{MIN} de la cellule 6T standard. Une excellente idée est apparu consiste a séparée le chemin de lecture de celui d'écriture, ceci grâce à l'utilisation d'une porte de lecture séparée au niveau de la cellule. Basé sur le fait que la marge de bruit statique d'une cellule 6T en mode de rétention est supérieure à celle en mode de lecture et qu'il a une limitation en terme de SNM puisqu'elle est inversement proportionnel à la stabilité en écriture WM. Les nouvelles architectures ayant une porte de lecture séparée ont une marge de stabilité en lecture

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

SNM_{READ} égale à celle en mode de maintien SNM_{HOLD} de la cellule SRAM 6T standard comme illustré dans la Figure 2-27.

La Figure 2-28 montre que la dégradation de la marge de lecture SNM dans la cellule 6T en réduisant la tension d'alimentation ce qui rend le fonctionnement à très basse tension susceptible aux erreurs même dans le cas ou des circuit d'assistance de lecture et d'écriture sont utilisées. La cellule 8T possède une marge de stabilité SNM_{READ} acceptable à très basse tension sans avoir besoin de circuit d'assistance de lecture ou d'écriture. Cependant une pénalité au niveau de la surface occupée par la cellule ayant un port de lecture séparé doit être tenu en considération. La tension d'alimentation minimale V_{MIN} est limitée principalement par la dégradation de WM et SNM dans la région de tension de sous le seuil. Plusieurs nouvelles cellules SRAM sont apparues pour rendre le fonctionnement à très basse tension possible. Les cellules SRAM 8T dans la Figure 2-29 (a) et celle en 2-29 (b) sont proposées comme solution pour éliminer la perturbation dans les cellules à moitié-sélectionnées durant les opérations de lecture et d'écriture. La cellule 8T (Figure 2-29 (b)) contient deux transistors PG en séries permettant la réduction des courants de fuite mais provoque la dégradation de la marge de stabilité en écriture WM et du courant de lecture dû au chemin résistive. La Figure 2-29(c) présente la cellule SRAM 7T [47] [48] dissymétrique et ayant un port de lecture séparé pour améliorer la stabilité en lecture SNM, cependant la marge de stabilité en écriture WM est faible dans cette cellule dû a l'accès d'une seul coté. La Figure 2-29(d) illustre la cellule 8T standard [49] [50] largement utilisée pour des applications à haute fréquence et aussi pour les applications très basse tension. L'inconvénient de cette cellule c'est la limitation du ratio I_{ON}/I_{OFF}. La Figure 2-29(e) présente la cellule SRAM 10T proposée dans [51], cette cellule contient deux ports de lecture séparés, l'inconvénient de cette cellule c'est l'impact négatif des courants de fuite à très basse tensionsur l'opération de lecture. La Figure 2-29(f) illustre une cellule 10T dissymétrique avec l'accès d'un seul coté pour l'opération de lecture, cette cellule utilise aussi un nouveau port de lecture permettant la réduction de l'impact des courants de fuite. Une cellule SRAM 8T appelée *ZIGZAG* est présentée dans [52] (Figure 2-29(g)) permettant l'amélioration du courant de lecture, cette cellule souffre d'un conflit entre le courant de fuite au niveau de la bitline et le courant de lecture. Figure 2-29(h) présente une autre cellule 10T VGND [45]. Plusieurs nouvelles architectures de cellule SRAM sont apparues durant ces deux dernières décennies permettant le fonctionnement de la mémoire SRAM en tension sous le seuil. La différence entre ces différentes architectures est la stabilité en écriture et en lecture (WM et SNM), l'immunité face aux perturbations (les cellules à moitié sélectionnées), le mode de lecture (« signle ended » ou « différentiel ») et la surface. D'une autre coté, les deux problèmes qui n'ont pas été évoqué dans la littérature sont la consommation parasite dans les cellules à moitié sélectionnées et

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

les erreurs de lecture à très basse tension due à la faiblesse du facteur I_{ON}/I_{OFF}, ces problématique vont être adressé dans le chapitre 3.

Le chapitre 2 s'est focalisé sur les avantages de la réduction de la tension d'alimentation afin d'atteindre le point optimum d'énergie dans la mémoire SRAM. Comme discuté précédemment, deux axes de recherche sont prédominants dans les travaux récents. Le premier consiste à l'utilisation de la cellule 6T dimensionnée pour favoriser la lecture ou l'écriture avec l'utilisation des techniques d'assistance. Alors que le deuxième axe consiste à utiliser d'autres architectures de cellule permettant la fonctionnalité à très basse tension sans besoin d'utiliser des circuits d'assistance. C'est dans cette direction que ce travail de thèse s'est focalisé. Des nouvelles architectures de cellules sont récemment proposées comme étant des solutions pour la limitation de la cellule 6T en terme de V_{MIN}. Malheureusement, ces nouvelles architectures sont limitées par : la faiblesse de leur rapport I_{ON}/I_{OFF} à très basse tension, l'échec de l'opération de lecture en tension sous le seuil et la faiblesse de leur immunité face aux perturbations.

Ce chapitre présent deux propositions de cellules (10T) fonctionnant à une tension d'alimentation égale à 300mV, ces cellules ont été fabriquées en technologie 28nm CMOS bulk. Les résultats de simulation sont confrontés aux résultats expérimentaux afin de valider les caractéristiques des cellules en tension sous le seuil. Les cellules sont proposées avec l'idée de séparée l'opération de lecture de celle d'écriture, ceci a permis d'apporter une solution à la limitation de la cellule (6T) en terme de stabilité de lecture et d'écriture (WM et SNM). Dans ce contexte, les cellules 8T [55], 9T [56], 10T [45] et 11T [57] sont apparus pour assurer la fonctionnalité à très basse tension. La cellule 8T a été proposée dans [58], cette cellule dispose d'un SNM en mode lecture égale à celui en mode de maintien. Une étude comparative a été proposée concernant les cellules [45], [59] et [52] respectivement (voir la Figure 3-1) fonctionnant à la tension sous le seuil en technologie 28nm CMOS bulk.

Dans les semi-conducteurs, la mobilité et la concentration de porteurs de charge sont dépendantes de la température. La Figure 3-2 illustre l'histogramme du courant de lecture, I_{CELL}, de la cellule SRAM 10T [45] pour différente valeur de température (-40°, 27° et 125°), obtenu par simulation Monte Carlo à très basse tension V_{DD}=300mV (Figure 3 (à gauche)) et à tension nominal V_{DD}=1V (Figure 3 (à droite)) respectivement. Le pic de la distribution dans la Figure 3-2 (à droite) présente la valeur de courant, I_{CELL}, la plus probable. La valeur de courant de lecture est plus petite à haute température à tension nominale. Ce phénomène est dû à la dégradation de la mobilité avec l'augmentation de la température. Cette tendance est inversée dans la plage de tension sous le seuil (voir la Figure 3 (à gauche)). La génération thermique de paires électron-trou contrebalance la dégradation de la mobilité

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

à très basse tension. La diffusion de la distribution est variable à très basse tension (ULV) alors qu'elle est constante pour V_{DD}=1V. La Figure 3-2 (à gauche) montre que le courant de lecture est très faible à basse température à basse tension.

Dans les applications haute fréquence, un amplificateur de lecture est utilisé quand la cellule SRAM est accédée, afin de détecter une certaine différence de voltage entre les deux bitlines dans le but d'accélérer l'opération de lecture. Malheureusement, avec la réduction de la tension d'alimentation les courants de fuite augmentent exponentiellement et provoquent la décharge des bitlines. L'opération de lecture se passe normalement quand le rapport I_{ON}/I_{OFF} est élevé. Autrement dit, le courant de lecture doit être capable de décharger la bitline plus rapidement que la somme des courants de fuite qui proviennent de la bitline vers les cellules SRAM non-sélectionnées dans une même colonne ($\Sigma I_{Leak}<I_{Read}$). La réduction de la tension V_{DD} réduit fortement la valeur moyenne du courant de lecture (Figure 3-3) et augmente l'impacte de la variation du process, tempréature et de la tension (PVT) sur les performances de la cellule. Les courants de fuite des cellules non-sélectionnées dans le domaine très basse tension représentent un obstacle important qui provoque un dysfonctionnement durant l'opération de lecture : pour cela une solution doit être fourni pour réduire les courants de fuite. La dégradation du courant de lecture dans une cellule donné par rapport aux courants de fuite générés par les cellules non-sélectionnés de la même colonne devient critique et provoque une défaillance dans le fonctionnement durant l'opération de lecture, en particulier à haute température et lorsque la tension d'alimentation est inférieure à 400 mV (voir Figure 3-2).

Un autre obstacle au bon fonctionnement de la mémoire SRAM à très basse tension est les « soft-error ». Dans les mémoires ou les circuits séquentielle, une perturbation « soft-error » (SE) est causée par une particule d'énergie qui entre dans la puce et génère suffisamment de charges libres pour basculer l'état d'une cellule [60]. La sensibilité aux SEs est directement liée à la capacité de la cellule: plus la capacité est petite, plus la sensibilité est grande [61]. La sensibilité aux SEs augmente avec la réduction de la tension d'alimentation [62]. Les SEs sont plus critiques en tension sous le seuil par rapport à la tension nominal. La technique la plus utilisée pour éviter les perturbations dûes aux SEs est d'entrelacer les bits (Bit-interleaving) comme indiqué sur la figure 4.3 (c), de telle sorte que les bits logiquement adjacents ne sont pas physiquement adjacentes. Comme déjà indiqué, la technique d'entrelacement de bits peut résoudre plusieurs SEs a très basse tension [45]. La Figure 3-5 décrit l'opération de lecture et d'écriture. Quand une cellule est sélectionnée soit pour la lecture ou d'écriture, toutes les cellules situées dans la même ligne (connectées au signal WWL) sont à moitié-sélectionnées et elles ont un potentiel flottant sur leurs bitlines. Ces cellules non sélectionnées souffrent de pertes parasites durant les opérations de lecture et d'écriture, qui sont dues à un courant de décharge. Le comportement parasite dans les cellules moitié-sélectionnées affecte directement la consommation

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

d'énergie dynamique, comme c'est le cas dans la cellule standard 8T [58] et de la cellule 10T [45]. Au cours d'une opération d'écriture, un courant parasite est injecté de V_{DD} à la bitline BLF comme indiqué sur la figure 3-5 (b). Pour éviter ce phénomène, on doit sélectionner un seul mot par ligne. Malheureusement, cette technique augmente le nombre de ligne, ce qui impacte négativement la complexité de l'architecture de la mémoire SRAM. La mise en œuvre d'une technique de multiplexage est proposée ici pour permettre d'avoir plusieurs mots par ligne sans avoir des pertes de consommation dans les cellules non sélectionnées. Une solution principale pour lutter contre le problème du courant parasite dans le port de lecture est d'utiliser une ligne de mot de lecture virtuelle (RWL_MUX) accessible par le port de lecture constitué d'un simple transistor (M9, M10 à la figure 3-6 (a)) dans une configuration multiplexé. Cette solution est censée offrir un temps de lecture minimum (port de lecture compsé d'un seul transistor).

La figure 3-6 (a) montre le schéma de la cellule proposée 10T-MUX SRAM. Cette cellule symétrique comprend deux inverseurs en tête-bêche et deux transistors d'accès en série de chaque coté permettant l'utilisation d'une technique d'entrelacement de bits. Ceci garantit un chemin résistif entre la bitlines et les nœuds de données internes. Les transistors de lecture sont utilisés pour transférer les données sur les lignes de bits de lecture (RBLT, RBLF). Figure 3-6 (b) montre le dessin de masque de la cellule proposée conçu en technologie CMOS 28nm occupant 0.8μm^2 de surface. Les transistors dans le chemin de lecture ont été dimensionnés afin d'obtenir une valeur de SNM a très basse tension acceptable. Afin d'obtenir un fonctionnement correct pendant l'opération d'écriture (pour éviter les courants parasites entre les lignes de bits et le signal RWL_MUX), la cellule 10T-MUX nécessite l'utilisation de quatre lignes de bits. Une technique de codage figée a été introduite pour résoudre la problème des cellules à moitié-sélectionnée [57]. Cette technique consiste à multiplexer le signal de RWL_MUX en fonction du nombre de mots par ligne. Cette option permet de sélectionner un seul mot pendant l'opération de lecture et d'éliminer les pertes dynamiques parasites des lignes de bits non sélectionnés (Figure 3-6 (c)).

Il y a donc une consequence évidente due à cette technique sur la complexité de la cellule et de la périphérie. Le principe de fonctionnement de la cellule mémoire 10T-MUX est illustré dans la figure 3-7 (a). En outre, la cellule 10T-MUX présente une limitation due à l'injection de charge de la ligne de lecture flottante. C'est donc la raison principale pour laquelle une autre solution de cellule est considérée. Figure 3-9 montre le schéma de différentes ports de lecture dans la littérature (port de lecture standard avec un seul transistor MOS, port de lecture ayant une masse virtuelle (VGND) et le port de lecture à sélection croisée XY proposé. La cellule mémoire 10T-MUX dans la figure 3-6 avec les signaux multiplexés RWL représente une solution pour éviter la consommation d'énergie due aux phénomènes parasites. Malheureusement, cette technique est complexe, car nous ne pouvons pas

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

empiler plus que deux mots par ligne et le fait que le décodage de ces signaux multiplexés va ajouter plus de contraintes sur le décodeur des lignes ce qui rend la tâche du concepteur assez difficile. Le schéma proposé de la cellule 10T-XY SRAM est présenté dans la Figure 3-16. Cette cellule se compose de deux inverseurs CMOS couplés en tête-bêche, deux transistors d'accès en série comme dans la cellule mémoire 10T-MUX sur la Figure 3-6 et un port de lecture en selection croisée XY (comme dans la Figure 3-9). Les transistors de lecture sont utilisés pour transférer les données internes vers la ligne de bit de lecture (RBLF). Comme représenté dans la Figure 14.03, l'impact du courant de fuite dans les celulles non sélectionnées au niveau d'une colonne devient négligeable, grâce à l'application d'une tension négative à la grille du transistor M10 (Figure 3-16) dans le port de lecture XY: ce qui a permet de resoudre les problèmes de défaillance Durant l'opération de lecture. Le principe de fonctionnement de la cellule 10T-XY est rapporté dans la figure 3-7 (b) avec les histogrammes des signaux de commande. Le tableau 3.2 présente le réglage des signaux de commande pour chaque mode de fonctionnement dans la cellule 10T-XY. Les deux cellules SRAM 10T proposées sont comparées à ceux de [45], [59] et [52] (Figure 3-1 respectivement (a), (b) et (c)). La cellule mémoire 10T-XY est la meilleure en terme d'énergie consommer puisque cette cellule permet d'éviter les pertes d'énergie dynamiques, réduit le temps de lecture et en même temps la consommation durant l'opération de lecture. La cellule 10T-XY est à ce jour la meilleure candidate en termes de courant de fuite par rapport aux autres cellules (72% de moins par rapport à la 10T VGND). Les cellules proposées ont été conçues et caractérisées sur une large plage de tension. Les mesures sont comparées aux résultats de simulation. La Figure 3-18 montre les courbes papillons expérimentaux pour la cellule 10T-XY pour divers tension d'alimentation à 25 °C. La Figure 19.3 présente les mesures de SNM en fonction de la tension d'alimentation pour différentes températures. Les cellules proposées ont plus de 50 mV de SNM à 25 °C pour une tension d'alimentation égale à 300mV, ce qui garantit une marge acceptable pour la lisibilité. Le fonctionnement des circuits en sous le seuil devient de plus en plus attractif grâce à la consommation ultra-faible à très basse tension. Malheureusement, la réduction de la tension d'alimentation s'accompagne d'une importante limitation dans le temps d'accès en lecture qui empêche le fonctionnement à haute fréquence et limite le champ d'applications possibles.

Le temps d'accès dans le domaine de l'ultra-basse tension est principalement dicté par le courant de lecture de la cellule SRAM et la capacité effective de la bitline. Ce chapitre présente deux contributions permettant un fonctionnement à très basse tension des cellules SRAM proposées au chapitre 3. Un amplificateur de lecture optimisé (SA) est d'abord détaillé. Un circuit de répliqua est ensuite introduit pour prendre soin des impacts de la variabilité et pour fournir des informations synchronisé dédié pour l'opération du SA. Les travaux sur les mémoires SRAM dans l'état de l'art ont

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

montré que la tension optimale qui permet d'avoir une énergie minimale est de l'ordre de 300mV pour les technologies Sub-90nm [15]. Dans le domaine de fonctionnement ultra-basse tension, la fréquence de fonctionnement est très faible, puisque Le temps d'accès en lecture est affecté négativement par une dégradation du courant de lecture de la cellule mémoire et puisque la variabilité augmente dans cette plage de tension. La technique la plus utilisée pour assurer l'opération de lecture est la technique « full-swing » [17]. Malheureusement, cette technique est un processus lent (grand retard). Il y a donc un intérêt pour optimiser les amplificateurs de lecture (SA) à la fois pour les cellules « single-ended » et différentielles dans la plage de tension sous le seuil. La littérature reporte que la tension de fonctionnement minimale des amplificateurs de lecture et au alentour de 500 mV à cause de l'augmentation du « mismatch » à très basse tension [54]. Dans ce chapitre, les avantages de la technologie FDSOI par rapport à la technologie CMOS bulk est détaillé en premier. La motivation et les principales limites de la détection des données en ULV sont ensuite présentées. Un amplificateur de lecture optimisé est illustré permettant d'améliorer le temps de lecture de la cellule proposé 10T-XY. Un nouveau circuit de répliqua avec une tolérance aux variations « PVT » est présenté et enfin une technique d'adaptation permettant d'optimiser l'opération de détection dans un SA est proposée. La technologie CMOS bulk est confrontée à de nombreux défis pour répondre aux exigences du nœud technologique 28nm. Deux principales limites sont la variabilité et l'électrostatique [72]. La technologie « Ultra-Thin Body and BOX, Fully Depleted SOI » (UTBB FDSOI) est apparu comme une alternative pour les futurs circuits intégrés avec moins de variabilité que dans le cas du CMOS bulk [72]. La polarisation arrière « back-biasing » consiste à appliquer une tension juste sous la BOX des transistors cibles [70]. Il en résulte la modification de la commande électrostatique des transistors et le décalage de leur tension de seuil. Ceci permet d'avoir plus de courant (rendement élevé), en contre partie le courant de fuite augmente (plus de consommation statique) comme le montre la figure 3.4 (c). Dans le cas de L'application du « Forward Body Bias» (FBB), la tension de seuil Vth diminue ce qui résulte à l'augmentation du courant (amélioration de la fréquence de fonctionnement), mais avec l'augmentation du courant de fuite (plus de puissance statique). Dans le cas de l'application du « Reverse Body Bias » (RBB), la tension de seuil Vth augmente ce qui provoque la dégradation du courant (dégradation de la fréquence), mais s'accompagne de la diminution du courant de fuite (réduction de la consommation statique).

La figure 4-5 présente les solutions qui permettent d'adapter le comportement des transistors en fonction des spécifications: les transistors LVT améliorent la fréquence de fonctionnement et les transistors HVT améliorent le rendement énergétique. Le fonctionnement de la mémoire SRAM dans le domaine de l'ultra-basse tension est très attractif pour réduire la consommation. Malheureusement, la plupart des applications n'ont pas besoin seulement de l'efficacité énergétique mais aussi d'une

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

plage acceptable de fréquence. Cependant, la fréquence de fonctionnement est limitée par des valeurs excessives de temps de lecture (comparée au temps d'écriture, Figure 4-8). La figure 4-9(a) représente un cycle de lecture d'une mémoire SRAM classique. Dans un premier temps, la capacité de la bitline est pré-chargée à travers un transistor PMOS. Ensuite, en fonction des données stockées, soit la ligne de la bitline de lecture est déchargée, soit elle reste chargée pendant l'opération de lecture. Enfin, l'amplificateur de lecture est activé et en fonction de la valeur dans les deux bitline de lecture, la donnée de sortie prend la valeur "0" ou "1". La figure 4-9 (b) illustre le temps de réduction cible dans les éléments de décharge et de détection. Nous ciblons d'optimiser deux composantes de temps: le temps de décharge et le temps de détection. Le délai de décharge acceptable de ligne de bit afin de garantir une opération de lecture réussie est difficile à prévoir. La Figure 4-10 présente les circuits de détection de l'opération de lecture pour les petits et les grands signaux [74]. Cependant, la dégradation du courant de lecture et l'augmentation de la variabilité provoquent la dimunition du temps de lecture ce qui limite la fréquence de fonctionnement possible dans la plage de tension sous le seuil.

C'est pour cette raison que l'utilisation de la technique de détection «full swing » en ULV n'est pas aussi avantageuse. Donc, l'utilisation d'un amplificateur de lecture est primordiale. L'équation 4.1 évalue le temps de décharge pour une tension différentielle, ΔV_{BL}, qui occupe une grande partie du temps de lecture (Figure 4-9). Le tableau 4-1 présente l'évolution du temps de décharge des bitlines pour atteindre une différence de voltage ΔV_{BL}, pour des tensions d'alimentation égalent à 300mV et 1V. Les résultats sont obtenus par simulation du chemin critique (une matrice de 128 cellules x 64 cellules différentielle 10T-XY, comme illustré à la Figure 4-11). Le tableau 4.2 illustre les valeurs de temps de lecture avec en utilisant un amplificateur et en utilisant la technique full-swing ». Si on résume, dans la plage de tension sous le seuil, il y a deux principales limitation pour l'opération de lecture en utilisant la technique « full-swing »: la limitation du nombre de cellules mémoire qu'on peut empilés dans une colonne à cause de la capacité de la bitline et la dégradation du courant de lecture, d'où la dégradation du temps de décharge. Ainsi, afin d'améliorer la fréquence de fonctionnement en ULV, nous nous sommes concentrés sur la solution qui consiste à l'utilisation d'un SA dans une plage de tension ultra-large [300mV, 1.3V].

L'amplificateur de lecture « SA » est le composant le plus critique dans la périphérie de la mémoire SRAM. Le rôle d'un SA est d'amplifier la tension d'entrée différentielle, ΔV_{BL}, développé entre les deux bitlines. La différence de voltage minimale, ΔV_{MIN}, est limitée par l'offset du SA. La tension d'offset est dû au « mismatch » dans les transistors à l'entrée du SA. Au moment du déclenchement, la tension d'entrée différentielle ΔV_{BL} doit être supérieure à l'offset: $\Delta V_{BL} \geq V_{OFFSET}$. Les performances de l'amplificateur impact directement le temps d'accès en lecture de la mémoire et la consommation dynamique. Plusieurs architectures d'amplificateurs ont été proposées dans l'état de l'art. Le SA à base

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

de « latch » [75] [76] est le plus utilisé dans les mémoires SRAM grâce à un compromis favorable en terme de la consommation d'énergie et de la vitesse (le temps de reaction du SA). La Figure 4-13 présente deux types d'amplicateurs. Le « VLSA » [77] (figure 4-13 (à gauche)) amplifie la difference de voltage, ΔV_{BL}, développée entre les deux bitlines. Le second, c'est l'amplificateur de détection de courant (CLSA) [78] (figure 4-13 (à droite)) qui amplifie une différence de courant crée par la différence de tension développée entre les deux bitlines. D'une part le VLSA est meilleur en terme de temps de réaction et occupe moins de surface que le CLSA (deux transistors NMOS supplémentaires). D'autre part la CLSA n'a pas la même contrainte que la VLSA en termes de précision nécessaire pour le signal SAEN afin de séparer clairement les nœud d'entrée et de sortie, les nœuds de sortie servent comme des nœuds d'entrée par le biais des transistors d'accès PMOS dans le VLSA (MP0 et MP3 dans la figure 4-13 (gauche)) [79].
Des simulations « Monte Carlo » sont appliquées aux amplificateurs VLSA et CLSA dans le domaine ultra-basse tension. Malheureusement, il n'y a pas de points de convergence détectés pour l'amplificateur de type CLSA car elle a beaucoup de défaillance à cause du « mismatch » qui provoque une large variation du courant de lecture. Cela rend l'opération du CLSA presque impossible en ULV. C'est la raison pour laquelle on s'est intéressé à l'amplificateur de lecture en mode tension.

L'amplificateur de lecture à base de « latch » (voir la Figure 4-15 (gauche)) est largement utilisé dans les mémoires standards car il présente une grande impédance d'entrée et il offre un gain en tension élevé avec un circuit simple. Dans ce qui suit, seul l'amplificateur de type VLSA est considéré. La cellule mémoire proposé 10T-XY posséde une configuration « single ended » (une seul bitline de lecture). Par conséquent, afin d'accéder aux données internes de la cellule en mode de lecture, l'utilisation d'un amplificateur de lecture déséquilibré est nécessaire (voir la figure 4-15 (à droite)), comme c'est le cas pour la mémoire DRAM. Selon l'architecture de la cellule, la détection du SA peut être différentielle comme c'est le cas pour la cellule 6T ou asymétrique similaire au cas de la cellule 8T standard (un seul port de lecture est utilisé). Afin de valider la faisabilité de la détection du SA en tension sous le seuil, Deux configurations d'amplificateur de lecture (différentiel et déséquilibrée VSA) ont été conçus pour les cellules ULV asymétriques et différentielles. La tension différentielle de la bitline minimale est limitée par l'offset du SA. La technique la plus utilisée pour réduire l'offset est assurée par l'augmentation des tailles des transistors critiques. D'abord, un amplificateur différentiel VLSA (la figure 4-15 (à gauche)) est conçu. Les signaux des nœuds internes sont présentés dans la figure 4-16 pour V_{DD} égale à 280 mV et 1V respectivement. Le « Forward Body-Biasing » est évalué à 0V, 280 mV et 1V respectivement. Comme le montre la figure 4-16(a) le temps de réaction du VLSA pour une tension d'alimentation de 280 mV est largement réduite (environ 80%) grâce au FBB.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Cependant, le FBB n'a pas d'impact sur le rendement du SA à la tension nominale comme montré sur la Figure 4-16 (à droite). La Figure 4-17 confirme l'avantage de l'application d'une polarisation en arrière pour améliorer les performances: la réduction du temps de réaction du SA de 80% à V_{DD}=280 mV et de 13% à V_{DD}=1V avec un FBB=1.2V. La probabilité de défaillance dépend beaucoup de la variation de la tension de seuil. Avec la réduction de la tension d'alimentation, la variation du Vt (σ_{Vth}) devient la plus grande source de défaillances durant l'opération de lecture. La tension d'offset correspondant à la tension d'entrée différentielle minimale pour laquelle le SA réalise l'opération de lecture avec zéro défaillance dans les pires conditions. Les Figure 4-18 (a) et (b) présentent la probabilité de défaillance d'un amplificateur différentiel de lecture optimisé à 280 mV et 1V respectivement. En conséquence, l'amplificateur différentiel optimisé en technologie 28nm FDSOI a une valeur d'offset égale à 40 mV à V_{DD}=280 mV et 30 mV à V_{DD}=1V, ceci reste inférieur à la valeur courante de tension d'offset dans l'état-of-the-art (50mV). Un amplificateur asymétrique VLSA est alors conçu et optimisé en 28nm FDSOI pour la cellule 10T-XY « single-ended » proposée (figure 4-20 (b)). Cet amplificateur est plus sensible à la variabilité à cause de son architecture déséquilibrée. Ceci explique l'augmentation de la tension d'offset comparée à l'amplificateur différentiel (comme le montre la Figure 4-19). La tension d'offset du SA asymétrique est égale à 100 mV à V_{DD}=280 mV et 60 mV à V_{DD} =1V. Afin de valider la fonctionnalité et pour estimer la consommation d'énergie des deux VSA (différentiel et déséquilibrée), des simulations Monte Carlo sont effectuées pour évaluer le chemin critique (avec divers scénarios de nombre de cellules empilés par colonne: 32, 64, 128, 512, 1024) avec la cellule 10T-XY « single ended » (figure 4-20 (b)) et la cellule 10T-XY différentielle (vpoir la figure 4-20 (a)).

Le tableau 4.3 présente l'évaluation de l'énergie pour les deux amplificateurs de lecture optimisés durant l'opération de lecture. Selon le tableau 4.3(a) l'application du FBB avec une valeur égale à 1V aura pour résultat l'augmentation de la consommation totale d'énergie par 36,15% par rapport au cas où la valeur du FBB est égale à 0V pour l'amplificateur différentiel VLSA et par 9,7% pour le « single ended ». Par conséquence, l'application du FBB permet d'améliorer le temps de réaction du SA (accès temps en lecture) mais en contre partie elle résulte à l'augmentation de la consommation d'énergie statique. Il y a donc un compromis entre la consommation d'énergie et la fréquence de fonctionnement qui doit être pris en compte.

L'équation (4-1) indique que la capacité de la bitline et le courant de lecture sont les deux principaux paramètres qui influent sur le temps de décharge de la bitline afin de développer une tension d'entrée différentielle, ΔV_{BL}, supérieure à la tension d'offset du SA. La capacitance de la bitline augmente avec l'augmentation du nombre de cellules empilées par colonne et le courant de lecture dépend du dimensionnement de la cellule. La figure 4-21 (à droite) présente la fonction de la densité de

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

probabilité (PDF) de la durée de lecture pour divers nombre de cellules 10T-XY empilés par colonne (32, 64, 128 et 256). Cette figure illustre l'augmentation de la variabilité et l'étalement du temps de lecture avec l'augmentation de la capacité de la bitline. Ceci s'explique pour deux raisons: La faiblesse du rapport I_{ON}/I_{OFF} et le l'augmentation du temps de lecture dans le domaine ultra-basse tension (voir tableau 4.2), ce qui conduit au choix de la limitation de nombre de cellule empilées par colonne à 64 dans l'architecture de la mémoire conçu, qui sera présenté dans le chapitre 5. Comme déjà mentionné au début de ce chapitre, le « back-biasing » peut être utilisés dynamiquement pour augmenter ou diminuer la tension de seuil. Une évaluation de l'impact de l'application de la polarisation arrière sur les performances de la cellule est présentée. L'application du FBB au niveau de la cellule diminue la tension de seuil mais augmente largement le courant de lecture ce qui se traduit par la diminution du temps de lecture. Comme le montre la Figure 4-22, l'application d'une valeur FBB = 1V permet de réduire le temps de lecture de 85% par rapport au cas où la valeur du FBB est égale à 0V. Malheureusement, le « FBB » provoque l'augmentation du courant de fuite total de la cellule SRAM. La figure 4-23 montre que l'application d'un FBB = 1.2V provoque l'augmentation du courant de fuite total de la cellule 10T-XY de 80% pour V_{DD} = 300 mV et de 56% pour V_{DD} = 1V par rapport au cas où le FBB=1V. La méthode traditionnelle pour limiter l'énergie consommée due à l'utilisation du FBB est de l'utiliser d'une façon dynamique. Par exemple le FBB sera appliqué uniquement pendant l'opération de lecture. Deuxièmement, il est possible de diviser la mémoire en deux, quatre ou plusieurs blocs de matrice de cellules où le FBB sera appliqué seulement dans le bloc sélectionné. Malheureusement, ces techniques sont limitées car l'énergie consommée reste significatif [81].

Une solution alternative doit être explorée. La Figure 4-24 (b) présente la technique proposée permettant la modulation de la tension de seuil dynamique par colonne, elle consiste à appliquer une polarisation FBB sur le NWELL, seulement pour les cellules dans la colonne sélectionnée. Cette technique permet de réduire considérablement l'énergie consommée par rapport à la technique de l'application du FBB dynamique par bloc (voir la figure 4-24 (a)). La technique proposée s'applique sur le caisson NWELL dans les colonnes sélectionnées; cependant, comme indiqué sur la figure 4-24(c): pour chaque cellule, il existe deux caissons NWELL. Ainsi, les deux colonnes qui sont adjacentes à la colonne sélectionnée seront à moitié-sélectionnées (un caisson NWELL non polarisé et l'autre polarisé à FBB). Une analyse de la marge au bruit statique d'une cellule moitié-sélectionnée (SNM_{HOLD}) a été effectuée. Cette étude confirme qu'il n'y a pas d'impact négatif dans le maintien de données dans les cellules à moitié-sélectionnées. D'autre part, il est essentiel de vérifier le comportement de la stabilité en lecture de la cellule SRAM face l'application du FBB. La Figure 4-25 présente l'évaluation du SNM pour la cellule 10T-XY pour différentes valeurs de FBB et de tensions d'alimentation. La figure montre que l'application de valeur excessive de FBB peut provoquer une

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

dégradation de la marge de bruit statique surtout à ultra-basse tension ce qui va générer des défaillances de lecture. La cellule mémoire 10T-XY souffre de défaillance lorsque le FBB>1V pour V_{DD}=350mV. Ce qui nous oblige à limiter la valeur de FBB à 1 V. La réduction de la tension d'alimentation à pour résultat l'augmentation de la variation de la tension de seuil. Ceci impacte directement les performances de la mémoire SRAM. Un amplificateur de lecture et une répliqua sont considérés comme une meilleure solution pour minimiser les impacts des variations PVT sur les performances. Dans la section précédente, un amplificateur déséquilibré de lecture est conçu ayant un V_{MIN}=280mV permettant d'améliorer le temps de lecture de la cellule asymétrique 10T-XY proposée. Cette section aborde le circuit de répliqua, essentiel pour émuler le chemin critique de l'opération de lecture. Le temps d'accès de la mémoire SRAM est le principal limitant de la fréquence de fonctionnement dans un SoC. Le temps d'accès de la mémoire dépend du courant de lecture de la cellule la plus faible et de la bitline ayant la capacité la plus grande. Afin d'améliorer les performances des SoCs, il est nécessaire d'avoir un temps d'accès rapide pour la SRAM [82]. Le temps de décharge de la capacité de la bitline par la cellule sélectionnée au cours de l'opération de lecture est une partie dominante dans le temps d'accès (Figure 4-9). Un amplificateur de lecture est utilisé afin de réduire le délai dû à la décharge de la bitline en amplifiant une petite différence de voltage développée entre les deux bitlines. Une décharge partielle de la bitline réduit le temps d'accès et la consommation dynamique. Un signal SAEN est nécessaire pour contrôler le moment de l'activation du SA. Comme représenté sur la figure 4-26, si le signal SAEN se déclenche avant que la tension différentielle d'entrée ne dépasse la tension d'offset du SA, $\Delta V_{BL}<V_{Offset}$, le SA amplifie aléatoirement ΔV_{BL} et peut donc causé des défaillances.

Dans le cas ou, le signal SAEN se déclenche trop tard ($\Delta V_{BL}>V_{Offset}$), le temps d'accès et la consommation dynamique augmentent. Il existe donc un moment optimal pour activer l'amplificateur de lecture. Ce moment optimal dépend des variations globale et locale [83]. Le circuit du répliqua est censé générer le signal SAEN dans le moment optimale à chaque condition PVT. Les circuits replica ont été fréquemment utilisé dans la mémoire SRAM pour le contrôle du SA. Ceci afin d'optimiser le temps de lecture en s'adaptant aux variations PVT. Dans cette section, nous allons analyser les différents problèmes liés aux circuits de répliqua et un nouveau circuit de répliqua sera proposé. Le circuit de répliqua a été introduit dans [84]. Plusieurs architectures de répliqua sont apparus dans l'état de l'art [83], [85], [86] pour émuler le chemin critique pour l'opération de lecture et d'écriture. La Figure 4-27 présente un circuit de répliqua classique dans une mémoire SRAM [87]. Les signaux sont comme suit (voir la Figure 4-26): l'adresse d'entrée est décodée et elle provoque l'activation de la ligne de mot correspondant à la cellule sélectionnée, ainsi que l'activation de la ligne réplica « replica WL ». La cellule SRAM sélectionnée commence à décharger la bitline de lecture. Dans le même

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

temps, un nombre fixe de cellules commencent à décharger la répliqua bitline: le signal de décharge générer par la bitline de répliqua est inversé et tamponné pour générer le signal SAEN qui déclenchera l'amplificateur de lecture. Enfin, le SA amplifie la différence de tension ΔV_{BL}, qui doit être supérieure à la tension d'offset. Le signal de répliqua de la bitline est également utilisé pour faire la remise à zéro du signal WL pour arrêter la décharge de la bitline et économiser l'énergie [83].

Le délai de l'opération de lecture est constitué d'une part du délai des portes logique dans le décodeur, le délai RC de la ligne de mots et finalement le délai de décharge de la bitline entraînés par la cellule SRAM sélectionnée. Le délai causé par les variations PVT dans les portes logique est différent de celui dans la bitline. En raison de la différence au niveau des V_{th} des transistors dans les portes logique et les cellules SRAM, le taux de variation du délai est différent en réduisant la tension d'alimentation. Il y a trois sources de délai dans l'opération de lecture: le délai RC dans la répliqua WL, le délai correspondant au décharge de la bitline et le délai pour générer le signal SAEN dans la partie control. La différence entre la variation du délai dans le réplica de celui dans la mémoire SRAM peut causer des défaillances ou une dégradation des performances (un moment de déclenchement de SAEN non optimum). Le délai de décharge de la bitline dans la répliqua causer par le « mismatch » peut être réduite en augmentant le nombre de cellules « conducteurs » dans la colonne de la répliqua [83]. La Figure 4-29 illustre la distribution de probabilité de délai de la bitline et du signal SAEN: un compromis entre la variation temporelle du signal SAEN due au « mismatch » et le délai de la bitline doit être fait pour éviter les échecs de l'opération de lecture et la dégradation des performances. Le circuit réplica conventionnel utilise un seul chemin pour acheminer le signal de déclenchement de réplica pour les opérations de lecture et d'écriture. Cependant, dans les cellules à double port et les cellules avec un port de lecture séparé, le chemin de lecture et d'écriture sont séparés. On se basant sur cette séparation, une nouvelle répliqua est proposée permettant d'éliminer la ligne réplica et les cellules « dummy » dans le décodeur, tout en assurant les mêmes fonctionnalitées que la répliqua standard avec des performances similaire. La figure 4-32 présente le principe de fonctionnement du circuit répliqua proposé. La Figure 4-33 montre les formes des signaux d'horloges de lecture et d'écriture générés par la répliqua proposée et la répliqua standard. La forme des deux signaux d'horloge montre qu'il y a un peu de délai du à l'effet capacitif. La Figure 4-34 illustre le schéma de la cellule standard 8T et la cellule double port 8T. La valeur de la capacité de la ligne de mot dans le chemin de lecture est différente de la valeur de la capacité dans la ligne de mots dans le chemin d'écriture. Cela se traduira par un délai de propagation différent entre les deux voies. Par conséquent, pour tenir compte de ce délai, un délai (1,25ns @VDD=500mV) a été arbitrairement ajouté. L'utilisation du circuit de répliqua proposé permet de réduire la surface de silicium de 10 à 20% par rapport au réplique standard, tout en assurant les mêmes fonctionnalitées et performances. Le temps de

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

réaction de l'amplificateur varie en fonction de la tension d'alimentation. Une largeur d'impulsion optimisée correspond au cas le plus défavorable (worst case) de la durée de réaction SA à 3σ. Dans les mémoires SRAM standard, le bloc de contrôle contient une seule chaîne de délai qui définit la largeur d'impulsion SAEN dans la plage de tension d'alimentation nominale. Toutefois, la variation du temps de réaction du SA et du délai dans la chaîne d'inverseurs ne sont pas identiques à ultra large gamme de tension (sub-V_T & above-V_T) due aux variations PVT. Par conséquent, nous devons optimiser une chaine de délai spécifique pour chaque gamme de tension correspondant au temps de reaction du SA dans cette plage. La cellule « canari » est le circuit qui permet de générer le signal SAEN. La Figure 4-35 présente la technique proposée pour optimiser la largeur de l'impulsion SAEN pour une plage de tension ultra-large. La figure 4-36 illustre le circuit proposé, qui permet la sélection d'une chaine de délai spécifique selon la tension d'alimentation en utilisant un détecteur de tension et un décodeur. Cette technique est basée sur l'introduction de diverses chaînes de délai dans les cellules canari optimisées pour chaque plage de tension. Chaque retard est caractérisé pour fournir une largeur d'impulsion optimisée pour la tension d'alimentation sélectionnée. Dans notre mémoire (chapitre 5), la sélection de la chaîne de délai est effectuée manuellement par le codage des bits dédiés en mode test.

Les circuits fonctionnant dans le domaine de l'ultra-basse tension introduit de nouveaux défis vis-à-vis des circuits de test. Pousser par l'avancement de la technologie et l'augmentation des besoins pour répondre aux applications haute fréquence et basse consommation introduisent de nouveaux défis en matière de test, les concepteurs de circuit intégrés et des cartes de test ont migrés ensemble vers le fonctionnement dans le domaine de l'ultra-basse tension [88]. La réduction de la tension d'alimentation a affaiblie l'immunité face au bruit : cela se traduit par l'augmentation de nombre de contraintes dans les testeurs pour les composants travaillant à très basse tension. Le testeur doit être capable de driver et de recevoir des signaux avec des marges plus petites que le testeur standard [88]. Malgré les techniques utilisées pour améliorer les testeurs pour satisfaire aux besoins des marges dans la plage de tension très basse, jusqu'aujourd'hui, il n'a y pas de testeur industriel capable de tester des circuits fonctionnant en tension de sous le seuil. Ce chapitre décrit un prototype de mémoire SRAM avec un BIST. Le prototype a été conçu avec succès et en cours de fabrication dans la technologie 28nm FDSOI. Un aperçu de la mémoire SRAM SYPHAX est fourni. Une évaluation de l'énergie consommée, de la fonctionnalité et des performances est présentée. Finalement, la méthodologie de test et le démonstrateur sont illustrés. La prétendue SYPHAX vise une mémoire SRAM pour des systèmes fonctionnant dans une plage très large de tension (UWVR), avec deux modes: le premier correspond au fonctionnement à haute performance (la tension d'alimentation est réglée à la valeur élevée) et le second mode correspond au fonctionnement à basse consommation (la tension d'alimentation est réglé sur une valeur de tension ultra-basse). Les applications cibles peuvent être les

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

systèmes de nœuds de capteurs sans fils, les implants biomédicaux et les objets connectés « internet of things ». Le tableau 5.1 présente les caractéristiques de la mémoire et le tableau 5.2 présente les conditions de fonctionnement de SYPHAX. Figure 5-1 montre le dessin de masque de la cellule proposée 10T-XY conçu en respectant les règles de dessin logique (DRC) en technologie 28nm FDSOI, dans la configuration « flip well » (chapitre 4) avec une taille égale à 0,62 μm^2: il y a une possibilité de gagner entre 30 à 40% en terme de surface de la cellule si nous appliquons l'optimisation avec les règle de dessin spécifique aux mémoires SRAM (partage de contacts ...) cela n'a pas été fait pour des contraintes de temps. Le tableau 5.3 présente les caractéristiques de la cellule 10T-XY en termes de marge de stabilité en lecture et en écriture et en termes de la tension d'alimentation minimale V_{MIN}. La faiblesse de la marge de stabilité en écriture à basse température limite la tension d'alimentation minimale à -40 ° C. C'est pourquoi nous avons une meilleure V_{MIN} dans la plage de température [0°C, 125°C]. Une mémoire SRAM fonctionnant dans une plage de tension ultra-large a été conçu contenant: la cellule 10T-XY asymétrique (chapitre 3), une technique de back-biasing » permettant la réduction de temps de lecture, Un amplificateur de lecture asymétrique optimisé avec un circuit de répliqua fonctionnant jusqu'à 300mV (chapitre 4), deux circuits de répliqua ont été implanté: une standard et la répliqua proposé présenté dans le chapitre 4 seront testés et les résultats de silicium seront confrontés aux résultats CAD.

La mémoire SYPHAX est composée de quatre blocs matrice comme le montre la Figure 5-2. Chaque bloc est composé de 64 cellules par ligne et de 128 cellules par colonne. La Figure 5-3 présente l'organisation du bloc matrice. Le bloc de la matrice est divisée en 16 sous-blocs (présentant 16 bits d'entrées / sorties). Huit colonnes composent chaque sous-bloc (MUX 8). Pendant le fonctionnement, l'un des deux blocs matrice est sélectionné (blocs dans le haut ou dans le bas comme le montre la Figure 5-2). Le circuit de répliqua standard est implémenté dans le bloc de matrice situé en bas à droite tandis que la répliqua proposé est implémenté dans le bloc de matrice situé en haut à droite. Un bit programmable est utilisé pour faire le choix entre les deux circuits répliqua. Le fonctionnement de la mémoire SRAM en tension sous le seuil est beaucoup plus sensible au « soft error ». Cela est dû à la tension d'alimentation inférieure dans les nœuds internes de la cellule en ULV [95]. La cellule 10T-XY utilise la technique d'entrelacement de bits « bit-interleaving » dans la structure de la colonne, cela permet le rejet de multiples « soft error » [96]. L'amplificateur de lecture asymétrique conçu pour la cellule 10T-XY est implémenté dans la mémoire SYPHAX comme le montre la figure 5-6. Le SA possède deux entrées: la première correspond à la tension de référence (dans notre cas, V_{DD}) qui est fourni par le RBL sélectionné (pré-chargée à VDD) des blocs non sélectionnés (TOP ou BOT). La seconde entrée correspond à la RBL de la colonne sélectionnée en lecture des blocs sélectionnés. Cette

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

méthode consiste à partager le SA déséquilibré entre les colonnes sélectionnées dans le bloc du haut et celui du bas. Cela réduit le nombre d'amplificateurs nécessaires ce qui permet de la réduction de la surface et la consommation d'énergie de la mémoire.

La Figure 5-8 présente les schématiques des « level shifters » (LS) implémentés dans la CUT SYPHAX à l'entrée et à la sortie afin de fournir une plage de tension ultra-basse. L'ajout des LS est indispensable car les testeurs ne fournissent pas de signaux de tension ultra-basse. Pour assurer une fonctionnalité efficace dans la plage de tension sous le seuil, des bits programmables ont été ajoutés. Le tableau 5.5 présente ces pins: le bit DC <0> est utilisé pour sélectionner l'opération avec ou sans « self-timing », le bit DC<1> sélectionne l'un des deux délais optimisés qui définissent la largeur d'impulsion SAEN pour la gamme de tension d'alimentation nominale et ULV. Les bits DC<2:3> sélectionne le nombre de cellules conducteur dans le circuit répliqua (possibilité de choisir entre 1 à 5 cellules conducteurs). Enfin DC<4> active le « level shifter » qui fournit des valeurs négatives pour les signaux XRWL dans les lignes non sélectionnées. Tous les résultats de simulation dans ce chapitre sont basés sur des simulations « full cut » avec un simulateur rapide.

La Figure 5-10 présente des signaux simulés durant l'opération de lecture/écriture à différentes lignes d'adresse dans la mémoire, en utilisant le net-liste complète de la mémoire extraite à partir du dessin de masque. La Figure 5-11 présente l'énergie totale par cycle et la fréquence de fonctionnement maximale de la mémoire SYPHAX pour différentes valeurs de tension d'alimentation. Ces résultats sont pris en TT et 25 ° C conditions. Le point d'énergie minimale de mémoire SYPHAX est égal à 2pJ à V_{DD}=400 mV avec une fréquence de fonctionnement de 25 MHz. Alors que pour V_{DD}=1.3V, l'énergie par cycle est égal à 15pJ et la fréquence de fonctionnement est égale à 1.5GHz. Il en résulte une réduction de l'énergie totale par cycle d'un facteur x8, entre les domaines à faible tension et de la tension nominale. Il est à noter que le temps de propagation et de la consommation d'énergie des « level shifters » sont prises en compte dans cette caractérisation. De l'autre côté, la consommation d'énergie du circuit générateur de « biasing » n'est pas prise en compte.

Le Tableau 5.7 compare les performances de la proposition mémoire SYPHAX n avec d'autres macros SRAM dans l'état de l'art ciblant les applications ULV dans les technologies CMOS avancées. La figure 5-12(a) représente la tension d'alimentation minimale .vs la technologie utilisée. Le V_{MIN} de la mémoire SYPHAX est égale à 300 mV, qui présente une valeur acceptable par rapport à d'autres designs ULV dans le-state-of-the-art. Le travail dans [27] est un peu meilleur en terme de fréquence de fonctionnement et au niveau de l'énergie consommé à 350 mV. Comme indiqué précédemment la faiblesse de la fréquence de fonctionnement en ULV est due à l'opération d'écriture: une correction doit être faite au futur afin d'améliorer ce point. Le travail dans [27] utilise une technique d'assistance en lecture originale. Cependant cette technique ne permet pas de résoudre les défaillances en dessus

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

de V_{DD}=400mV. SYPHAX est meilleur en termes d'économie d'énergie et de la fréquence de fonctionnement à la tension nominale grâce à l'utilisation de SA déséquilibré et l'utilisation la cellule 10T-XY à très faible courant de fuite (comme le montre la figure 5-12(e)).

Le test des mémoires SRAM dans le domaine de l'ultra-basse tension est l'un des plus importants défis dans la prochaine génération de mémoire faible consommation, ceci est en raison de la faiblesse des testeurs industriels [88] : qui sont incapable de tester des IPs dans le domaine de tension sous le seuil [88]. La méthode de test « DMT » est la technique qui utilise un appareil d'essai externe permettant l'accès aux nœuds internes de la mémoire. Le test est effectué sur les pins d'E/S. Le testeur est responsable de la génération de pattern de test dans la mémoire, et de décoder les signaux de sortie. Le testeur va alors faire la comparaison des données. L'avantage de cette méthode est que l'utilisateur peut facilement changer les paternes de test de l'extérieur. Le test « DMT » est la méthode la plus utilisée pour le test de dispositifs ULV. Cependant, cette méthode est limitée en termes de couverture de défauts. Le built-in self-test (BIST) est le mécanisme de test de mémoire permettant d'avoir diagnostic efficace. Un BIST spécifique a été utilisé pour tester la mémoire SYPHAX en tension sous le seuil et dans la plage de tension nominale. La cellule 10T-XY proposé et les techniques proposées (le chapitre 4) sont implémentées dans le CUT SYPHAX. Cette mémoire est intégrée dans un démonstrateur conçu en technologie 28nm FDSOI. L'objectif de ce démonstrateur est de valider la fonctionnalité de notre mémoire dans une large plage de tension (300mV<V_{DD}<1.3V) et de comparer les résultats en termes de consommation, de la fréquence de fonctionnement et du rendement par rapport aux produits SRAM industriels en technologie 28nm FDSOI. Ce travail cible la conception de mémoire SRAM dans une large plage de tension avec trois objectifs: d'abord éviter les limitations concernant les cellules fonctionnant en ultra-basse tension dans l'état de l'art en termes de consommation statique et fonctionnalité. 2ème, trouver des nouvelles architectures de cellule qui peuvent être utilisées pour concevoir une mémoire SRAM avec une large plage de tension de fonctionnement. Quelles techniques peuvent être utilisées pour assurer la fonctionnalité en tension sous le seuil? Et enfin, comment la mémoire SRAM UWVR conçu peut être testée?

