TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Technische Elektronik

Power Efficient and Robust Sense Amplifiers for Embedded Non-Volatile Memories in High-Speed Microcontrollers for Automotive Applications

Mihail Valer'evic Jefremow

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Georg Sigl

Prüfer der Dissertation:

1. Univ.-Prof. Dr. rer. nat. Doris Schmitt-Landsiedel
2. Univ.-Prof. Dr.-Ing. Heinrich Klar,
   Technische Universität Berlin

Die Dissertation wurde am 23.12.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 03.11.2014 angenommen.

Preface

Automotive microcontroller units (MCUs) have highly complex system on a chip (SoC) architectures. Besides CPU cores and memory they comprise a variety of peripherals like different ADC architectures and communication interfaces like CAN and FlexRay to interconnect e.g. the huge amount of controllers in a car. A main feature and also differentiator of the modern MCU is the non-volatile embedded memory to support fast and secure autonomous operation. The high end applications like powertrain, transmission and advanced driver assistance systems (ADAS) are the MCU performance drivers and require a steady increase of the real-time performance. To support these needs multi-core architectures were introduced for the MCU. It allows scalable performance increase depending on the number of CPU cores without an increase in system clock frequency enabling low power operation. The next main architectural evolution step would be the 64bit architecture [4]. The overall system performance heavily depends on read speed of the embedded non-volatile memory containing instruction code and calibration data. The read speed, data throughput and size of the embedded non-volatile memory have to increase accordingly to avoid a bottleneck for the system performance. But at the same time the number of MCUs per car is rapidly increasing. For example a premium car comprises today more than 100 MCUs [5] requiring a considerable amount of power for memory operation. Therefore new design approaches in MCUs and in embedded memory modules are required to allow further performance increase and higher memory sizes with even lower power consumption to support the performance/power demands of automotive industry.

This work addresses the circuit design of the read path for the embedded non-volatile memory, in particular the design of the sense amplifier (SA). The SA determines read speed and power consumption of the memory module during the read operation. This thesis also provides an overview on the existing solutions from both practical and analytical perspective. Various circuit concepts are discussed and design guidelines are given, pros and cons for particular solutions are assessed. Based on the outcome of this analysis the novel concept and circuit implementation allowing bitline-capacitance-cancellation for higher memory density under low power and high speed sensing operation is introduced. It is analyzed analytically and results are proven by measurement on two different chips. Future trends in memory technology development and implications on circuit design are investigated exemplary for embedded STT-MRAM in a similar manner. To fulfill the specific memory requirements a new circuit exploiting a time-differential approach is developed to provide higher read yield and robustness for STT-MRAM. In addition all

newly developed circuits in this work are compared to state-of-the-art approaches by measurement based on testchips and products. Moreover the content of this work is not only limited to embedded memories in automotive applications but can also be deployed for all other types of applications as industrial and consumer memory solutions. To cover these different application fields, circuit solutions not only for high speed but also for low power and low voltage constraints are exploited. The circuits and architectures presented are also applicable for various kinds of chips with array organizations like stand-alone NAND- or NOR-Flash memories, CMOS imagers or FPGAs.

The basic memory cell structure, its characteristics and the array organization of the embedded Flash is discussed in chapter 2. Chapter 3 covers the the state-of-the art sensing approaches for embedded Flash. Starting from the implications on a memory cell the sense amplifier concepts and circuits are discussed from a practical point of view followed by deriving an analytical circuit model for prediction of speed and power consumption. Chapter 4 introduces the bitline-capacitance-cancellation sensing circuit for embedded Flash memory developed in this work and uses the same analytical approach as in the previous chapter. Chapter 4 finishes with measurement results obtained by the testchip confirming the high speed capability of this circuit. The embedded Flash memory macro design for high speed and low power operation is discussed in chapter 5. The multi voltage domain global to local multiplexer and the local ground read circuit design are introduced in combination with the proposed sense amplifier circuit to achieve best in class read throughput in conjunction with low power operation. Finally the memory module developed in this work is compared by measurement results to the state-of-the-art approaches to demonstrate the respective benefits. The last chapter focuses on the sense amplifier circuit design for STT-MRAM, which is one of the promising candidates for the future non-volatile memories. Based on the analysis of the STT-MRAM-cell characteristics, a time-differential sense amplifier circuit is developed, confirming higher read yield compared to the state-of-the-art approach by measurements.

Contents

Abbreviations and Symbols **1**

1 Introduction **9**

2 Embedded NOR-Flash **13**

2.1 Embedded Flash Memory Cell 13

2.2 NOR-Flash Array Architecture 14

2.3 Summary 16

3 Voltage vs. Current Sensing Scheme **17**

3.1 Concepts 17

3.2 Sensing on Drain- or Source-Side 19

3.3 Sensing Circuits 21

3.3.1 Voltage Sensing 21

3.3.2 Current Sensing 25

3.4 Read Access Time Considerations 29

3.4.1 Read Path Model 29

3.4.2 Voltage Sensing 31

3.4.3 Current Sensing 34

3.4.4 Comparison 36

3.5 Power Considerations 38

3.5.1 Voltage Sensing 39

3.5.2 Current Sensing 40

3.5.3 Comparison 41

3.6 Summary and Conclusion 42

4 Bitline-Capacitance-Cancellation Sensing Scheme **45**

4.1 Time Domain Sensing Concept . . . 45
4.2 Voltage to Time Converter Circuits . . . 48
4.2.1 Comparator Based Approach . . . 48
4.2.2 Charge Transfer Sense Amplifier . . . 50
4.2.3 Common Gate Sense Amplifier . . . 51
4.3 Slope Detection Circuit . . . 54
4.4 Bitline-Capacitance-Cancellation Sense Amplifier Circuit . . . 62
4.5 Source-Side Sensing for 65nm HS3P eFlash . . . 67
4.6 Sensing Scheme Architecture and Measurement Results . . . 70
4.7 Power Considerations . . . 75
4.8 Comparison . . . 78
4.9 Summary and Conclusion . . . 80

5 Embedded Flash Macro Design Aspects 82
5.1 Macro Architecture . . . 82
5.2 Memory Array . . . 82
5.3 Multi Voltage Domain Global to Local Multiplexer . . . 84
5.4 Local Ground Referenced Read Circuit Design . . . 86
5.5 Measurements and Sensing Schemes Comparison . . . 86

6 Sensing for Embedded STT-MRAM 92
6.1 Embedded STT-MRAM . . . 92
6.2 State-of-the-Art Sensing Schemes . . . 94
6.3 Time-Differential Sensing Scheme . . . 96
6.3.1 Organization . . . 96
6.3.2 Time-Differential Sense Amplifier Circuit . . . 97
6.4 Sensing Scheme Comparison and Measurement Results . . . 100
6.5 Summary . . . 101

7 Conclusion and Outlook 103

A Author Publications 105

B Slope Detection Transfer Function 106

C Acknowledgment 108

Bibliography **111**

Abbreviations and Symbols

a	number of physical sectors in a non-volatile memory
A_0	small-signal gain of an operational amplifier utilized in current sense amplifier circuit of Type D or B [6]
A_{CT}	voltage change at the output of the charge transfer sense amplifier (see figure 4.5)
ADC	analog to digital converter
ASB	assembly buffer latch required for write operation in non-volatile memory macro
ATD	address transition detection block in figure 5.4
b	bitline multiplexer factor in a non-volatile memory
BL	bitline
BL_{REF}	reference bitline
c	number of wordlines (wordline width) in a non-volatile memory
C_{BL}	bitline capacitance
C_{BLd}	capacitance per unit length of a bitline
C_F	capacitance used in slope dependent current source circuit in figure 4.13, 4.15, 4.19 and 4.21
CG	control gate terminal of the HS3P-cell
C_{GBL}	global bitline capacitance of a memory array utilizing the hierarchical bitline architecture
C_{GES}	sum of the coupling capacitance values from floating gate to each terminal of a non-volatile memory cell transistor
C_{INT}	reference capacitance used for discharge timing generation for local ground referenced read circuits in figure 5.4
C_{load}	load capacitance of a sense amplifier, comparator or an operational amplifier
C_{par}	overall capacitance of the LVLS node in the level-shifter of the slope detection circuit in figure 4.19, which is dominated by a parasitic capacitance
CPU	Central processing unit
CS	common terminal of all HS3P-eFlash cells in one physical sector
d	number of sense amplifiers (read word width) in a non-volatile memory

$DISGEN$	discharge time generator generating the discharge timing for the bitline-capacitance-cancellation sensing scheme in figure 5.4
FOM	figure of merit
FOM_{PS}	figure of merit for power efficiency and speed of a sense amplifier circuit
$FOMSP$	figure of merit for sensing performance for a sensing scheme scaling the access time to the global bitline length
GBL	global bitline of a memory array utilizing the hierarchical bitline architecture
GBL_{REF}	global reference bitline
GBL_{REP}	replica of the global bitline used in the address transition detection block in figure 5.4
GBW	gain-bandwidth product of an operational amplifier
$GLOLO$	global to local multiplexer for the hierarchical bitline architecture multiplexing several local bitlines LBL to one global bitline GBL
g_m	small-signal transistor transconductance
g_{m*}	sum of small-signal transconductances for the bulk-source (g_{mb}) and the gate-source (g_m) voltage of a transistor
g_{mb}	small-signal transconductance for the bulk-source of a transistor
$HS3P$	hot source triple poly embedded Flash memory cell
HV	high voltage domain
I/O	input/output circuits used in a pads of a chip
I_{BIAS}	sense amplifier or a comparator circuit bias current
$I_{BIAS,OP}$	operational amplifier bias current required for Type D (see figure 3.10) or B (see figure 3.12) current sense amplifier circuit [6]
I_{BL}	bitline current
$I_{BL'0'}$	bitline current for the '0'-cell state access
$I_{BL'1'}$	bitline current for the '1'-cell state access
$I_{BL,charge}$	mean current required to charge the bitline during a read cycle
$I_{CELL'0'}$	cell current in read mode corresponding to low resistive state
$I_{CELL'1'}$	cell current in read mode corresponding to high resistive state
I_{CELL}	non-volatile memory cell current in read mode
I_{CF}	slope dependent current
i_{CF}	small-signal slope dependent current
i_{CF0}	initial condition for i_{CF} in slope dependent current source in figure 4.14
I_{CONST}	constant bias current for slope dependent current source
DAC	digital to analog converter
I_{DS}	drain to source current of a non-volatile memory cell
I_{GBL}	current on the global bitline

I_{INT}	reference current used for discharge timing generation for local ground referenced read circuits in figure 5.4
I_{LVLS}	level-shifter bias current in the slope dependent current
I_{PRE}	precharge current of a bitline or a global bitline due to bitline capacitance charging
I_{REF}	reference current required for a read operation of a memory cell
$Ivdd$	mean current consumption of a sensing circuit during the read access cycle
$Ivdd, static_{I,A}$	calculated static current consumption for the current sensing approach shown in figure 3.10 comprising Type A sense amplifier circuit 3.13, 3.14
$Ivdd_{dynCGSDSoSiSA}$	dynamic current consumption of the source side common gate amplifier enhanced by slope detection circuit
$Ivdd, pre_{I,A}$	calculated dynamic current consumption (during the precharge phase) for the current sensing approach shown in figure 3.10 comprising Type A sense amplifier circuit 3.13, 3.14
$Ivdd_{staticCGSDSoSiSA}$	current consumption of the source side common gate amplifier enhanced by slope detection circuit during the sense phase
$Ivdd_V$	calculated worst case current consumption for the voltage sensing approach shown in figure 3.8
λ_n	small-signal channel-length modulation coefficient of a NMOS transistor
λ_p	small-signal channel-length modulation coefficient of a PMOS transistor
LBL	local bitline of a memory array utilizing the hierarchical bitline architecture
$ls, \tau_{CGDraSiSA}$	large-signal time delay of the common gate drain-side sense amplifier in figure 4.7
$ls, \tau_{CGSoSiSA}$	large-signal time delay of the common gate source-side sense amplifier in figure 4.9
$ss, \tau_{CGSoSiSA}$	small-signal time delay of the common gate source-side sense amplifier in figure 4.9
ls, τ_{COMP}	large-signal time delay of the comparator circuit in figure 4.4
MCU	microcontroller unit
MTJ	magnetic tunnel junction
MV	mdeium voltage domain
$PSRR$	power supply rejection ratio
$Pvdd$	mean power consumption of a sense amplifier circuit during the read access cycle
$Pvdd_{CGSDSoSiSA}$	power consumption of the source side common gate amplifier enhanced by slope detection circuit
$Pvdd_{COMPSoSiSA}$	power consumption of the source side comparator based sense amplifier circuit
$Pvdd_{CGSoSiSA}$	power consumption of the source side common gate amplifier

$Pvdd_{I,A}$	calculated worst case power consumption for the current sensing approach shown in figure 3.10 comprising Type A sense amplifier circuit 3.13, 3.14
$Pvdd_{I,Alv}$	calculated worst case power consumption for the current sensing approach shown in figure 3.10 comprising Type A sense amplifier circuit for very low voltage operation 3.14
$Pvdd_{I,DB}$	calculated worst case power consumption for the current sensing approach shown in figure 3.10 comprising Type D (see figure 3.10) or B (see figure 3.12) current sense amplifier circuit [6]
$Pvdd_{limitTiDo}$	lower limit for a power consumption of a time domain sensing circuit
$Pvdd_V$	calculated worst case power consumption for the voltage sensing approach shown in figure 3.8
Q_{FG}	charge on the floating gate of a non-volatile memory cell transistor
R_{BL}	bitline resistance
R_{BLd}	resistance per unit length of a bitline
R_{BLpath}	sum resistance of the bitline path
R_{CELL}	equivalent resistance for modeling a linear region biased memory cell
$R_{MTJ,MAX}$	maximum MTJ resistance of the embedded STT-MRAM cell
$R_{MTJ,MIN}$	minimum MTJ resistance of the embedded STT-MRAM cell
r_{ds}	small-signal output resistance of a transistor
$REFSA$	reference sense amplifier circuit in figure 5.4 used for global time generation
Rel_{Iwin}	relative cell current window
R_{MUX}	multiplexer resistance
r_{out}	small-signal output resistance of an output node in a sense amplifier or a comparator circuit
$RRAM$	resistive random access memory
R_{SA}	sense amplifier input resistance
S	voltage slope of the input signal for slope detection circuit or a slope dependent current source
SA	sense amplifier
SG	select gate terminal of the HS3P-cell
SL	source line terminal of the embedded STT-MRAM cell
SoC	system on a chip
$ss, \tau_{CGDraSiSA}$	small-signal time constant of the common gate drain-side sense amplifier in figure 4.7
ss, τ_{COMP}	small-signal time constant of the comparator circuit in figure 4.4
STT	spin-torque-transfer
T_{acc}	access time of a sensing scheme

$T_{accCGSDSoSiSA}$	access time of the source side common gate amplifier enhanced by slope detection circuit
$T_{accCGSoSiSA}$	access time of the source side common gate amplifier
$T_{accCOMPSoSiSA}$	access time of the source side comparator based sense amplifier circuit
$T_{accI,A}$	calculated access time for the current sensing approach shown in figure 3.10 comprising Type A current sense amplifier circuit [6]
$T_{accI,DB}$	calculated worst case read access time for the current sensing approach shown in figure 3.10 comprising Type D (see figure 3.10) or B (see figure 3.12) current sense amplifier circuit [6]
$T_{acclimitTiDo}$	lower physical limit for a read access time of a time domain sensing scheme
T_{accV}	calculated access time for the voltage sensing approach shown in figure 3.8
τ_{BL}	time constant of a bitline
τ_{LVLS}	small-signal time constant of the level-shifter in the slope detection circuit in figure 4.19
$\tau, max_{CGDraSiSA}$	worst case delay of the common gate drain-side sense amplifier in figure 4.7
$\tau, max_{CGSoSiSA}$	worst case delay of a common gate source-side sense amplifier
τ, max_{COMP}	maximum time delay of the comparator circuit in figure 4.4
τ, min_{COMP}	minimum time delay of the comparator circuit in figure 4.4
TDC	time to digital converter
T_{limit}	lower physical limit for a sense delay of a current integrating sensing scheme
TMR	Tunnel-Magneto-Resistance value of the MTJ element in the STT-MRAM cell
T_{pre}	precharge time of a sensing scheme
$T_{preCGSDSoSiSA}$	precharge time for the bitline-capacitance-cancellation sensing circuit in source-side sensing configuration
$T_{preCGSoSiSA}$	precharge time for a common gate sensing circuit in source-side sensing configuration
$T_{preCOMPSoSiSA}$	precharge time for a comparator based sensing circuit in source-side sensing configuration
T_{preI}	precharge time for current sensing scheme predicted by model 3.16
$T_{preTiDo}$	precharge time for time a domain sensing scheme
T_{preV}	precharge time for voltage sensing scheme predicted by model 3.7
T_{REF}	reference time for time domain sensing concept (see figure 4.1
$T_{senseCGDraSiSA}$	sense delay for of the common gate drain-side sense amplifier in figure 4.7
$T_{senseCGSDSoSiSA}$	sense delay for the bitline-capacitance-cancellation sensing circuit in source-side sensing configuration

$T_{senseCGsoSiSA}$	sense delay for of the common gate source-side sense amplifier in figure 4.9
$T_{senseCOMP,LIN}$	calculated sense delay for the time domain sensing approach based on comparator sense amplifier shown in figure 4.4 reading a Flash cell biased in linear region
$T_{senseCOMP,SAT}$	calculated sense delay for the time domain sensing approach based on comparator sense amplifier shown in figure 4.4 reading a Flash cell biased in saturation region
$T_{senseI,A}$	calculated sense delay for the current sensing approach shown in figure 3.10 comprising Type A current sense amplifier circuit [6]
$T_{senseI,DB}$	calculated sense delay for the current sensing approach shown in figure 3.10 comprising Type D or B current sense amplifier circuit [6]
T_{senseV}	calculated sense delay time for the voltage sensing approach shown in figure 3.8
$T_{senseV,SAT}$	calculated sense delay for the voltage sensing approach shown in figure 3.8 reading a Flash cell biased in saturation region
T_{sense}	sense delay of a sensing scheme
T_{win}	time window defining the robustness of a time domain sensing scheme
V_B	bulk voltage of a floating gate based non-volatile memory cell transistor
V_{BL}	bitline voltage
V_{BLmax}	maximum bitline voltage of a sense amplifier
$V_{BL,pre}$	bitline voltage after the precharge phase for voltage sensing scheme predicted by model 3.8
V_{BLREF}	reference bitline voltage
V_{W0}	bitline to source line threshold voltage of the embedded STT-MRAM cell for switching from high to low resistive cell state
V_{W1}	bitline to source line threshold voltage of the embedded STT-MRAM cell for switching from low to high resistive cell state
V_{BOOST}	high voltage domain usually generated by an on-chip charge pump
V_{CELL}	voltage across the memory cell in simplified bitline path of figure 3.16 and 3.17
v_{CF0}	initial condition for v_{CF} in slope dependent current source in figure 4.14
V_{CLAMP}	clamp voltage to regulate the bitline potential used in sense amplifier circuit shown in figure 3.13, 3.15 or to set a bitline voltage threshold for circuit in figure 4.5
V_D	drain voltage of a floating gate based non-volatile memory cell transistor
V_{DD}	digital supply voltage domain
VDD_{MIN}	minimal supply voltage required for proper circuit operation

V_{DDVIO}	I/O supply voltage domain
V_{DScell}	drain-source voltage of a memory cell
V_G	gate voltage of a floating gate based non-volatile memory cell transistor
V_{GBL}	global bitline voltage
V_{GScell}	gate-source voltage of a memory cell
V_{IN}	input voltage for a slope detection circuit or a slope dependent current source
v_{in}	small-signal input voltage for a slope detection circuit or a slope dependent current source
V_{out}	output voltage of a voltage sense amplifier, time domain sense amplifier or a comparator circuit
V_{outmax}	desired maximum voltage level output of a comparator, common-gate circuit or a charge transfer amplifier
SR	slew rate of a comparator or common-gate circuit
V_{outmin}	minimum voltage level output of a comparator, common-gate circuit or a charge transfer amplifier
$VPROG$	write voltage domain for a non-volatile memory macro utilizing the HS3P memory cell
V_{OV}	overdrive voltage of a MOS transistor biased in saturation region
V_{PRE}	bitline precharge voltage potential for time domain sensing concept (see figure 4.1
V_R	bitline to source line voltage of the embedded STT-MRAM cell for read operation
V_{read}	voltage potential applied to a gate of a non-volatile memory cell during read operation
V_{RSA}	voltage across the sense amplifier input resistance in simplified bitline path of figure 3.17
V_S	source voltage of a floating gate based non-volatile memory cell transistor
V_{SA}	current sense amplifier output voltage applied to a bitline or a global bitline during the read operation
V_{sense}	minimum input voltage difference or threshold voltage required for robust sensing operation for a voltage sense amplifier or time domain sense amplifier respectively
$V_{sense,LIN}$	calculated input voltage difference for a voltage sense amplifier reading a Flash cell biased in linear region
$V_{sense,SAT}$	calculated input voltage difference for a voltage sense amplifier reading a Flash cell biased in saturation region
V_{SL}	source line terminal voltage of the embedded STT-MRAM cell
V_{SS}	on-chip ground potential
V_{TH}	threshold voltage of a MOS transistor
V_{THcell}	threshold voltage of a memory cell

$V_{THshift}$	threshold voltage shift of a floating gate based non-volatile memory cell transistor due to electron injection into the floating gate
WL	wordline
W_{MAR}	read window margin of a time domain sensing scheme

Chapter 1

Introduction

In contrast to standalone NAND-Flash memory which is used for mass storage applications featuring several GBytes of memory capacity and read access times of several tens of microseconds [7], the embedded NOR-Flash memory usually offers only several MBytes of memory but with tight access times of only several tens of nanoseconds[8],[9]. Fast read access in embedded Flash is mainly driven by microcontroller applications where code memory with direct code execution and code fetch from the embedded Flash is required in real time applications. Important market segments for microcontrollers are: consumer, industrial and automotive. Most of today's microcontrollers independent of their application comprise an embedded NOR-Flash memory to fulfill the requirements for complex algorithms and flexibility in the development process.

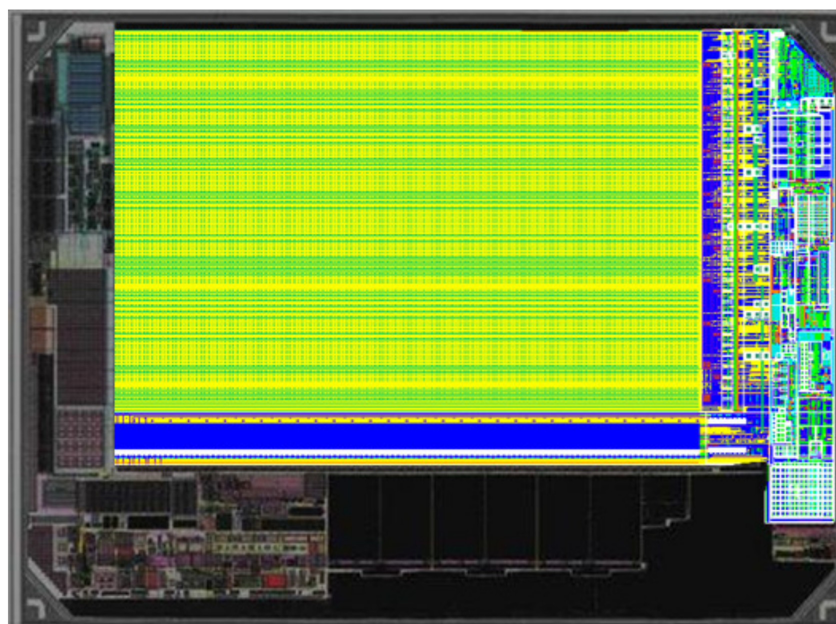

Figure 1.1: Chip Micrograph of Infineon SLC14MCO480 SIM-Card Microcontroller with Highlighted Layout of the Non-Volatile Memory Macro

The most widely used embedded Flash memory is produced for mobile applications. Figure 1.1 shows the die photo of Infineons SLC14MCO480 microcontroller with 480kByte embedded NOR-Flash memory (highlighted layout) fabricated in 65nm technology. The microcontroller features an ARM $Cortex^{TM}$-M0 CPU core operating at 32MHz clock frequency and 12kByte RAM. This microcontroller targets the SIM card market requiring low cost and low power products. The Flash memory has free partitioning between code and data offering 100k cycles endurance. The read operation is done on 32bit in parallel at a 50ns read access time. The ambient operating temperature for the microcontroller ranges from -25°C to +85°C.

Infineons XMC 1000 microcontroller fabricated in 65nm as shown in figure 1.2 fits the requirements for low-end industrial applications like LED lighting, motor control or digital power conversion. It compromises 200kByte embedded Flash memory, which is

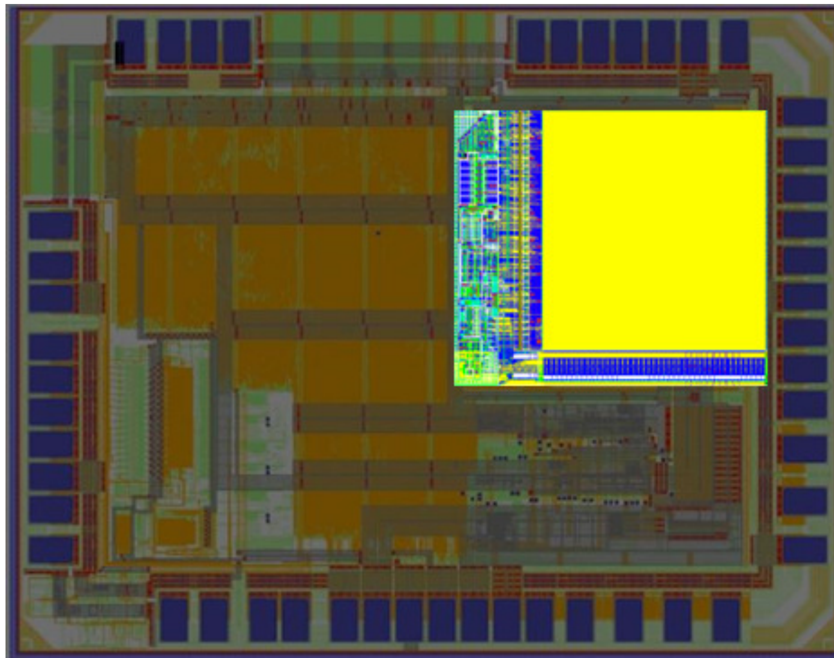

Figure 1.2: Chip Micrograph of Infineon XMC 1000 Industrial Microcontroller with Highlighted Layout of the Non-Volatile Memory Macro

even smaller compared to the SIM-card microcontroller, but the system on a chip (SoC) itself includes a large set of peripherals like multiple ADCs. And the ambient temperature range is increased from -40°C up to 105°C. The access time, the read throughput and the CPU core are the same as for the SIM-card microcontroller to support the reuse of designs for lower development cost and faster time to market.

The most critical and tight requirements for the embedded Flash memory modules with respect to reliability and performance are derived from automotive applications like powertrain, chassis control and safety. Automotive applications especially for the powertrain segment require robust high performance microcontrollers including on-chip non-volatile memory. Due to low read access times the NOR-Flash memory is preferably used in today's SoCs for automotive applications [10]. Because of its high temperature

reliability, up to 20 years data retention and sub 30ns random read access time, every modern microcontroller for automotive applications in the powertrain segment comprises an on-chip NOR-Flash memory for code and data storage. The real time performance of a microcontroller heavily depends on the read access time of the memory module. Therefore the design of the non-volatile memory macro and especially of the read path is one of the main design differentiators in today's automotive microcontrollers.

Figure 1.3 shows the layout of Infineons TC27x automotive microcontroller fabricated in 65nm with highlighted embedded Flash module. The microcontroller includes 4MByte

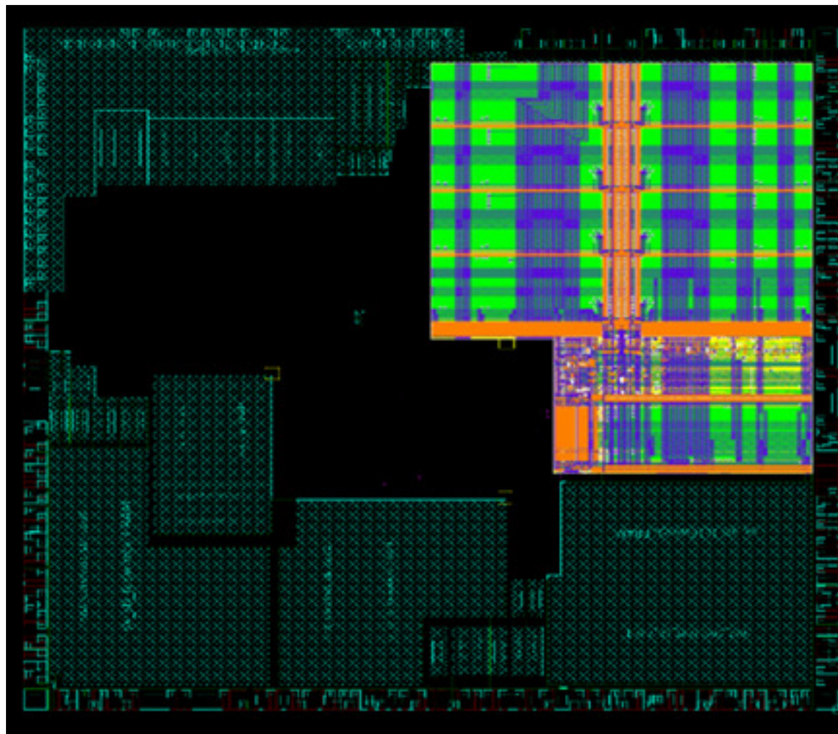

Figure 1.3: Toplevel-Layout of Infineon $AURIX^{TM}$ TC27x Automotive Microcontroller with Highlighted Layout of the Non-Volatile Memory Macro

program and 64kByte data embedded Flash memory. The program memory is used for code execution and data memory for data storage. The data is processed by a triple core CPU featuring the $Tricore^{TM}$ architecture with 200MHz core frequency. In addition the microcontroller comprises 472kByte RAM and a large number of ADCs, DACs and interface blocks.

The data memory is optimized for write operation and retention, allowing 500k cycles but it has low read speed with 100ns read access time reading 64bit in parallel. The read access time for the program memory is only 30ns and 512bit of data can be read out in parallel achieving a read throughput of 2.13GByte/s. Due to this high read throughput the power efficiency is very important to fulfill the low power requirements. In addition the read circuit has to provide high power supply rejection ratio since there are many blocks connected to the same power supply generating severe supply noise. Even more the automotive applications require the widest operating temperature range from -40°C

up to 125°C (ambient). The die temperature itself may be even increased up to 170°C.

The following work focuses on read scheme and read circuit design for automotive applications. The power efficiency, circuit area and robustness is not only very important for the automotive segment but also for industrial and consumer applications. This means that the design approaches developed in this work can be also applied for all other applications. Therefore the different design approaches will be evaluated with respect to their power consumption, speed and robustness to propose the best solution in terms of power efficiency and speed for the given requirements.

Chapter 2

Embedded NOR-Flash

2.1 Embedded Flash (eFlash) Memory Cell

There is a variety of eFlash cell concepts used in todays microcontrollers. However the majority use the basic idea of a floating gate. The basic structure of a floating gate cell is shown in figure 2.1. It consists of an NMOS transistor with two gates: the control gate

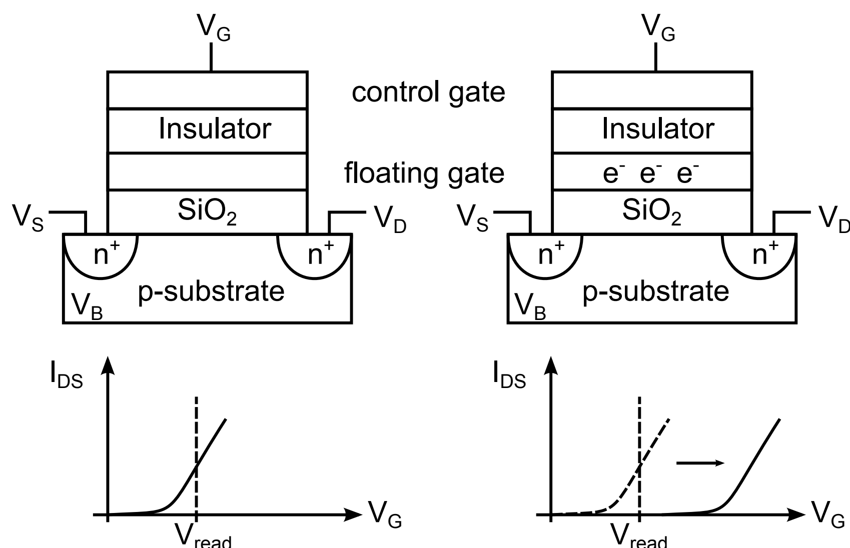

Figure 2.1: Basic Flash Cell Working Principle

and the floating gate. The floating gate has no direct connection to any terminal of the cell and is capacitively coupled to the control gate. It is insulated by the silicon oxide to the p-substrate forming a conventional NMOS transistor. To prevent any charge transfer between the two gates the second insulator layer is inserted.

By applying high voltages at the control gate ($V_G > 10V$) electrons are injected by Fowler-Nordheim tunneling ($V_B = V_S = V_D < 0V$) through the silicon oxide or by channel hot electrons ($V_B = V_S = 0V$; $V_D > 3.2V$) into the floating gate during the write operation. By this means the transfer curve of the Flash transistor is shifted to higher

V_G voltage as illustrated in figure 2.1 corresponding to a threshold voltage shift in the NMOS transistor. The threshold voltage shift $V_{THshift}$ can be calculated as:

$$V_{THshift} = \frac{Q_{FG}}{C_{GES}} \tag{2.1}$$

where Q_{FG} is the floating gate charge and C_{GES} is the sum of coupling capacitances between the floating gate and all other nodes of the Flash transistor. During the erase operation ($V_G < -10V$; V_S or V_D or $V_B > 6V$) electrons are tunneling from the floating gate to source or to the p-substrate. Hot hole injection ($V_G < -10V$; $V_B = V_S = 0V$; $V_D > 3.2V$) is also possible to erase the Flash cell but is only used in charge trapping cell types [11]. The specific write and erase method is depending on cell architecture: one-transistor-cell, two-transistor-cell or split-gate cell. The references [10] and [11] give a detailed overview on today used eFlash cell types and their characteristics. Since this work focuses on design related items the one transistor memory cell symbol is used in following without loss of generality to indicate eFlash memory cell.

To read out the state of a Flash cell the following read bias conditions are applied: $V_D \approx 1V$, $V_S = 0V$, $V_G = V_{read} \approx 4V$. During the read operation the eFlash cell is usually biased in between the linear and the saturation region however some split gate cell concepts allow to bias the memory cell in saturation region only with $V_{read} < 1.5V$ [17],[12],[8]. To cover all possible cell structures the two extreme cases will be considered in the following chapters: the linear region and the saturation region biased cell. The Flash transistor with shifted threshold voltage is non conductive (usually $I_{DS} < 1\mu A$) corresponding to the one state cell (high resistive). The zero state (low resistive) cell with no threshold voltage shift (or slightly positive threshold voltage shift) in the Flash transistor has usually drain-source current of $I_{DS} \approx 20\mu A$. The threshold shift is adjustable depending on the voltage and time in the program operation by this means multiple threshold voltage levels can be defined to store more than two state in one cell (multilevel cell) which is extensively used in NAND Flash architectures. Because of the reliability and retention issues, today's automotive eFlash memories are using single level cells (two states in one cell) [10],[11]. Since the focus of this work is the read (sensing) operation and the circuit concepts, the following chapters address the circuit blocks and concepts determining the read access time. The reference [7] gives a detailed overview of the Flash technology and related circuit concepts.

2.2 NOR Array Architecture

The NOR-Flash is a specific array structure for Flash transistors as shown in figure 2.2. Flash cells are located at the wordline-bitline intersections in the array. All cells are connected with their source terminal to the common plate line which is usually connected to V_{SS} during the read operation. The drain terminals are connected to a corresponding bitline in the array. The read word width of typically 128bit or 256bit (2^d-bits in figure 2.2) determines the number of bitlines/cells accessed simultaneously and therefore the number of read (sense) amplifiers required in the design. The typical macro size in current

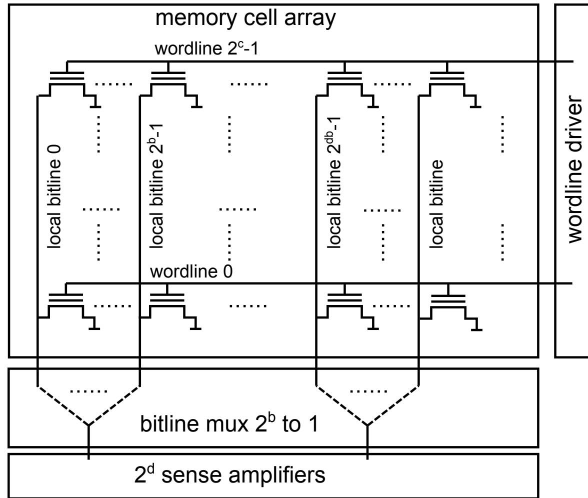

Figure 2.2: Basic NOR-Flash Architecture

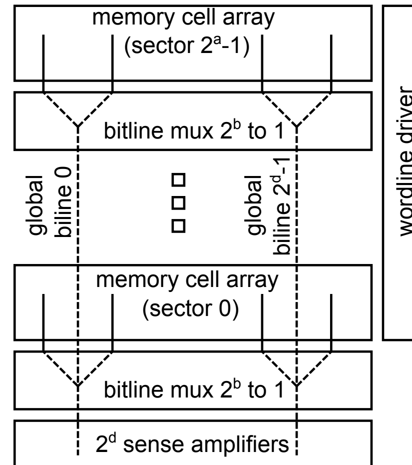

Figure 2.3: Hierarchical Bitline Architecture

eFlash designs varies from 512kByte to several megabyte. For example a 2MByte NOR-Flash macro with 4096 wordlines (corresponds to 2^c in figure 2.2) contains 4096 bitlines (corresponds to 2^{db} in figure 2.2). For a 256bit read word width the multiplexer factor of $4096/256 = 16$ ($2^{db}/2^d = 2^b$) has to be considered. In case the specific split gate cell concept allowing low voltage read path is not used [17],[8],[12] the high voltage on the drain/source terminals for erase operation of the Flash cell requires thick oxide devices in the bitline path. Since the devices are usually utilized in the bitline multiplexer its resistance is several orders of magnitude higher compared to other types of memory like DRAM or SRAM. The high multiplexer resistance sets a lower limit for a read access time as will be shown in the next chapter.

One of the first order parameters for the access time (as will be explained in the next section) is the bitline capacitance C_{BL} seen by the sense amplifier during the sensing phase. In sub 65nm technologies C_{BL} is mainly dominated by metal to metal capacitance between neighboring bitlines. To minimize the bitline length and hence the bitline capacitance the number of cells connected to one wordline (wordline width) has to be as high as possible. Usually the wordline width is equal or below the bitline width (number of cells connected to one bitline) to ensure the propagation delay of the wordline signal is not a limiting factor for the read access time. To further reduce the bitline capacitance without increasing the wordline width the hierarchical concept comprising the global bitline is used as shown in figure 2.3. The memory cell array is divided in 2^a sectors with smaller bitline capacitance. Every sector includes its own bitline multiplexer connecting the selected local bitline to a corresponding global bitline. Every global bitline is directly connected to a sense amplifier. The capacitance of the global bitline is much lower compared to a local one since there are only 2^d global bitlines and therefore the distance from a global bitline to the next one (bitline pitch) is much longer compared to the local bitlines. The drawback of this approach is the additional bitline multiplexers consuming additional chip area. There is a sweet spot between the number of sectors and bitline capacitance which is technology and design dependent, since every additional sector switch

introduces parasitic capacitance to the global bitline which must be low compared to the local bitline capacitance. Despite the usage of the global bitline concept the bitline capacitance in eFlash macros is still higher compared other types of embedded memory like eDRAM or SRAM.

2.3 Summary

- The basic eFlash cell structure consists of a NMOS transistor with two gates: the control gate and the floating gate to store the cell state information.
- The memory cell operation region (linear or saturation) depends on cell concept and technology.
- Hierarchical bitline architecture is used to reduce bitline capacitance.
- The bitline path contains thick oxide devices to block the high voltages during erase operation if the specific split gate cell concept allowing low voltage read path [17],[8],[12] is not used.

Chapter 3

Voltage vs. Current Sensing Scheme

In this section the state-of-the-art voltage and current sensing schemes for eFlash are discussed from design point of view.

3.1 Concepts

Figure 3.1 illustrates the voltage sensing concept used for embedded non-volatile NOR-type memories. The READ_START signal starts the read out operation activating the

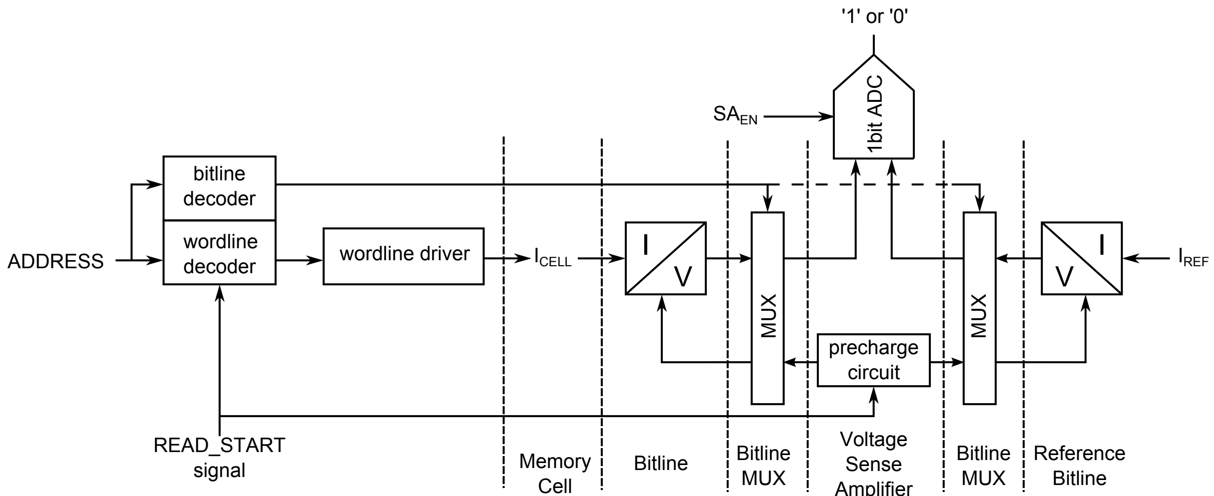

Figure 3.1: Voltage Sensing Concept

wordline and the bitline decoder respectively. The read operation consists at least of two phases: precharge and sense. During the precharge the wordline driver applies the gate voltage to the selected cell delivering the cell current I_{CELL}. Simultaneously the bitline multiplexer connects the selected bitline and the so called reference bitline to the sense amplifier circuit. The reference bitline is typically matched in terms of capacitance and resistance to the common bitline in the cell array to provide fully symmetrical conditions. This symmetry also provides very good power supply rejection ratio ($PSRR$). The sense

amplifier itself has to provide read biasing conditions to the selected bitline during the precharge phase by a dedicated precharge circuit. The precharge phase is finished when the bitline and wordline are charged to the desired values, usually reaching more than 90% of the end value is sufficient to end the precharge phase. In the subsequent sensing phase the bitline and the reference bitline converts the cell current I_{CELL} and the reference current I_{REF} to a voltage. This current to voltage conversion is done by integrating the corresponding current on the bitline capacitance. After a sufficient voltage difference between the two bitlines has been developed the sense amplifier is activated by the SA_{EN} signal. The sense amplifier performs an one-bit (for a single level cell) analog to digital conversion. For voltage sensing the analog value is a bitline voltage difference between the selected memory bitline and the reference bitline. The main difference between a sense amplifier circuit and an analog to digital converter (ADC) is the interaction/dependence of the sensing circuitry on the memory cell and array structure, since the sense amplifier has to provide biasing conditions for the bitline in the precharge phase and converts the cell current to a digital output in the sensing phase. The sense amplifier which is optimized to the specific memory structure and cell physics will achieve the best power efficiency for the required read performance.

The basic current sensing concept is shown in figure 3.2. Compared to voltage sensing

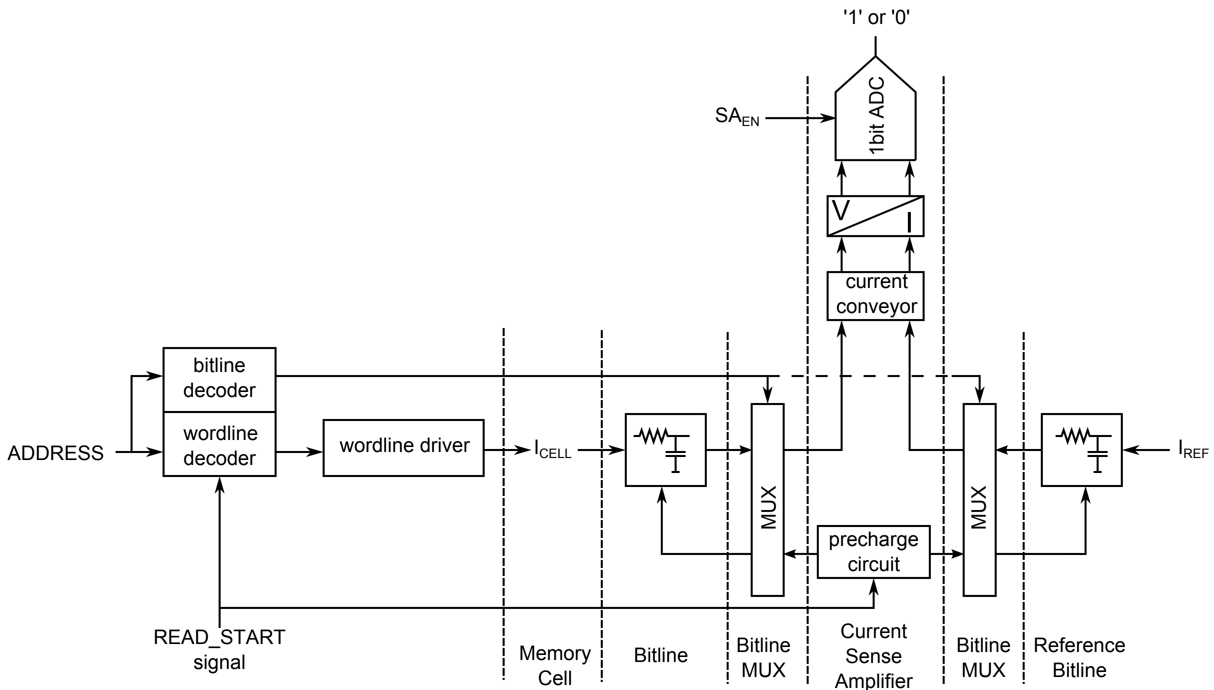

Figure 3.2: Current Sensing Concept

(figure 3.1) the cell current I_{CELL} is not converted to a voltage by the bitline capacitance. The cell current is continuously flowing on the bitline acting as a passive low pass filter and flows directly through the current sense amplifier. To support this static current flow the sense amplifier utilizes the current conveyor circuit with low input impedance providing static bias conditions for the bitline voltage. The current conveyer generates an output current which is equal to its input current. The output current of the conveyor

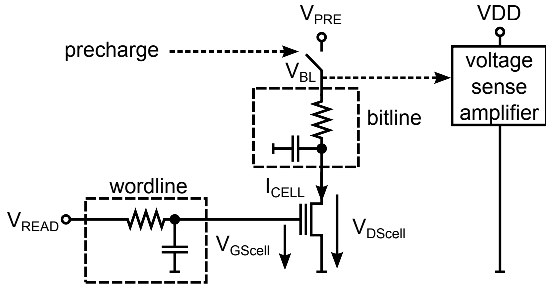

Figure 3.3: Drain-Side Voltage Sensing

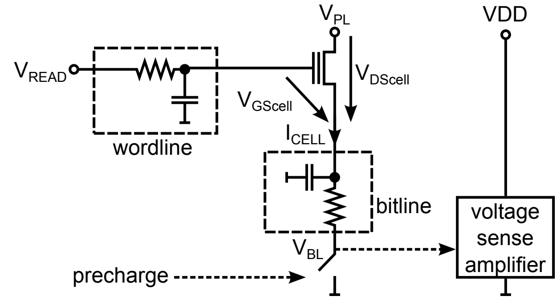

Figure 3.4: Source-Side Voltage Sensing

is converted to voltage by a subsequent current to voltage converter. The converted voltage difference is amplified to a full CMOS level by the 1bit *ADC*. Compared to the voltage sense amplifier setup (as shown in figure 3.1) the current sense amplifier comprises two additional building blocks: the current conveyor and the current converter. By this means every current sense amplifier comprises a voltage sense amplifier [6]. Since there is static current flow from the sense amplifier towards the memory cell, the current sensing approach usually results in a higher current consumption compared to the voltage sensing approach. The drawback of higher complexity and power consumption in the current sense amplifier circuit compared to voltage sensing is compensated by a significant read performance increase for low cell currents (will be shown in section 3.4) since the bitline voltage is forced to a fixed potential during the read operation, which is independent on the cell current.

3.2 Sensing on Drain- or Source-Side

As already mentioned for the voltage sensing setup the cell current to voltage conversion is performed by integrating the cell current on the bitline capacitance. The bitline voltage change is sensed and amplified to a full CMOS level. This sensing concept is suitable for very low supply voltages since the voltage sense amplifier is not in the cell current path. The memory cell and the sense amplifier are connected in parallel hence the maximum voltage headroom (which is the supply voltage) is available for the drain to source voltage of the cell and for the sense amplifier circuit respectively. There are two possible sense sides for a memory cell: the electrical source and drain of the memory transistor. For the drain-side sensing setup (fig. 3.3) the maximum level for the bitline precharge voltage V_{PRE} is V_{DD}. During the sensing phase the bitline voltage V_{BL} is discharged by the cell current I_{CELL} towards V_{SS}. If the memory transistor cell is biased in saturation region ($V_{DScell} \geq V_{GScell} - V_{THcell}$) the cell current I_{CELL} is nearly independent of the bitline voltage as long the saturation condition is valid. Therefore the bitline voltage during the sensing phase will have linear discharge slope with respect to the sensing time. In case the cell is biases in linear region ($V_{DScell} < V_{GScell} - V_{THcell}$) the bitline discharge slope is not constant (RC like discharge) due to the cell current reduction during the sensing phase, slowing down the discharge speed and therefore the sensing performance. The parasitic bitline resistance will reduce the maximum drain to source voltage of the cell V_{DScell}, and

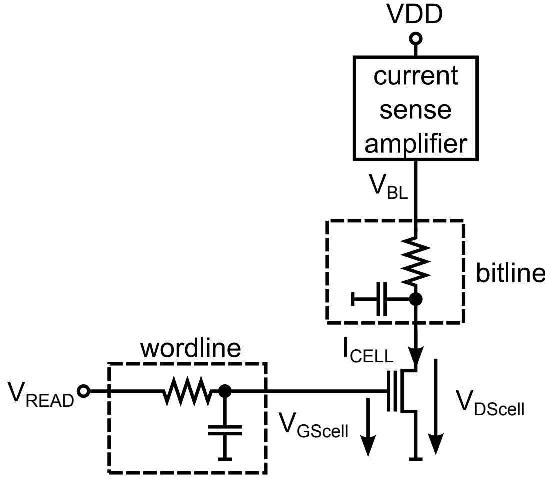

Figure 3.5: Drain-Side Current Sensing

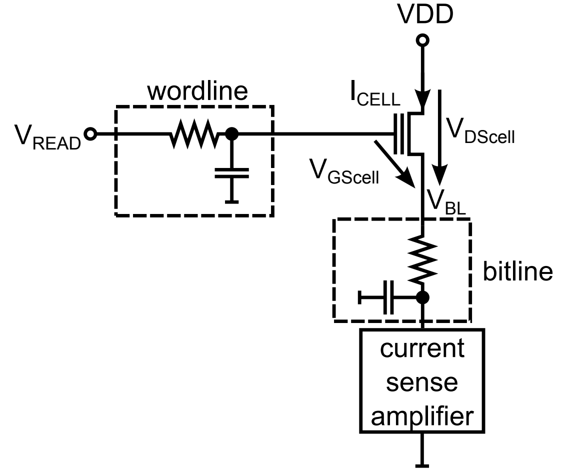

Figure 3.6: Source-Side Current Sensing

therefore the linear region biased cell is more sensitive to high bitline resistance compared to a cell which is biased in saturation region. In case the sense amplifier is placed at the electrical source of the memory transistor (fig. 3.4) the bitline voltage increase during the sensing phase will decrease the cell current independent of the operation region of the memory transistor due to V_{GScell} voltage reduction. In addition the bitline voltage increase leads to higher threshold voltage of the memory transistor due to the body effect.

Due to constant threshold voltage of the memory transistor during the sensing phase and low dependence on the bitline voltage in case of the memory cell biased in saturation, the drain-side sensing is the concept of choice in embedded NOR-Flash memories comprising voltage sense amplifiers [12], [8]. The maximum bitline voltage level is defined by the memory transistor reliability, since the high V_{DScell} may cause hot electron injection damaging the gate oxide between the floating gate and the channel. This can lead to significantly reduced reliability. In addition the injected electrons will decrease the threshold voltage of the memory transistor and therefore also the cell current.

For the current sensing operation the sensing circuit is placed in the cell current path. The sense amplifier and the memory cell are series connected sharing the available supply voltage headroom. In contrast to voltage sensing the sense amplifier regulates the bitline voltage to a fixed potential during the whole sensing phase which is more suitable for a linear region biased cell. Similar to voltage sensing the current sensing operation can be performed at the drain and the source of the memory transistor as illustrated in figure 3.5 and 3.6. The drain-side current sensing comprising bitline voltage regulation is the most frequently used approach since as mentioned before the transistors overdrive voltage ($V_{GScell} - V_{THcell}$) is not dependent on the bitline resistance.

The source-side current sensing leads to higher threshold voltage of the memory transistor since the bitline voltage V_{BL} is higher than the bulk potential of the memory transistor (body effect). However the threshold voltage increase may be even required for some applications to suppress the leakage current of deselected cells on the bitline at high temperatures.

The voltage headroom of the current sense amplifier defines the maximum bitline voltage. To increase the bitline voltage headroom the folded cascode approach is used [13], [6] as illustrated in figure 3.7. The maximum voltage headroom for the cell is $V_{DD} - V_{OV}$,

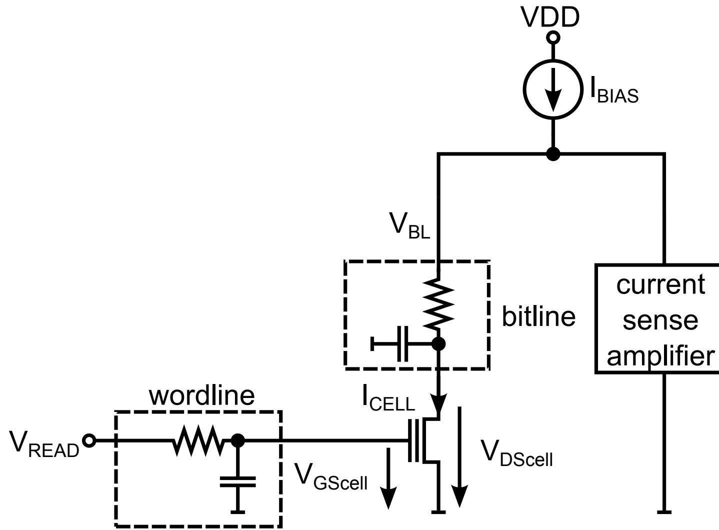

Figure 3.7: Low Voltage Drain-Side Current Sensing

where V_{OV} is the minimum required overdrive voltage (about 200mV) for the I_{BIAS} current source. This approach requires more current compared to a conventional current sensing, since the bias current I_{BIAS} has to supply two current branches.

3.3 Sensing Circuits

3.3.1 Voltage Sensing

Figure 3.8 shows a typical voltage sensing setup used in read path for eFlash [12], [8]. The voltage sense amplifier requires full symmetrical implementation to achieve high power supply rejection ratio ($PSRR$), as it is required in automotive applications for robust and reliable operation. Therefore every sense amplifier requires a reference path comprising a reference bitline BL_{REF} which is fully matched in terms of capacitance and resistance to a memory bitline BL in the cell array. To minimize the area overhead the bitline of an inactive sector is used for reference path [12], [8] as it is done in DRAM type memories. The latch-type sense amplifier itself as shown in figure 3.8 occupies very small area due to its simple structure. The reference side requires the reference current source I_{REF} as it does not have an active cell. The usage of preconditioned NOR-cells (reference cells) [9], [14], [15], [16] on the reference bitline instead of the current source I_{REF} allows further matching between the cell and the reference path in terms of the temperature dependence of I_{CELL}. This approach is usually chosen if the current difference between the low resistive cell (cell state '0') and reference current is very low. The drawback of this approach is the possible read disturb of the reference cells since they are accessed more

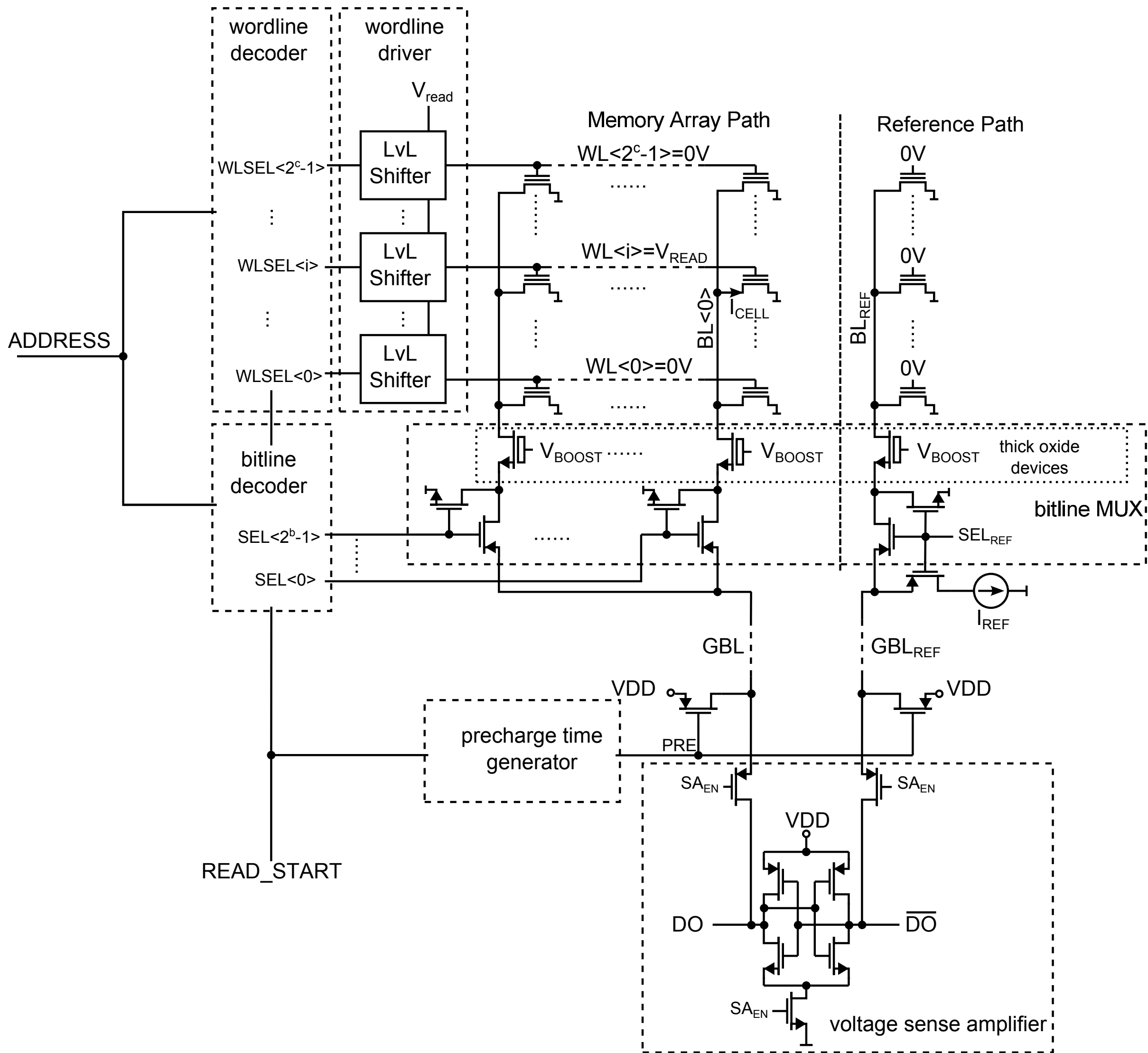

Figure 3.8: eFlash Read Path for Voltage Sensing

frequently compared to the memory cells. In addition the preconditioning and monitoring of the reference cells requires high design overhead and is very time consuming in the test, the fixed reference current I_{REF} is used in most of today's eFlash memories.

As already mentioned in the previous section the bitline multiplexer requires thick oxide devices as shown in figure 3.8 to block the high voltages on the bitlines during erase or write operation in the memory array if the specific split gate cell allowing low voltage read path [17],[8],[12] is not used. The sense amplifier itself is usually using logic domain devices to reduce the area and the power consumption in the sense circuit. In order to decrease the on-resistance of these thick oxide devices, they are constantly biased by the high gate voltage V_{BOOST} generated by an on-chip charge pump during the read operation. Despite this gate boosting the on-resistance of these devices is still in the range of several kOhm, which is almost one magnitude higher compared to other switches in the read path. This is one of the major read access time limiters, as will be shown in the

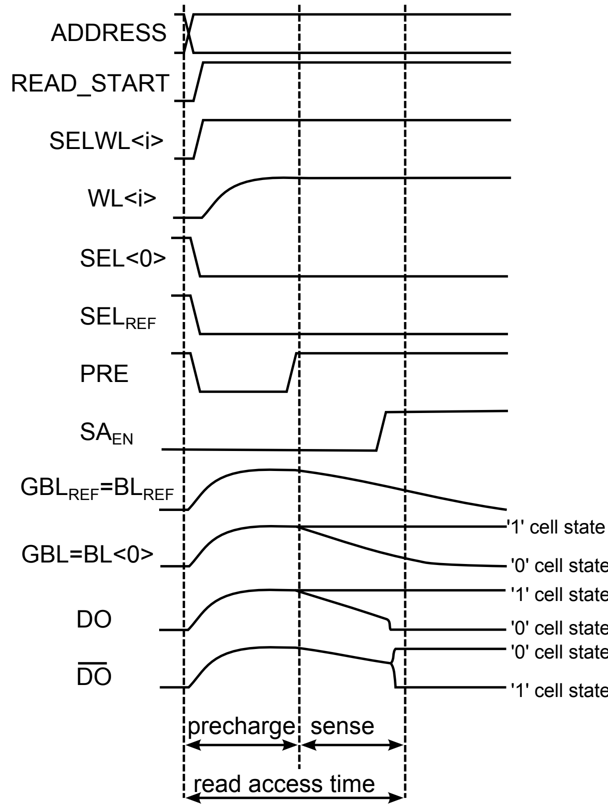

Figure 3.9: Timing Diagram of Figure 3.8

next section. In addition the thick oxide devices consume the main part of the bitline multiplexer area due to strongly increased drain/source pull back and active area spacing to fulfill the high voltage design rules. The logic PMOS transistors in the multiplexer are switching the selected bitline to the sense amplifier input if the corresponding select signal is low whereas the deselected bitlines are discharged by the logic NMOS transistors to ground. For very area critical designs with reduced read access time requirements the logic PMOS and NMOS transistors are skipped and the multiplexing is done in the high voltage domain by thick oxide devices only. This approach causes a significant power consumption in the charge pump circuit providing the V_{BOOST} voltage due to the gate capacitance charging of the thick oxide devices. Besides the increased power consumption the switching speed is very limited since high voltage level shifters are required in the bitline address decoder circuit.

The read operation starts when the ADDRESS signal changes its state and the READ_START signal goes high activating the address decoder and the precharge circuit as shown in figure 3.9. The decoded address selects the corresponding WLSEL signal (in this case $WLSEL$<i>) activating the level shifter (LvL Shifter) which starts charging the selected wordline. Simultaneously the corresponding SEL signal (in this case SEL<0>) goes low switching the selected bitline to the global bitline GBL and SEL_{REF} switches the reference bitline BL_{REF} to the global reference bitline GBL_{REF}. As already mentioned the read out operation consists of two phases precharge and sense as depicted in figure 3.9.

During the precharge phase the wordline is charged to V_{read} voltage and bitlines are set to initial conditions. As the PRE signal goes low the GBL and the GBL_{REF} are charged up to V_{DD} and the reference current source I_{REF} is switched to the reference bitline (see figure 3.9). The precharge is finished as soon as both GBLs reach about 99% of V_{DD}. During the precharge phase the gate voltage (wordline) of the accessed cell has to reach about 90% of V_{read} voltage, since if the on-resistance of the cell changes significantly during the subsequent sense phase it may cause read out fails. This criteria determine the wordline propagation delay and hence the maximum possible wordline length in the specific Flash memory comprising voltage sensing scheme.

During the sense phase the bitline and the reference bitline are discharged by I_{CELL} and I_{REF} respectively as shown in figure 3.9. The reference path discharges with constant voltage slope since the constant current reference is applied to the bitline capacitance which has no or very little voltage dependence. The selected bitline is discharged through the accessed eFlash transistor to ground in case of '0' state cell. If the eFlash transistor is biased in linear region the discharge voltage slope is not constant leading to a slow discharge speed. In addition the cell current changes considerably during the bitline discharge and is not suitable for reference concept using fixed reference current. Hence the voltage sensing approach as shown in figure 3.8 is well suited only for memory cell types which remain in saturation over a wide bitline voltage range. In case the '1' state cell is accessed the bitline voltage ideally remains unchanged.

Automotive applications require full speed functionality at 170°C. Due to high temperature the threshold voltage of the (deselected) '0' state cells may be strongly decrease leading to a leakage current on the bitline slowly discharging the bitline capacitance. The threshold voltage of '0' state cells has to be accurately controlled to prevent read out fails at high temperatures. As soon as the voltage difference between the bitline and reference path is high enough (usually more than 100mV) the sense amplifier is enabled by SA_{EN} signal delivering the digital output DO. By activating the SA_{EN} the GBL and GBL_{REF} with high capacitive load are decoupled from the DO and $\bar{DO}$ nodes during the latch phase of the voltage sense amplifier to ensure high speed swing on the output nodes. The discharging speed of the GBL is defined by its capacitance and the '0' cell state current. The lower bound for the voltage difference between the two bitlines required for robust read operation is defined by the sense amplifier offset caused by transistor mismatch and the $PSRR$ of the latch comparator circuitry.

Several solutions were proposed to reduce the latch comparator offset. In [18] two capacitances together with several switches are inserted into the latch comparator circuit for offset compensation, significantly reducing the input offset. However this approach requires additional timing overhead for the offset compensation procedure during the read routine degrading the read access time. Another approach to reduce the input offset is presented in [8] adding an offset current to the bitline depending on the internal offset of the sense amplifier. This is done only during the initialization phase of the chip and has no impact on read access time but the offset is reduced in the best case only by a factor two.

3.3.2 Current Sensing

Figure 3.10 illustrates the drain-side current sensing scheme for the embedded NOR-Flash [9]. Compared to the voltage sensing setup in figure 3.8 the read circuits and multiplexer

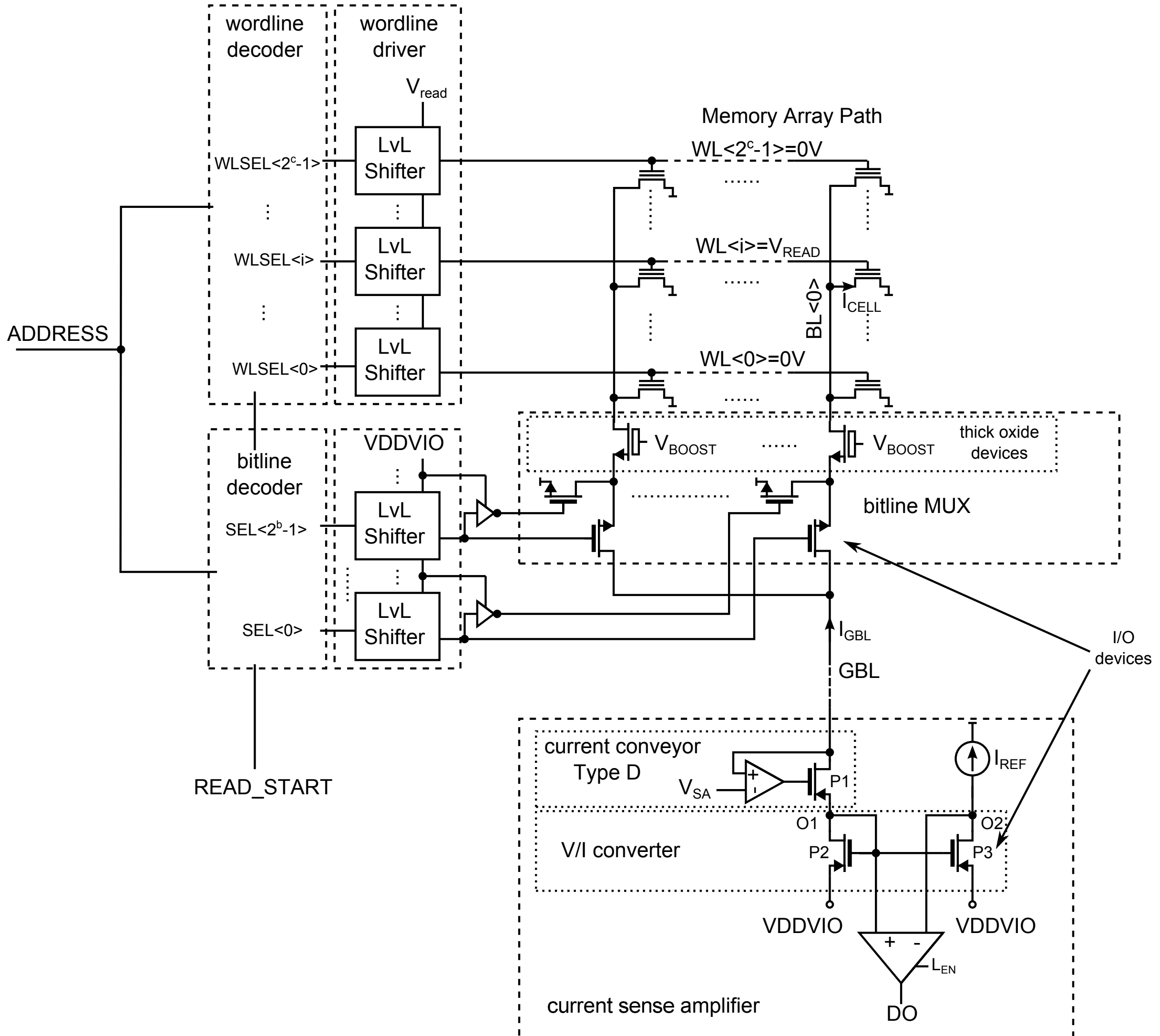

Figure 3.10: eFlash Read Path for Current Sensing

are built in the V_{DDVIO} (I/O) voltage domain (3.3V in this design) to provide sufficient voltage headroom for the bitline voltage adjustment and the sensing circuits. For this reason I/O devices are used in the sense amplifier circuit. They have higher threshold voltage and thicker gate oxide leading to lower transconductance and worse matching characteristics compared to the V_{DD} domain devices. The usage of I/O devices induces higher area consumption of the sensing circuits compared to V_{DD} domain sense amplifier. Compared to low voltage realizations for the current sense amplifier [13] this approach has a higher voltage range for the bitline voltage adjustment and achieves higher sensing speed as will be shown in section 3.4.

To reduce power and area consumption the reference path is usually not symmetrically built for a fully differential signal [9] as in voltage sensing setup. Instead only the reference current I_{REF} is provided to the sense amplifier which is usually set to the half of the low resistive cell state current to provide enough margin for the read operation. Despite this single ended approach the area of the sensing circuit is much higher compared to voltage sense amplifier (see figure 3.8) due to additional opamp and current to a voltage conversion stage. The single ended approach suffers from reduced power supply rejection ratio ($PSRR$). To reduce the $PSRR$ requirements the I/O supply for the sense amplifier is usually provided on an extra pin of the chip strongly reducing the noise on the sense amplifier supply.

The bitline current I_{GBL} is flowing through the current conveyor and is mirrored by P2, P3 current mirror to the O2 node. The current conveyor decouples the high capacitive GBL node from the internal nodes allowing fast swing O1 and O2 during read operation. The reference current I_{REF} is supplied to all sense amplifiers and fed into the O2 node. The O2 node is a high impedance node compared to O1 node. The voltage swing of O2 with respect to O1 node is amplified by a subsequent differential latch comparator. Figure 3.11 shows the timing diagram for the current sensing setup. The address change triggers the READ_START signal activating the wordline and bitline decoder. The selected wordline is driven by the corresponding level shifter (LvL Shifter) to V_{read} similar to voltage sensing. Compared to the voltage sensing setup 3.8 the bitline decoder requires level shifters since the path uses I/O type NMOS transistor to connect the bitline with the current sense amplifier. This I/O level shifter introduces additional area and power overhead and slows down the decoder speed compared to voltage sensing path. As soon as the selected bitline (in this case SEL<0>) is connected to the global bitline a voltage drop occurs on GBL due to charge sharing between the two lines. The current conveyor circuit senses this voltage bump and drives P1 out of its steady state (large signal) operation point by setting its gate voltage to V_{SS} driving the GBL voltage to a target voltage potential V_{SA}. As the bitline and wordline voltage rise, the cell current I_{CELL} continuously increases till the final voltage levels are reached. The bitline charging current I_{PRE} increases as the bitline is charging and it decreases when the bitline voltage approaches the final voltage level. The global bitline current I_{GBL} seen by the current sense amplifier is the sum of cell current and the charging current $I_{GBL} = I_{CELL} + I_{PRE}$. If the low resistive cell ('0' cell state) is accessed the bitline current will settle above the reference current level I_{REF}. The P3 will stay conducting, the node O2 will stay at V_{DDVIO} and O1 node will settle one threshold and one overdrive voltage of P2 below V_{DDVIO}. For this case the digital output DO does not change its state if the output latch comparator is enabled by the L_{EN} signal, hence the '0' cell state is not critical for read access time for current sensing. In case the high resistive cell is accessed ('1' cell state), I_{GBL} will decrease ideally to 0μA. As the I_{GBL} crosses the I_{REF} current level the current sense amplifier enters the steady state (small signal) operation and the high impedance node O2 will be discharged by I_{REF} towards V_{SS}. The O1 node will be charged up too, but it will settle one threshold voltage (of PMOS) below V_{DDVIO}, since the P2 is diode connected. As soon as enough voltage difference between O1 and O2 is reached, the output latch comparator is enabled by L_{EN}. In contrast to voltage sensing, here the read

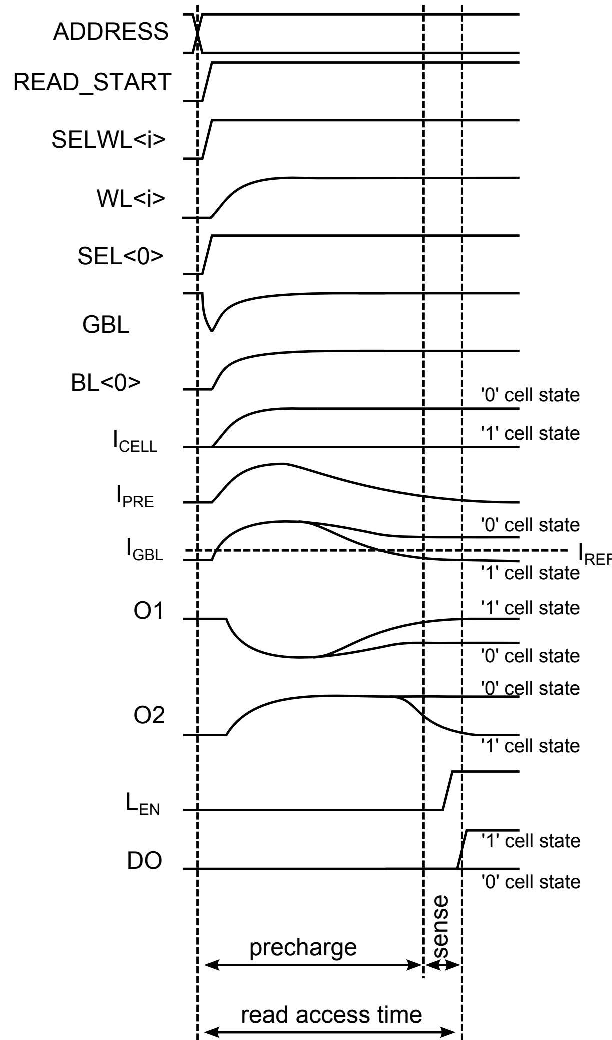

Figure 3.11: Timing Diagram of Figure 3.10

access on high resistive cell ('1' cell state) cell is critical for the read access time.

In this particular circuit the conveyor is built by a regulated transistor P1 with low input impedance and fast settling time. Additional precharge circuitry can be skipped since during the precharge phase the gate of P1 is pulled to ground by the opamp achieving very low input impedance. In [6] B.Wicht gives a detailed overview and classification of current conveyor circuits in Type A, B, C and D which will be used in following. The current sensing circuit in figure 3.10 is classified as a Wilson Type (Type D) which has the fastest current step response among all current conveyor circuits. However replacing P1 by an NMOS transistor (Type B) as shown figure 3.12 is also a good approach if a low voltage operation is not desired. It exhibits the fastest response on large bitline voltage drop (large signal behavior) during the precharge phase due to the source follower configuration. For power critical designs the opamp is replaced by a fixed gate voltage (Type A) as illustrated in figure 3.13 sacrificing some read speed (see section 3.4). The area of Type A is smaller compared to Type B or D since the opamp is no more used, however the Type A circuit requires additional precharge circuitry consuming additional

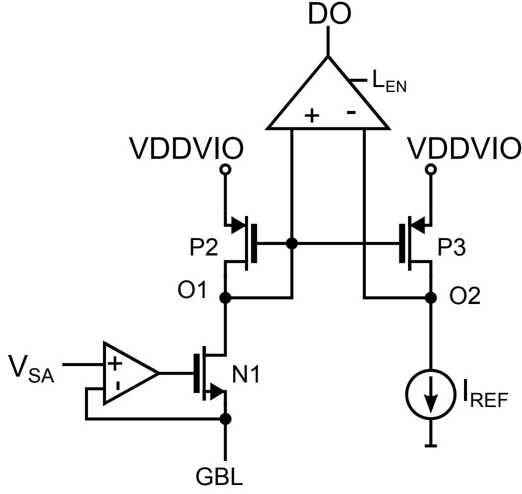

Figure 3.12: Current Sense Amplifier with Type B Current Conveyor

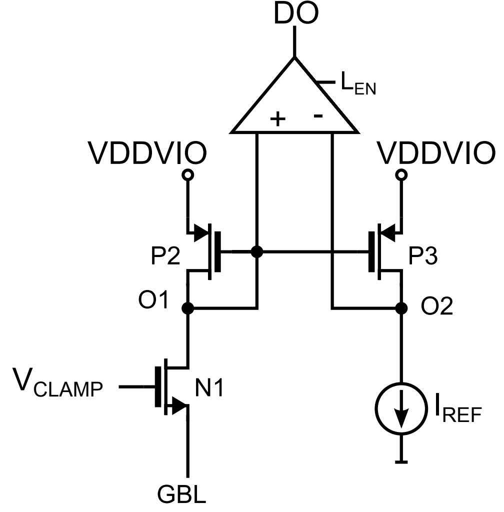

Figure 3.13: Current Sense Amplifier with Type A Current Conveyor

area. This approach was extensively used in standalone NOR Flash memories [19], [20]. For low voltage applications the bitline regulation is skipped and the diode connected PMOS is connected directly to the bitline (Type A) [15], [21], [22] [23] as shown in figure 3.14. The minimum supply voltage for the low power Type A Conveyor is:

$$VDD_{MIN} = V_{GBL} + |V_{TH,P2}| + |V_{OV,P2}| \tag{3.1}$$

where the $V_{TH,P2}$ and $V_{OV,P2}$ is the threshold and the overdrive voltage of P2. By this means the bitline voltage level will have strong temperature dependence and is not suitable for automotive application requiring high temperature operation for memories utilizing eFlash transistors biased in linear region. In addition the $PSRR$ is nearly zero, since the bitline voltage is directly related to the supply voltage. For a very low voltage and low power realization the folded cascode approach (Type A) is used as illustrated in figure 3.15 requiring a sophisticated precharge circuitry consuming considerable amount of area for a fast read operation [13]. The minimum supply voltage can now be reduced to:

$$VDD_{MIN} = V_{GBL} + |V_{OV,P2}| \tag{3.2}$$

However the global bitline voltage V_{GBL} has to be higher than the gate source voltage of N1:

$$V_{GBL} \geq V_{TH,N1} + V_{OV,N1} \tag{3.3}$$

The dotted PMOS transistor can be inserted for bitline voltage regulation forming a source follower structure similar to figure 3.13. This circuit exhibits good $PSRR$ due to symmetrical biasing scheme.

The latch comparator as shown in figure 3.8 is not suitable for the voltage amplifier stage in the current sense amplifier since it actively draw current at its inputs in the sense phase. The simplest voltage comparator for the voltage sensing stage is realized by placing an inverter to the O2 node in figure 3.10 [9]. It is a good approach for Type A, B and D current conveyors [9]. For low voltage swing on O2 and O1 nodes the high

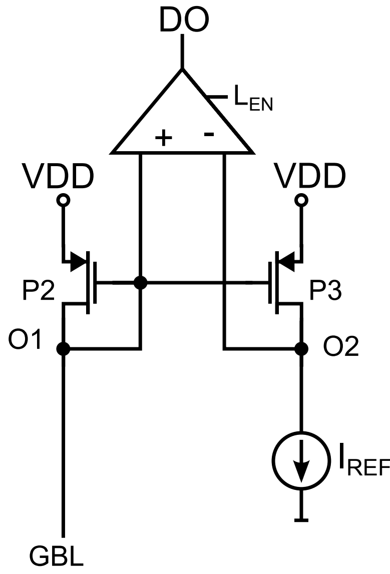

Figure 3.14: Current Sense Amplifier with Type A Current Conveyor for Low Voltage

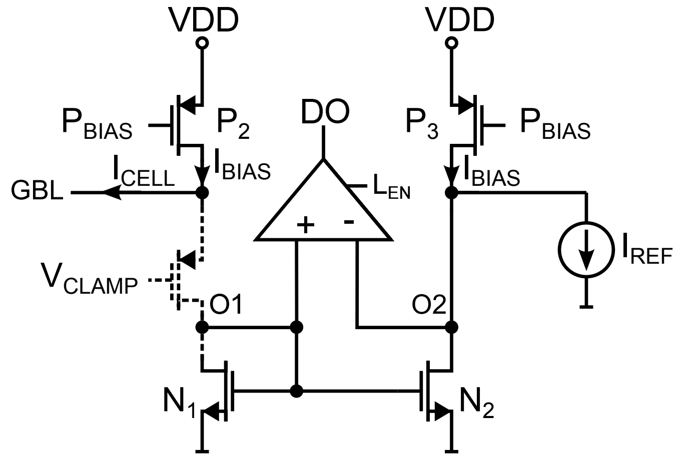

Figure 3.15: Current Sense Amplifier with Type A Current Conveyor for Very Low Voltage

impedance latch-type sense amplifier [6] is used. For example in case of the Type A current conveyor for low voltage (see figure 3.14) the latch-type sense amplifier is a good choice. For low supply voltage applications the double-tail approach was proposed [24]. The Type A conveyor for very low voltage (see figure 3.15) is a good use case for the double-tail sense amplifier.