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Contents

RESUME **3**

AUTHOR'S PUBLICATIONS AND PATENTS **8**

RESUME ETENDU (FRANÇAIS) **9**

ABBREVIATIONS **36**

LIST OF FIGURES **37**

LIST OF TABLES **41**

INTRODUCTION **42**

1. TECHNOLOGY LIMITATION AND ALTERNATIVE TECHNOLOGIES **43**

2. DYNAMIC VOLTAGE SCALING **49**

2.1 STATIC ENERGY REDUCTION 49

2.2 DYNAMIC ENERGY REDUCTION 50

3. PREVIOUS ULV DESIGN WORKS **50**

4. THESIS CONTRIBUTION **51**

STATE OF THE ART OF ULV SRAM **52**

1. INTRODUCTION **52**

2. MISMATCH **53**

3. POWER CONSUMPTION IN SRAM BITCELL **54**

4. ENERGY PER OPERATION **55**

5. SRAM METRICS **57**

5.1. 6T SRAM BITCELL 57

5.2. STATIC NOISE MARGIN (SNM) 58

5.3. READ MARGIN (RM) 59

5.4. WRITE MARGIN (WM) 60

5.5. DATA RETENTION VOLTAGE 62

6. SRAM BITCELL OPTIMIZATION **63**

7. FAILURES IN SRAM BITCELL OPERATION **63**

7.1. READABILITY FAILURE 64

7.2. READ ABILITY FAILURE 64

7.3. WRITE ABILITY FAILURE 64

7.4. HOLD FAILURE 64

7.5. PVT IMPACT ON THE SRAM BITCELL OPERATION 64

8. LIMITATION IN THE 6T SRAM BITCELL IN TERM OF VDDMIN **66**

9. STATE-OF-THE-ART IN ULTRA-LOW VOLTAGE SRAM **67**

9.1. WRITE-PREFERRED BITCELL AND READ-ASSIST CIRCUIT 68

9.2. READ-PREFERRED BITCELL AND READ-ASSIST CIRCUITS 69

9.3. PREVIOUS ULTRA-LOW VOLTAGE BITCELL ARCHITECTURES 71

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

10. LIMITATIONS IN ULTRA-LOW VOLTAGE SENSE 74

11. CONCLUSION 75

ULTRA-LOW VOLTAGE BITCELLS 77

1. INTRODUCTION 77

2. CONSTRAINTS AND LIMITATIONS OF AVAILABLE BITCELLS 78

2.1. EFFECTS OF TEMPERATURE AND DOPING ON MOBILITY 78

2.2. I_{ON}-TO-I_{OFF} RATIO 79

2.3. SOFT ERROR DISTURBANCE 80

2.4. DYNAMIC LOSSES 81

3. PROPOSED "10T-MUX" BITCELL 82

4. PROPOSED XY READ/WRITE BITCELL 84

4.1. STANDARD READ-PORT 84

4.2. V_{GND} READ-PORT 86

4.3. PROPOSED XY READ PORT 87

4.4. PROPOSED XY BITCELL 88

5. COMPARISON WITH STATE-OF-THE-ART BITCELLS 89

6. SILICON VS SIMULATION EVALUATION IN 28NM LP BULK 92

7. CONCLUSION 94

SOLUTIONS ENABLING UWVR 95

1. INTRODUCTION 95

2. FDSOI TECHNOLOGY BENEFIT 46

3. READ VS. WRITE OPERATION 96

4. LIMITATION IN READ OPERATION TIMING IN UWVR 98

5. UWVR SMALL-SIGNAL SENSING SCHEME 101

5.1 SENSE AMPLIFIER 101

5.2 OPTIMIZED UWVR VOLTAGE SENSE AMPLIFIER 103

6. DISCHARGE TIME OF THE BITLINE 107

6.1 IMPACT OF THE CAPACITANCE ON THE DISCHARGE TIME: 108

6.2 DYNAMIC MODULATION OF V_{TH} IN 28NM FDSOI TECHNOLOGY 108

7. REPLICA CIRCUIT 111

7.1 STATE-OF-ART 112

7.2 REPLICA CIRCUIT 115

7.3 CONFIGURABLE SA PULSE WIDTH 118

8. CONCLUSION 119

TEST-CHIP AND SIMULATION RESULTS 120

1. ULV 32KB "SYPHAX" SRAM MEMORY 120

1.1 CONTRIBUTION TO DESIGN INNOVATION 122

1.2 DESIGN 122

1.2.1 Memory Floorplan 122

1.2.2 Decoder structure 124

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

1.3 SHARING UNBALANCED VSA 125
1.4 LEVEL SHIFTERS 126
1.5 PROGRAMMABLE PINS 127
1.6 LOGIC BEHAVIOR 127

2. SIMULATIONS RESULTS 128

3. BENCHMARK 131

4. PROBLEMS AND CHALLENGES THAT HAVE BEEN OVERCOME 133

5. PROTOTYPE OF UWVR TEST METHODOLOGY APPLIED TO THE SYPHAX MEMORY 133

6. CONCLUSION 139

CONCLUSION 140

PERSPECTIVES 142

REFERENCE 144

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Abbreviations

SRAM	Static Random Access Memory
CMOS	Complementary metal oxide semi-conductor
ULV	Ultra-low voltage
UWVR	Ultra-wide voltage range
IC	integrated circuit
PU	Pull Up
PD	Pull Down
PG	Pass Gate
BLT	Bit-Line True
BLF	Bit-Line False
RBL	Read Bit-Line
RDF	Random dopant fluctuation
SA	Sense Amplifier
DRT	Data retention voltage
V_{MIN}	minimum supply voltage
SNM	Static Noise Margin
WM	Write Margin
WT	Write time
RT	Read time
I_{read}	Bit-Cell read Current
I_{Leak}	Bit-Cell leakage current
PVT	Process Voltage Temperature
MC	Monte Carlo
V_{th}	Threshold Voltage
DRV	Data Retention Voltage
RBB	Reverse Body Bias
FBB	Forward Body Bias
VLSA	Voltage latch-type sense amplifier
CLSA	Current latch-type sense amplifier
ΔV_{BL}	Bit-Line Voltage difference
MC	Monte Carlo
σ	Standard deviation
μ	Average Value
RYL	Read Operation Yield
MSB	Most Significant Bit
LSB	Least Significant Bit
DRC	Design Rule Check
LVS	Layout Versus Schematic
BIST	Built-in self-test

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

List of figures

Figure 1-1: SRAM memory architecture 43
Figure 1-3: Trends in gate oxide thickness, threshold voltage and power supply voltage VDD, versus channel length CMOS logic technologies [5] 44
Figure 1-4: Active power density and subthreshold leakage power density versus channel length [5] 45
Figure 1-5: Transistor architectures: (a) bulk silicon, (b) FDSOI and (c) FINFET. 45
Figure 1-6: Static energy reduction techniques (a) Stack effect, (b) sleep transistors and (c) body bias effect ... 50
Figure 2-1: Schematic representation of random dopant fluctuation in a NMOSFET 53
Figure 2-2: Simulation of drain current, ID vs. VGS for various W/L values in 28 CMOS LP 54
(1024 Monte-Carlo runs, 25°C, |VGS| = 1V) 54
Figure 2-3: Schematic of charge and discharge operations in a standard CMOS inverter 54
Figure 2-4: Summary of leakage current mechanisms in deep submicrometer transistors 55
Figure 2-5: Energy profiles of the 90nm carry-ahead adder with respect to VDD [32] 56
Figure. 2-6. Measured energy characteristics across wide voltage range IA-32 56
Processor [33] 56
Figure 2-7: The 6T SRAM bitcell 57
Figure 2-8: Seevinck's schematic setup [34] 58
Figure 2-9: Typical voltage transfer curve for read, write and retention mode 59
Figure 2-10: Read operation: read 1 (left) and read 0 (right) 59
Figure 2-11: Butterfly curves during read operation a standard 6T SRAM bitcell in 28 CMOS LP 60
Figure 2-12: Circuit for WSNM when writing '1'. (left), WSNM when writing '1': width of the smallest embedded square at the lower-right side (right). 61
Figure 2-13: Voltage Transfer Characteristic (VTC) of SRAM cell to evaluate write margin 61
by the BL sweeping method 61
Figure 2-14: Circuit for write margin from WL sweeping (left), Write margin (VWL) is defined as the difference between VDD and the WL voltage when nodes Q and QB flip (right). 61
Figure 2-15: Butterfly curves during hold mode for a 6T SRAM bitcell in 28 CMOS LP 62
Figure 2-16: Simulated static power vs. supply voltage for a 6T bitcell (0.12 µm, 28 CMOS bulk) 62
Figure 2-17: Read margin curves of the 6T SRAM cell under PVT variations in 28nm FDSOI technology 65
Figure 2-18: Butterfly curves of two different 6T bitcells with an area : (a) 0.197µm2, (b) 0.120µm2 65
(1024 MC runs) 65
Figure 2-19: WM histograms for two different 6T bitcells, (a) 0.12µm (b) 0.197µm (1024 MC runs) 66
Figure 2-20: SNM (Average -4 σ) versus supply voltage (FS, 125 °C)(VDDMIN, SNM = 613 mV) 66
Figure 2-21: WM (Average -4 σ) versus supply voltage (SF, -40 °C) (VDDMIN, WM= 566 mV) 67
Figure 2-22: Tuning VDD or VSS [40] 68
Figure 2-23: WL voltage drop 69
Figure 2-24: Negative bitline voltage (write 0) 70
Figure 2-25: Word line voltage boost [45] 70
Figure 2-26: Rising V_{SS} or lowering V_{DD} 71
Figure 2-27: Butterfly curves of (a) 6T bitcell in read mode (b) 6T bitcell in hold mode and (c) standard 8T bitcell in read mode (TT, 1V, 25°C: PVT conditions &1024 MC runs) 71
Figure 2-28: Butterfly curves of 6T bitcell (left) and 8T bitcell (right) in read mode with 72
V_{DD} = 1V and 350 mV (TT, 25°C, 1024 MC runs) 72

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Figure 2-29: SRAM bitcell architectures : (a) 8T bitcell, (b) CR8T [49], (c) 7T [47], (d) 8T [50], (e) differential 10T [51], (f) 9T[51], (i) ZIGZAG8T [52] and (k) VGND10T [45]....73
Figure 3-1: (a) The 10T IK JOON bitcell [45], (b) 10TVGND [59] and (c) the zig-zag Z8T [52]....78
Figure 3-2: Monte-Carlo simulation of Iread (1024 runs), (TT Corner Process) (left) at VDD=300 mV and (right) at VDD=1V 79
Figure 3-3: Read current degradation in presence of variation with respect to VDD scaling 80
(Monte-Carlo 1024 runs,TT and 25°C) 80
Figure 3-4: Three different scenarios of soft errors [63] 80
Figure 3-5: Behavior of the 10T bitcell [45]: (a) Read and (b) Write operations....81
Figure 3-6: Proposed 10T-MUX ULV bitcell: (a) schematic, (b) Layout and (c) hard coding technique 82
Figure 3-7: Simulated waveforms of main control signals in operation of the 10T-MUX ULV bitcell (a) and the 10T-XY ULV bitcell (b)....83
Figure 3-8: Schematic for the charge/discharge issue in the proposed 10T-MUX bitcell in matrix configuration 84
Figure 3-9: Read port configurations....84
Figure 3-10: Bitlines behavior in 10T-VGND bitcell in [45] (Figure 3-1(b))....85
Figure 3-11: Bitlines behavior in proposed 10T-XY bitcell at VDD = 380 mV 85
Figure 3-12: comparaison of the read bitlines' behavior in the 10T-VGND bitcell and the differential 10T-XY bitcell respectively (100 MC runs, corner FF, 125°C, 64 cells per column) 86
Figure 3-13: Leakage and read current behaviour with the VGND read port 86
Figure 3-14. Behaviour of currents in the XY read port without underdrive (a)....88
and using under-drive technique (b) 88
Figure 3-15. I_{DS} versus V_{GS} at two different drain voltages for 250 × 40 nm n-channel transistor in a 28nm CMOS process. 88
Figure 3-16: Proposed 10T bitcell with XY read port 89
Figure 3-17: (a) SNM (Average – 3 σ), (b) WM (Average – 3 σ), (c) Write time (Average + 3 σ) and (d) Leakage current (Average + 3 σ) for state-of-the-art ULV bitcells and proposed 10T bitcells 90
Figure 3-18: Experimental Butterfly curve for the 10-XY bitcell for various supply voltages....92
Figure 3-19: SNM measurement versus supply voltage and temperature variation for the 10T-MUX bitcell (a) and the 10T-XY bitcell (b)....92
Figure 3-20: Silicon vs simulation standby current for various supply voltages for the 10T-MUX bitcell (a) and the 10T-XY bitcell (b) 93
Figure 3-21: Read current measurement for various supply voltage and temperature for the 10T-MUX bitcell (a) and the 10T-XY bitcell (b)....93
Figure 4-1: Circuit schematic of conventional domino single-ended sensing (full swing) [71]....96
Figure 4-2: TEM cross section of the hybrid FDSOI/bulk cointegration in a SRAM cut periphery. The BOX thickness is 25nm [73] 46
Figure 4-3: Back biasing concept 47
Figure 4-4 Well layer configurations and Body bias (a) CMOS bulk, (b) Regular-Well [RVT] and (c) Flip-Well [LVT]....48
Figure 4-5: Several VTH flavors available for logic device in each family (e.g. HVT, RVT, LVT)....49
Figure 4-6: Waveforms of main control signals in operation of the 6T bitcell....97
Figure 4-7: Write (left) and read (right) current path in the bitcell using a read port....97
Figure 4-8: Write time WT and Read time RT @3σ for 10T XY bitcell (64L, 128C) in 28FDSOI (TT Process, 25C PVT Conditions) 97
Figure 4-9: Critical path in an SRAM during read operation: (a) standard timing, (b) targeted reduction timing 98

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Figure 4-10: (a) Small signal sensing scheme, (b) large signal sensing scheme with multiplexing in local bitline, (c) large signal sensing scheme with multiplexing in global bitline. [72] 99
Figure 4-11: The simulation setup of a 128-bit x 64-bit SRAM array 100
Figure 4-12: Behavior of bitline voltages during read operation (as schematicaly described in Figure 4-9) 101
Figure 4-13: Latch type SA: (left) voltage sense amplifier [77] and (right) current sense amplifier [78] 102
Figure 4-14: Read current degradation in the 10T-XY bitcell in presence of variation with respect to VDD scaling in 28 FDSOI (Monte-Carlo 1024run,TT and 25°C) 102
Figure 4-15: Schematic of the Unbalanced VLSA 103
Figure 4-16: Waveforms of internal nodes of the 28 FDSOI differential VSA in Figure 4-12(left) @280mV (left) @1V supply voltage (right) (corner SS, -40°C process and temperature conditions) 103
Figure 4-17. Pulse width of differential VSA in Figure 4-14 (left) (a) @280mV and (b) @1V supply voltage 104
Figure 4-18: Probability of failure for the differential VSA versus effective ΔV at VDD (a) 280mV and (b) 1V power supply (1024 MC Runs) 105
Figure 4-19 Probability of failure for the unbalanced VSA versus ΔV at (a) 280mV (b) 1.2V power supply ... 106
(1024 MC Runs) 106
Figure 4-20: schematic of the differential 10T-XY bitcell (a) and the single ended 10T-XY bitcell (b) 106
Figure 4-21. Probability density function of the read time for various scenarios depending on the number of cells per column (Monte-Carlo 1024 runs,VDD= 300mV, TT and 25°C) 108
Figure 4-22. Probability density function of the read time for various voltage value of body bias 109
(10T-XY bitcell, Monte-Carlo 1024 runs, VDD= 350mV, SS and -40°C, corner conditions) 109
Figure 4-23: Simulation of the total leakage current in the 10T-XY bitcell at 300mV and 1V supply voltage respectively in the case of two body bias values (0V and 1.2V) (Monte-Carlo 1024 runs, FF and 125°C corner conditions) 109
Figure 4-24: (a) FBB per block, (b) Proposed dynamic threshold voltage modulation per column and (c) layout view of NWELL-PWELL intersection 110
Figure 4-25. Static noise margin fo the 10-TXY bitcell with various value of FBB in the case of 111
0.35V, 0.5V and 1V supply voltage respectively(-40C_FS_MC1024) 111
Figure 4-26: Scenario of SA activation with SEAN signal 111
Figure 4-27: Conventional replica circuit 112
Figure 4-28: Increase of the access time due to the increase in the SAEN variation 113
Figure 4-29 Probability distribution of BL and SAEN delay [83] 114
Figure 4-30: Conventional timing replica circuit and SEA timing waveform [82] 114
Figure 4-31 (a) Conventional RBL replica with 3 fixed driver cells, (b) Conventional RBL replica with 5 potential driver cells, (c) Probability distribution of the replica BL delay [83] 115
Figure 4-32 Proposed replica circuit 116
Figure 4-33 Waveforms for the read and write clocks generated by the proposed replica (top) and the standard replica (bottom) (TT, 500mV, 25°C PVT conditions, FBB=0V) 117
Figure 4-34. Capacitance modeling in the read and write path in the standard 8T bitcel (left) and dual port 8T bitcell (rigth) 117
Figure 4-35: Proposed replica circuit with adapted delay 118
Figure 4-36. Proposed adaptive sensing time technique 118
Figure 5-1. Schematic and layout of the proposed 10T-XY bitcell (0.62 μm^2) 121
Figure 5-2. Architecture of 32kbit SYPHAX memory 123
Figure 5-3. Matrix block organization 124
Figure 5-4. Address bit organization 124
Figure 5-5. X-Decoder structure 125
Figure 5-6. Unbalanced voltage sense amplifier share 126
Figure 5-7. Simulation of the SA internal nodes in the full memory 126

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

(extracted CUT, TT, 500 mV, 25°C: PVT conditions) 126
Figure 5-8. Schematic of the Level shifters (a) High-to-Low and (b) Low-to-High 127
Figure 5-9. SYPHAX CUT Layout (XY=487.912µm x 113.2µm) 128
Figure 5-10. Simulation of the read and the write operation for the full memory cut in full swing read mode using xa simulator (350mV, TT, 25 °C PVT conditions) 129
Figure 5-11. Energy per cycle and operating frequency profile of SYPHAX CUT versus supply voltage) 130
Figure 5-12. Comparison of the SYPHAX memory with the state-of-the-art cuts in terms of VDD, MIN (a), access energy per bit (b) and maximum operating frequency at ULV (c) and nominal supply voltage (d) 132
Figure 5- 13. View of our demonstrator 134
Figure 5-14. Layout of our demonstrator 135
Figure 5-15. Finite State Machine (FSM) of the proposed BIST 135
Figure 5-16. Timing chronogram simulation of our demonstrator in Scan-in state (S1) 136
Figure 5-17. Timing chronogram simulation of our demonstrator in configuration state (S2) 136
Figure 5-19. Timing chronogram simulation of our demonstrator in check state (S7) 137
Figure 5-20. Timing chronogram simulation of our demonstrator in Scan-out state (state 8) 138
Figure 5-21. TESTCHIP 138
Figure 5-22. Test equipement 139

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

List of tables

TABLE 2.1 SUMMARY OF DIFFERENT SRAM BITCELL TOPOLOGIES 74
TABLE 3.1 SUB THRESHOLD CURRENT EVALUATION VERSUS VGS AND VDS VALUES 87
TABLE 3.2 CONFIGURATION OF CONTROLS SIGNALS IN THE 10T-XY BITCELL 89
TABLE 3.3 SUMMARY OF DIFFERENT SRAM BITCELL TOPOLOGIES 91
TABLE 4.1 EVOLUTION OF THE DISCHARGING TIME FOR DIFFERENTIAL VOLTAGE SENSE AMPLIFIER 100
TABLE 4.2 EVOLUTION OF THE READ TIME FOR LARGE AND SMALL SIGNAL SENSING SCHEMES 100
TABLE 4.3 TOTAL ENERGY CONSUMPTION EVALUATION OF OPTIMIZED SAS 107
TABLE 5.1 SYPHAX MEMORY SPECIFICATIONS 120
TABLE 5.2 OPERATING CONDITIONS 121
TABLE 5.3 PERFORMANCES OF THE 10T-XY-BITCELL AT 4*SIGMA IN 28 FDSOI 121
TABLE 5.4 EMERGING IMPLANTABLE BIOMEDICAL DEVICES 122
TABLE 5.5 DEBUG PINS 127
TABLE 5.6 TRUTH TABLE 128
TABLE 5.7 COMPARISON WITH OTHER STATE-OF-THE-ART MEMORY CUTS 131

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Chapter 1

Introduction

Since 1960, date of the first MOSFET circuit designed by D.Khang and M.Atalla, [1] CMOS technology scaling is going ever thinner and follows Gordon Moore‘s famous observation law (1965) [2]. This law states that the number of transistors per integrated circuit (IC) doubles every year and that processor speeds, or overall processing power for computers, doubles every 18 months. After half a century, this law is still relevant and performances of ICs are significantly better. This explains the increase in the operating frequency and the complexity of embedded functions into chips. Unfortunately for every new process technology the threshold voltage (V_{TH}) and the gate oxides become smaller. This causes increase in the static power consumption due to leakage currents. Besides the evolution of portable electronic systems has been very fast. These systems have become an essential part of human daily life. These devices range from mobile phones, which contain a lot of Gadgets, operate at high frequency and consume increasingly, to medical systems (e.g pacemaker) and wireless sensor nodes that require that their operation time between two charges is large enough. These situations raise more challenges, which must be addressed, mainly at the levels of silicon area, performances and energy consumption. The battery capacity has significantly improved in the last few years, however this improvement is still not enough to satisfy the low power applications energy budget requirement. Energy harvesting systems add more constrains at the power consumption level since the amount of harvested energy from the environment is very limited and uncertain. This requires that the circuits used in a lot of systems must be very energy efficient. System-on-Chips (SoC) contain many complex functions, what increases the request in term of battery energy budget. Particularly the Static Random Access memory (SRAM) is an indispensable part in SoCs and takes a big part in terms of silicon area and total power consumption. Designing of a low power SRAM is today mandatory but a challenging task and a large effort is made in this way for few years.

SRAM memories are mixed-signal circuits containing bitcells, sense amplifier, digital logic, read/write assist circuits and timing generators. There are several SRAM topologies depending on their purpose. Nevertheless, SRAM architectures can be divided into four main blocks: bitcells array, Row & Column decoder, Timing control block, replica tracking circuit and IOs block.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

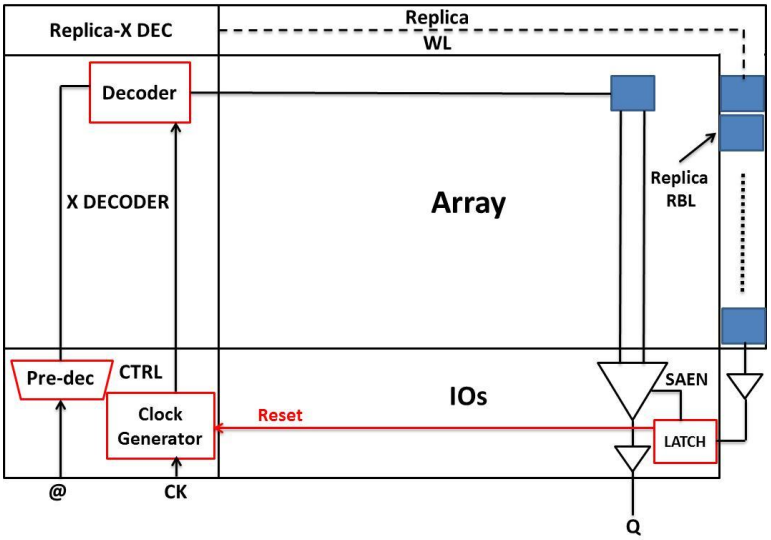


Figure 1-1: SRAM memory architecture

Figure 1-1 shows the basic architecture of SRAM memory. The input address is decoded and leads to the assertion of the word line corresponding to the selected bitcell and the assertion of the replica wordline, assisting read or write operation.

1. Technology limitation and alternative technologies

Figure 1-2 illustrates the typical structure of the MOSFET. The main parameters that characterize the CMOS transistors are: the oxide capacitance, C_{ox}; the gate-source voltage, V_{GS}; the drain-source voltage, V_{DS} and the Drain-source current I_{DS} and the threshold voltage V_{TH}.

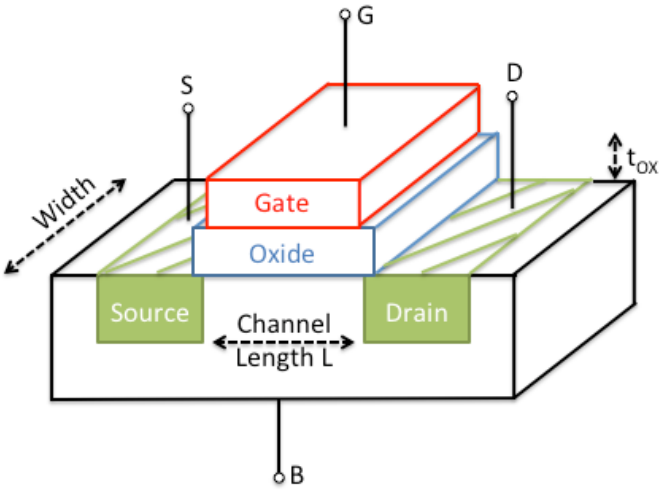


Figure 1-2: Typical structure of a Mosfet

Scaling the physical size in CMOS transistors is approaching its boundary and it is expected to reach limitations in the 22nm technology by 2018 node according to the ITRS [3]. The main challenges and limitations that could prevent CMOS to be used in the future are [4]:

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

- Physical limitations due to the increase in tunneling and leakage currents as the devices are becoming smaller, thus an impact on the performances and functionality of CMOS devices.
- Technological challenges: these involve the inability of lithography-based techniques to provide the resolution below the wavelength of visible light to manufacture CMOS devices.
- Economical challenges: the significant massive increase in the cost of production, fabrication and testing makes the investment in a new technology quite unaffordable

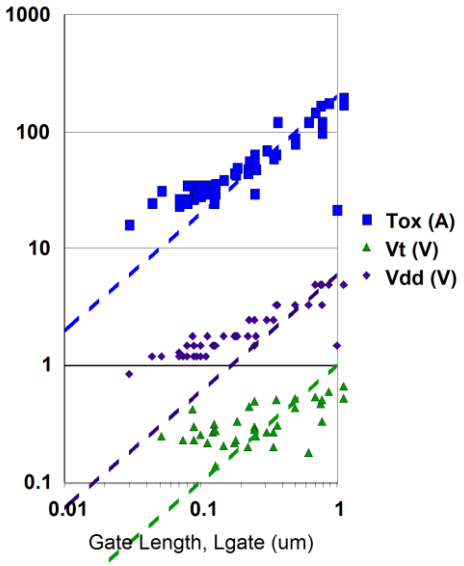


Figure 1-3: Trends in gate oxide thickness, threshold voltage and power supply voltage VDD, versus channel length CMOS logic technologies [5]

The trends in the gate oxide thickness, threshold voltage and supply voltage versus channel length are presented in Figure 1-3. T_{ox} stopped scaling due to the atomic limitation (there is only few atomic layers). In the same way the threshold voltage V_{TH} stops to scale due to the increase in the leakage current: $I_{LEAK} \sim exp(-V_{TH} \times q/nKT)$. Standby power limitation bounds the threshold voltage to a minimum value around 200mV at normal temperature [6]. Supply voltage scaling has stopped as well since kT/q doesn't scale and the threshold voltage scaling causes additional power penalty. On the other side if the supply voltage stops to scale, the energy will scale weakly. Finally the V_{DD}/V_{TH} ratio defines the gate speed.

Figure 1-4 presents the active power density and the subthreshold leakage power density trends versus the channel length calculated from industry trends [5] for a junction temperature at 25°C in a standard transistor. Empirical extrapolations indicate that subthreshold power will equal the active power at gate length equal to 20nm. Power starts to be a dominant issue in advanced technology nodes. Thus the power per chip continues to increase.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

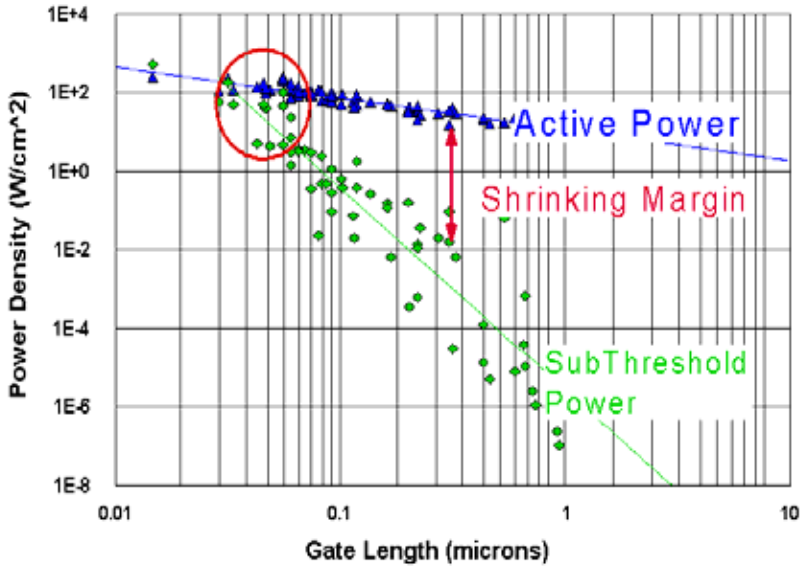


Figure 1-4: Active power density and subthreshold leakage power density versus channel length [5]

Advanced technologies impose new limitations on the circuit designer. Leakage power is a major constraint for low-power battery-operated applications. A lot of innovative low power design techniques have been proposed in the last decade in order to satisfy this request. A complementary process technology is required to enable the functionality at ultra-low voltage supply range: the search for devices with the best possible performances while maintaining extremely low power consumption [7].

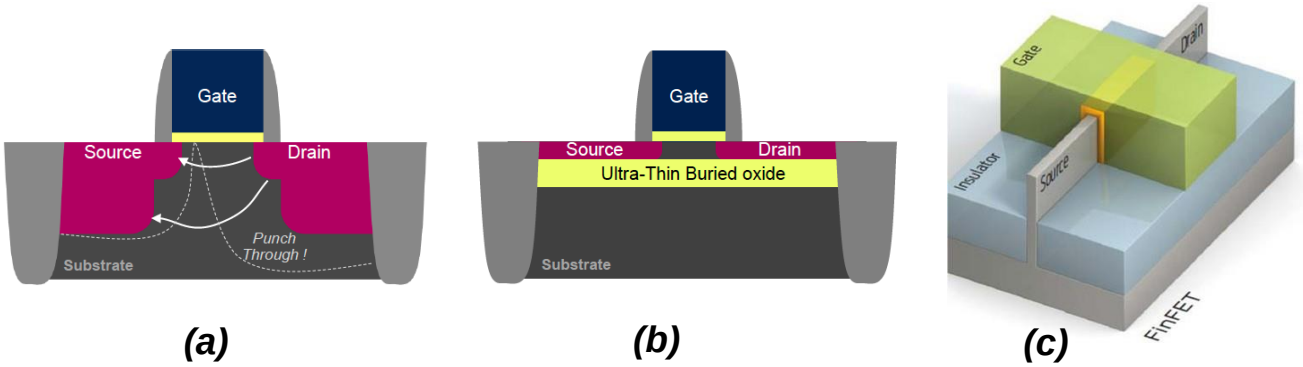


Figure 1-5: Transistor architectures: (a) bulk silicon, (b) FDSOI and (c) FINFET.

Figure 1-5 illustrates a schematic comparison of bulk, FDSOI and FINFET transistors. FDSOI and FINFET technologies appear as an alternative to overcome the limitations in the bulk technology. These two technologies are the only ones able to scale down beyond 28nm. The benefits of FDSOI technology versus bulk silicon technology have been presented in [8], [9], [10]. Planar Fully-Depleted SOI (FDSOI) demonstrates around 30% more performances compared to bulk technology at 28nm node [11]. On the other side FinFET technology has been recently promoted in order to gain on cost and compatibility with CMOS technology. The 22nm node is the first commercially available FINFET bulk technology, which opens a new era of 3D CMOS for low power applications. The FINFET

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

technology offers a new pathway for Moore's Law beyond 20nm since it features better performances at a given power consumption compared to planar CMOS technology. A 16nm/14nm FinFET process can potentially offer a 40-50% performance increase and a 50% power reduction compared to a 28nm process [12]. A comparison between FINFET and FDSOI is illustrated in [13]. Alternatively the 28nm FDSOI technology is part of the context of the work presented in this manuscript. Benefits of FDSOI technology will be summarized in the following.

2. FDSOI technology benefit

Planar CMOS bulk technology is facing a lot of challenges to meet the requirements of 28nm node. Two main limitations are the transistor variability and electrostatics [14]. Ultra-Thin Body and BOX, Fully Depleted SOI (UTBB FDSOI) technology appears as an alternative for future integrated circuits with less variability than in bulk counterparts [14]. The very thin body used in FDSOI transistor ensures that all electrical paths between source and drain are very close to the gate (Figure 4-2). This allows an excellent electrostatic control over the channel and results in an improvement of the sub-threshold slope, Drain-induced barrier lowering DIBL[1] and other short channel effects (SCE) [15].

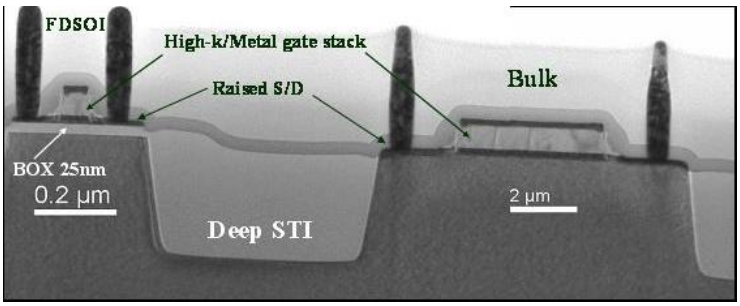


Figure 4-2: TEM cross section of the hybrid FDSOI/bulk cointegration in a SRAM cut periphery. The BOX thickness is 25nm [16]

In addition, the planar FD technology does not demand doping or pocket implants in the channel to control the transistor behavior like tuning the threshold voltages [15] [17]. Therefore, the major issue of random dopant fluctuation mostly disappears. The absence of doping also helps reaching good performances since high channel doping induces reduced carrier mobility.

[1] DIBL: is a short-channel effect in MOSFETs referring originally to a reduction of threshold voltage of the transistor at higher drain voltages.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

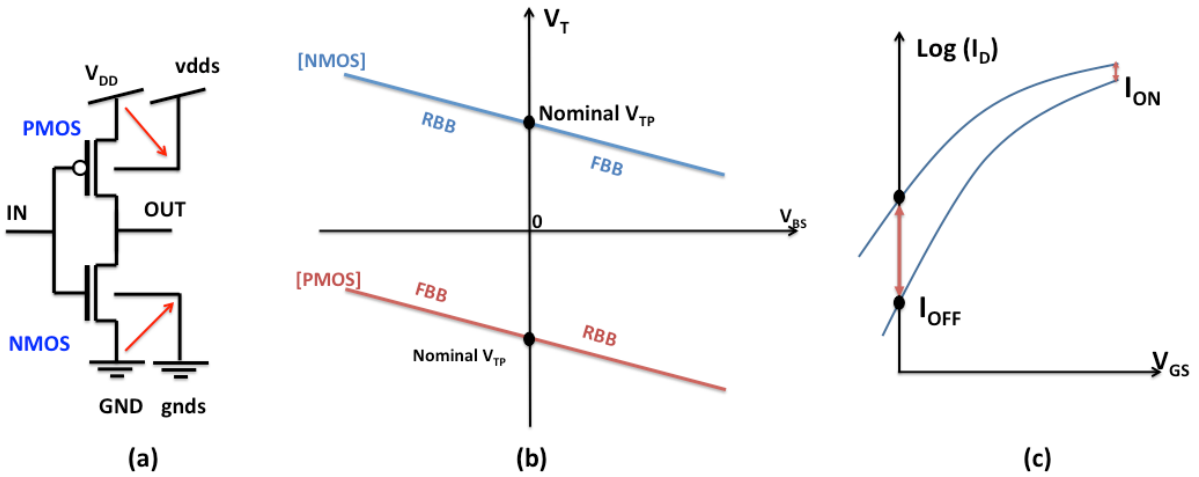


Figure 4-3: Back biasing concept

Back biasing consists of applying a voltage just under the BOX of target transistors [17]. This resulted in changing of the electrostatic control of the transistors and shifts their threshold voltage. This allows having more drive current (higher performance) at the expense of increased leakage current (more static consumption) as shown in Figure 4-3(c).

Applying back biasing results in changing the electrostatic control of the channel and to increase or decrease the threshold voltage. In the case of Forward Body Bias (FBB), V_{th} decreases what results in the increase in the driving current (improving of the operating frequency) but the increase in the leakage current (more static power consumption penalty). In the case of Reverse Body Bias (RBB), V_{th} increases what causes the degradation of the driving current (degradation of the frequency) but the decrease in the leakage current (reduction of the static power consumption). As shown in Figure 4-3 For PMOS, RBB means $V_{BS} > 0V$ and FBB means $V_{BS} < 0V$ while for NMOS, RBB means $V_{BS} < 0V$ and FBB means $V_{BS} > 0V$. In order to limit the power consumption penalty due to the use of back basing, it can be considered with a dynamic application and block-by-block application (only for selected blocks). Body biasing can be applied during a limited period of time to improve performances when the static power consumption is not a main limitation. Or body biasing can be used to reduce static power consumption when improving performances is not a priority. Hence back-biasing presents a new control button allowing the designer to make a trade-off between speed and power consumption according to specifications and time during operation. Figure 4-4 illustrates the Well layer configurations and the authorized range of body bias voltage for CMOS bulk technology, regular well (RVT) and flip-well (LVT) for FDSOI technology.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

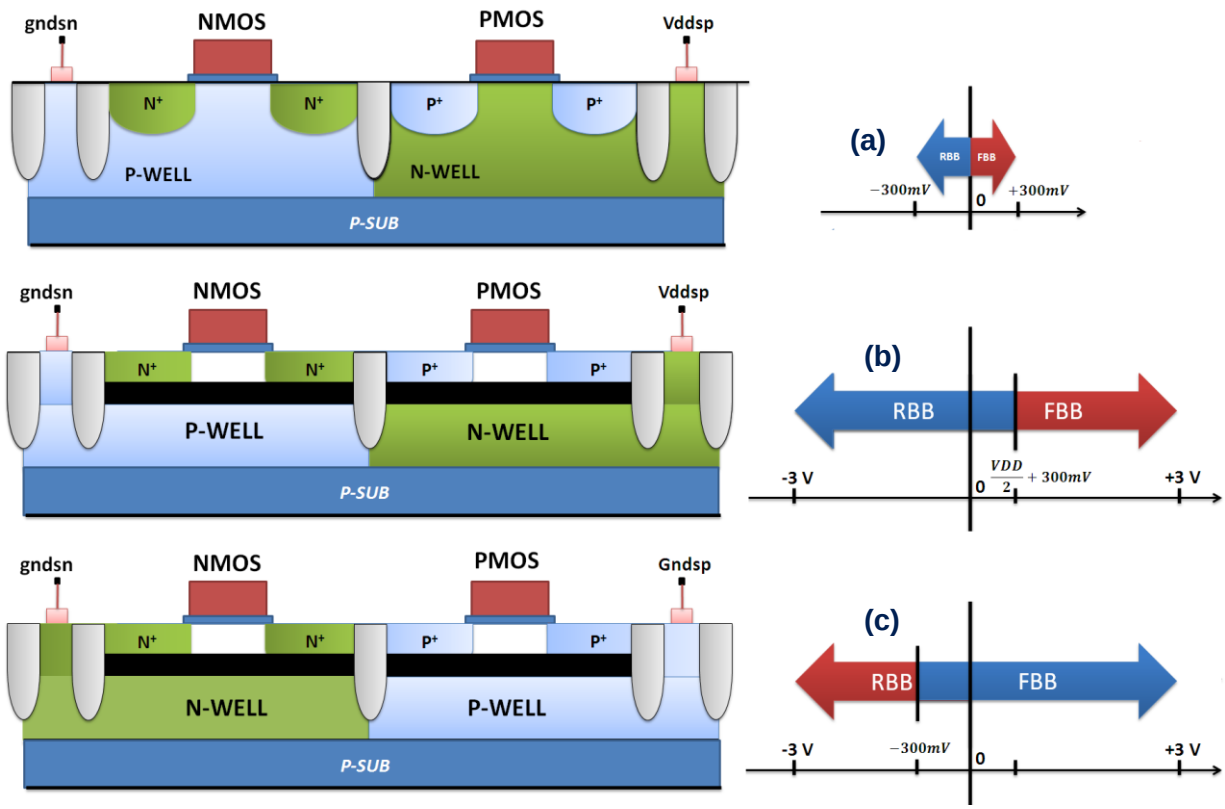


Figure 4-4 Well layer configurations and Body bias (a) CMOS bulk, (b) Regular-Well [RVT] and (c) Flip-Well [LVT]

As shown in Figure 4-4(a) in the bulk technology, the body bias range is limited to -300mV in RBB due to GIDL while FBB is limited to +300mV due to source-drain-well junction leakage and a rise in latch-up probability at higher voltage and temperature. In Figure 4-4(b) and (c), FDSOI technology allows using an extended body bias range. In regular well configuration, RBB extends from -3V to +300mV and FBB extends from +300 mV to +3V. In Flip-well configuration, RBB, extends from -3V to -300 mV and FBB extends from -300 mV to +300 mV [17]. This provides designers with another space to improve performances and energy efficiency of circuits. The high-V_{TH} (HVT) transistors help reduce the static energy consumption at the cost of reduced operating frequency. The low-V_{TH} (LVT) transistors help increase the frequency at the cost of higher static consumption.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

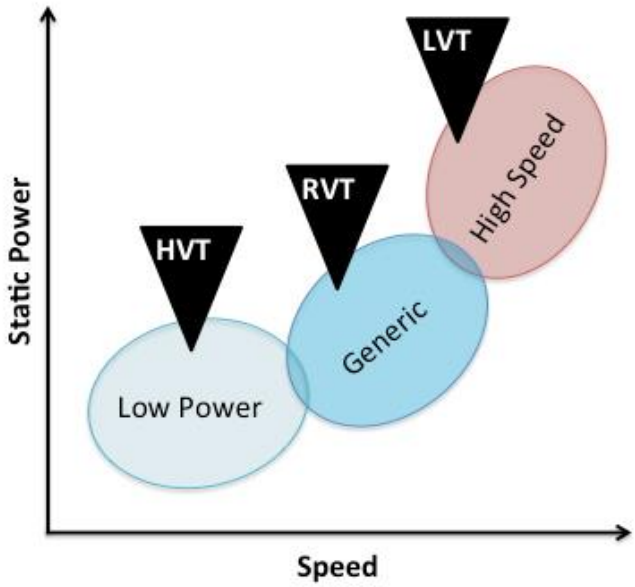


Figure 4-5: Several VTH flavors available for logic device in each family (e.g. HVT, RVT, LVT)

Figure 4-5 presents solutions, which allow adapting the behavior of transistors according to the specifications: LVT transistors improve the operating frequency and HVT transistors improve the energy efficient.

2. Dynamic Voltage Scaling

Various techniques have been proposed in the state of the art to reduce the total power consumption for ICs in sub-100nm technologies.

2.1 Static energy reduction

In order to reduce the growth of the leakage current in the advanced technologies nodes, many techniques have been developed. Researches have focused on two main axes: the modulation of the threshold voltage in the transistor and the power management at the IC level. Two main ways can be used to modulate the threshold voltage in the transistor. The first one consists in the adjustment of V_{TH} by the modulation of oxide gate thickness, the channel length or the channel doping [18]. Many semiconductor companies use three types of transistors (Low VT, Regular VT and High VT) in their technology platforms. The threshold voltage level characterizes these transistors: in the digital circuits LVT transistors are used to improve the frequency while the RVT and HVT transistors are used to reduce the leakage current and to minimize the static power consumption. The key techniques are to use the low-*V*t device to gain performance and high-*V*t devices to cut off the leakage paths [19] [20]. The second method to modulate the threshold voltage consists in back biasing the transistors [21]. In Reverse Body Bias (RBB), the PMOS transistor body, initially connected to V_{DD}, is supplied with

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

higher voltage. The NMOS transistor body, initially connected to GND, is supplied with a lower voltage. The initial power voltage (V_{DD} and GND) is kept unchanged. Figure 1-6 presents techniques to reduce the static energy. The stacked transistor allows reducing leakage current by 2-10x factors. The adjustment of V_{TH} by body biasing results in reduction of leakage current by 2-1000x and finally the use of sleep transistors results in reduction of leakage current by 5-10x.

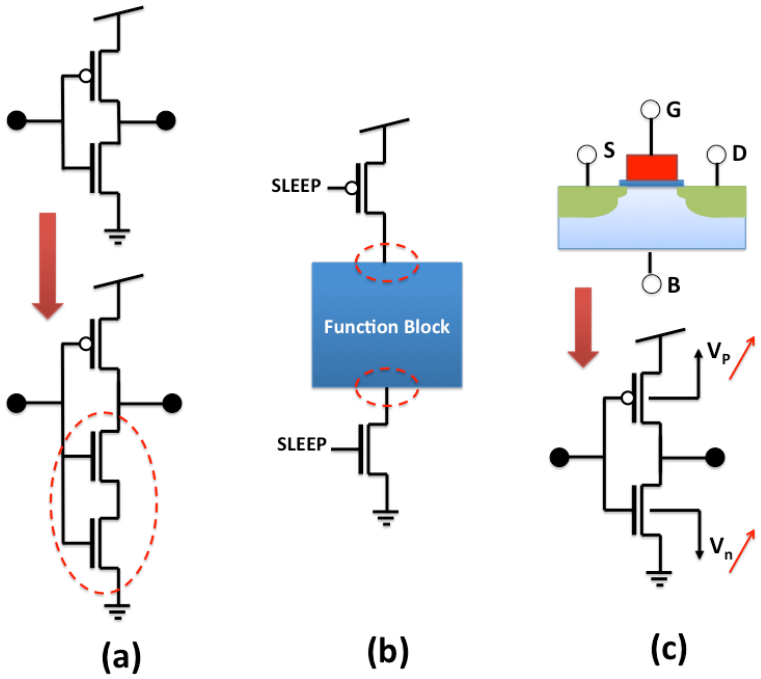


Figure 1-6: Static energy reduction techniques (a) Stack effect, (b) sleep transistors and (c) body bias effect

2.2 Dynamic energy reduction

Lowering the supply voltage is the most effective way to reduce the energy consumption. Scaling down of the supply voltage results in a quadratic saving of dynamic energy and a linear saving of static energy. The minimum functional supply voltage of the SoC is limited by the minimum V_{DD} value (V_{MIN}) of the SRAM for two reasons. First along with scaling down V_{DD}, the delay in SRAM increases in a larger manner than the CMOS logic delay and second the degradation of the SRAM bitcell at low voltage results in functional failures. Hence the design of SRAM operating at ultra-low voltage is of utmost importance to develop an Ultra-Low Power system.

3. Previous ULV design works

The study of operating digital circuits in near and subthreshold supply voltage has been addressed firstly in [22]. Recently several processor designs operating in sub-V_{TH} supply voltage range are made to meet the requirement of Internet of Things, wireless sensor nodes and biomedical applications. In [23], a sub-V_T processor has been designed in 130nm technology. This processor consumes 11nW @ V_{DD}=160mV and 3.5pJ/inst @ V_{DD}=350mV. An ultra low voltage FFT processor operating down to

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

180mV is presented in [24]. However, SRAM is the critical part which limits the energy efficiency of these processors since $VDD_{MIN, SRAM} > VDD_{MIN, LOGIC}$. [25] states that the optimum energy point of the SRAM memories is located in the subthreshold supply voltage range. Hence ultra-low voltage presents an interesting research axis for building ultra-low power SRAMs. A lot of previous ULV SRAM designs in sub-100nm technologies achieve energy saving. However most of these research works can be organized into two categories. The first one involves the SRAM operating only in above V_{TH} supply voltage range ($V_{DD} > V_{TH}$) like in [26], [27]. The second category concerns SRAM operating only in the sub-V_T voltage range ($V_{DD} < V_{TH}$) like in [28] and [29]. Most of these designs targeted only sub-V_T or above-V_T supply voltage operating range. However there is not enough works that address ultra-wide voltage range applications like in [30], [31] and [28]. To design this kind of memories a new techniques (at process and design levels) must be used and explored to ensure an efficient functionality down to sub-V_T and above-V_T supply voltage range. This thesis addresses this research axis and proposes new techniques to enable the UWVR functionality.

4. Thesis contribution

This thesis focuses on UWVR SRAM Memory design. Many points and research axes have been explored and some innovative techniques have been proposed in literature. Two research axes exist in order to design an ultra-low power SRAM memory: the first one consists in using the standard 6T bitcell but with read and write assist techniques allowing for operation at lower V_{DDMIN}; the second one consists in looking for another bitcell architecture that is able to operate in subthreshold domain without the need for any assist techniques. This is the research axis focused in this thesis. Chapter 2 reviews the state-of-the-art of the near and subthreshold SRAMs and presents the various techniques used in order to reduce the power consumption of memory in advanced technology nodes. Chapter 3 presents a comparative study of performances of various architectures of Ultra-Low Voltage bitcells in the state-of-the-art in 32nm and 28nm CMOS Bulk technology with the proposal of two original 10T ultra-low leakage bitcells. The different options and solutions allowing the efficient functionality under ULV and improving performances are described in Chapter 4. These proposed techniques are built on a test-chip for an Ultra-wide voltage range, 32kb SRAM, described in Chapter 5. Finally conclusion in Chapter 6 presents the perspective of proposed prototype and near future prospectives

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Chapter 2

State of the art of ULV SRAM

Introduction

Recently a lot of research works focused on energy efficiency in SoC. The increase in the number of cores per processor on a single die requires an optimization and new solutions must be used in order to improve the energy consumption issue [29]. The growing demand related to mobile electronic systems in terms of energy budget – related to the slow improvement of the batteries' capacity that fails to satisfy this need - requires that Ultra-power design techniques must be applied to enable longer battery life. Scaling down the supply voltage is a de-facto approach used to decrease the total power consumption. Static Random Access Memory (SRAM) in a System-on-a-Chip comprises 20% to 80% of the total chip transistor count in average. It is thus one of the significant sources of energy consumption in the SoC operation. Scaling down V_{DD} is unfortunately accompanied by the degradation of bitcell stability criteria what limits the ULV operation feasibility. The state of the art of SRAM is described in the perspective of limitation: the increase in process variability with scaling, the SRAM bitcell performance limitations, alternative solutions to enable an efficient operation of a SRAM bitcell in near and subthreshold range and the limitations of ULV bitcell in term of operating frequency.

An analysis of mismatch is presented in Section II. Section III illustrates the various sources of power consumption in SRAM bitcell: the benefit of scaling down the supply voltage V_{DD} at the level of static and dynamic power consumption is analyzed. The main metrics allowing the sizing and the characterization of SRAM bitcells are detailed in Section IV. In section V the common techniques to optimize the SRAM bitcell in read and write operations are presented. The various sources of operating failures in SRAM bitcells are illustrated in Section VI. Main limitations of the standard 6T bitcell are discussed in Section VII: degradation of write and read margins at ULV. An analysis and discussion of the state-of-the-art of ultra-low voltage SRAM are presented in section VIII. Finally the limitations of sense amplifier in sub-VT range are discussed in Section XI.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Mismatch

Technology scaling is limited by mismatch that affects directly the ICs' yield; this limitation becomes more significant in advanced technologies nodes and causes many financial losses for semiconductor companies. Mismatch is one of the most critical phenomena in mixed-signal design circuits and especially in SRAMs. In 1989, Marcel Pelgrom [30] proposed the famous modeling law for mismatch: according to this law the mismatch between two equal transistors of dimensions W x L separated by a distance D is presented as follows:

Random mismatch → **Systematic mismatch** ←

$$\sigma_{V_T} = \frac{A_{V_T}}{\sqrt{W \times L}} + S_{V_T}^2 D^2$$
$$\frac{\sigma_\beta}{\beta} = \frac{A_\beta}{\sqrt{W \times L}} + S_\beta^2 D^2 \qquad (2\text{-}1)$$
$$\frac{\sigma_\gamma}{\gamma} = \frac{A_\gamma}{\sqrt{W \times L}} + S_\gamma^2 D^2$$

Where $\beta = \mu C_{ox} \frac{W}{L}$ is the current factor, a combination of four variables parameters where μ is the channel mobility and C_{ox} is the gate oxide capacitance. V_{Th} is the threshold voltage $V_T = V_{T0}$

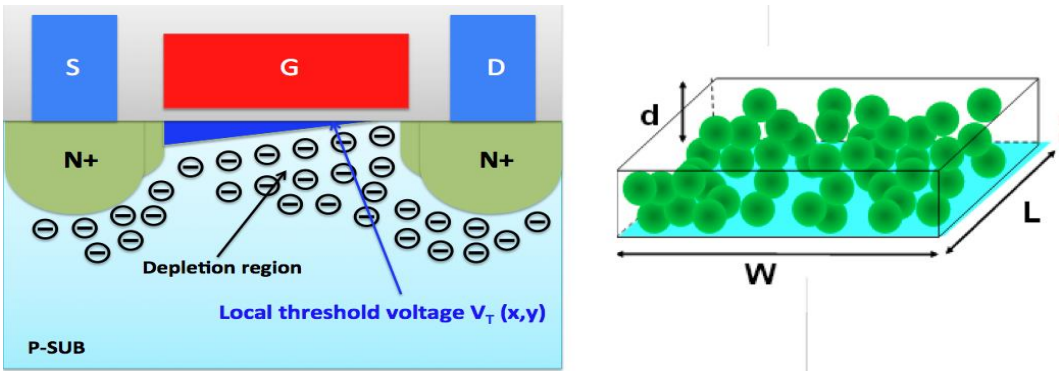


Figure 2-1: Schematic representation of random dopant fluctuation in a NMOSFET

The main source of variation in CMOS devices is the Random dopant fluctuation (RDF): The variation of number of dopants in the depletion region (Figure 2-1). In general the number of dopants in a volume is controlled by the average implementation dose, Na. This number is variable from one volume to another due to the random processes. Pelgrom's law indicates that in order to get more and more accuracy, the area should be large enough what increases the capacitive load and causes the increase in the power consumption. Hence a trade-off must be settled between area and accuracy. Figure 2-2 presents the simulated drain current behavior for various W/L ratios in NMOS and PMOS transistors in 28 CMOS LP. The results confirm the impact of area in the mismatch.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

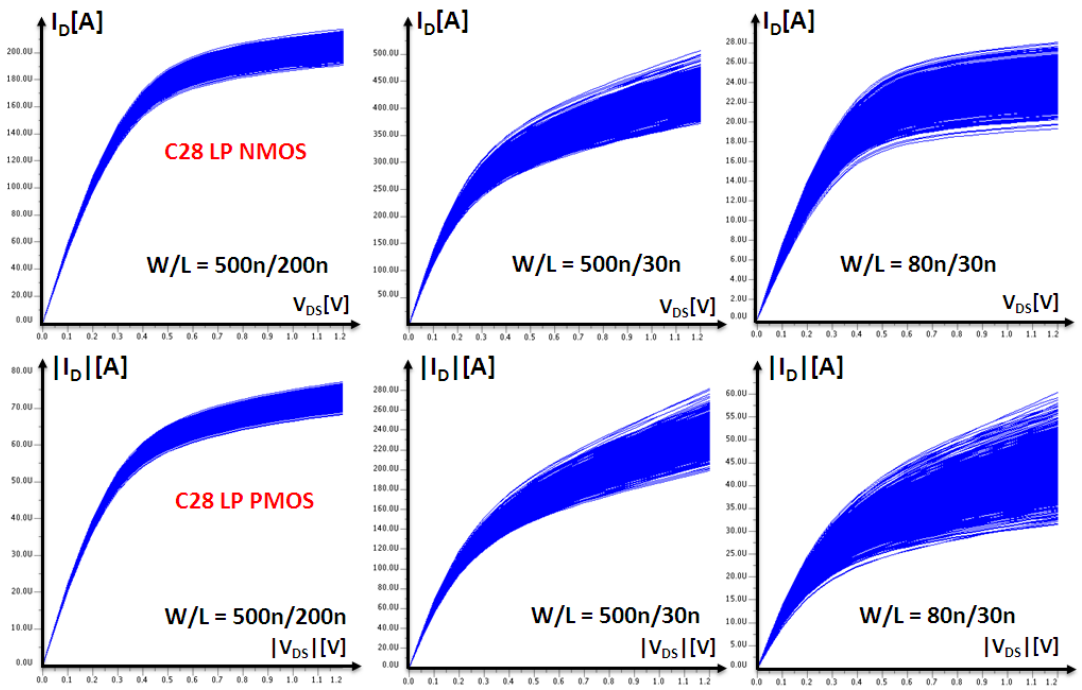


Figure 2-2: Simulation of drain current, ID vs. VGS for various W/L values in 28 CMOS LP (1024 Monte-Carlo runs, 25°C, |VGS| = 1V)

Mismatch is an important parameter, which impacts directly the yield and the performances of the SRAMs. This is why memory designers must give a considerable attention in order to built an accurate memory.