3.4 Read Access Time Considerations

3.4.1 Read Path Model

In order to further understand the strengths and weaknesses of the different sensing approaches the basic model for read access time in NOR-Flash is provided in this chapter. The bitline path has to be modeled since it defines the read access time of the memory. The wordline path modeling is omitted since its time constant is usually smaller to ensure a nearly constant gate source voltage of the memory cell during the sensing phase. The figure 3.16 shows a simplified bitline path for a NOR-Flash memory. The cell is

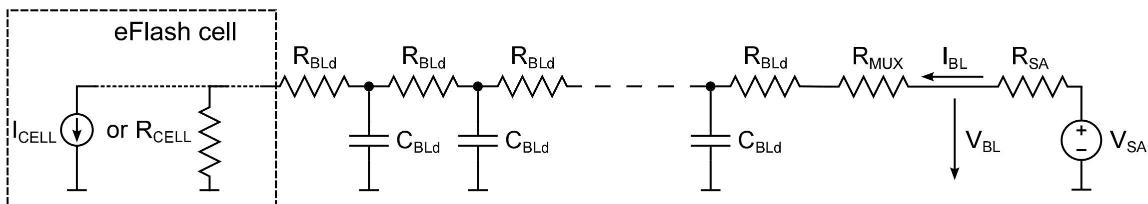

Figure 3.16: Simplified Bitline Path for NOR-Flash Memory

modeled either by the resistance R_{CELL} or by constant current source I_{CELL} depending on the operation region of the eFlash cell. The capacitance and resistance of the bitline is modeled by the distributed RC delay line with R_{BLd} for resistance value and C_{BLd} for capacitance value per unit length. The sense amplifier is modeled by its input resistance R_{SA} and the bitline multiplexer is represented by R_{MUX}. The overall bitline path resistance is in the range of several kilo-ohms compared to several hundreds of ohms for other types of memories like SRAM, DRAM or MRAM. It is due to the fact that the multiplexer requires a thick gate oxide device with high on-resistance to withstand high voltages (higher 10V) required to write or erase the Flash cells. The specific split gate cell allows the usage of low voltage read path without the thick gate oxide device in the bitline multiplexer [17],[8],[12].

For voltage sensing the input resistance of the sensing circuit is several orders of magnitude higher compared to the values of other resistances in the read path. For this reason R_{SA} is modeled as an open for the voltage sensing approach [6]. The typical value for the input resistance of a current sense amplifier is about 1kΩ [6].

The time constant τ_{BL} for the for distributed RC line is approximated by [25], [26] as

$$\tau_{BL} = \frac{R_{BL}C_{BL}}{2} \tag{3.4}$$

where R_{BL} and C_{BL} represent the lumped resistance and capacitance value of the bitline. By this means the read path can be simplified to the circuit shown in Figure 3.17 where the distributed RC line is replaced by the resistance $R_{BL}/2$ and capacitance C_{BL}. The

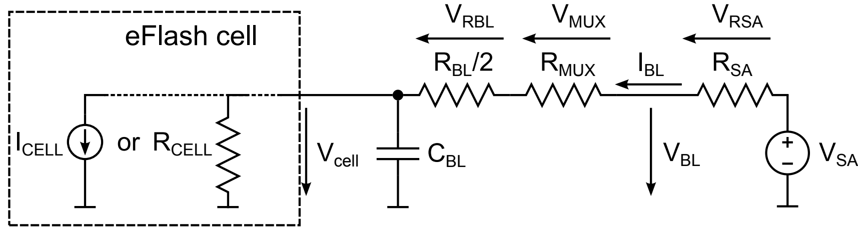

Figure 3.17: Simplified Read Path for NOR-Flash memory

typical values of the bitline capacitance and resistance for an embedded NOR-FLASH memory are about 1pF and 2kΩ respectively. This corresponds to τ_{BL} of 1ns which is a first order approximation for the sensing speed of the current sensing scheme used in SRAM type memories [26], [6], calculated for a continuous (nontruncated) linear-ramp signal or by applying the small signal analysis to the simplified read path (see figure 3.17). However the reported read access times for current sensing based embedded NOR-Flash memories are in a range of several tens of nanoseconds [9], [13], [22], [27]. This discrepancy occurs because the large signal bitline charging phase is not considered, which is typical for Flash based memories. During this precharge phase the sense amplifier circuitry enters the large-signal operation introducing strong non-linear effects to the transfer function of the circuit. Therefore, the equation 3.4 describes the current sensing delay only for a steady state operation point of the current sense amplifier with negligible multiplexer resistance if no large-signal charging effects of the bitline capacitance are considered.

These effects are getting more important if the resistance of the bitline multiplexer is in the same order of magnitude as the intrinsic bitline resistance and if the bitline voltage changes its value during the precharge phase considerably, which is the case for the Flash based memories.

3.4.2 Voltage Sensing

Figure 3.18 illustrates the access time partitioning for the voltage sensing. As already

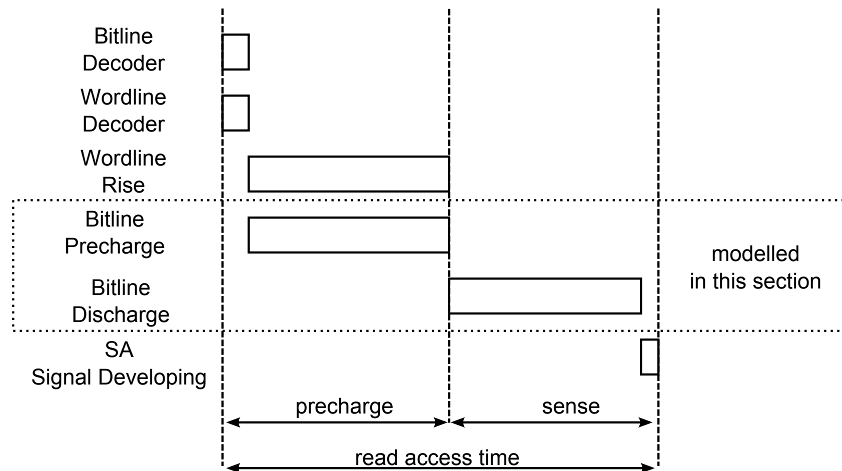

Figure 3.18: Access Time Partitioning for Voltage Sensing

mentioned in section 3.1 the sensing operation consists of two phases: precharge and sense. During the precharge phase the decoder tree addresses the selected bitline and wordline. Subsequently the bitline and wordline are charged simultaneously. The bitline is driven by the precharge circuit from ground to a fixed potential, which is typically about $V_{SA} = 1V$. The wordline voltage has to reach its final value before the sensing phase starts, since the cell current must be nearly constant during the bitline discharge. In the sense phase the bitline and the reference bitline are discharged by the cell and reference current respectively. After sufficient voltage difference is achieved between the reference and a memory bitline, the sense amplifier is activated amplifying the input voltage difference and latching the digital output. The delay of the decoder will be omitted in the following since it usually consumes a very small portion of the access time budget.

If the bitline capacitance is even not fully charged but the V_{BL} voltage (see figure 3.17) reached the precharge level the precharge phase can be terminated. The subsequent charge sharing between the sense amplifier input and the bitline capacitance does not impact the sensing accuracy since the voltage sensing is usually using full differential approach by matching the memory and the reference path in terms of resistance and capacitance [12]. By this means the charge sharing occurs in both paths like common mode offset and does not disturb the bitline voltage difference between the bitline and the reference bitline. The resistance of the voltage sense amplifier during the precharge phase is given by the on resistance of the precharge switch devices (see chapter 2).

The sum of all resistances in the bitline path R_{BLpath} is

$$R_{BLpath} = R_{BL}/2 + R_{SA} + R_{MUX} \tag{3.5}$$

In the following calculations the impact of R_{CELL} or I_{CELL} on the charging speed of the bitline capacitance and the final V_{CELL} voltage level is neglected, since for the linear region biased cell the cell resistance R_{CELL} is more than one order of magnitude higher compared to the R_{BLpath} and for an in saturation biased cell I_{CELL} will cause only a small voltage drop ($\approx 40mV$) on the bitline path resistance.

The bitline voltage during the precharge phase is calculated by

$$\begin{aligned} V_{BL} &= V_{SA} - V_{RSA} \\ &= V_{SA}\left(1 - \frac{R_{SA}}{R_{BLpath}}\, e^{\frac{-t}{R_{BLpath}C_{BL}}}\right) \end{aligned} \tag{3.6}$$

The precharge time T_{preV} needed for the V_{BL} to reach 99% of the V_{SA} is

$$T_{preV} = R_{BLpath}C_{BL}ln\left(\frac{R_{SA}}{0.01R_{BLpath}}\right) \tag{3.7}$$

The resistance $R_{SA} = 100\Omega$ for voltage sensing during the precharge can be easily achieved, since the precharge device acts as a digital switch and does not require any regulation loop. For the typical values of the resistances and capacitances in the read path for voltage sensing during the precharge phase ($R_{SA} = 100\Omega, C_{BL} = 1pF, R_{BL} = 2k\Omega, R_{MUX} = 2k\Omega$) the model 3.7 predicts a precharge time of 3.6ns. The bitline voltage after the precharge phase $V_{BL,pre}$ is calculated as follows:

$$V_{BL,pre} = V_{SA}\left(1 - e^{\frac{-T_{preV}}{R_{BLpath}C_{BL}}}\right) \tag{3.8}$$

For the precharge time of 3.6ns and the typical values ($C_{BL} = 1pF$, $R_{CELL} = 50k\Omega$, $V_{sense} = 100mV$, $R_{BLpath} = 3.1k\Omega$) $V_{BL,pre}$ is 0.7V which is significantly lower compared to the bitline voltage at the end of the precharge phase (0.99V). This is caused by the voltage drop on multiplexer and bitline resistance due to the remaining charging current for the bitline capacitance. This will cause a negative voltage bump on V_{BL} after the precharge phase. Since the voltage sensing circuits are usually realized in full symmetrical manner (see chapter 2) this bump will be seen as a common mode offset in the sense amplifier. It has to be mentioned that the bitline path is modeled by a lumped capacitance and resistance which represents the worst case for a voltage bump after a capacitance charging. For eFlash cell biased in saturation region this bump is not important but for linear region biased cell this is critical due to negative feedback on the cell current.

The sense delay for the voltage sensing T_{senseV} is the time required until the voltage difference between the memory bitline and the reference bitline reach the voltage V_{sense}

required for robust operation of the comparator latch. The bitline voltage difference V_{sense} for the cell biased in saturation is calculated as following:

$$\begin{aligned} V_{sense,SAT} &= (V_{BL,pre} - V_{BLREF}) - (V_{BL,pre} - V_{BL}) \\ &= \frac{T_{senseV,SAT} I_{REF}}{C_{BL}} - \frac{T_{senseV,SAT} I_{CELL}}{C_{BL}} \\ &= \frac{T_{senseV,SAT}\,(I_{CELL} - I_{REF})}{C_{BL}} \end{aligned} \tag{3.9}$$

And hence the sense delay is:

$$T_{senseV,SAT} = \frac{C_{BL} V_{sense}}{I_{CELL} - I_{REF}} \tag{3.10}$$

For $V_{sense} = 100mV$ and typical values ($C_{BL} = 1pF$, $I_{CELL} = 20\mu A$, $I_{REF} = 10\mu A$) the sense delay is 10ns and the bitline will be discharged by 200mV during the sense phase. Of course the sense delay can be significantly reduced if the V_{sense} voltage becomes smaller. The lower limit for V_{sense} is given by the transistor mismatch in the sense amplifier circuit and the overall system noise. V_{sense} of 100mV is already hard to achieve [6].

In case the linear region biased memory transistor is discharging the bitline with respect to a constant reference current source on the reference bitline the calculation for the voltage difference V_{sense} changes as following:

$$V_{sense,LIN} = V_{BLREF} - V_{BL} = V_{BL,pre} - \frac{T_{senseV,SAT} I_{REF}}{C_{BL}} - V_{BL,pre}\, e^{-\frac{T_{senseV,SAT}}{R_{CELL} C_{BL}}} \tag{3.11}$$

There is no explicit solution for this equation with respect to the sense time $T_{senseV,SAT}$ but the numerical solution shows that for the bitline precharge level $V_{BL,pre}$ of 0.7V, $V_{sense,LIN}$ will never reach 100mV, since the bitline discharge is RC like. For $V_{BL,pre}$ of 0.9V the sense delay is about 22ns which is more than a factor of two higher with respect to the delay if a cell is used which is biased in saturation region. The bitline precharge time for $V_{BL,pre} = 0.9V$ is 7.2ns which is also a factor of two higher compared to 3.6ns calculated for a cell biased in saturation.

The access time for the voltage sensing T_{accV} using the typical values given above is 13.6ns for an in saturation biased cell and 29.2ns for a cell biased in linear region. The linear region biased cell heavily degrades the access time if a voltage sensing scheme as shown in figure 3.8 is used, and it is not suitable for high speed and reliable read operation, hence only the in saturation region biased cell will be considered in the following for the voltage sensing scheme. The read access time for voltage sensing is calculated as following:

$$T_{accV} = T_{preV} + T_{senseV} = R_{BLpath} C_{BL} ln\left(\frac{R_{SA}}{0.01 R_{BLpath}}\right) + \frac{C_{BL} V_{sense}}{I_{CELL} - I_{REF}} \tag{3.12}$$

The latching time of the voltage comparator is omitted in this model, since it is in a range of several tens of picoseconds [6] which is more than two orders of magnitude lower compared to the read access time.

3.4.3 Current Sensing

Figure 3.19 illustrates the corresponding access time partitioning for the current sensing.

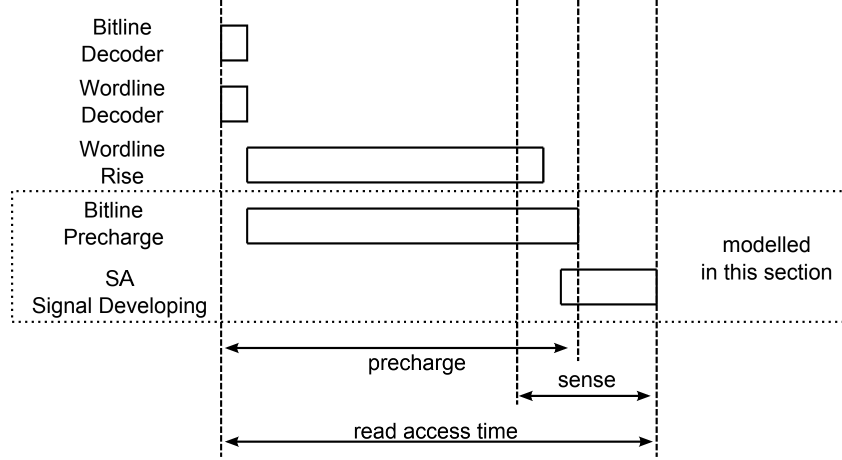

Figure 3.19: Access Time Partitioning for Current Sensing

Similar to the voltage sensing approach the bitline and wordline voltages are charged during the precharge phase. The bitline is driven by the sense amplifier from ground to a fixed potential, which is typically about $V_{SA} = 1V$. There is no absolutely defined border between the precharge and sense phase, since the bitline is still charging in the sense phase by the sense amplifier and the internal signal development starts.

To obtain the read access time model for the current sensing the bitline current instead of the bitline voltage is taken into account. The bitline current I_{BL} during the precharge phase is the sum of the precharge current I_{PRE} required to charge the bitline capacitance and the cell current I_{CELL}:

$$I_{BL} = I_{PRE} + I_{CELL} \tag{3.13}$$

In the following calculations the impact of R_{CELL} or I_{CELL} on the charging current and speed of the bitline capacitance is neglected similar to the voltage sensing model. Since the current sense amplifier forces the bitline voltage to a fixed potential there is no difference between a cell biased in linear or in saturation region with respect to the access time. In the following only a memory transistor biased in linear region is considered for current sensing without loss of generality, where:

$$R_{CELL} = \frac{V_{SA}}{I_{CELL}} \tag{3.14}$$

The precharge current is defined by the bitline capacitance which is charged by the sense amplifier through the multiplexer device.

$$I_{PRE} = \frac{V_{SA}}{R_{BLpath}} \, e^{\frac{-t}{R_{BLpath}C_{BL}}} \tag{3.15}$$

The precharge time T_{preI} needed to decrease the precharge current I_{PRE} to a given value is calculated by

$$T_{preI} = R_{BLpath} C_{BL} ln \left(\frac{V_{SA}}{R_{BLpath} I_{PRE}} \right) \tag{3.16}$$

Compared to voltage sensing the precharge time for the current sensing is directly proportional to the path resistance R_{BLpath}. This is especially important for the sensing operation in Flash type memories, since as already mentioned the resistance of the bitline multiplexer is relatively high compared to other memory types. For reliable read operation using current sensing the precharge current I_{PRE} during the sensing phase has to be well below the reference current level I_{REF} for robust read operation, e.g. $I_{PRE} = I_{REF}/2$. As already mentioned in section 3.3.2, I_{REF} is usually set to half of the value of the low resistive cell current. This relationship implies:

$$I_{PRE} = \frac{1}{2} I_{REF} = \frac{1}{2}\frac{1}{2} I_{CELL} = \frac{V_{SA}}{4R_{CELL}} \tag{3.17}$$

yielding the following expression for the precharge time:

$$T_{preI} = R_{BLpath} C_{BL} ln \left(\frac{4R_{CELL}}{R_{BLpath}} \right) \tag{3.18}$$

For the typical values of the resistances and capacitances in the read path ($R_{SA} = 1k\Omega$, $C_{BL} = 1pF$, $R_{BL} = 2k\Omega$, $R_{MUX} = 2k\Omega$, $R_{CELL} = 50k\Omega$) the model 3.18 predicts a precharge time for the current sensing of 15.6ns, which is far above the intrinsic delay of the RC line ($\tau_{BL} = \frac{R_{BL} C_{BL}}{2} = 1ns$). To speed up the precharge phase the current sense amplifier usually includes a dedicated precharge circuitry [6] reducing the input resistance of the sense amplifier. Assuming $R_{SA} = 500\Omega$ which is already hard to achieve within a typical power and area specifications, the precharge time T_{preI} for current sensing is lowered only to 14.2ns, which is still more than a factor of three higher compared to the precharge time for the voltage sensing ($T_{preV} = 3.6ns$). To significantly reduce the precharge time of current sensing, the path resistance, which is dominated by the multiplexer on resistance, has to be reduced. The proposed principle of the multiplexer compensation [6] is a good approach for reducing the impact of the multiplexer resistance on precharge and sensing phase. It requires a second path to the bitline which has a significant area impact in Flash memories due to high voltage design rules in the bitline multiplexer.

After the precharge phase is completed, the sensing phase starts. For the current sensing the delay of the sense amplifier circuit based on Wilson Type (Type D as shown in figure 3.10) or regulated cascode (Type B as shown in figure 3.12) architecture is derived by small signal analysis in [6]. The small signal analysis is a valid approach for the analysis of the sense phase, since the sensing circuits are almost in their final operation point. The sense delay for the Type D or B single ended current sensing $T_{senseI,DB}$ depends on bitline capacitance, transconductance of the input device g_m and the gain of the feedback amplifier A_0. It can be approximated as:

$$T_{senseI,DB} = \frac{C_{BL}}{A_0 g_m} \tag{3.19}$$

For the typical values ($C_{BL} = 1pF$, $A_0 = 10$, $g_m = 200uA/V$) equation 3.19 yields a sense time of 500ps, which indeed is a very low value compared to the precharge time T_{preI}. For the current sense amplifier using cascode or diode structure (Type A as shown in figure 3.13, 3.14, 3.15) the delay time of the current conveyor stage is approximated by [6] as:

$$T_{senseI,A} = \frac{C_{BL}}{g_m} \tag{3.20}$$

Compared to equation 3.19 Type A has a factor of $A_0 = 10$ higher delay. $T_{senseI,A} = 5ns$ in case the typical values are used, which is already more than one-third of the precharge time T_{preI}.

The access time for the current sensing scheme utilizing Type B or D current conveyor circuit $T_{accI,DB}$ can be approximated as:

$$T_{accI,DB} = T_{preI} + T_{senseI,DB} = R_{BLpath}C_{BL}ln\left(\frac{4R_{CELL}}{R_{BLpath}}\right) + \frac{C_{BL}}{A_0 g_m} \tag{3.21}$$

And for sensing scheme using Type A current conveyor the access time $T_{accI,A}$ is approximated as:

$$T_{accI,A} = T_{preI} + T_{senseI,A} = R_{BLpath}C_{BL}ln\left(\frac{4R_{CELL}}{R_{BLpath}}\right) + \frac{C_{BL}}{g_m} \tag{3.22}$$

For typical values in embedded NOR-Flash memory ($R_{SA} = 1k\Omega$, $C_{BL} = 1pF$, $R_{BL} = 2k\Omega$, $R_{MUX} = 2k\Omega$, $R_{CELL} = 50k\Omega$, $A_0 = 10$, $g_m = 200uA/V$) the access time is 14.7ns for Type D or B and 19.7ns for Type A circuits.

3.4.4 Comparison

By comparing the calculated values for an access time of the current sensing (14.2ns or 19.7ns) and the voltage sensing (13.6ns) by using typical values occuring in NOR-Flash memories, it becomes obvious that the voltage sensing is fully competitive with the current sensing approach. Of course all previous calculations are first order approximations since they are based on simplified array and sensing circuits. But they provide good insight on the strengths and weaknesses for both sensing concepts.

Figure 3.20 illustrates the access time for both sensing approaches predicted by 3.21, 3.22 and 3.12 with respect to the bitline capacitance. The bitline resistance is proportionally scaled to

$$R_{BL} = \frac{C_{BL}}{1pF} 2k\Omega \tag{3.23}$$

It is clear that all equations are proportional to the bitline capacitance C_{BL}. The current sensing will not gain any speed advantage over voltage sensing approach if the bitline capacitance increases, as has been stated earlier in [26] and [6]. It must be mentioned that these previous works were suited for the SRAM memory where the bitline multiplexer resistance is several magnitudes smaller compared to Flash type memories.

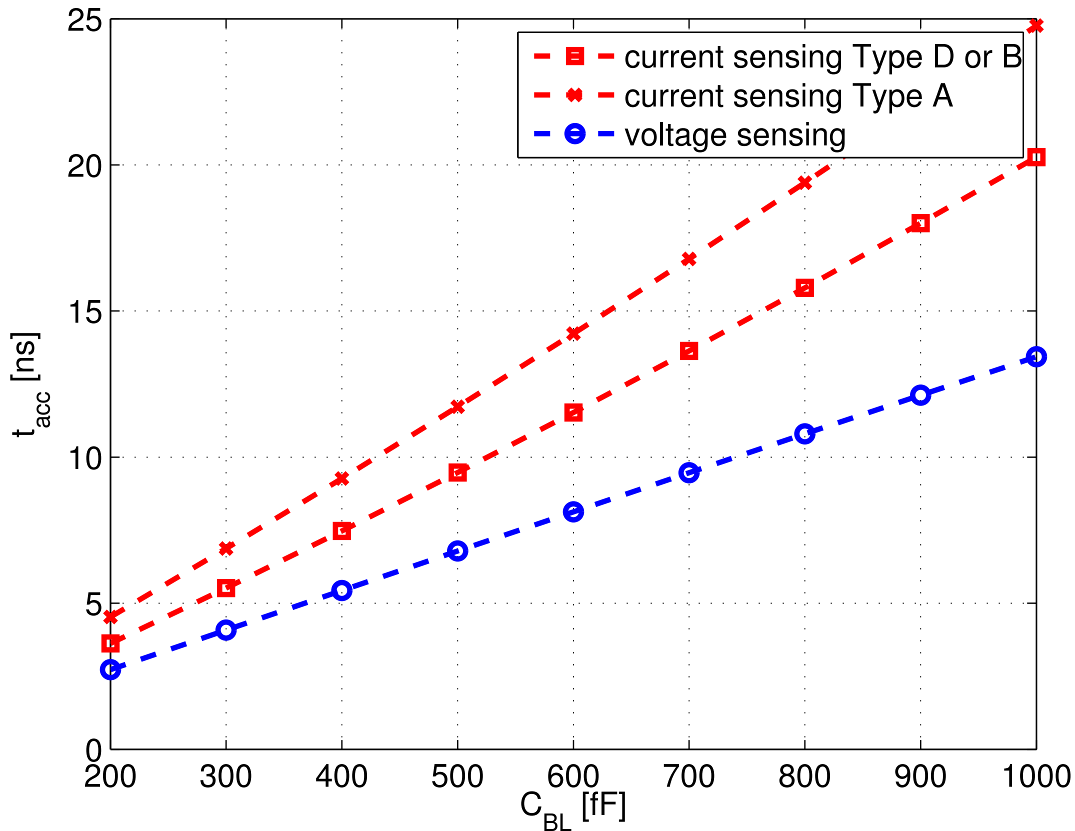

Figure 3.20: Calculated Access Time vs. Bitline Capacitance

Figure 3.21 shows the access time for voltage and current sensing derived from the equations 3.21, 3.22 and 3.12 versus the cell current for three different values of the multiplexer resistance R_{MUX}. The current sensing outperforms the voltage sensing if the cell current significantly decreases (below $10\mu A$), as shown in the figure 3.21. There is a crossover point of the access time curves for each value of R_{MUX}. For low values of R_{MUX} e.g. 100Ω the crossover occurs at a cell current value of about $45\mu A$ for Type D or B and $20\mu A$ in case Type A current sensing is used. The crossover moves to lower cell currents if the multiplexer resistance increases. This is one of the main reasons why the standalone NAND Flash memories which suffer from low cell currents (below $1\mu A$) due to series connection of the memory cells in the NAND-string switched from voltage to current sensing utilizing the all bitline architecture (ABL) [28]. For typical values in Flash-memories, the crossover point appears for $I_{CELL} \approx 17\mu A$ (for Type D and B) and $I_{CELL} \approx 12\mu A$ (for Type A) which is below the typical cell current of $20\mu A$.

Lowering or increasing the multiplexer resistance will speed up or slow down the current sensing significantly compared to voltage sensing. It means the only design parameter having different influence on the two sensing schemes is the bitline path resistance R_{BLpath} which is usually dominated by the multiplexer resistance R_{MUX}. As already mentioned the bitline multiplexer usually requires thick oxide devices to block the high voltages du-

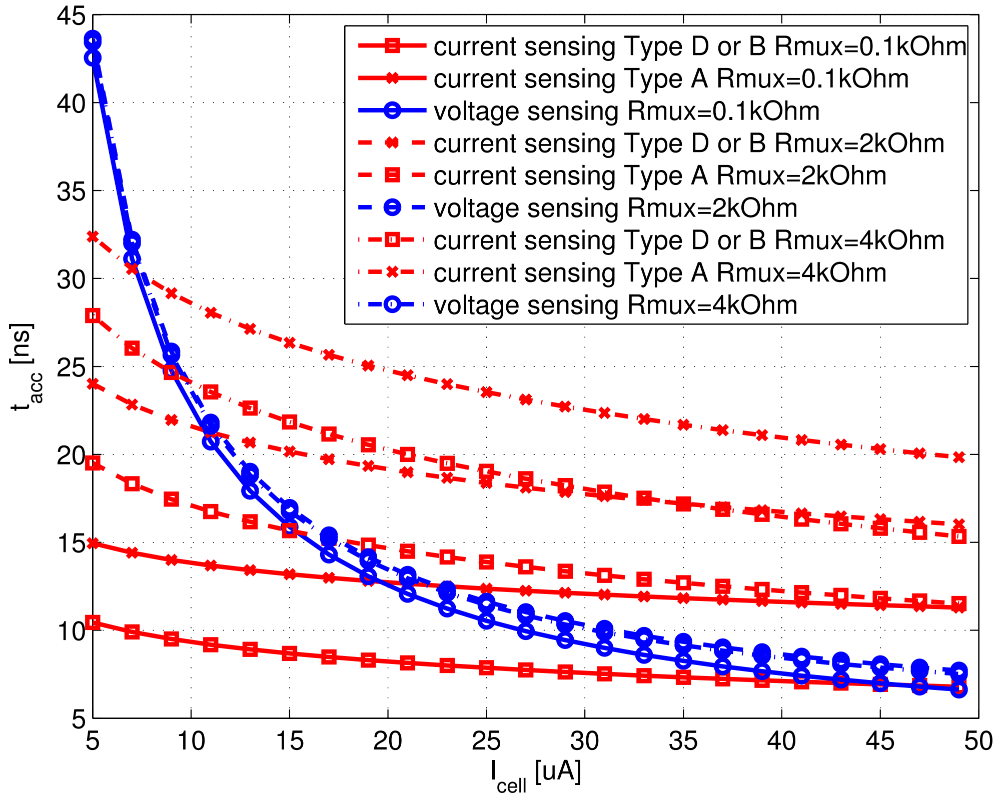

Figure 3.21: Calculated Access Time vs. Cell Resistance

ring erase and program operation in the Flash memory. Increasing the transistor widths of the bitline multiplexer will lead to a significant area increase on the memory macro setting the lower bound of the multiplexer resistance.

For NOR-Flash typical values of the bitline capacitance and read path resistances the read access time model derived in this chapter predicts the voltage sensing approach to be fully competitive with current sensing. The voltage sensing approach (as shown in figure 3.8) is only suitable for a memory cell operating in saturation otherwise it will suffer from strong read speed degradation and robustness issues (see section 3.4.2), whereas the current sensing is independent on memory cell operation region.

3.5 Power Considerations

The power and current consumption of the sensing scheme is a very important design parameter in today's embedded memories, since they require high read throughput of several GByte/s [9] [8]. This is achieved by reading more than hundred sense amplifiers simultaneously leading to high power consumption in the read path of the memory macro.

The power considerations in this work are done from the system level perspective. Because the total current and therefore the power drained out of the supply domain (e.g.

V_{DD}) is one of the main parameters in a memory module specification. The equation for power calculation:

$$P = \frac{\int_0^{T_{acc}} U(t)\, I(t)\, dt}{T_{acc}} \tag{3.24}$$

is approximated without loss of generality by:

$$Pvdd = V_{DD} Ivdd \tag{3.25}$$

where V_{DD} is a constant voltage, as it is usually provided by the systems voltage regulator or from an external solution. The current $Ivdd$ is the mean current during the read access cycle and is the sum of constant bias and dynamic charging current. The mean current required to charge a bitline is calculated by:

$$I_{BL,charge} = \frac{C_{BL} V_{DD}}{T_{acc}} \tag{3.26}$$

To understand the strengths and weaknesses of the different sensing approaches with respect to their power consumption the basic power models for the different sensing approaches and circuits are provided in this section. It the it will be assumed that the high current cell state is continuously read with maximum read speed during the read operation, which is the worst case for the power consumption.

3.5.1 Voltage Sensing

The dynamic latch-type sense amplifier shown in figure 3.8 is extensively used in SRAM and DRAM memory types too due to its low power consumption and fast amplification speed [6]. The cross currents during the latching phase are usually negligible compared to the charging current required during the precharge phase of the bitline. In addition the cell and the reference currents are flowing to ground through the precharge devices in the precharge phase. In the following the simplified assumption is made that the cell and the reference current are flowing during the whole precharge phase. This is a worst case consideration, since the cell current will only flow after the wordline reaches a voltage level above the threshold voltage of a memory cell. After the latching phase is completed the latch remains in its state without any active current flow from power supply to ground if the leakage current is ignored. Hence the basic model for the current consumption of the voltage sensing utilizing dynamic latch type sense amplifier $Ivdd_V$ is defined by the charge required to drive the reference and the memory bitline in addition to cell and reference current flow during the precharge operation:

$$Ivdd_V = \frac{V_{SA} \cdot 2C_{BL}}{T_{accV}} + (I_{CELL} + I_{REF}) \frac{T_{preV}}{T_{accV}} \tag{3.27}$$

The first addend of the equation calculates the current required to charge the two bitlines and the second one the cell and the reference current flowing during the precharge phase. Since the bitline and reference bitline capacitance are discharged to ground there is no

active current path between V_{DD} and V_{SS} in the sense phase (see section 3.1). The power consumption is calculated as:

$$Pvdd_V = \left(\frac{V_{SA} \cdot 2C_{BL}}{T_{accV}} + (I_{CELL} + I_{REF})\frac{T_{preV}}{T_{accV}}\right) V_{DD} \tag{3.28}$$

where V_{DD} is the core logic supply. For typical values in NOR-Flash ($C_{BL} = 1pF$, $V_{SA} = 1V$, $V_{DD} = 1.2V$, $I_{REF} = 0.5 \cdot I_{CELL} = 10\mu A$), the values for the precharge phase $T_{preV} = 3.6ns$ and the access time $T_{accV} = 13.6ns$ as derived in the previous section the power required per one sense amplifier is $Pvdd_V = 157\mu A \cdot 1.2V \approx 188mW$. To reduce the power consumption of the voltage sense amplifier the bitline capacitance per sense amplifier has to be as small as possible. The proposed voltage sensing scheme in [8] utilizes small sector sizes with local sense amplifiers sacrificing module area versus low power consumption and high sensing speed.

3.5.2 Current Sensing

Compared to the voltage sense amplifier the current sense amplifier requires a static reference and cell current flow during the read operation due to the current conveyor circuit and the subsequent current to voltage conversion stage. The Type A current conveyor as shown in figure 3.13 and 3.14 requires a static current flow $Ivdd, static_{I,A}$ of:

$$Ivdd, static_{I,A} = I_{REF} + I_{CELL} \tag{3.29}$$

during the whole read operation. In addition the precharge current $Ivdd, pre_{I,A}$ required for charging the bitline (similar to voltage sensing) has to be taken into account, however only one bitline per sense amplifier is charged during the precharge phase, as the current sensing is realized in single ended manner (see section 3.3.2). The average precharge current is:

$$Ivdd, pre_{I,A} = \frac{V_{SA}C_{BL}}{T_{accI,A}} \tag{3.30}$$

The overall power consumption is calculated as:

$$Pvdd_{I,A} = \left(I_{REF} + I_{CELL} + \frac{V_{SA}C_{BL}}{T_{accI,A}}\right) V_{DD} \tag{3.31}$$

where V_{DD}=1.2V is the core logic supply yielding $Pvdd_{I,A} \approx 97mW$ in case the low voltage realization is used (see figure 3.14); or V_{DD} is the I/O voltage supply (typically $V_{DDVIO} = 3.3V$) with $Pvdd_{I,A} \approx 267mW$ if the bitline clamp transistor is used for bitline regulation (see figure 3.13).

If a high bitline voltage voltage is desired under low voltage operation the current sense amplifier with Type A current conveyor for very low voltage (see figure 3.15) should be used. The folded cascode approach is biasing the current branches by a constant current I_{BIAS} during the whole read operation. To guarantee a robust circuit operation I_{BIAS} is typically set slightly higher than I_{CELL}:

$$I_{BIAS} = 1.2I_{CELL} \tag{3.32}$$

The overall power consumption of the low voltage folded cascode Type A current sense amplifier is:

$$Pvdd_{I,Alv} = \left(2 \cdot 1.2 I_{CELL} + \frac{V_{SA} C_{BL}}{T_{accI,A}}\right) V_{DD} \tag{3.33}$$

Because this circuit is taken for very low voltage current sensing the core logic supply is taken for calculation obtaining $Pvdd_{I,Alv} \approx 118mW$ for typical values in NOR Flash memories.

If sufficient supply voltage headroom is available the regulated cascode (Type B see figure 3.12) or the Wilson Type (Type D see figure 3.10) are used regulating the bitline voltage by a local opamp from the I/O supply domain requiring a significant amount of power. These approaches require an additional current compared to Type A due to the opamp for clamp device regulation. The opamp bias current $I_{BIAS,OP}$ is usually in the range of $I_{BIAS,OP} \approx I_{CELL}$. The power consumption is obtained by adding the $I_{BIAS,OP}$ current to the current consumption of Type A sense amplifier:

$$Pvdd_{I,DB} = \left(I_{REF} + 2I_{CELL} + \frac{V_{SA} C_{BL}}{T_{accI,DB}}\right) V_{DD} \tag{3.34}$$

For $V_{DD} = 3.3V$ and typical values for NOR-Flash $Pvdd_{I,DB} \approx 333mW$ which is the highest value for power consumption compared to the other types of sense amplifiers.

3.5.3 Comparison

For a fair comparison of different sensing approaches and circuits the values for the power consumption calculated in the previous sections are scaled with the corresponding access time calculated in section 3.4. The inverse value of this product is taken to define the figure of merit for power efficiency and speed (FOM_{PS}):

$$FOM_{PS} = \frac{1}{PowerConsumption \times ReadAccessTime} \tag{3.35}$$

Table 3.1 summarizes the values for different sensing approaches obtained by the models derived in the previous sections for the read access time, the power consumption and FOM_{PS}. These models predict that for access times below 15ns the current sensing approach (Type B or D) requires about two times higher power. However the voltage sensing is only suitable for cells biased in saturation, since it will suffer from strong read speed degradation and robustness issues (see section 3.4.2), if a cell operating in the linear regime is used. The power reduction for the current sensing approach by skipping the local opamp (Type A) heavily degrades the read access time. The best FOM_{PS} value is achieved by the Type A for low voltage, but this approach has very limited bitline voltage headroom and the bitline voltage level is strongly dependent on temperature. Therefore it is not suited for robust read operation at high temperatures required by automotive applications. The Type A for very low voltage applications (see figure 3.15) achieves a good compromise between bitline voltage range and speed. In addition the usage of clamp device allows temperature dependent V_{CLAMP} voltage adjustment to provide stable bitline voltage over a wide temperature range.

Table 3.1: Voltage versus Current Sensing

	Voltage Sensing	Type A	Type A for low voltage	Type A for very low voltage	Type B	Type D
Access Time	13.6ns	19.7ns	19.7ns	19.7ns	14.7ns	14.7ns
Power per Sense Amplifier	188mW	267mW	97mW	118mW	333mW	333mW
FOM_{PS} in $\frac{10^6}{Ws}$	391	190	523	430	204	204

It has to be mentioned here that the models derived in the previous sections are only a first order approximation to provide a rough estimation since they don't take into account the speed and power consumption of the precharge circuitry required for Type A as well as the current consumption of the reference current distribution.

3.6 Summary and Conclusion

The voltage and current sensing approaches and corresponding circuits have been analyzed and evaluated with respect to speed, power, low voltage operation, area, power supply rejection ratio ($PSRR$), cell type compatibility and cell current dependence. Figure 3.22 summarizes the findings of the previous sections.

The voltage sensing approach exhibits a very good overall tradeoff. The voltage sensing provides:

- fastest read access time due to lower precharge time compared to current sensing
- low power consumption due to zero DC current flow during the sensing phase
- scalable power consumption and access time by adjusting the bitline capacitance value
- parallel connection of cell current path and voltage sensing circuit offering a maximum available voltage headroom required for low voltage applications
- low area consumption due to utilization of core logic devices, simple precharge and sensing circuit structure
- good power supply rejection ratio due to a fully symmetrical structure

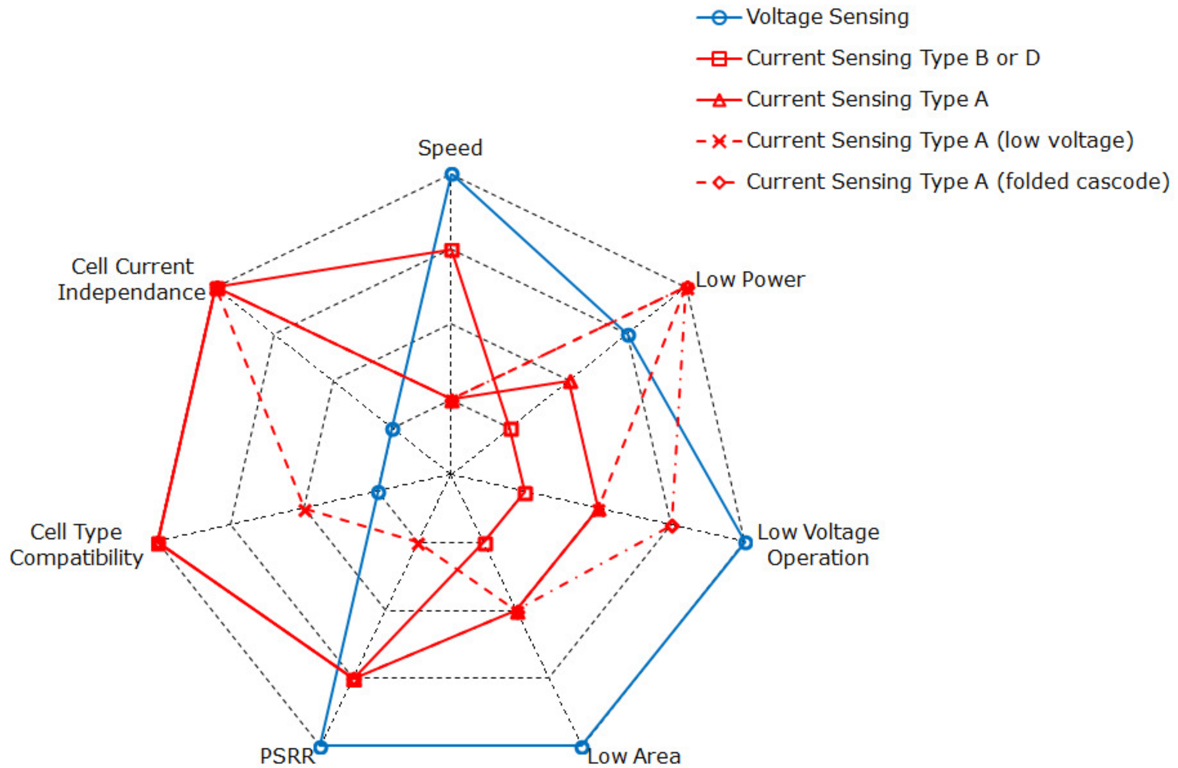

Figure 3.22: Voltage vs. Current Sensing Comparison

However voltage sensing reveals strong dependance on cell current and thus is only suitable for cell types biased in saturation.

The current sensing is not dependent on cell type and it has lower speed degradation with respect to lower cell current compared to voltage sensing. The current sensing approach exhibits against voltage sensing:

- high sensing speed at the cost of a high amount of power (Type B or D)
- lower power consumption with reduced sensing speed (Type A)
- series connection of sense amplifier and memory cell that requires higher supply voltage for same bitline voltage headroom as voltage sensing
- higher area consumption due to current conveyor circuit and current to voltage conversion stage
- lower power supply rejection ratio due to single ended approach for power and area reduction

For embedded Flash cell types biased in saturation the voltage sensing delivers faster higher performance at lower power consumption as the current sensing approach (Type B or D). But if an embedded Flash cell exhibits strong drain voltage dependance the

voltage sensing will strongly degrade in speed and robustness compared to the current sensing.

Chapter 4

Bitline-Capacitance-Cancellation Sensing Scheme

In this section a novel sensing scheme developed in this work is introduced which combines the strengths of voltage and current sensing to achieve high speed performance at low power consumption and has low dependence on the particular cell type. This is achieved by using time a domain sensing scheme [29], [30] as the basic approach in combination with a novel sense amplifier comprising slope detection circuit [1, Jefremow].