Power consumption in SRAM bitcell

The power consumption in SRAM includes two contributions: the dynamic power consumption and the static power consumption. The dynamic power consumption results in the charge and discharge of capacitances during read and write operations and the static power is consumed due to the leakage current.

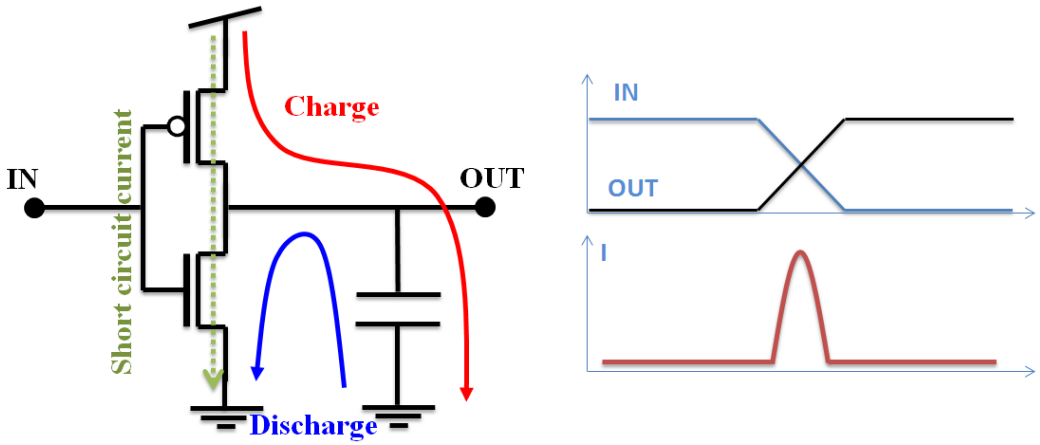


Figure 2-3: Schematic of charge and discharge operations in a standard CMOS inverter

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

As shown in Figure 2-3 the dynamic power consumption results in the switching and short-circuit power. These power components depend on circuit activity and operating frequency (see equation 2-2).

$$P_{Dyn} \sim \alpha . C_{EFF} . V_{DD}^2 . F_{REQ} \quad (2\text{-}2)$$

Where α is the activity factor, C_{EFF} is the effective switching capacitance, V_{DD} is the supply voltage and F_{REQ} is the operating frequency.

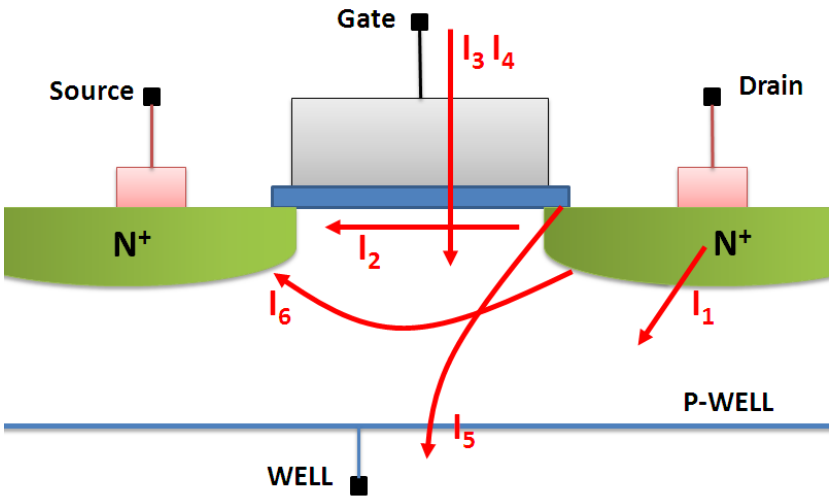


Figure 2-4: Summary of leakage current mechanisms in deep submicrometer transistors

Figure 2-4 shows the different leakage current sources in deep submicron technologies. I_1 is the reverse-bias pn-junction leakage; I_2 is the subthreshold leakage; I_3 is the oxide tunneling current; I_4 is the gate current due to hot-carrier injection; I_5 is the GIDL[2] component; I_6 is the channel punch-through current [31]. In another side, the increase in the leakage current in advanced technology nodes makes the static power consumption became a significant contributor to the power dissipation in ICs.

$$P_{Static} \sim I_{Leakage} . V_{DD} \quad (2\text{-}3)$$

Where $I_{Leakage}$ is the total off current of the circuit.

Energy per operation

The energy per operation metrics is widely used to evaluate the energy efficiency in SRAMs presented by the following equation:

$$E_{Total} \sim I_{Leakage} . V_{DD} . T_{CLK} + \alpha . C_{TOT} . V_{DD}^2 \quad (2\text{-}4)$$

One of the most effective ways to reduce the energy consumption is to lower the supply voltage, as this significantly decreases the dynamic power dissipation and the leakage power. As shown in Figure

[2] Gate-induced Drain Leakage

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

2-5, voltage down scaling reduces the dynamic energy as a square law but linearly increases the leakage power losses. With scaling down of V_{DD} delay increases exponentially, resulting in leakage energy increases and dominates the total energy consumption at low V_{DD}.

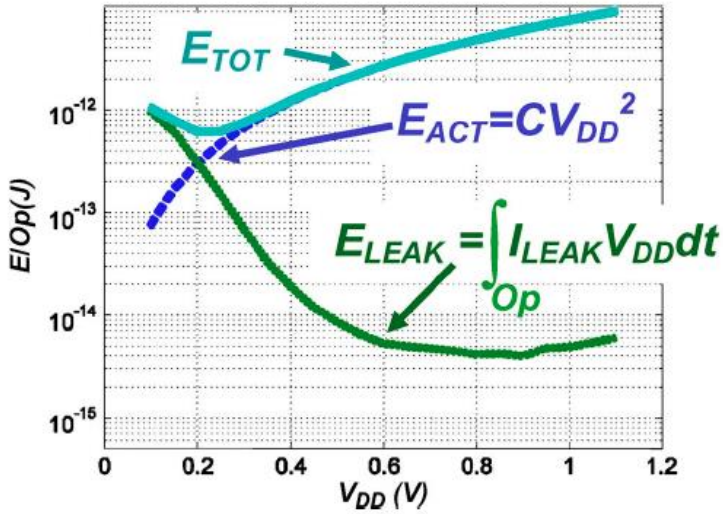


Figure 2-5: Energy profiles of the 90nm carry-ahead adder with respect to VDD [32]

Figure. 2-6 shows the total power of logic and memory across the voltage range in [26], this work describes a processor fabricated in 32nm CMOS technology presenting a V_{DDMIN} equal to 550mV for the SRAM memory part. A lot of memory design works like in [26] and [32] confirm that there is an optimum voltage, V_{Opt} that allows reaching minimum energy consumption per operation, near the MOSFET threshold region.

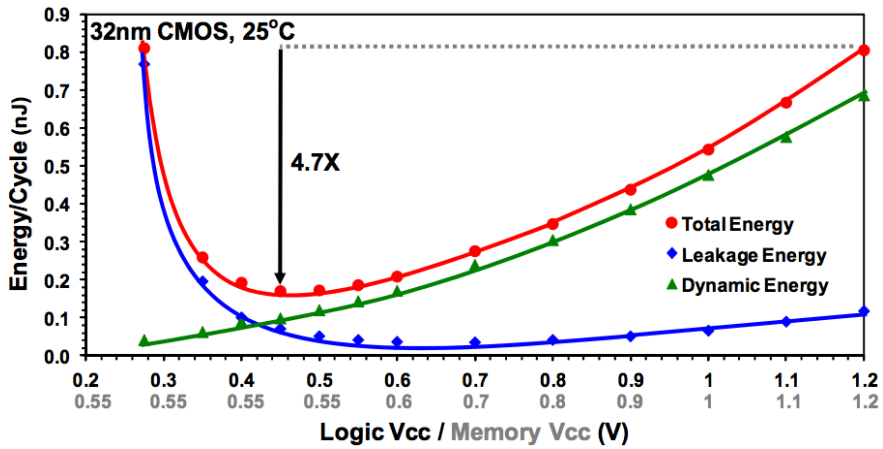


Figure. 2-6. Measured energy characteristics across wide voltage range IA-32 Processor [33]

However in this operating region and given the large number of bitcells in a SRAM cut, it is essential to design a bitcell affected by the lowest possible leakage current. In the following we will discuss the different metrics, which are used to characterize and size the SRAM bitcells.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

SRAM Metrics

Area and stability are two main important parameters for the characterization of SRAM bitcell: the bitcell area defines more than two-third of the total SRAM cut size and L1[3] caches occupy a big part of many SoCs. The cell stability defines the sensitivity faced to process, voltage and temperature variation (PVT). These two aspects are interdependent as shown by Pelgrom law (increasing the bitcell area allows to decrease the sensitivity to process variability). In advanced CMOS technology nodes bitcells become very sensitive to the various sources of noises due to the increase in variability. The most challenging issue in sub-threshold SRAM is increasing reliability during read and write operations. A good metric for read/write margins is critically important to all kinds of SRAM designs [33]. Various metrics are defined in order to characterize the bitcell stability. In the next section the 6T SRAM bitcell will be presented and the metrics will be described.

1.1. 6T SRAM bitcell

Figure 2-7 shows the conventional 6T SRAM bitcell well-known architecture. The 6T SRAM memory cell is composed of two cross-coupled CMOS inverters with two pass-transistors connected to complementary bitlines. The two pass-gate transistors (PG0 and PG1) are controlled by the wordline (WL) signal to perform the read and write accesses depending on the bitlines. The columns bitline true (BLT) and bit line false (BLF) act as input/output nodes carrying the data from internal nodes to the sense amplifier during the read operation, or from write circuitry to the memory cells during write operation.

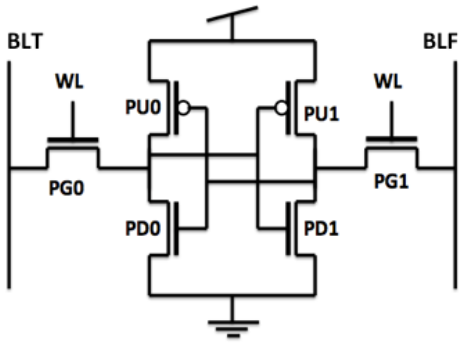


Figure 2-7: The 6T SRAM bitcell

[3] L1 is a cache used by the central processing unit (CPU) of a computer to reduce the average time to access data from the main memory. The cache is a smaller, faster memory which stores copies of the data from frequently used main memory locations. Most CPUs have different independent caches, including instruction and data caches, where the data cache is usually organized as a hierarchy of more cache levels (L1, L2 etc.)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

1.2. Static noise margin (SNM)

The static noise margin quantifies the amount of voltage noise required at the internal nodes of a bitcell to flip the cell's contents. The degradation of SNM can limit voltage scaling for SRAM designs above this minimum voltage. The SNM approach appears with the first studies performed at the level of infinite logic gate chain in order to define the absolute minimum voltage (noise) to apply at the input to cause the flip of the output state. A mathematical equivalence is derived allowing establishing the static noise margin of the bitcell [34] [35]. Many methods exist in the literature to calculate the SNM. Seevinck method [36] becomes the de-facto approach in industrial SRAM design flow used to evaluate SNM. Figure 2-8 illustrates the setup circuit allowing the achievement of butterfly curves. The well-known butterfly curve is a graphical representation of the worst-case SNM.

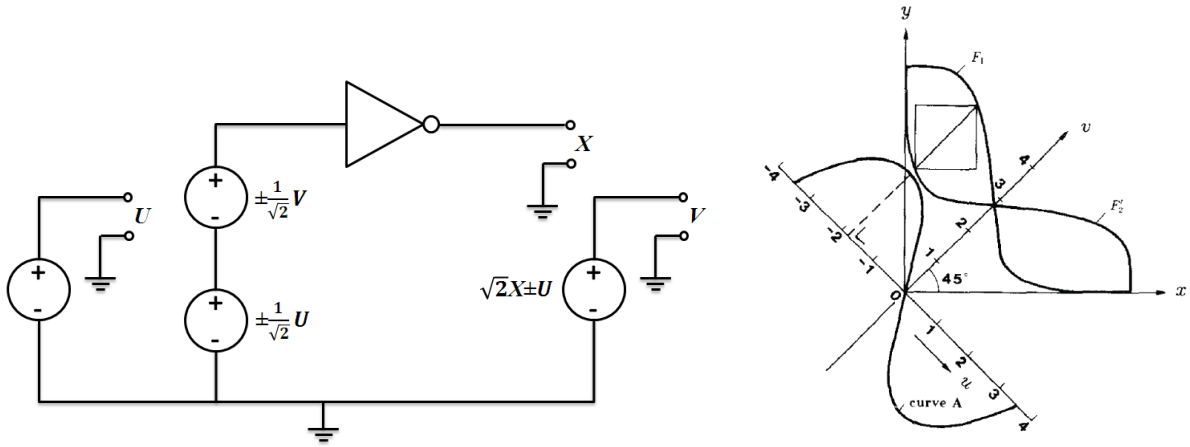


Figure 2-8: Seevinck's schematic setup [34]

In this schematic, U is an independent voltage source that is swept to implement a standard rotation transformation of 45 degrees [34]. When performed twice (or once for two inverters) with appropriate sign selection, plotting U versus V yields a butterfly curve from which SNM can be readily calculated by way of subtraction. It is of utmost importance to point out that this technique is open loop for an SRAM cell in the sense that each inverter is isolated from its cross-coupled partner. The feedback seen in the equation of the schematic simulation is an artifact of the rotation transformation and as such, does not intrinsically capture the closed-loop dynamics of an actual SRAM cell [34]. This method can be used for read, write and hold mode in order to establish the static noise margin for the write (WSNM), the static noise margin for the read (SNM_{READ}) and the static noise margin during the standby mode (SNM_{HOLD}) corresponding to the side of the maximum square which can be inserted in the butterfly curve as shown in Figure 2-9.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

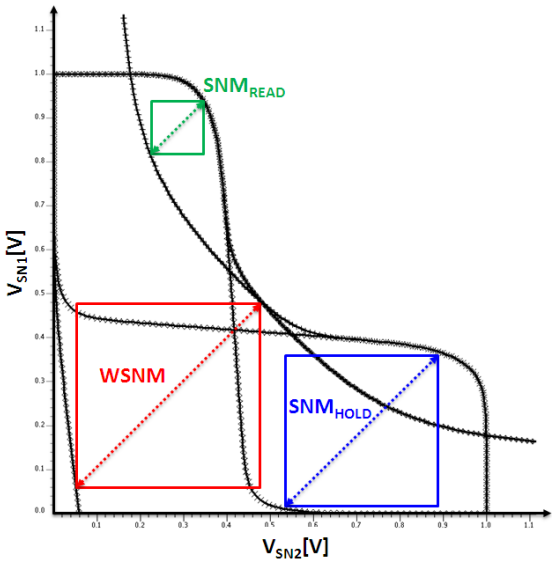


Figure 2-9: Typical voltage transfer curve for read, write and retention mode

1.3. Read margin (RM)

During the read operation in the 6T cell, the word line, WL, is activated. The two bitlines are floating and they converge into a state depending on the configuration of the internal nodes of the selected bitcell as shown in the Figure 2-10. The differential voltage developed by the two bitlines is sensed by the sense amplifier. Bitline's capacity and the read current are the main parameters which define the read time of the selected bitcell (see equation 2-5).

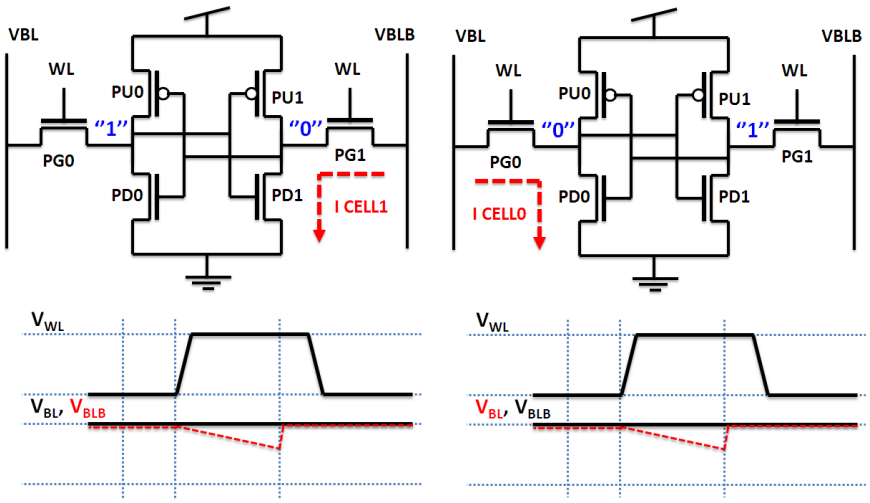


Figure 2-10: Read operation: read 1 (left) and read 0 (right)

$$\Delta V_{BL} = I_{cell}.\Delta t/C_{BL} \qquad (2\text{-}5)$$

The read operation is the most destructive in term of information: successful read operation is related to the read stability and the read current which guarantee data sensing and reduction of read power consumption. The Static Noise Margin is an important parameter for memory allowing characterizing the bitcell read stability. In this mode, SNM is the minimum amount of noise that can corrupt the data

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

stored in the bitcell. It has been shown that during read the SNM takes its lowest value and the cell is in its weakest state.

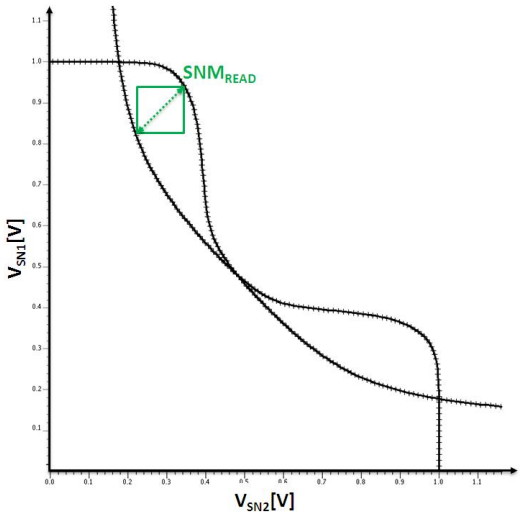


Figure 2-11: Butterfly curves during read operation a standard 6T SRAM bitcell in 28 CMOS LP

Hence, the read operation is more critical in term of stability. This is why SNM usually correspond to the SNM calculated in the read mode (see Figure 2-11).

1.4. Write margin (WM)

During the write operation in the 6T bitcell, the write circuit assigns V_{DD} and GND for the bitlines (BL and BLB) depending on the value of the data input, DIN. The word-line connected to the selected bitcells is activated at the same time and the value of the internal nodes will flip based on the state of the bitlines. The write static noise margins (WSNM) are widely used as the criteria of write ability. Figure 2-12 shows the circuit when writing “1”. The write margin corresponds to the width of the smallest square that can be introduced between the lower-right half of the butterfly curves [33]. For a successful write, only one cross-point should be found in the butterfly curves, indicating that the cell is mono-stable, as shown in Figure 2-12(right). The final write margin for the bitcell is the minimum between the WSNM for writing “1” and for writing “0”.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

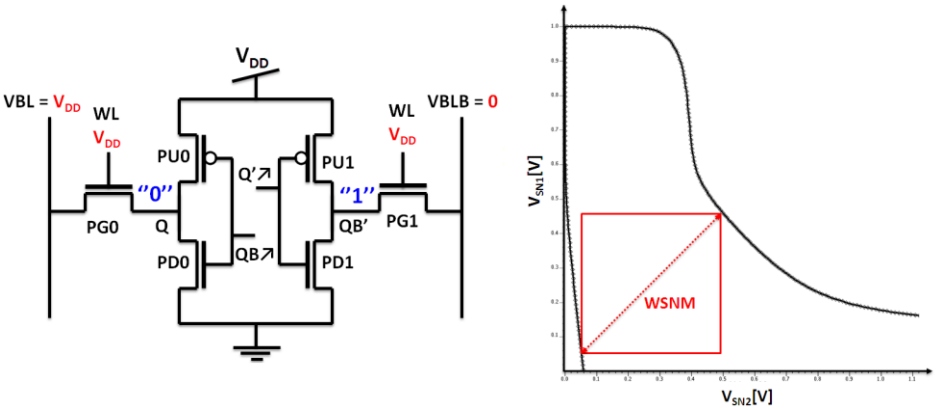


Figure 2-12: Circuit for WSNM when writing '1'. (left), WSNM when writing '1': width of the smallest embedded square at the lower-right side (right).

There exist also several other static metrics along the measurement of WM, such as BL and WL margins required for flipping the state of the bitcell, where the write margin is defined as the minimum bitline voltage (Figure 2-13) and the minimum WL voltage respectively (Figure 2-14).

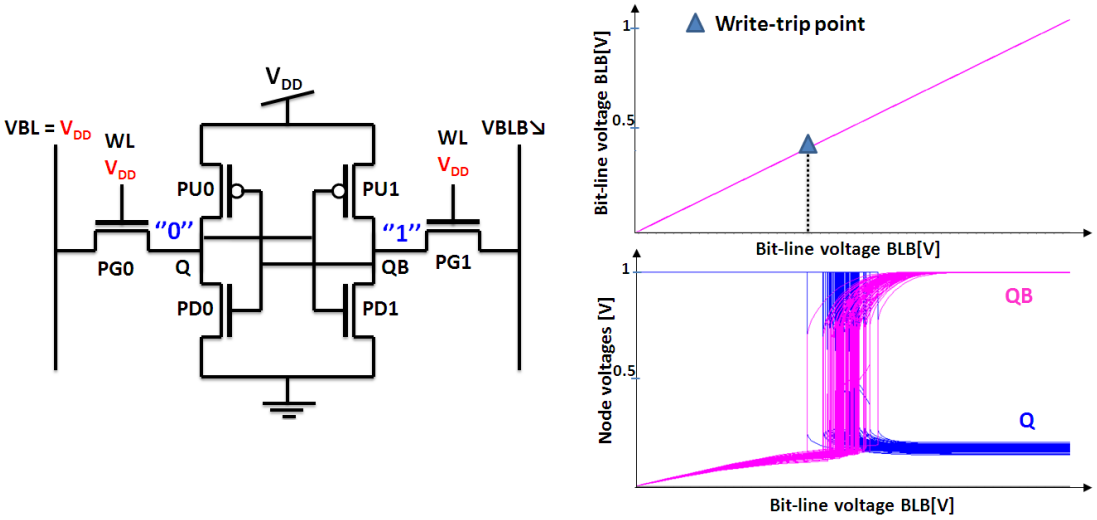


Figure 2-13: Voltage Transfer Characteristic (VTC) of SRAM cell to evaluate write margin by the BL sweeping method

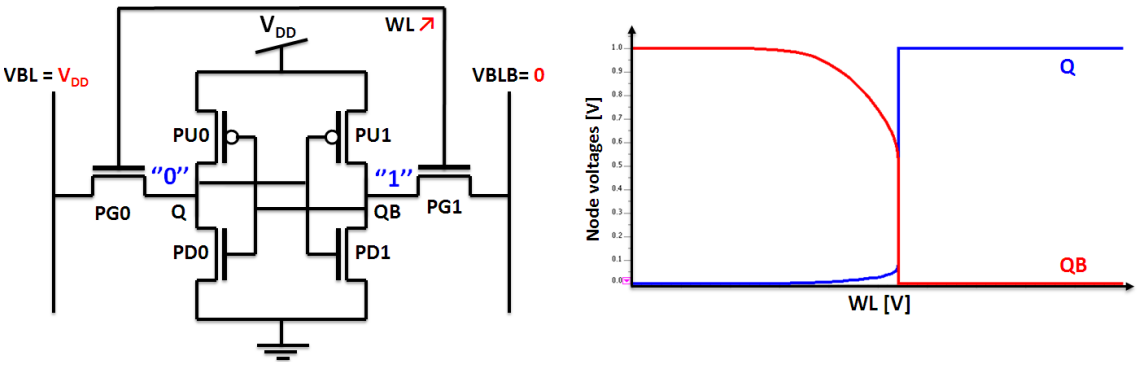


Figure 2-14: Circuit for write margin from WL sweeping (left), Write margin (VWL) is defined as the difference between VDD and the WL voltage when nodes Q and QB flip (right).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

1.5. Data retention voltage

The hold-margin or the data retention voltage is the minimum voltage for retaining data in the internal nodes in the bitcell in standby mode (Figure 2-15).

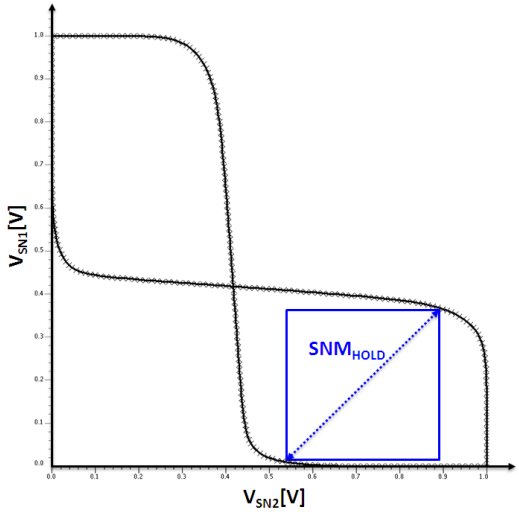


Figure 2-15: Butterfly curves during hold mode for a 6T SRAM bitcell in 28 CMOS LP

The static power consumption has widely increased in advanced CMOS technology nodes. During the hold mode, the static power consumption is the main contributor to the total power consumption of the SRAM (Equation 2-2). In data retention mode, the access transistors are turned off, and the supply voltage is decreased down to a certain value (Data Retention Voltage (DRV)) in order to reduce static power consumption [37]. The influence of the supply voltage on the leakage power of a 6T bitcell in data retention mode is illustrated in Figure 2-16 (6T bitcell with 0.12 µm area in 28nm CMOS bulk).

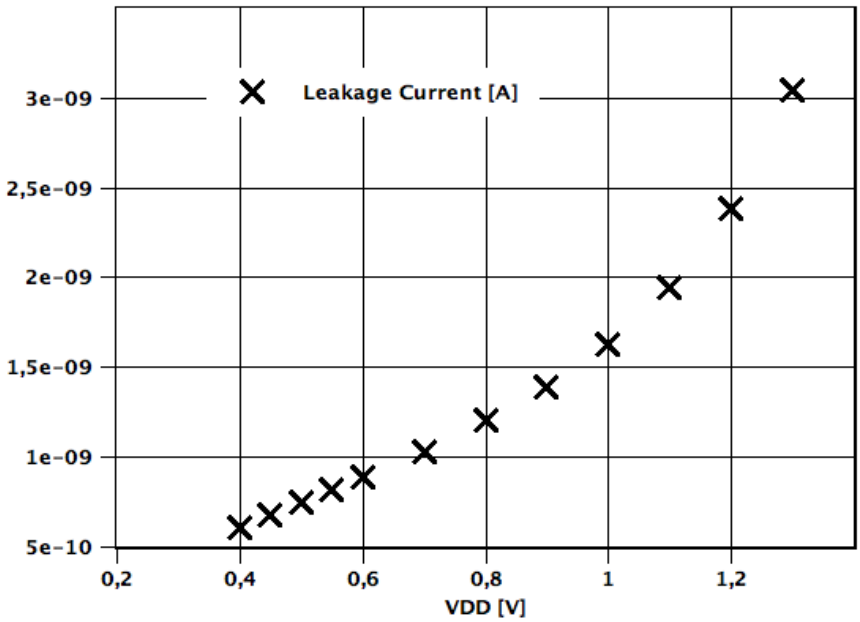


Figure 2-16: Simulated static power vs. supply voltage for a 6T bitcell (0.12 µm, 28 CMOS bulk)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

As shown in Figure 2-16, scaling down V_{DD} by 40% (from 1V to 0.6V) results in 70% reduction in the static power consumption in worst-case process and temperature conditions. Therefore reduction of supply voltage in standby mode is a very beneficial and desired option allowing power saving. But this option causes the destruction of the internal data since lowering supply voltage below DRV results in hold failures. A tradeoff must be settled between power saving and reliability.

SRAM bitcell optimization

Read and write margins are the main metrics that define the SRAM bitcell stability. These two margin criteria dependent on the strength of α, β and γ ratios.

Where $\boldsymbol{\alpha} = \frac{(W/L)_{PG}}{(W/L)_{PU}}$, $\boldsymbol{\beta} = \frac{(W/L)_{PD}}{(W/L)_{PG}}$ *and* $\boldsymbol{\gamma} = \frac{(W/L)_{PD}}{(W/L)_{PU}}$

To design a bitcell with a good write-ability, the PU transistors in Figure 2-7 must be smaller than the PG transistors (increase α ratio) and similarly the PD transistors must be bigger than the PG transistors to guarantee an acceptable readability (increase β ratio).

$$(W/L)_{PD} > (W/L)_{PG} > (W/L)_{PU} \quad \textbf{(2-6)}$$

The relation (2-6) illustrates that the promotion of the write-ability by increasing the PG transistors compared to PU transistors causes the degradation of the readability. Therefore increasing write-ability must be followed by an increase in readability. So, if we summarize, improving WM is limited by the degradation of SNM_{READ} and improving readability is limited by the PD transistors size, which defines indirectly the bitcell area that should be as small as possible due to its impact on the overall area of a SoC.

Failures in SRAM bitcell operation

Optimization of the bitcell design must be performed while targeting minimal failures. Considering the impact of the bitcells in the area of SoC, a lot of trade-offs must be taken into account for example the design requirements between WM and SNM. This makes the design of an optimized bitcell a challenging task. However a sizing that does not take into account these trade-offs makes the bitcell very sensitive and increases the probability of failures. Even if the bitcell is optimized, it is affected by other sources of failure due to the non-idealities in silicon like PVT variations, the random telegraph noise, aging … these perturbation sources affect directly the offset voltage of the sense amplifier and the timing delay what causes failures at the bitcell level. In order to design an SRAM with an acceptable yield, all these effects should be anticipated as margins. There are four metrics to

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

characterize the bitcell failures: the write failures, the read failures, the read stability failures and the hold failures.

1.6. Readability Failure

The readability failure consists in the inability for reading the bitcell internal data during the read cycle time. Successful read operation in the 6T bitcell is detected when the bitline holding GND is more than $\Delta V \geq V_{OFFSET}$ below the bitline holding a V_{DD} when WL turns ON. V_{OFFSET} is the offset voltage of the Sense Amplifier (SA).

1.7. Read ability Failure

A read failure occurs when the internal data change during the read cycle time. This kind of failure is due to degradation in the bitcell SNM. The worst case of the read stability is when there is a multiple read operation in a same row of the memory. Since the weakest bitcell can flip during the second cycle, hence stability is detected after an additional read operation.

1.8. Write ability Failure

The write ability failure consists in the inability for changing the bitcell internal data during the write cycle time. Checking the write ability failure is pessimistic if measurement of the internal nodes is performed immediately after the word line is turned-off and it is optimistic if measurement is done after a long period of time.

1.9. Hold Failure

As scaling down is the best way to reduce the static energy consumption during standby mode, decreasing the supply voltage below the data retention voltage DRV results in hold failures because the feedback provided by the cross-coupled inverters becomes too weak and the bitcell flips.

1.10. PVT impact on the SRAM bitcell operation

Process, Voltage and Temperature variability (PVT) are the main sources that impact directly the reliability of the integrated circuits and especially the SRAM bitcells as shown in Figure 2-17. The spread of local mismatch in advanced technology nodes is wider, also there is a lot of request to reduce power consumption by scaling down the supply voltage and the temperature effect which presents a different behavior between the nominal and at ultra-low voltage: this will result in the decrease in SRAM reliability.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

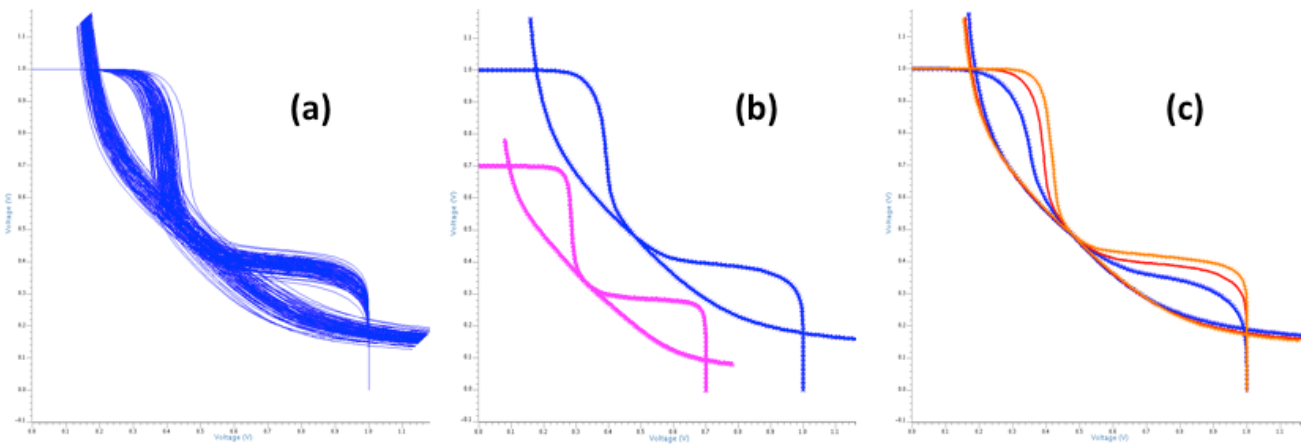


Figure 2-17: Read margin curves of the 6T SRAM cell under PVT variations in 28nm FDSOI technology

Butterfly curve in Figure 2-17(a) is obtained by a Monte Carlo simulation (1024 runs) at 1V and 25°C. Figure 2-17 (b) illustrates the butterfly of the 6T bitcell at 1V and 0,7V supply voltage and finally Figure 2-17 (c) presents the butterfly curves for various temperature (-40°C, 25°C and 125°C).

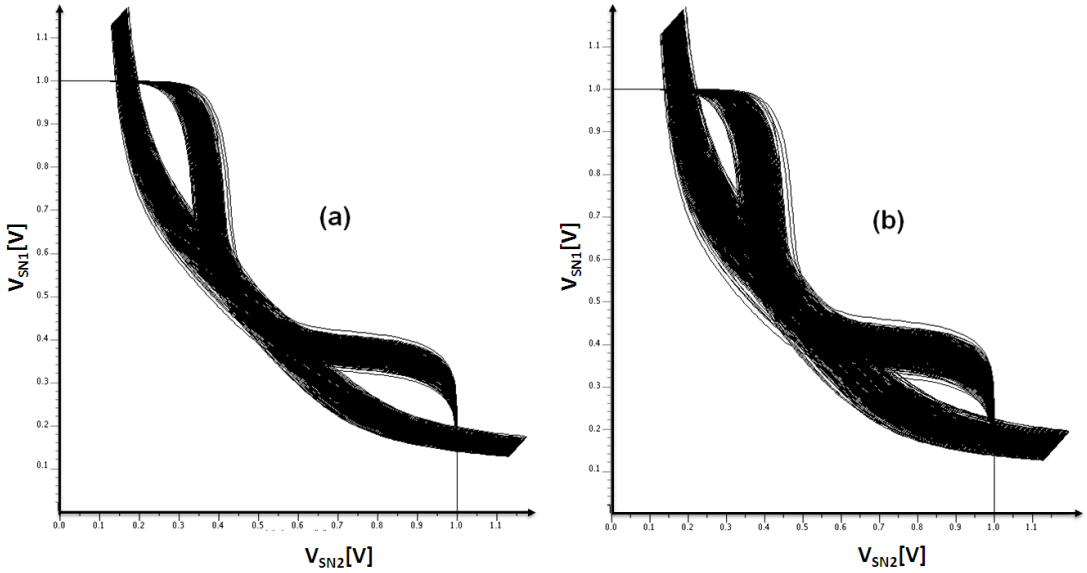


Figure 2-18: Butterfly curves of two different 6T bitcells with an area : (a) 0.197µm2, (b) 0.120µm2 (1024 MC runs)

Figure 2-18 presents Monte Carlo simulations of two 6T SRAM bitcells with different size (0.197µm and 0.12µm). According to Pelgrom law, increasing the bitcell area causes the reduction of the σ_{VT} and the increase in the read stability and vice versa. Butterfly curves show that the larger bitcell has less variability and more SNM value. Similarly, the histograms of the WM in Figure 2-19 illustrate that the spread of the distribution is thinner for the larger bitcell (0,197µm). Simulation results confirm that there is more variability in the smaller bitcell (0,12µm): σ_{VT} (a) =22.5 mV ≥ σ_{VT} (b) =15.4 mV.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

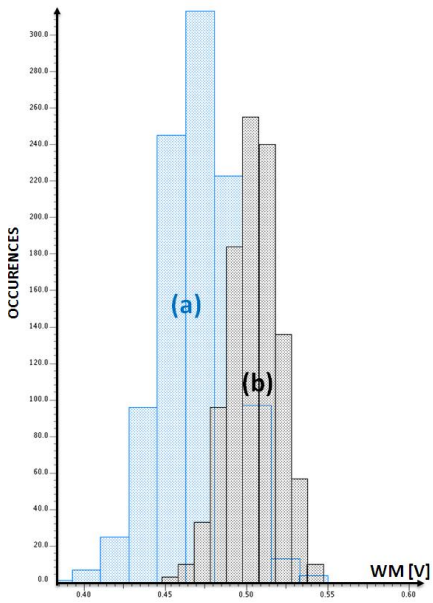


Figure 2-19: WM histograms for two different 6T bitcells, (a) 0.12µm (b) 0.197µm (1024 MC runs)

Limitation in the 6T SRAM bitcell in term of VDDmin

As explained in Chapter 1, scaling down the supply voltage is a de-facto approach used in order to decrease the total power consumption in SRAMs. The minimum supply voltage (V_{MIN}) is the V_{MIN} in the SoC. V_{MIN} of SRAMs limits the lowest operating voltage in the SoC. In SRAM, V_{MIN} is limited mainly by the bitcell read stability (SNM) and by its write-ability (WM). Since variability increases with scaling down V_{DD} and given that the SNM and the WM in the standard 6T bitcell are inversely related metrics, this creates a limitation in term of minimum operating voltage.

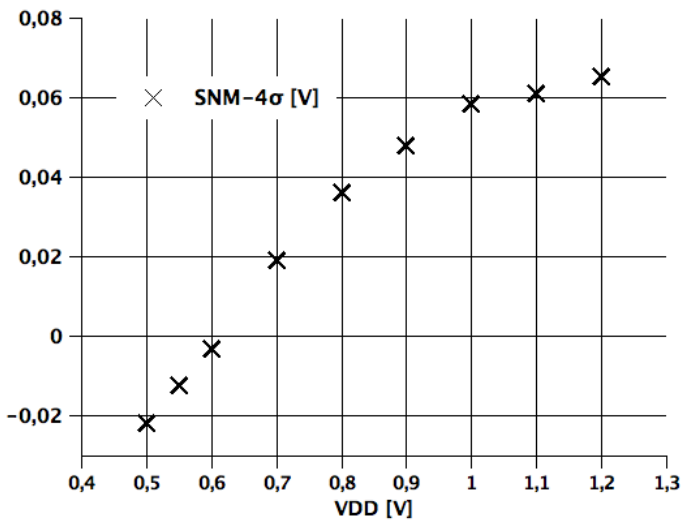


Figure 2-20: SNM (Average -4 σ) versus supply voltage (FS4, 125 °C)(VDDMIN, SNM = 613 mV)

[4] Fast Slow

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

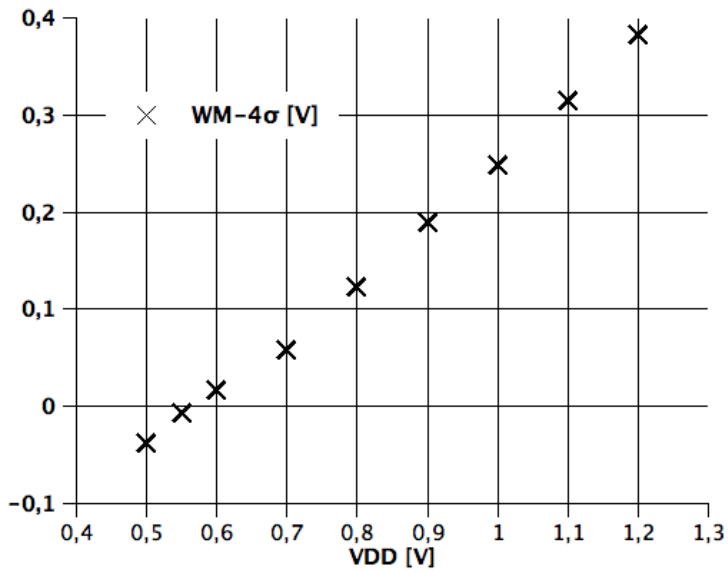


Figure 2-21: WM (Average -4 σ) versus supply voltage (SF5, -40 °C) (VDDMIN, WM= 566 mV)

$$VDD_{MIN} = MAX\left(VDD_{MIN,SNM}, VDD_{MIN,WM}\right) \quad (2\text{-}7)$$

The minimum supply voltage in a bitcell is the maximum between $VDD_{MIN,\ SNM}$ and $VDD_{MIN,\ WM}$ where $VDD_{MIN,\ SNM}$ and $VDD_{MIN,\ WM}$ present the minimum supply voltage for which SNM ≥ 0 mV and WM ≥ 0 mV respectively. Operating the conventional 6T bitcell at Low Voltage (LV) with a good yield is challenging, since the bitcell read-stability and its write-ability get degraded and cannot be both optimized at the same time for a given area, since they have conflicting design requirements. Research and prediction indicate that VDDMIN of the 6T bitcell is around ~0.6V [38]. Figures 2-20 and 2-21 show an evaluation in terms of SNM and WM for the 6T bitcell (0.197µm2) in 28nm CMOS bulk. The results indicate that VDDMIN is limited to 613 mV at 4σ due to degradation of the SNM.

State-of-the-Art in ultra-low voltage SRAM

Over last years two research directions are predominant in order to design an ultra-low-power SRAMs. The first one consists in using the standard 6T bitcell but with read and write assist techniques allowing operations at lower VDD_{MIN}. However the using of read and write assist circuits result in more complexity at the periphery level and over consumption. The second direction consists in looking for other bitcell architectures that will be able to operate in sub-threshold domain without the need for any assist techniques. In order to solve the limitations in WM and SNM at low voltage range due to the conflicting design requirements, Preferred bitcells are used, based on the improvement of the Write ability WM (or SNM) at the cost of the degradation of readability SNM_{READ} (or WM) and vice versa by adjusting the strength of bitcell transistors by transistor size (width and

[5] Slow Fast

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

length) and V_{TH}. This is achieved through the adjustment of the dimensions and the threshold voltage V_{TH} of transistors in the bitcell [39]. At the same time in order to compensate for the degradation of the write-ability or the readability, read and/or write-assist circuits must be used. The different techniques allowing designing a preferred bitcell are now described.

1.11. Write-Preferred bitcell and Read-Assist Circuit

In order to improve the write-ability in the bitcell, α ratio must be set to a higher value (refer to relation 2-6). This can be obtained by increasing the strength of PG transistors or by making the PU transistors weaker.

Using PU transistors with high V_{TH0} (HVT MOS)	This technique presents a barrier for SRAM operating in sub and near threshold range.
Increasing the width of PG transistors	It's a good technique allowing the improvement of WM and read current but with area penalty [39].
Using PG transistors with low V_{TH0} (LVT MOS)	It's a very beneficial technique to increase the WM and read current without any area penalty. Unfortunately this approach causes an increase in I_{OFF} current and so reduces the number of bitcells that can be stacked per column.

In order to compensate for the degradation in readability in the write-preferred bitcell, a read assist technique must be used. The most common read assist techniques are now described:

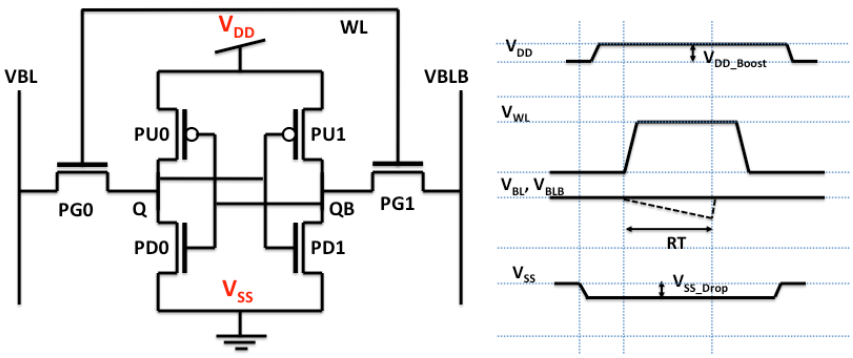


Figure 2-22: Tuning VDD or VSS [40]

Increasing V_{DD} or lowering V_{SS} at the bitcell level results in the increase in the strength of PD transistors and thus sufficiently increases the SNM (Figure 2-22). However to prevent the large energy consumption due to this method, there is a requirement for a column-by-column control of local bitcell V_{DD} (or V_{SS}) to improve SNM of all bitcells in read operation mode and only these ones. Several circuits have been proposed in [40] [41] [42]. The drawback of this technique is the need for more specific circuitry assuring the raise and lowering of V_{DD} and for V_{SS} what increases the complexity of

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

the periphery.

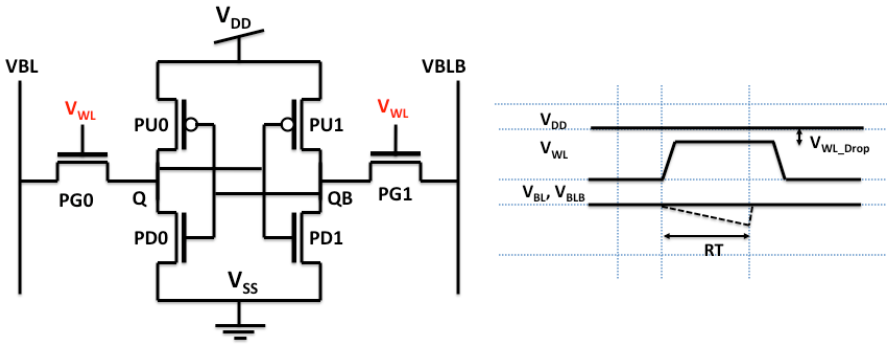


Figure 2-23: WL voltage drop

Lowering the WL signal (Figure 2-23) for the selected bitcell during read operation allows reducing the amount of charge injection between pre-charged bitlines and the lower internal node. The SNM is improved [43] [44] without using options like enlarging the PD transistors to get bigger β ratio.

1.12. Read-Preferred bitcell and Read-Assist Circuits

As already explained, getting a read-preferred bitcell is achieved when β and γ ratios are large. This can be obtained by making the PD transistors stronger or by reducing the strength of PG and PU transistors. Several approaches used to improve SNM_{READ} are detailed now.

Using PG and PD transistors with high V_{TH0} (HVT MOS)	This technique is beneficial since it hasn't any area penalty and since SNM improves significantly. The implication of this method is the degradation of read current what causes a negative impact on read time and I_{ON} /I_{OFF} ratio.
Decreasing the width of PG transistors	The drawback of this method is the decrease in the read current while decreasing PG width.
Decreasing the width of PU transistors	This technique improves the SNM_{READ} but not with the same significance as the other methods.
Increasing the length of PG transistors and the width of PD transistors	This approach allows increasing readability by making the PG transistors weaker. The increase in PD transistors is for the purpose of compensating for the decrease in the read current due to the increase in PG lengths.
Increasing the width of the PD transistors	This technique allows not only to improve the readability but also to increase the read current. The drawback is an area penalty [39].

To improve the weakness of write-ability in the read-preferred bitcell, various write assist techniques are used which are described now.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

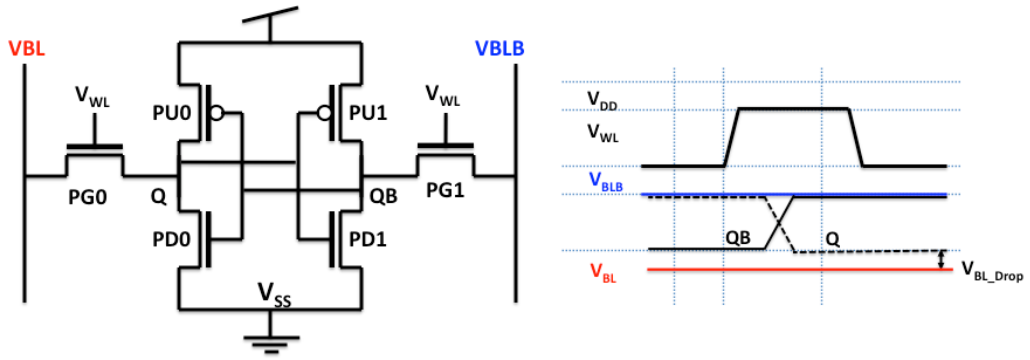


Figure 2-24: Negative bitline voltage (write 0)

Negative bitline is an efficient write-assist technique allowing the improvement of the WM for read-preferred bitcells (Figure 2-24). Scaling down the supply voltage, V_{GS} of the pass-gate transistors decreases and the mismatch due to the variability of V_{TH} increases. As a result writing 0 becomes critical and the write failure probability increases. Applying a negative bitline, V_{GS} increases and so writing 0 into the bitcell becomes more efficient. Several circuits have been proposed in previous SRAM designs to overdrive the bitline [44] [40]. The drawback of this approach is an area penalty due to the need of capacitance; also caution should be exercised since excessive negative voltage in BL can cause a data-flip in unselected bitcells on the selected column (disturbed bitcells). The on-set timing of the negative bitline must be adjusted carefully.