4.1 Time Domain Sensing Concept

Figure 4.1 shows the basic difference between voltage and time domain sensing. During the precharge phase a bitline and a reference bitline are precharged to V_{PRE} and in the subsequent sensing phase discharged by a high ('1') or low ('0') resistive cell state current and reference current respectively. Hence these two concepts are similar with respect to bitline voltage behavior. The main difference between the voltage and time domain sensing is in the post-processing of the bitline voltage signal. The voltage sensing compares the voltage difference between the bitline ($V_{'1'}$ or $V_{'0'}$) and the reference bitline at a given time T_{REF} by enabling the voltage sense amplifier (see section 3.3.1). For the time domain sensing not the absolute value of the bitline voltage but the time difference is evaluated for reaching the defined voltage threshold V_{sense} between the bitline ($T_{'0'}$ or $T_{'1'}$) and the reference bitline generating the reference time T_{REF} as shown in figure 4.1. This approach is more suitable for linear region biased cells compared to voltage sensing, since the bitline voltage change required for sensing operation can be adjusted by V_{sense}-level and does not depend on the discharge speed of the bitline with respect to the reference bitline (see section 3.4.2). For fast sensing time and low cell current dependance the V_{sense} must be set slightly below the precharge level V_{PRE}. Hence then the required bitline voltage change during the sensing phase is small leading to fast sensing time. This is also favorable for the linear region biased cell as the low bitline voltage swing ensures minimal cell current change during the sensing operation. The lower border for the voltage difference $V_{PRE} - V_{sense}$ is defined by the sense amplifier mismatch and overall

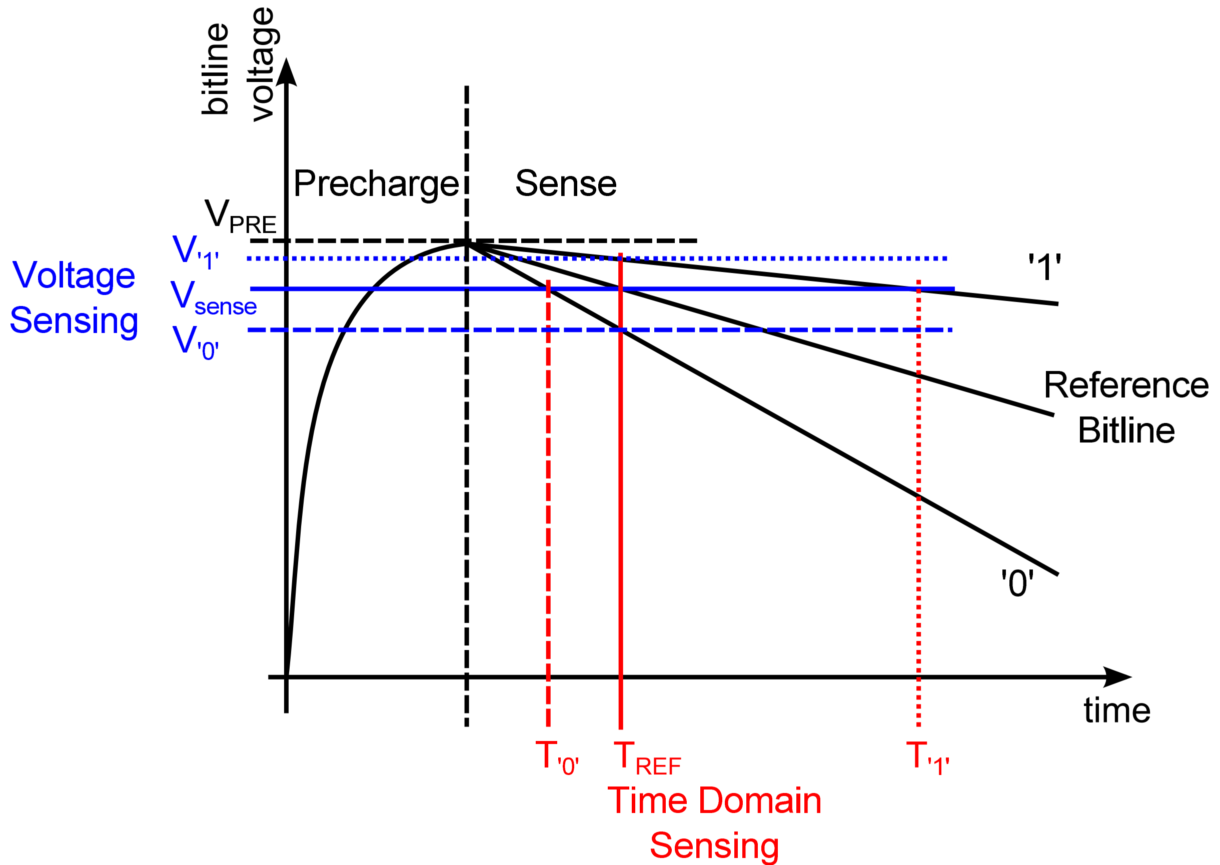

Figure 4.1: Time Domain vs. Voltage Sensing Basics

supply noise of the system.

The sense amplifier is active during the whole sensing phase and generates a self-timed digital output signal as soon as the bitline potential reach the V_{sense} voltage. Hence the conventional latch type sense amplifiers used for voltage sensing (see section 3.3.1) are not suitable for the time domain sensing. Instead a time to digital (TDC) converter is required to detect $T_{'0'}$ or $T_{'1'}$ with respect to T_{REF}.

Figure 4.2 illustrates the basic time domain sensing concept. This approach is very similar to the voltage sensing (see figure 3.1). During the precharge phase the bitlines, wordline and the single reference bitline per block of bitlines multiplexed to a common sense amplifier are charged. In the sense phase the cell current is integrated on the bitline capacitance performing the current to voltage conversion. The symmetry between memory and reference path provides very good power supply rejection ratio ($PSRR$) as it is the case for the voltage sensing (see figure 3.8). This approach is very similar to the the voltage sensing (see figure 3.1). During the precharge phase the bitline and wordline voltage are charged and in the sense phase the cell current is integrated on bitline capacitance performing the current to voltage conversion. The symmetry with reference bitline provides very good power supply rejection ratio ($PSRR$) as it is the case for the voltage sensing (see figure 3.8).

The reference information is now represented in time domain by a digital signal with

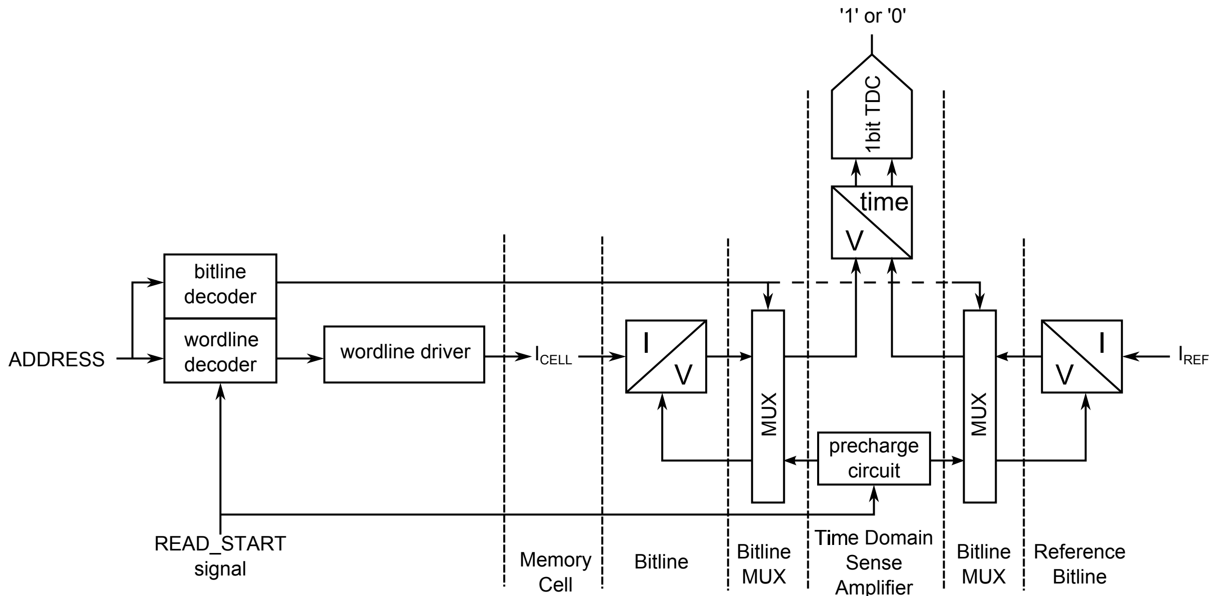

Figure 4.2: Time Domain Sensing Concept

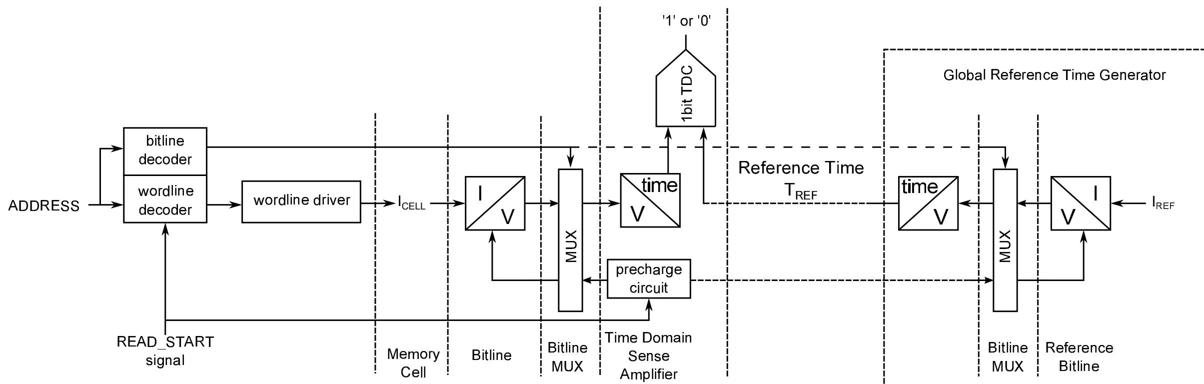

Figure 4.3: Time Domain Sensing Concept with Global Reference Time Generator

corresponding timing. The time domain coding is more robust and requires less power to distribute the reference signal compared to a voltage or current domain signal coding [31]. Hence, the reference signal can be distributed without any sophisticated circuitry and the global reference time generation can be used without sacrificing the symmetry with reference bitline as shown in figure 4.3. This approach also reduces significantly the power consumption during the precharge phase compared to voltage sensing, since only one reference bitline is charged instead of charging a reference bitline in every sense amplifier during the precharge phase.

The sense amplifier comprises the voltage to time conversion stage followed by an one bit *TDC*. It does not require a dedicated enable signal since the voltage to time and one bit time to digital conversion is self timed (see section 4.6) reducing the control overhead for the sensing circuit.

4.2 Voltage to Time Converter Circuits

4.2.1 Comparator Based Approach

For the voltage to time conversion a comparator circuit is required, comparing the bitline voltage V_{BL} to a V_{sense} potential (see figure 4.1) generating a digital output signal as soon as V_{BL} reaches V_{sense}. The conventional latch type sense amplifiers (see figure 3.8) are not suited for the voltage to time conversion since they evaluate the voltage difference at the input at a predefined time point triggered by the corresponding enable signal. To continuously sense the input voltage difference during the whole sensing phase the static voltage comparator circuit is required. Figure 4.4 shows the conventional comparator circuit realized as a single ended OTA [32], [33], [34]. The dashed circuitry can be added to improve gain and speed of the comparator but it requires additional power and area and is usually omitted due to power and area constraints [29]. The sensing speed of the

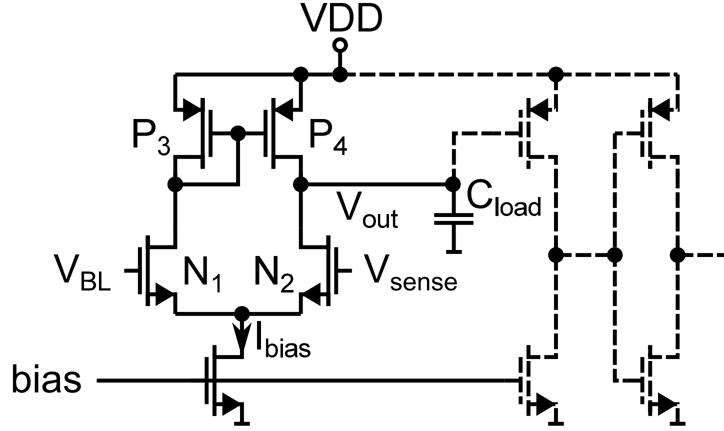

Figure 4.4: Conventional Comparator Circuit for Time Domain Sensing

non-dashed circuit heavily depends on operation conditions and input signal range [32], [33], [34]. To predict the comparator speed the large- and small-signal delay has to be considered [32], [33], [34]. The small-signal time constant of the comparator circuit is defined by it's bandwidth which is determined by the time constant of the dominant pole:

$$ss, \tau_{COMP} = C_{load} r_{out} = \frac{C_{load}}{(\lambda_{n,N2} + \lambda_{p,P4}) \frac{I_{BIAS}}{2}} \tag{4.1}$$

where C_{load} is the output load capacitance, λ_n and λ_p is the small-signal channel-length modulation coefficient of NMOS and PMOS transistors N_2 and P_4 respectively. For a bias current I_{BIAS}=10μA, λ_p=λ_n=0.4V^{-1} and a load capacitance C_{load} of 5fF the small-signal time constant ss, τ_{COMP} is about 1.3ns. In case a large input signal step is applied the comparator circuit is slewing and the large signal operation has to be considered [32], [33], [34]. The large-signal delay of the comparator circuit is calculated as:

$$ls, \tau_{COMP} = \frac{V_{outmax} - V_{outmin}}{SR} = \frac{(V_{outmax} - V_{outmin})\, C_{load}}{I_{BIAS}} \tag{4.2}$$

where SR is the slew rate of the comparator circuit. For the supply voltage of V_{DD}=1.2V and the values mentioned above the large-signal delay is about 0.6ns. However in most cases the comparator output signal is post-processed by a digital circuit which has a threshold level of a digital inverter (about half of the supply voltage). Since the V_{outmax}-$V_{outmin} \approx \frac{V_{DD}}{2}$, ls, τ_{COMP} is reduced to 0.3ns. If an input signal voltage step is applied to a comparator, the circuit changes it's operation region from large-signal to small-signal [33]. The large-signal behavior dominates first. After the output voltage is close to the desired output value the small-signal operation takes over. Hence the minimal delay (time constant) of the comparator is the small-signal delay:

$$\tau, min_{COMP} = ss, \tau_{COMP} \tag{4.3}$$

and the maximum delay can be approximated as:

$$\tau, max_{COMP} = ss, \tau_{COMP} + ls, \tau_{COMP} \tag{4.4}$$

Hoverer the ss, τ_{COMP} considers only the delay required to reach about 63.5% of the desired output value, to reach more than 95% of the desired output voltage 3xss, τ_{COMP} are needed. For the typical values $\tau, max_{COMP} = 1.3ns + 0.3ns = 1.6ns$. For further considerations only the maximum comparator delay τ, max_{COMP} is used, as it represents the worst case for the access time.

The drawback of the comparator circuit is the static bias current I_{BIAS} drawn during the whole sense operation. A further disadvantage is the significantly higher sensing delay compared to a latch type sense amplifier (see figure 3.8), which exhibits a sense delay of only several tens of picoseconds [6], [24].

The sense phase of the comparator based time domain sensing is determined by the discharge speed of the bitline (see figure 4.1) and the delay of the comparator:

$$T_{senseCOMP,SAT} = \frac{C_{BL}\left(V_{PRE} - V_{sense}\right)}{I_{CELL}} + \tau, max_{COMP} \tag{4.5}$$

For typical values (C_{BL}=1pF, V_{PRE}=V_{DD}=1.2V, V_{sense}=1.1V and I_{CELL}=20μA), $T_{senseCOMP,SAT}$ is 6.6ns in case of a cell biased in saturation. For cell biased in linear region the sense delay can be calculated as:

$$T_{senseCOMP,LIN} = -ln\left(\frac{V_{sense}}{V_{PRE}}\right) R_{CELL}C_{BL} + \tau, max_{COMP} \tag{4.6}$$

yielding ≈6.8ns for $T_{senseCOMP,LIN}$ using typical values and R_{CELL}=$\frac{1.2V}{20\mu A}$=60$k\Omega$. The difference between $T_{senseCOMP,SAT}$ and $T_{senseCOMP,LIN}$ is rather small (below 5%), since the bitline voltage change required for sensing operation does not depend on the discharge speed of the bitline with respect to the reference bitline, which makes the time domain sensing suitable for linear region biased cells (see section 4.1). Hence, in the following there will be no distinction between a cell biased in saturation or in linear region.

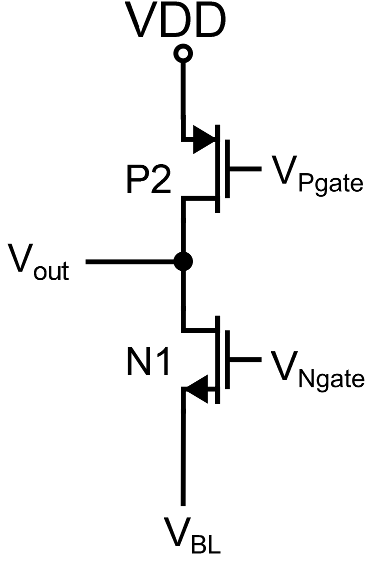

Figure 4.5: Charge Transfer Sense Amplifier

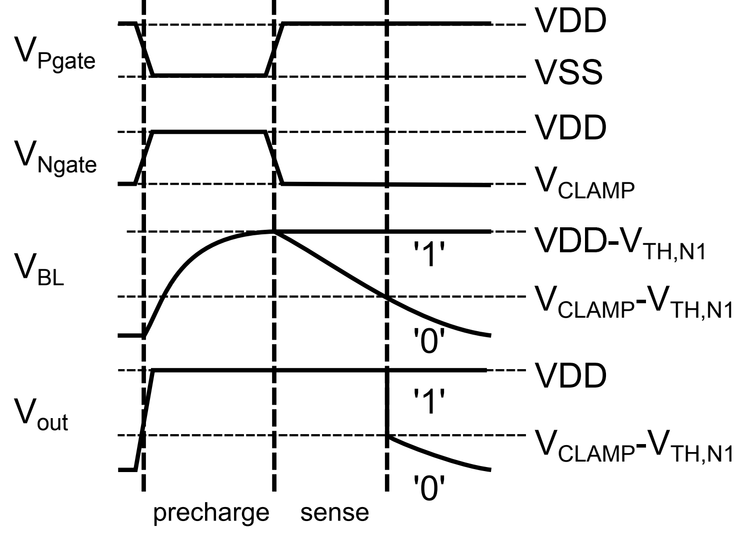

Figure 4.6: Timing Diagram of Figure 4.5

4.2.2 Charge Transfer Sense Amplifier

The voltage sensing method as used in standalone NAND-Flash memories deploys the common gate structure as shown in figure 4.5 to obtain a bitline voltage gain by charge transfer from the parasitic capacitance of the output node V_{out} to the bitline capacitance C_{BL} [35], [36], [37], [38], [23]. The corresponding timing diagram is shown in figure 4.6. In the precharge phase the precharge signal V_{Pgate} goes low and charges the output node V_{out} to V_{DD}. The V_{Ngate} voltage is set e.g. to V_{DD} and charges the bitline to about V_{DD}-$V_{TH,N1}$, where $V_{TH,N1}$ is the threshold voltage of the N1. In the subsequent sensing phase V_{Pgate} is set to V_{DD} and the V_{Ngate} voltage is set to V_{CLAMP} potential. The bitline V_{BL} is discharged by the cell current in case the low resistive cell ('0' cell state) is accessed. As soon as the bitline voltage V_{BL} reaches V_{CLAMP}-$V_{TH,N1}$, N1 is conducting leading to a charge sharing between the V_{out} and V_{BL} nodes. The V_{out} is discharged to V_{BL} since the parasitic capacitance at V_{out} is much lower compared to the bitline capacitance. In case the high resistive cell state is accessed the V_{BL} remains at precharged potential and V_{out} equals V_{DD} (V_{out}=V_{DD}).

This approach has no static current flow from V_{DD} to V_{SS} in the circuit during the sense phase. There is still power dissipated due to capacitance discharge to V_{SS} but the power supply domain does not need to provide any current during this phase. However this approach limits the maximum voltage headroom to

$$V_{BLmax,CT} = V_{DD} - V_{TH,N1} \tag{4.7}$$

which is not suitable for low voltage applications. In addition the output V_{out} has no rail to rail voltage swing which requires an additional amplification stage consuming additional power and area. The voltage change at the output of the charge transfer sense amplifier

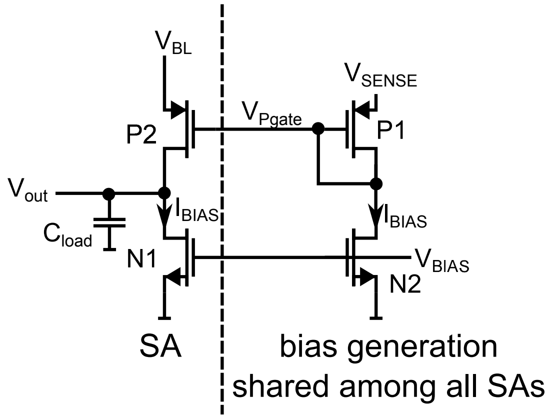

Figure 4.7: Drain-Side Common Gate Sense Amplifier

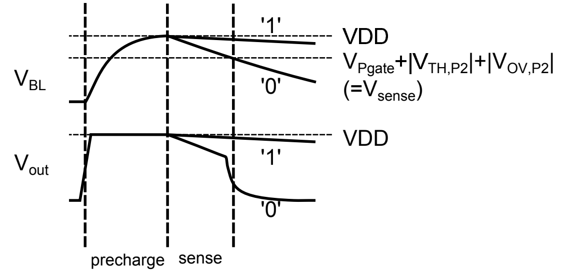

Figure 4.8: Timing Diagram of Figure 4.7

A_{CT} is defined by the voltage difference:

$$A_{CT} = V_{PRE} - (V_{CLAMP} - V_{TH,N1}) \tag{4.8}$$

To achieve significant output swing the $V_{CLAMP} - V_{TH,N1}$ potential has to be set slightly above V_{SS}. Low V_{CLAMP} voltage leads to a low bitline voltage during the sense operation diminishing the cell current and therefore the sensing speed.

4.2.3 Common Gate Sense Amplifier

By folding the common gate structure of the charge transfer sense amplifier (see figure 4.5) and adding an active bias current source as shown in figure 4.7 the drain-side common gate sense amplifier is realized. It allows the maximum bitline voltage headroom V_{BLmax}=V_{DD} required for low voltage applications and offers rail to rail output voltage swing while maintaining the low current consumption during the sense phase. The corresponding bias generation can be realized by two transistors P1 and N2, which are fully matched to transistors P2 and N1. The P1 is in diode configuration and is connected with its source to the V_{sense} voltage and with its drain to the transistor N2. The transistor N2 acts like a current source providing the current I_{BIAS}.

Figure 4.8 shows the corresponding timing diagram for the drain-side common gate sense amplifier. In the precharge phase the bitline voltage V_{BL} and V_{out} are charged to V_{DD} similar to the charge transfer approach. In the subsequent sensing phase the BL is discharged by the cell current and the bias current I_{BIAS} in parallel. This leads to a bitline discharge by I_{BIAS}<<I_{CELL} even in case the non conductive cell ('1' cell state) is accessed as shown in figure 4.8. For proper circuit operation the width to length ratio of P2 must be high compared to N1 and the bias current I_{BIAS} has to be in a range of 1μA. P2 would operate in weak inversion for the given bias current introducing a sharp threshold for its operation region at the V_{sense} voltage:

$$V_{sense} = V_{Pgate} + |V_{TH,P1}| + |V_{OV,P1}| = V_{Pgate} + |V_{TH,P2}| + |V_{OV,P2}| \tag{4.9}$$

Therefore V_{out} will follow the bitline voltage V_{BL} for $V_{BL} > V_{sense}$, because P2 operates in the linear region, as its absolute overdrive voltage value is higher compared to one of P1 which is in saturation. And as soon as V_{BL} reaches the sense voltage V_{sense}, P2 enters the saturation region and the circuit P2, N1 operates as a common gate amplifier. The further bitline voltage change is amplified to:

$$\begin{aligned} V_{out} &= (V_{sense} - V_{BL})\, g_{m,P2} \cdot r_{out} \\ &= (V_{sense} - V_{BL})\, g_{m,P2} \cdot (r_{ds,P2} || r_{ds,N1}) \\ &= (V_{Pgate} + |V_{TH,P2}| + |V_{OV,P2}| - V_{BL})\, \frac{2I_{BIAS}}{|V_{OV,P2}|} \cdot \frac{1}{(\lambda_{n,N1} + \lambda_{p,P2})\, I_{BIAS}} \\ &= (V_{Pgate} + |V_{TH,P2}| - V_{BL})\, \frac{2}{|V_{OV,P2}|\,(\lambda_{n,N1} + \lambda_{p,P2})} \end{aligned} \tag{4.10}$$

The main difference of this approach compared to other sense amplifier types is the additional active bias current I_{BIAS} in parallel to the cell current discharging the bitline capacitance during the sense phase. This current reduces the relative cell current window Rel_{Iwin}:

$$Rel_{Iwin} = \frac{I_{BL'0'} - I_{BL'1'}}{I_{BL'0'}} \tag{4.11}$$

where the $I_{BL'0'}$ and $I_{BL'1'}$ are the bitline currents for the '0'- and '1'-cell state access respectively. For the conventional sense amplifier, which is high ohmic during the bitline discharge, and typical cell currents for eFlash cell ($I_{CELL'0'} \approx 20\mu A$ and $I_{CELL'1'} \approx 0\mu A$) the Rel_{Iwin} is at it's maximum of 100%. For the common gate sense amplifier with only $1\mu A$ bias current ($I_{BIAS} = 1\mu A$) the Rel_{Iwin} is already reduced to

$$\frac{21\mu A - 1\mu A}{21\mu A} \approx 95\% \tag{4.12}$$

For higher I_{BIAS} values the Rel_{Iwin} is significantly lowered requiring higher sense amplifier accuracy to resolve the effective cell current window. Therefore I_{BIAS} must be set as low as possible to reduce its impact on the cell current window. The lower bound for the bias current is given by the sense amplifier speed, which is defined by the parasitic capacitance of the V_{out} node. The delay of the common gate amplifier formed by P2 and N1 is calculated in the same way as for the comparator circuit (see figure 4.4). The small-signal portion of the common gate drain-side sense amplifier (CGDraSiSA) is calculated as:

$$ss, \tau_{CGDraSiSA} = C_{load} r_{out} = \frac{C_{load}}{(\lambda_{n,N1} + \lambda_{p,P2})\, I_{BIAS}} \tag{4.13}$$

The small-signal delay is by a factor of two lower compared to a comparator circuit (see equation 4.1), because the I_{BIAS} current is now contributing by its full magnitude since the common gate circuit (N1 and P2) is a single input amplifier. The large-signal delay is similar to a comparator circuit and is described by:

$$ls, \tau_{CGDraSiSA} = \frac{V_{sense} - V_{outmin}}{SR} = \frac{(V_{sense} - V_{outmin})\, C_{load}}{I_{BIAS}} \tag{4.14}$$

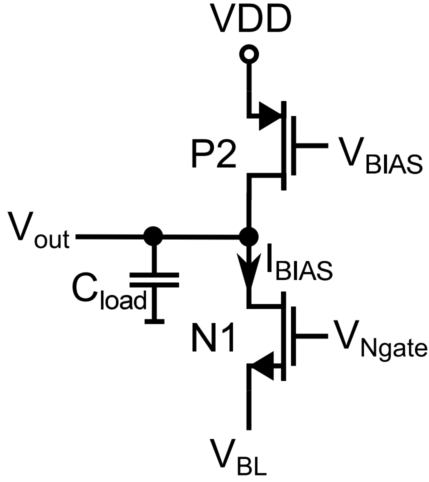

Figure 4.9: Source-Side Common Gate Sense Amplifier

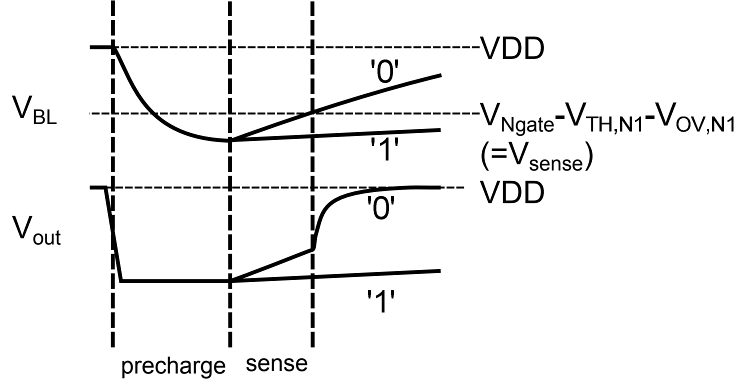

Figure 4.10: Timing Diagram of Figure 4.9

In case the output node V_{out} (see figure 4.7) serves as an input for a digital circuit, V_{outmin} is the inverter threshold voltage. The worst case delay for the common gate circuit is the sum of the large- and small-signal portion:

$$\tau, max_{CGDraSiSA} = ls, \tau_{CGDraSiSA} + ss, \tau_{CGDraSiSA} \tag{4.15}$$

To assess the overall sense delay of this sensing approach the additional discharge current during the sense phase has to be considered. The sense delay of the common gate circuit can be calculated as:

$$T_{senseCGDraSiSA} = \frac{C_{BL}\left(V_{PRE} - V_{sense}\right)}{I_{CELL} + I_{BIAS}} + \tau, max_{CGDraSiSA} \tag{4.16}$$

If the source-side sensing is required (see chapter 3.2) for the specific eFlash cell structure the common gate sense amplifier as shown in figure 4.9 can be utilized. Here the source-side of the memory cell transistor is connected to the bitline BL. The corresponding timing diagram is shown in figure 4.10 and 4.23. The bitline BL and V_{out} are precharged to V_{SS} during the precharge phase. In the subsequent sensing phase BL is charged by the cell current towards V_{DD}. The V_{out} node first follows the bitline potential since N1 is in linear region, but as soon as V_{BL} reaches $V_{Ngate} - V_{TH,N1} - V_{OV,N1}$ N1 enters the saturation region and the circuit N1 and P2 operates as an common gate amplifier. The relatively small parasitic capacitance of the V_{out} node compared to the bitline capacitance is rapidly charged by the I_{BIAS} current to V_{DD}. The V_{sense} voltage is now defined as:

$$V_{sense} = V_{Ngate} - V_{TH,N1} - V_{OV,N1} \tag{4.17}$$

The delay of the sense circuit can be calculated in the same way as for the drain-side scheme (see equations 4.13 - 4.16). The small-signal delay portion is equal to the drain-side approach:

$$ss, \tau_{CGSoSiSA} = ss, \tau_{CGDraSiSA} = C_{load} r_{out} = \frac{C_{load}}{\left(\lambda_{n,N1} + \lambda_{p,P2}\right) I_{BIAS}} \tag{4.18}$$

For the large-signal delay the equation 4.14 has to be changed to:

$$ls, \tau_{CGSoSiSA} = \frac{V_{outmax} - V_{sense}}{SR} = \frac{(V_{outmax} - V_{sense})\, C_{load}}{I_{BIAS}} \tag{4.19}$$

Since the output voltage is charged from V_{sense} towards V_{DD}. The worst case delay of the common gate source-side sense amplifier is the sum of the two portions:

$$\tau, max_{CGSoSiSA} = ls, \tau_{CGSoSiSA} + ss, \tau_{CGSoSiSA} \tag{4.20}$$

The overall sense delay is similar to equation 4.16:

$$T_{senseCGsoSiSA} = \frac{C_{BL} V_{sense}}{I_{CELL} + I_{BIAS}} + \tau, max_{CGSoSiSA} \tag{4.21}$$

For typical values ($C_{load} = 5fF$, $C_{BL} = 1pF$, I_{CELL}=20μA, $V_{PRE}-V_{sense} = |100mV|$, $V_{DD} = 1.2V$, $\lambda_p = \lambda_n = 0.4V^{-1}$) and the bias current $I_{BIAS} = 1\mu A$ the sense delay of the common gate amplifier is $T_{senseCGDraSiSA} = T_{senseCGsoSiSA} \approx 13.6ns$, which is about a factor of two higher compared to the conventional comparator circuit (6.6ns). This is due to the relatively large delay of 8.8ns for the common gate circuit (1.6ns for the comparator delay). By increasing the bias current to $5\mu A$ the common gate sense amplifier delay is reduced to $1.8ns$ and the overall sense delay is now $5.8ns$, which is more than 10% lower compared to the conventional approach with low current consumption in the sense phase.

As already mentioned the drawback of the common gate sense amplifier compared to the state-of-the-art (comparator based sense amplifier) is the reduced effective cell current window, diminishing the robustness of the sensing scheme if the cell current difference between high and low resistive cell state is low. For $5\mu A$ bias current the relative cell current window Rel_{Iwin} is reduced to 80%, which can be insufficient for a robust operation required in automotive applications.

But the main advantages are the low current consumption during the sense phase and very simple circuit structure (only two transistors) requiring less area. In addition the access time can be even reduced below the lower limit for the state-of-the-art sensing approaches, as the sense amplifier is not high ohmic during the sense phase but is actively discharging (or charging for the source-side sensing) the bitline capacitance by it's bias current.

4.3 Slope Detection Circuit

As described in the previous section the drawback of the common gate sense amplifier is the cell current window reduction (see equation 4.11) due to static bias current I_{BIAS} (see figure 4.7, 4.9). This limits the maximum value for I_{BIAS} which determines the sensing speed. The influence on the cell current window is getting more important if the current window is lowered. E.g. for $I_{CELL'0'}$=10μA and $I_{CELL'1'}$=0μA and typical values:

- $C_{load} = 5fF$

- $C_{BL} = 1pF$
- $V_{PRE} - V_{sense} = |100mV|$
- $V_{DD} = 1.2V$
- $\lambda_p = \lambda_n = 0.4V^{-1}$

and $I_{BIAS} = 5\mu A$ the sense delay time $T_{senseCGDraSiSA} = T_{senseCGsoSiSA}$ is reduced to 8.4ns which is more than 25% faster compared to the state-of-the-art approach ($T_{senseCOMP,SAT}$ ≈11.6ns) and is even 16% lower compared to the lower physical limit of the state-of-the-art sensing which is not considering any comparator delay:

$$T_{limit} = \frac{C_{BL} V_{sense}}{I_{CELL}} \tag{4.22}$$

But the relative cell current window is only $Rel_{Iwin} \approx 66.7\%$. This rather low value is not acceptable for automotive conditions to support reliable read operation.

To enable reliable read operation, the bias current I_{BIAS} must be changed adaptively depending on the cell current. For the access on the low resistive state cell the I_{BIAS} must be at its maximum value $I_{BIAS,max}$, and for high resistive cell the I_{BIAS} must be at its minimum value $I_{BIAS,min}$, e.g. for $I_{BIAS,max} = 5\mu A$ and $I_{BIAS,min} = 0\mu A$ the relative cell current window for the previous example is expanded from 66.7% to its full value of 100% without reducing the read access time.

To obtain the information for the I_{BIAS} adaptation an additional input parameter is required. Therefore within this work, an additional function block was developed. The conventional voltage sensing only processes the information about the voltage level. But besides the voltage level the bitline voltage also changes with respect to time, which is the voltage slope. Figure 4.11 and 4.12 show the common gate sense amplifier for conventional voltage sensing which is enhanced by a slope detection unit. The slope detection unit senses the bitline slope and adjusts I_{BIAS} current by regulating the V_{BIAS} node.

To generate the current from the voltage the circuit shown in figure 4.13 can be used. The input voltage V_{IN} is applied to the gate of N2 which is connected with the source terminal to the capacitance C_F. The drain terminal of N2 is connected to the PMOS P1 which is in diode connection forming an active load for N2. The input voltage is propagated to the capacitance C_F as N2 is in source follower configuration. For proper function of the source follower the input potential must remain higher than the threshold voltage of N2 ($V_{TH,N2}$). A fast V_{IN} voltage ramp will charge up C_F rapidly requiring higher current I_{CF} compared to a slow voltage ramp at the input. To calculate the current I_{CF} of this slope dependent current generator the equivalent circuit for small-signal operation shown in figure 4.14 is considered. Since the voltage ramp applied to the capacitance C_F determines the circuit behavior the function for i_{CF} can be calculated by using Laplace transform for $V_{IN} > V_{TH,N2} - V_{CF}$ (N2 is conducting). The small signal transfer function for i_{CF} is

$$i_{CF}(s) = s v_{in}(s)\, g_{m,N2} \left(\frac{g_{m,N2}}{C_F} + s \right)^{-1} \tag{4.23}$$

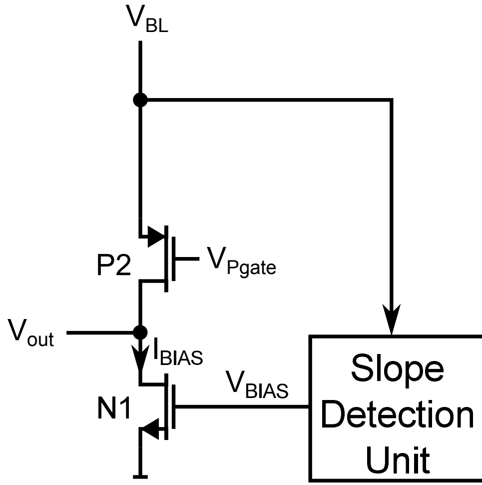

Figure 4.11: Drain-Side Common Gate Sense Amplifier Enhanced by Slope Detection Unit

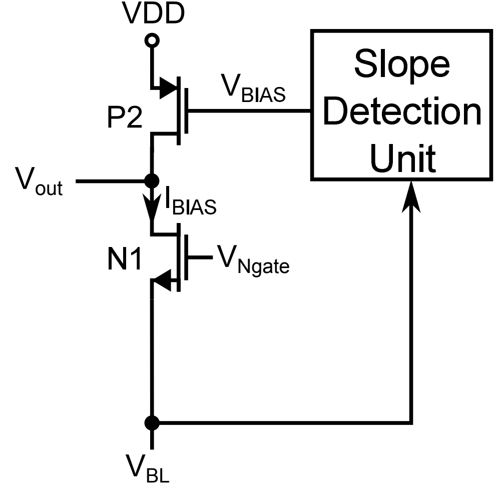

Figure 4.12: Source-Side Common Gate Sense Amplifier Enhanced by Slope Detection Unit

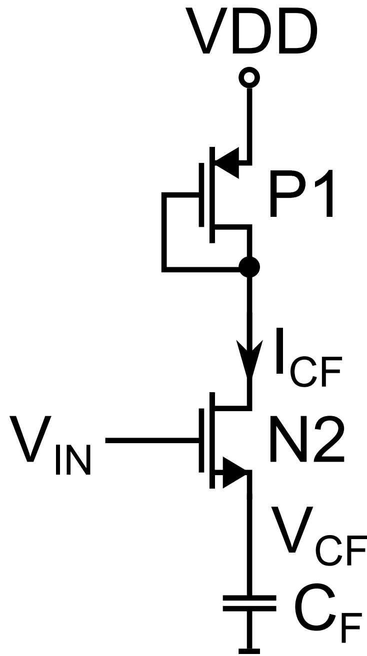

Figure 4.13: Slope Dependent Current Source

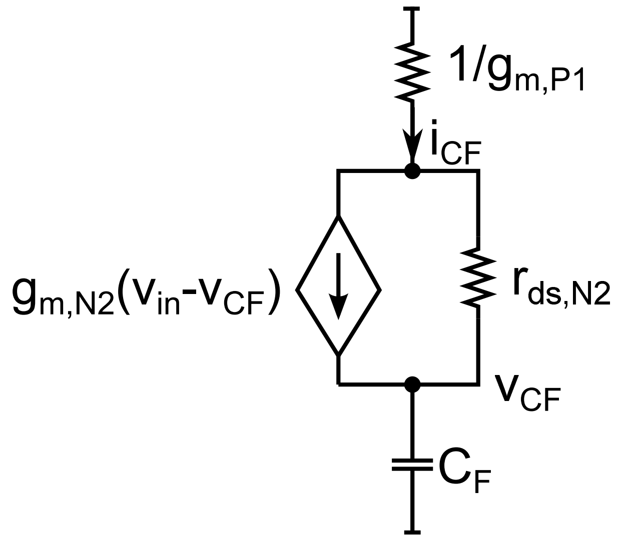

Figure 4.14: Equivalent Circuit for Figure 4.13 for Small-Signal Operation

(for more details see Appendix B). The transfer function has one pole which has the time constant $\tau = \frac{g_{m,N2}}{C_F}$ and is the source follower frequency response [32], [33], [34]. To get more insight into the circuit operation a small signal input voltage ramp is applied at the input:

$$V_{IN}(t) = t \cdot S \tag{4.24}$$

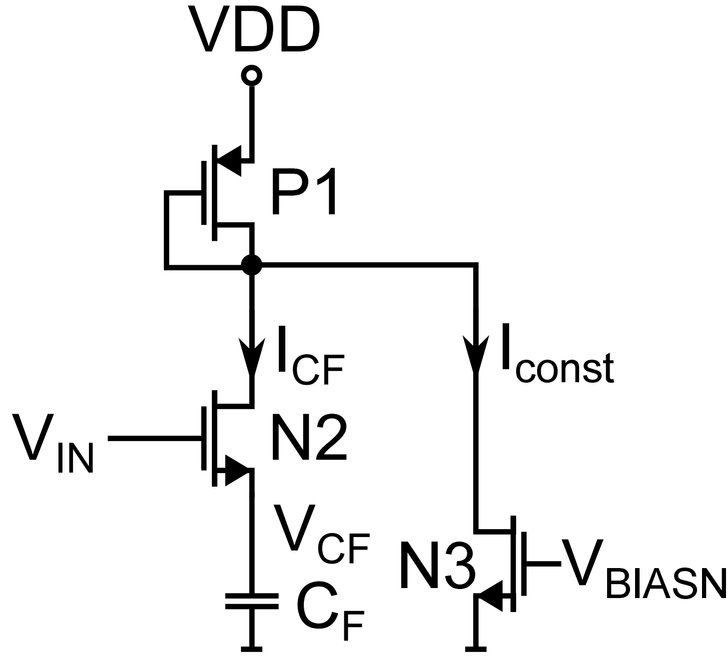

Figure 4.15: Slope Dependent Current Source with Biasing

where S is the magnitude of the voltage slope in volt per second. The corresponding Laplace transform for $V_{IN}(t)$ is:

$$V_{IN}(s) = \frac{1}{s^2}S \quad (4.25)$$

Hence the small-signal output current can be calculated by the reverse Laplace transform:

$$i_{CF}(t) = S \cdot C_F \left(1 - e^{-\frac{g_{m,N2}}{C_F}t}\right) \quad (4.26)$$

To calculate the overall current behavior of the circuit the initial condition for the current $i_{CF}(t=0) = i_{CF0}$ has to be taken into account, which is defined by the initial voltage condition v_{CF0} of the capacitance C_F. If the V_{IN} voltage will remain constant i_{CF0} will decrease to zero (C_F is charged to $V_{IN} - V_{TH,N2}$) with the time constant $\tau = \frac{g_{m,N2}}{C_F}$. Therefore the overall small-signal current transfer can be calculated as:

$$i_{CF}(t) = S \cdot C_F \left(1 - e^{-\frac{g_{m,N2}}{C_F}t}\right) + i_{CF0}e^{-\frac{g_{m,N2}}{C_F}t} \quad (4.27)$$

The current $i_{CF}(t)$ is proportional to the slope of the input voltage ramp. However there is the second portion ($i_{CF0}e^{-\frac{g_{m,N2}}{C_F}t}$) which is defined by the initial conditions and can be adjusted to a desired value by design. This factor dominates the behavior of the circuit at the beginning.