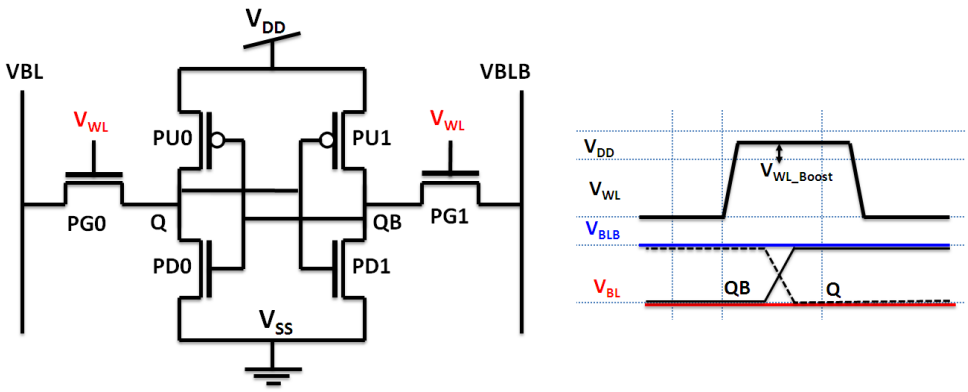


Figure 2-25: Word line voltage boost [45]

Boosting WL during a write operation (Figure 2-25) makes V_{GS} of the PGs transistors raise and hence writing a data into the bitcell becomes easier. This is an efficient method to compensate for the weakness of write-ability in the read-preferred bitcell however dual voltages and specific circuits must be used what costs an area penalty and causes more congestion at SRAM architecture level. Moreover increase in V_{WL} results in degradation of the SNM_{HOLD} for half-selected bitcells in the same row.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

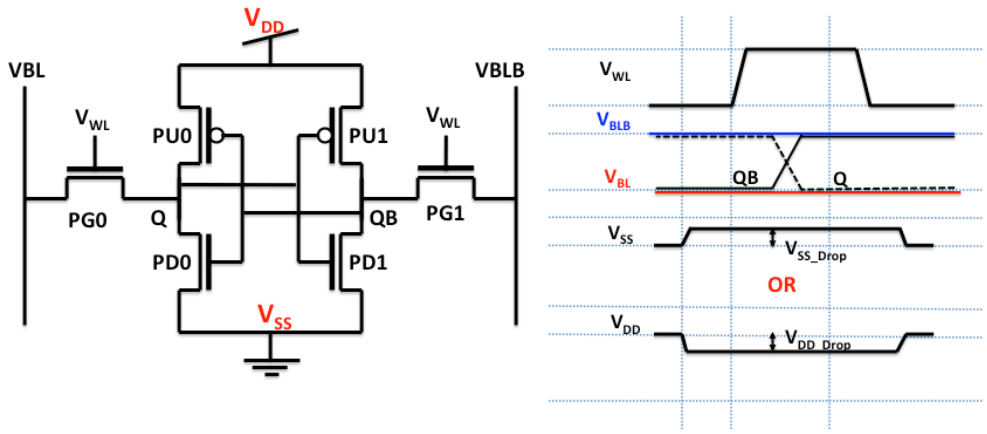


Figure 2-26: Rising V_{SS} or lowering V_{DD}

Increasing V_{SS} and lowering the supply voltage (Figure 2-26) are widely used as write-assist techniques to improve the WM. A safety voltage margin must be taken in order to avoid data flipping in half-selected bitcells. This method requires also column-by-column control of V_{SS} (or V_{DD}) lines to avoid the increase in power consumption. Increasing V_{SS} (or lowering V_{DD}) is required only for the selected column [39].

1.13. Previous ultra-low voltage bitcell architectures

Various new bitcell architectures have been proposed in order to solve the limitations in the standard 6T bitcell in term of minimum supply voltage.

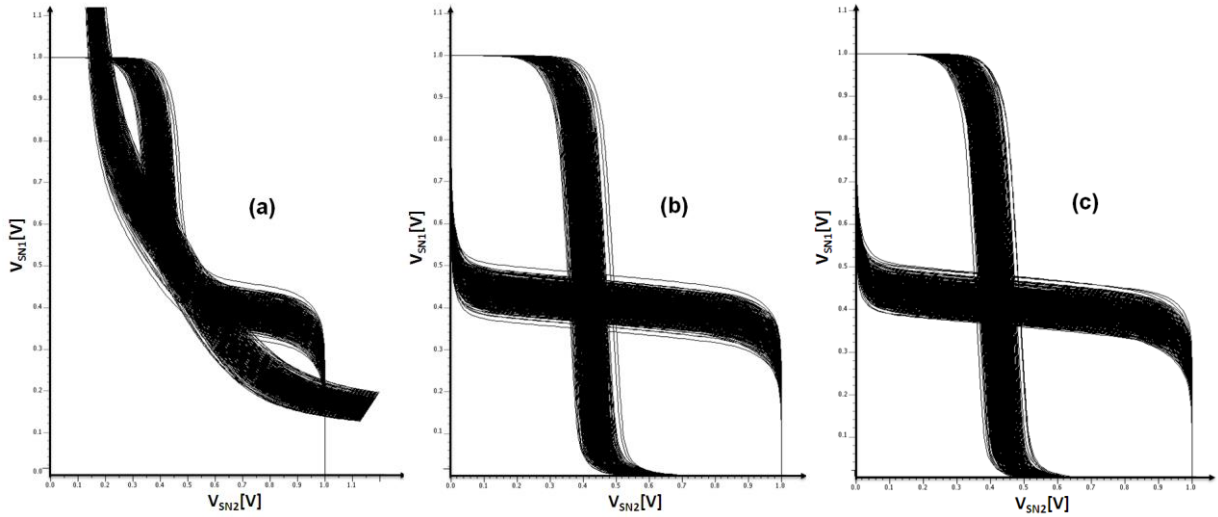


Figure 2-27: Butterfly curves of (a) 6T bitcell in read mode (b) 6T bitcell in hold mode and (c) standard 8T bitcell in read mode (TT, 1V, 25°C: PVT conditions &1024 MC runs)

Excellent ideas have appeared considering the separation of read and write paths using a separated read port. Based on the fact that the Static noise margin of a 6T bitcell in retention mode is given quite higher than that in read mode and that it has a limitation in term of SNM due to the conflicting design requirement of the read versus write margins in the 6T bitcell. As a result SNM of bitcells with

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

separated read port is almost the same as the hold SNM of the conventional 6T SRAM as illustrated with the 8T bitcell in Figure 2-27.

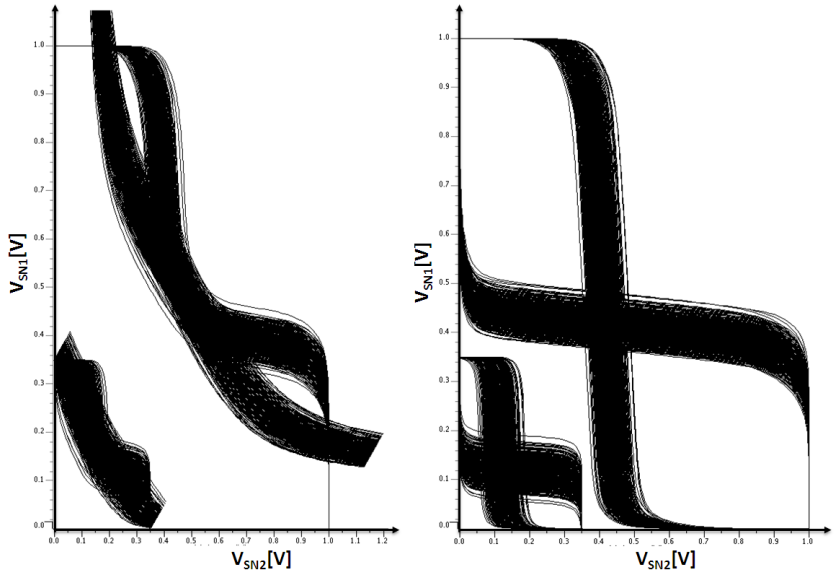


Figure 2-28: Butterfly curves of 6T bitcell (left) and 8T bitcell (right) in read mode with V_{DD} = 1V and 350 mV (TT, 25°C, 1024 MC runs)

The butterfly curves in Figure 2-28 illustrate the significant degradation of SNM in the standard 6T bitcell while scaling down V_{DD} what makes the operation in ultra-low voltage a challenging task even when using read and write assist options. The 8T standard cell offers always enough read stability margin under ULV without the need for any assist techniques. However an area penalty in the bitcell using separated read-port must be taken into consideration. The minimum operating supply voltage V_{MIN} is mainly limited by the degradation of SNM and WM in sub and near threshold voltage range. A lot of new bitcell architectures have been proposed with the aim of making the ultra-low voltage operation feasible. The proposed 8T bitcells in Figures 2-29(a) and 2-29 (b) [46] allow eliminating the disturbance in the half-selected bitcells during read and writing operations. This approach results in improvement of SNM_{READ}. The 8T bitcell in Figure 2-29 (b) [46] uses two PG transistors in series allowing the reduction in the leakage current but it is also accompanied by a degradation of WM and read current due to the resistive path. Figure 2-29(c) shows the single ended 7T bitcell [47] [48] using a read-port to improve the read margin. However the WM is weak due to the dissymmetry. Figure 2-29(d) presents the standard 8T bitcell [49] [50], which is widely used for high, speed application and also in ULV range. The drawback of this bitcell is the limitation in I_{ON}/I_{OFF} ratio. As shown in Figure 2-29(e), the proposed differential 10T bitcell in [51] uses two separated read ports but leakage issues also limit it. In [51] a single ended 10T bitcell (Figure 2-29(f)) has been proposed using a special read port allowing the reduction in leakage current. In [52] an original differential *ZIGZAG* 8T bitcell has been presented (Figure 2-29(g)) allowing to improve the read current but suffering from conflicts between the leakage and the read current during read operation.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

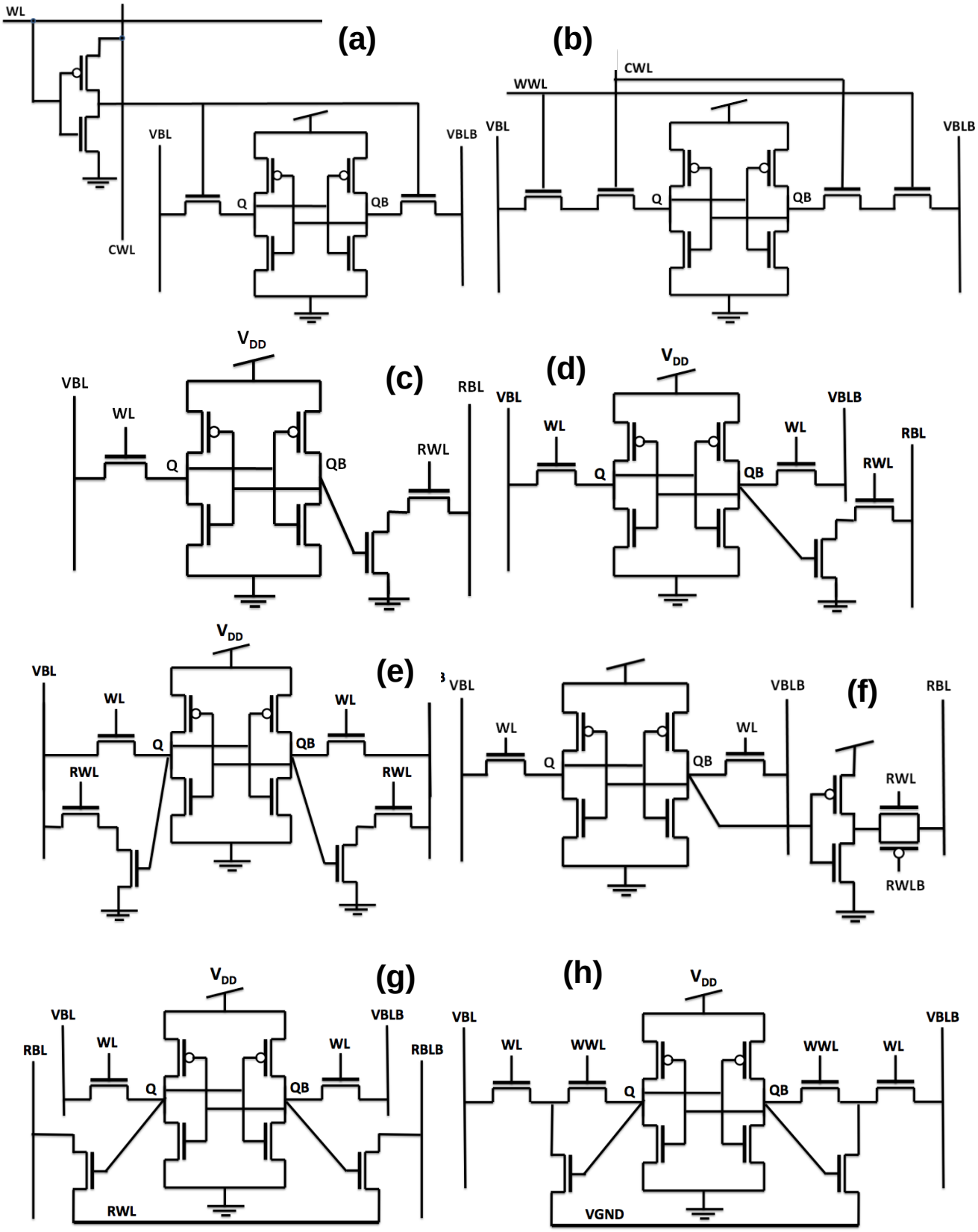


Figure 2-29: SRAM bitcell architectures : (a) 8T bitcell, (b) CR8T [49], (c) 7T [47], (d) 8T [50], (e) differential 10T [51], (f) 9T[51], (i) ZIGZAG8T [52] and (k) VGND10T [45]

Figure 2-29(h) presents another interesting VGND 10T bitcell [45]. A lot of new bitcell architectures have been appeared in the last years enabling the functionality of SRAM at ULV. What makes the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

difference between these bitcells is the stability, the immunity facing perturbations (half-selected bitcells), the sensing mode (single ended or differential) and the area. In the other hand, two issues that have not been addressed in the state of the art relates to the parasitic power consumption and the read failures due to the weakness of the I_{ON}/I_{OFF} ratio at ULV, these issues will be addressed in the next chapter.

TABLE 2.1 SUMMARY OF DIFFERENT SRAM BITCELL TOPOLOGIES

	8T (a)	8T (b) [49] [50]	7T (c) [47] [48]	8T (d) [49] [50]	10T (e) [51]	10T (f) [51]	Z8T (g) [52]	10T (h) [45]
Controls signals	**2**	**2**	**2**	**2**	**2**	**3**	**2**	**3**
Sensing	**Differential**	**Differential**	**Single**	**Single**	**Differential**	**Single**	**Differential**	**Differential**
Ion/Ioff ratio issue @ ULV	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**Yes**
WM	**High**	**Low**	**Very Low**	**High**	**High**	**High**	**High**	**Low**
SNM	**Low**	**Low**	**Medium**	**High**	**High**	**High**	**High**	**Very High**
Half-selected immunity	**High**	**High**	**Very Low**	**Very Low**	**Very Low**	**Very Low**	**Very Low**	**Very Low**
Leakage	**Medium**	**Very Low**	**High**	**Very High**	**Very High**	**medium**	**medium**	**Very Low**

Limitations in Ultra-low voltage sense

Static Random Access Memory (SRAM) faces an important limitation in read cycle time that prevents high frequency operation and the possible applications at ULV. The full swing sensing is a practical approach to circumvent the poor performances of sense amplifiers (SA) under ULV operation. Besides ULV induces a limit in the performance in frequency of SRAMs. The read cycle time is negatively affected by degradation in the bitcell read current and the increase in variability. The most common technique used with single-ended bitcells is called *full-swing sensing* [53] which is a slow process (large delay). There is an interest to optimize sense amplifiers (SAs) for both full-swing and single-ended bitcells under ULV. State-of-the-art SAs are reported for supply down to 500 mV due to the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

increase in local and global variation with scaling down V_{DD} [54]. Lowering V_{MIN} down to 300mV is detailed in chapter 4 along with review of the state of the art in sense amplifiers.

Conclusion

This chapter illustrates the main limitations of sub and near-threshold SRAM bitcell design. The increase in the threshold voltage variation has a negative impact on mismatch and the bitcell performances. Ultra-low voltage bitcells should then have a relatively larger area to avoid failures compared to bitcells for standard power supply conditions. Two main research axes are discussed in order to solve the limitations in WM and SNM in ultra-low voltage range due to the conflicting design requirements (tradeoffs must be settled between α and β ratios). As discussed in the chapter, a lot of previous works are based on the use of read- or write- preferred 6T bitcells. However this technique needs additional assist-circuits, what increases the complexity of the SRAM periphery while not always be efficient in term of VDD_{MIN}. The use of new bitcell architectures with separated read ports appears as a promising alternative that enables the feasibility of efficient operation at ULV. Unfortunately this kind of bitcells (Fig. 28, using read ports) suffers from parasitic phenomena like the degradation in the I_{ON}/I_{OFF} ratio resulting in read failures and the parasitic power consumption in the half-selected bitcells. The read-time is largely decreased at ULV what limits the operating frequency and eventually the possible applications. So to improve the read-time, a sense amplifier should be used. Unfortunately data sensing during the read operation is a challenging task at ULV since the VDD_{MIN} of a sense amplifier is limited to ~500 mV beyond 65nm CMOS [54].

The thesis topic was clearly to evaluate new bitcell architectures to limit the usage of assist-circuits and to demonstrate solutions in 28nm CMOS bulk. The operation of the SRAM bitcell in deep ULV conditions has been studied and the main limitations have been highlighted. This step is presented in Chapter 3 and led to the selection of a 10T bitcell with modified read ports. The degradation of the read current at ULV limits the operating frequency (small current to charge/discharge a significant bitline capacitor). Forward body biasing has been studied to boost the read current (modulation of V_{TH}) and an improved subthreshold sense amplifier has been designed (high-speed sense amplifier down to 280 mV power supply) in chapter 4.

As a summary:

- Chapter 3 details the various constraints and limitations of previous ULV bitcells. Bitcell candidates are proposed and validated (silicon results in 28 nm CMOS bulk) that avoid the main limitations in sub-threshold operation.
- Chapter 4 discusses several options used in order to enable operation in ultra-wide voltage range for a cut of 32kb SRAM memory.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

- Finally Chapter 5 presents the architecture of a 32kb SRAM memory and a test chip in 28 nm FDSOI.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Chapter 3

Ultra-low voltage bitcells

1. Introduction

The chapter 2 highlights the advantages of scaling down supply voltage in order to reach the minimum energy point in SRAM memory. The minimum supply voltage of Systems-on-Chip is limited by the VDD_{MIN} of the SRAM. As discussed previously, two research directions are predominant in order to design ultra-low-power SRAMs. The first one consists in using the standard 6T bitcell but with read and write assist techniques allowing for operation at lower VDD_{MIN}. The second direction consists in looking for other bitcell architectures that will be able to operate in sub-threshold domain without the need for any assist techniques. This chapter focuses the latter direction. New SRAM bitcell architectures have been proposed recently as solutions to the limitations of the six-transistor (6T) SRAM bitcell in term of minimum supply voltage, V_{MIN}. There is no demonstrated bitcell as superior under ultra-low supply voltage like the 6T bitcell at nominal voltage. Main limitations concern first the ratio between the read current and the standby current at the lowest operating voltage, Second the bitcell robustness to perturbations and third the data sensing sensitivity, among other but minor limitations. The chapter presents two proposals of ten-transistor (10T) Ultra-Low-Voltage bitcells for 0.3V operation and processed in 28nm LP CMOS bulk. Simulation results are compared to experimental results to demonstrate a satisfying operation at Ultra-Low supply voltage. Two 10T SRAM bitcells are proposed here with the idea of gains in layout and flexibility when decoupling the data-write path from the data-read path. This scheme seems to offer acceptable performances under Ultra-Low-Voltage (ULV) whereas the 6T bitcell is obviously limited in SNM and WM. Recently in this context, various SRAM bitcells have been proposed (8T [55], 9T [56], 10T [45] and 11T [57]) to enhance stability for robust ultra-low voltage and power operation. 8T bitcells have been proposed [58] that feature a read-mode SNM equal to the hold-mode SNM. A comparative study is proposed that concerns the SRAM bitcells reported in [45], [59] and [52] respectively (Figure 3-1), working under ULV in 28nm bulk CMOS, i.e. the technology selected for the experimental vehicles reported in this manuscrit. In [52] a separate data-read path is added to the original 6T SRAM bitcell with independent data-lines. In [59] additional pass-gate transistors separate even more the operation of the data-read path. A more complex architecture is proposed in [55] for that purpose.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

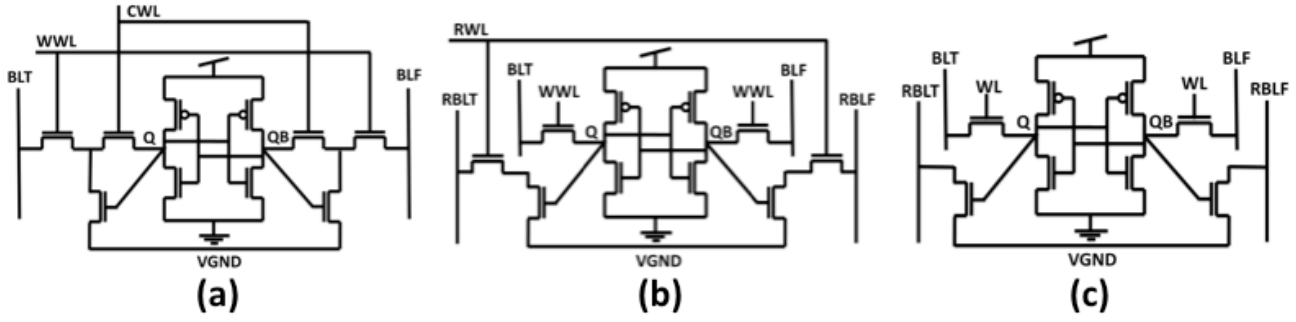


Figure 3-1: (a) The 10T IK JOON bitcell [45], (b) 10TVGND [59] and (c) the zig-zag Z8T [52]

An analysis of the various constraints and limitations of available ULV bitcells is presented in Section II: particularly the effects of temperature, the I_{ON}/I_{OFF} ratio, the dynamic losses and soft error disturb. A first proposal of 10T bitcell is presented in Section III: a multiplexing technique enables to reduce significantly the effect of parasitic currents in un-selected bitcells. Section IV details the second proposal of 10T XY bitcell. An original data-read port is introduced to limit the effect of leakage current. Comparison with state-of-the-art bitcells is proposed in Section V in terms of SNM, WM, leakage current and WT. Measurements on silicon vehicles are presented in Section VI along with simulation results in 28nm CMOS. The simulation results confirm the interest of these new 10T bitcells to be processed in 28nm CMOS technology.

Constraints and limitations of available bitcells

A conventional 6T SRAM bitcell is built with two cross-coupled CMOS inverters, the contents of which can be accessed through two nMOS access transistors. However operating a conventional 6T at Low Voltage (LV) with a good yield is challenging, since the bitcell read-stability and its write-ability get degraded and cannot be both optimized at the same time for a given silicon area, since they have conflicting design requirements. This section reviews the main common limitations of SRAM bitcells in ULV operations.

1.1. Effects of temperature and doping on Mobility

In a semiconductor both mobility and charge carrier concentrations are temperature dependent. Figure 3-2 shows the histograms of the read current, I_{CELL}, of the 10T bitcell in [45] that is described later-on, at three different temperatures (-40°, 27° and 125°), as obtained with a Monte Carlo (MC) scheme at ULV (Figure 3 (left)) and under 1V (Figure 3 (right)) respectively.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

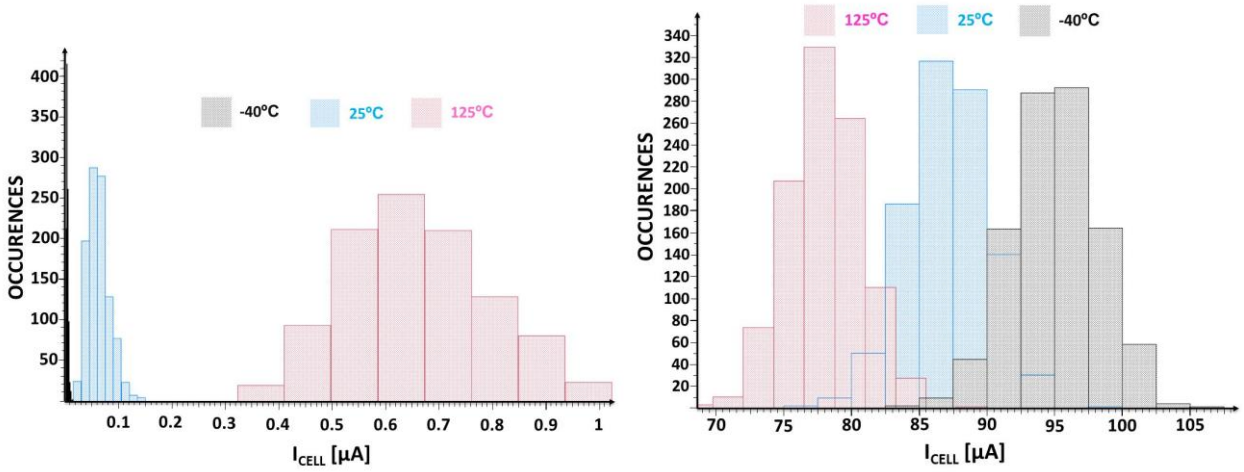


Figure 3-2: Monte-Carlo simulation of Iread (1024 runs), (TT Corner Process) (left) at VDD=300 mV and (right) at VDD=1V

The peak of the distribution in Figure 3-2 (right) gives the most probable I_{CELL} value. It is noticeable that the current value is smaller with a higher temperature. This phenomenon is related to the channel mobility degradation with temperature. This tendency is reversed at ULV (Figure 3 (left)). The carrier-pair thermal generation counterbalances the mobility degradation at ULV. The spread of the distribution is variable at ULV while it is rather constant at V_{DD}=1V. Figure 3-2 (left) gives evidence of a low I_{CELL} current value at ULV at low temperature.

1.2. I_{ON}-to-I_{OFF} ratio

In high-speed operations, it is common to use a sense amplifier when a bitcell is accessed, to detect differentially the drop on one of the read bitlines with respect to the other one, to quickly evaluate the data. However the leakage currents of the access transistors discharge both bitlines. A differential readout works if the read-current of the accessed-cell is able to discharge a bitline more quickly than the aggregate leakage-current of all the other bitcells tied to the same bitline in the same column.

As expected, scaling down V_{DD} strongly reduces the mean read current (Figure 3-3) and increases the further degradation from variation. The leakage current of the un-accessed bitcells in ultra-low-voltage domain represents a big barrier causing read operation failure: a solution must be provided to reduce the leakage current. The degradation of the read current for a given bitcell regarding the leakage current generated by unselected bitcells in the same column, becomes critical and causes failure in read operation, especially at high temperature and when the supply voltage is less than 400 mV (Figure 3-2). Reducing V_{DD} degrades the so-called I_{ON}-to-I_{OFF} ratio, which is equal to I_{CELL}/ (N_r*$I_{LEAK, BL}$), where I_{CELL} is the read current of a given bit cell, $I_{LEAK, BL}$ is the leakage current of a non-accessed bitcell in the same column and Nr the number of non-accessed bitcells. V_{DD} sets an upper limit to the number of bitcells, Nr, which can be stacked in a column

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

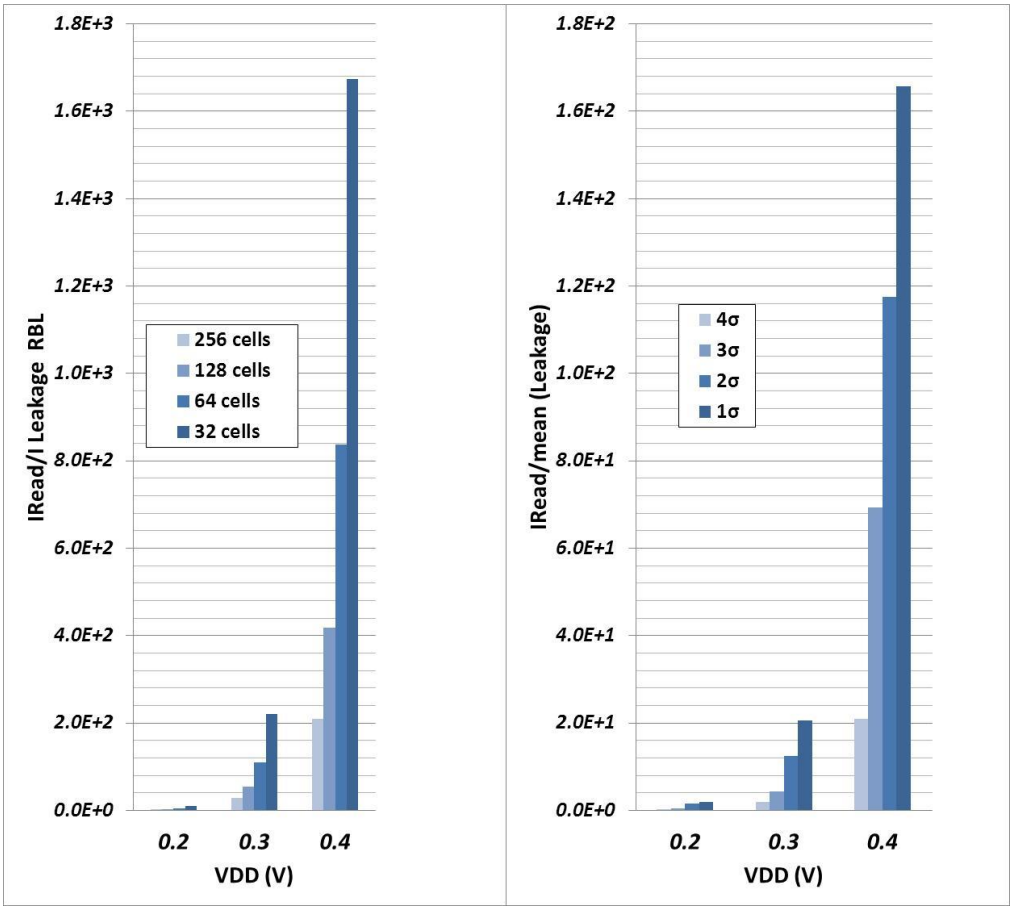


Figure 3-3: Read current degradation in presence of variation with respect to VDD scaling (Monte-Carlo 1024 runs,TT and 25°C)

1.3. Soft Error disturbance

In memory circuits or sequential logic, a soft-error disturbance (SE) is caused by an energetic particle that enters the chip and generates enough free charges to toggle the state of a latch [60].

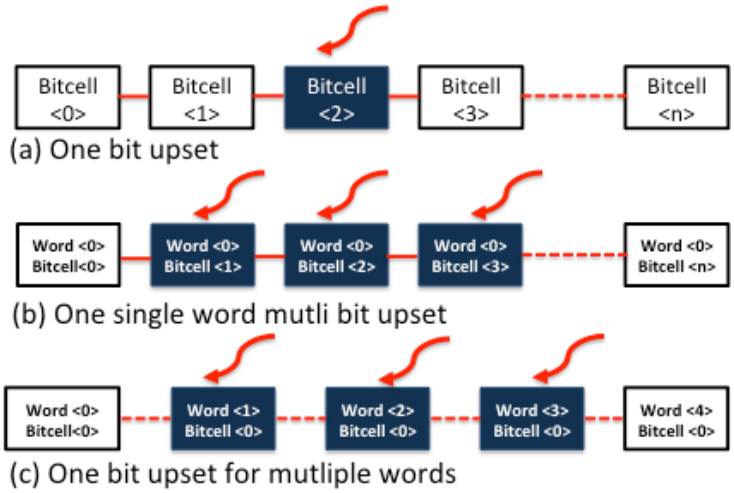


Figure 3-4: Three different scenarios of soft errors [63]

The sensitivity to SEs is directly related to the cell capacitance: the smaller the capacitance, the larger the sensitivity [61]. At each new technology node, due to the surface shrink and its related capacitance

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

shrink, bitcells become more sensitive to SEs. Since the sensitivity increases as the voltage is scaled down [62], SEs are more critical for sub-threshold SRAMs than for standard voltage SRAMs. In advanced SRAM design, soft error became a challenging phenomenon, which directly impact the reliability [64]. Figure 3-4 shows three possible scenarios of soft errors after a particle attack. First scenario in Figure 3-4(a) shows that only one upset bit was occurred due to a particle hit: in this case the soft error can be fixed with Hamming Single Error Correction/Double Error Detection codes (SECDED) [65]. Figure 3-4(b) and Figure 3-4(c) illustrate single event multi-bit upsets, which become predominant in advanced technologies nodes with smaller bitcells [66]. As shown in Figure 3-4(b), many bits in the same word are corrupted: SECDED is unable to correct such error type. The most used technique to avoid single word multi-bit upsets is to interleave bits as shown in Figure 3-4(c), such that logically adjacent bits are not physically adjacent. When soft error results in a multi-bit upset in a bit-interleaved array structure [45] [57], the SECDED is able to solve this error type since each word contains only a single upset bit.

1.4. Dynamic losses

As already stated, the bit interleaving technique can resolve multiple bit soft-errors at ULV [45]. Figure 3-1(a) shows the implementation of a 10T bitcell using two pass-gate transistors and Figure 3-5 details the read- and write-operations. When a bitcell is selected either for read or write operation, all bitcells located in the same Write Word-Line (WWL) are half-selected and have a floating potential on their bit lines. The unselected bitcells suffer during read and write operations from additional losses, which are due to a current discharge. The parasitic components' behavior in half-selected bitcells directly affects the dynamic energy consumption, as it is the case in the standard 8T bitcell [58] and the 10T bitcell [45].

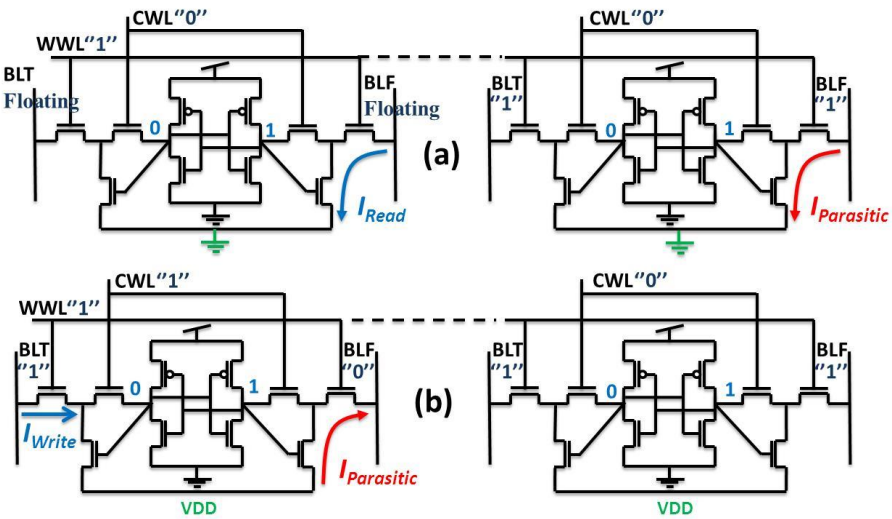


Figure 3-5: Behavior of the 10T bitcell [45]: (a) Read and (b) Write operations

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

During a write operation, a parasitic current is also injected from V_{DD} to the bitline BLF as indicated in Fig 3-5(b). To avoid this phenomenon, the easiest way is to use one word per line. There are no more unselected bitcells in the selected line. Unfortunately this technique increases the number of bitlines with a negative impact on the complexity of the SRAM memory architecture. Moreover the number of bitlines is limited and finally the gain on the latter dynamic losses is very small: it is not worth the complexity. The implementation of a multiplexing technique is proposed here to allow multiple words per line but with limited effects from the unselected bitcells.

Proposed "10T-MUX" bitcell

One primary solution to fight the leakage current issue in read port is to consider a virtual read word line (RWL_MUX) accessed through simple one-transistor read ports (M9, M10 in Figure 3-6(a)) in a multiplexed configuration. This solution is supposed to offer the minimum delay in read port access and thus to provide the minimal read-time.

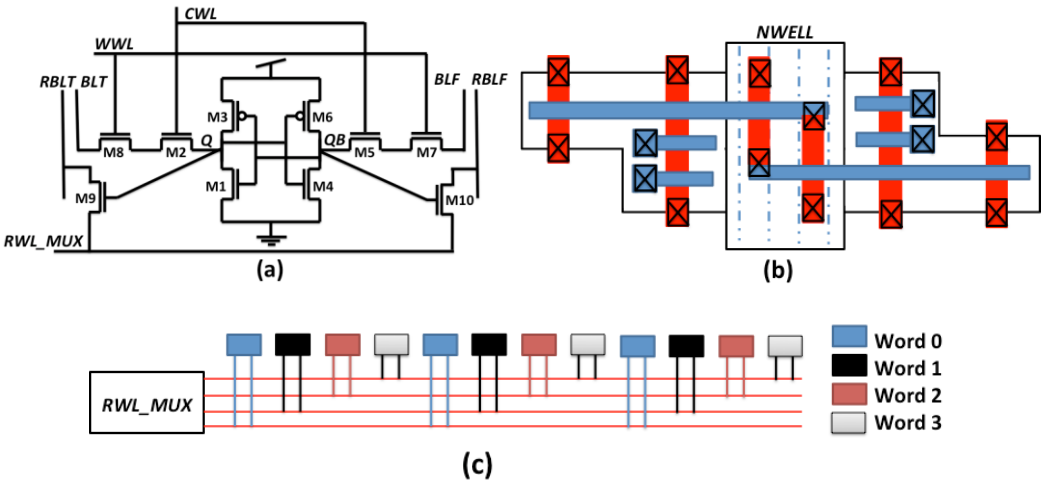


Figure 3-6: Proposed 10T-MUX ULV bitcell: (a) schematic, (b) Layout and (c) hard coding technique

Figure 3-6(a) shows the schematic of the proposed 10T-MUX SRAM bitcell. This symmetrical cell comprises two cross-coupled CMOS inverters and two pass-gate transistors in series to allow the use of a bit-interleaving technique. This guarantees a high resistive path between bitlines and internal data nodes. The read transistors are used to transfer the data to the read bitlines (RBLT, RBLF). Figure 3-6(b) shows the layout of the proposed bitcell as implemented in 28nm bulk CMOS with 0.8µm^2 area. The transistors in the read path have been sized in order to obtain an acceptable SNM value and read current value under worst case conditions at 300 mV. In order to obtain a proper functionality during the write operation (to avoid parasitic current between the RWL_MUX and bitlines), the 10T-MUX bitcell requires the use of four bitlines.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

A hard coding technique is introduced to solve the half-selected-cell bitlines issue [57]. The idea of this technique is to multiplex the RWL_MUX signal according to the number of words per line. This allows selecting only one word during the read operation and hence eliminates the issue of the parasitic dynamic losses from the unselected bitlines (Figure 3-6(c)). There is an obvious impact on complexity at bitcell and periphery levels. The principle of operation of the 10T-MUX bitcell is reported in Figure 3-7(a) with the waveforms of main control signals for a write 0, hold, read 0, write 1, hold and read 1.

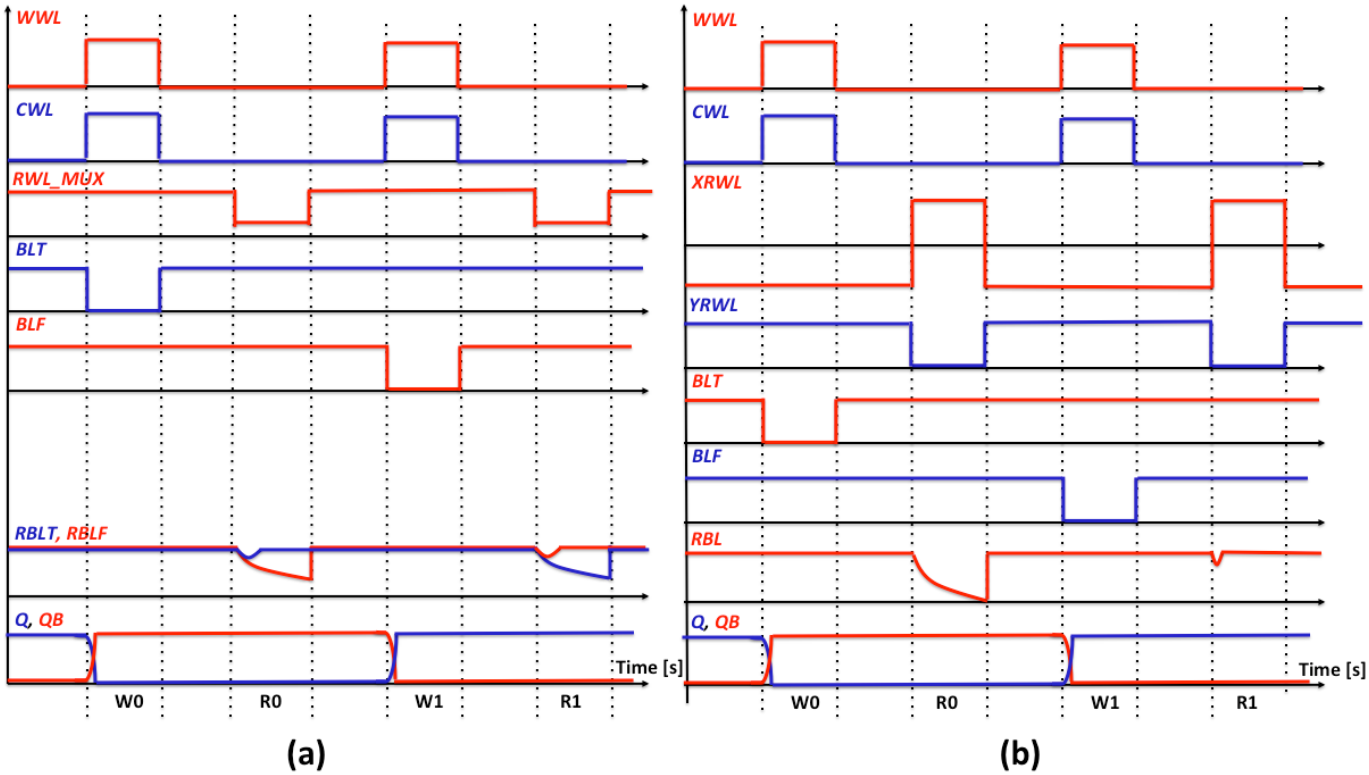


Figure 3-7: Simulated waveforms of main control signals in operation of the 10T-MUX ULV bitcell (a) and the 10T-XY ULV bitcell (b)

Moreover the proposed 10T-MUX bitcell presents a limitation due to injection of charge in the floating read line.

As shown in Figure 3-8, the capacitance in the RWL_MUX is C_{RWL_WORD} = C0 = Cn = 13.44 fF (with 32-bit word and C_{RWL_BIT} =0.42 fF, extracted value from Electrical-Rules-Check operation). The RWL_MUX driver is connected only to one word thanks to the multiplexed RWL technique. The drawback is the charge and discharge of the RWL_MUX capacitance (Cn) for the unselected lines because they are in a floating state what results in additional dynamic losses. Further simulations have been performed to quantify the additional losses in a matrix configuration (X/Y column and row dimension).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

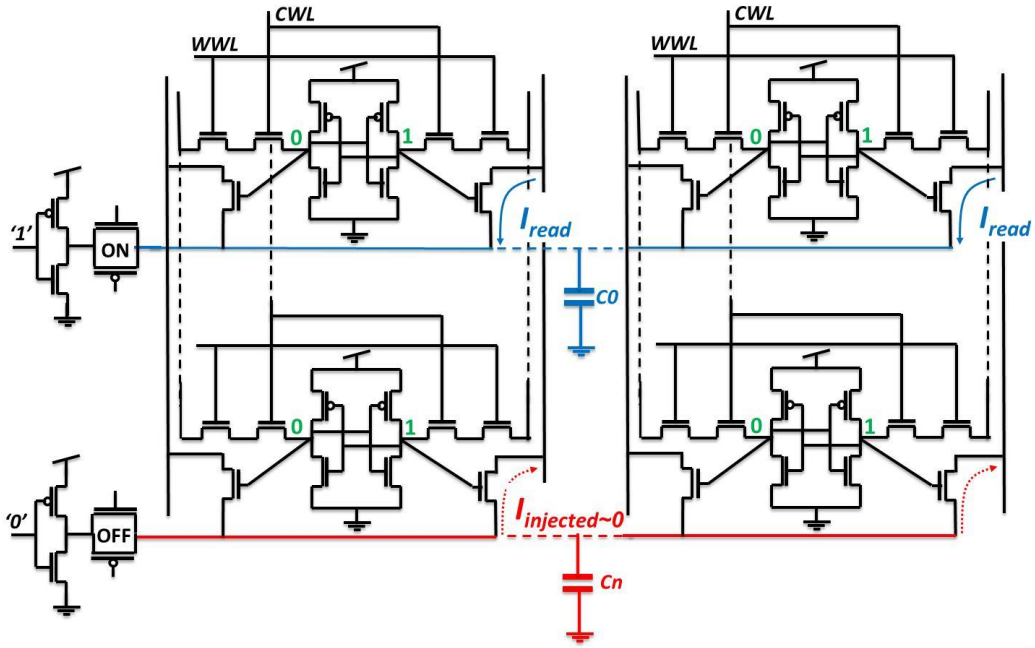


Figure 3-8: Schematic for the charge/discharge issue in the proposed 10T-MUX bitcell in matrix configuration

In 28nm CMOS bulk it appears a read failure for X=32 and Y=128, what means that the capacitance C_0 is too large. Hence, there is a hard limitation (limitation in the number of bitcells stacked per row). Besides the obvious complexity introduced by the separate read ports and the multiplexing technique may not be quantified at periphery level. It is thus the main reason why another bitcell solution is considered.

Proposed XY Read/write bitcell

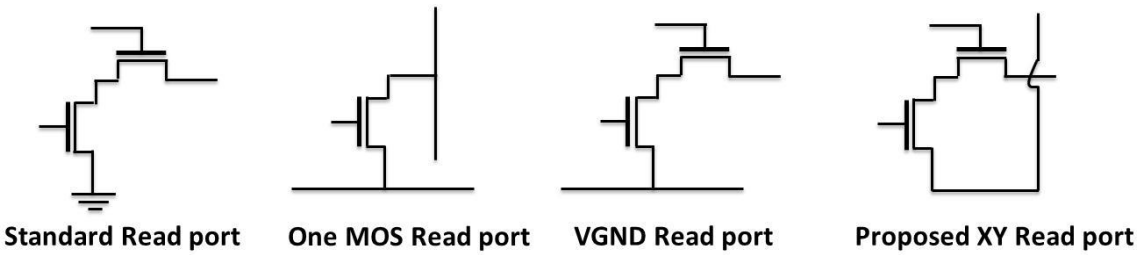


Figure 3-9: Read port configurations

Figure 3-9 shows the schematic of various read port in the state-of-the-art (standard, one MOS and virtual ground (VGND) read port) and the proposed XY selection read port that is detailed in following Section.

1.5. Standard read-port

Figure 3-10 shows the behavior of the 10T bitcell in [45] in 128 columns x 64 row critical path. The bitcell has been designed in 28nm CMOS technology.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

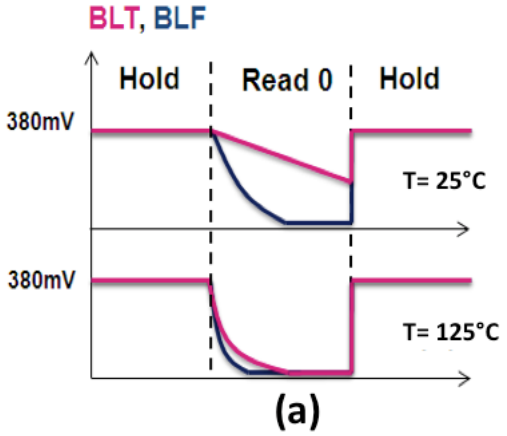


Figure 3-10: Bitlines behavior in 10T-VGND bitcell in [45] (Figure 3-1(b))

As shown in Figure 3-10, the impact of the leakage current from unselected bitcells in the selected column becomes critically dominant and causes read operation failures especially in Fast-Fast, 125°C corner conditions. As a summary, there are two principal drawbacks in a bitcell using a standard read port configuration: the read failure caused by the weakness of the I_{ON}/I_{OFF} ratio in ULV condition and the dynamic parasitic consumption due to pseudo-read in the unselected words in the selected row. A read boost circuit is proposed in [67] as a solution to the read failures caused by the leakage effect. The circuit in [67] was designed in order to achieve two functions: detecting the falling bitline to speed-up the pull-down, and to enhance the opposite pull-up to overcome the leakage effect. There are two drawbacks: first the circuit is useful only for the differential bitcell sense configuration. Second the gate length of the read-port nMOS transistors must be increased in order to reduce the leakage current so as to overcome the worst case situation in FF and 125°C corner conditions. That will result in the reduction of the read current (increased read time) with more silicon area penalty. The proposed 10T-XY bitcell has also been designed in 28nm CMOS technology. Figure 3-11 shows the typical behavior of the bit lines. The unselected bitcells have not the same negative impact and fewer read failures are awaited.

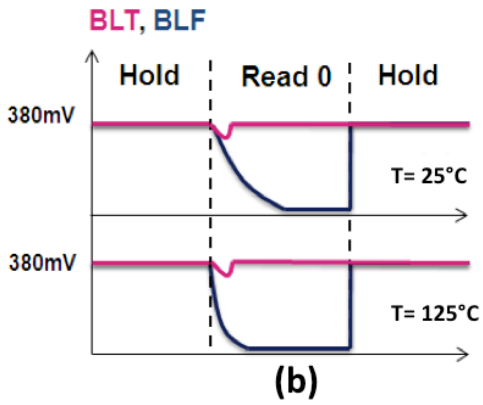


Figure 3-11: Bitlines behavior in proposed 10T-XY bitcell at VDD = 380 mV

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Figure 3-12 shows a comparative simulation of both bitcells (in [45], i.e. 10T-VGND and 10T-XY). The behavior of the read bitcells at ULV shows that the read port in the 10T-XY attenuates significantly the leakage current from the unselected bitcells in the selected column.

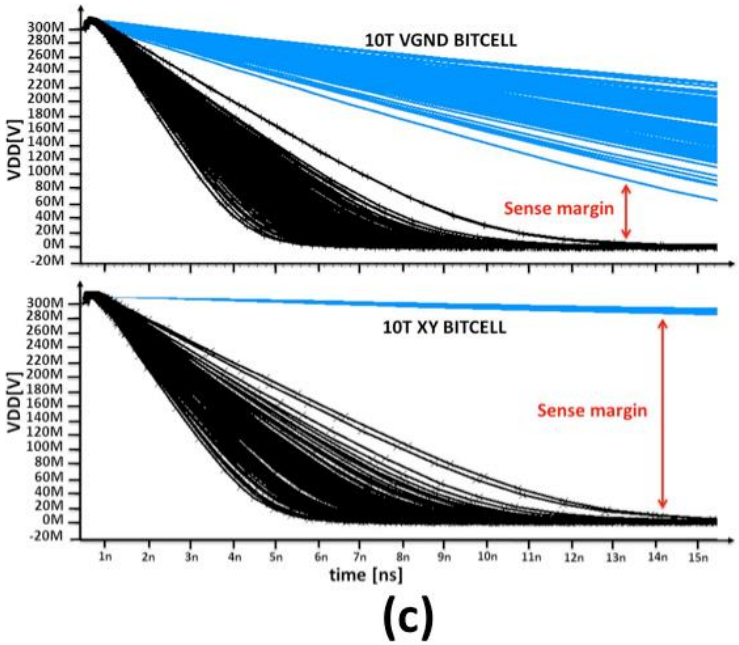


Figure 3-12: comparaison of the read bitlines' behavior in the 10T-VGND bitcell and the differential 10T-XY bitcell respectively (100 MC runs, corner FF, 125°C, 64 cells per column)

1.6. V_{GND} read-port

Bitcells using a read port with zero voltage conditions (V_{GND}) allow reducing the I_{OFF} current significantly and hence solving the weakness of the I_{ON}/I_{OFF} ratio under ULV (Figure 3-13).

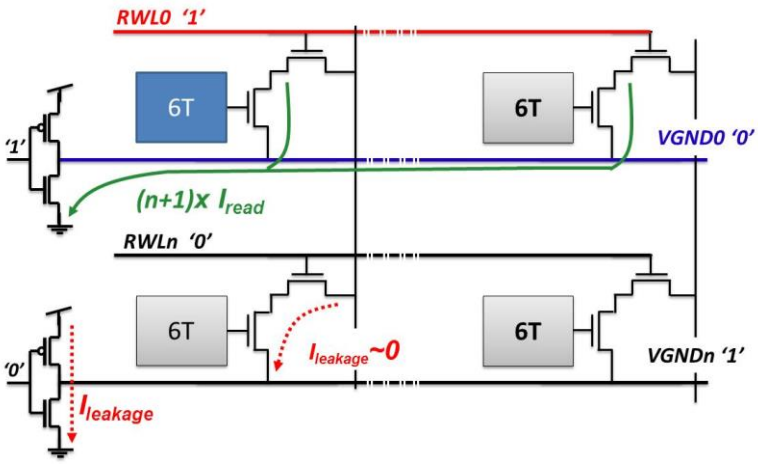


Figure 3-13: Leakage and read current behaviour with the VGND read port

However, there are two drawbacks in such a read port: first the dynamic parasitic consumption due to pseudo-read in the unselected words. Second a charge pump should be used in order to increase the VGND-line driver [68]. VGND driver must sink the read-current from all bitcells in the accessed row

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

and it must draw the leakage-current from all unselected rows. This results in a problem to set VGND to GND and causes a high probability of read failures.