To improve the transient behavior the additional current sink is added as shown in figure 4.15. This current sink (transistor N3) adjusts the biasing conditions of P1 in case

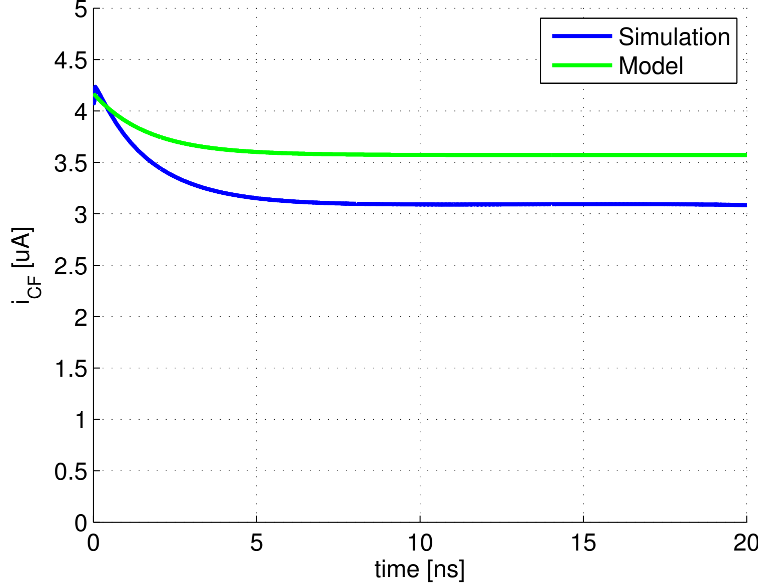

Figure 4.16: Simulation vs. Model for Slope Dependent Current Source with Biasing

the input voltage V_{IN} is not rising and i_{CF} is zero. In this case the current sink still provides I_{CONST} forcing the gate to source voltage of P1 to be around its threshold. By this means P1, which will be later used as a part of a current mirror (see figure 4.22), has faster response to input voltage ramp.

The circuit shown in figure 4.15 is simulated in a 65nm technology with $g_{m,N2} = 75\mu A/V$, $I_{CONST} = 1\mu A$, $i_{CF0} \approx 4.2\mu A$, $v_{CF0} = 0V$ and $C_F = 125fF$. Figure 4.16 shows the calculated (by equation 4.27) and simulated current response $i_{CF}(t)$ for the input voltage ramp $V_{IN}(t) = t \cdot S$ with $S = \frac{10\mu A}{350fF}$. The simulation confirms the desired operation of the slope dependent current source 4.15 generating a constant current i_{CF} for a constant voltage slope applied at the input. The simulation results in a lower current $i_{CF}(t)$ as predicted by equation 4.27. This is due to the fact that the body-effect of N2 (see figure 4.13) was not taken into account so far. Figure 4.17 shows the corresponding equivalent small-signal circuit considering the body-effect of N2. The overall current transfer function of equation 4.23 is now updated to:

$$i_{CF}(s) = sv_{in}(s)\, g_{m,N2} \left(\frac{g_{m,N2} + g_{mb,N2}}{C_F} + s \right)^{-1} \tag{4.28}$$

And the corresponding time response for the voltage ramp at the input becomes

$$i_{CF}(t) = SC_F \frac{g_{m,N2}}{g_{m,N2} + g_{mb,N2}} \left(1 - e^{-\frac{g_{m,N2}+g_{mb,N2}}{C_F}t} \right) + i_{CF0} e^{-\frac{g_{m,N2}+g_{mb,N2}}{C_F}t} \tag{4.29}$$

The system's time constant has now changed to $\tau = \frac{g_{m,N2}+g_{mb,N2}}{C_F}$ and the effective capacitance C_F is now reduced by the factor $\frac{g_{m,N2}}{g_{m,N2}+g_{mb,N2}}$, where $g_{mb,N2}$ is the small-signal transconductance for the bulk-source potential. Figure 4.18 confirms the $g_{mb,N2}$ dependence of the circuit, by taking the $g_{mb,N2} \approx 12\mu A/V$ into account.

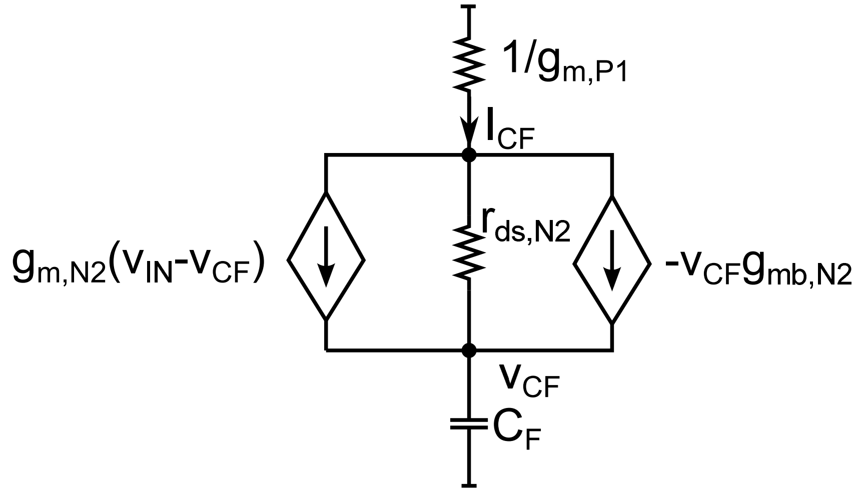

Figure 4.17: Equivalent Circuit for Figure 4.13 for Small-Signal Operation Considering Body-Effect of N2

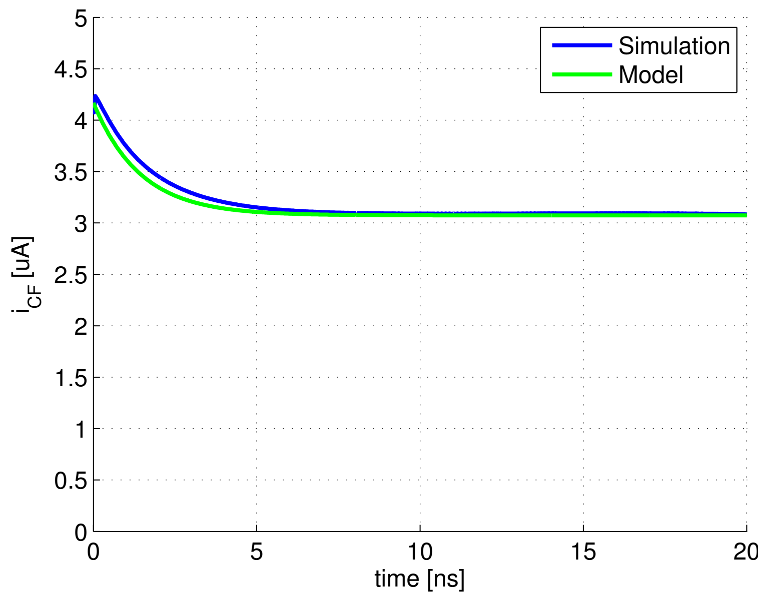

Figure 4.18: Simulation vs. Model for Slope Dependent Current Source with Biasing Considering Body-Effect

The circuit as shown in figure 4.13 requires a positive voltage ramp, which is required to improve the source-side sense amplifier (see figure 4.12). However the input voltage level has to be higher than the threshold voltage of N2 for proper circuit operation. But the sensing operation requires the circuit to operate for zero volt input, because the bitline capacitance is discharged to V_{SS} during the precharge phase and charged towards V_{DD} during the sense phase (see figure 4.10). To support an input voltage close to V_{SS} a level-shifter stage has to be added to the slope detection unit, as shown in figure 4.19. The input voltage V_{IN} is shifted by transistor P2 to $V_{IN} + |V_{TH,P2}| + |V_{OV,P2}|$, where

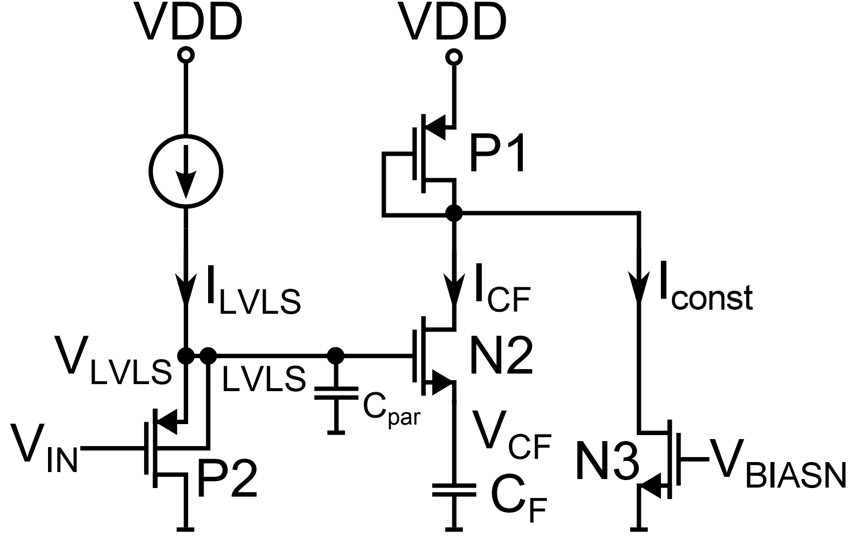

Figure 4.19: Slope Detection Circuit for Source-Side Sensing

$|V_{TH,P2}|$ and $|V_{OV,P2}|$ is the threshold voltage and overdrive voltage of P2 respectively. To overcome the body-effect of the transistor P2, its bulk and source are connected together. The level shifter, which is a source follower stage [32], [33], [34], introduces the second pole to the current transfer function of i_{CF} by its time constant $\tau_{LVLS} = \frac{g_{m,P2}}{C_{par}}$ (C_{par} is the overall capacitance of the LVLS node dominated by a parasitic capacitance):

$$i_{CF}\left(s\right) = sv_{in}\left(s\right) g_{m,N2} \left(\frac{g_{m,N2} + g_{mb,N2}}{C_F} + s\right)^{-1} \frac{g_{m,P2}}{C_{par}} \left(\frac{g_{m,P2}}{C_{par}} + s\right)^{-1} \tag{4.30}$$

The time response for the voltage ramp at the input has now changed to:

$$\begin{aligned} i_{CF}\left(t\right) &= S\frac{g_{m,N2}}{g_{m*,N2}}C_F + S\frac{g_{m,N2}g_{m,P2}C_F}{g_{m*,N2}C_{par} - g_{m,P2}C_F}\left(\frac{C_F}{g_{m*,N2}}e^{-\frac{g_{m*,N2}}{C_F}t} - \frac{C_{par}}{g_{m,P2}}e^{-\frac{g_{m,P2}}{C_{par}}t}\right) \\ &+ i_{CF0}e^{-\frac{g_{m*,N2}}{C_F}t} \end{aligned} \tag{4.31}$$

$$g_{m*,N2} = g_{m,N2} + g_{mb,N2} \tag{4.32}$$

The circuit shown in figure 4.19 was designed in 65nm technology with $g_{m,P2} \approx 82\mu A/V$, the other design parameters were the same as for the slope dependent current source with biasing (see figure 4.15). Figure 4.20 compares the simulated behavior for the slope dependent current generator with and without level-shifter. The simulation reveals a low impact of the level-shifter on circuit operation in current design, due to low parasitic capacitance on LVLS node. The effect of gate-source and gate-channel capacitance of P2 and N2 is very low due to source-follower configuration [32], [33], [34]. To assess the impact of the level-shifter on circuit operation additional, 50fF were added to the LVLS node. Figure 4.20 shows the simulation results for circuit 4.19 with additional 50fF on LVLS node and the current response predicted by equation 4.31 with $C_{par} = 50fF$. Equation 4.31, which is based on level-1 transistor parameters gives already a very

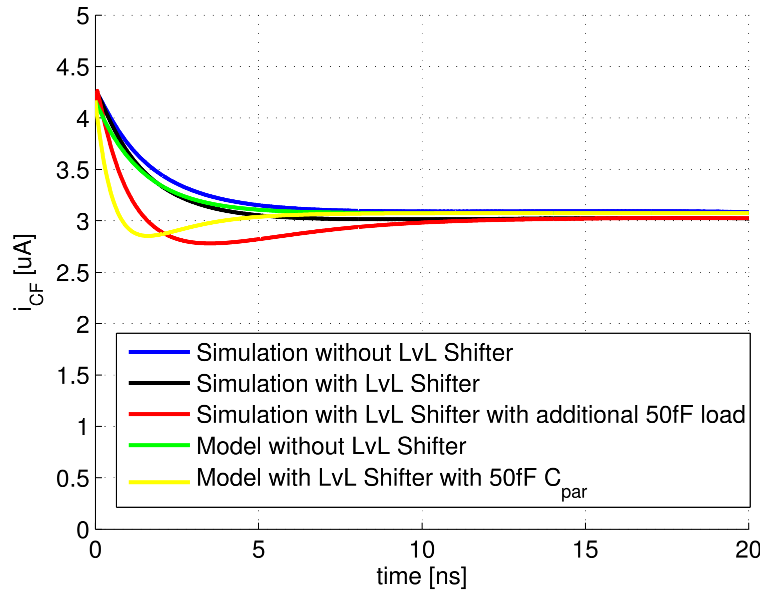

Figure 4.20: Simulation vs. Model for Slope Detection Circuit for Source-Side Sensing

good approximation of the circuit behavior. The parasitic capacitance of the LVLS node introduced by layout should be kept as small as possible to guarantee the desired circuit operation for a predefined level-shifter current I_{LVLS}.

In case a drain-side sensing is required the slope detection circuit accordingly figure 4.21 can be implemented. The corresponding equation for the slope dependent current

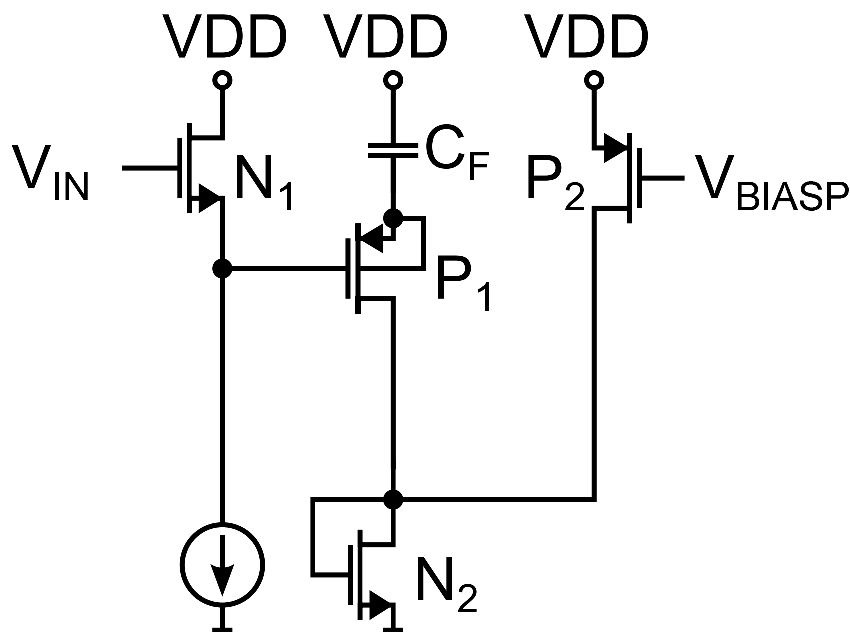

Figure 4.21: Slope Detection Circuit for Drain-Side Sensing

can be obtained by the same way as in the source-side case.

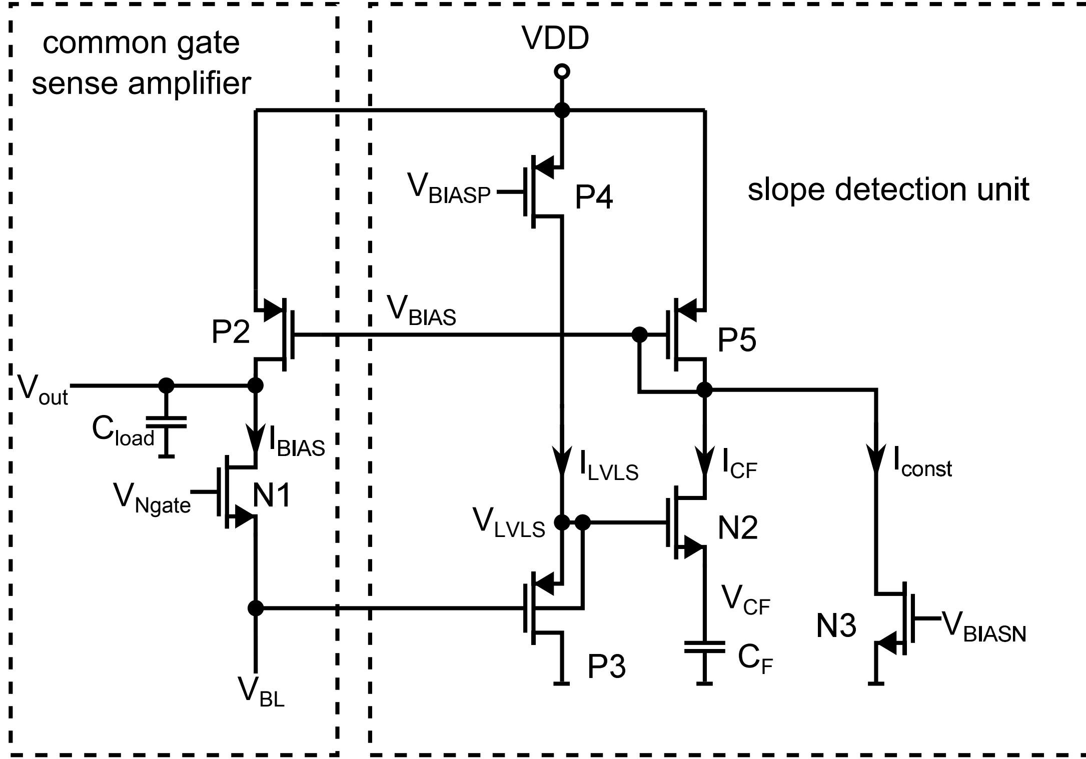

Figure 4.22: Bitline-Capacitance-Cancellation Sensing Circuit for Source-Side Sensing

4.4 Bitline-Capacitance-Cancellation Sense Amplifier Circuit

By combining the slope detection circuit 4.21 or 4.19 discussed in the previous section with the common gate circuits 4.7 and 4.9 respectively, the bitline-capacitance-cancellation circuit is created. Figure 4.22 shows the bitline-capacitance-cancellation sensing circuit for source-side sensing. The level-shifter bias current is now generated by P4, acting as a constant current source. Transistors P5 and P2 form a 1:1 current mirror and the bias current I_{BIAS} for the common gate circuit is defined by:

$$I_{BIAS} = I_{CF} + I_{CONST} \tag{4.33}$$

I_{CONST} is a constant current drawn by N3 added to I_{CF}, which is the slope depended current portion. To reduce the impact of the constant bias current portion on the relative cell current window the I_{CONST} has to be kept below $1\mu A$ (see equation 4.12).

Figure 4.23 shows the transient simulation of the bitline-capacitance-cancellation circuit for source-side sensing. The sensing period comprises two phases. During the first sensing phase the transistor N1 is in linear region and the V_{out} node follows V_{BL}. In the second sensing phase V_{BL} reaches V_{sense} potential and N1 enters the saturation region. The V_{BL} change is now amplified by $g_m r_{ds}$ on the V_{out} node.

To better understand the operation principle of the common gate sense amplifier enhanced by the slope detection unit the transfer functions have to be considered. For

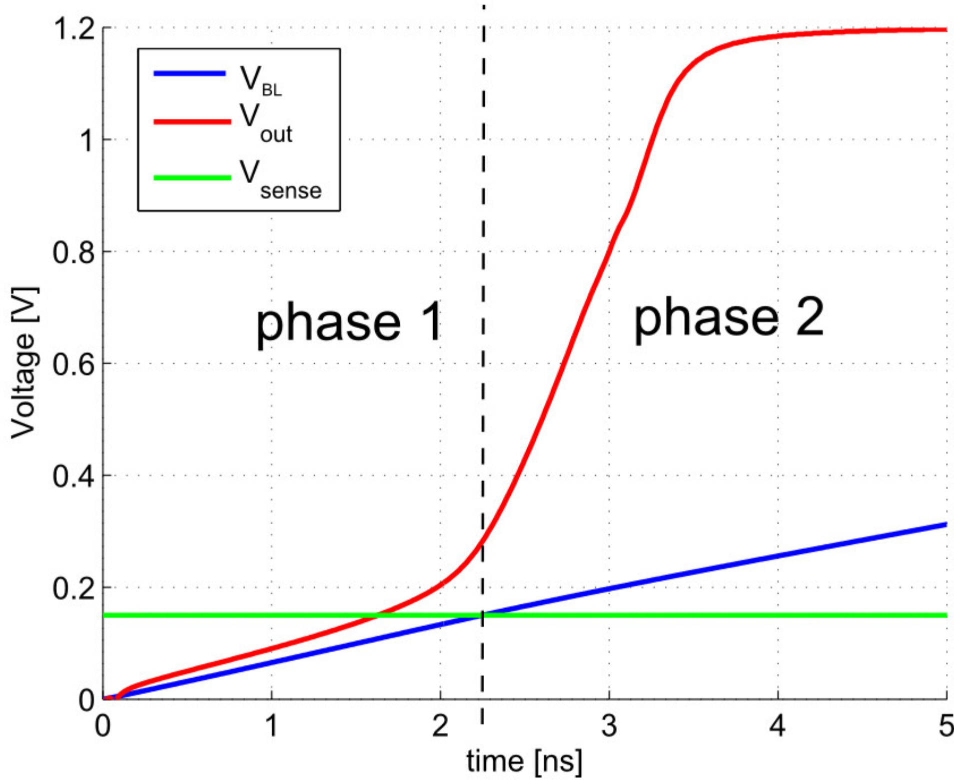

Figure 4.23: Simulation of Common Gate Sense Amplifier Utilizing Slope Detection

low V_{sense} voltage the transfer function obtained by small-signal analysis (see previous section) remains valid through the whole sensing period. Considering only low frequency behavior ($\frac{g_{m,N2}+g_{mbN2}}{C_F} >> s$ and $\frac{g_{mP2}}{C_{par}} >> s$) and neglecting the body-effect of the source follower transistors in the slope detection circuit ($g_{mbN2} << g_{m,N2}$) the transfer function of the slope depended current source 4.30 (source-side case) is simplified to:

$$I_{CF}\left(s\right) \approx sV_{IN}\left(s\right)C_F \tag{4.34}$$

Figure 4.24 shows the closed loop block diagram for the sensing scheme utilizing the slope detection (figure 4.22) during the sense phase before V_{BL} reaches the V_{sense} voltage (first part of the sensing phase). The cell current I_{CELL} and the bias current I_{BIAS} are integrated on the bitline capacitance C_{BL} generating the output voltage V_{out}. Before the bitline potential reaches V_{sense} voltage the gain of the common gate sense amplifier is one and V_{out} follows the V_{BL} potential. The bias current is directly proportional to the bitline voltage slope:

$$I_{BIAS} = sC_F V_{BL} + I_{CONST} \tag{4.35}$$

And therefore the current to voltage transfer function of the system is:

$$\frac{V_{out}}{I_{CELL} + I_{CONST}} = \frac{\frac{1}{sC_{BL}}}{1 - \frac{sC_F}{sC_{BL}}} = \frac{1}{s\left(C_{BL} - C_F\right)} \tag{4.36}$$

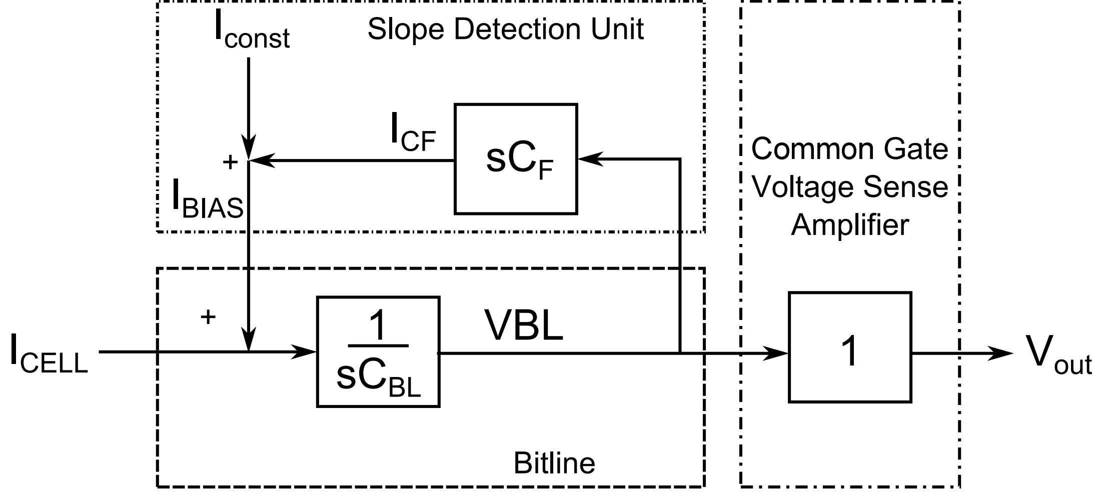

Figure 4.24: Block Diagram for Common Gate Sense Amplifier Utilizing Slope Detection ($V_{BL} > V_{sense}$ for drain-side; $V_{BL} < V_{sense}$ for source-side)

The cell current I_{CELL} and the constant bias current I_{CONST} are integrated on the capacitance $(C_{BL} - C_F)$, hence the effective bitline capacitance during the sense phase seen by the sense amplifier is now $(C_{BL} - C_F)$. The slope detection unit partially cancels the bitline capacitance C_{BL} by the C_F capacitance. The integration phase and therefore the sensing speed can now be adjusted by choosing the factor C_F and is no longer limited by the bitline capacitance C_{BL}. Thus the common gate sense amplifier enhanced by the slope detection circuit developed in this work, has the unique ability to overcome the physical limit of the state-of-the-art approach (see equation 4.22).

When V_{BL} reaches the V_{sense} potential the transistor N1 (see figure 4.22) enters the saturation changing the common gate stage from voltage follower to an amplifier. The I_{BIAS} current is no longer applied by its full value to the bitline, since it charges the parasitic capacitance of the V_{out} node till the supply voltage is reached. In addition the I_{BIAS} current is decreasing with increasing V_{out} potential due to finite output resistance of P2. Therefore the current feedback loop is no longer properly maintained after V_{BL} reaches the V_{sense} potential. To assess the worst case delay of the common gate amplifier the feedback is assumed to be open loop as shown in figure 4.25. The common gate stage amplifies V_{BL} by the factor $g_m r_{ds}$, where the I_{BIAS} current generated by the slope detection circuit determines the amplification speed. The bias current adaptively changes with cell current generating high bias current for high cell current and the bias current is at it's lowest value ($I_{BIAS} \approx I_{CONST}$) for zero cell current. Hence it significantly improves the speed of the common gate sense amplifier without disrupting the effective cell current window.

The sensing delay of the common gate sense amplifier enhanced by a slope detection can be calculated as follows:

$$T_{senseCGSDSoSiSA} = \frac{(C_{BL} - C_F)\, V_{sense}}{I_{CELL} + I_{CONST}} + \tau, max_{CGSoSiSA} \tag{4.37}$$

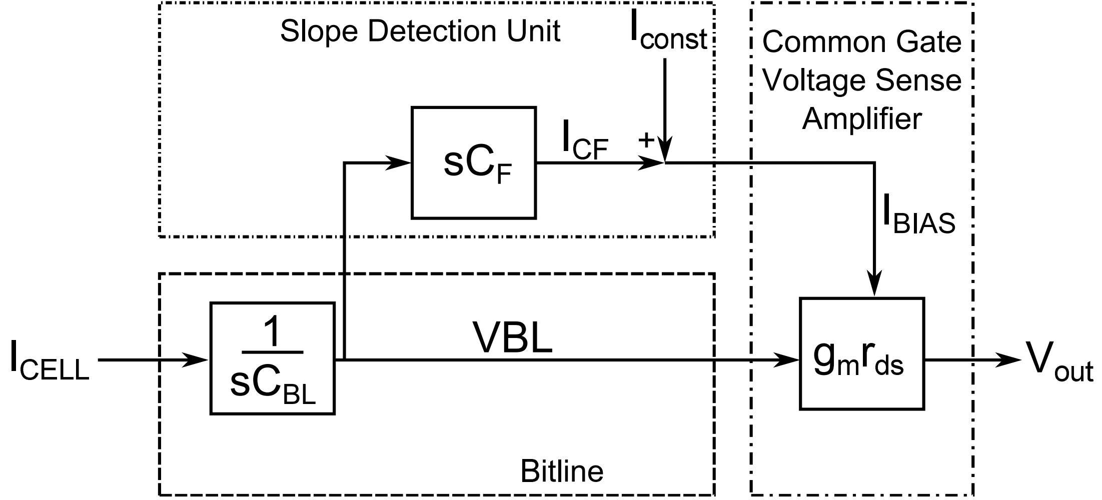

Figure 4.25: Block Diagram for Common Gate Sense Amplifier Utilizing Slope Detection ($V_{BL} < V_{sense}$ for drain-side; $V_{BL} > V_{sense}$ for source-side)

And the bias current can be calculated from equation 4.35 as:

$$I_{BIAS} = \frac{C_F}{C_{BL}} I_{CELL} + I_{CONST} \tag{4.38}$$

Compared to a common gate amplifier without slope detection enhancement (see equation 4.21) the proposed sensing circuit has significantly faster integration time due to partial effective bitline capacitance cancellation and higher speed of the common gate stage.

For better accuracy the body effect of the NMOS transistor (N2 in figure 4.22) in the slope detection circuit has to be considered:

$$T_{senseCGSDSoSiSA} = \frac{\left(C_{BL} - \frac{g_{m,N2}}{g_{m,N2}+g_{mb,N2}} C_F\right) V_{sense}}{I_{CELL} + I_{CONST}} + \tau, max_{CGSoSiSA} \tag{4.39}$$

$$I_{BIAS} = \frac{\frac{g_{m,N2}}{g_{m,N2}+g_{mb,N2}} C_F}{C_{BL}} I_{CELL} + I_{CONST} \tag{4.40}$$

Figure 4.26 compares the calculated sensing delay for the conventional comparator based approach (see equation 4.5), the common gate sense amplifier (see equation 4.21) and the bitline-capacitance-cancellation circuit (see equation 4.39) for source-side sensing. In addition the simulated delay of the bitline-capacitance-cancellation circuit is plotted. The simulation and the calculations were done for a 65nm technology.

The model based on level one transistor parameters gives already a very good approximation for the simulated delay. The common-gate circuit biased with $I_{BIAS} = 10\mu A$ has the same delay for $I_{CELL} = 20\mu A$ as the bitline-capacitance-cancellation sensing circuit but the delay of the common-gate circuit is not significantly increasing for low I_{CELL} values due to the constant $10\mu A$ bias current, which is applied directly to the bitline. As already mentioned this rather high bias current disrupts the relative cell current window

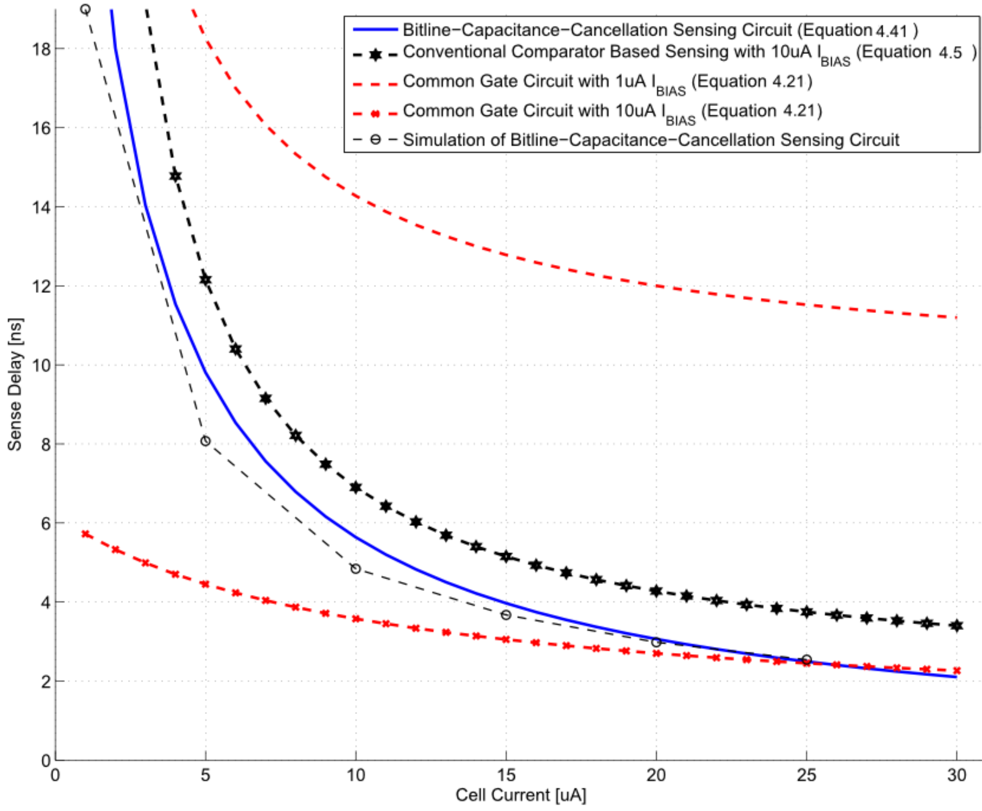

Figure 4.26: Simulation of Bitline-Capacitance-Cancellation Sensing Circuit vs. Model with following parameters: $g_{m,N2} = 75\mu A/V$, $g_{mb,N2} = 12\mu A/V$, $C_F = 125fF$, $C_{BL} = 350fF$, $V_{sense} = 0.15V$, $I_{CONST} = 1\mu A$, $C_{load} \approx 5fF$, $V_{outmax} = 0.8V$, $\lambda_n = \lambda_p \approx 0.4V^{-1}$.

Rel_{Iwin} (see equation 4.11) and is not acceptable for a robust read operation. Therefore the bitline-capacitance-cancellation sensing circuit was developed, which maintains the relative cell current window of about 95% for $I_{CONST} = 1\mu A$ (see equation 4.11) by adapting I_{BIAS} with respect to the cell current by the slope detection circuit. The common-gate amplifier biased with $I_{BIAS} = 1\mu A$ delivers the same relative cell current window but is more than two times slower compared to the bitline-capacitance-cancellation sensing circuit for cell current higher than $10\mu A$. The best relative cell current window of 100% is given by a conventional comparator based approach but this approach is about 40% slower for $I_{CELL} = 20\mu A$ compared to the bitline-capacitance-cancellation sensing circuit for the same bias current.

The minimal supply voltage for the sense amplifier is defined by the I_{CF}-path of the slope dependent current source circuit and the maximum required C_F voltage, which is usually $\approx V_{sense}$:

$$VDD_{MIN} = V_{OV,N2} + |V_{OV,P5}| + |V_{TH,P5}| + V_{sense} \tag{4.41}$$

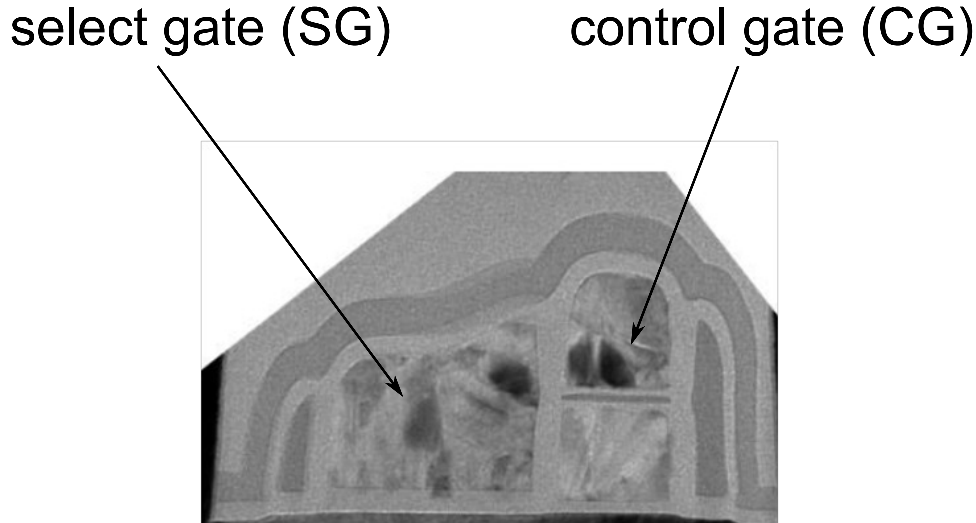

Figure 4.27: Cross Section of the HS3P-cell

Since V_{sense} is usually set below 0.2V, the minimum supply voltage is lower or equal to the comparator based approach [32], [33], [34] and comparable with the Type A current sense amplifier approach for very low voltage (see figure 3.15).

4.5 Source-Side Sensing for 65nm HS3P eFlash

Automotive applications require highly reliable memories providing more than 10 years data retention and full operation functionality up to 170°C. One of the promising candidates for sub-65nm technology is the HS3P eFlash [39]. It is a 2 transistor split-gate eFlash cell suitable for a fast and low power write operation by source side injection and a fast read operation by providing a cell current of more than $20\mu A$. Figure 4.27 shows the cross section of the HS3P-cell with dedicated select transistor controlled by the select gate and a memory transistor comprising the control and floating gate. The write operation is performed by source-side injection [7], injecting hot electrons into the floating gate to shift the transfer curve of the cell (see chapter 2.1). Electrons are accelerated by the high electric field from source to drain and injected into the floating gate by applying high gate voltage as shown in figure 4.28. The select gate overdrive voltage allows to control the programming current for low power operation. The electrons are extracted from the floating gate into the bulk during the erase operation by Fowler-Nordheim tunneling [7] as shown in figure 4.29.

From design point of view the cell shown in figure 4.27 can be modeled as two transistors in series as shown in figure 4.30. The select gate transistor is acting like a switch for the eFlash cell allowing to suppress the leakage of deselected cells on the bitline during read and write operation [39]. The control gate transistor is the actual storage element which has a fixed gate bias during read operation. There are four different arrangements with respect to the electrical drain or source of the cell connected to the bitline and the arrangements of the two transistors to each other. Figure 4.30 and 4.31 show the so called

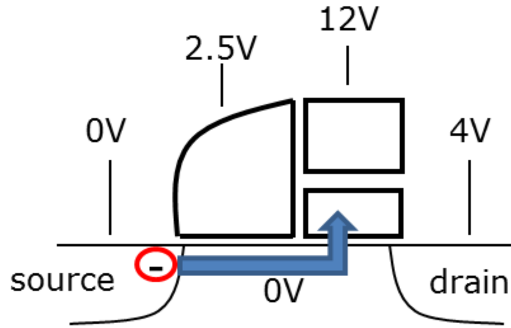

Figure 4.28: Typical Write Operation for HS3P-cell

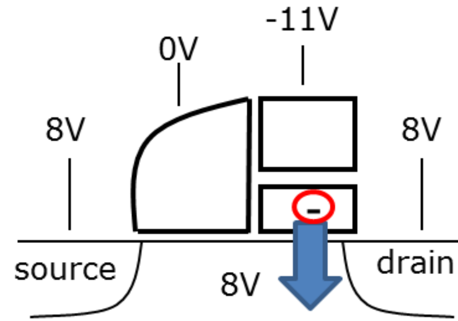

Figure 4.29: Typical Erase Operation for HS3P-cell

forward connected cell where the select gate is connected to the electrical drain of the cell. The reverse connected cell is shown in figure 4.32 and 4.33.

For the forward connected cell the control gate transistor is biased in linear region and the select gate transistor is in saturation. Therefore the control gate transistor acts as a resistor placed at the source of the select gate transistor reducing its gate-source voltage. The select gate transistor has higher threshold voltage compared to the control gate transistor. Therefore the forward connection has lower read current compared to reverse connection of the HS3P-cell where the select gate is biased with the maximum overdrive voltage. The single cell current measurements for read biasing conditions with V_{CG} increased by constant voltage steps $\triangle V$ shown in figure 4.34 confirm this behavior revealing more than 50% cell current increase for the reverse connection (green dashed lines) of the cell. Hence the reverse read connected cell is required for the high speed read operation.

As already stated in chapter 3.2 the drain-side sensing is the best choice for embedded memories (see section 3.2). But the reverse connected HS3P cell for drain-side sensing (see figure 4.32) exhibits one important drawback. The HS3P cell transistor capacitance seen from the bitline side (drain-side capacitance of control gate) is depending on its threshold voltage and therefore on the cell state. For a low resistive cell state the threshold voltage of the control gate is below 0V, it drives the control gate transistor into inversion with inversion charge in the transistor channel for the read condition. And for high resistive state the control gate transistor is in depletion region with no charge in the transistor channel. Therefore the bitline capacitance is varying depending on the pattern stored in the memory for drain-side sensing of a reverse connected cell.