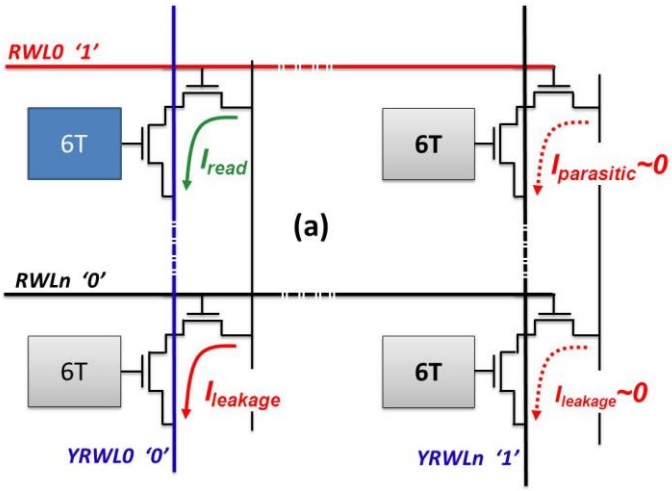
1.7. Proposed XY read port

An XY read-port is proposed in Figure 3-8: There are two control signals in this read-port: the horizontal one, RWL, is set to V_{DD} for a selected row and to GND for unselected rows. The vertical one, YRWL, is set to GND for the selected column and to V_{DD} for unselected columns. As shown in Figure 3-14(a) the proposed XY read-port avoids the power consumption from parasitic phenomena but still suffers from leakage effect issue. Figure 3-15 shows a typical curve of the drain current, I_D, of an NMOSFET versus gate voltage, V_{GS}. The transistor off-state current, I_{OFF}, is the drain current when the gate voltage is set to zero.

TABLE 3.1 SUB THRESHOLD CURRENT EVALUATION VERSUS VGS AND VDS VALUES

V_D	*100mV*	*1V*
I_{SUB} @ V_{GS}= -300mV	**0.14pA**	**10pA**
I_{SUB} @ V_{GS}= 0V	**30pA**	**144pA**

Applying a negative supply to the gate allows to decrease the sub-threshold current by 99.5% at V_D=100 mV and by 93% at V_D=1V (See Table 3.1). So applying a negative supply to RWL in unselected rows within the proposed XY read-port allows avoiding the leakage effect phenomena as shown in Figure 3-14(b).



Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

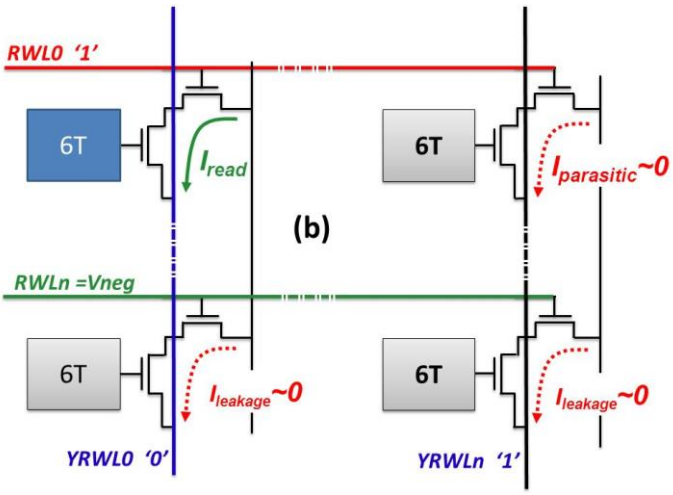


Figure 3-14. Behaviour of currents in the XY read port without underdrive (a) and using under-drive technique (b)

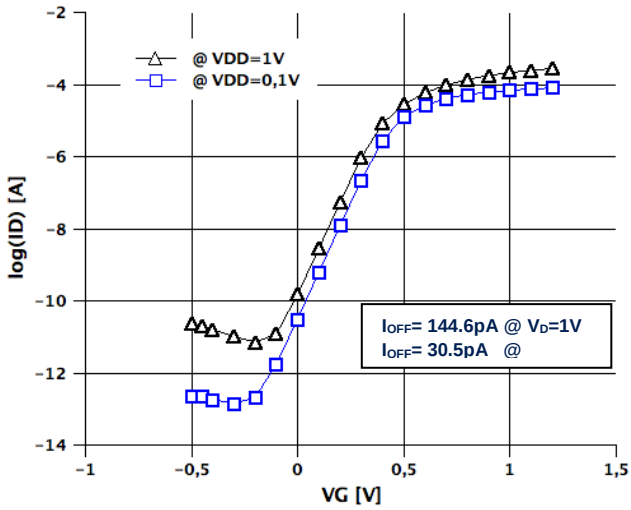


Figure 3-15. I_{DS} versus V_{GS} at two different drain voltages for 250 × 40 nm n-channel transistor in a 28nm CMOS process.

1.8. Proposed XY bitcell

The 10T-MUX bitcell in Figure 3-6 with multiplexed RWL signals represents a solution to avoid the power consumption from parasitic phenomena. Unfortunately this technique is complex since we cannot stack more than two words per line in the layout area and the fact that decoding such multiplexed signals will add more constraints on the X-decoder what makes the task of the designer quite hard.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

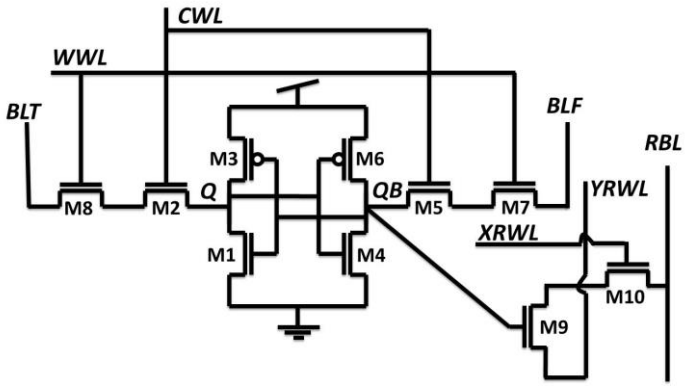


Figure 3-16: Proposed 10T bitcell with XY read port

The schematic of the proposed XY 10T SRAM bitcell is presented in Figure 3-16. This cell comprises two cross-coupled CMOS inverters, two pass-gate transistors in series like in the 10T-MUX bitcell in Figure 3-6 and the XY selection read-port (as in Figure 3-9). The read transistors are used to transfer the internal data to the read bitline (RBLF). As shown in Figure 3-14 the impact of the leakage current of unselected bitcells in a selected column becomes negligible thanks to the applied under-drive technique in the gate of transistor M10 (Figure 3-16) in the XY read-port: this solves read failure issues. The principle of operation of the 10T-XY bitcell is reported in Figure 3-7(b) with the waveforms of main control signals. Table 3.2 presents the setting control signals for each operation mode in the 10T-XY bitcell.

TABLE 3.2 CONFIGURATION OF CONTROLS SIGNALS IN THE 10T-XY BITCELL

Mode/Signals	WWL	CWL	XRWL	YRWL	BLT	BLF	RBL
Hold	GND	GND	V_{DD}	$-V_{DD}$	V_{DD}	V_{DD}	V_{DD}
Read	GND	GND	V_{DD}	GND	V_{DD}	V_{DD}	Floating
Write	V_{DD}	V_{DD}	$-V_{DD}$	V_{DD}	D_{IN}	$/D_{IN}$	V_{DD}

5. Comparison with state-of-the-art bitcells

The proposed 10T SRAM bitcells are compared to the ones in [45], [59] and [52] (respectively Figure 3-1(a), (b) and (c)). All simulation results are obtained using the 28nm CMOS International Semiconductor Development Alliance (ISDA) model card. The device sizing was performed and optimized in the same way for all bitcells. The bitcells have been evaluated on the basis of SNM, WM, Write-Time (WT), Read current and standby leakage current. All Monte-Carlo simulations are performed with 1024 runs at 300 mV supply voltage and in worst-case temperature [-40°, 125°] and process corners. All criteria have been evaluated at three-sigma process variability.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

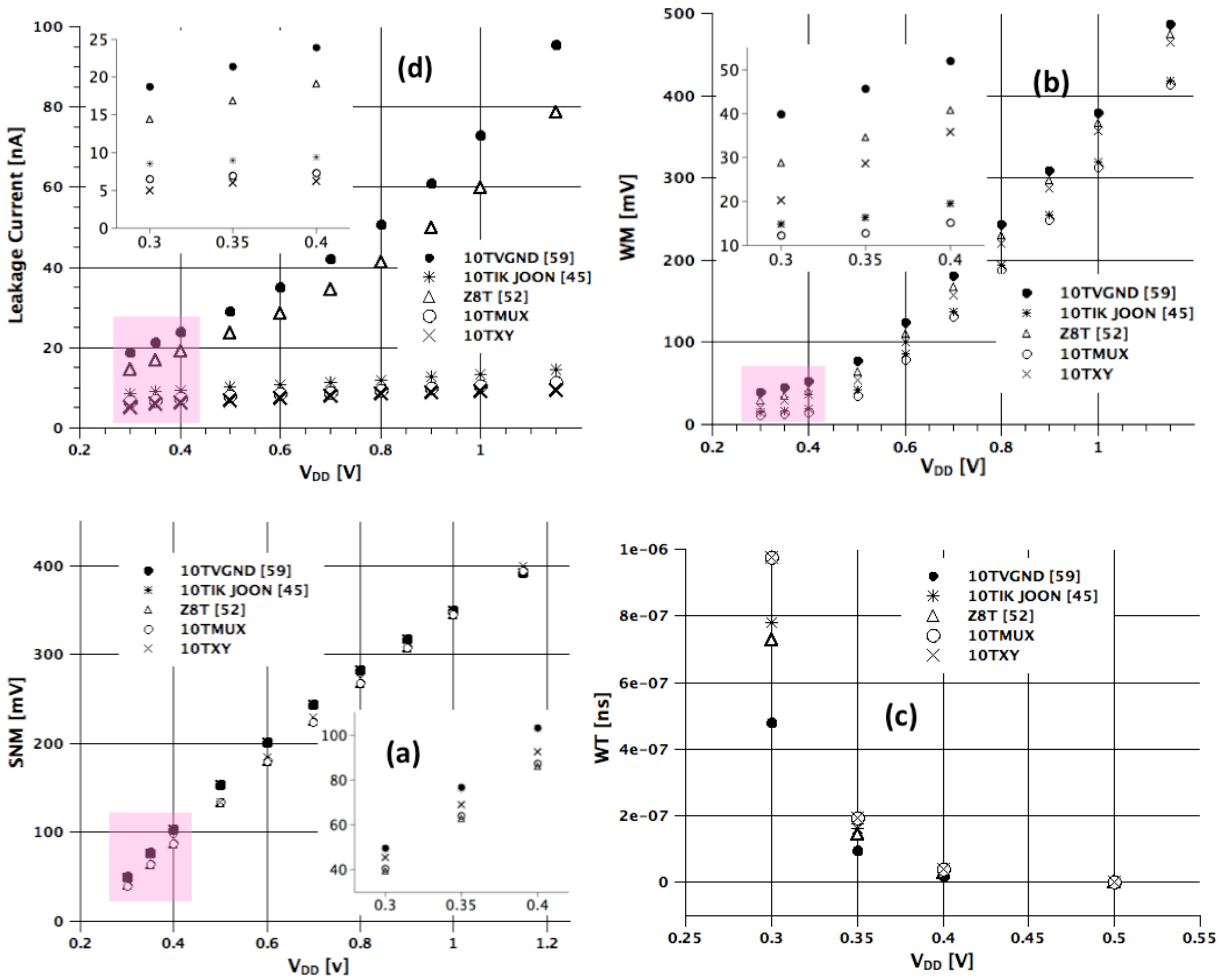


Figure 3-17: (a) SNM (Average – 3 σ), (b) WM (Average – 3 σ), (c) Write time (Average + 3 σ) and (d) Leakage current (Average + 3 σ) for state-of-the-art ULV bitcells and proposed 10T bitcells

Figure 3-17 (a) and (b) show that the two proposed bitcells have comparable performances in terms of SNM and WM with respect to the other bitcells considered in this work. In Figure 3-17(c) a small degradation in term of WT is observed. This is due to the resistive path formed by the two series pass-gate transistors (M5 and M7) or (M2 and M8) in Figure 3-1(a), Figure 3-6 and Figure 3-16 respectively. The analysis of the leakage current in the different bitcells shows that there are two major contributors: the current flowing from the bitline pre-charged to V_{DD} to the internal node equal to GND and the current through the read port. Sources of leakage are weakened in the proposed bitcells thanks to the resistive path formed by the two series pass-gate transistors and the transistor in the read port that has its source and drain at the same potential (RBL = RWL_MUX= YRWL=V_{DD}). As shown in Figure 3-17(d), the proposed bitcells are better in term of minimal leakage current. It is observed that the leakage current is almost constant as function of the supply voltage. Hence the proposed bitcells are able to operate on a wide voltage range without the penalty of a significant leakage current. The 10T-XY bitcell is a good candidate for Ultra-Wide-Voltage-Range SRAM design since it overcomes parasitic power consumptions and the leakage effect (the weakness of the Ion-to-Ioff ratio

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

at ULV) without static power penalty. Simulation results at 300 mV and worst-case conditions are summarized in Table 3.3. A comprehensive comparison is built from quantitative values. The 10T-XY bitcell is highly functional in terms of Ion-to-Ioff ratio and negligible dynamic power losses with respect to the unselected word lines. The 10T-MUX is a differential bitcell with similar performances but better read-time value at the penalty of a larger complexity at bitcell and periphery levels compared to the 10T-XY. The selected state-of-the-art bitcells are less complex, what means less dynamic energy consumption per operation but they are not fully functional in worst-case conditions, i.e. below 400 mV power supply. Read failures are related to the unsatisfying values of the read current (I_{CELL}) what leads to unsatisfying Ion-to-Ioff ratio. The 10T-XY is finally preferable to the 10T-MUX. All bitcells in Table 3.3 have been designed in 28nm CMOS Bulk but only the 10T-XY and 10T-MUX have been experimentally verified here. Table 3.3 presents the total energy per cycle of critical path (array 128x128) of different bitcells at 300mV supply voltage. No peripheral circuits have been included in the simulation. These peripheral circuits may introduce additional power consumption and area penalties.

TABLE 3.3 SUMMARY OF DIFFERENT SRAM BITCELL TOPOLOGIES

	10TVGND [59]	**10TIK JOON [45]**	**Z8T [52]**	**10T-MUX [This work]**	**10T-XY [This work]**
Cell Area: the 6T layout considered as a reference (x1)	**3.5x**	**3.5x**	**2.27x**	**3.5x**	**3.5x**
Controls signals	**3**	**2**	**2**	**3**	**4**
Sensing	**Differential**	**Differential**	**Differential**	**Differential**	**Single**
Ion/Ioff ratio issue @ ULV	**Yes**	**Yes**	**Yes**	**No**	**No**
Writability [mV]	**29.8**	**14.7**	**28.8**	**12.3**	**20.25**
Half-selected immunity	**Low**	**High**	**Low**	**High**	**Very High**
Leakage [nA]	**18.7**	**8.51**	**14.4**	**6.52**	**5.37**
WT [µs]	**0.47**	**0.77**	**0.72**	**0.97**	**0.9747**
SNM [mV]	**39.7**	**49.3**	**39.4**	**40.6**	**45.552**
I_{CELL}[nA]	**1.41**	**1.83**	**5.46**	**5.52**	**5.11**
$E_{Write, Total}$ (pJ)	**0.34**	**0.7**	**0.9**	**0.95**	**0.46**
$E_{Read, Total}$ (pJ)	**1.05**	**1.53**	**2.07**	**1.9**	**0.7**

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

The results must be analyzed on a comparative basis. The 10T-XY bitcell is the best one in term of energy saving since this bitcell avoids the dynamic energy losses and reduces the read time and at the same time the read power consumption. The 10T-XY bitcell is so far the best candidate in the term of leakage current compared to the other bitcells (72% less leakage compared to the 10T VGND).

6. Silicon vs simulation evaluation in 28nm LP bulk

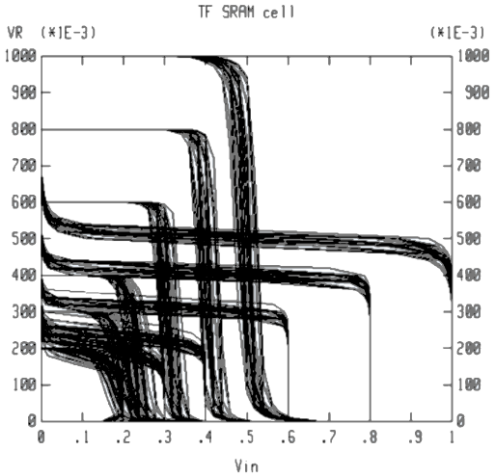


Figure 3-18: Experimental Butterfly curve for the 10-XY bitcell for various supply voltages

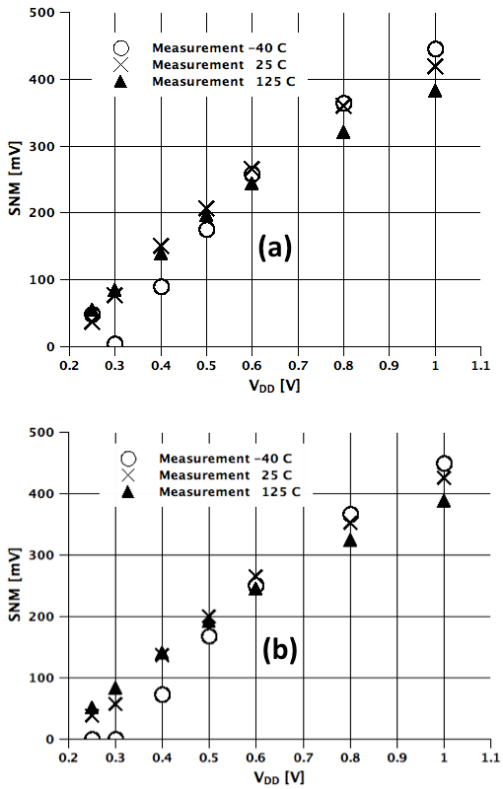


Figure 3-19: SNM measurement versus supply voltage and temperature variation for the 10T-MUX bitcell (a) and the 10T-XY bitcell (b)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

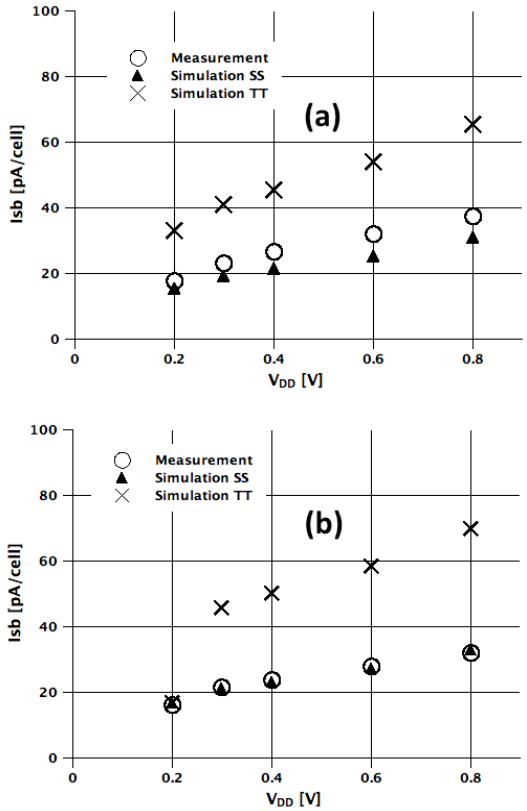


Figure 3-20: Silicon vs simulation standby current for various supply voltages for the 10T-MUX bitcell (a) and the 10T-XY bitcell (b)

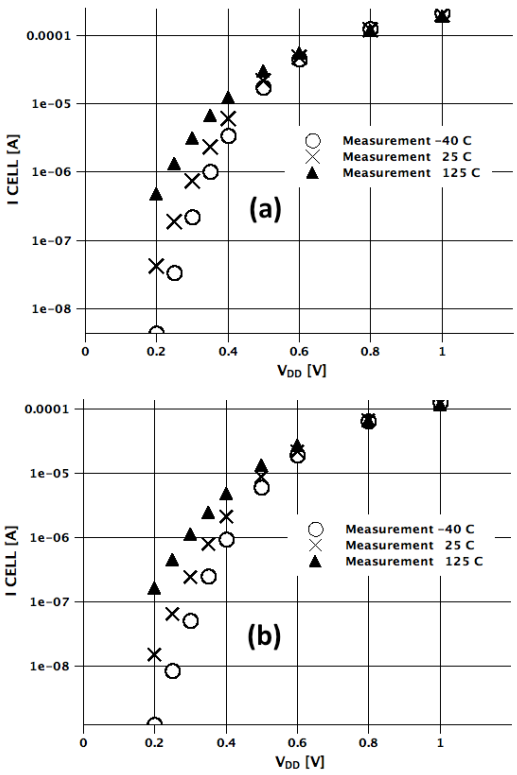


Figure 3-21: Read current measurement for various supply voltage and temperature for the 10T-MUX bitcell (a) and the 10T-XY bitcell (b)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

The proposed bitcells have been processed and characterized over a wide voltage range in terms of SNM, read current and leakage current. Measurements are compared to simulation results. Figure 3-18 shows experimental butterfly curves for the 10T-XY bitcell for various supply voltage at 25°C. Figure 3-19 presents SNM measurements versus supply voltage for various temperatures. The proposed bitcells have more than 50 mV SNM value at 25°C, what guarantees an acceptable margin for readability. The temperature inversion phenomenon appears at a supply voltage around 0.6 V. Figure 3-20 shows the leakage current measurement. Results are in line with simulations in the Slow/Slow corner and current values at ULV are better than the ones of the best in classical bitcells in literature. The temperature variation effect is more critical at ultra-low-voltage as shown in Figure 3-21 with the level of the read current: it results in different behaviors between low and high temperatures. More efforts are required in order to make the functionality independent of the temperature variation.

7. Conclusion

In this chapter two 10T bitcells have been proposed as candidate to alleviate main limitations of in state-of-the-art architectures (high leakage current and dynamic parasitic power consumption) but to offer similar performances in terms of stability (SNM, WM) as the previously proposed ULV bitcells. Dynamic parasitic power consumption is avoided in the half-selected bitcells, thanks to a proposed hardcoding technique and a XY read-port. A comparative study was performed and simulations in worst-case corners yield pertinent results. Silicon evaluation has been performed in 28nm CMOS. The next step is the silicon validation of a 32kb UWVR L1 Cache designed with the proposed10T-XY bitcell.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Chapter 4

Solutions enabling UWVR

Sub-threshold operation of circuits becomes more and more attractive thanks to the ultra-low power consumption at ULV. Unfortunately scaling down the supply voltage of SRAM faces an important limitation in read access time that prevents high frequency operation and the field of possible applications. The access time under ultra-low voltage operation is mainly dictated by the available read current of the SRAM bitcell and the bitline effective capacitance. This chapter presents two contributions enabling ULV operation of the SRAM bitcells in chapter 3. An optimized sense amplifier (SA) is first detailed. A replica circuit is then introduced to take care of the impacts of variability and to provide dedicated timing information for the SA operation.

1. Introduction

Recent literature on ULV SRAMs has shown that the optimal voltage to reach a minimum energy operation is around 300 mV for sub-90nm technologies [69]. Besides ULV induces a limit in the performances in frequency of SRAMs. The read access time is negatively affected by a degradation of the bitcell effective read current and the increase in variability. The most common technique used for read operation is the full swing sensing technique at ULV [70] that is unfortunately a slow process (large delay).

Figure 4-1 presents an example of single ended circuit [71] used to ensure full swing read mode. This technique consists in looking forward to the discharge of the read bitline: the read bitline is pulled down if Q=’1’ (selected bitcell ‘internal data) or the read bitline stays high when Q=’0’. Multiplexer allows to assign one of the bitlines to local Bitline BLX. The discharge of BLX results in the discharge of the global bitline (GBL). There is an interest to optimize sense amplifiers (SAs) for both full-swing and single-ended bitcells under ULV. State-of-the-art SAs are reported for supply down to 500 mV due to the increase in local and global variability with scaling down V_{DD} [54].

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

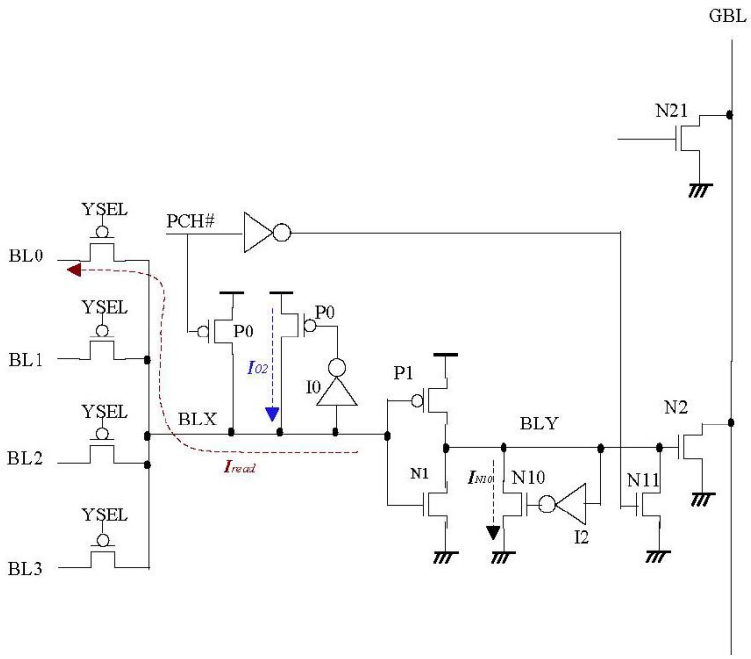


Figure 4-1: Circuit schematic of conventional domino single-ended sensing (full swing) [71]

In this chapter the motivation and the main limitations of sensing data at ULV are presented. An optimized voltage sense amplifier is illustrated allowing to improve the read time of the proposed 10T-XY bitcell. A new replica circuit with PVT tolerance is presented and finally an adaptive technique allowing to optimize the sensing operation is proposed.

2. Read Vs. write operation

During a read operation, the two read bitlines (RBLT and RBLF) of the bitcell are initially pre-charged. After the pre-charge state, both read bitlines are maintained in a floating state and the wordline signal, WL, is pulled high. Depending on the data stored in the internal nodes of the bitcell, one of the read bitlines is driven low while the other read bitline stays high. During the write operation one of the two bitlines is driven to the level of the desired data. WL signal is pulled high, exposing the storage nodes to both bitlines. Since pass-gate transistors can drive a ‘0’ more easily, the write operation begins by writing a ‘0’ first. After Writing ‘0’, the internal feedback between the inverters in the bitcell forces the other node to ‘1’ (Figure 4-6).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

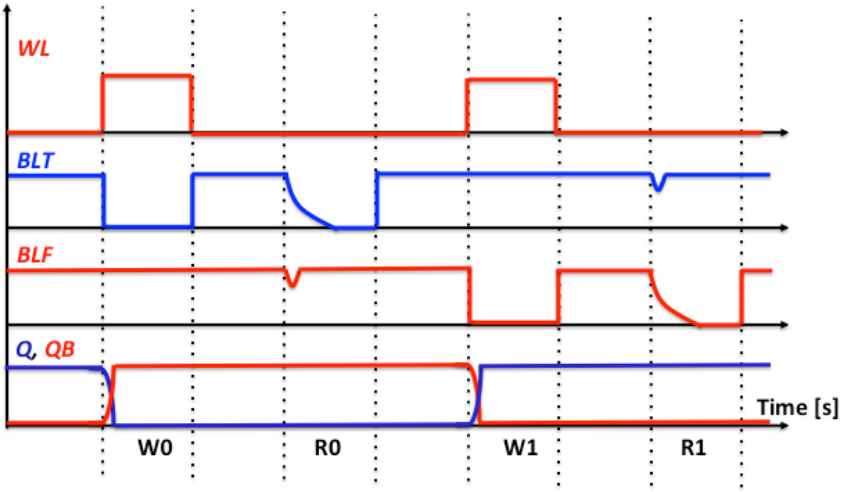


Figure 4-6: Waveforms of main control signals in operation of the 6T bitcell

Figure 4-7 summarizes the behavior of the bitcell using a read port during the read and the write operations. The discharge of the large capacitance of the floating read bitline makes the read operation slower than the write operation.

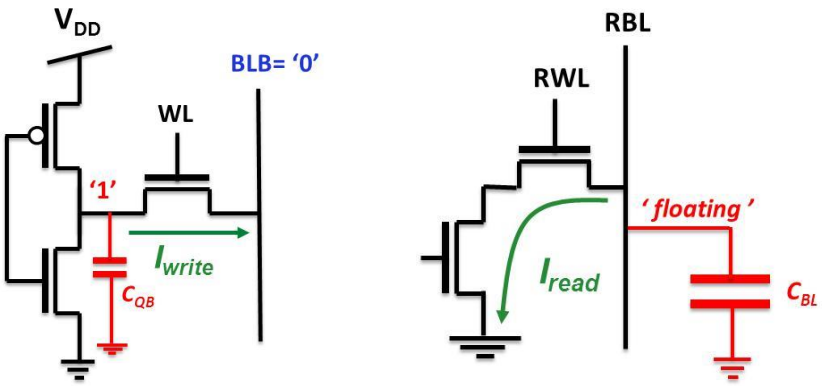


Figure 4-7: Write (left) and read (right) current path in the bitcell using a read port

Figure 4-8 presents a 1024-point Monte Carlo (MC) simulation of the write time and the read time at 3 σ for the critical path (XY=128x64) using the 10 XY bitcell. Figure 4-8 shows that the read time is larger than the write time by 78 % at 300mV and by 70 % at 1V supply voltage. These simulations indicate that the bitcell faces an important limitation in read access time that prevents high frequency operation and the possible applications under ULV.

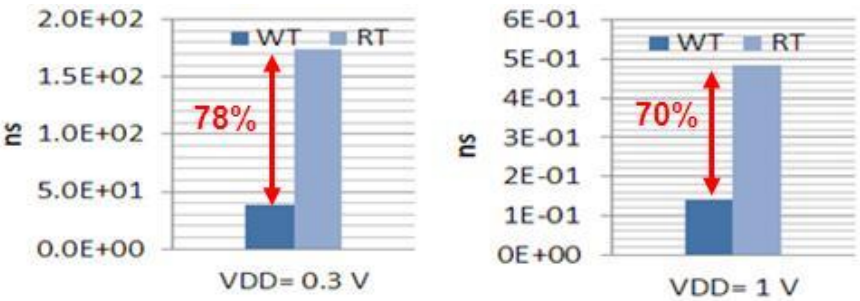


Figure 4-8: Write time WT and Read time RT @3σ for 10T XY bitcell (64L, 128C) in 28FDSOI (TT Process, 25C PVT Conditions)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

3. Limitation in read operation timing in UWVR

Operation of SRAM under ULV is very attractive to reduce consumption. Unfortunately most applications need not only energy efficiency but also acceptable range of speed. However operation frequency is limited by excessive values of read time (compared to values of write time, Figure 4-8). Figure 4-9(a) depicts a read cycle in a conventional SRAM. First the bitline capacitance is pre-charged through a PMOS transistor. Then depending on the stored data, the read bitline is discharged or stays high during the read operation. Finally the sense amplifier is activated and depending on the value of both read bitlines, the output data takes the value “0” or “1”. Figure 4-9(b) illustrates the targeted reduction time in the discharge and sensing components. We target to optimize two time contributions: the discharge and the sensing times.

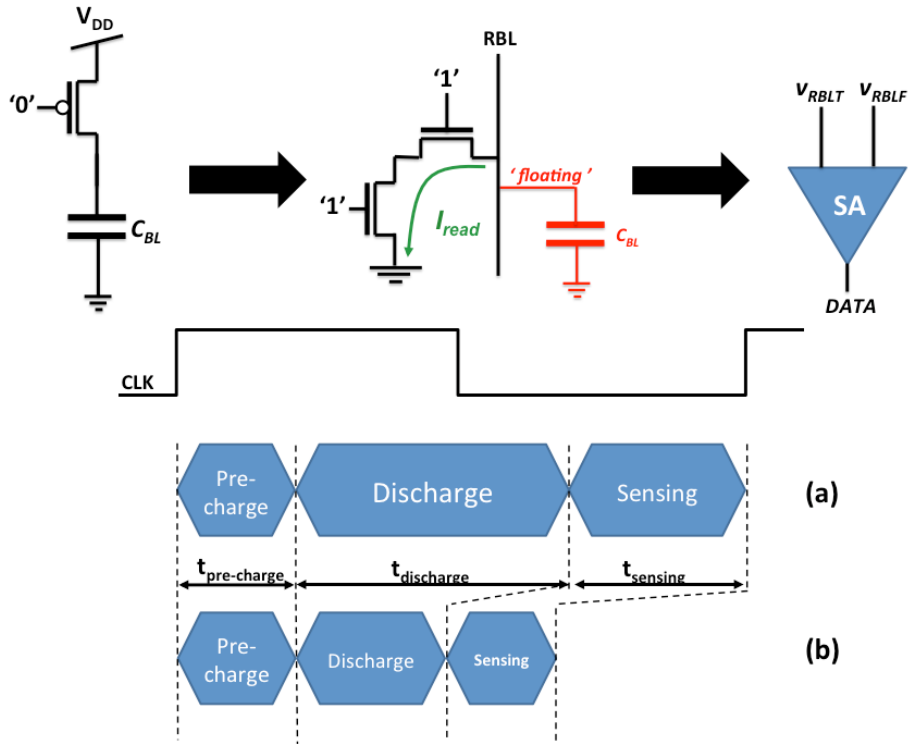


Figure 4-9: Critical path in an SRAM during read operation: (a) standard timing, (b) targeted reduction timing

The acceptable amount of bitline discharge to guarantee a successful read operation is a difficult to predict. Figure 4-10 presents the sensing schemes for small and large signals as examined in [74].

The conventional differential small signal-sensing scheme (Figure 4-10(a)) allows stacking a large number of bitcells per column and therefore presents a large bitline capacitance, thus the need for a sense amplifier capable to operate on a small difference voltage, ΔV (but higher than the offset voltage). At ultra-low supply voltage, the weakness of the I_{ON}/I_{OFF} ratio limits the number of bitcells to the stacked around 128 bitcells per column and the increase in mismatch causes larger offset voltage in a SA what makes the design of a ULV sense amplifier very challenging.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

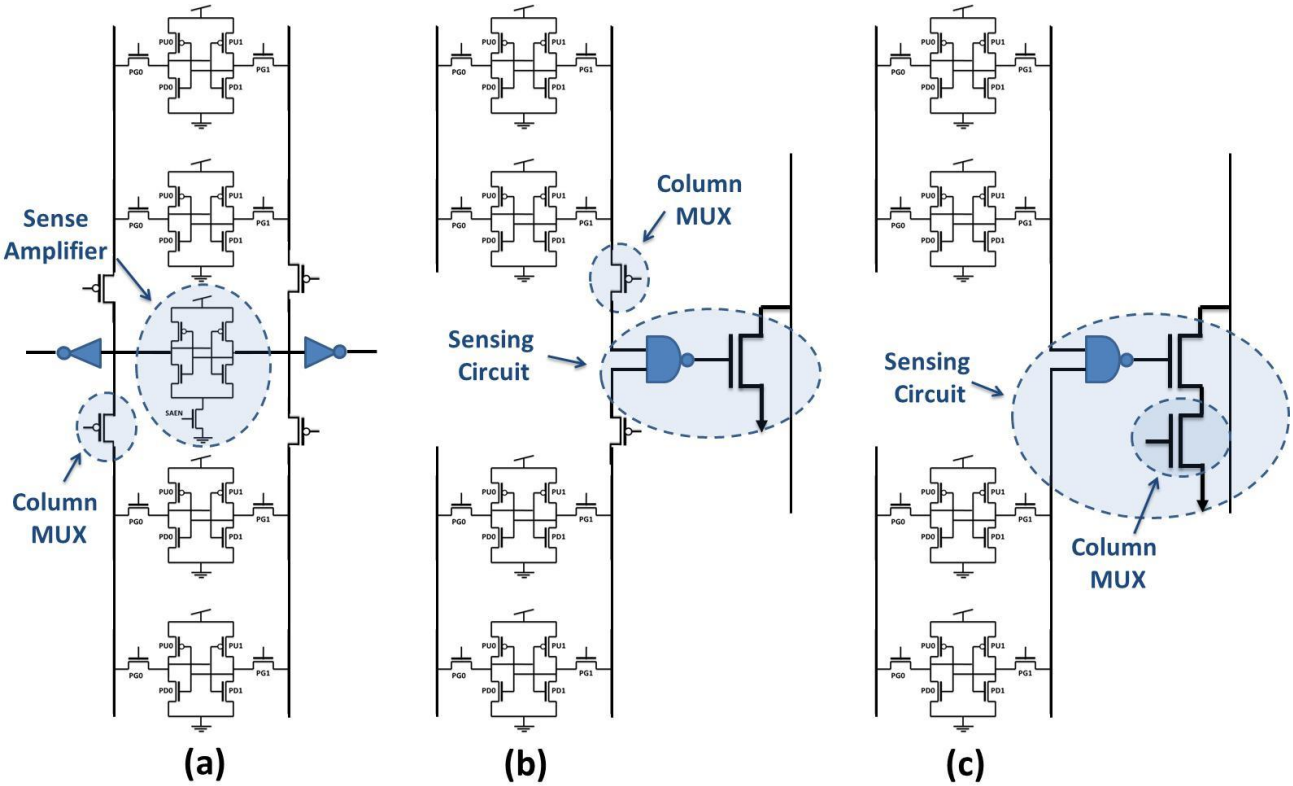


Figure 4-10: (a) Small signal sensing scheme, (b) large signal sensing scheme with multiplexing in local bitline, (c) large signal sensing scheme with multiplexing in global bitline. [72]

In the case of large signal sensing scheme (Figure 4-10(b) and (c)), the read bitline discharges from the supply voltage to ground. The discharge time is correlated to the bitline capacitance, hence the idea to limit the number of bitcells to be stacked. An acceptable read time could be reached. This technique is advantageous since just a simple logic port is needed to recover the stored data what allows to avoid the consumption of SA at nominal voltage. However in near and sub-threshold voltage range, the degradation of the read current and the increase in the variability make the read time impact larger what limits the possible operating frequency. This is why the use of the large signal sensing technique at ULV is not beneficial. The use of an amplifier is then primordial.

$$\Delta t = C_{BL}.\Delta V_{BL} / I_{cell} \quad (4\text{-}1)$$

Equation 4-1 evaluates the discharging time for a differential voltage, ΔV_{BL}. This is a large part in the read time (Figure 4-9). Table 4-1 presents the evolution of the discharging time for differential voltage sense amplifier with various values of ΔV_{BL}, with respect to 300mV and 1V supply voltage. Results have obtained by simulation of critical path (128-bit x 64 bit SRAM array) using a differential 10-XY bitcell as shown in Figure 4-11.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

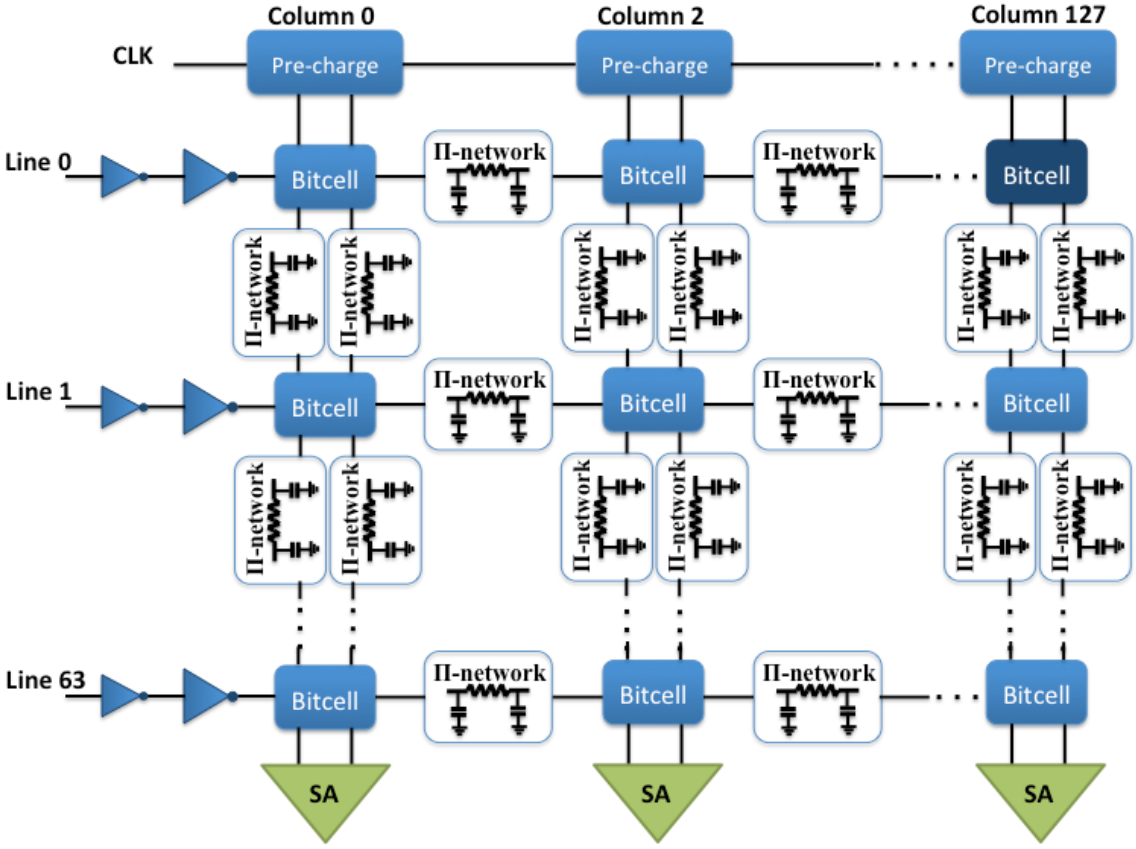


Figure 4-11: The simulation setup of a 128-bit x 64-bit SRAM array

TABLE 4.1 EVOLUTION OF THE DISCHARGING TIME FOR DIFFERENTIAL VOLTAGE SENSE AMPLIFIER

ΔV_{BL} (mV)	50	60	70	80	90	100
Sensing time (ns) ***(TT, 300mV, 25°C)***	*50*	*54.4*	*58.2*	*63.5*	*67*	*70.17*
Sensing time(ns) ***(TT, 1V, 25°C)***	*0.5*	*0.536*	*0.56*	*0.62*	*0.71*	*0.8*

TABLE 4.2 EVOLUTION OF THE READ TIME FOR LARGE AND SMALL SIGNAL SENSING SCHEMES

Large signal sensing	$T_{READ} = T_{PRECH} + C_{BL}.V_{DD}/\ I_{cell}$
Small signal sensing	$T_{READ} = T_{PRECH} + C_{BL}.\Delta V_{BL}/\ I_{cell} + T_{Sensing}$

Table 4.2 illustrates the read time value in the case of the large signal and the small signal-sensing scheme. If we summarize, in near and sub-threshold voltage range, there are two main limitations for large signal sensing technique: first one is the limitation in the number of bitcells to be stacked per column due to the bitline capacitance and the second one is the degradation of the discharge time due to weakness of the read current. So in order to improve the operating frequency at ULV, we have

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

focused solutions and techniques to extend the use of the small signal sensing technique in ultra-wide voltage range [300mV, 1.3V].

4. UWVR small-signal sensing scheme

The scheme in Figure 4-10(a) suffers limitations at ULV. This section introduces solutions to alleviate these limitations: i.e. to enable a satisfying read time without a constraint on the number of bitcells per column.

4.1 Sense amplifier

The sense amplifier is among the most critical component in the SRAM peripheries. The role of the SA is to amplify the differential input voltage, ΔV, developed between the two bitlines of a selected column of bitcells. The minimum signal swing, ΔV_{MIN}, is limited by the offset of the SA. The offset voltage is due to transistor mismatch in the supposedly identical matched transistor pair at the input of the SA. At the time of triggering, the SA differential input voltage should be higher than the offset: $\Delta V \geq V_{OFFSET}$. The performances of the SA directly impact the access read time of the memory and the dynamic energy consumption. As shown in Figure 4-12 the read time with a small-signal sensing technique is composed of two main parts. The bitline discharge time, t_{BL}, depends mainly on the read current, the bitline capacitance and the offset voltage of the SA. A second part of time delay consists in the reaction time of the sense amplifier.

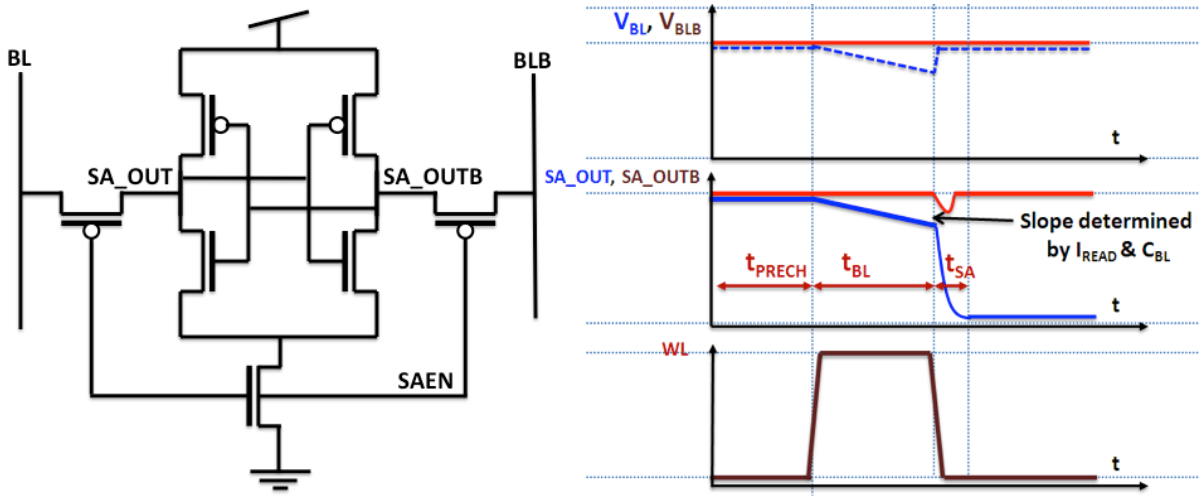


Figure 4-12: Behavior of bitline voltages during read operation (as schematicaly described in Figure 4-9)

Several architectures of SA have been proposed in the state-of-the-art. The latch-type SA [75] [76] is the most widely used in SRAM memory thanks to a favorable trade-off in terms of power consumption and speed (reaction time). Figure 4-13 presents two latch-types SAs. The first one is the VLSA [77] (Figure 4-13(left)) that amplifies the voltage difference, ΔV_{BL}, developed between the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

bitlines. The second one is the current sense amplifier (CLSA) [78] (Figure 4-13(right)) that amplifies a current difference created by the voltage difference developed between the two bitlines.

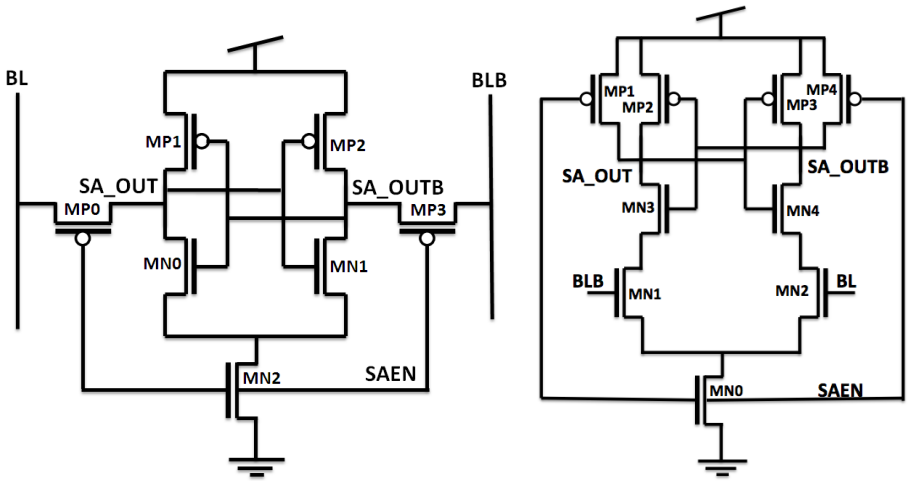


Figure 4-13: Latch type SA: (left) voltage sense amplifier [77] and (right) current sense amplifier [78]

On the one hand the VLSA is better in term of sensing speed and has less layout area than the CLSA (additional two NMOS transistors). On the other hand the CLSA hasn't the same constraint as the VLSA in term of precision needed for the SAEN signal in order to clearly separate the input and output nodes since the output nodes serve as input nodes through PMOS access transistors (MP0 and MP3 in Figure 4-13(left)) [79]. Scaling down the supply voltage the read current decreases exponentially with a large process variation impact as shown in Figure 4-14.

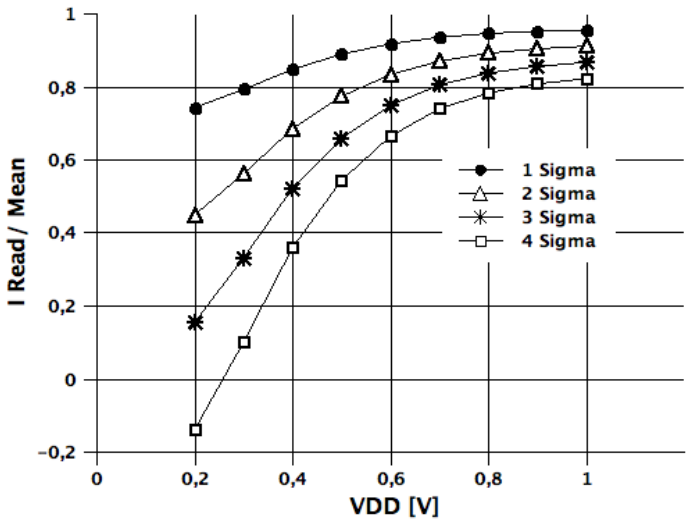


Figure 4-14: Read current degradation in the 10T-XY bitcell in presence of variation with respect to VDD scaling in 28 FDSOI (Monte-Carlo 1024run,TT and 25°C)

Monte Carlo simulations are applied to the ULV sense-amplifiers. Unfortunately there is no convergence point that has been detected for the CLSA because it is prone to more failure probability than the VLSA due to the read current mismatch. This makes the operation of CLSA almost impossible at ULV. This is the reason of interest for the voltage sense amplifier.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

4.2 Optimized UWVR Voltage Sense amplifier

The latch type sense amplifier (Figure 4-15(left)) is widely used in conventional memories since it presents high input impedance to the bitlines and it offers a high voltage gain with a simple circuit. In the following only the latch type VSA is considered.

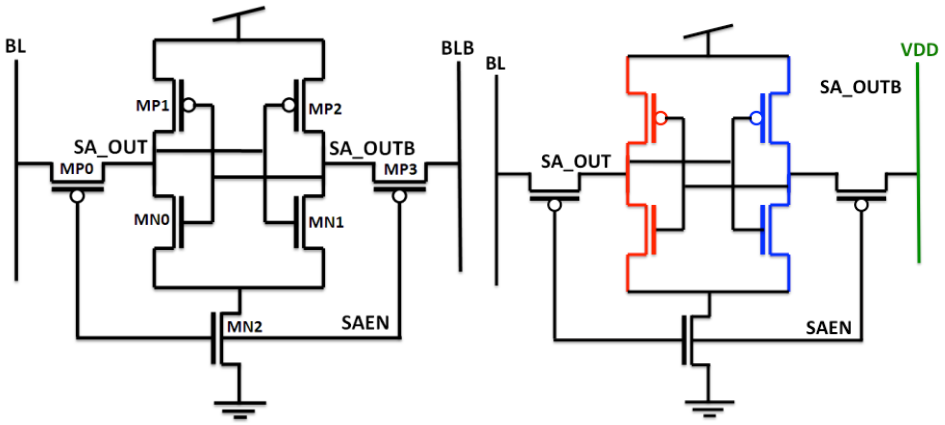


Figure 4-15: Schematic of the Unbalanced VLSA

The proposed 10T XY bitcell has a single ended configuration (one read bitline). Hence, in order to access the internal data of the bitcell in read mode, the use of a single-ended sense amplifier is needed (Figure 4-15 (right)), as it is the case for the DRAM memory. Depending on the architecture of the bitcell, the sensing can be differential as it is the case for the 6T bitcell or single-ended similar to the case of the 8T bitcell (only one read port is used as read path). In order to validate the fesability of the small-signal sensing at subthreshold range, two configurations of the latch-type sense amplifier (diferential and unbalanced VSA) have been designed for single-ended and differential ULV bitcells. The minimum bitline differential voltage is limited by the mismatch offset of the SA. The most widely used technique to reduce the mismatch is ensured by upsizing the critical transistors.