The different bitline capacitance values will lead to different charging speeds among the bitlines during the read operation, causing an access time spread among the bitlines. It can be handled by taking the additional capacitance into account. This generates an additional drawback for current integrating schemes which require a reference bitline which should be ideally C-matched to the selected bitlines. The reference bitline serves as a reference for many selected bitlines, where each bitline can comprise a variation of high/low resistive cells. The bitline capacitance mismatch will reduce the effective voltage or time window and hence the robustness of the sensing scheme. It should be mentioned

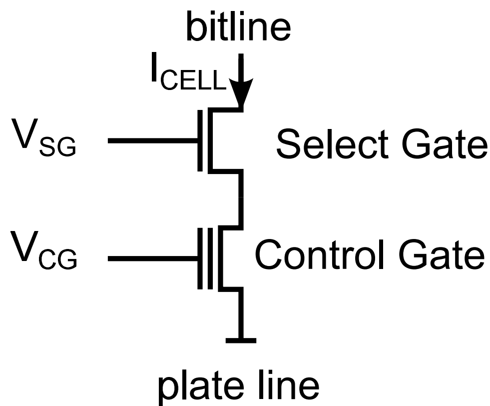

Figure 4.30: HS3P Cell in Forward Connection for Drain-Side Sense Amplifier

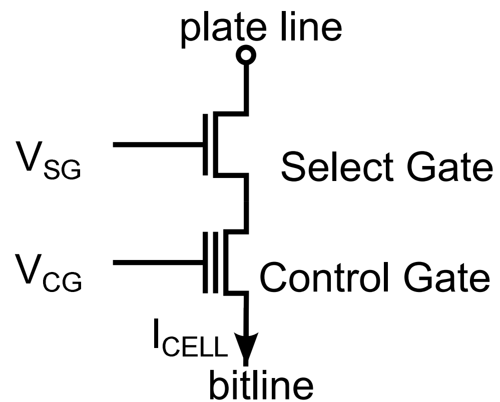

Figure 4.31: HS3P Cell in Forward Connection for Source-Side Sense Amplifier

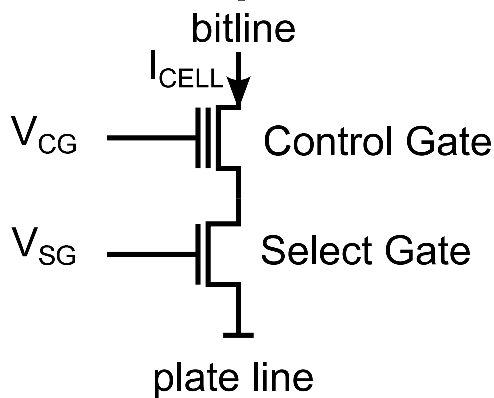

Figure 4.32: HS3P Cell in Reverse Connection for Drain-Side Sense Amplifier

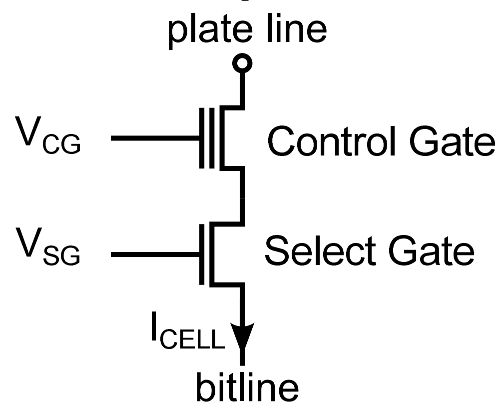

Figure 4.33: HS3P Cell in Reverse Connection for Source-Side Sense Amplifier

here that for sub 65nm technologies the bitline capacitance is mainly dominated by the metal to metal capacitance and therefore the mentioned drawback is relatively small, but should be considered as it defines the robustness and the accuracy of the voltage or time domain sensing scheme.

It is possible to overcome this drawback if the time domain sensing is used for the reverse connected cell for source-side (see figure 4.33) sensing. Because the cell capacitance seen from the bitline is the source-side capacitance of the select gate transistor, it is independent on the cell state. The cell current is integrated on the bitline capacitance and the bitline voltage which is the source voltage of the cell is increasing during the sense phase. The bitline voltage increase reduce the cell current due to reduction of the gate to source voltage and body effect of the select gate transistor. Therefore this cell configuration is not well suited for the conventional voltage sensing, because the absolute bitline voltage change during the sense phase can not be controlled or adjusted. This is due to the fact that only the relative voltage change of the selected bitline with respect to the reference bitline is monitored during the sense phase (see chapter 3.4.2). This would

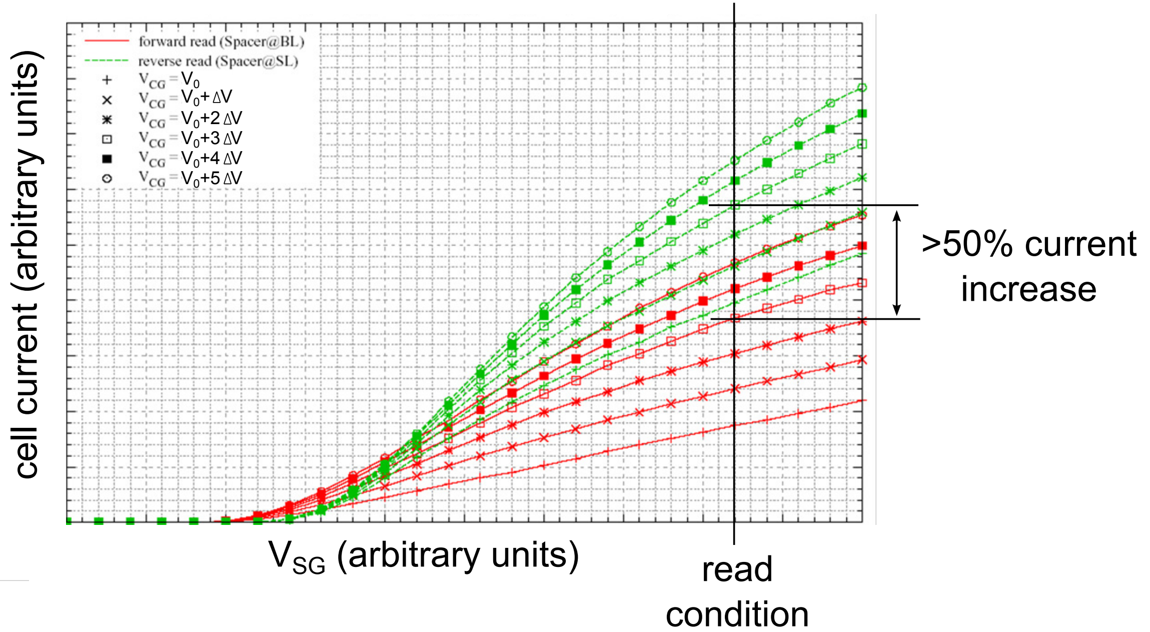

Figure 4.34: Read Current Measurement for HS3P eFlash Cell

lead to strong cell current reduction reducing the sense speed and the robustness of the sensing scheme.

The time domain sensing allows to control the required absolute bitline voltage change during the read operation by adjusting the V_{sense} voltage. It should be set as low as possible to minimize the cell current reduction due to bitline voltage increase and reduce the integration time and the sense delay.

4.6 Sensing Scheme Architecture and Measurement Results

Figure 4.35 shows the 512kByte demonstrator chip fabricated in 65nm technology. The memory array is split in two wings each comprising 256kB. 1024 wordline drivers are located between the two memory wings and the sense amplifiers are placed at the bottom of each wing. 128 sense amplifiers correspond to one 256kB wing, to enable simultaneous read out of 256bits. The schematic of one 256kB memory wing is shown in Figure 4.36. The sense amplifier (SA) acts as a voltage level detector, generating a digital output signal (SO), if the bitline potential reaches the V_{sense} voltage, which is set to 100mV. Each sense amplifier on the cell array side is multiplexed to 16 bitlines (BL) by the select<15:0> signals. The bitline multiplexer comprises high voltage protection circuitry to protect the V_{DD} domain circuits from high voltage, which is required during write and erase operation. The multiplexer concept will be discussed in more detail in section 5.1. On the cell array side, one bitline is connected to the sense amplifier and the 15 deselected

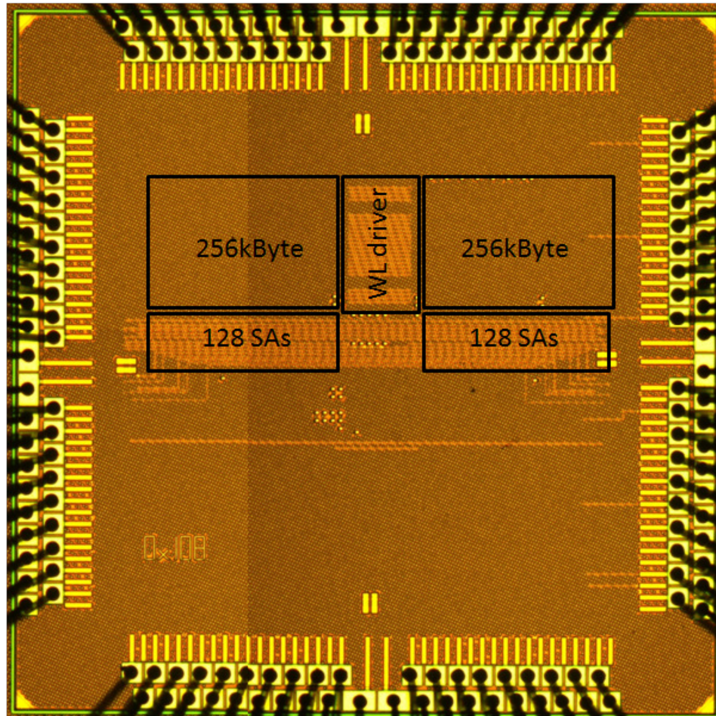

Figure 4.35: Die Photograph of the 512kB Demonstrator Chip

bitlines are continuously charged to core supply voltage V_{DD}. By using this continuous precharge scheme, the bitline capacitance C_{BL} of all deselected bitlines acts as a buffer cap for the V_{DD} supply. This way the supply noise is suppressed, enabling low-voltage swing sensing phase for low V_{sense} values, and there is no need for a precharge buffer [9]. The reference path includes one sense amplifier that is connected to one reference bitline BL_{REF}. To reduce the area and power overhead only one global reference time generator (reference sense amplifier) is used for each wing (see section 4.1). The output of the reference sense amplifier is used to strobe and latch the output signals of all sense amplifiers on the cell array. The drain-side of the eFLASH cells on this reference bitline (BL_{REF}) is floating. Therefore no cell current occurs, resulting in a matched C_{BL} between BL_{REF} and the bitlines of the cell array. A constant reference current (I_{REF}) charges C_{BL} and thus defines the charging speed of the BL_{REF} with respect to the memory bitlines BL on the array side.

Measured internal signals are shown in Figure 4.37 and the corresponding timing diagram in Figure 4.38. After address (addr) change, the precharge signal (PRE) goes high and connects the selected bitlines and BL_{REF} to ground. The wordline (WL) rises to 4.2V (not shown in Figure 4.37) and the control gate in the floating memory cell is set in a range of 3V to 4V. After precharge time of 5ns (T_{pre}) BL and BL_{REF} are precharged to V_{SS} and the PRE signal changes to low. The selected bitlines and BL_{REF} are then charged by I_{CELL} and I_{REF}, respectively. The bitline where a low-resistance cell is selected is charged faster (BL(0)) than a bitline with a high-resistance cell (BL(1)). The voltage on the BL_{REF} reaches the V_{sense} voltage level later than BL(0) and earlier than BL(1). The corresponding signal REFSO changes its state and latches the output

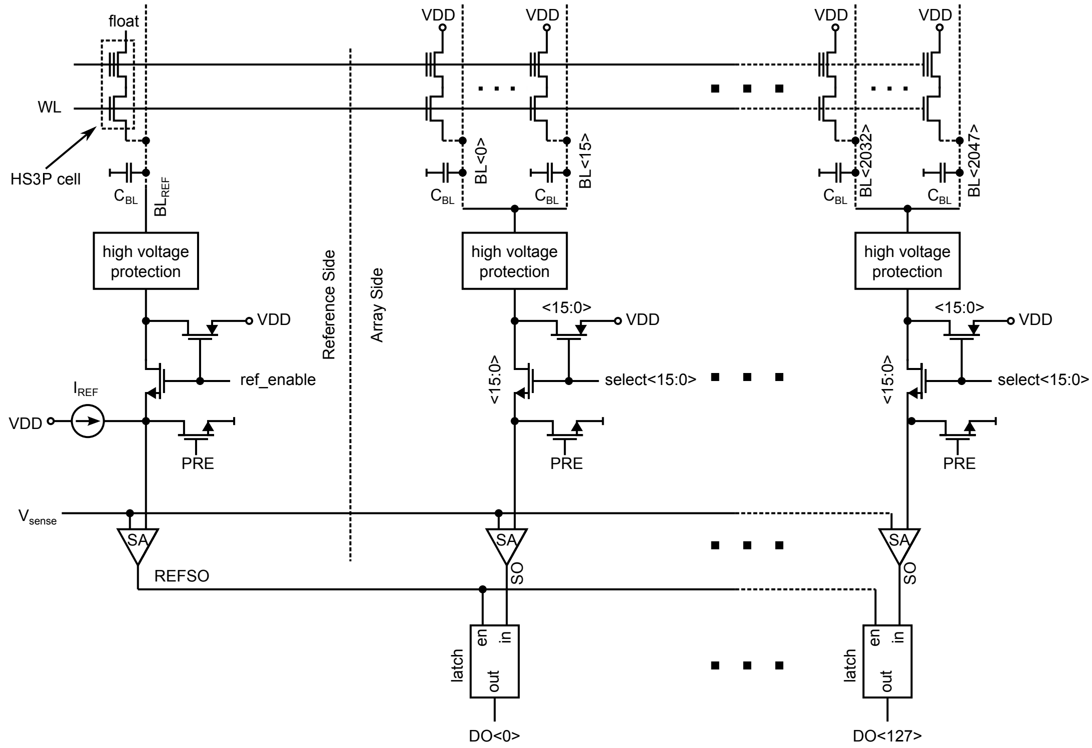

Figure 4.36: Memory Wing Architecture of Figure 4.35

signals of all sense amplifiers. The minimum read-access time (T_{acc}) is defined by the time when the sense amplifier reading a low-resistance state generates the digital output signal (SO(0) goes low in Figure 4.37, 4.38). After this time it is possible to distinguish between a high- and low-resistance cell state (see Figure 4.38). The read window which is measured by time for time domain sensing (T_{win}) is defined as the time period between high-low transitions of the signals SO(0) (low-resistance state) and signal SO(1) (high-resistance state).

The bitline-capacitance-cancellation sense amplifier (SA) schematic comprising the slope detection circuit is shown in figure 4.39. The sense amplifier comprises the two circuit blocks discussed in the previous sections: the common gate sense amplifier and the slope dependent current source, which are combined in the bitline-capacitance-cancellation sensing circuit for source-side sensing as shown in figure 4.22. To generate the V_{Ngate} voltage, additional bias circuitry is added. The NMOS pair N4 and N1 acts as current mirror. The source of N4 is connected to V_{sense}, which is regulated by the global OpAmp and local N5 transistor. The transistor P6 acts as a bias source for N4 and N5. The global V_{sense} regulator uses 3.3V devices and is connected to the I/O supply. If the bitline potential is below V_{sense}, N1 is conducting and the node V_{int} follows the bitline potential. As soon as the bitline voltage reaches the V_{sense} potential, N1 and P2 become a common gate amplifier (see section 4.2.3). Further bitline voltage increase is amplified at the node V_{int} to V_{DD}, forcing the digital output signal SO to the low state. The C_F capacitance is discharged during the precharge phase by an NMOS transistor to V_{SS} potential to guarantee the same initial conditions for every read operation and correct circuit operation of the slope detection unit circuitry.

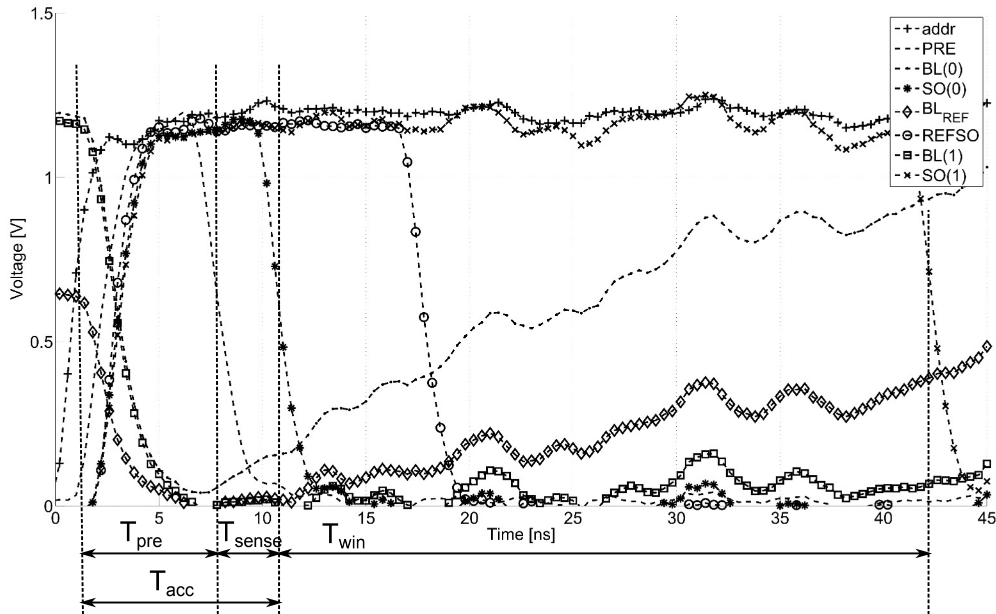

Figure 4.37: Measured Internal Signals at 25°C and Nominal Supply Voltage of Circuit in Figure 4.36

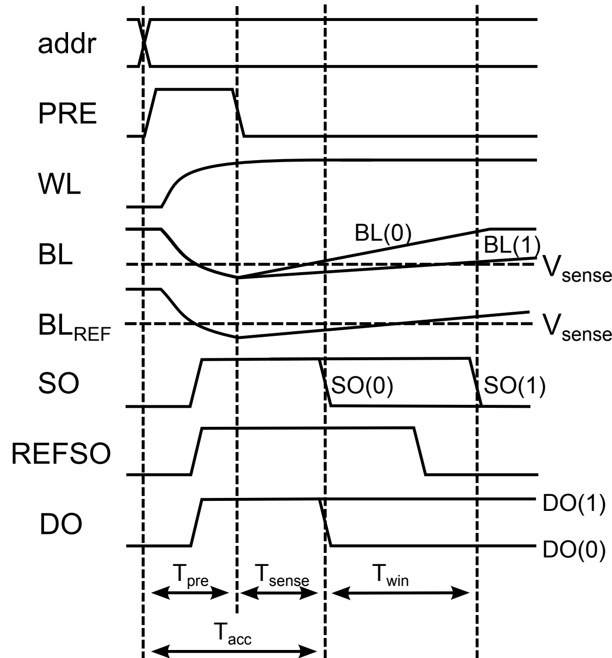

Figure 4.38: Timing Diagram for Internal Signals of Circuit in Figure 4.36

To disable the slope detection circuit two additional NMOS transistors where added which are controlled with sl_en and its complement signal sl_en_b. By setting the signal sl_en to low, the slope detection circuit can be deactivated for sensing only by the common gate stage. In this mode, the current I_{CONST}, which is the bias current for the common gate amplifier, can be varied and the PRE signal inside the sense amplifier remains at

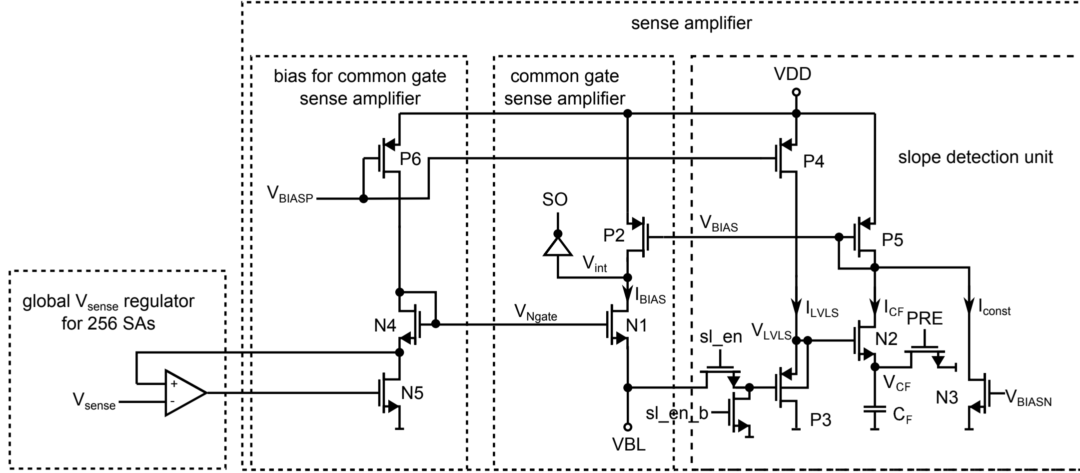

Figure 4.39: Bitline-Capacitance-Cancellation Sense Amplifier Circuit

V_{DD} to ensure no current flow in the I_{CF}-current-path.

Figure 4.40 shows the measured sense delay T_{sense} and read window T_{win} for different C_F values with activated slope detection circuit and I_{CONST}=1μA. The same measurements are made with slope detection circuit off for different I_{CONST} currents, which corresponds to a conventional common gate sense amplifier circuit. T_{sense} with I_{CONST} of $4\mu A$ (slope detection circuit off) is comparable to T_{sense} with C_F of 100fF (slope detection circuit on). But T_{win} is reduced by more than a factor of 3 if the slope detection circuit is off, leading to a read failure if the difference of I_{CELL} between high- and low-resistance cells is small. As already mentioned it is due to the rather high constant bias current I_{CONST} which is integrated on the bitline capacitance together with the cell current, reducing the relative cell current window Rel_{Iwin}. In case of the enabled slope detection circuit the bitline-capacitance-cancellation circuit adjusts its bias current I_{BIAS} depending on the cell current to ensure that Rel_{Iwin} is not reduced. The C_F of 100fF occupies 20% of the SA area.

To evaluate the quality of the slope detection circuit, a read window margin is defined as:

$$W_{MAR} = \frac{T_{win}}{T_{sense}} \tag{4.42}$$

Measured W_{MAR} is plotted in figure 4.41. When the slope detection circuit is on, W_{MAR} is between 7 and 8 and stays constant if the effective C_{BL} is reduced, proving the expected function of the circuitry. In contrast, W_{MAR} drops significantly if the slope mode is off and I_{CONST} is increased leading to reduced Rel_{Iwin}.

Figure 4.42 shows a T_{acc} shmoo with bit yield for the bitline-capacitance-cancellation sensing scheme. The read access time of the sensing scheme is measured by reading out the full 512kByte memory. At nominal supply voltages 1.2V V_{DD} and 3.3V V_{DDVIO}, T_{acc} is 9.1ns and T_{win} is 32ns (see figure 4.42) giving high robustness. T_{acc} is still below 11ns at 160°C and 20% lower supply voltages (0.9V V_{DD} and 2.64V V_{DDVIO}). By reading 256 sense amplifiers simultaneously the maximum read throughput is 2.9 GByte/s.

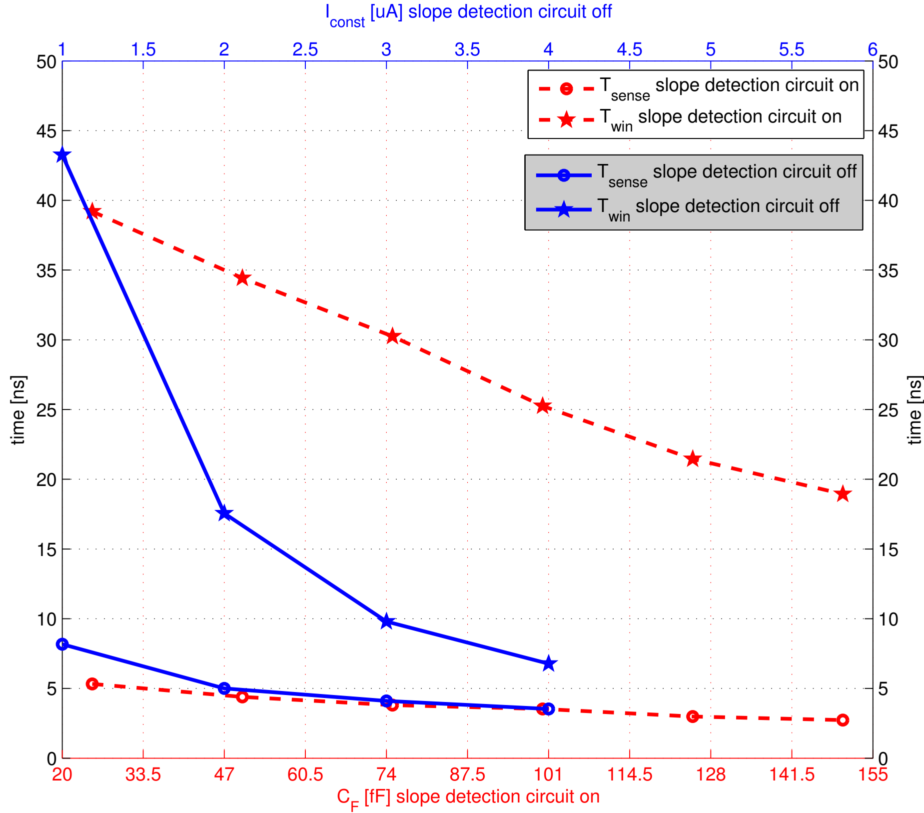

Figure 4.40: Measurement of Sense Delay T_{sense} vs. Read Window T_{win} for the Bitline-Capacitance-Cancellation Sense Amplifier and the Common Gate Sense Amplifier

4.7 Power Considerations

The dynamic current consumption of the proposed sensing circuit (see figure 4.39) is defined by the precharge current of the bitline capacitance, the cell current I_{CELL} and the bias current I_{BIAS} which flows during the whole precharge phase (see section 3.5), and the current I_{CF} charging the capacitance of the slope detection circuit (see figure 4.39). I_{BIAS} remains constant ($=I_{CONST}$) during the precharge phase since the bitline voltage is not rising. The total average current is therefore:

$$\begin{aligned} Ivdd_{dynCGSDSoSiSA} &= \frac{V_{DD} \cdot C_{BL}}{T_{accCGSDSoSiSA}} + (I_{CELL} + I_{CONST}) \frac{T_{preCGSDSoSiSA}}{T_{accCGSDSoSiSA}} \\ &+ \frac{V_{sense} \cdot C_F}{T_{accCGSDSoSiSA}} \end{aligned} \tag{4.43}$$

The bitline charging current portion is considered only for one bitline as the time domain sensing allows the usage of a global reference time generator (see figure 4.3). The C_F capacitance is charged close to the V_{sense} potential during the sense phase. $T_{preCGSDSoSiSA}$ is the precharge time of the time domain sensing circuit. It is not the same as for the voltage sensing. This is because for time domain sensing the bitline voltage is evaluated

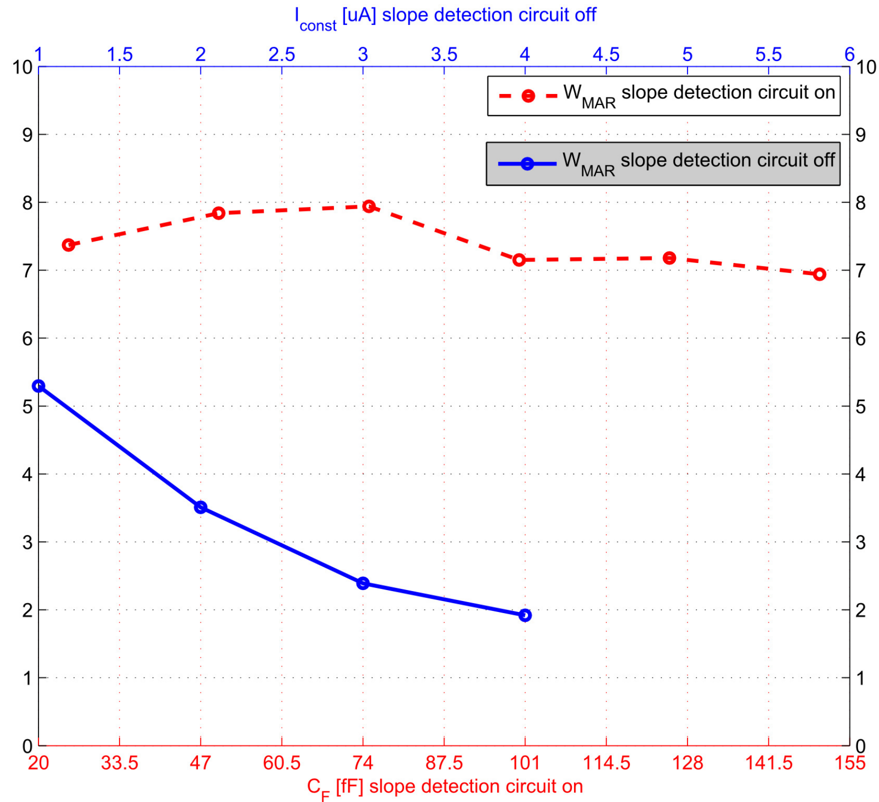

Figure 4.41: Measurement of Read Window Margin W_{MAR} for the Bitline-Capacitance-Cancellation Sense Amplifier and the Common Gate Sense Amplifier

against a fixed voltage reference (V_{sense} in figure 4.1). The voltage bump on the bitline due to charge sharing (see chapter 3.4.2) has to be low compared to the V_{sense} voltage for proper sensing operation. As a rule of thumb the voltage bump on the bitline is acceptable if the bitline capacitance in the lumped bitline model in figure 3.17 is charged lower than $0.5V_{sense}$ for source side or higher than V_{DD}-$0.5V_{sense}$ for drain-side sensing. Hence the

Temp	VDD	VDDVIO	8.8	9.1	9.4	9.7	10.0	10.3	10.6
25°C	1.43V	3.63V	98.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
25°C	1.2V	3.3V	98.5%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
25°C	1.04V	2.64V	85.1%	99.6%	100.0%	100.0%	100.0%	100.0%	100.0%
25°C	0.9V	2.64V	61.3%	99.9%	100.0%	100.0%	100.00%	100.00%	100.00%
160°C	1.43V	3.63V	0.55%	9.21%	24.93%	69.82%	98.19%	100.00%	100.00%
160°C	1.2V	3.3V	0.99%	33.65%	56.52%	81.27%	99.99%	100.00%	100.00%
160°C	1.04V	2.64V	8.61%	12.48%	53.82%	75.25%	99.96%	100.00%	100.00%
160°C	0.9V	2.64V	0.01%	15.52%	34.12%	63.89%	96.93%	98.78%	100.00%

Access time [ns]

Figure 4.42: Read Access Time Shmoo of the Demonstrator Chip Shown in Figure 4.35

precharge time for time domain sensing utilizing source-side sensing can be calculated as:

$$T_{preCGSDSoSiSA} = -R_{BLpath}C_{BL}ln\left(\frac{0.5V_{sense}}{V_{PRE}}\right) \tag{4.44}$$

For typical values of resistances and capacitances in the read path for voltage sensing during the precharge phase ($R_{SA} = 100\Omega, C_{BL} = 1pF, R_{BL} = 2k\Omega, R_{MUX} = 2k\Omega$), V_{PRE}=V_{DD}=1.2V and V_{sense}=100mV) the equation 4.44 predicts a precharge time of 9.9ns. This is more than factor of two higher compared to voltage sensing but still lower with respect to current sensing (see section 3.4). The access time $T_{accCGSDSoSiSA}$ of the common gate sense amplifier enhanced by the slope detection circuit for source-side sensing setup comprises precharge and sense time:

$$T_{accCGSDSoSiSA} = T_{preCGSDSoSiSA} + T_{senseCGSDSoSiSA} \tag{4.45}$$

Capacitance C_F should be well below the bitline capacitance C_{BL} to guarantee robust read operation. As a rule of thumb, C_F should be below half of the bitline capacitance. Assuming 500fF for C_F the access time for the typical values in eFlash memories ($R_{SA} = 100\Omega, C_{BL} = 1pF, R_{BL} = 2k\Omega, R_{MUX} = 2k\Omega, V_{PRE} = V_{DD} = 1.2V$ and $V_{sense} = 100mV$) and the circuit parameters used in previous sections ($g_{m,N2} = 75\mu A/V$, $g_{mb,N2} = 12\mu A/V$, $I_{CONST} = 1\mu A$, $C_{load} \approx 5fF$, $V_{outmax} = 0.6V$, $\lambda_n = \lambda_p \approx 0.4V^{-1}$) $T_{accCGSDSoSiSA} = 9.9ns + 3.6ns = 13.5ns$ which is quite the same value as for the voltage sensing approach (see table 3.1).

The static current consumption portion of the common-gate sense amplifier is defined by the bias current in the slope detection circuit:

$$Ivdd_{staticCGSDSoSiSA} = I_{CONST} + I_{LVLS} \tag{4.46}$$

Typical values as used in figure 4.26 for I_{LVLS} and I_{CONST} current are $4\mu A$ and $1\mu A$ respectively. The current consumption of the bias circuit for the common gate sense amplifier (see figure 4.39) is not taken into account, because the bias circuits for the voltage and the current sensing approaches are also omitted (see section 3.5.3). The power consumption of the V_{sense} regulator (see figure 4.39) is negligible compared to the power consumption of 256 sense amplifiers.

The overall power consumption of the common gate sense amplifier enhanced by the slope detection circuit for the source-side sensing setup is:

$$Pvdd_{CGSDSoSiSA} = (Ivdd_{dynCGSDSoSiSA} + Ivdd_{staticCGSDSoSiSA})\,V_{DD} \tag{4.47}$$

For typical values used before the power consumption is $Pvdd_{CGSDSoSiSA} \approx 136mW$.

For the common gate sense amplifier without the slope detection circuit there is no static current portion (see chapter 4.2). Therefore it's power consumption can be approximated as follows:

$$Pvdd_{CGSoSiSA} = \left(\frac{V_{DD} \cdot C_{BL}}{T_{accCGSoSiSA}} + (I_{CELL} + I_{BIAS})\frac{T_{preCGSoSiSA}}{T_{accCGSoSiSA}}\right)V_{DD} \tag{4.48}$$

Table 4.1: Time Domain Sensing Approach Comparison using typical values for the eFlash memory ($R_{SA} = 100\Omega$, $C_{BL} = 1pF$, $R_{BL} = 2k\Omega$, $R_{MUX} = 2k\Omega$, $V_{PRE} = V_{DD} = 1.2V$ and $V_{sense} = 100mV$) and transistor parameters ($g_{m,N2} = 75\mu A/V$, $g_{mb,N2} = 12\mu A/V$, $I_{CONST} = 1\mu A$, $C_{load} \approx 5fF$, $V_{outmax} = 0.6V$, $\lambda_n = \lambda_p \approx 0.4V^{-1}$)

	Bitline-Capacitance-Cancellation Sense Amplifier	Common Gate Sense Amplifier I_{BIAS}=1μA	Comparator Based Approach I_{BIAS}=10μA	Lower Limit for state-of-the-art (not considering any sense amplifier circuitry)
Access Time	13.5ns	23.4ns	16.5ns	14.9ns
Power per Sense Amplifier	134mW	73mW	114mW	68mW
FOM_{PS} in $\frac{10^6}{Ws}$	553	585	532	987

The power consumption of the comparator based time domain sensing approach (see chapter 4.2) is defined by the bitline capacitance charging current the static power consumption of the comparator circuit:

$$Pvdd_{COMPSoSiSA} = \left(\frac{V_{DD} \cdot C_{BL}}{T_{accCOMPSoSiSA}} + I_{CELL} \frac{T_{preCOMPSoSiSA}}{T_{accCOMPSoSiSA}} + I_{BIAS} \right) V_{DD} \quad (4.49)$$

The precharge time for common gate and the comparator based sense amplifier are equal to the precharge time calculated for the bitline-capacitance-cancellation sense amplifier:

$$T_{preCGSoSiSA} = T_{preCOMPSoSiSA} = T_{preCGSDSoSiSA} = T_{preTiDo} \quad (4.50)$$

4.8 Comparison

Table 4.1 compares the three different sensing approaches. The figure of merit for power efficiency and speed (FOM_{PS}) is used for a fair comparison for different approaches (see chapter 3.5.3). The power consumption is scaled with the corresponding access time and the inverse value of this product is taken to define the FOM_{PS} (higher value is better). In addition to the three sensing schemes discussed before the lower physical limit for the state-of-the-art current integrating sensing scheme approach is added to the table. The lowest achievable access time for the conventional current integrating schemes is defined

by the precharge time for time domain sensing $T_{preTiDo}$ (see equation 4.50) and the lower physical limit for a sense delay of a current integrating sensing scheme T_{limit} (see equation 4.22):

$$T_{acclimitTiDo} = T_{preTiDo} + T_{limit} \quad (4.51)$$

The lower limit for power consumption for a time domain sensing scheme is defined by the charging current for the bitline capacitance and the cell current flow during the precharge phase:

$$Pvdd_{limitTiDo} = \left(\frac{V_{DD} \cdot C_{BL}}{T_{acclimitTiDo}} + I_{CELL}\frac{T_{preTiDo}}{T_{acclimitTiDo}}\right) V_{DD} \quad (4.52)$$

The lowest read access time of 13.5ns with moderate power consumption of 134mW is given by the bitline-capacitance-cancellation sensing scheme comprising the common gate amplifier enhanced by the slope detection circuit. The slope detection circuit allows to overcome the lower bound for the conventional time domain sensing approach. If the slope detection enhancement is not utilized the common gate amplifier suffers from strong speed degradation. The access time increases by more than 70% to 23.4ns (calculated for I_{BIAS}=1μA) but it demonstrates the lowest power consumption among the three sensing approaches. It is mainly defined by the charging current required to charge the bitline capacitance during the precharge operation with low current consumption in the sense phase. The increased bias current for the common gate amplifier reduces the access time but the relative cell current window Rel_{Iwin} significantly decreases (see section 4.2) which makes this approach not suitable for robust read operation. The conventional time domain sensing with comparator based approach is more than 20% slower compared to the bitline-capacitance-cancellation sensing scheme with even lower figure of merit FOM_{PS}. By increasing the bias current of the comparator based approach it will never reach the speed of the bitline-capacitance-cancellation sensing scheme due its physical limit of 14.9ns.

Compared to the voltage or current sensing approaches (see Table 3.1) the time domain sensing schemes reaches significantly higher figure of merit values for power efficiency and speed FOM_{PS}. This is due to the usage of the global reference time generator concept (see figure 4.3) which does not require to charge two bitline capacitances during the precharge phase (as for voltage sensing) or any static reference current (as for current sensing).

For high speed applications the the bitline-capacitance-cancellation sensing scheme is the best choice. It is more than 20% faster compared to the next best approach and even faster than the limit for state-of-the-art time domain sensing schemes. And it has the second best FOM_{PS} value among the three time domain sensing schemes. For ultra low power application with relaxed access time requirements the common gate sense amplifier is the best choice, since it achieves the lowest power consumption among all sensing approaches with the best FOM_{PS} value for power efficiency and speed. In addition the area consumption is the smallest among all sense amplifier types, since it requires only two transistors (see figure 4.9). The area consumption of the bitline-capacitance-cancellation sense amplifier is comparable to the comparator based approach and is lower compared to the current sensing schemes, caused by the fact that it does not require any current distribution or additional precharge circuitry.

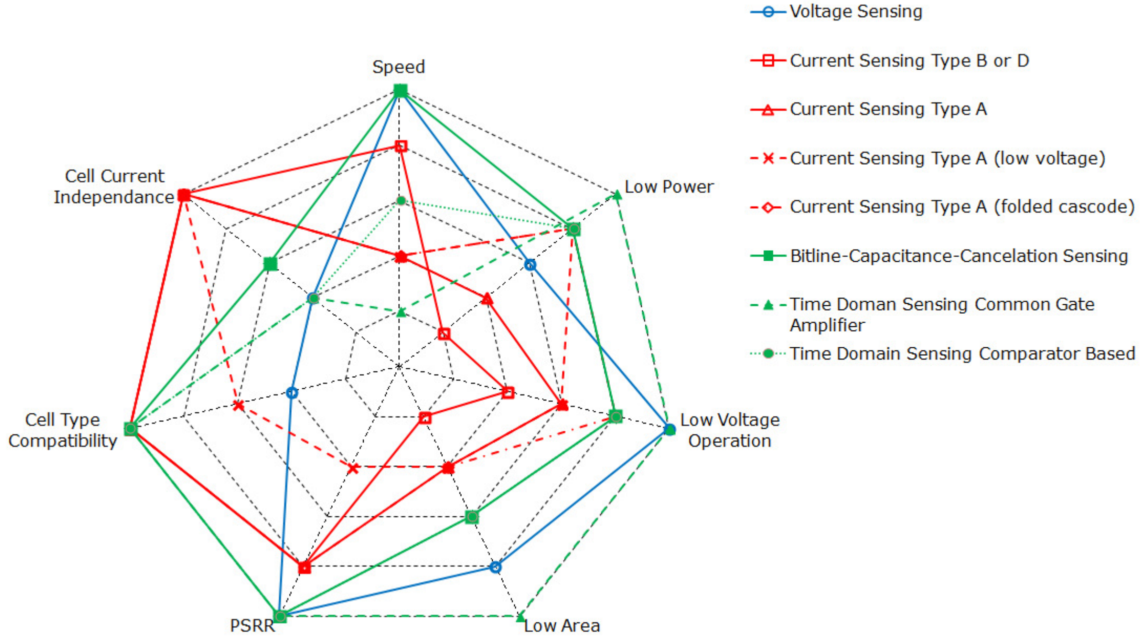

Figure 4.43: Sensing Approach Comparison

4.9 Summary and Conclusion

The accurately defined low voltage swing of the bitline during the read operation makes the time domain sensing concept suitable for the linear region biased cell. The slope detection enhancement introduced in this work, makes the bitline-capacitance-cancellation sensing scheme, as developed in this work, a superior sensing concept compared to voltage, current and state-of-the-art time domain sensing schemes for the eFlash memories. Figure 4.43 summarizes the findings of the previous sections and compares all the sensing approaches discussed. The bitline-capacitance-cancellation sensing scheme significantly overcomes the state-of-the-art approaches due to:

- global time reference concept allowing to reduce the bitline charging current compared to voltage sensing
- high $PSRR$ achieved by symmetrical reference circuit and digital coded timing for reference signal
- low power and low area required for time to voltage comparison stage is realized by the common gate amplifier
- slope detection circuit enhancement of the common gate amplifier allowing to overcome the lower physical limit for the read access time T_{limit} of the state-of-the-art integration scheme
- low static current consumption of the slope detection circuit part

The measurement results of a 512kByte demonstrator test chip utilizing the bitline-capacitance-cancellation sensing scheme fabricated in 65nm reveal a read throughput of 2.9GByte/s and a worst case read access time of 11ns measured at 160°C.