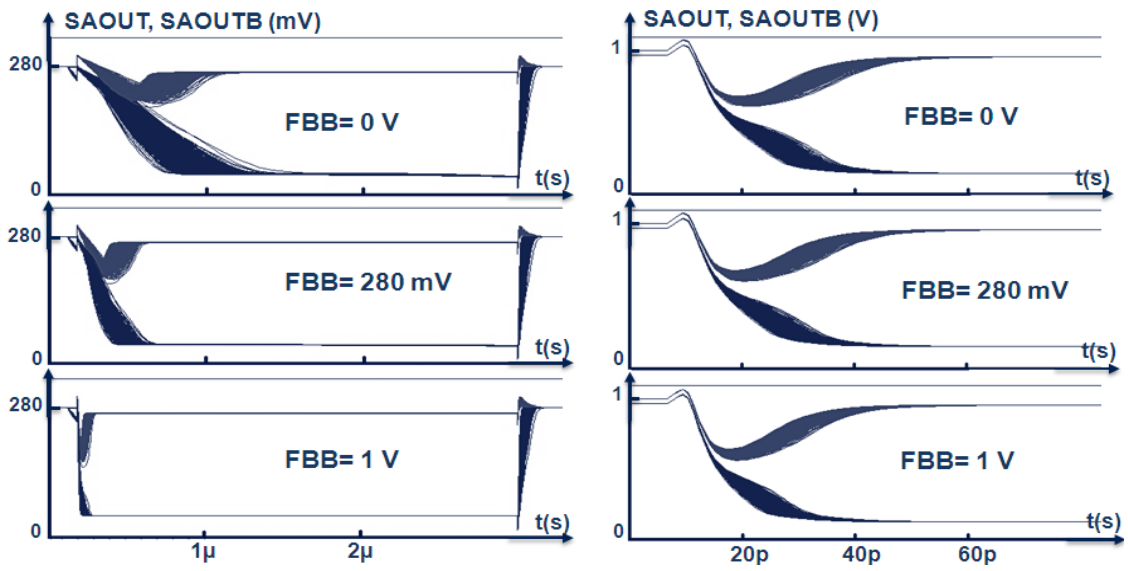


Figure 4-16: Waveforms of internal nodes of the 28 FDSOI differential VSA in Figure 4-12(left) @280mV (left) @1V supply voltage (right) (corner SS, -40°C process and temperature conditions)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

First a ULV differential VSA is considered as in Figure 4-15(left). Internal node waveforms are presented in Figure 4-16 at V_{DD} of 280 mV and 1V respectively. Forward Body-Biasing is evaluated at 0V, 280 mV and 1V respectively. As shown in Figure 4-16(a) the reaction time of the VSA at 280 mV supply voltage is largely reduced (around 80%) thanks to a large FBB at the NWELL of the NMOS transistors level. FBB at nominal supply voltage as shown in Figure 4-16(right) does not impact the SA yield.

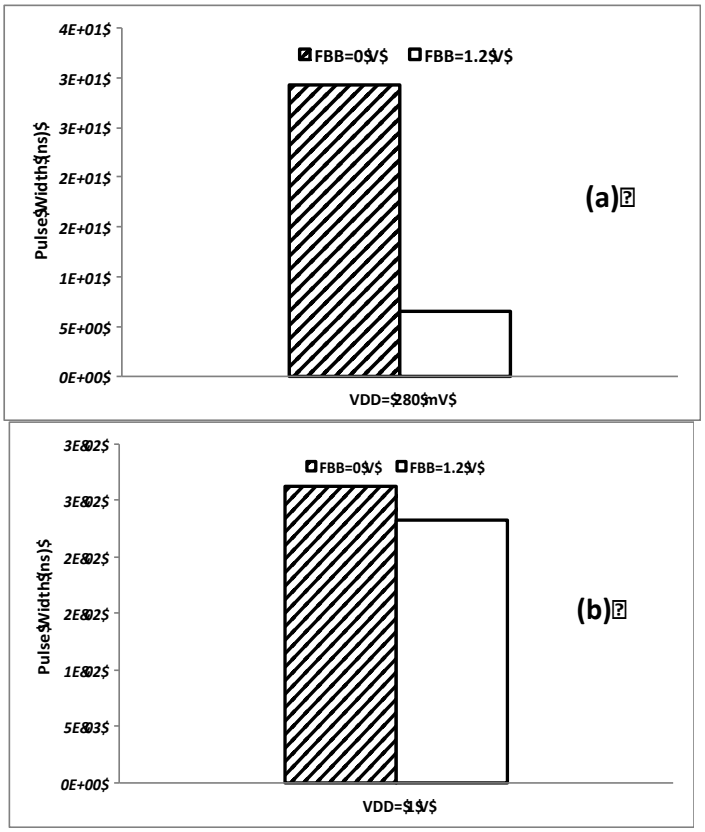


Figure 4-17. Pulse width of differential VSA in Figure 4-14 (left) (a) @280mV and (b) @1V supply voltage

Figure 4-17 confirms the benefit of applying a body bias for improving performances : reduction of the sensing time by 80% at V_{DD}=280 mV and by 13% at V_{DD}= 1V with FBB=1.2V.This illustrates the benefit of body biasing in improving performances of the ultra low voltage circuits.

To estimate the failure probability of a sense amplifier, Monte-Carlo simulations are performed, over 1024 draws. For a fixed targeted value of V_{BL}, the correct sensing operation occurs when ΔV_{BL} is larger than V_{OS} [80]. Then, Failure probability of the SA is estimated as:

$$P(F) = \frac{number\ of\ incorrect\ sensing\ operations}{total\ MC\ simulation\ trials} \qquad (4\text{-}2)$$

The probability of failure depends significantly on the threshold voltage variation. Scaling down the supplly voltage, the Vt variation (σ_{Vth}) becomes larger resulting in more sensing failures. The offset

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

voltage corresponds to the minimum differential input voltage for which the SA achieves the read operation with zero failure in the worst-case conditions. The upsizing of the critical transistors resulted (MN0, MN1 and MN2 in figure 4-15 (left)) eliminates the read failures due to the increase in the mismatch offset at low supply voltage. Figure 4-18(a) and (b) present the Probability of failure for an optimized differential sense amplifier at 280 mV and 1V respectively. As a result the optimized differential SA in 28nm FDSOI has an offset value equal to 40 mV at V_{DD}=280 mV and 30 mV at V_{DD}=1V which remains below to the common standard offset voltage value in the state-of-the-art (50mV).

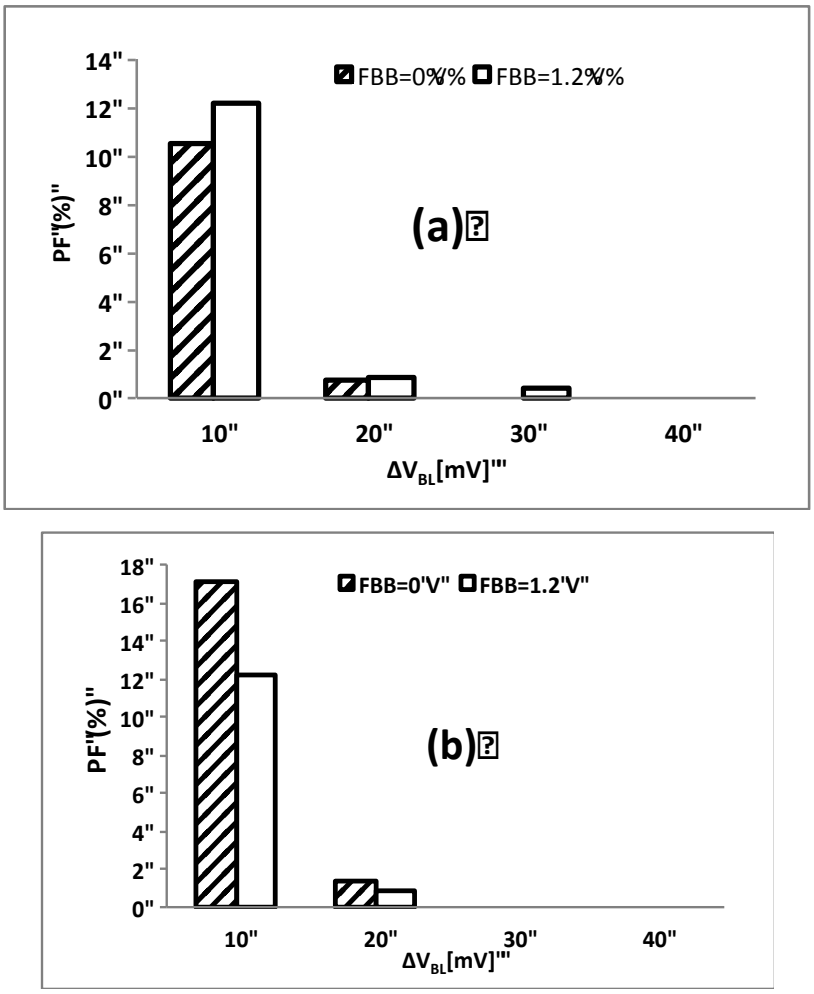


Figure 4-18: Probability of failure for the differential VSA versus effective ΔV at VDD (a) 280mV and (b) 1V power supply (1024 MC Runs)

An unbalanced single-ended VSA is then designed and optimized in 28nm FDSOI for the proposed single ended 10T-XY bitcell (Figure 4-20(b)). Unbalanced VSA is more sensitive to variability due to its dissymmetrical architecture. This explains the increase in V_{Offset} compared to the differential VSA as shown in Figure 4-19. The SA single-ended offset value is equal to 100 mV at V_{DD}=280 mV and 60 mV at V_{DD}=1V when the circuit is optimized in 28nm FDSOI technology.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

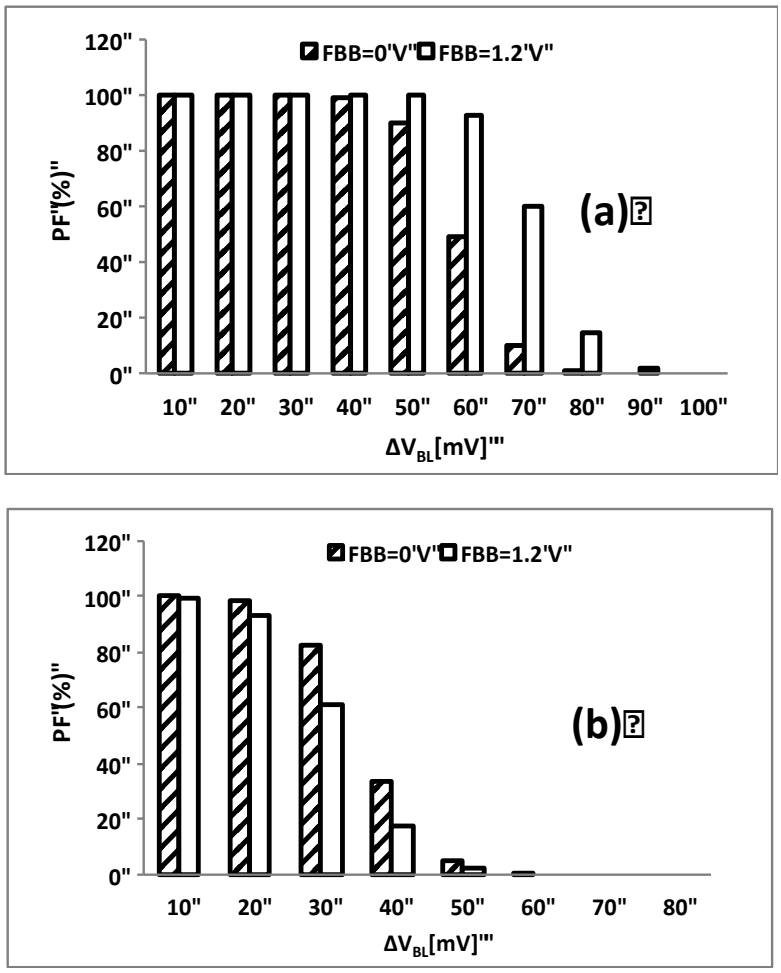


Figure 4-19 Probability of failure for the unbalanced VSA versus ΔV at (a) 280mV (b) 1.2V power supply (1024 MC Runs)

In order to validate the functionality and to estimate the energy consumption of both VSAs (differential and unbalanced VSA), Monte Carlo simulations are performed to evaluate a critical path (with various scenarios of number of bitcells stacked per column: 32, 64, 128, 512, 1024) in the ULV single ended 10T-XY bitcell (Figure 4-20(b)) and the differential 10T-XY bitcell (figure 4-20(a)).

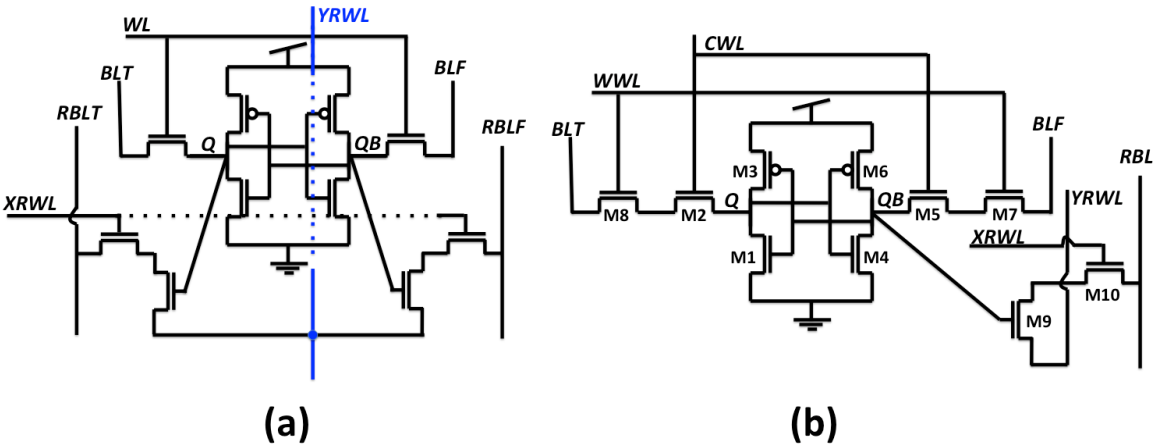


Figure 4-20: schematic of the differential 10T-XY bitcell (a) and the single ended 10T-XY bitcell (b)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Table 4.3 presents the energy evaluation of both optimized SAs during the read operation. According to table 4.3(a) applying FBB=1V results in increasing the total energy consumption by 36,15% compared to the case when FBB=0V for the differential VLSA and by 9,7%for the single ended VLSA. Consecontely, applying FBB allows to improve the reaction time of the SA (read acces time) but it also costs an increase in the energy consumption. Hence there is a trade-off between energy consumption and frequency that must be taken into account.

TABLE 4.3 TOTAL ENERGY CONSUMPTION EVALUATION OF OPTIMIZED SAS

Architectures	**(a) Total energy consumption for different body bias value (fJ)**			
	FBB = 0	**FBB = 0.5**	**FBB = 1**	**FBB = 1.2**
Differential VLSA	**0.39**	**0.425**	**0.531**	**0.615**
Single ended VLSA	**0.462**	**0.471**	**0.507**	**0.579**

TT 25 0.3V (C_{BL}=0.256pF & ΔV= 100 mV)

Architectures	**(b) Total energy consumption for various bitline capacitances (fJ)**				
	C_{BL}=0.256pF	**C_{BL}=0.512pF**	**C_{BL}=1.024pF**	**C_{BL}=2.048pF**	**C_{BL}=4.096pF**
Differential VLSA	**0.531**	**0.572**	**0.593**	**0.61**	**0.75**
Single ended VLSA	**0.507**	**0.594**	**0.615**	**0.69**	**0.71**

TT 0.3V 25 (FBB= 1V & ΔV= 100 mV)

Architectures	**(c) Total energy consumption for different V_{DD} (fJ)**			
	V_{DD} = 0.3	**V_{DD} = 0.6**	**V_{DD} = 0.8**	**V_{DD} = 1**
Differential VLSA	**0.531**	**1.74**	**2.8**	**4.1**
Single ended VLSA	**0.507**	**1.866**	**2.88**	**4.14**

TT 25 (C_{BL}=0.256pF & FBB= 1V & ΔV= 100 mV)

According to table 4.3(b) increasing the number of bitcells stacked per bitlines, result in the rising of bitline capacitance and hence the increase of the discharging time to reach ΔV_{BL}= 100 mV: consecontely the energy consumption increase. Table 4.3(c) shows that scaling donw the suplly voltage from V_{DD}=1V to V_{DD}=300mV result in reduction of energy consumption by factor of 8,2x for the differential VLSA and by 8,16x in the case of the single ended VLSA.

5. Discharge time of the bitline

Equation (4-1) indicates that the bitline's capacitance and the read current are both main parameters that impact the discharge time of the bitline in order to develop a given differential input voltage to the SA higher than its offset voltage.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

5.1 Impact of the capacitance on the discharge time:

The bitline capacitance rises with the increase of the number of bitcells stacked per column and the read current depends on the size of the bitcell. Figure 4-21(right) presents the probability density function (PDF) of the read time for various number of 10T-XY bitcells stacked per column (32, 64, 128 and 256 bitcells). This figure illustrates the increase in the variability and the spread of the read time with the increase in the read bitline capacitance. For two reasons. The weakness of the I_{ON}/I_{OFF} ratio and the large read time (Table 4.2) value at ULV, leads to a choice to limit the number of bitcells stacked per column to 64 in the design of ULV SRAM, which will be presented in the chapter 5.

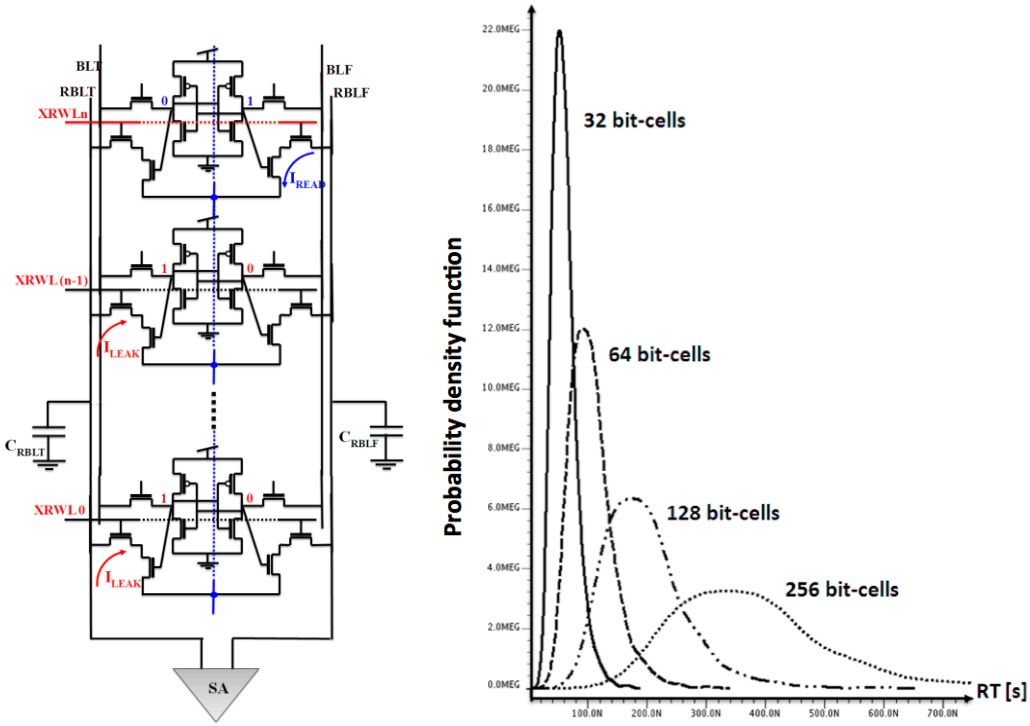


Figure 4-21. Probability density function of the read time for various scenarios depending on the number of cells per column (Monte-Carlo 1024 runs,VDD= 300mV, TT and 25°C)

5.2 Dynamic modulation of V_{TH} in 28nm FDSOI technology

As already mentioned at the beginning of this chapter, back biasing can be used dynamically to increase or decrease the threshold voltage. An evaluation of the impact of applying forward body bias on the bitcell performances is presented. Applying FBB at the bitcell level decreases the threshold voltage but the read current is largely increased what results in the decrease in the read time (rising of the driving current). As shown in Figure 4-22, applying FBB = 1V allows to reduce the read time by 85% compared to the case where FBB voltage value is to 0V.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

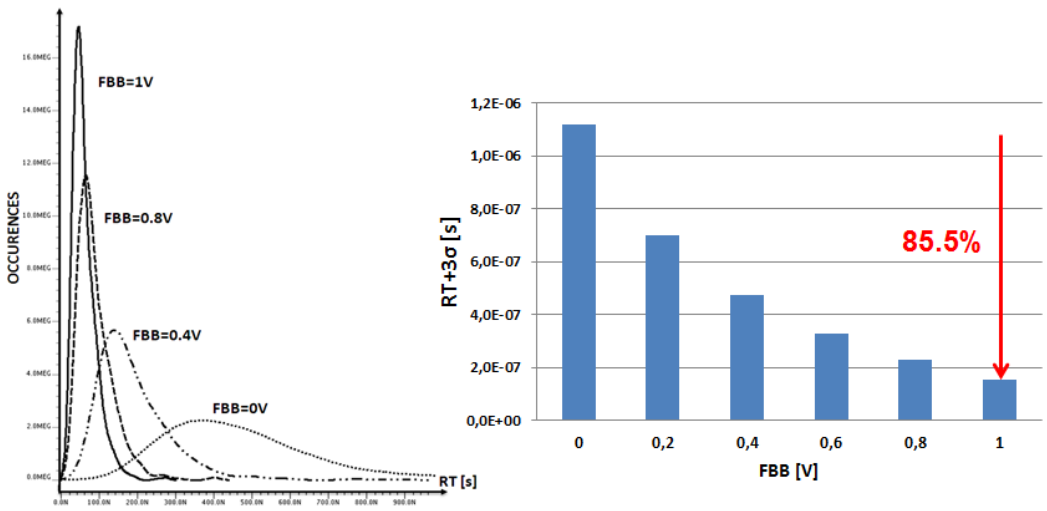


Figure 4-22. Probability density function of the read time for various voltage value of body bias (10T-XY bitcell, Monte-Carlo 1024 runs, VDD= 350mV, SS and -40°C, corner conditions)

Unfortunately, forward body bias technique causes the increase in the total leakage current in the bitcell. Figure 4-23 illustrates that applying FBB = 1.2V causes the increase in the total leakage current of the 10T-XY bitcell by 80% at V_{DD}=300mV and by 56% at V_{DD} = 1V compared to the case where FBB voltage value is 0V. So, if we summarize forward body bias impacts, it appears that the threshold voltage decreases what results in the increase in the driving current (improving of frequency) but the leakage current increases what will result in static power consumption penalty. The traditional ways to limit the power consumption penalty due to the use of forward body bias is firstly to consider a dynamic technique. For example FBB will be applied only during the read operation. Second it is possible to split the memory into 2, 4 or more matrix blocks of bitcells where FBB will be applied only for the selected block. These techniques are limited since the power consumption penalty is still significant [81]. An alternative solution must be explored.

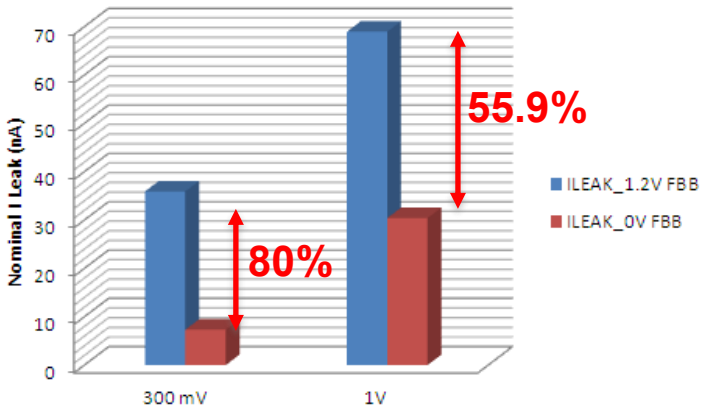


Figure 4-23: Simulation of the total leakage current in the 10T-XY bitcell at 300mV and 1V supply voltage respectively in the case of two body bias values (0V and 1.2V) (Monte-Carlo 1024 runs, FF and 125°C corner conditions)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

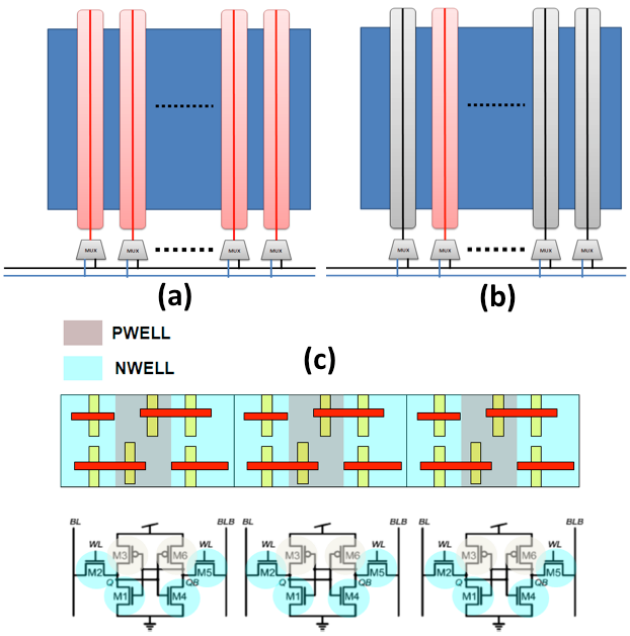


Figure 4-24: (a) FBB per block, (b) Proposed dynamic threshold voltage modulation per column and (c) layout view of NWELL-PWELL intersection

Figure 4-24(b) presents the proposed dynamic threshold voltage modulation per column, which consists in applying forward body bias on the NWELL only for the bitcells in the selected column. This technique is achieved through the combinations of select in multiplexers and a read clock, what allows selecting the straps corresponding to the NWELL layer for the selected column. This technique allows reducing significantly the power consumption penalty compared to the dynamic FBB technique, which apply FBB by block (Figure 4-24(a)). The proposed technique applies FBB on the NWELL strap in the selected columns: however as shown in Figure 4-24(c) for each bitcell, there are two NWELL straps. Hence two columns that are close to the selected column will be half-selected (one NWELL strap without FBB and the other one with FBB). An analysis of the static noise margin of the half-selected bitcell (SNM_{HOLD}) in term of FBB has been performed. This study confirms that there is no negative impact in maintaining data in the half-selected bitcells. On the other hand, it is essential to check the behavior of read stability of the selected bitcell facing the application of forward body bias. Figure 4-25 presents the evaluation of the SNM for the 10T-XY bitcell for various values of FBB and supply voltages. The Figure demonstrates that excessive value of FBB may cause a degradation of the static noise margin especially at ultra-low voltage what will generate read failures. The 10T-XY bitcell suffers from read failures when FBB > 1V for 350mV supply voltage. This settles FBB value to 1V over UWVR.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

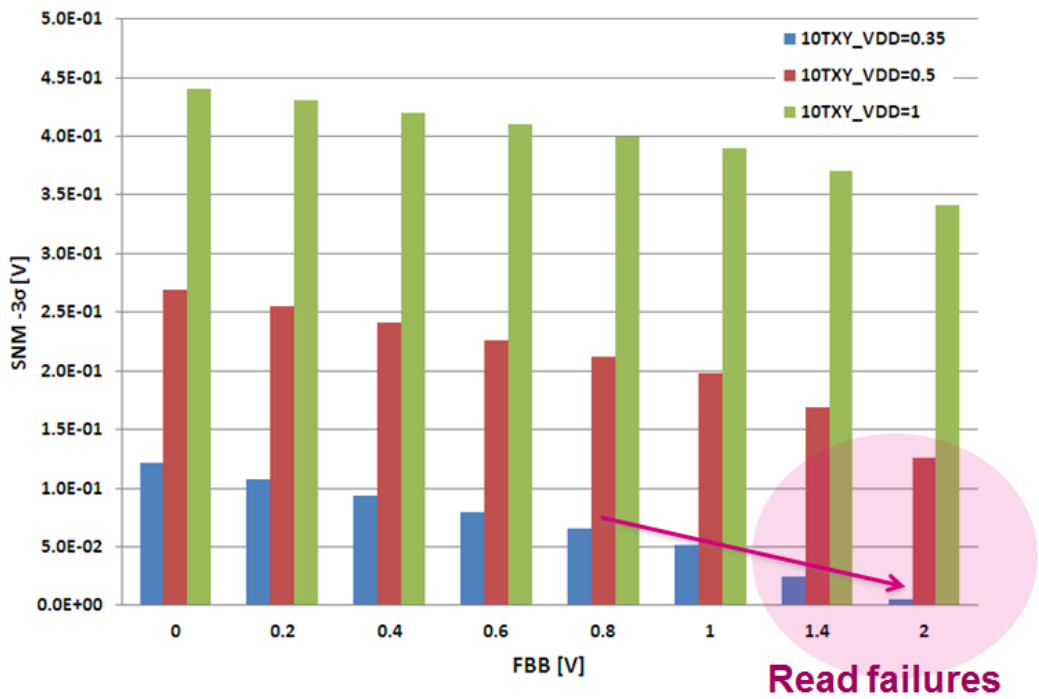


Figure 4-25. Static noise margin fo the 10-TXY bitcell with various value of FBB in the case of 0.35V, 0.5V and 1V supply voltage respectively(-40C_FS_MC1024)

6. Replica circuit

Scaling down supply voltage results in increase in the threshold voltage variation what directly impact the SRAM performances. A sense amplifier and a replica circuits are considered as a best solution to minimize PVT variation impacts on the speed and performances. In the previous section an unbalanced voltage sense amplifier is designed operating down to 280 mV supply voltage allowing to improve the read time of the proposed single-ended 10T-XY bitcell. This section addresses the replica circuit, which is essential to emulate the worst-case timing path for read operation.

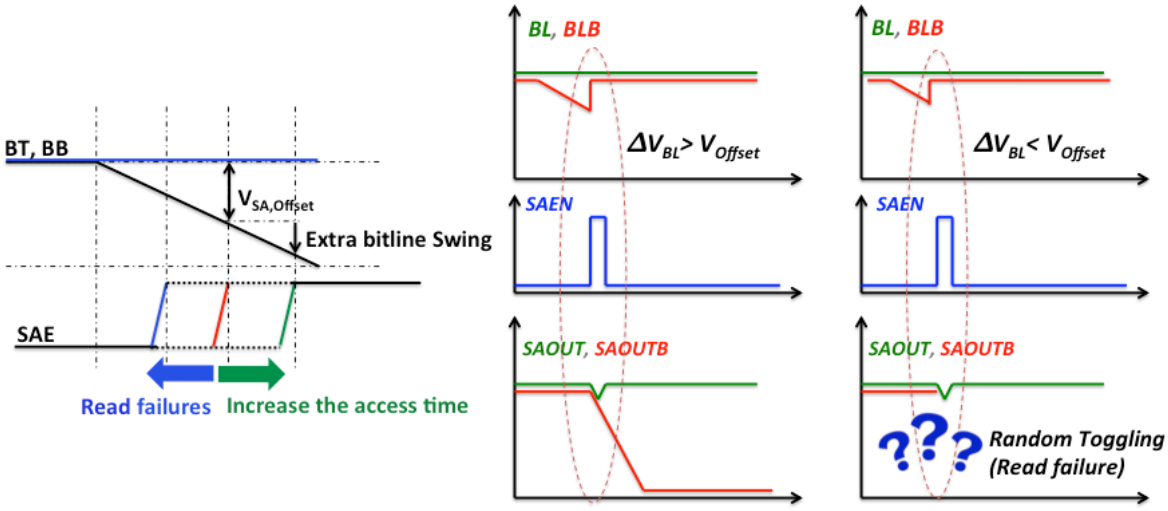


Figure 4-26: Scenario of SA activation with SEAN signal

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

The access time of the SRAM memory is the main limit of the performance in the SoC. The SRAM access time depends on the read current of the weakest cell and the largest bitline capacitance. In order to improve SoC performances, a fast SRAM access time is required [82]. The discharging time of the bitline capacitance by the selected bitcell during the read operation is a dominant part in the access time (Figure 4-9). A sense amplifier is used in order to reduce the bitline related delay amplifying the little voltage difference developed by the bitlines. A small bitline swing reduces the access time and the dynamic power consumption. An SA-enable (SAEN) signal is needed to control the activation timing of the SA. As shown in Figure 4-26, if the SAEN signal fires before the differential input voltage exceeds the SA offset voltage, $\Delta V_{BL} < V_{Offset}$, the SA will randomly amplify the difference voltage and may causes read failures. In the other case, if the SAEN fires too late ($\Delta V_{BL} > V_{Offset}$), then the dynamic power consumption and access time increase. Hence there is an optimal timing to enable the sense amplifier. This optimal timing depends on the global and local PVT variations process, voltage and temperature (PVT) [83]. The replica bitline technique is supposed to generate the optimal SAEN timing at each PVT condition. Replica circuit schemes have been frequently used in embedded SRAM for word-line pulse and sense amplifier Controls, to reduce the timing skew in data sensing for synchronous SRAMs under PVT. In this section we will analyze the different issues related to the replica circuits and a new replica circuit will be proposed.

6.1 State-of-art

Replica tracking circuit is firstly introduced in [84]. Several replica architectures are presented in the state of the art [83], [85], [86] to emulate the worst-case timing path for read and write operation. Figure 4-27 presents a conventional SRAM replica circuit [87].

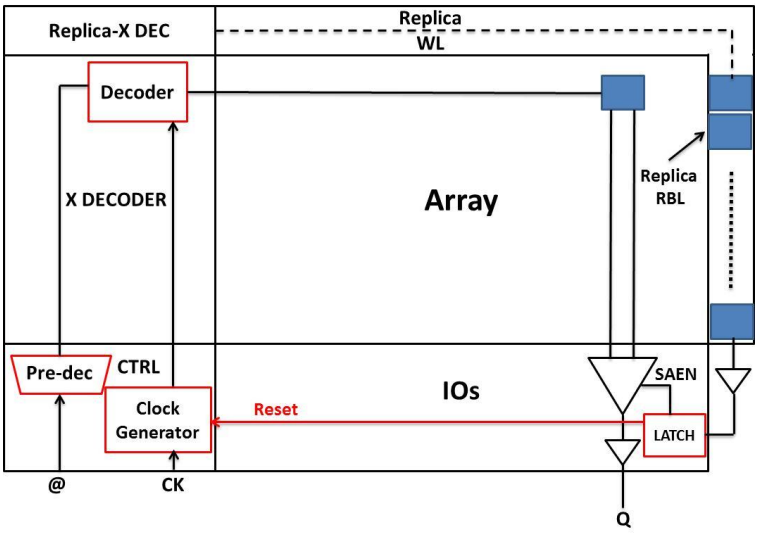


Figure 4-27: Conventional replica circuit

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

The signals are as follows (Figure 4-26): the input address is decoded and leads to the assertion of the word line corresponding to the selected bitcell and the assertion of the replica wordline. The selected bitcell starts to discharge the read bitline. In the same time a fixed number of bitcells start to discharge the replica bitline: the read bitline signal is then inverted and buffered to generate the SAEN signal that will trigger the SA. Finally the SA amplifies the difference voltage ΔV_{BL}, which must be higher than the SA offset voltage. The read bitline replica signal is also used to turn-off the active WL to stop the bitline swing and save power [83]. The conventional replica column architecture uses replica bitcells, which are identical to the core bitcells (Figure 4-27). These replica bitcells are arranged for replicating bitlines capacitance and to emulate the discharge current. Firstly the replica read bitline is pre-charged to V_{DD}. Then WL signal is activated and a fixed number of bitcells discharges the replica bitline. At the same time, the normal selected bitline is also discharged through the accessed bitcell [82]. The timing for SAEN is defined when the difference between the pair of slowest bitline voltages in the selected column (critical path) becomes larger than the offset voltage of the sense amplifier. Usually the impact of the random threshold voltage variation at the level of the bitcell and the replica bitline is different. As a result the delay variation in the replica bitline due to V_{TH} variation increases the SRAM access time. As shown in Figure 4-28 there is a significant impact of the SAEN timing variation performances as it increases the access time and bitline power consumption [82].

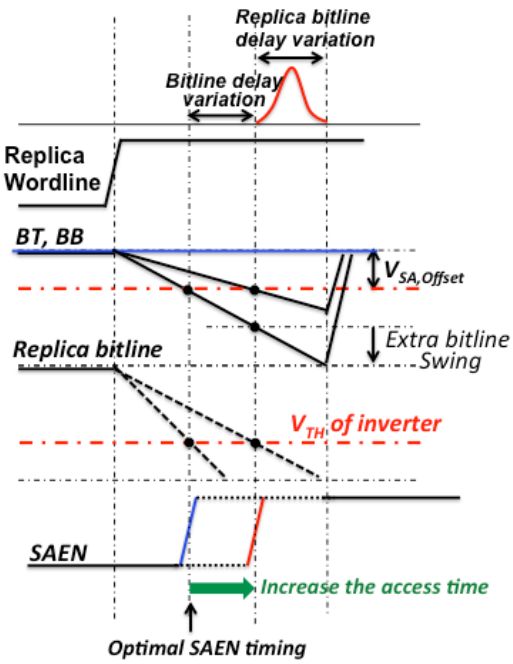


Figure 4-28: Increase of the access time due to the increase in the SAEN variation

The delay for the read operation is composed firstly by the logic gate delay in the decoder, the RC delay in the wordline and finally the bitline discharge delay driven by the selected bitcell. The shift delay in the logic gate caused by the PVT variation is different compared to the bitline delay. Due to the V_{TH} difference among transistors in logic gates and bitcells, a different rate of change appears in

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

delay with respect to V_{DD}. There are three delay paths in the read operation: RC delay in the replica WL, read path to discharge the bitline and a control path to generate SAEN. The delay mismatch of these paths mays induces a fail of read operation or degradation of performances. The replica bitline discharge delay caused by the mismatch can be reduced by increasing the number of so-called driver bitcells in the replica column [83]. Figure 4-29 illustrates the probability distribution of delay in the bitline and SAEN signals: a trade-off between the SAEN timing variation due to the local transistor mismatch and the bitline delay must be made to avoid the read failures and the degradation of performances.

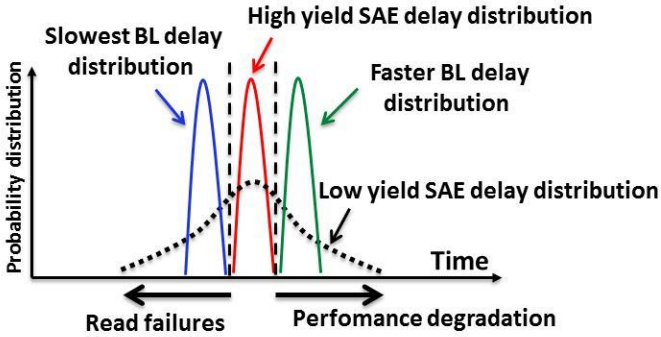


Figure 4-29 Probability distribution of BL and SAEN delay [83]

Figure 4-30 presents a configurable replica bitline (C_{RBL}) technique proposed in [83] for controlling SA enable, allowing the cancelation of local mismatch (RBL delay variation).

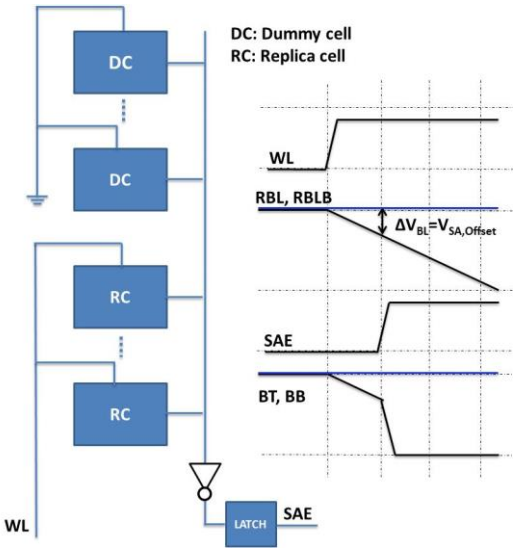


Figure 4-30: Conventional timing replica circuit and SEA timing waveform [82]

The principle of replica bitline technique is that the control path has a delay driven by bitcells as the replica bitline delay. The delay driven bitcells in the control path must have the same length as the read path. Therefore the delay shift of control path according to PVT variation yields the same ratio as the one of the read path [82]. The replica bitline technique attains self-timed tracking with optimal SAE timing according to PVT variation. The replica bitline allows emulating the worst-case delay of

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

discharge of the read bitline capacitance. The word line replica allows emulating the RC delay in the line. The combination of the replica WL and the replica RBL allows tracking the worst case in read and writing timing path in order to provide an optimized timing for SAE and RESET signals. Figure 4-31 (c) presents the normalized delay distributions of the read bitline in [83] that select 3 bitcells among 5 potential driver bitcells (Figure 4-31 (b)) and 3 among 10 potential driver bitcells that best cancel the mismatch compared to the conventional RBL with 3 fixed driver bitcells (Figure 4-31 (a)). As shown in Figure 4-31 (c) the standard deviation of RBL delay decreases exponentially with the number of configurable driver bitcells.

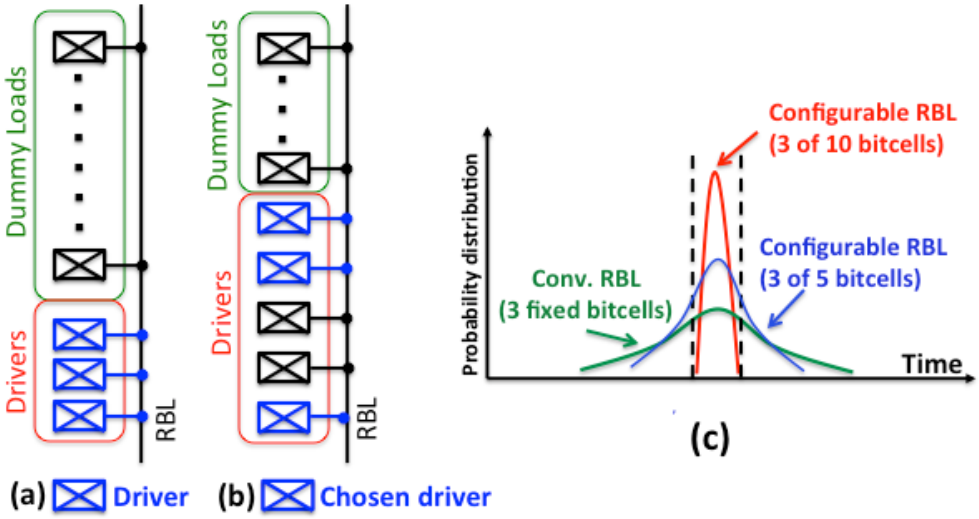


Figure 4-31 (a) Conventional RBL replica with 3 fixed driver cells, (b) Conventional RBL replica with 5 potential driver cells, (c) Probability distribution of the replica BL delay [83]

6.2 Replica circuit

Replica column schemes have been frequently used in embedded compatible SRAM for word line pulse and sense amplifier controls to reduce the timing skew in data sensing for synchronous SRAMs due to the PVT variability. The conventional replica column schemes generate a single set of timing applied to both read and write operations. However in dual port and bitcells with separated read port, the read and write paths are separated. Based on this separation, a new replica is proposed as follows to remove the wordline replica and the decoder dummy cells while ensuring the same functionality as the standard tracking replica with similar performances. The write path, used in read operation is considered in the replica scheme.

Figure 4-32 shows the operating principle of the proposed tracking replica circuit. First the wordline replica is removed, considering that the bitcell has two signals, rwl (for read) and wl (for write) (as the case of the standard 8T bitcell in Figure 4-34) to activate the read and the write operations respectively. Four multiplexors at the inputs of the top two wordlines in the matrix are used. It exists two scenarios that define the operation of the proposed replica:

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

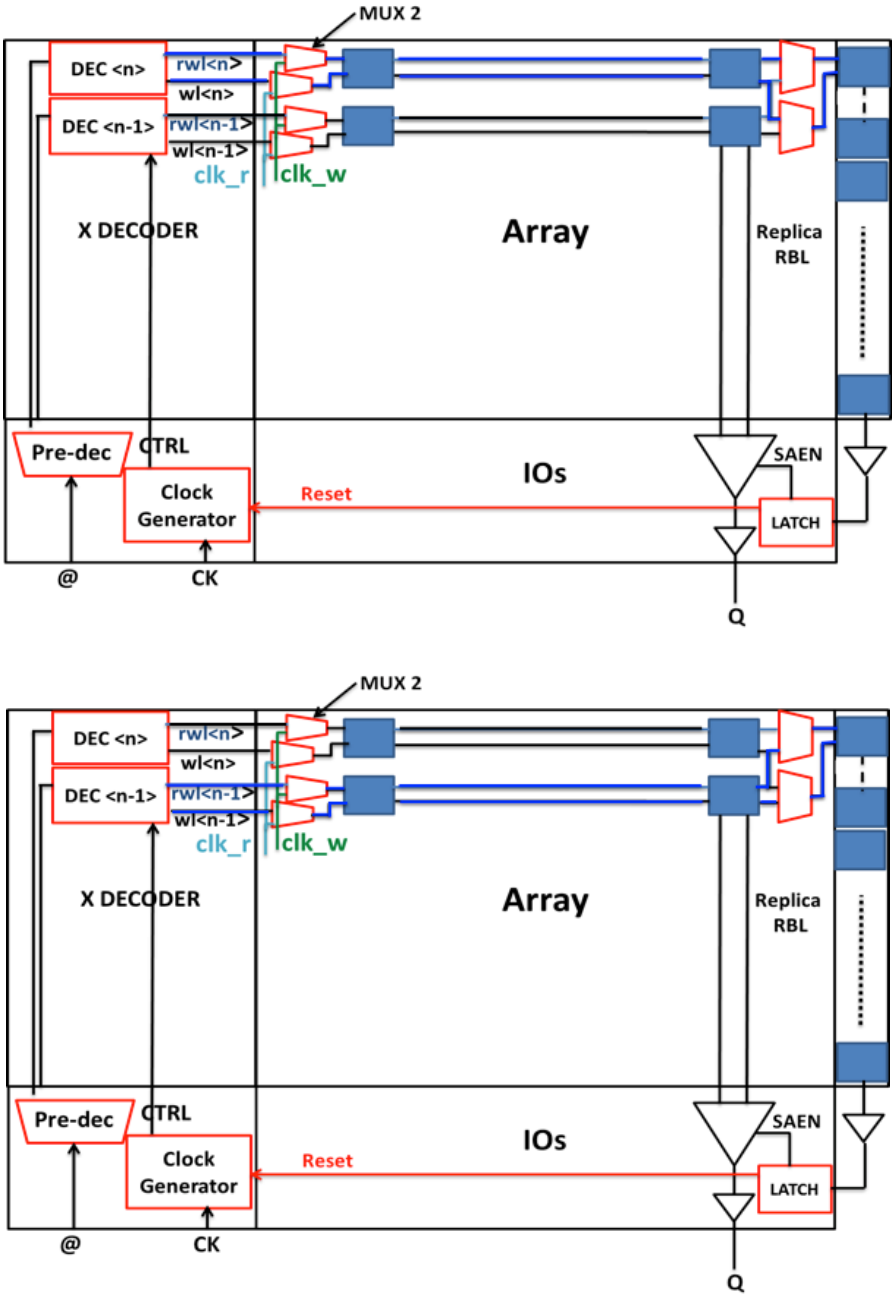


Figure 4-32 Proposed replica circuit

- If the top line <n> is not selected then it is used as a wordline replica as illustrated in Figure 4-32 (top). In the case of read operation, the read clock is propagated in the write path (wl) to provide the rwl_dum signal (which corresponds to the propagated signal in the dummy wordline), which activates the discharge of the read bitline signal in the replica column. Similarly in the case of write operation, the write clock is propagated in the read path (rwl) to provide the wl_dum signal, which activates the discharge of the write bitline dummy signals in the replica column.
- If the top line <n> is selected then the line <n-1> is used as a wordline replica as illustrated in Figure 4-32 (bot): in the case of the read operation, the read clock is propagated in the write

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

path (wl) to provide the rwl_dum signal which activates the discharge of the read dummy bitline in the replica column. Similarly in the case of write operation, the write clock is propagated in the read path (rwl) to provide the wl_dum signal, which activates the discharge of the write dummy bitline in the replica column.

Two multiplexors are used at the end of the top lines in order to select the right signals coming from the wordlines<n> and <n-1>, allowing the activation of the replica column during read or write operations.

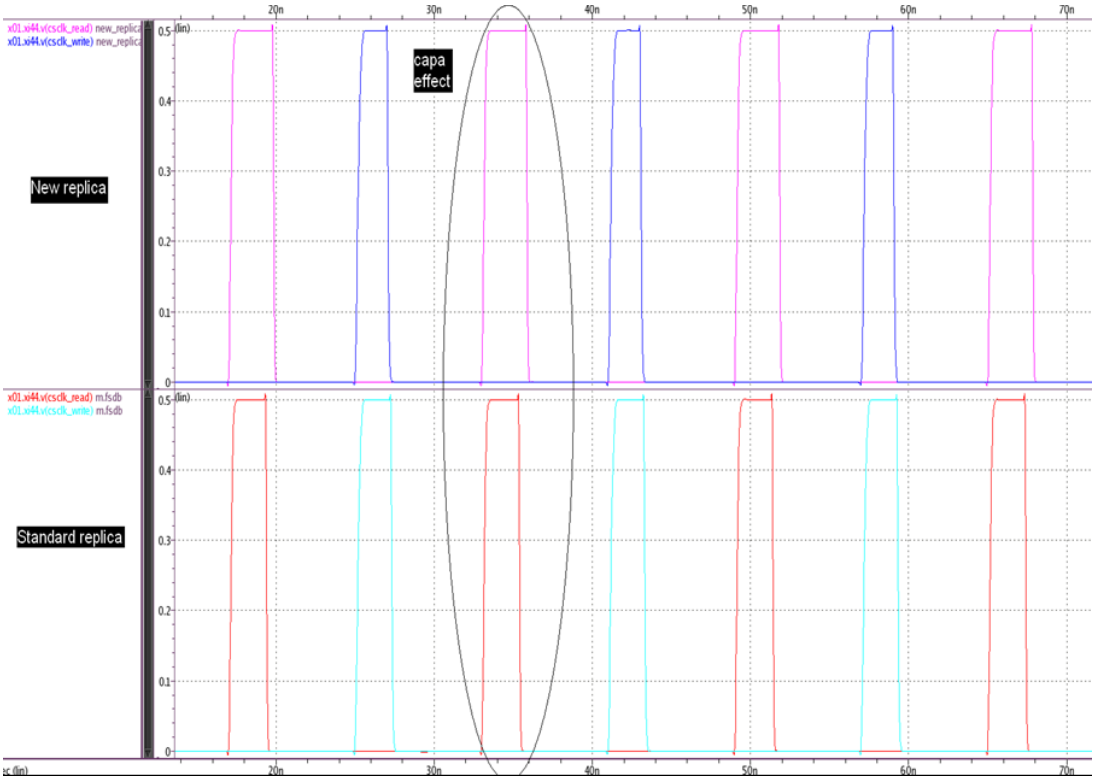


Figure 4-33 Waveforms for the read and write clocks generated by the proposed replica (top) and the standard replica (bottom) (TT, 500mV, 25°C PVT conditions, FBB=0V)

Figure 4-33 shows the waveforms of the read and write clocks generated by the proposed replica and the conventional replica. As shown in these waveforms there is a little shift delay between the two clocks generated by both replica-circuits due to capacitive effect.

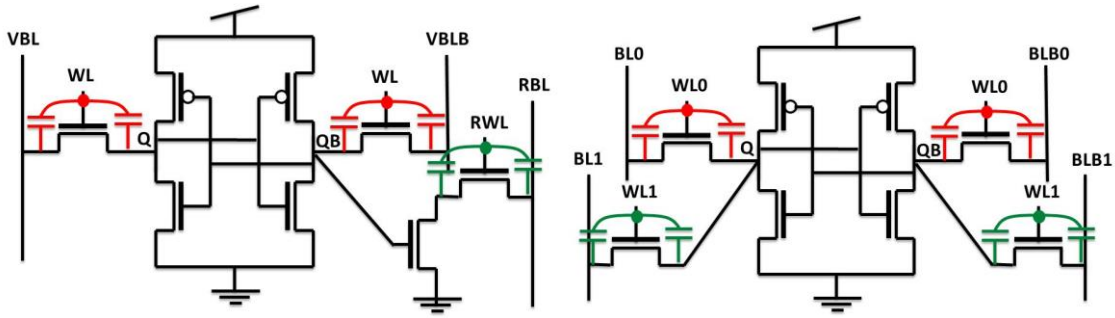


Figure 4-34. Capacitance modeling in the read and write path in the standard 8T bitcel (left) and dual port 8T bitcell (rigth)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Figure 4-34 illustrates the schematics of the standard 8T bitcell and the dual port 8T bitcell. The capacitance value in the read wordline is different from the capacitance value in the write wordline. This will result in a different propagation delay between the two paths. As discussed previously in Figure 4-32 the “wl_dum” signal is propagated in the read wordline path and the “rwl_dum” signal is propagated in the write wordline path in the selected replica line (line<n> or line<n-1>). This inversion may cause a little shift delay since the two delay paths are different as shown in Figure 4-32. Hence to take into account this matter, a little delay (1,25ns @500mV supply voltage) has been arbitrarily added in the path, featuring the smallest propagation delay to match both delays as shown in Figure 4-35.

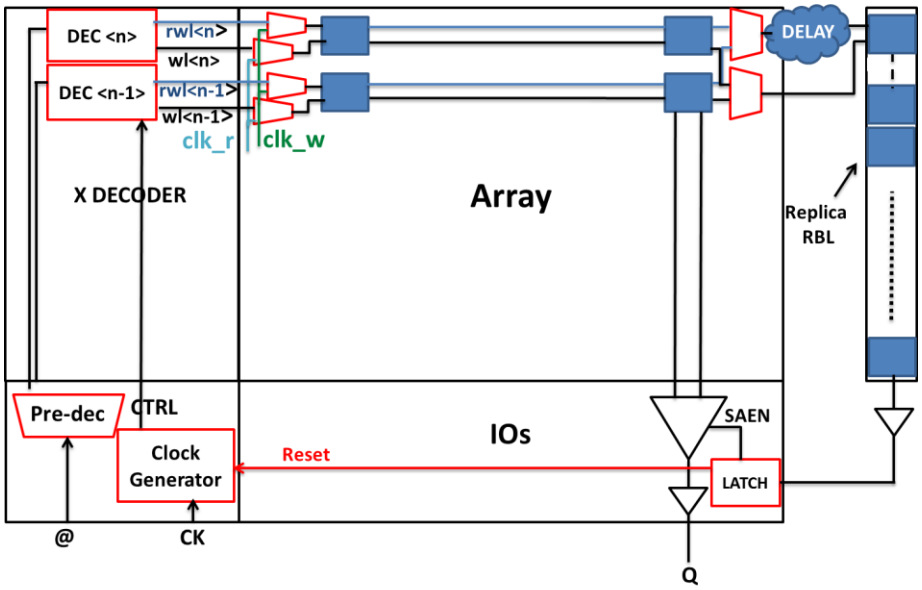


Figure 4-35: Proposed replica circuit with adapted delay

The use of the proposed replica technique allows to reduce the silicon area penalty by 10 to 20% compared to the standard replica scheme while ensuring better functionality and performances.

6.3 Configurable SA pulse width

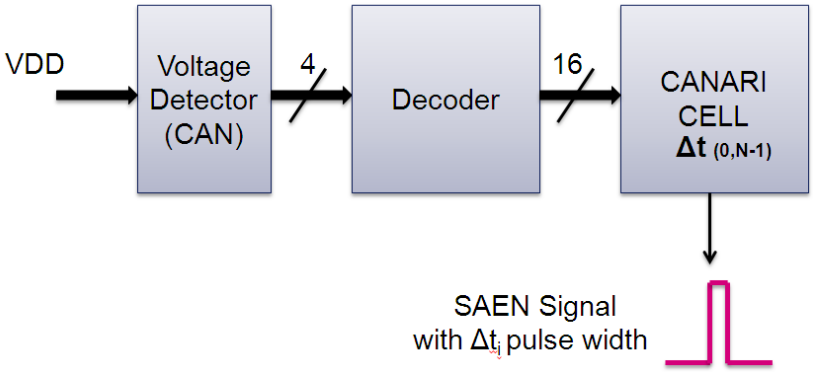


Figure 4-36. Proposed adaptive sensing time technique

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

The reaction time (RT) of the sense amplifier varies according to the supply voltage. An optimized pulse width corresponds to the worst case of the SA reaction time at 3σ. In standards SRAM the control timing block contain only one inverter chain delay that define the SAEN pulse width in the nominal supply voltage range. However the variation of the SA reaction time and the inverter chain delay are not the same at ultra wide voltage range due to PVT variability. Hence we need to optimize a specific inverter chain delay for each voltage range corresponding to the RT in this voltage range.

A canari cell is introduced in correspondence to the timing control block that generates the SAEN pulse. Figure 4-35 presents the proposed technique to optimize the pulse width signal for ultra-wide voltage range. Figure 4-36 illustrates the proposed circuit, which allows according to V_{DD} the selection of a delay automatically by using a voltage detector and a decoder. This technique is based on the introduction of various inverter chains in the canary cell that are optimized for each voltage range. Each delay is characterized to provide an optimized pulse width for the selected supply voltage. In our design (chapter 5), the selection of the inverter chain delay specific for each voltage range is performed manually through the coding of bits dedicated to test mode.

7. Conclusion

The chapter illustrates the main limitation and the motivation for sensing data in subthreshold domain. A differential and an unbalanced voltage sense amplifier is designed working down to 280 mV supply voltage to improve the read time of the proposed 10T-XY bitcell. A new replica circuit with PVT tolerance is presented allowing the reduction in silicon area penalty by 10 to 20% compared to a standard replica and finally an adaptive technique allows optimizing the sensing timing. The SA trigger signal SAEN, is of pulse shape. The variation of the pulse width of the sense amplifier with scaling down the supply voltage is critical. A canary cell includes the chain inverter delay allowing the generation of a fixed pulse width, suitable for the ultra-wide voltage range.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Chapter 5

Test-Chip and simulation results

Ultra-low voltage devices introduce new challenges with respect to circuit test. Driven by the technology scaling and the increase in the request for high speed and low power integrated circuits, circuit and board designers are moving toward near and subthreshold operating voltage [88]. Scaling down the supply voltage endangers the immunity to noise: this results in larger constraints on the tester of ultra low voltage devices. The tester must be able to drive and receive signals within a smaller margin than the standard testers [88]. Despite the techniques used to improve testers to meet the low-voltage margins, there is still today no industrial tester able to test a circuit operating in near and subthreshold supply voltage range. This chapter describes a prototype of SRAM macro with a proposal of BIST block. The prototype has been successfully designed and is under fabrication in 28nm FDSOI technology. An overview of the 32kbit SRAM is provided. Next evaluation of energy consumption and functional and performance benchmarks are presented and finally the test methodology is illustrated and the prototype circuit is detailed.

1. ULV 32kb “SYPHAX” SRAM memory

The so-called SYPHAX IC targets a memory for UWVR systems, with two modes: the first one corresponds to the operation for high performance operation (the supply voltage is set to the high value) and the second one corresponds to a low power consumption operation (the supply voltage is set to an ultra-low voltage value). The target applications can be wireless sensor node systems or biomedical implants or Internet of Things. Table 5.1 summarizes the memory specifications and Table 5.2 presents the operating conditions of the SYPHAX CUT.

TABLE 5.1 SYPHAX MEMORY SPECIFICATIONS

Technology	28 nm FDSOI
Word length	32
MUX	8
Memory size	1Kword, 32Kbit
Cell type	10T XY SRAM bitcell
Cell size (Logic DRC rules)	0.62 µm^2
Memory core size	55231.63 µm^2 (X*Y=487.912*113.2)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

TABLE 5.2 OPERATING CONDITIONS

Process	FFA/FSA/SFA/SSA Corners
Temperature	[-40°C, 125°C]
Supply voltage	[300mV, 1.3V]

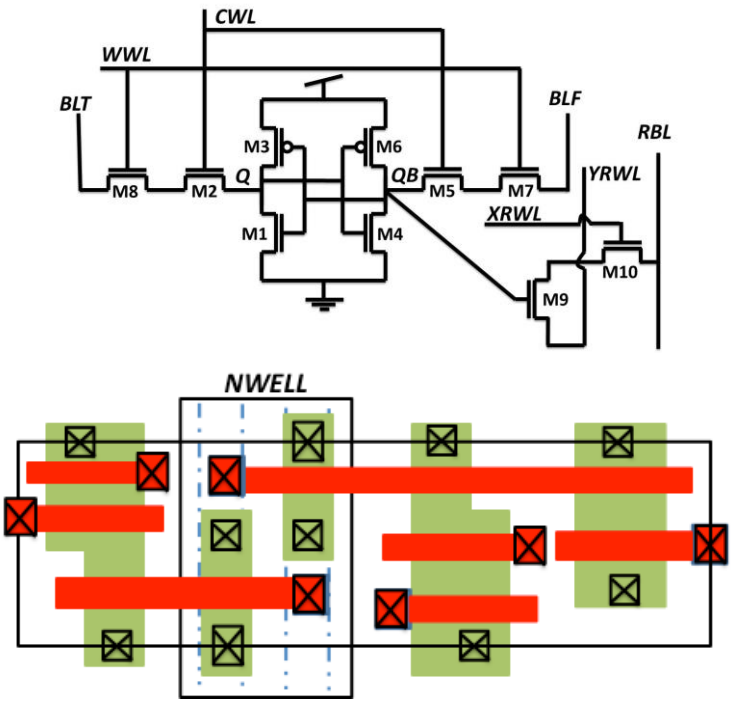


Figure 5-1. Schematic and layout of the proposed 10T-XY bitcell (0.62 µm^2)

Figure 5-1 shows the layout of the proposed 10T-XY bitcell designed with respect to logic DRC rules in 28FDSOI, in flip well configuration (chapter 4) with a size equal to 0.62 µm^2: there is a possibility to gain between 30 to 40% in area in term of layout of the bitcell if we apply the SRAM rule optimization (share contact…) what has not been done for time constraints. Table 5.3 shows the characteristic of the 10T-XY bitcell in terms of read and write margins and in term of the minimal supply voltage. The weakness of the write margin at low temperature limit the minimal supply voltage at -40°C. This is why we have better V_{MIN} in [0°C, 125°C] temperature range.

Temperature range	SNM [V]	WM [V]	V_{MIN} [V]
[-40°C, 125°C]	0.266	0.349	0.349
[0°C, 125°C]	0.266	0.240	0.266

TABLE 5.3 PERFORMANCES OF THE 10T-XY-BITCELL AT 4*SIGMA IN 28 FDSOI

So, according to results in table 5.3 and depending on the target applications, it is possible to get better V_{MIN} if we limit to [0°C, 125°C] temperature range as it is the case for biomedical applications [0°C, 80°C] (see Table 5.4).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

TABLE 5.4 EMERGING IMPLANTABLE BIOMEDICAL DEVICES

Applications	Power	Frequency	Energy
Pacemaker – defibrillator [89] [90]	<10 µW	1kHz DSP	10 years battery lifetime
Hearing aid & cochlear implant [91] [92] [93]	100-2000 µW	32kHz-1MHz DSP	One week battery lifetime
Body-area monitoring [94]	140 µW	<10 MHz DSP	External battery

Table 5.4 illustrates the emerging implantable biomedical applications. The implantable character of this kind of systems makes energy consumption a challenging constraint. The battery lifetime of applications presented in Table 5.4, illustrate the time between surgical replacements.

1.1 Contribution to design innovation

An ultra-wide voltage range SRAM has been designed including:

- An original single-ended 10T bitcell with XY read and write selection, with an ultra-low leakage current and avoiding the parasitic dynamic power consumption as discussed in chapter 3.
- An original dynamic back-biasing technique is implemented allowing the improvement in read time as presented in chapter 4.
- An optimized unbalanced sense amplifier with a replica operating down to 0.3V as discussed in chapter 4
- Two replica circuits have been implemented: the standard one and the proposed replica presented in chapter 4 will be tested and silicon results will be compared to simulation.

1.2 Design

1.2.1 Memory Floorplan

The SYPHAX memory cut is composed of four matrix blocks as shown in Figure 5-2. Each block is composed of 64 bitcells by row and 128 bitcells by column. Figure 5-3 presents the top-level matrix block organization. The matrix block is divided into 16 sub-blocks (presenting 16 bits of inputs/outputs). Eight columns compose each sub-block (MUX 8). During operation, one of the two matrix blocks is selected (blocks in the top or blocks in the bottom as shown in Figure 5-2). The standard replica circuit is implemented in the matrix block located in bottom-right while the proposed replica is implemented in the matrix block located in top-right. A programmable bit is used to make the choice between the two replicas.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

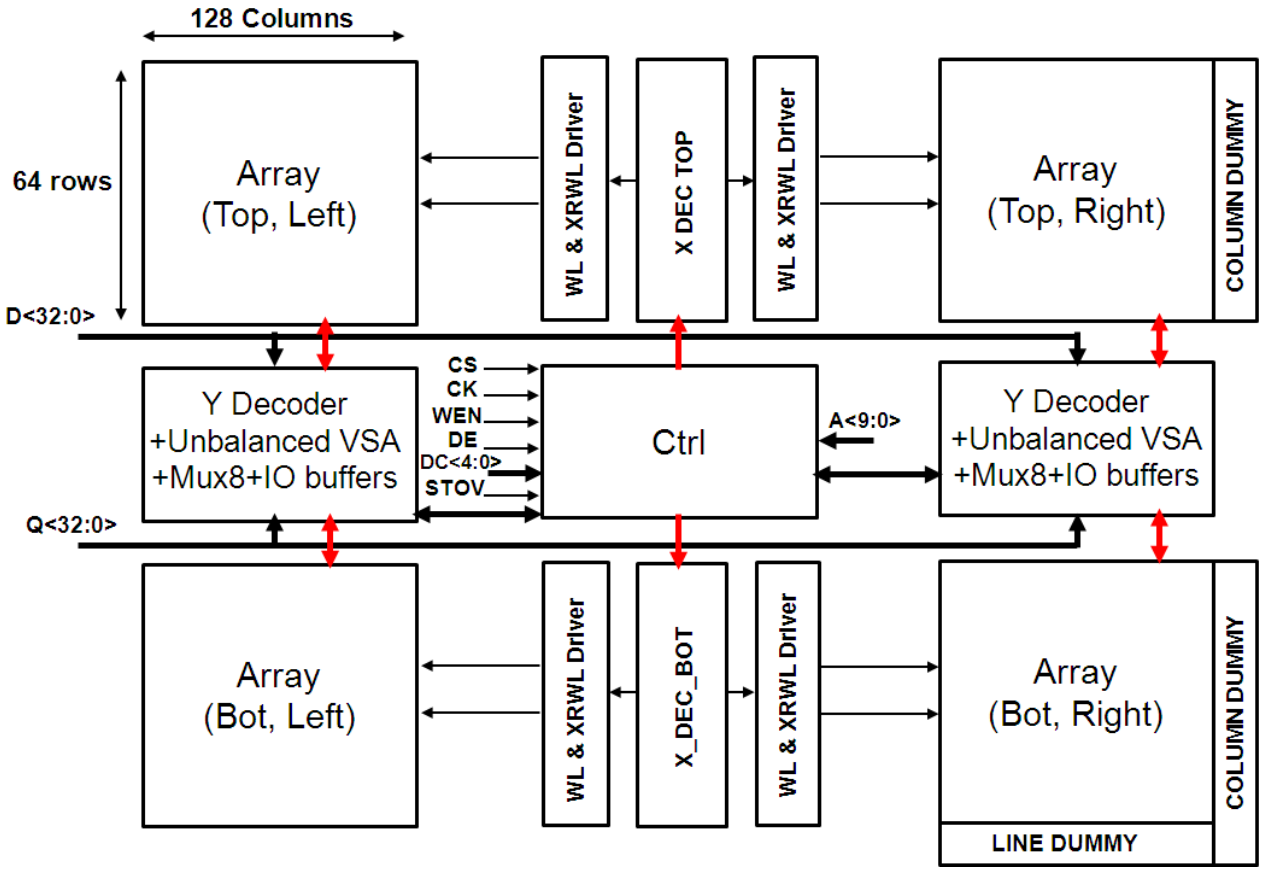


Figure 5-2. Architecture of 32kbit SYPHAX memory

SRAM operation in subthreshold voltage is significantly more sensitive to soft-error. This is due to the lower supply voltage in the internals nodes of the bitcell at ULV [95]. The 10T-XY subthreshold bitcell, with the bit-interleaving technique in the column structure, allows rejecting multiple soft errors [96]. As shown in Figure 5-3, on the one hand, the wordline (WWL) signal is shared by the bitcells in a row. On the other hand, the column wordline (CWL) signal is shared by the bitcells in the column. Both of them control the write path. In the same way, the XRWL signal is shared by the bitcells in a row and the YRWL is shared by the bitcells in column. These signals control the read path. Therefore both WWL and CWL must be set to V_{DD} to ensure the write operation. So each column is individually selected depending on the value of CWL. Similarly both XRWL and YRWL must be activated to allow the read operation. In the same way each column is individually selected depending on the value of YRWL. During the write or read operations, the selected 10T-XY bitcells do not disturb the stability of adjacent bitcells in the same column.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

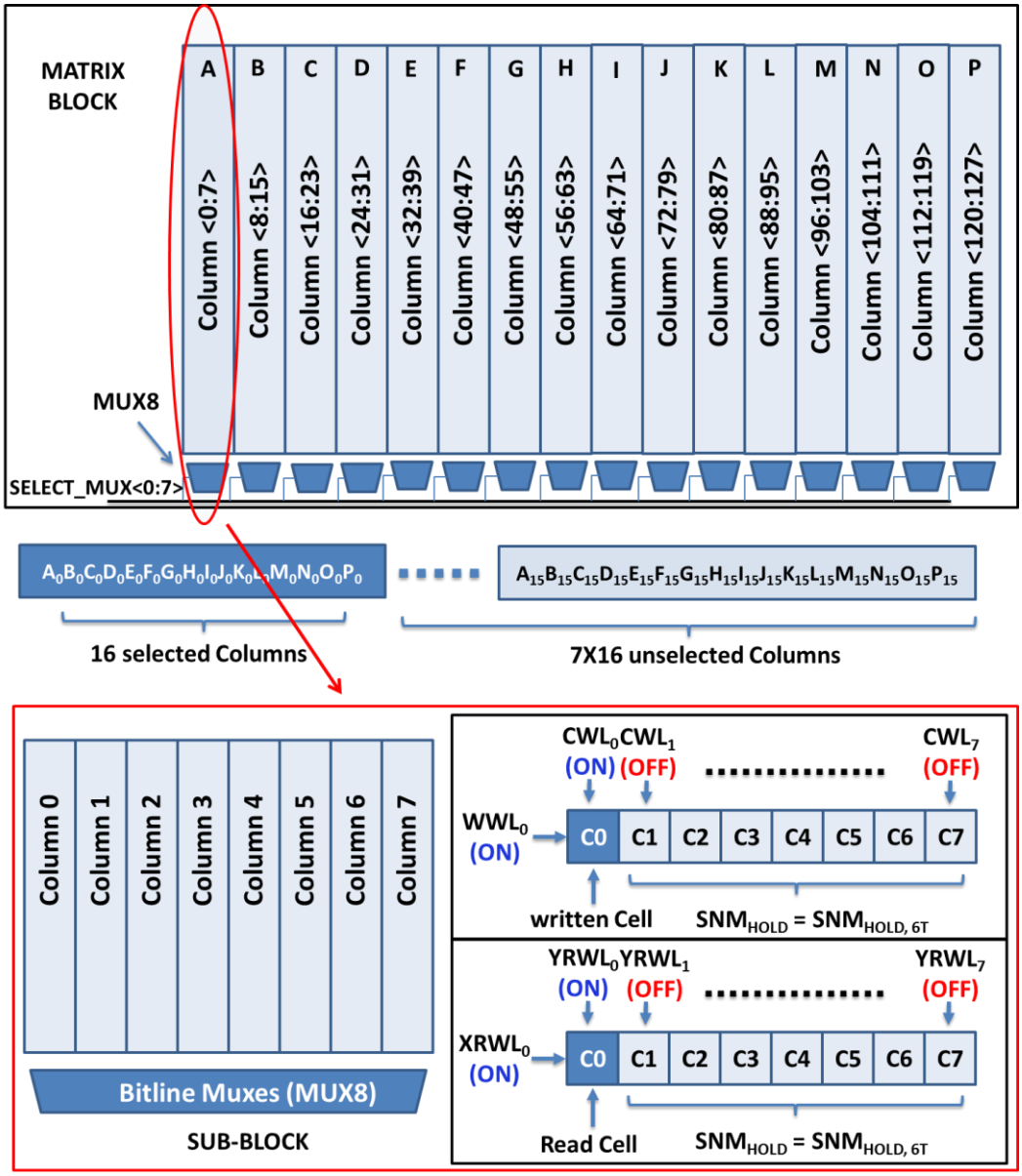


Figure 5-3. Matrix block organization

1.2.2 Decoder structure

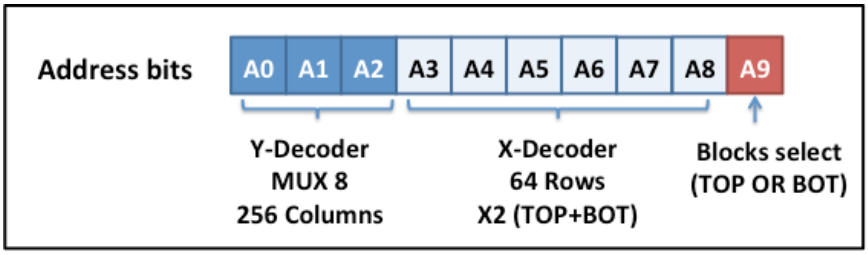


Figure 5-4. Address bit organization

Figure 5-4 shows the organization of the 10 address bits allowing to select 1K*words. The first three address bits, A<0:2>, are used to generate the SELECT_COL<0:7> for the selection of 256 columns. The address bits, A<3:8>, are used for the Y-DECODER. These bits are decoded through two stages. The logic scheme of the X-decoder is presented in Figure 5-5. The first stage allows the generation of

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

64 signals. The 64 wordline signals, WL<0:63>, applied to the pass-gate transistors, are generated by the second stage of the decoding logic, which combines DEC and CLK_W (the clock of the write operation). Similarly the 64-wordline signals (XRWL<0:63>) applied to the read ports are generated by the second stage of the decoding logic and a level shifter, which combine DEC and CLK_R (the clock of the read operation). The use of the level shifter is required by the negative supply voltage (under-drive) in the unselected rows to reduce the I_{OFF} current. In addition as explained in chapter 3, the use of an under-drive technique avoids read failures at ultra low voltage range thanks to the compensation of the I_{ON}/I_{OFF} ratio weakness.

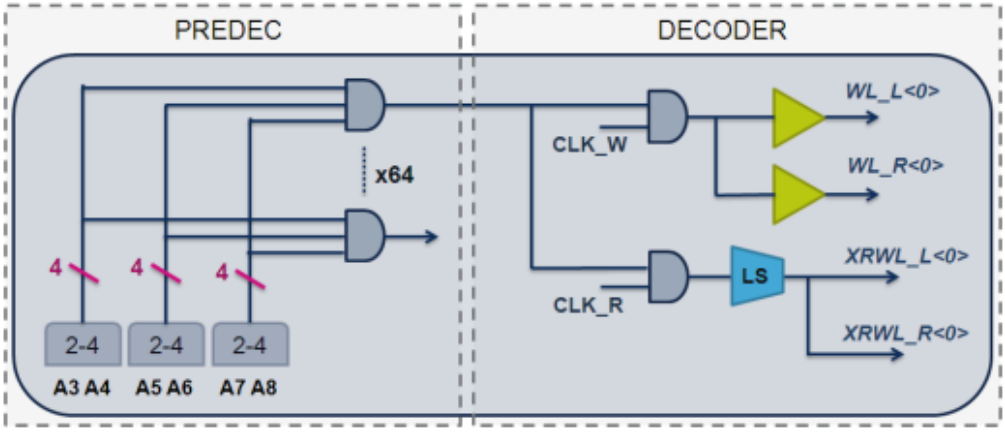


Figure 5-5. X-Decoder structure

1.3 Sharing Unbalanced VSA

The unbalanced sense amplifier, designed for the single-ended 10T-XY bitcell, is implemented in the SYPHAX memory as shown in Figure 5-6. The SA has two inputs: the first one corresponds to the reference voltage (in our case, V_{DD}), which is provided by the selected RBL (pre-charged to V_{DD}) in the unselected blocks (TOP or BOT). The second input corresponds to the RBL of the selected column in read mode in selected blocks. This method consists in sharing unbalanced SA between the selected columns in the TOP block and similarly in the bottom one. This reduces the number of necessary SAs so it reduces the area and the power consumption of the memory cut.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

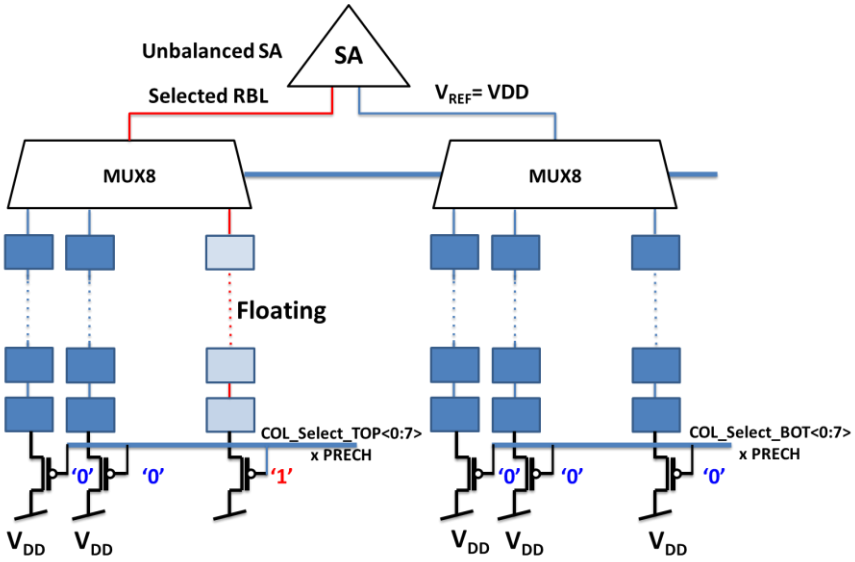


Figure 5-6. Unbalanced voltage sense amplifier share

Figure 5-7 shows the simulated waveforms of the input/output signals of the optimized unbalanced sense-amplifier. The sensing data starts with the rising edge of the SAEN pulse. The internals nodes of the unbalanced sense amplifier are placed in the input of an RS latch to get the output read signal (DOUT). The pulse width of the SAEN signal is defined by the optimized delay chain in the timing control block (canary cell), as explained in chapter 4.

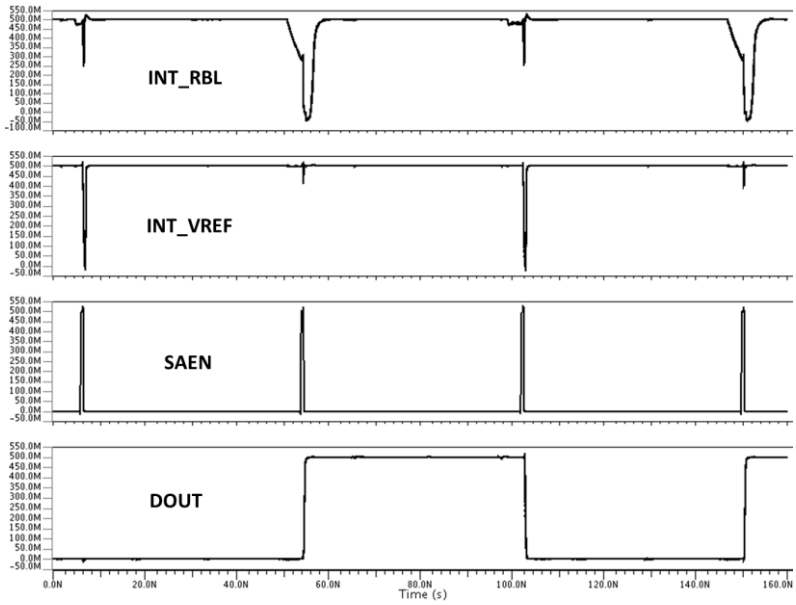


Figure 5-7. Simulation of the SA internal nodes in the full memory

(extracted CUT, TT, 500 mV, 25°C: PVT conditions)

1.4 Level Shifters

Figure 5-8 presents the schematic of the level shifters (LS) implemented in the SYPHAX CUT at the INPUT and OUTPUT to provide an ultra-low voltage gap. We should add the Level shifters since major testers do not provide ultra-low voltage signals.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

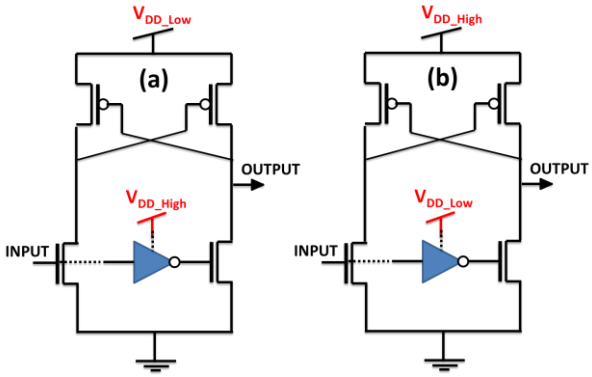


Figure 5-8. Schematic of the Level shifters (a) High-to-Low and (b) Low-to-High

1.5 Programmable pins

To ensure an efficient functionality down to sub-V_T and above-V_T supply voltage range, programmable pins have been added. Table 5.5 presents these programmable pins: DC<0> is used to select the operation with or without self-timing, DC<1> selects one of the two optimized delays which define the SAEN pulse width for nominal and ULV supply voltage range. DC<2:3> selects the fixed cell driver in the replica circuit (possibility to select between 1 to 5 fixed drivers. Finally DC<4> activates the level shifter, which provides negative values for the XRWL signals in the unselected rows.

TABLE 5.5 DEBUG PINS

Debug pins	
DE	Debug enable
STOV	Allows to select the operation with or without self-timing
DC<0>	Allows to select one among the two implemented replica circuits
DC<1>	Allows to introduce Delay in the RBL_DUM
DC<2:3>	Allows to accelerate read and write inside replica circuit
DC<4>	Allows to activate level shifter in the read path

1.6 Logic behavior

Table 5.6 presents the truth table, which describes the functionality of the SYPHAX memory in various modes. The read operation starts with every rising edge of the clock (CK). The latched decoded address selects one of the memory locations in the memory core. Then the stored data is ready to be read by the unbalanced sense amplifier. Sensing data starts with the rising edge of SAEN pulse as discussed in previous chapter. The write operation starts with every rising edge of the write clock (CK). The addresses, the value of Write Enable (WEN), the data available on the data bus and the bit write information are latched at the rising edge of the clock. The latched decoded address

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

selects one of the memory locations in the memory core. Then the stored data is ready to be written in the selected location.

TABLE 5.6 TRUTH TABLE

Function	CSN	CK	WEN	D	TBYPASS	INITN	Q	Action on array
Standby	H	X	-	-	L	H	Q-1	No Change
Read	L	↑	L	X	L	H	Mem	No Change
Write	L	↑	H	Din	L	H	Q-1	Din
Memory Bypass	X	X	X	Din	H	H	Din	No Change
Initialization	X	X	X	X	X	X	L	X

Figure 5-9 shows the layout of the 32kb SYPHAX memory. The size of this memory is equal to 55142 μm^2 in 28nm FDSOI.

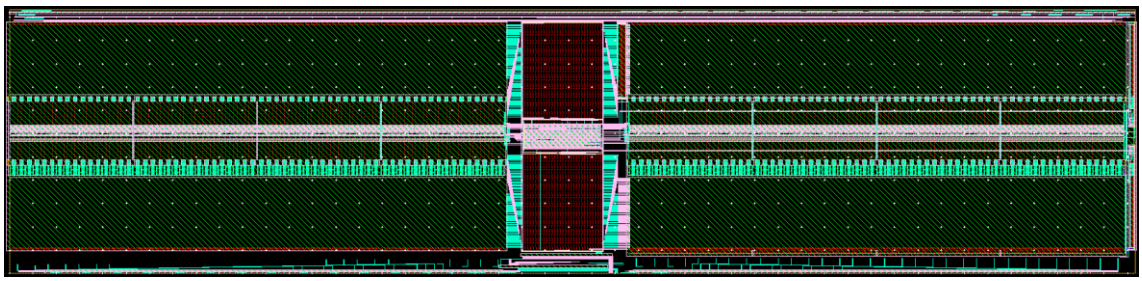


Figure 5-9. SYPHAX CUT Layout (XY=487.912µm x 113.2µm)

2. Simulations Results

All simulation results in this chapter are based on full cut using the extracted net-list ensured by a fast simulator. Figure 5-10 presents simulated waveforms of the read/write operations at different address lines in the memory cut, using the full cut net-list extracted from the layout, given that the extracted net-list includes parasitic devices (RC: resistance-capacitance). The read delay is the delay between the CK positive edge and the transition on the output signal Q, in read mode (WEN=1). The write delay is the delay between the CK positive edge and the transition on the internal data of the bitcell. Figure 5-11 presents the total energy per cycle and the maximum operating frequency of the full cut for various supply voltage values.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

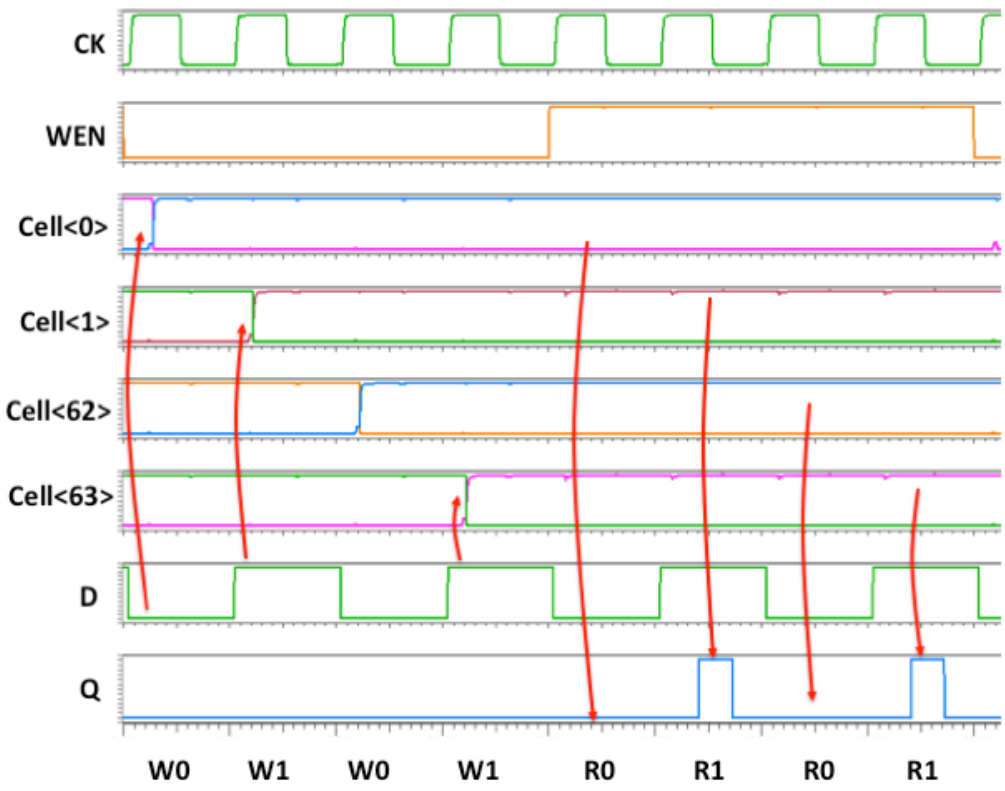


Figure 5-10. Simulation of the read and the write operation for the full memory cut in full swing read mode using xa simulator (350mV, TT, 25 °C PVT conditions)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

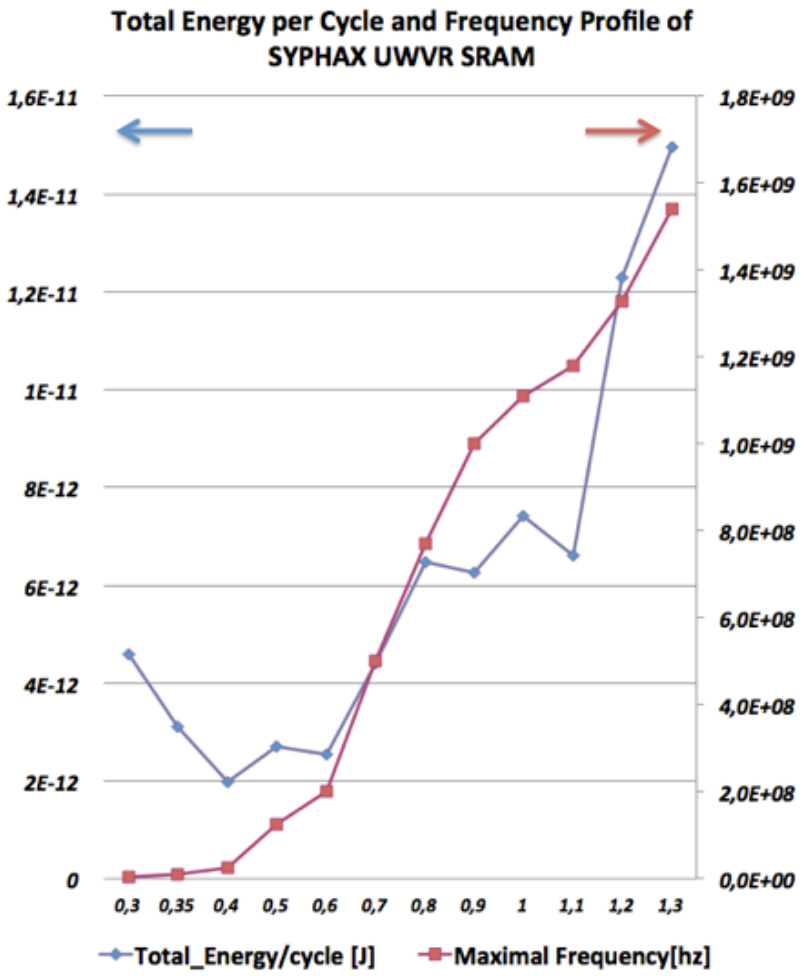


Figure 5-11. Energy per cycle and operating frequency profile of SYPHAX CUT versus supply voltage)

These results are taken in TT and 25°C conditions. The minimum energy point of SYPHAX memory is equal to 2pJ at V_{DD}= 400 mV with 25 MHz operating frequency. At 1.3V supply voltage, the energy per cycle is equal to 15pJ and the operating frequency is equal to 1,5GHz. This results in a reduction of the total energy per cycle by 8, between nominal and ultra low voltage domains. It should be noted that the propagation time and the power consumption of the level shifters are taken into account in this characterization. On the other side, the power consumption of body-bias generator circuit (providing of negative and positive programmable supply voltage) has not taken into account (the body-bias generator block will be used by the others circuit in the SOC).

An optimization has been done at the level of the read operation using sense amplifiers and applying forward body bias techniques. This results in a significant, improvement in the read access time. On the other hand, there is no write assist technique to improve the write access time. Hence the access time of the SYPHAX memory in ultra-low voltage range is limited by the write time. There is a possibility to gain 45% to 60% in term of frequency, at ULV, with adequate corrections in the design (using a write assist technique).

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

3. Benchmark

Table 5.7 compares the performances of the proposed ultra-wide voltage range SRAM memory with other state-of-the-art SRAM macros targeting ULV applications in advanced technologies nodes.

TABLE 5.7 COMPARISON WITH OTHER STATE-OF-THE-ART MEMORY CUTS

Design	Technology	Capacity	Operating V_{DD}	Word length	Frequency (MHz)	Access energy (Access energy per bit)
Ming et al [Ming-Hsien] [24]	65nm CMOS	72Kbit	(0.35-1.2)V	32	200 @1V	4.5pJ @ 0.5V (140.62 fJ @ 0.5V)
Suziki et al [Suzuki] [25]	130nm CMOS	32Kbit	(0.3-1.5)V	32	6.8-960	-
Verma et al [Verma] [26]	65nm CMOS	256Kbit	(0.35-0.5)V	128	0.03-1	30pJ @ 0.5V (238.37 fJ @ 0.5V)
Takeda et al [Takeda] [97]	90 nm CMOS	64Kbit	(0.44-1)V	16	50-833	-
Sinangil et al [Sinan] [68]	65nm CMOS	64Kbit	(0.4-1.2)V	128	0.02-200	23pJ @ 0.8V (179.6875 @ 0.8V)
Fady et al [Fady] [27]	28nm FDSOI	32Kbit	(0.35-1.3)V	32	13-1020	1.15pJ @ 0.3V 22.8pJ @ 1.2V (35.93 fJ 0.3V) (712.5 fJ @1.2V)
Chien-Fu et al [Chien] [28]	65nm CMOS	16Kbit	(0.18-0.3)	16	4.8-48	-
This work	28nm FDSOI	32Kbit	(0.3-1.3)V	32	8-1510	3pJ @ 0.3V 2pJ @0.4V 15pJ @ 1.3V (62.5fJ @ 0.3V) (468.75 fJ @ 1.3V)

Figure 5-12(a) shows the minimal supply voltage vs. technology node. The $V_{DD,\ MIN}$ of SYPHAX memory is equal to 300mV, which presents an acceptable value compared to other ULV design works in the-state-of-the-art. The work in [27] is a little bit better in term of operating frequency and power consumption at 350 mV. As discussed previously the weakness of operating frequency at ULV is due to the write operation: a correction must be done in the future in order to improve this point. The work in [27] use an interesting read assist technique. However this technique does not allow avoiding the

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

I_{ON}/I_{OFF} weakness above 400mV. SYPHAX CUT is better in term of power consumption and operating frequency at nominal voltage thanks to the use of the unbalanced SA and the use of the ultra low leakage 10T-XY bitcell (as shown in Figure 5-12(e)). As shown in figure 5-12(c) and (d), the circuit is able to operate at 8MHz at ULV, and up to 1.5 GHz at nominal voltage. As shown in Figure 5-12(b) the memory presents low power consumption at ULV and nominal voltage compared to other works.

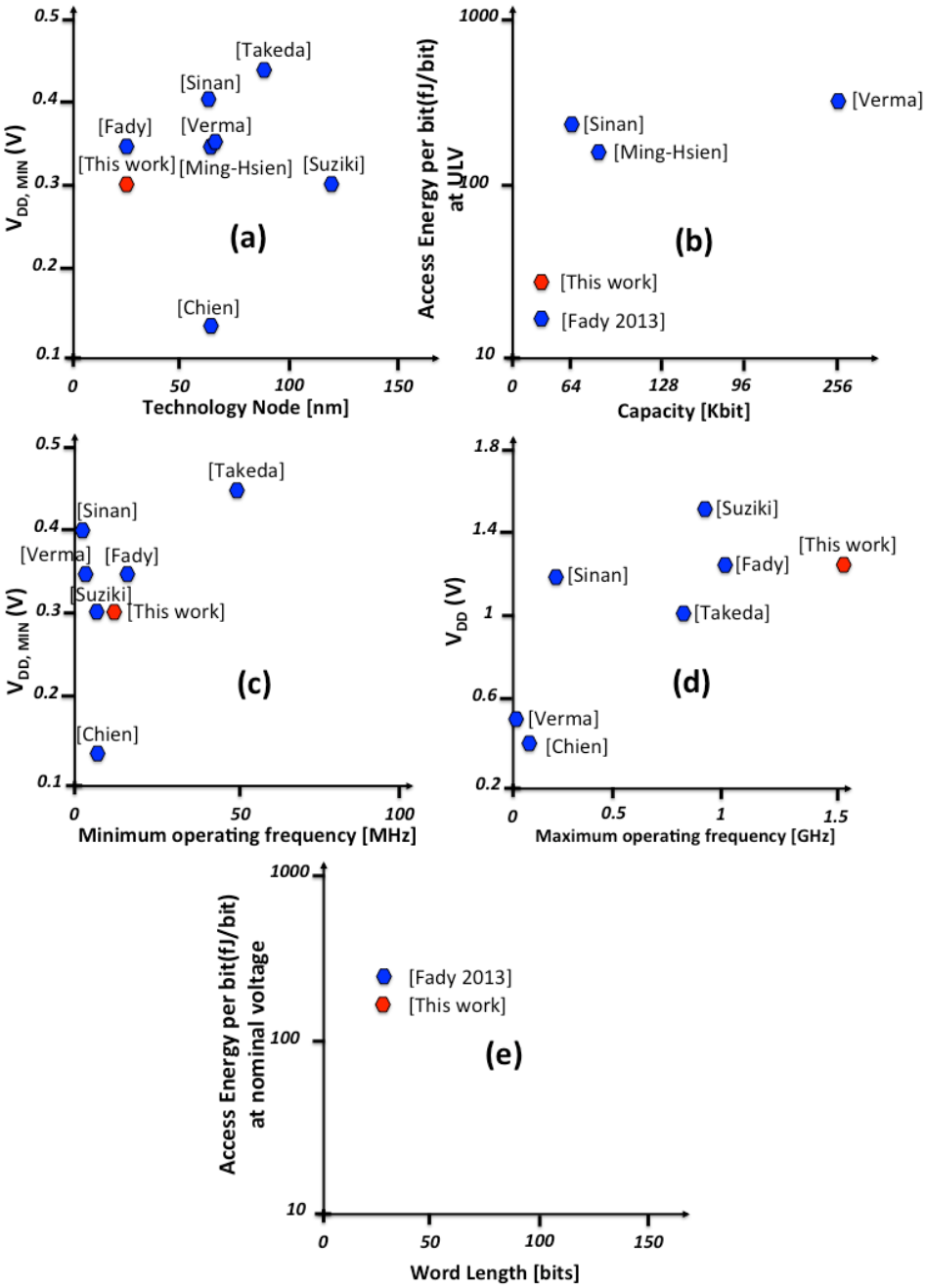


Figure 5-12. Comparison of the SYPHAX memory with the state-of-the-art cuts in terms of VDD, MIN (a), access energy per bit (b) and maximum operating frequency at ULV (c) and nominal supply voltage (d)

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

4. Problems and challenges that have been overcome

Many challenging problems have been overcome during the SYPHAX memory design:

- At the level of bitcells: the proposed 10-XY bitcell allows to significantly reduce the leakage current thanks to the resistive write path (using two pass-gates) and the read-port. Read failures due to the weakness of the I_{ON}-to-I_{OFF} ratio is one of the main limitations of SRAM functionality at ultra-low voltage. This issue is solved thanks to an under-drive at the level of read-port in the unselected rows. Finally, parasitic power consumption was avoided thanks to the XY configurations.
- At the level of decoder: under-drive of the unselected rows during read operation is needed only below 400 mV supply voltage. A level shifter (LS) has been used to ensure the progress of the negative voltage to read-ports. To avoid significant power consumption due to the use of the negative supply voltage, logic controlled a programmable bits was implemented to make the LSs inactive and replace the negative supply by ground in the unselected rows (XRWL signal).
- To overcome the important limitation in read access time at ultra-low voltage range which limits the frequency operation and the field of possible applications: first, an optimized sense amplifier is designed operating down to 280 mV and second a dynamic modulation technique of V_{TH} has been used to boost the read time in the selected column.
- To avoid the large variation of the pulse width at ultra wide voltage range due to PVT variations, a configurable SA pulse width technique is adopted and two delay chains have been implemented in the timing control block. A programmable bit was used to select the specific delay for each supply voltage range (sub-V_T or above-V_T).
- 32kbit SRAM architecture has been defined taking into account the limitation of number of bitcells stacked per bitline and per row.

5. Prototype of UWVR Test Methodology applied to the SYPHAX memory

The test of ultra-low voltage SRAM memories is one of the big challenging issues in the next generation of low power SRAMs due to the weakness of industrial testers [88] unable to ensure operation near subthreshold domains [88]. Direct memory test (DMT) is the method, which uses an external tester allowing the access to the internal memory. The test is performed over the I/O pins. The tester is responsible for writing the testing pattern into memory, and reading back the information stored. The tester will then make the comparison of the data, which has been reading from the memory

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

cell to what value should have been there. The advantage of using this method is that the user who should have a good idea of how the tester works can easily change the testing patterns. The Direct Memory Test is the most popular method for the test of ULV devices. However this method is limited in terms of covered faults. The built-in-self-test (BIST) is besides the mechanism for testing memories and providing an efficient fault diagnostic. A specific BIST has been used to test SYPHAX memory in sub-threshold and above threshold supply voltage range. Our proposed 10T-XY bitcell and proposed techniques (discussed in chapter 4) are included in the SYPHAX CUT. This memory is embedded in a demonstrator designed in 28nm FDSOI technology. The aim of this demonstrator is to validate the functionality of our memory in sub-V_T and above-V_T range and to compare the results in terms of power consumption, operating frequency and yield compared to industrial SRAM products in 28nm FDSOI.

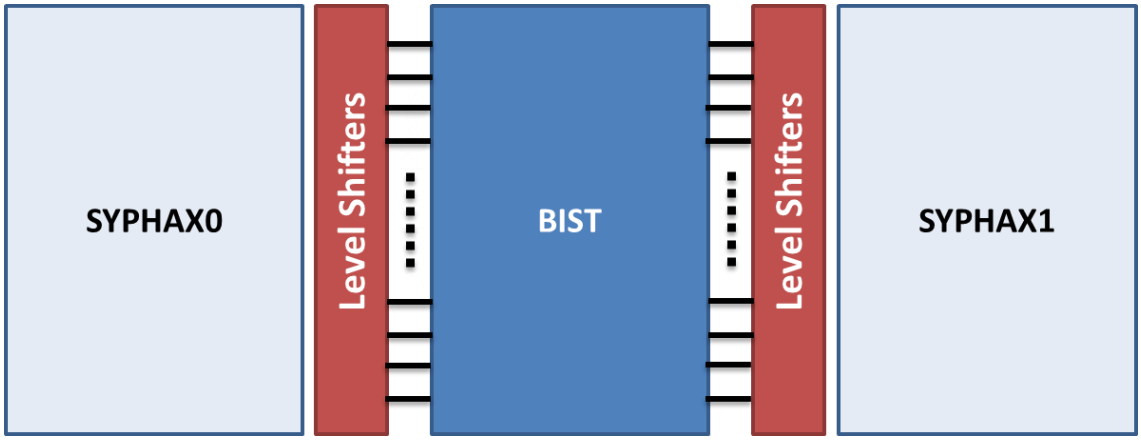


Figure 5- 13. View of our demonstrator

Figure 5-13 presents the symbol view of the demonstrator including a proposed BIST and two instances of SYPHAX CUT to increase the statistical data. Two main supply voltages are used: the first one corresponds to the nominal voltage used to power the BIST and the level shifters. The second supply voltage corresponds to the ultra low voltage supply used to power both memory instances. The layout of the demonstrator is shown in Figure 5-14 designed in 28nm FDSOI.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

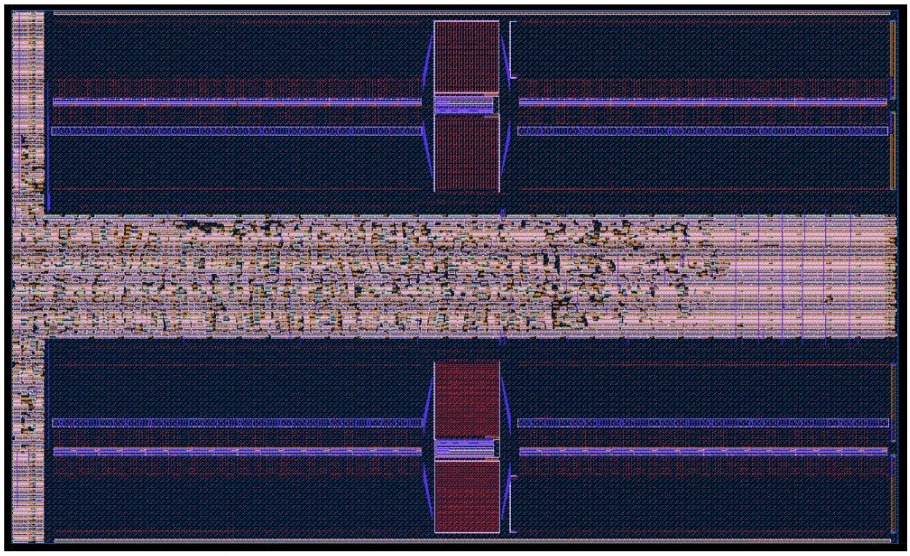


Figure 5-14. Layout of our demonstrator

The Inputs/Outputs (IOs) and the BIST of the demonstrator are embedded in a Test-chip, powered by the nominal domain voltage. The level shifters embedded in SYPHAX CUT ensure an interface between the nominal and the ultra-low voltage domains. Figure 5-15 presents the finite state machine (FSM) of the BIST, ensuring the test of our ULV SRAM memory. The BIST presents three- loop finite state machines (FSM) and it is controlled by input signals. The first loop performs the scan-in and the scan-out, while the second loop allows the "scan in, initialization, check and scan out" and finally the third loop performs the "scan-in, initialization, test, check and scan-out". The BIST uses two clocks: the high-speed clock is used for the scan-in and the scan-out state while the clock of our memory is used over the other states.