Chapter 5

Embedded Flash Macro Design Aspects

This section discusses the design of the embedded non-volatile macro for automotive applications utilizing the HS3P-eFlash technology [39]. The focus of this section is on optimizations in the read path circuits of a non-volatile memory macro to achieve power efficient and robust high speed read operation fulfilling the automotive requirements. To compare the advantages of the macro design developed in this work, which is discussed in the next sections, it will be compared to the memory macro which was fabricated in the same 65nm technology comprising the HS3P-cell but utilizes a conventional approach (see figure 3.10) based on a previous design [9].

5.1 Macro Architecture

Figure 5.1 shows a chip micrograph with removed metalization of the 4MB embedded Flash macro fabricated in 65nm technology [2, Jefremow]. The memory array is split in two memory banks 2MB each. To achieve high read throughput every memory bank has its own read path and comprises 280 sense amplifiers (SAs) which can be read out in parallel. The analog block with charge pumps and reference circuits is located at the bottom of the macro.

5.2 Memory Array

The proposed macro utilizes the $HS3P$-cell [39] (see section 4.5) with dedicated select transistor controlled by the select gate (see figure 4.27) and a memory transistor comprising the control gate (see figure 4.27). $HS3P$-cell allows fast write operation due to low current and efficient source side injection (see section 4.5).

To minimize the RC-delay of the wordline signal the memory array of each bank is split in two tiles 1MB each with wordline drivers in between. In addition the SG-path

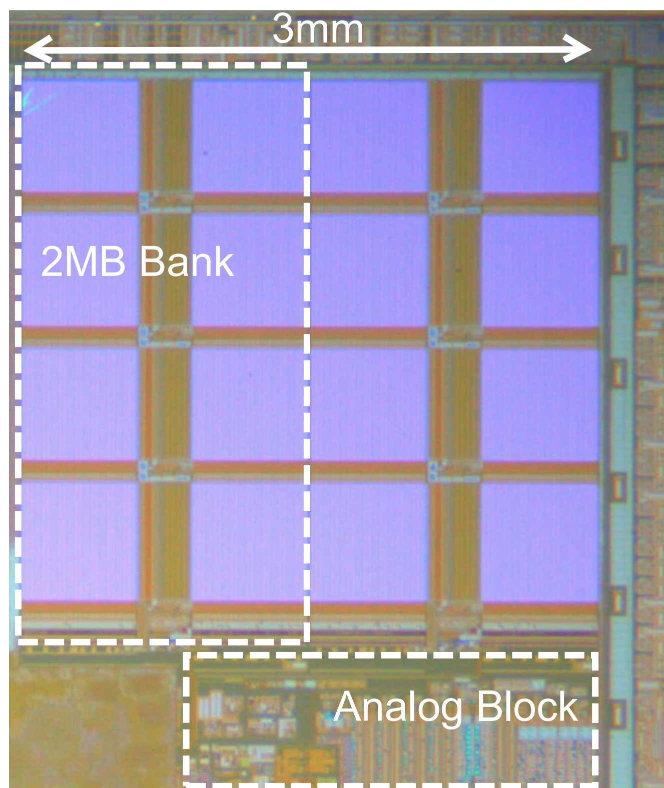

Figure 5.1: 4MB Memory Macro Chip Micrograph

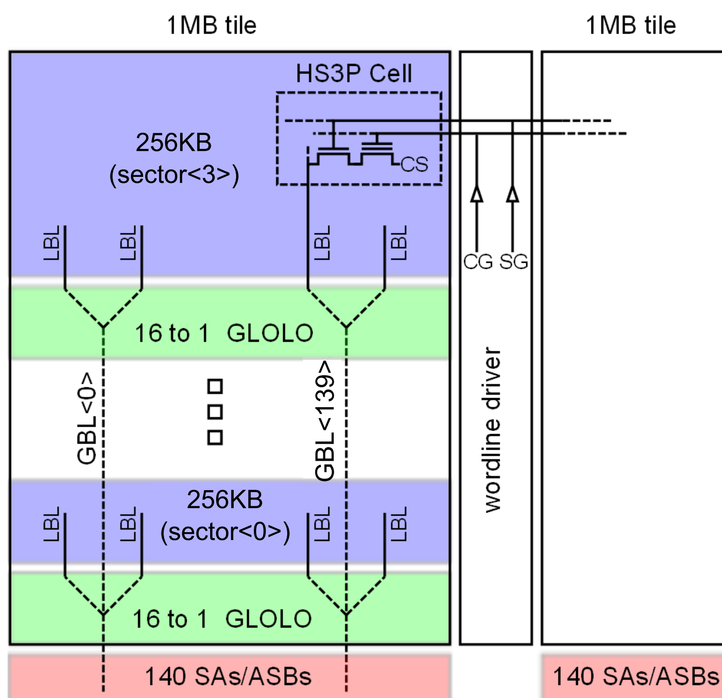

Figure 5.2: 2MB Memory Array Architecture

is realized with I/O devices which are smaller and faster compared to the high voltage transistors used in the CG-path. This allows sub 5ns wordline (SG) rise time with low dynamic power consumption since the CG-path is statically biased during the whole read operation.

To reduce the RC-delay of the bitline the hierarchical bitline architecture (global to local switches) [9] has to be used as illustrated in figure 5.2. The 1MB tile contains four 256KB physical sectors with corresponding global to local multiplexers ($GLOLO$s). All cells in one sector share the common source (CS) line. 1024 HS3P cells are connected to one local bitline (LBL). The $GLOLO$ multiplexes 16 LBLs to one global bitline (GBL). The 140 GBLs are directly connected to the 140 sense amplifiers (SAs) and assembly buffer latches (ASBs) required for the write operation.

5.3 Multi Voltage Domain Global to Local Multiplexer

Different from the conventional approach (see figure 3.10) based on a previous design [9] a multi voltage domain global to local multiplexer is implemented in this work to enable a low voltage bitline path to decrease the area and power consumption of the macro [2, Jefremow]. Figure 5.3 shows a simplified schematic of the bitline path. The $GLOLO$ com-

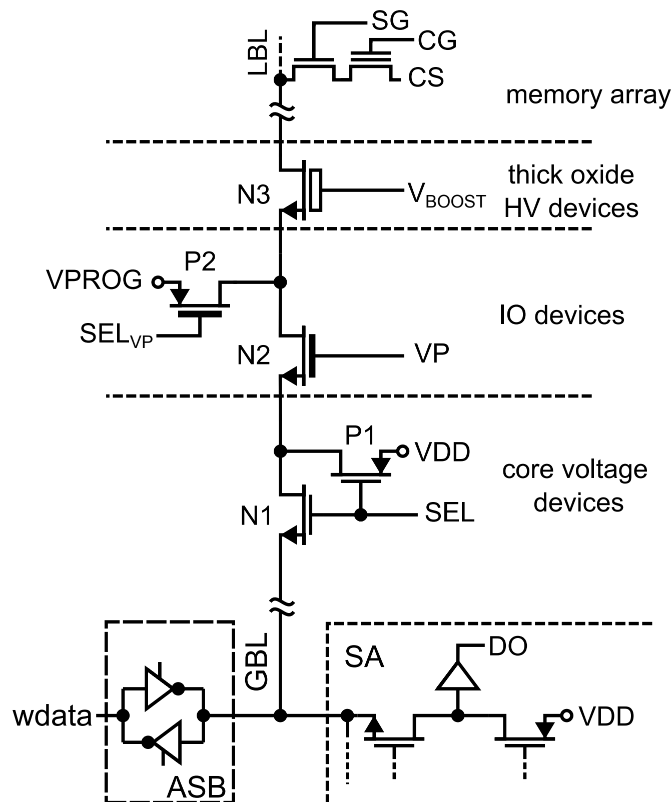

Figure 5.3: Multi Voltage Domain Global to Local MUX

Table 5.1: biasing conditions for figure 5.3

	READ	WRITE	ERASE
CS	V_{DD}	MV	HV
LBL	$\leq V_{DD}$	V_{SS}/MV	HV
V_{BOOST}	HV	HV	V_{SS}
VPROG	V_{DDVIO}	MV	V_{DDVIO}
SEL_{VP}	V_{DDVIO}	MV/V_{SS}	V_{SS}
VP	V_{DDVIO}	V_{DD}	V_{SS}
SEL	V_{DD}/V_{SS}	V_{DD}/V_{SS}	V_{SS}

prises three different kinds of transistors for hierarchical multi voltage domain switching. The thick gate oxide high voltage transistor N3 separates the high voltage cell array domain and the write voltage domain $VPROG$. The core voltage V_{DD} domain is separated from the $VPROG$ domain by the I/O device. The biasing conditions for the different operation modes in the memory array are summarized in table 5.1. During read and write operation transistor N3 is continuously boosted by high voltage (HV) to minimize its on-resistance, since a high read path resistance heavily degrades the read access time (see section 3.4). During write operation N2 is continuously biased by V_{DD} protecting the core devices from the medium voltage (MV) of the $VPROG$ domain. By this means the core devices N1 and P1 are used to select (SEL=V_{DD}, LBL=GBL) or deselect (SEL=V_{SS}, LBL=MV) the LBL during the write operation. The write data is provided by the wdata signal to the assembly buffer latch. For wdata='1' the selected LBL is pulled to ground and the programming current flows through the corresponding $HS3P$-cell (source-side injection programming). After 5us the write operation is completed and the cell is deselected by setting SEL to V_{SS}. In case wdata='0' the voltage of the selected LBL remains unchanged and the corresponding cell is excluded from write operation.

In read mode N2 is biased by V_{DDVIO}=3.3V (I/O supply voltage) to minimize its on-resistance. The GBL selection is again performed by N1 and P1. Thus, contrary to the conventional approach (see figure 3.10) based on a previous design [9] the core domain devices are used for selection in read and write operation and no assembly buffer and sense amplifier circuitry in the I/O voltage domain is required.

5.4 Local Ground Referenced Read Circuit Design

To further reduce the power and area consumption of the bitline path compared to a conventional approach (see figure 3.10) based on the previous design [9] and increase the read performance the bitline-capacitance-cancellation sensing scheme (see section 4.6) is utilized. The proposed flash macro contains 560 SAs which can operate simultaneously and therefore generate supply noise [2, Jefremow]. In addition the microcontroller comprises a variety of high speed digital (e.g. multi-core architecture) and analog building blocks (e.g. multiple ADCs and DACs) adding significant supply noise too. To cope with the ground noise of the SoC the local ground referenced read circuits are implemented as shown in figure 5.4 [2, Jefremow]. Each group of 140 sense amplifiers (SAs) require one reference sense amplifier $REFSA$ (see figure 5.4), and all sense amplifiers are built identically. The 2MB bank comprises one discharge time generator $DISGEN$ (see figure 5.4) utilizing the same common gate amplifier structure (N1, N2 and P1, P2 in figure 5.4). All circuits used in read operation have the same basic structure and therefore have the same characteristics with respect to system noise. The timing diagram of the read phase is shown in figure 5.5. With an address change the address transition detection ATD (see figure 5.4) drives the signal PREINT to ground starting the integration of the current I_{INT} on the C_{INT} cap at node GBL_{REP}. Simultaneously the $DISGEN$ control logic (see figure 5.4) sets the signal PRESA to V_{DD} starting the discharge of the C_{GBL} caps of the GBL and GBL_{REF} nodes. When GBL_{REP} reaches V_{INT} potential the PRESA signal is set to ground starting the integration phase at the C_{GBL} caps on the GBL and GBL_{REF} nodes. If a low resistive cell is accessed, the GBL node will reach the V_{sense} voltage level earlier with respect to the GBL_{REF} indicating the '0' cell state. In case the '1' cell state is read, the digital output DO remains unchanged. A low voltage swing on the GBL is required for high speed read operation (V_{sense} has to be as low as possible). Thus, the system's ground noise becomes the limiting factor for the read speed. The common gate structure (N1, N2 and P1, P2) shows an excellent power supply rejection ratio ($PSRR$) with respect to ground noise if the gate terminal of the two NMOS transistors is referenced to ground. To increase the $PSRR$ an additional capacitance C_{VREF} (see figure 5.4) is applied in all common gate circuits. By this means every common gate structure is referenced to its own local ground. This allows V_{sense} to be set only 50mV above the ground level.

5.5 Measurements and Sensing Schemes Comparison

To measure the system's supply noise we applied the worst case power pattern to the microcontroller. The measurements were performed on wafer level to further increase the supply noise due to the additional inductance of the needle card. The measured V_{DD} and ground (V_{SS}) levels are shown in figure 5.6. The measured voltage level is reduced by about 20% due to 50Ω input resistance of the measurement setup. The voltage difference V_{DD}-V_{SS} indicates a high system noise level of more than 30mV in less than 20ns. Nevertheless the read operation shows no fails in the flash macro for the applied noise pattern even

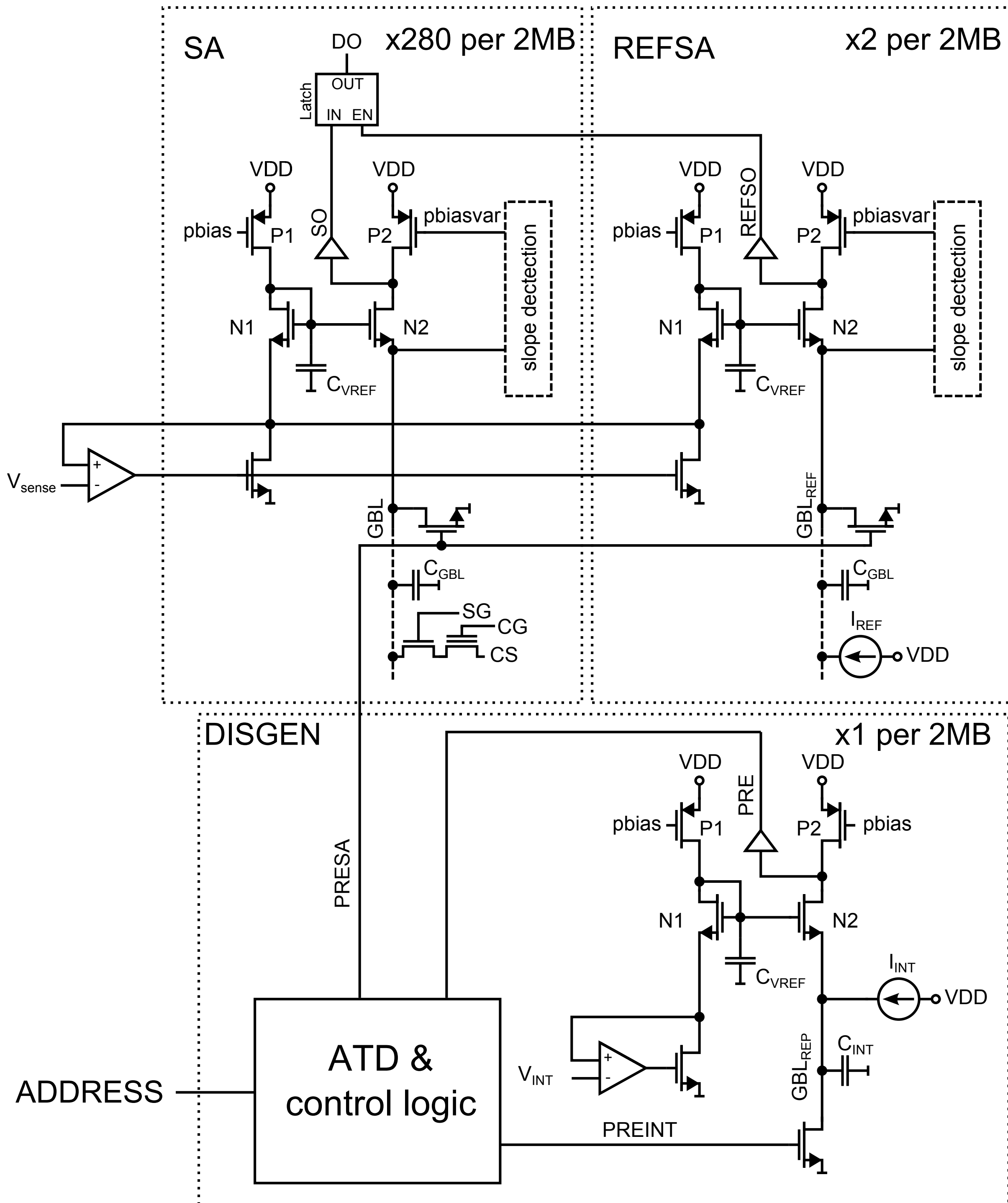

Figure 5.4: Local Ground Referenced Read Circuits

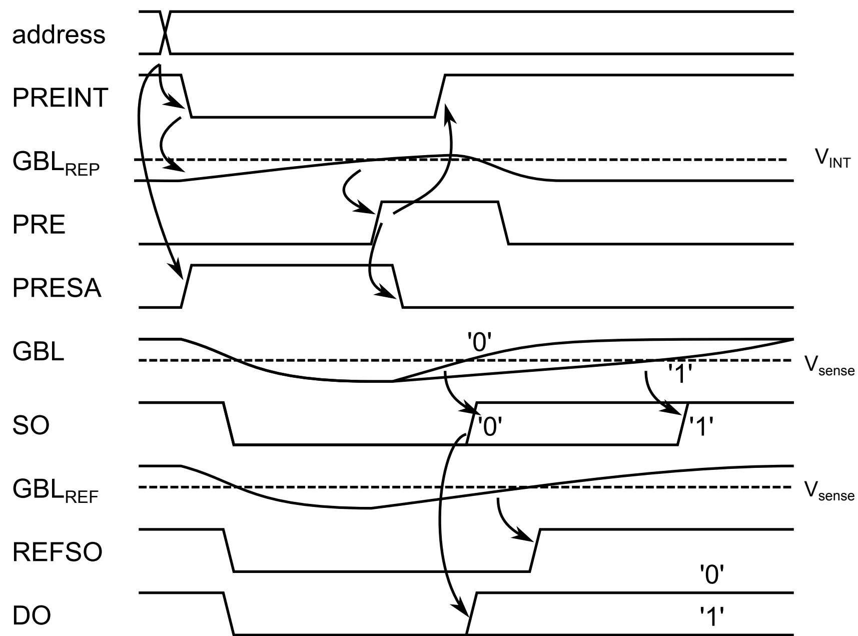

Figure 5.5: Read Access Timing on '0' and '1' Cell State (ref. to Fig. 5.4)

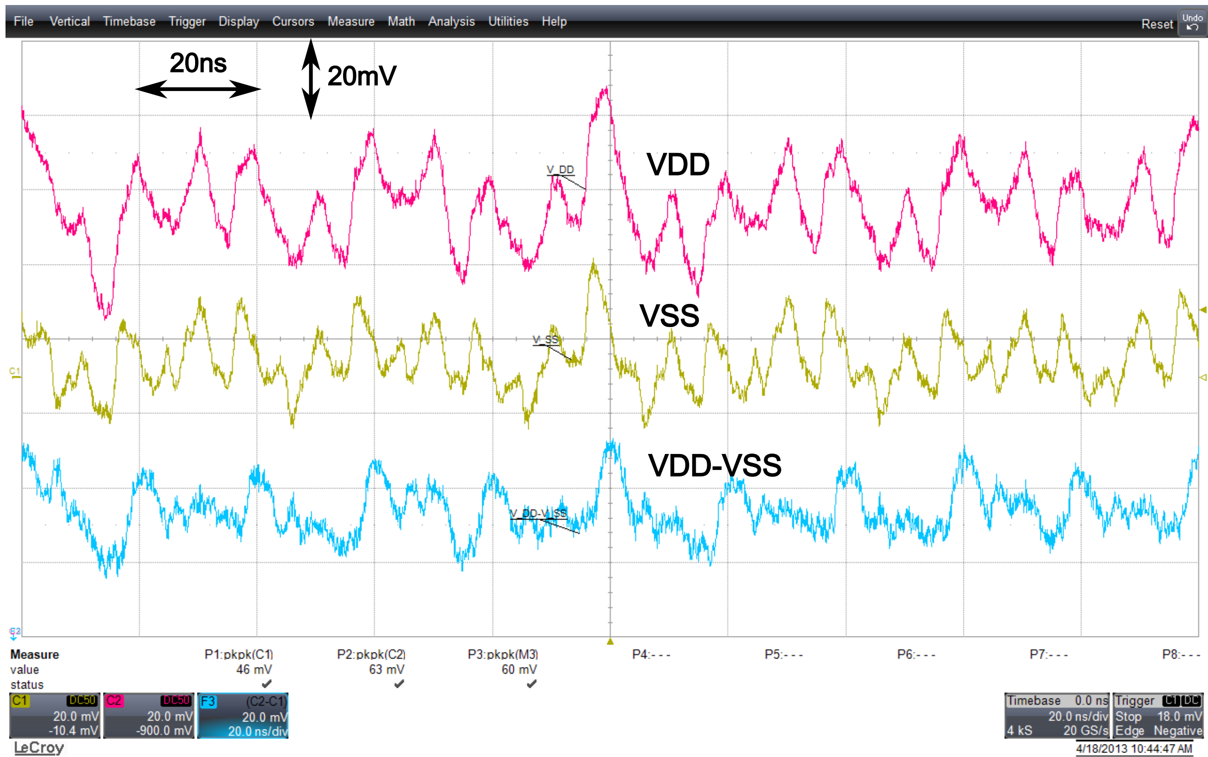

Figure 5.6: System Noise Measurement

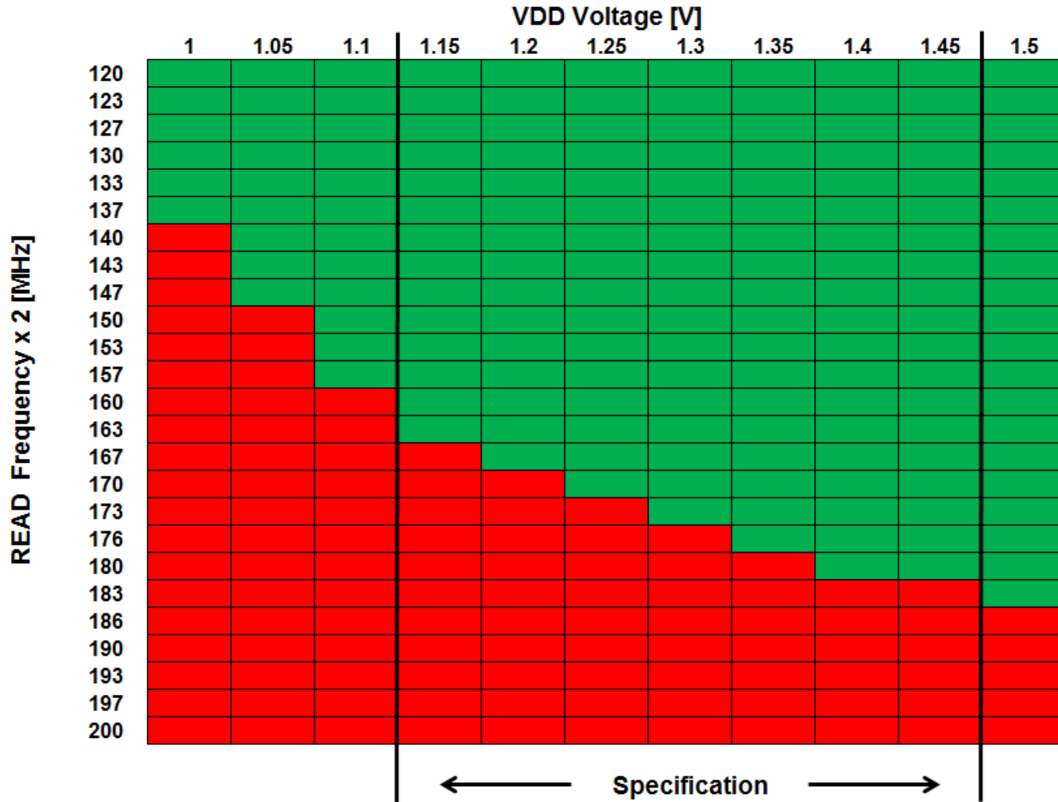

Figure 5.7: Read Access Time Measurement at 170°C

when lowering V_{sense} to only 50mV above ground.

The read access time measurement performed at the worst case temperature of 170°C for V_{sense}=50mV is shown in figure 5.7. For the worst case V_{DD} supply voltage of 1.15V, lowered by more than 10% with respect to the nominal value of 1.3V, a read access time of 2/163MHz≈12.3ns is achieved. By reading the 560 SAs in parallel a read throughput of 5.7GByte/s is achieved.

Table 5.2 compares this work to the previous design in [9] and the recently published one [8]. The proposed flash macro achieves 1.4MByte/s write throughput which is about one order of magnitude higher compared to previous designs [8], [9].

For a comparison of the sensing schemes a figure of merit (FOM) for sensing performance $FOMSP$ is needed. The sensing circuits of this three designs have different bitline capacitances and therefore different bitline RCs. Based on a common FOM in analog design, scaling the gain-bandwidth product GBW of an operational amplifier with its load capacitance [33], the read access time is scaled to the number of cells on the global bitline (GBL width) as it is directly proportional to the GBL capacitance. The sense amplifier design allowing a higher GBL width for a specified read access time is more area efficient or it is faster for a specified GBL width. The $FOMSP$ in the proposed design is:

$$FOMSP = READ\ Frequency \cdot GBL\ width = 81.5MHz \cdot 4KB = 326MHzKB \quad (5.1)$$

Table 5.2: Design Comparison

	This Work	Previous Design [9]	ISSCC 2013 [8]
Process	65nm	130nm	40nm
Cell Type	HS3P-FLASH	UCP-FLASH	SG-MONOS
Capacity (code flash)	4MB	2MB	4MB
SA Supply	1.3V	3.3V	1.25V
FOMSP	326MHzKB	172MHzKB	40MHzKB
Operating Temp.	-40°C/170°C	-40°C/150°C	-40°C/170°C
Read Throughput	5.7GB/s	2GB/s	5.1GB/s
Write Throughput	1.4MB/s	≈0.1MB/s	≈0.17MB/s[40]
Rewrite Time	3.5s/2MB	≈70s/2MB	≈25s/2MB

The proposed multi voltage domain global to local multiplexer scheme allows the usage of core voltage devices for accessing the memory cells in read and write mode, decreasing the area and power consumption of the bitline multiplexer and the assembly buffer latch. To further reduce power consumption the time domain source side sense amplifier scheme is utilized. Figure 5.8 shows the area and power savings due to the multi voltage domain global to local multiplexer scheme. It allows to save more than 57% overall power of the memory macro in read mode and more than 54% of the read path area compared to a conventional approach (see figure 3.10) based on a previous design [9], which was fabricated in the same 65nm technology comprising the $HS3P$-cell.

The proposed local ground referenced circuit design technique is used in all performance critical read circuits to further increase the performance and robustness of the sensing scheme. Compared to the demonstrator test chip (see section 4.6) the V_{sense} voltage is lowered from 157mV to 50mV. The measurement performed on the memory macro reveals 38% higher read frequency compared to the read frequency for 150mV V_{sense}.

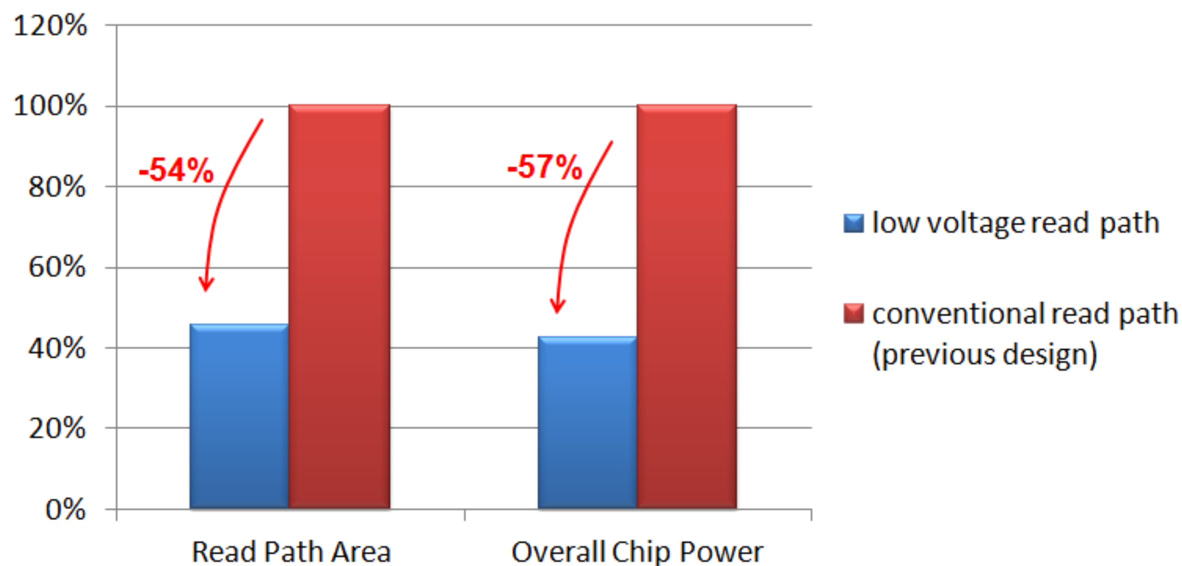

Figure 5.8: Area and Power Saving due to Multi Voltage Domain Global to Local Multiplexer Scheme utilizing the Bitline-Capacitance-Cancellation Sensing Scheme

The 4MB embedded flash macro fabricated in 65nm technology outperforms the previous design [9] by more than a factor of 2 higher read throughput even with reduced power and area consumption. It achieves 5.7GByte/s read throughput which is more than 10% faster with respect to a comparable design (see table 5.2). The figure of merit for the sensing performance $FOMSP$ is more than 80% higher compared to previous design and more than 8 times higher with respect to the recent publication [8].

Chapter 6

Sensing for Embedded STT-MRAM

This chapter discusses the design of sense amplifier circuits for the emerging non-volatile magnetic random access memory MRAM. Compared to Flash based memories MRAM has nearly unlimited endurance and a competitive cell size [41]. In addition MRAM write operation is performed using core voltage domain devices only and the basic MRAM process requires only two additional mask layers (eFlash typically more than ten). This makes it a promising candidate for embedded nonvolatile memory in next generation microcontrollers.

However the field switching based Stoner-Wohlfarth MRAM [42] requires high current (about 10mA) for write operation which consumes high power and chip area in the driver circuits [43],[44]. In addition the Stoner-Wohlfarth MRAM has a half-select read disturb issue, which was relaxed by introducing the Toggle MRAM [45]. The invention of the spin-torque-transfer (STT) MRAM solved the issue of the high write current by exploiting the electron spin polarization for switching the cell magnetization. The drawback of this approach is the read disturb due to the same current flow direction/path in read and write mode [46] and the relatively low resistance difference (smaller than factor two) [47] between high and low resistive cell state.

6.1 Embedded STT-MRAM (eSTT-MRAM)

Figure 6.1 shows the schematic representation of the embedded STT-MRAM cell. The cell comprises the NMOS transistor acting as selecting element controlled by the wordline signal WL. The NMOS transistor is usually a low voltage device with minimal channel length since the read and write operation of the STT-MRAM cell is compatible to low voltage. The magnetic tunnel junction MTJ is located above the select transistor and built by the so called free and fixed layer. The free layer changes its magnetic polarization depending on the programmed cell state, and the fixed layer is used for electron spin polarization. During the read operation the source line SL is set to V_{SS} and V_{DD} is applied to the WL terminal of the selected cell.

Figure 6.2 shows the overall cell resistance with respect to the bitline to source line

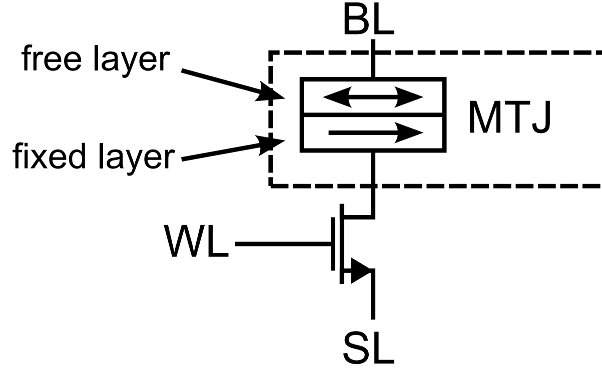

Figure 6.1: Basic STT-MRAM Cell Structure

voltage difference $V_{BL} - V_{SL}$ of the selected eSTT-MRAM cell. For positive V_{BL} to V_{SL}

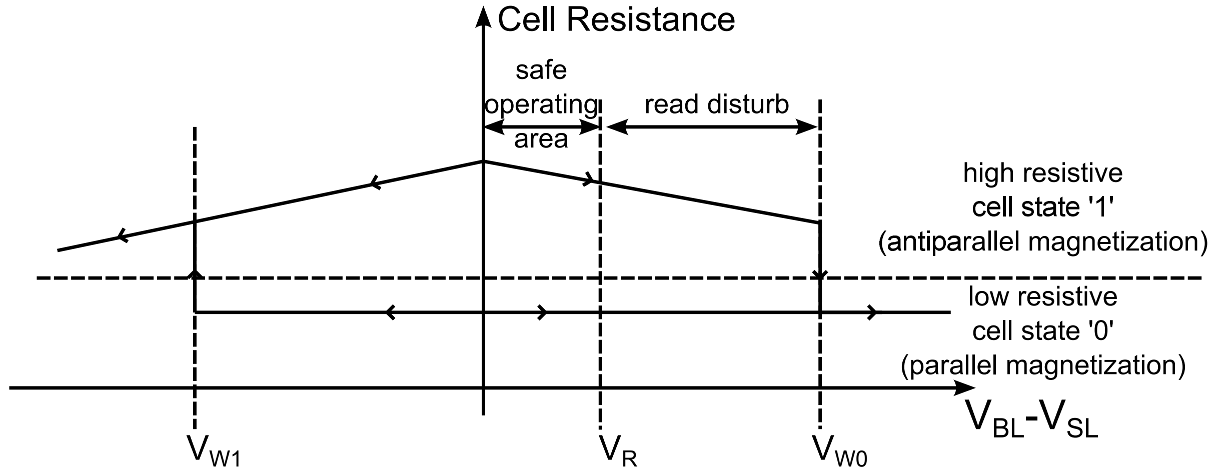

Figure 6.2: Resistance Characteristic of the STT-MRAM Cell

voltage difference the cell resistance of the high resistive cell state is continuously reduced by increasing the bitline voltage due to MTJ resistance reduction. If the bitline threshold voltage V_{W0} is reached, the MTJ element changes its state from anti-parallel (free layer anti-parallel to fixed layer) to parallel configuration. By further increasing the V_{BL} voltage the STT-MRAM cell resistance remains constant at low resistive cell state. In case the selected cell is in low resistive state it remains unchanged. The difference between read and write operation is only defined by the bitline voltage level assuming V_{SL}=0V. To minimize the read failure rate the V_{BL} voltage has to be as low as possible. The required maximum read failure rate defines the safe operating area for the bitline voltage (V_{BL}-V_{SL}<V_R) during read operation [46], which has to be guaranteed by the sense amplifier circuit.

To switch the cell from low to high resistive state the V_{BL}-V_{SL} must be more negative than V_{W1}. Therefore V_{BL} voltage is set to V_{SS} and V_{SL} has to be higher than the absolute value of voltage threshold $|V_{W1}|$. In case the selected cell is in high resistive state it remains unchanged.

The TMR (Tunnel-Magneto-Resistance) ratio is one of the main characteristics of

the eSTT-MRAM cell. It is defined as:

$$TMRratio = \frac{R_{MTJ,MAX} - R_{MTJ,MIN}}{R_{MTJ,MIN}} \tag{6.1}$$

Typical values for the TMR ratio of the MTJ element are in the range of 100% [41]. This means the resistance difference of the MTJ between a low and a high resistive cell state is only a factor of two. In contrast, the embedded Flash cell has the ratio of about 20 ($I_{CELL'0'}$=20μA to $I_{CELL'1'}$=1μA). By taking into account the on-resistance of the select transistor and the series resistance of the bitline and source line the effective resistance difference between a low resistive and a high resistive embedded STT-MRAM cell is well below the factor of two [47].

The reported resistance values [47] in the bitline path show only about 32% relative resistance change of the bitline path resistance. Assuming a current sensing scheme applying 100mV bitline voltage during the read operation, the sense amplifier has to distinguish between 6.6μA and 4.5μA. By defining the reference level in the middle of the read window which is 5.55μA the sense amplifier has to distinguish a current difference of only 1μA. Taking the local cell variations into account the current difference becomes less than 100nA [47]. Therefore the sense amplifier has to detect current changes of 100nA at operating point of 5.45μA. Compared to eFlash where the sense amplifier has to detect the cell current difference of about 10μA in a 10μA operating point, the eSTT-MRAM requires two orders of magnitude higher accuracy of the sensing scheme.

6.2 State-of-the Art Sensing Schemes

As already explained in the previous section two major drawbacks of the STT-MRAM technology are the small read window because of the low tunnel magnetic resistance (TMR) ratio and the read disturb, which is proportional to the bitline voltage. To maximize the read window and minimize read disturb, the bitline voltage has to be as low as possible, since the TMR ratio is increased (see figure 6.2) and the read disturb is reduced with decreasing bitline potential.

Due to these requirements the conventional voltage sensing scheme is not suitable for the STT-MRAM, since voltage sensing requires a considerable voltage difference between the array bitline and the reference bitline for robust read operation (see chapter 3) and the required bitline voltage for read operation is usually below 200mV [48].

The conventional sense amplifier circuit for MRAM memory [48],[49],[44],[45] is shown in figure 6.3. It consists of three PMOS transistors P1, P2, P3 and three NMOS transistors N1, N2, N3. The NMOS transistors are acting as source followers for V_{CLAMP} voltage adjusting the bitline potential. To match the bitline path in terms of capacitance and resistance the reference side comprises at least two reference cells which are located on the same wordline. The source terminals of N2 and N3 are connected together averaging the current of these two reference bitlines. The two reference cells are preconditioned in two different cell states, high and low resistive. Therefore the average current flowing

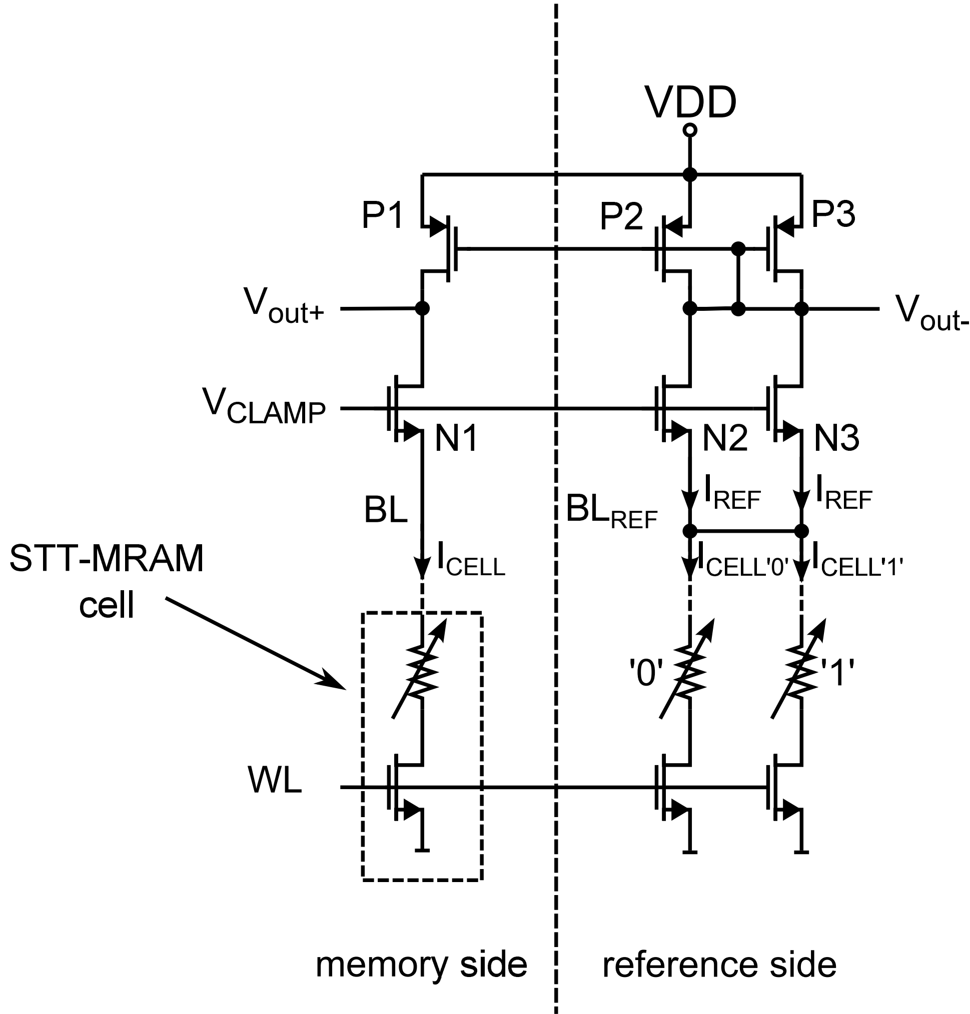

Figure 6.3: State-of-the-Art Sense Amplifier Circuit for *STT*-MRAM

through N2 and N3 is:

$$I_{REF} = \frac{I_{CELL'0'} + I_{CELL'1'}}{2} \tag{6.2}$$

P2 and P3 are diode connected as a current mirror together with P1 mirroring I_{REF} into V_{out+} node. The output node V_{out-} is a low impedance and V_{out+} is a high impedance node. If the cell current I_{CELL} is greater than the reference current I_{REF}, node V_{out+} is discharged and the voltage difference between V_{out+} and V_{out-} becomes negative. It becomes positive if the cell current is lower than the reference current. The outputs nodes V_{out+} and V_{out-} are connected to a latch type amplifier to obtain a full CMOS level signal [49],[45], [46] or even a second amplifier stage can be inserted [48].