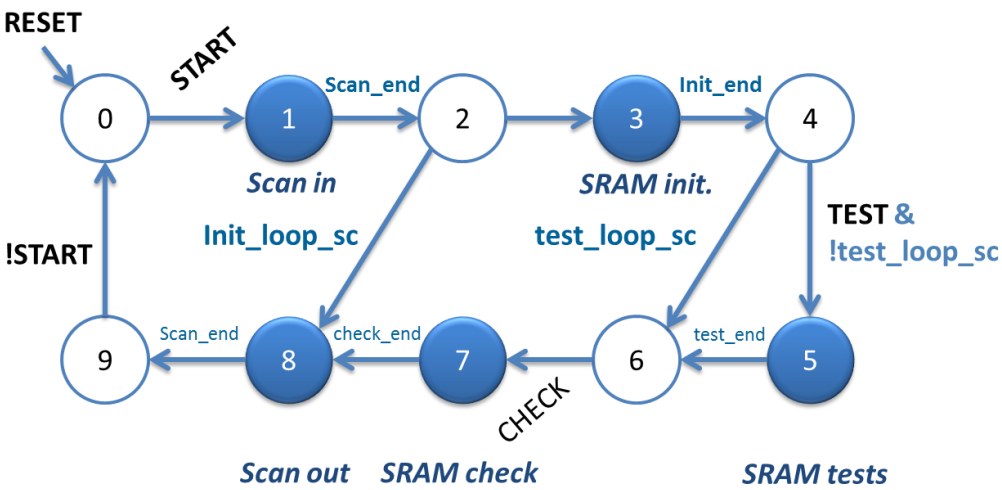


Figure 5-15. Finite State Machine (FSM) of the proposed BIST

The scan starts when scan-begin flips to V_{DD} and the scan ends when scan-end flips to V_{DD}. As shown in Figure 5-16, for scan-in a new data must be delivered at each clock falling edge.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

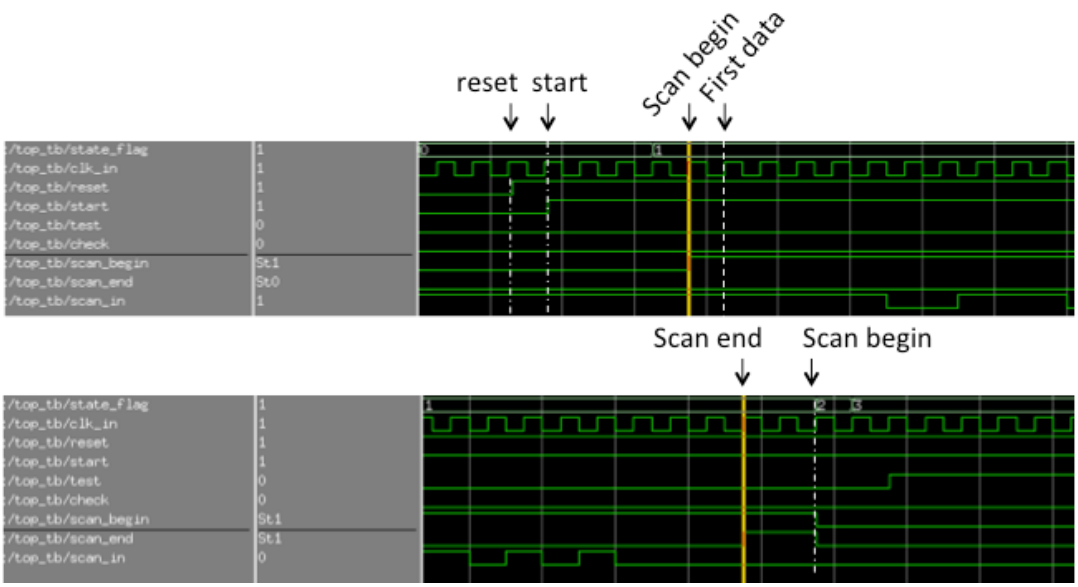


Figure 5-16. Timing chronogram simulation of our demonstrator in Scan-in state (S1)

The system is asynchronously reset. The configuration state allows setting the programmable pins according to the voltage operating range sub-V_T or above-V_T as shown in Figure 5-17 (for our design it's related to DE, STOV and DC<0:4> programmable pins). This is for adapting the memory to operate in the selected supply voltage range.

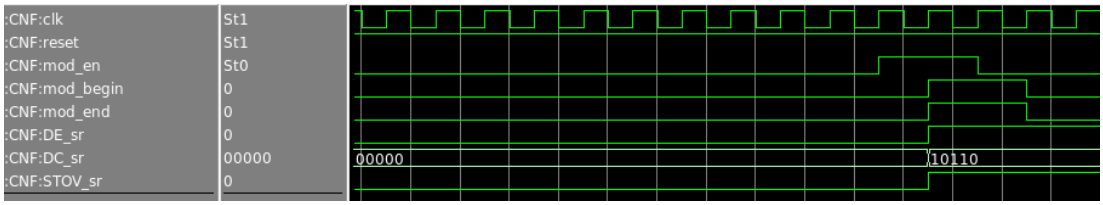


Figure 5-17. Timing chronogram simulation of our demonstrator in configuration state (S2)

The initialization state performs a first write operation to initialize the data in all bitcells before the start of the test. The SRAM initialization is performed at nominal voltage as a safe mode. Once the SRAM initialization is over, the memory supply voltage can be adjusted for the SRAM test and then the TEST signal can be turned on. The SRAM tests are performed from the initial test address "addr_sr_tst_0_sc" to the final address "addr_sr_tst_end_sc" with an increment of specific value "addr_sr_tst_inc_sc" as illustrates in Figure 5-18. Four modes are available to generate test patterns: write the same value, read only, write a dual value and alternate read/write operations.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

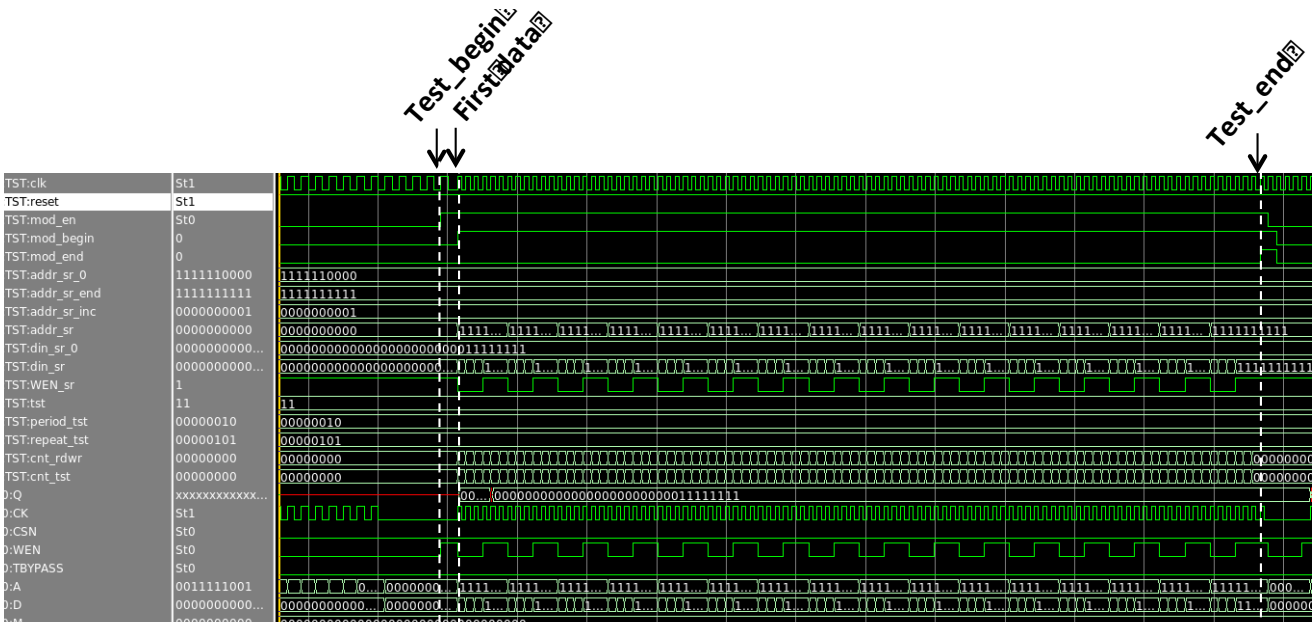


Figure 5-18. Timing chronogram simulation of our demonstrator in test state (S5)

Once the SRAM tests are done, the state machine goes in S7 and waits for the CHECK command. In S7, the SRAM is supplied with the nominal voltage as a safe mode before the data are checked. The timing chronogram during the check states is presented in figure 5-19.

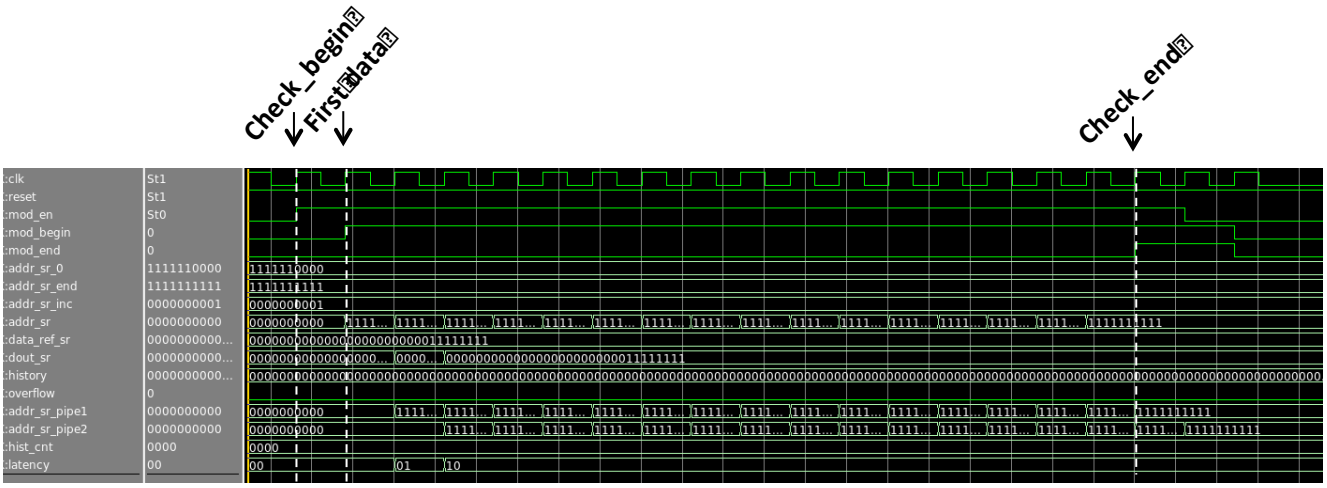


Figure 5-19. Timing chronogram simulation of our demonstrator in check state (S7)

For scan out a new data is collected at each clock falling edge as shown in Figure 5-20.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

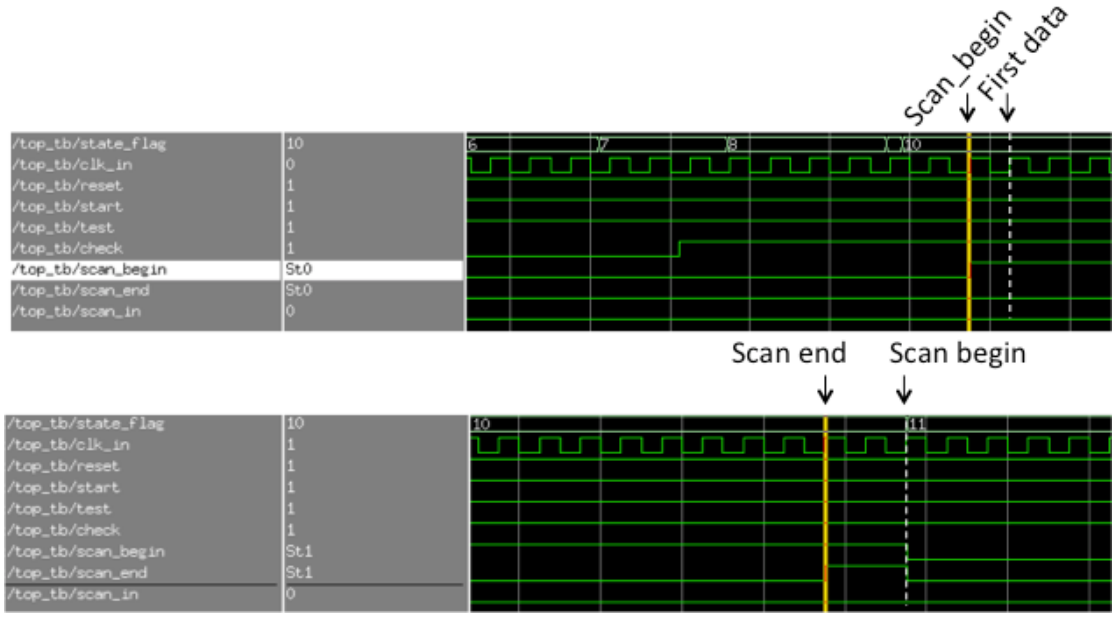


Figure 5-20. Timing chronogram simulation of our demonstrator in Scan-out state (state 8)

The demonstrator is implemented in 8 rows of scribes (2222um x 480um) in a test-chip. This test-chip has been designed with LVT transistors offered by the design kit. Standard DRC has been applied. This test-chip is under manufacturing in 28FDSOI technology. Silicon results are not available at the time of this manuscript. Digital and analog IO buffers have been implemented to interface the IOs and the demonstrator. The test-chip including our demonstrator, analog/digital buffers and IOs pads is shown in Figure 5-21.

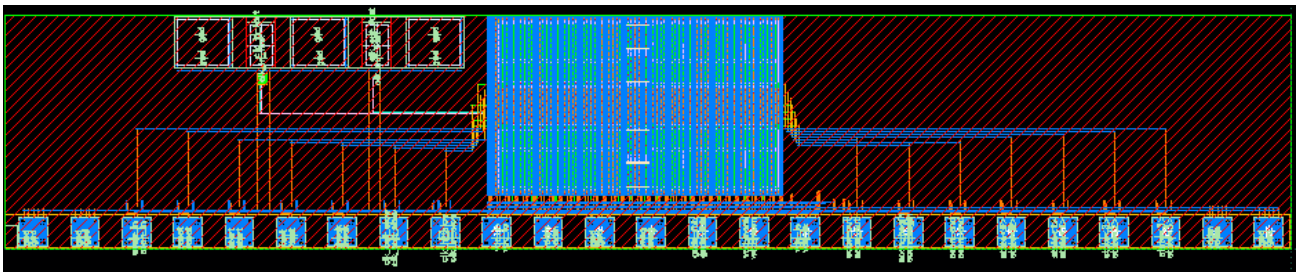


Figure 5-21. TESTCHIP

Analog and digital buffers have been introduced between the IOs and the demonstrators to amplify the current. Figure 5-22 illustrates the test equipment using PXI interface, which will be used to test our demonstrator.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

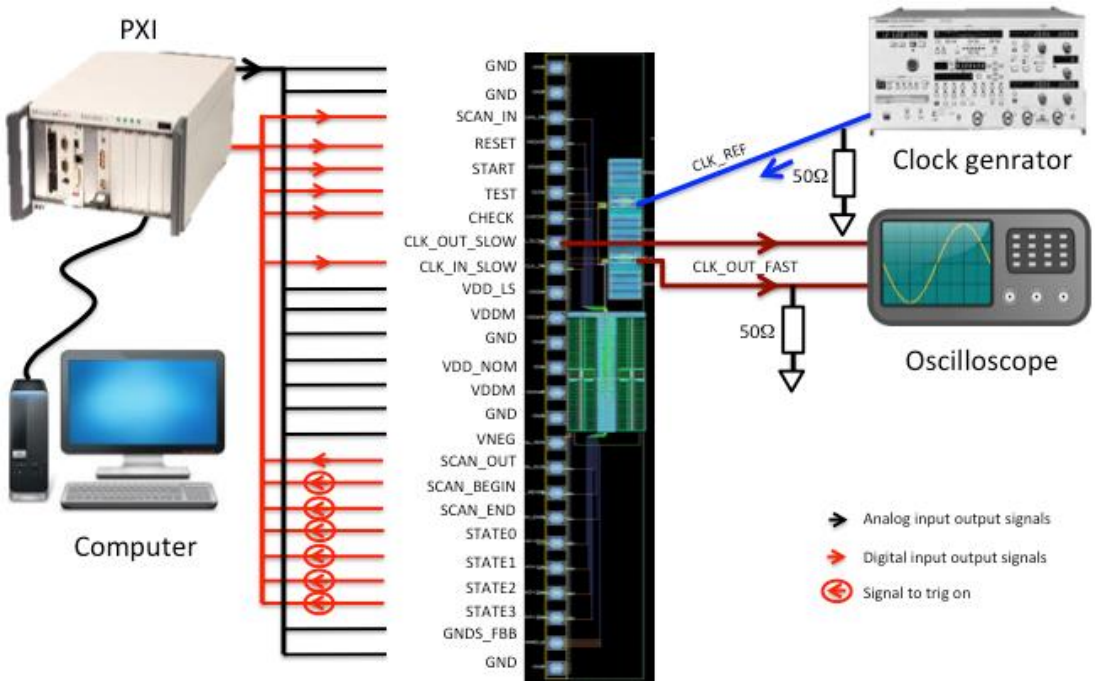


Figure 5-22. Test equipement

6. Conclusion

An ultra-wide voltage memory cut is presented, including the proposed 10T-XY bitcell and other techniques. The 32kbit memory is designed in 28nm FDSOI technology. The performances in terms of operating frequency and power consumption are evaluated and compared to other ULV design works in the state-of-the-art. The results confirm that our proposal is competitive. SYPHAX CUT with a proposed BIST has been implemented in a demonstrator to evaluate the functionality, performances and yield at UWVR. Simulation results are functionally correct.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Chapter 6
Conclusion

This work focuses on sub-V_T and above-V_T SRAM memory design with three objectives: first to avoid the limitations of state-of-the-art ultra-low voltage bitcells in terms of static power consumption and functionality in order to find a new bitcell architecture to resolve these main limitations. Second to found new bitcell architectures that can be used to design an Ultra-wide voltage range SRAM memory. Which techniques can be used to ensure the functionality on sub-V_T and above-V_T? And finally, how the designed UWVR SRAM designed memory can be tested? There is no industrial tester able to cover tests for ULV circuits.

The details of these contributions are summarized below:

- The main limitations of sub and near-threshold SRAM bitcell design were illustrated. The increase in the threshold voltage variation has a negative impact on mismatch and the bitcell performances. Ultra-low voltage bitcells should then have a relatively larger area to avoid failures compared to bitcells for standard power supply conditions. Two main research axes are discussed in order to solve the limitations in WM and SNM in ultra-low voltage range due to the conflicting design requirements (tradeoffs must be settled between α and β ratios). A lot of previous works are based on the use of read- or write- preferred 6T bitcell. However this technique needs additional assist-circuit, which increases the complexity of the SRAM periphery while not being always efficient in term of VDD_{MIN}. The use of new bitcell architectures with separated read ports appears as a promising alternative that enables the feasibility of efficient operation at ULV. Unfortunately this kind of bitcells suffers from

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

parasitic phenomena like the degradation in the I_{ON}-to-I_{OFF} ratio resulting in read failures and the parasitic power consumption in the half-selected bitcells.

- The selection of a proposed 10T bitcell with modified read ports has been done. The degradation of the read current at ULV limits the operating frequency (small current to charge/discharge a significant bitline capacitor). Forward body-biasing has been studied to boost the read current (modulation of V_{TH}) and an improved subthreshold sense amplifier has been designed (high-speed sense amplifier down to 280 mV power supply).
- The benefits of FDSOI technology compared to CMOS bulk technology has been introduced. The main limitation and the motivation for sensing data in subthreshold domain have been discussed. A differential and an unbalanced voltage sense amplifiers are designed working down to 280 mV supply voltage to improve the read time of the proposed 10T-XY bitcell. A new replica circuit with PVT tolerance is presented to reduce silicon area penalty by 10 to 20% compared to standard replica and finally an adaptive technique allows optimizing the sensing timing.
- An ultra-wide voltage memory cut is presented, including the proposed 10T-XY bitcell and other techniques. The 32kbit memory is designed in 28nm FDSOI technology. The performances in terms of operating frequency and power consumption are evaluated and compared to other ULV design works in the state-of-the-art. The results confirm that our proposal is competitive. SYPHAX CUT with a proposed BIST has been implemented in a demonstrator to evaluate the functionality, performances and yield at UWVR. Simulation results are functionally correct.

The ultra-low voltage SRAM design challenges and limitations were studied and addressed. The interactions between, the power consumption and the density increase the main limitations in SRAMs designs in sub-V_T and above-V_T. These interactions increase also the possibility to solve and meet objectives. The main objectives of this work are to overcome some limitations, which are not

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

addressed in the stat-of-the art. For example, the read failures due to the weakness of the I_{ON}-to-I_{OFF} above 400mV, which prevents the functionality at ULV. This work has provided a solutions at the level of bitcell architecture, peripheral circuit assist options and test methodologies in order to enable the possibility to consider UWVR SRAM as an efficient feasible solution by semiconductor industry to target future applications.

Perspectives

The Proposed 10T-XY bitcell is implemented in scribes (matrix of bitcells) in 28nm FDSOI technology. These scribes have been tested. The test results confirm the performances of proposed bitcell in terms of stability and leakage at ultra wide voltage range. SYPHAX memory has been extracted and validate per simulation with eldo simulator. The test-chip is under processing and it will be tested in the near future. The test allows the validation of the compliance of prediction in term of performances (functionality in sub-V_T and above-V_T, energy consumption, comparison of standard and proposed replica circuit, operating frequency, body bias benefit, the pertinence of proposed techniques). Regarding the memory development, although the simulation appears as conclusive, it remains to be confirmed by measuring silicon. The layout of the proposed 10T-XY bitcell designed with respect to logic DRC rules in 28FDSOI, in flip well configuration with a size equal to 0.62 μm^2: On additional work must be done by resizing the proposed bitcell aggressively and by applying the SRAM rule optimization (share contact…) which will offer the possibility to gain between 40 to 50% in area in term of layout of the arrays.

The aim of the thesis is to explore the feasibility and possibilities of operation of the synchronous SRAM memory under ultra-wide voltage range. However the work, which has been done on the SYPHAX memory, can give best result using alternative options as write and read assist circuits, which allow increasing performances at ultra-low voltage. In the other hand, the using of single-ended bitcell and unbalanced sense amplifier may be revisited: another SRAM design can be planned based

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

on differential 10T-XY bitcell and differential voltage sense amplifier to offer better performances. Finally, there are a lot of research axes that can be explored for the master of the ULV domain and to improve the energy efficiency as the leakage recycling in sub-V_T, the improvement of the test methodology at ULV and by exploring new options, which are offered by the advanced new technology like the 14nm FDSOI and FINFET technologies.

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Reference

[1] *D. Khang, and M. M. Atalla, IRE Solid-State Devices Res. Conf., Carnegie Institute of Technology, Pittsburg, Pa. (1960)*

[2] *http://www.intel.com/content/www/us/en/history/museum-gordon-moore-law.html*

[3] *IntI. Technology Roadmap for Semiconductor 2007 Edition Executive Summary. http://www.itrs.netiLinks/2007ITRS/ExecSum2007.pdf*

[4] *Haron, N.Z.; Hamdioui, S., "Why is CMOS scaling coming to an END?," Design and Test Workshop, 2008. IDT 2008. 3rd International , vol., no., pp.98,103, 20-22 Dec. 2008*

[5] *E.J. Nowak, "Maintaining the Benefits of CMOS Scaling When Scaling Bogs Down", IBM JRD, vol 46, no 2/3, 2*

[6] *Mark White, Yuan Chen Scaled « CMOS Technology Reliability Users Guide » National Aeronautics and Space Administration. JPL Publication 08-14 3/08*

[7] *Steven A. Vitale, Peter W. Wyatt, Member IEEE, Nisha Checka, Jakub Kedzierski, and Craig L. Keast "FDSOI Process Technology for Subthreshold-Operation Ultralow-Power Electronics"*

[8] *A. Uchiyama, S. Baba, Y. Nagatomo, and J. Ida, BFully depleted SOI technology for ultra low power digital and RF applications,in 2006 IEEE Int. SOI Conf. Proc., Oct. 2006, pp. 15–16.*

[9] *A. Ebina, T. Kadowaki, Y. Sato, and M. Yamaguchi, BUltra low-power CMOS IC using partially-depleted SOI technology, in Proc. Custom Integrated Circuits Conf., May 2000, pp. 57–60.*

[10] *J. L. Pelloie, C. Raynaud, O. Faynot, A. Grouillet, and J. Du Port de Pntcharra, BCMOS/SOI technologies for low-power and low-voltage circuits,[Microelectronic Engineering, vol. 48, pp. 327–334, 1999.*

[11] *N. Planes et al, p333, VLSI 2012*

[12] *Mayur Bhole, Aditya Kurude, Sagar Pawar « FinFET- Benefits, Drawbacks and Challenges » November, 2013, IJESRT*

[13] *Hook, T.B., "Fully depleted devices for designers: FDSOI and FinFETs," Custom Integrated Circuits Conference (CICC), 2012 IEEE , vol., no., pp.1,7, 9-12 Sept. 2012*

[14] *Naran Sirisantana, Liqiong Wei, Kaushik Roy, High-Performance Low- Power CMOS Circuits Using Multiple Channel Length and Multiple Oxide Thickness, ICCD, 2000, pp. 227-232*

[15] *Takeo Yamashita et al., A 450MHz 64b RISC Processor using Multiple Threshold Voltage CMOS, ISSCC, 2000, session 25.3*

[16] *Meeta Srivastav, Prof. S.S.S.O. Rao, Himanshu Bhatnagar, Power Reduction Technique using Multi-vt libraries, IDEAS, 2005, pp. 363-367*

[17] *Tadahiro Kuroda et al., A 0.9V 150MHz 10mW 4mm. 2-D Discrete Cosine Transform Core Processor with Variable-Threshold-Voltage Scheme, ISSCC, 1996, pp. 166-168*

[18] *Swanson, R.; Meindl, J., "Ion-implanted complementary MOS transistors in low-voltage circuits," Solid-State Circuits Conference. Digest of Technical Papers. 1972 IEEE International, vol.XV, no., pp.192,193, 16-18 Feb. 1972*

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

[19] *Hanson, S.; Bo Zhai; Mingoo Seok; Cline, B.; Zhou, K.; Singhal, M.; Minuth, M.; Olson, J.; Nazhandali, L.; Austin, T.; Sylvester, D.; Blaauw, D., "Performance and Variability Optimization Strategies in a Sub-200mV, 3.5pJ/inst, 11nW Subthreshold Processor," VLSI Circuits, 2007 IEEE Symposium on , vol., no., pp.152,153, 14-16 June 2007*

[20] *Wang, A.; Chandrakasan, A., "A 180mV FFT processor using subthreshold circuit techniques," Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International , vol., no., pp.292,529 Vol.1, 15-19 Feb. 2004*

[21] *Verma, N.; Kwong, J.; Chandrakasan, A.P., "Nanometer MOSFET Variation in Minimum Energy Subthreshold Circuits," Electron Devices, IEEE Transactions on , vol.55, no.1, pp.163,174, Jan. 2008*

[22] *Sinangil, M.E.; Verma, N.; Chandrakasan, A.P., "A Reconfigurable 8T Ultra-Dynamic Voltage Scalable (U-DVS) SRAM in 65 nm CMOS," Solid-State Circuits, IEEE Journal of , vol.44, no.11, pp.3163,3173, Nov. 2009*

[23] *Sharma et al., 8T SRAM with Mimicked Negative Bitlines and Charge limited Sequential Sense Amplifier for Wireless Sensor Nodes. Proceedings of IEEE European Solid-State Circuits Conference (ESSCIRC), pp. 531–534, Sept 2011b*

[24] *Ming-Hsien Tu; Jihi-Yu Lin; Ming-Chien Tsai; Chien-Yu Lu; Yuh-Jiun Lin; Meng-Hsueh Wang; Huan-Shun Huang; Kuen-Di Lee; Wei-Chiang Shih; Shyh-Jye Jou; Ching-Te Chuang, "A Single-Ended Disturb-Free 9T Subthreshold SRAM With Cross-Point Data-Aware Write Word-Line Structure, Negative Bitline, and Adaptive Read Operation Timing Tracing," Solid-State Circuits, IEEE Journal of , vol.47, no.6, pp.1469,1482, June 2012*

[25] *Suzuki, T.; Yamagami, Y.; Hatanaka, I.; Shibayama, A.; Akamatsu, H.; Yamauchi, H., "0.3 to 1.5V embedded SRAM with device-fluctuation-tolerant access-control and cosmic-ray-immune hidden-ECC scheme,"Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International, vol., no., pp.484,612 Vol. 1, 10-10 Feb. 2005*

[26] *Verma, N.; Chandrakasan, A.P., "A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy," Solid-State Circuits, IEEE Journal, vol.43, no.1, pp.141,149, Jan. 2008*

[27] *Abouzeid, F.; Bienfait, A.; Akyel, K.C.; Feki, A.; Clerc, S.; Ciampolini, L.; Giner, F.; Wilson, R.; Roche, P., "Scalable 0.35 V to 1.2 V SRAM Bitcell Design From 65 nm CMOS to 28 nm FDSOI,"Solid-State Circuits, IEEE Journal of , vol.49, no.7, pp.1499,1505, July 2014*

[28] *Chien-Fu Chen; Ting-Hao Chang; Lai-Fu Chen; Meng-Fan Chang; Yamauchi, H., "A 210mV 7.3MHz 8T SRAM with dual data-aware write-assists and negative read wordline for high cell-stability, speed and area-efficiency," VLSI Technology (VLSIT), 2013 Symposium on , vol., no., pp.C130,C131, 11-13 June 2013*

[29] *Multi-Core Processors: New Way to Achieve High System Performance Gepner, P. ; Kowalik, M.F. Parallel Computing in Electrical Engineering, 2006. PAR ELEC 2006. International Symposium on Digital, 2006*

[30] *Pelgrom, M.J.M.; Duinmaijer, Aad C J; Welbers, A.P.G., "Matching properties of MOS transistors," Solid-State Circuits, IEEE Journal of , vol.24, no.5, pp.1433,1439, Oct 1989*

[31] *Roy, K.; Mukhopadhyay, S.; Mahmoodi-Meimand, H., "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," Proceedings of the IEEE , vol.91, no.2, pp.305,327, Feb 2003*

[32] *Jain, S.; Khare, S.; Yada, S.; Ambili, V.; Salihundam, P.; Ramani, S.; Muthukumar, S.; Srinivasan, M.; Kumar, A; Gb, S.K.; Ramanarayanan, R.; Erraguntla, V.; Howard, J.; Vangal, S.; Dighe, S.; Ruhl, G.; Aseron, P.;*

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

Wilson, H.; Borkar, N.; De, V.; Borkar, S., "A 280mV-to-1.2V wide-operating-range IA-32 processor in 32nm CMOS," Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International , vol., no., pp.66,68, 19-23 Feb. 2012

[33] *Jiajing Wang; Nalam, S.; Calhoun, B.H., "Analyzing static and dynamic write margin for nanometer SRAMs," Low Power Electronics and Design (ISLPED), 2008 ACM/IEEE International Symposium on , vol., no., pp.129,134, 11-13 Aug. 2008*

[34] *Wieckowski, M.; Sylvester, D; Blaauw, D.; Chandra, V.; Idgunji, S.; Pietrzyk, C.; Aitken, R., "A black box method for stability analysis of arbitrary SRAM cell structures," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2010 , vol., no., pp.795,800, 8-12 March 2010*

[35] *Hauser, John R., "Noise margin criteria for digital logic circuits," Education, IEEE Transactions on, vol.36, no.4, pp.363,368, Nov 1993*

[36] *Seevinck, E.; List, F.J.; Lohstroh, J., "Static-noise margin analysis of MOS SRAM cells," Solid-State Circuits, IEEE Journal of , vol.22, no.5, pp.748,754, Oct 1987*

[37] *Vatajelu, E.I.; Figueras, J., "Statistical analysis of 6T SRAM data retention voltage under process variation," Design and Diagnostics of Electronic Circuits & Systems (DDECS), 2011 IEEE 14th International Symposium on , vol., no., pp.365,370, 13-15 April 2011*

[38] *Yamaoka, M.; Maeda, N.; Shinozaki, Y.; Shimazaki, Y.; Nii, K.; Shimada, S.; Yanagisawa, K.; Kawahara, T., "Low-power embedded SRAM modules with expanded margins for writing," Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International , vol., no., pp.480,611 Vol. 1, 10-10 Feb. 2005*

[39] *Young Hwi Yang; Jisu Kim; Hyunkook Park; Wang, J.; Yeap, G.; Seong-Ook Jung, "SRAM bitcell design for low voltage operation in deep submicron technologies," IC Design & Technology (ICICDT), 2011 IEEE International Conference on , vol., no., pp.1,4, 2-4 May 2011*

[40] *Shibata, N.; Kiya, H.; Kurita, S.; Okamoto, H.; Tan'no, M.; Douseki, T., "A 0.5-V 25-MHz 1-mW 256-kb MTCMOS/SOI SRAM for solar-power-operated portable personal digital equipment - sure write operation by using step-down negatively overdriven bitline scheme," Solid-State Circuits, IEEE Journal of , vol.41, no.3, pp.728,742, March 2006*

[41] *Hirabayashi, O.; Kawasumi, A.; Suzuki, A.; Takeyama, Y.; Kushida, K.; Sasaki, T.; Katayama, A.; Fukano, G.; Fujimura, Y.; Nakazato, T.; Shizuki, Y.; Kushiyama, N.; Yabe, T., "A process-variation-tolerant dual-power-supply SRAM with 0.179μm^2 Cell in 40nm CMOS using level-programmable wordline driver," Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International , vol., no., pp.458,459,459a, 8-12 Feb. 2009*

[42] *Yamaoka, M.; Osada, K.; Ishibashi, K., "0.4-V logic library friendly SRAM array using rectangular-diffusion cell and delta-boosted-array-voltage scheme," VLSI Circuits Digest of Technical Papers, 2002. Symposium on , vol., no., pp.170,173, 13-15 June 2002*

[43] *Ohbayashi, S.; Yabuuchi, M.; Nii, K.; Tsukamoto, Y.; Imaoka, S.; Oda, Y.; Yoshihara, T.; Igarashi, M.; Takeuchi, M.; Kawashima, H.; Yamaguchi, Y.; Tsukamoto, K.; Inuishi, M.; Makino, H.; Ishibashi, K.; Shinohara, H., "A 65-nm SoC Embedded 6T-SRAM Designed for Manufacturability With Read and Write Operation Stabilizing Circuits," Solid-State Circuits, IEEE Journal of , vol.42, no.4, pp.820,829, April 2007*

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

[44] *Nii, K.; Yabuuchi, M.; Tsukamoto, Y.; Ohbayashi, S.; Oda, Y.; Usui, K.; Kawamura, T.; Tsuboi, N.; Iwasaki, T.; Hashimoto, K.; Makino, H.; Shinohara, H., "A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment," VLSI Circuits, 2008 IEEE Symposium on , vol., no., pp.212,213, 18-20 June 2008*

[45] *Ik Joon Chang; Jae-Joon Kim; Sang Phill Park; Roy, K., "A 32kb 10T Subthreshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90nm CMOS," Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International , vol., no., pp.388,622, 3-7 Feb. 2008*

[46] *Yabuuchi et al., A 45 nm 0.6 V Cross-Point 8T SRAM with Negative Biased Read/Write Assist, Symposium on VLSI Circuits Digest of Technical Papers, pp. 158–159 (2009)*

[47] *S. Tawfik and V. Kursun, "Low power and robust 7T dual-VT SRAM circuit," in Proc. Int. Symp. Circuits Syst., 2008, pp. 1452–1455.*

[48] *T. Suzuki, H. Yamauchi, Y. Yamagami, K. Satomi, and H. Akamatsu, "A stable 2-port SRAM cell design against simultaneously read/writedisturbed accesses," IEEE J. Solid-State Circuits, vol. 43, no. 9, pp. 2109–2119, Sep. 2008.*

[49] *Chang, L.; Fried, D.M.; Hergenrother, J.; Sleight, J.W.; Dennard, R.H.; Montoye, R.K.; Sekaric, Lidija; McNab, S.J.; Topol, A.W.; Adams, C.D.; Guarini, K.W.; Haensch, W., "Stable SRAM cell design for the 32 nm node and beyond," VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on , vol., no., pp.128,129, 14-16 June 2005*

[50] *Morita, Y.; Fujiwara, H.; Noguchi, H.; Iguchi, Y.; Nii, K.; Kawaguchi, H.; Yoshimoto, M., "An Area-Conscious Low-Voltage-Oriented 8T-SRAM Design under DVS Environment," VLSI Circuits, 2007 IEEE Symposium on , vol., no., pp.256,257, 14-16 June 2007*

[51] *Noguchi, H.; Okumura, S.; Iguchi, Y.; Fujiwara, H.; Morita, Y.; Nii, K.; Kawaguchi, H.; Yoshimoto, M., "Which is the best dual-port SRAM in 45-nm process technology? — 8T, 10T single end, and 10T differential —," Integrated Circuit Design and Technology and Tutorial, 2008. ICICDT 2008. IEEE International Conference on , vol., no., pp.55,58, 2-4 June 2008*

[52] *Jui-Jen Wu; Yen-Huei Chen; Meng-Fan Chang; Po-Wei Chou; Chien-Yuan Chen; Hung-Jen Liao; Chu; Wen-Chin Wu; Yamauchi, H., "A Large σVTH /VDD Tolerant Zigzag 8T SRAM With Area-Efficient Decoupled Differential Sensing and Fast Write-Back Scheme," Solid-State Circuits, IEEE Journal of , vol.46, no.4, pp.815,827, April 2011*

[53] *K. Zhang, K. Hose, V. De, and B. Senyk, "The scaling of data sensing schemes for high speed cache design in sub-0.18µm technologies," in Proc. IEEE Symp.VLSI Circuits, June 2000, pp. 226–227.*

[54] *Optimization of a Voltage Sense Amplifier operating in Ultra Wide Voltage Range with Back Bias Design Techniques in 28nm UTBB FD-SOI Technology*

[55] *Chiu YW, Lin JY, Tu MH, Jou SJ, Chuang CT. 8T single-ended sub-threshold SRAM with cross-point data-aware write operation. In: Proc IEEE Symp on Low Power Electronics and Design ISLPED 2011; 2011. p.169-74.*

[56] *Tu MH, Lin JY, Tsai MC, Lu CY, Lin YJ, Wang MH, et al. A single-ended disturb-free 9T subthreshold SRAM with cross-point data-aware write word-line structure, negative bitline, and adaptive read operation timing tracing. IEEE J. Solid-State Circ 2012; 47(6):1469-82.*

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

[57] *Chiu YW, Hu YH, Tu MH, Zhao JK, Jou SJ, Chuang CT. A 40nm 0.32V 3.5MHz 11T Single-Ended Bit-Interleaving Subthreshold SRAM with Data-Aware Write-Assist. In: Proc IEEE Symp on Low Power Electronics and Design ISLPED 2013; 2013. p.51-6.*

[58] *Calhoun BH, Chandrakasan A. A 256 kb subthreshold SRAM using 65 nm CMOS. In: Proc IEEE Int. Solid-State Circuits Conf ISSCC 2006; 2006. p. 628–29.*

[59] *Song T, Kim S, Lim K, Laskar J. Fully-gated ground 10T-SRAM bitcell in 45 nm SOI technology. Electronics Letters 2010;46(7):515 -16.*

[60] *Slayman C. Soft Errors – Past History and Recent Discoveries. In: Proc IEEE International Integrated Reliability Workshop IIRW 2010; 2010. p.25-30.*

[61] *Geppert L. A static RAM says goodbye to data errors [radiation induced soft errors]. IEEE Spectrum 2004;41(2):16-17.*

[62] *Chandra V, Aitken R. Impact of Technology and Voltage Scaling on the Soft Error Susceptibility in Nanoscale CMOS. In: Proc IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems DFTVS '08; 2008. p.114-22.*

[63] *Daeyeon Kim; Chandra, V.; Aitken, R.; Blaauw, D.; Sylvester, D., "Variation-aware static and dynamic writability analysis for voltage-scaled bit-interleaved 8-T SRAMs," Low Power Electronics and Design (ISLPED) 2011 International Symposium on , vol., no., pp.145,150, 1-3 Aug. 2011*

[64] *V. Chandra, R. Aitken, "Impact of voltage scaling on nanoscale SRAM reliability," Design, Automation & Test in Europe, pp. 387- 392, Apr, 2009*

[65] *R.W. Hamming, "Error detecting and error correcting codes," Bell System Technical Journal, Vol. 29, pp. 147-160, Apr. 1950*

[66] *R. Naseer, J. Draper, "Parallel double error correcting code design to mitigate multi-bit upsets in SRAMs," European Solid-States Circuits Conference, pp. 222-225, Sep. 2008*

[67] *67 Abouzeid F, Clerc S, Pelloux-Prayer B, Roche, P. 0.42-to-1.20V read assist circuit for SRAMs in CMOS 65nm. In: Proc IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference S3S 2013; 2013. p.1-2.*

[68] *68 Sinangil ME, Verma N, Chandrakasan AP. A reconfigurable 65nm SRAM achieving voltage scalability from 0.25–1.2V and performance scalability from 20kHz–200MHz. In: Proc 34th European Solid-State Circuits Conference ESSCIRC 2008; 2008. p.282-85.*

[69] *Planar fully depleted silicon technology to design competitive SOC at 28nm and beyond, www.soitec.com/*

[70] *Flatresse, P.; Giraud, B.; Noel, J.; Pelloux-Prayer, B.; Giner, F.; Arora, D.; Arnaud, F.; Planes, N.; Le Coz, J.; Thomas, O.; Engels, S.; Cesana, G.; Wilson, R.; Urard, P., "Ultra-wide body-bias range LDPC decoder in 28nm UTBB FDSOI technology," Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International , vol., no., pp.424,425, 17-21 Feb. 2013*

[71] *Evaluation of Differential vs. Single-Ended Sensing and Asymmetric Cells in 90nm Logic Technology for On-Chip Caches*

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

[72] *Hu, V. P-H.; Fan, M.-L.; Su, P.; Chuang, C.-T., "Threshold Voltage Design of UTB SOI SRAM with Improved Stability/Variability for Ultra-Low Voltage Near Subthreshold Operation," Nanotechnology, IEEE Transactions on , vol.PP, no.99, pp.1,1, 0*

[73] *Fenouillet-Beranger, C.; Perreau, P.; Pham-Nguyen, L.; Denorme, S.; Andrieu, F.; Tosti, L.; Brevard, L.; Weber, O.; Barnola, S.; Salvetat, T.; Garros, X.; Casse, M.; Leroux, C.; Noel, J.P.; Thomas, O.; Le-Gratiet, B.; Baron, F.; Gatefait, M.; Campidelli, Y.; Abbate, F.; Perrot, C.; De-Buttet, C.; Beneyton, R.; Pinzelli, L.; Leverd, F.; Gouraud, P.; Gros-Jean, M.; Bajolet, A.; Mezzomo, C.; Leyris, C.; Haendler, S.; Noblet, D.; Pantel, R.; Margain, A.; Borowiak, C.; Josse, E.; Planes, N.; Delprat, D.; Boedt, F.; Bourdelle, K.; Nguyen, B.Y.; Boeuf, F.; Faynot, O.; Skotnicki, T., "Hybrid FDSOI/bulk High-k/metal gate platform for low power (LP) multimedia technology," Electron Devices Meeting (IEDM), 2009 IEEE International , vol., no., pp.1,4, 7-9 Dec. 2009*

[74] *Zhang, K.; Hose, K.; De, V.; Senyk, B., "The scaling of data sensing schemes for high speed cache design in sub-0.18 /spl mu/m technologies," VLSI Circuits, 2000. Digest of Technical Papers. 2000 Symposium on , vol., no., pp.226,227, 15-17 June 2000*

[75] *R. Sarpeshkar, J. L. Wyatt, N. C. Lu, and P. D. Gerber, "Mismatch sensitivity of a simultaneously latched CMOS sense amplifier," IEEE J. Solid-State Circuits, vol. 26, no. 10, pp. 1413–1422, Oct. 1991.*

[76] *B. Wicht, T. Nirschl, and D. Schmitt-Landsiedel, "Yield and speed optimization of a latch-type voltage sense amplifier," IEEE J. Solid-State Circuits, vol. 39, no. 7, pp. 1148–1158, Jul. 2004.*

[77] *J. R. Cavaliere, W. J. Scarpero,"Sense Amplifier," US Patent, N. 3879621, April 22, 1975*

[78] *Kobayashi, T.; Nogami, K.; Shirotori, T.; Fujimoto, Y., "A current-controlled latch sense amplifier and a static power-saving input buffer for low-power architecture," Solid-State Circuits, IEEE Journal of , vol.28, no.4, pp.523,527, Apr 1993*

[79] *Woo, S.-H.; Kang, H.; Park, K.; Jung, S.-O., "Offset voltage estimation model for latch-type sense amplifiers," Circuits, Devices & Systems, IET , vol.4, no.6, pp.503,513, November 2010*

[80] *Comparative Study of Various Latch-Type Sense Amplifiers*

[81] *Muhammad M. Khellah, Dinesh Somasekhar, Yibin Ye, Ali R. Farhang, Gunjan H. Pandya, Vivek K. De "SRAM with forward body biasing to improve read cell stability" US 6985380 B2*

[82] *Komatsu, S.; Yamaoka, M.; Morimoto, M.; Maeda, N.; Shimazaki, Y.; Osada, K., "A 40-nm low-power SRAM with multi-stage replica-bitline technique for reducing timing variation," Custom Integrated Circuits Conference, 2009. CICC '09. IEEE , vol., no., pp.701,704, 13-16 Sept. 2009*

[83] *Arslan, U.; McCartney, M.P.; Bhargava, M.; Xin Li; Ken Mai; Pileggi, L.T., "Variation-tolerant SRAM sense-amplifier timing using configurable replica bitlines," Custom Integrated Circuits Conference, 2008. CICC 2008. IEEE , vol., no., pp.415,418, 21-24 Sept. 2008*

[84] *Amrutur, B.S.; Horowitz, M., "Techniques to reduce power in fast wide memories [CMOS SRAMs]," Low Power Electronics, 1994. Digest of Technical Papers., IEEE Symposium , vol., no., pp.92,93, 10-12 Oct. 1994*

[85] *Gupta, S.; Rana, P.K., "A 28nm 6T SRAM memory compiler with a variation tolerant replica circuit," SoC Design Conference (ISOCC), 2012 International , vol., no., pp.458,461, 4-7 Nov. 2012*

[86] *Amrutur, B.S.; Horowitz, M.A., "A replica technique for wordline and sense control in low-power SRAM's," Solid-State Circuits, IEEE Journal of , vol.33, no.8, pp.1208,1219, Aug 1998*

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

[87] *Zhongyuan Wu; Zhiqiang Gao; Xiangqing He, "A high performance embedded SRAM compiler," ASIC, 2003. Proceedings. 5th International Conference on , vol.1, no., pp.470,473 Vol.1, 21-24 Oct. 2003*

[88] *Chris Jacobsen, Tony Saye, Ken Parker "In-circuit Testing of Low Voltage Devices Technical Paper " Agilent Technologies, Inc.*

[89] *L. Wong, S. Hossain, A. Ta, J. Edvinsson, D. Rivas, and H. Naas, "A very lowpower CMOS mixed-signal IC for implantable pacemaker applications," IEEE Journal of Solid-State Circuits, vol. 39, no. 12, pp. 2446–2456, 2004.*

[90] *L. Padeletti and S. S. Barold, "Digital technology for cardiac pacing," The American Journal of Cardiology, vol. 95, no. 4, pp. 479–482, Feb. 2005.*

[91] *S. Kim, N. Cho, S.-J. Song, D. Kim, K. Kim, and H.-J. Yoo, "A 0.9-V 96-µW digital hearing aid chip with heterogeneous $-# DAC," in Proc. IEEE Symp. VLSI Circuits, June 2006, pp. 55–56.*

[92] *H. Neuteboom, B. M. J. Kup, and M. Janssens, "A DSP based hearing instrument IC," IEEE Journal of Solid-State Circuits, vol. 32, no. 11, pp. 1790–1806, Nov. 1997.*

[93] *J. Georgiou and C. Toumazou, "A 126-µW cochlear chip for a totally implantable system," IEEE Journal of Solid-State Circuits, vol. 40, no. 2, pp. 430–443, Feb. 2005.*

[94] *B. Gyselinckx, C. Van Hoof, J. Ryckaert, R. Yazicioglu, P. Fiorini, and V. Leonov, "Human++: autonomous wireless sensors for body area networks," in Proc. IEEE Custom Integrated Circuits Conference, 2005, pp. 13–19.*

[95] *Ik Joon Chang; Jae-Joon Kim; Sang Phill Park; Roy, K., "A 32kb 10T Subthreshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90nm CMOS," Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International , vol., no., pp.388,622, 3-7 Feb. 2008*

[96] *Maiz, J.; Hareland, S.; Zhang, K.; Armstrong, P., "Characterization of multi-bit soft error events in advanced SRAMs," Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International , vol., no., pp.21.4.1,21.4.4, 8-10 Dec. 2003*

[97] *Takeda, K.; Hagihara, Y.; Aimoto, Y.; Nomura, M.; Nakazawa, Y.; Ishii, T.; Kobatake, H., "A read-static-noise-margin-free SRAM cell for low-VDD and high-speed applications," Solid-State Circuits, IEEE Journal of , vol.41, no.1, pp.113,121, Jan. 2006*

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

FOLIO ADMINISTRATIF

THÈSE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE LYON

NOM : **FEKI** DATE de SOUTENANCE : 29-05-2015
Prénom : **ANIS**

TITRE : Conception **d'une mémire SRAM en tension sous le seuil pour des applications biomédicales et les noeuds de capteurs sans fils en technologies CMOS avancées.**

NATURE : **Doctorat** Numéro d'ordre : 2015ISAL0018

Ecole doctorale : **ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE (E.E.A)**
Spécialité : **Micro et nano-électronique**

RESUME :

L'émergence des circuits complexes numériques, ou System-On-Chip (SOC), pose notamment la problématique de la consommation énergétique. Parmi les blocs fonctionnels significatifs à ce titre, apparaissent les mémoires et en particulier les mémoires statiques (SRAM). La maîtrise de la consommation énergétique d'une mémoire SRAM inclue la capacité à rendre la mémoire fonctionnelle sous très faible tension d'alimentation, avec un objectif agressif de 300 mV (inférieur à la tension de seuil des transistors standard CMOS).
Dans ce contexte, les travaux de thèse ont concerné la proposition d'un point mémoire SRAM suffisamment performant sous très faible tension d'alimentation et pour les nœuds technologiques avancés (CMOS bulk 28nm et FDSOI 28nm). Une analyse comparative des architectures proposées dans l'état de l'art a permis d'élaborer deux points mémoire à 10 transistors avec de très faibles impacts de courant de fuite. Outre une segmentation des ports de lecture, les propositions reposent sur l'utilisation de périphéries adaptées synchrones avec notamment une solution nouvelle de réplication, un amplificateur de lecture de données en mode tension et l'utilisation d'une polarisation dynamique arrière du caisson SOI (Body Bias).

MOTS-CLÉS: **Ultra-low voltage, SRAM, Subthreshold, bitcells, UWVR, Sense amplifier, circuit replica.**

Laboratoire (s) de recherche : **Laboratoire Ampère (INSA-Lyon)**

Directeur de thèse : Professeur **Bruno ALLARD (INSA-Lyon)**
: Mr. **David TURGIS SRAM MEMORY Expert (STMicroelectroncis)**

Président de jury : Professeur **Jean-Michel PORTAL**
Composition du jury : Professeur **Jean-Michel PORTAL**
Professeur Pascal NOUET
Assistant Professeur Luca LARCHER
Professeur Bruno ALLARD
David TURGIS
Dr Olivier Thomas

Cette thèse est accessible à l'adresse : http://theses.insa-lyon.fr/publication/2015ISAL0018/these.pdf
© [A. Feki], [2015], INSA Lyon, tous droits réservés