The advantage of this circuit is the simple structure, which comprises only two transistors per bitline. Therefore a low count of transistor-pairs, which suffer from mismatch, determines the sense amplifier accuracy. In addition it is suitable for low voltage operation since as already mentioned the required bitline voltage during the read operation is usually below 200mV [48] to fulfill the read disturb requirements. Therefore the PMOS transistors in the current mirror and the NMOS transistors in the source follower have

sufficient voltage headroom.

As already described lowering the bitline voltage reduces the read disturb and increases the read window. The decisive factor for the lower limit of the bitline voltage is the accuracy of the sensing scheme [48]. The read signal, which is usually the cell current, decreases with lower bitline potential. Therefore the accuracy of the sensing scheme is the main design goal for the eSTT-MRAM memory. To be able to sense a 100nA current change in a bias point of several μA all the sources of mismatch have to be minimized in the circuit.

Various approaches have been proposed to improve the accuracy of the sensing scheme. The straight forward approach is to trim the reference current path or the memory path individually (local trimming) by manual adjustment of the transistor sizes [48] or the reference cells used in read operation [50]. This approach leads to a long test time and is not suitable for high sense amplifier counts required for high speed memories. In addition the adjustment of the PMOS current mirror ratio [51],[45] or increasing the number of reference bitlines used for reference current generation [49] may lead to slight accuracy improvement. But the best overall tradeoff in accuracy and cost is achieved reducing the sense amplifier mismatch adaptively. The proposed current-sampling-based sense amplifier technique [23],[22] utilizes sampling technique to cancel the mismatch between the PMOS current mirrors. However this approach was proposed for NOR-Flash memory where no bitline voltage regulation was traded against accuracy increase. This approach is not suitable for eSTT-MRAM memory, since it will lead to read disturb and large operating point variance diminishing the accuracy, since no bitline clamp devices are present. The read disturb can also be reduced by adaptive bitline voltage control [52] using a negative resistance sensing approach. But this requires a complex sense amplifier design and is more vulnerable to variations, which again diminishes the read window.

6.3 Time-Differential Sensing Scheme

6.3.1 Organization

A time-differential sensing scheme developed in this work uses a time-differential sense amplifier to increase the sensing accuracy versus the state-of-the-art sense amplifier circuit (see figure 6.3) [3, Jefremow]. This allows a sub 80mV bitline voltage potential for the STT-MRAM memory.

Figure 6.4 shows the organization of the 8Mbit eSTT-MRAM test macro fabricated in 40nm CMOS technology. It consists of four tiles of 2Mbit each. One tile has 32 sense amplifiers (SAs) that are read out simultaneously and multiplexed to the I/O. The digital portion and the analog block for reference voltage and current generation are located in the center of the macro. Figure 6.5 shows the architecture of one tile. Each of the 32 SAs is connected to 64 multiplexed bitlines and one reference bitline with 1024reference cells. The sense amplifier has a precharge path and an active load configurable for two operation modes. The 32 reference bitlines are connected together on the drain nodes of

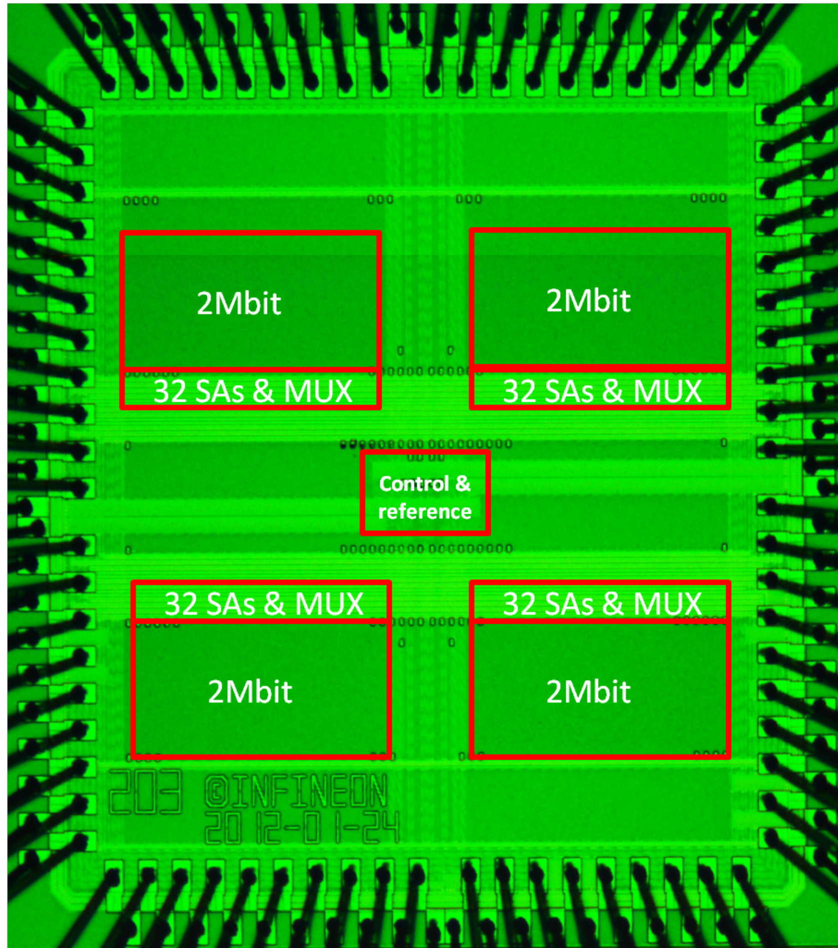

Figure 6.4: Die Photograph of a 8MBit eSTT-MRAM Demonstrator Chip Developed in this Work Fabricated in 40nm CMOS

the cells. The voltage applied to these reference bitlines is regulated by a global regulator and can be adjusted from 80mV to 140mV in 10mV steps. The regulator controls the gate voltage V_{CLAMP} of the NMOS transistors in each sense amplifier (SA). Every SA has a digital latch that amplifies the output voltage difference between V_{out+} and V_{out-} to a full CMOS level and drives the data output.

6.3.2 Time-Differential Sense Amplifier Circuit

Figure 6.6 shows the sense amplifier circuit developed in this work. It is built by two PMOS transistors P1, P2 and two NMOS transistors N1, N2, capacitance C1 and 8 switches S1-S8. All switches except S5 are realized as NMOS transistors with minimal gate length controlled by a digital control logic. The switch S5 is realized as NMOS transistor with a half-sized dummy switch structure [53] to minimize the clock feedthrough and channel charge injection on capacitance C1.

In the default configuration mode, switches S6 and S7 are permanently closed and S8 is open, as shown in Figure 6.6. Thus the capacitance C1 is not active. This configuration

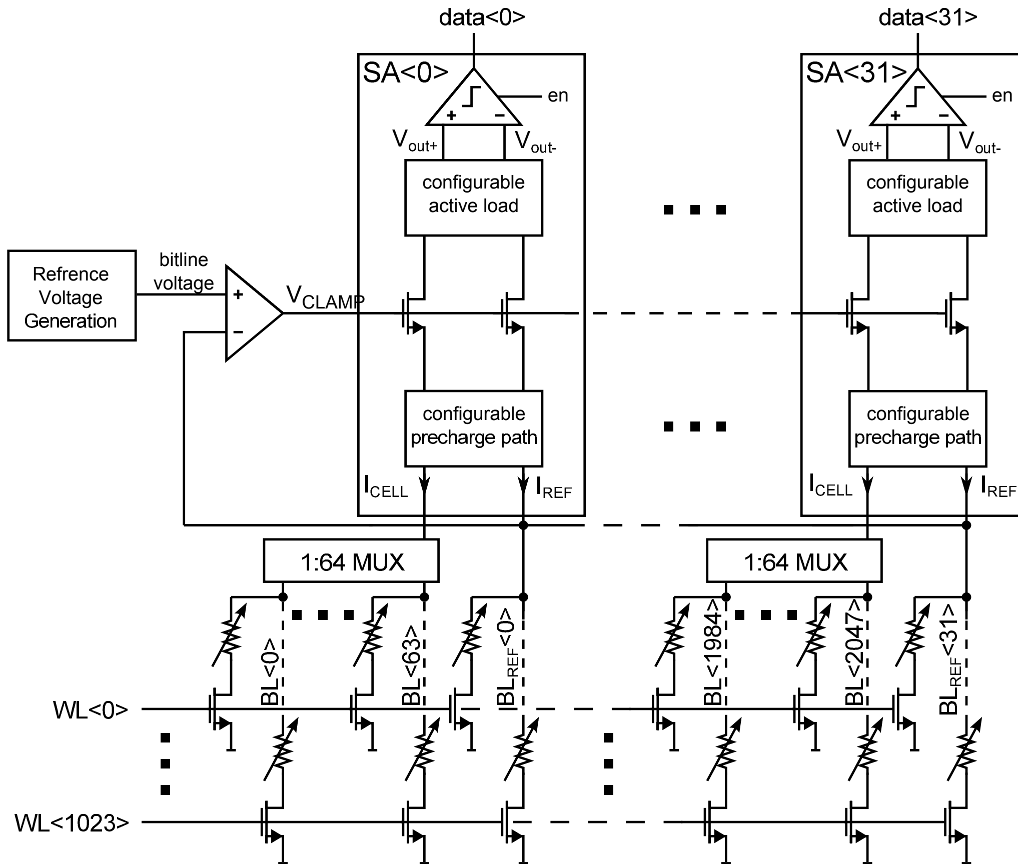

Figure 6.5: Overview of One Memory Tile

corresponds to the conventional sense amplifier configuration as shown in figure 6.3.

Figure 6.9 shows the timing diagram for the sense amplifier in default configuration. During the precharge phase the switches S1-S5 are closed, equalizing the two output nodes V_{out+} and V_{out-} and the source potentials of N1 and N2. The bitline BL and the reference bitline BL_{REF} are connected together. The bitline BL is charged up by N1, N2 and by the charge of the reference bitline BL_{REF}, causing a voltage drop on BL_{REF} due to this charge sharing. The precharge phase is finished as the bitline voltage BL settles.

In the sense phase, switches S2, S3 and S5 are open. The cell current I_{CELL} is flowing through N1 and P1 and the reference current I_{REF} through N2 and P2 as shown in figure 6.6. Transistors N1 and N2 are in cascode configuration for regulating the bitline potential by voltage V_{CLAMP}. P1 and P2 form a current mirror. Output V_{out+} is a high-impedance node compared to node V_{out-}. If the cell current is greater than the reference current, node V_{out+} is discharged and the voltage difference between V_{out+} and V_{out-} becomes negative. It becomes positive if the cell current is lower than the reference current. The accuracy of the SA is determined by the matching of N1, N2 and P1, P2. Increasing the size of the transistors leads to higher accuracy, but the speed degrades due to increased parasitic capacitance. Therefore the time-differential sensing approach developed in this work uses the same circuit elements for sensing of the reference and the signal currents, minimizing matching problems. The mode of the SA (see figure 6.6) can be switched

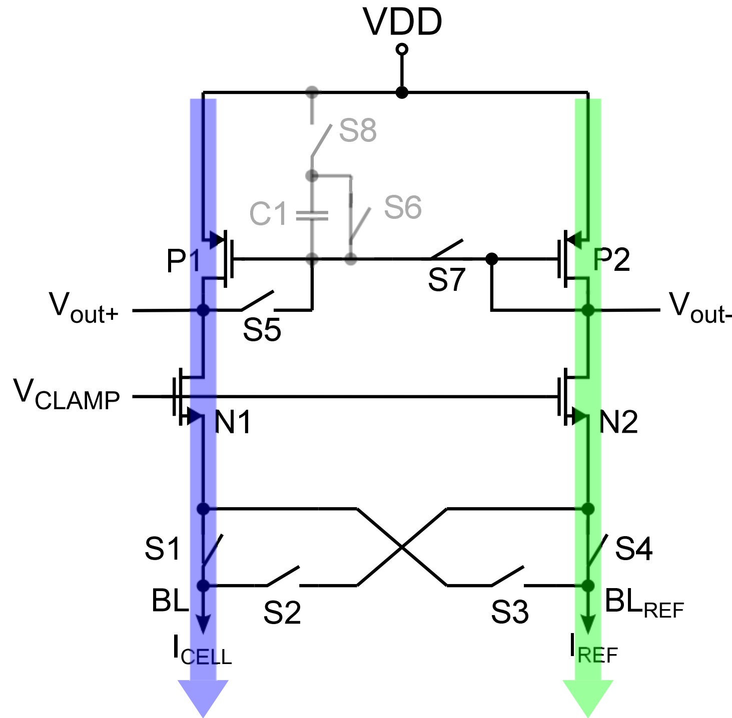

Figure 6.6: Schematic of the *SA* Developed in this Work in Default Configuration in Sense Phase

to time-differential mode without any extra transistor. By this means the two sensing schemes can be compared in terms of accuracy and speed on the same chip using identical transistor sizes.

In the time-differential configuration, shown in figure 6.7, 6.8, the switches S6 and S7 are open and the switch S8 closed. The V_{out-} node is now assigned to the gate of P1 by a multiplexer (not shown in the figure).

Figure 6.10 shows the timing diagram for the sense amplifier in time-differential configuration. This sensing scheme consists of 3 phases: precharge, sample and sense.

During precharge the switches S1-S5 are closed, same as in the precharge phase for the default configuration (see figure 6.7).

In the sample phase, switches S1 and S4 are open and the others remain in the previous state, as shown in figure 6.7. The reference current flows through N1 and P1, the cell current flows through N2 and P2. The gate of P1 is connected to the capacitance C1. Thus the gate-source voltage of P1 as determined by Iref is sampled to C1. The cell path is precharged through P2 and N2.

In the sense phase S5 is opened and the sampled voltage V_{out-} on C1 remains applied to the gate of P1, as shown in figure 6.8. The switches S2, S3 are now opened and S1, S4 closed, connecting P1, N1 to the cell and the P2, N2 to the reference cell. If the cell current is higher/lower than the reference current, the voltage difference between V_{out+} and V_{out-} becomes negative/positive. It becomes positive if the cell current is lower than the reference current. So the cell is read in the sense phase by the same circuit (P1, N1)

as the reference cell in the sample phase. This time-differential sensing scheme almost completely eliminates sense amplifier mismatch. The reference bitline is biased by P2, N2 for the next read cycle.

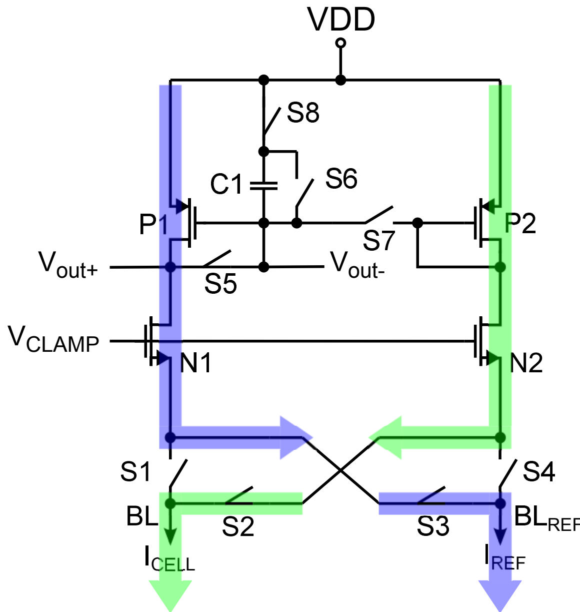

Figure 6.7: Schematic of the SA Developed in this Work in Time-Differential Configuration During Sample Phase

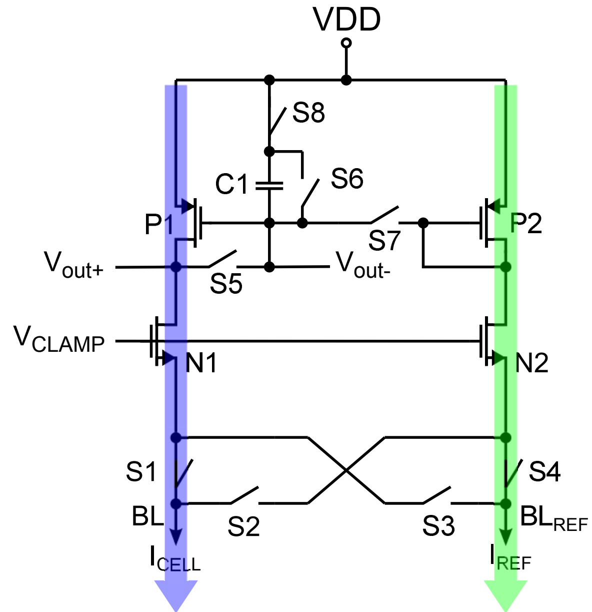

Figure 6.8: Schematic of the SA Developed in this Work in Time-Differential Configuration During Sense Phase

6.4 Sensing Scheme Comparison and Measurement Results

To measure the accuracy of the sensing scheme the signal-margin-screening approach [48] is used, measuring the optimum bitline voltage for read operation. There are two limiting factors for the bitline voltage. The upper threshold is determined by the TMR ratio reduction with increasing bitline voltage (see figure 6.2). The lower threshold is determined by the sense amplifier accuracy, since the read signal, which is the cell current, is reduced with decreasing bitline potential. Figure 6.11 shows the measured weak read bits versus bitline voltage of the 8Mbit macro. The weak read bit fails are caused by the memory cells, which cell to the reference current difference is smaller than the sense amplifier accuracy. All bit counts are plotted relative to the measured value for the time-differential configuration at 80mV bitline potential. The optimum value for the default sensing scheme is 130mV, and the count of weak read bit fails rapidly increases for lower bitline voltages. In contrast, for the time-differential sensing scheme the weak read bit fails are reduced as the bitline voltage is lowered. Even for 80mV, which is the lowest setting for the bitline voltage, a minimum fail count is achieved. Compared to

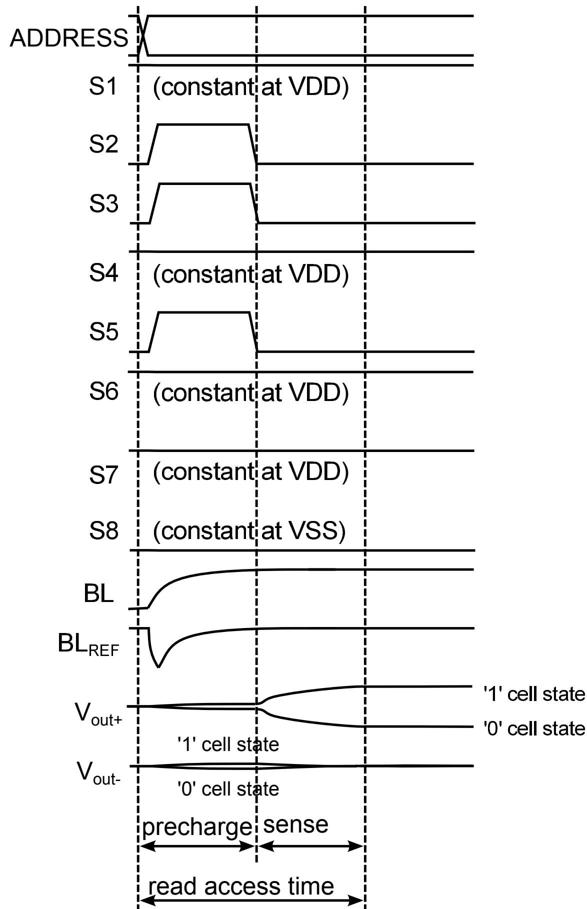

Figure 6.9: Timing for the SA Developed in this Work in Default Configuration (see figure 6.6)

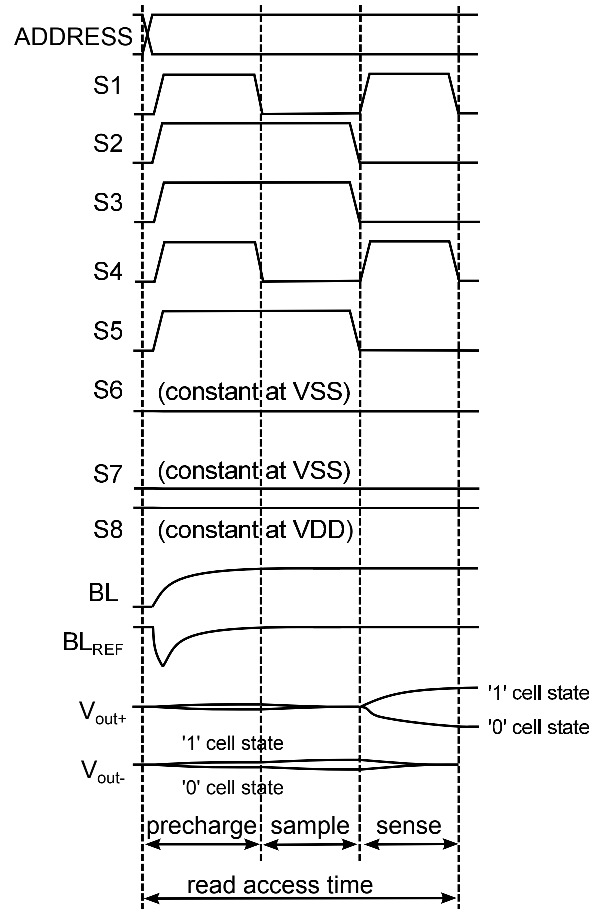

Figure 6.10: Timing for the SA Developed in this Work in Time-Differential Configuration (see figure 6.7, 6.8)

the default configuration the time-differential approach results at 80mV in a read yield improvement, confirming the higher accuracy of this sensing scheme compared to the default configuration. The read-access time shmoo of the two sensing schemes is plotted in figure 6.12. The weak read bit fails, which are mostly independent of the read access time, are not taken into account for this measurement. The higher sensing accuracy of the time-differential sensing comes with a slightly reduced sensing speed, since it requires three sensing phases. For a nominal supply voltage of 1.1V, the sensing speed is 23ns, which is 1ns longer compared to the default (state-of-the-art) configuration.

6.5 Summary

The nearly unlimited endurance and competitive cell size makes STT-MRAM a promising candidate for non-volatile memory in future microcontrollers. But to overcome the drawbacks like the small read window and read disturb the sensing scheme has to provide highest possible read accuracy in combination with bitline voltage regulation.

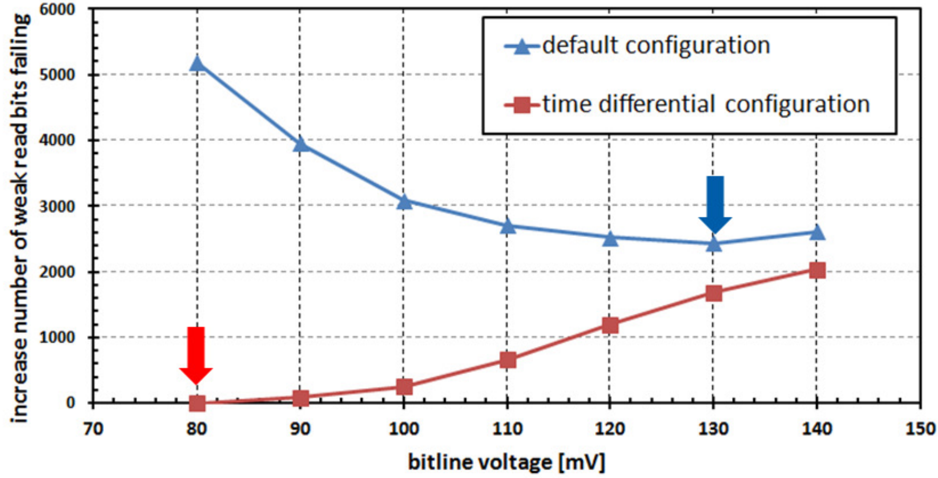

Figure 6.11: Measured Weak Read Bits Versus Bitline Voltage of the 8Mbit Macro

SA type	VDD	27ns	26ns	25ns	24ns	23ns	22ns	21ns	20ns
default	1V	pass	pass	pass	pass	pass	pass	fail	fail
default	1.1V	pass	pass	pass	pass	pass	pass	fail	fail
default	1.2V	pass	pass	pass	pass	pass	pass	pass	fail
time differential	1V	pass	pass	fail	fail	fail	fail	fail	fail
time differential	1.1V	pass	pass	pass	pass	pass	fail	fail	fail
time differential	1.2V	pass	pass	pass	pass	pass	pass	pass	fail

Figure 6.12: Read Access Time Shmoo of the Proposed Sensing Scheme for Two Different Configuations

The time-differential sensing scheme developed in this work overcomes the limitations of the state-of-the-art sense amplifier circuit by exploiting the same devices for reading the array and the reference path.

The measurement results of an 8Mbit STT-MRAM macro prove that the differential-sensing concept developed in this work achieves a bitline voltage reduction of more than 38% compared to the state-of-the-art concept, allowing sub-80mV-BL-potential operation together with significant read yield improvement and with low impact on sensing speed. In addition the power consumption is reduced compared to the state-of-the-art scheme, due to the bitline voltage reduction. At a 1.1V supply voltage the measured access time for the sensing scheme developed in this work is 23ns.

Chapter 7

Conclusion and Outlook

In this work the voltage, current and time domain sensing in dependance of the relevant NOR-Flash parameters like bitline RC-values and cell current were analyzed and evaluated. The current sensing is very sensitive on multiplexer resistance and its speed is strongly decreasing for high multiplexer resistance. Therefore it is less sensitive on cell current range compared to voltage sensing. For the typical NOR-Flash bitline RC-values and cell current parameters the voltage sensing is preferable for high speed and the current sensing for low power applications. The major drawback of voltage sensing is the strong dependence on the cell output characteristics and it is not suitable for all cell types.

The time domain sensing scheme which is based on a voltage sensing approach overcomes the cell characteristic dependence of voltage sensing by a defined bitline voltage level threshold. In addition it allows significantly lower power consumption due to the robust global reference time generator approach. The novel design approach developed in this work [1, Jefremow] takes into account the bitline slope information and was implemented in the bitline-capacitance-cancellation sense amplifier circuit for time domain sensing. It provides lower power consumption compared to voltage or current sensing also for low speed and low power applications by utilizing only the common gate amplifier stage. The bitline-capacitance-cancellation sense amplifier allows adjustable access time with respect to the bitline capacitance and overcomes the physical limit for the state-of-the-art current integrating sensing schemes.

The 65nm HS3P eFlash memory macro [2, Jefremow], which was developed in this work utilizes the bitline-capacitance-cancellation circuit achieving best in class read throughput of 5.7GByte/s. The multi voltage domain global to local multiplexer and the local ground referenced read circuit design developed in this work allows best in class figure of merit for the performance of the sensing scheme. By comparing two memory macros fabricated in the same technology the proposed macro design saves 54% of read path area and 57% overall chip power in read mode compared to the existing current sensing approach.

There is still room for improvement to further reduce the area and power consumption of the proposed sensing scheme and the sense amplifier circuit respectively. For example the bias circuitry for the common gate sense amplifier (see figure 4.39) and the sensing

scheme interaction with the microcontroller core offers potential to further reduce the power and area of the memory macro.

The ongoing scaling and associated issues for productivity increase encourages the development of alternatives for the Flash based memory types due to limited scaling capability of the charge based memories [10]. A promising candidate for future embedded memories is the emerging *STT*-MRAM technology, which offers nearly unlimited endurance and competitive cell sizes. The novel time-differential sensing scheme developed in this work [3, Jefremow] addresses the primary weaknesses of the *STT*-MRAM technology like read disturb and small read window by allowing sub 80mV bitline voltage during the sense operation due to a high precision sense amplifier circuit with minimum area overhead. The embedded *STT*-MRAM macro developed in this work [3, Jefremow] was fabricated in 40nm CMOS technology. The measurements reveal significant read yield improvement and power reduction due to higher sensing accuracy of the time-differential sensing circuit compared to a state-of-the-art sense amplifier.

Another promising candidate for future non-volatile memories is the resistive RAM (*RRAM*) [54]. *RRAM* has lower endurance compared to *STT*-MRAM but it is suitable for a multilevel cell approach [55] and offers very competitive cell sizes due to cross point array architecture [56]. Hoverer this memory type has nearly the same drawbacks as the *STT*-MRAM technology like small read window due to wide resistance distribution and read disturb. Therefore the proposed time-differential sense amplifier fits also very well to this cell type but has to be further optimized to support cross point array structure and the respective parasitic effects such as sneak current.

Appendix A

Author Publications

[1] M. Jefremow, T. Kern, U. Backhausen, C. Peters, C. Parzinger, C. Roll, S. Kassenetter, S. Thierold, and D. Schmitt-Landsiedel, "Bitline-capacitance-cancelation sensing scheme with 11ns read latency and maximum read throughput of 2.9 GB/s in 65nm embedded Flash for automotive," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2012, pp. 428–430.

[2] M. Jefremow, T. Kern, U. Backhausen, J. Elbs, B. Rousseau, C. Roll, L. Castro, T. Roehr, E. Paparisto, K. Herfurth *et al.*, "A 65nm 4MB embedded Flash macro for automotive achieving a read throughput of 5.7 GB/s and a write throughput of 1.4 GB/s," in *Proc. European Solid-State Circuits Conference*. IEEE, 2013, pp. 193–196.

[3] M. Jefremow, T. Kern, W. Allers, C. Peters, J. Otterstedt, O. Bahlous, K. Hofmann, R. Allinger, S. Kassenetter, and D. Schmitt-Landsiedel, "Time-differential sense amplifier for sub-80mV bitline voltage embedded STT-MRAM in 40nm CMOS," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2013, pp. 216–217.

Appendix B

Slope Detection Transfer Function

To calculate the transfer function for i_{CF} of the slope dependent current source (see figure 4.13) the Kirchhoff Law is applied to node v_{CF} in the equivalent circuit for small-signal operation (see figure 4.14):

$$i_{CF}(s) - g_{m,N2}(v_{in}(s) - v_{CF}(s)) + \frac{v_{CF}(s) + \frac{i_{CF}(s)}{g_{m,P1}}}{r_{ds,N2}} = 0 \tag{B.1}$$

where

$$v_{CF} = \frac{i_{CF}}{sC_F} \tag{B.2}$$

Therefore the equation B.1 now can be expressed as:

$$\begin{aligned} g_{m,N2}v_{in}(s) &= i_{CF}(s) + g_{m,N2}\frac{i_{CF}(s)}{sC_F} + \frac{i_{CF}(s)}{sC_F r_{ds,N2}} + \frac{i_{CF}(s)}{g_{m,P1}r_{ds,N2}} \\ &= i_{CF}(s)\frac{1}{s}\left(s + \frac{g_{m,N2} + \frac{1}{r_{ds,N2}}}{C_F} + \frac{s}{g_{m,P1}r_{ds,N2}}\right) \\ &= i_{CF}(s)\frac{1}{s}\left(s\left(1 + \frac{1}{g_{m,P1}r_{ds,N2}}\right) + \frac{g_{m,N2} + \frac{1}{r_{ds,N2}}}{C_F}\right) \end{aligned} \tag{B.3}$$

With $g_m >> \frac{1}{r_{ds}}$ and $g_m r_{ds} >> 1$ [32], [33], [34], the equation B.3 can be simplified to:

$$g_{m,N2}v_{in}(s) = i_{CF}(s)\frac{1}{s}\left(s + \frac{g_{m,N2}}{C_F}\right) \tag{B.4}$$

yielding the final expression for i_{CF}:

$$i_{CF}(s) = sv_{in}(s)\,g_{m,N2}\left(\frac{g_{m,N2}}{C_F} + s\right)^{-1} \tag{B.5}$$

To obtain the overall transfer function of the slope detection circuit for the source-side sensing (see figure 4.19), the additional transfer function of the PMOS level shifter

has to be considered:

$$\begin{aligned} g_{m,P1}\left(v_{LVLS}\left(s\right)-v_{in}\left(s\right)\right)+\frac{v_{LVLS}\left(s\right)}{r_{ds,P2}}+sv_{LVLS}\left(s\right)C_{par} &= 0 \\ v_{LVLS}\left(s\right)\left(g_{m,P1}+\frac{1}{r_{ds,P2}}+sC_{par}\right) &= g_{m,P1}v_{in}\left(s\right) \\ v_{LVLS}\left(s\right)=\frac{g_{m,P1}}{g_{m,P1}+\frac{1}{r_{ds,P2}}+sC_{par}}v_{in}\left(s\right) & \end{aligned} \quad \text{(B.6)}$$

With $g_m >> \frac{1}{r_{ds}}$ [32], [33], [34], the equation B.6 can be simplified to:

$$v_{LVLS}\left(s\right)=\frac{g_{m,P1}}{C_{par}}v_{in}\left(s\right)\left(\frac{g_{m,P1}}{C_{par}}+s\right)^{-1} \quad \text{(B.7)}$$

The over all transfer function can now be calculated by combining the equation B.5 and B.7:

$$\begin{aligned} i_{CF} &= \frac{i_{CF}\left(s\right)}{v_{LVLS}\left(s\right)}\frac{v_{LVLS}\left(s\right)}{v_{in}\left(s\right)} \\ i_{CF} &= sv_{in}\left(s\right)g_{m,N2}\left(\frac{g_{m,N2}}{C_F}+s\right)^{-1}\frac{g_{m,P1}}{C_{par}}\left(\frac{g_{m,P1}}{C_{par}}+s\right)^{-1} \end{aligned} \quad \text{(B.8)}$$

Appendix C

Acknowledgment

Alles Alte, soweit es Anspruch darauf hat, sollen wir lieben, aber für das Neue sollen wir recht eigentlich leben. (Theodor Fontane)

Die vorliegende Dissertation wurde während meiner Zeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Technische Elektronik der Technischen Universität München unter der Leitung von Frau Professorin Dr. rer. nat. Doris Schmitt-Landsiedel in Zusammenarbeit mit der ATV ATM Abteilung an der Infineon Technologies AG, unter der Anleitung von Doktor Thomas Kern angefertigt.

Mein besonderer Dank gebührt meiner Doktormutter Frau Professorin Schmitt-Landsiedel. Sie gewährte mir die nötigen Freiräume, ermöglichte mir einen regen Ideenaustausch und trat meiner Arbeit stets mit Interesse gegenüber. Das auf dieser Weise geschaffene, ausgewogene Klima war ideal für den Entstehungsprozess meiner Dissertation. Darüberhinaus wurde jede Phase dieser Arbeit von Ihr intensiv und voller Empathie begleitet. Ohne Ihren förderlichen Anregungen sowie Ihre stete Diskussions- und Hilfsbereitschaft wäre das Gelingen dieser Dissertation nicht möglich gewesen.

Mein besonderer Dank gebührt auch Herrn Doktor Thomas Kern, der mich unermüdlich unterstützte, mir die erforderliche Selbständigkeit gewährte und mich, wenn nötig, durch konstruktive Kritik motivierte. Unsere Zusammenarbeit war stets vom Streben nach der besseren Lösung geprägt. Seine menschliche Unterstützung und sein Vertrauen, waren wesentliche Pfeiler bei der Fertigstellung dieser Arbeit.

Ebenso danken möchte ich Herrn Doktor Christian Peters für die immerwährende Möglichkeit, Ihn um Rat zu fragen, seine fachkundige Beratung, seine tatkräftige Unterstützung bei den Mess-Setups der “Dose“ sowie den Laboranalysen.

Im Weiteren gilt mein großer Dank Herrn Jayachandran Bhaskaran, Herrn Doktor Klaus Öttinger und Herrn Rainer Bartenschlager. Sie ermöglichten mir meine Konferenzreisen und traten auch bei anderen organisatorischen Anliegen immer für mich ein.

Außerdem sollen jene Mitarbeiter namentlich erwähnt werden, die für mich eine enorme Unterstützung im Design, beim Layout sowie in der Analyse waren. Zu danken ist

in diesem Zusammenhang Herrn Professor Dr. rer. nat. Ullrich Menczigar, Herrn Christoph Roll, Herrn Ulrich Backhausen, Herrn Doktor Johannes Elbs, Herrn Doktor Thomas Nirschl, Herrn Benoit Rousseau, Frau Stefanie Thierold, Herrn Christoph Parzinger, Herrn Doktor Wolf Allers, Herrn Doktor Karl Hofmann, Herrn Doktor Leonardo Castro, Herrn Doktor Thomas Röhr, Herrn Doktor Jan Otterstedt, Herrn Robert Allinger und Herrn Jörg Weller.

Auch an alle Kollegen und Kolleginnen am Lehrstuhl für Technische Elektronik an der Technischen Universität München und der ATV ATM Abteilung an der Infineon Technologies AG geht mein ausdrücklicher Dank. Die freundschaftliche Atmosphäre, die interessante Diskurse sowie das allseits entgegengebrachte Verständnis waren für mich unerlässlich, insbesondere in der Endphase meiner Dissertation.

Danke auch an meine Eltern Valerij Jefremow und Tatjana Jefremowa, die mir das Ingenieursein in die Wiege gelegt haben.

Zuletzt möchte ich mich bei Herrn Professor Dr.-Ing. Heinrich Klar bedanken, der sich die Zeit genommen hat, meine Arbeit als Zweitprüfer zu begutachten.

Diese Arbeit widme ich meiner Frau Sima und meiner Tochter Raisa

Bibliography

[4] D. Schaefer-Siebert, “Electronic controls between the poles of automotive demands, consumer electronics solutions and development costs,” *ATA & SAE International Conference ”The Convergence of Systems Towards Sustainable Mobility”*, 2012.

[5] H. Hidaka, “Embedded Flash Memory Design,” *International Symposium on VLSI Technology, Systems and Applications*, 2012.

[6] B. Wicht, *Current Sense Amplifiers: For Embedded SRAM in High-performance System-on-a-chip Designs.* Springer, 2003, vol. 12.

[7] M. G. e. a. Joe E. Brewer, *Nonvolatile Memory Technologies with Emphasis on Flash*, J. E. Brewer and M. Gill, Eds. John Wiley & Sons, 2008.

[8] T. Kono, T. Ito, T. Tsuruda, T. Nishiyama, T. Nagasawa, T. Ogawa, Y. Kawashima, H. Hidaka, and T. Yamauchi, “40nm embedded SG-MONOS flash macros for automotive with 160MHz random access for code and endurance over 10M cycles for data,” in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers.* IEEE, 2013, pp. 212–213.

[9] C. Deml, M. Jankowski, and C. Thalmaier, “A 0.13μm 2.125 MB 23.5ns Embedded Flash with 2GB/s Read Throughput for Automotive Microcontrollers,” in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers.* IEEE, 2007, pp. 478–617.

[10] R. Strenz, “Embedded Flash technologies and their applications: Status & outlook,” in *IEEE International Electron Devices Meeting.* IEEE, 2011, pp. 9–4.

[11] H. Hidaka, “Evolution of embedded flash memory technology for MCU,” in *IEEE International Conference on IC Design & Technology.* IEEE, 2011, pp. 1–4.

[12] M. Hiraki, T. Tanaka, Y. Shinagawa, K. Suzukawa, M. Fujito, Y. Kawail, D. Mishina, T. Ohshima, S. Abe, H. Kubota, T. Yamaki, S. Takuma, K. Shiba, K. Kuroda, H. Ohsuga, K. Masujima, and K. Matsubara, “A 3.3V 90 MHz flash memory module embedded in a 32b RISC microcontroller,” in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*, 1999, pp. 116–117.

[13] A. Conte, G. L. Giudice, G. Palumbo, and A. Signorello, “A high-performance very low-voltage current sense amplifier for nonvolatile memories,” *IEEE Journal of Solid-State Circuits*, vol. 40, no. 2, pp. 507–514, 2005.

[14] M.-F. Chang and S.-J. Shen, “A process Variation Tolerant Embedded Split-Gate Flash Memory Using Pre-Stable Current Sensing Scheme,” *IEEE Journal of Solid-State Circuits*, vol. 44, no. 3, pp. 987–994, 2009.

[15] T. Tanzawa, Y. Takano, T. Taura, and S. Atsumi, “Design of a sense circuit for low-voltage flash memories,” *IEEE Journal of Solid-State Circuits*, vol. 35, no. 10, pp. 1415–1421, 2000.

[16] R. Micheloni, L. Crippa, M. Sangalli, and G. Campardo, “The flash memory read path: building blocks and critical aspects,” *Proceedings of the IEEE*, vol. 91, no. 4, pp. 537–553, 2003.

[17] T. Tanaka, H. Tanikawa, T. Yamaki, Y. Umemoto, A. Kato, Y. Shinagawa, and M. Hiraki, “A 512 kB MONOS type flash memory module embedded in a microcontroller,” in *2003 Symp. on VLSI Circuits Digest of Technical Papers*. IEEE, 2003, pp. 211–212.

[18] J. Javanifard, T. Tanadi, H. Giduturi, K. Loe, R. L. Melcher, S. Khabiri, N. T. Hendrickson, A. D. Proescholdt, D. A. Ward, and M. A. Taylor, “A 45nm Self-Aligned-Contact Process 1Gb NOR Flash with 5MB/s Program Speed,” in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2008, pp. 424–624.

[19] D. Elmhurst and M. Goldman, “A 1.8-V 128-Mb 125-MHz multilevel cell flash memory with flexible read while write,” *IEEE Journal of Solid-State Circuits*, vol. 38, no. 11, pp. 1929–1933, 2003.

[20] M. Taub, R. Bains, G. Barkley, H. Castro, G. Christensen, S. Eilert, R. Fackenthal, H. Giduturi, M. Goldman, C. Haid *et al.*, “A 90 nm 512 Mb 166 MHz multilevel cell flash memory with 1.5 MByte/s programming,” in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2005, pp. 54–584.

[21] M.-K. Seo, S.-H. Sim, M.-H. Oh, H.-S. Lee, S.-W. Kim, I.-W. Cho, G.-H. Kim, and M.-G. Kim, “A 130-nm 0.9-V 66-MHz 8-Mb (256K$\times$ 32) local SONOS embedded flash EEPROM,” *IEEE Journal of Solid-State Circuits*, vol. 40, no. 4, pp. 877–883, 2005.

[22] M.-F. Chang, S.-J. Shen, C.-C. Liu, C.-W. Wu, Y.-F. Lin, S.-C. Wu, C.-E. Huang, H.-C. Lai, Y.-C. King, C.-J. Lin *et al.*, “An offset-tolerant current-sampling-based sense amplifier for Sub-100nA-cell-current nonvolatile memory,” in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2011, pp. 206–208.

[23] M.-F. Chang, S.-J. Shen, C.-C. Liu, C.-W. Wu, Y.-F. Lin, Y.-C. King, C.-J. Lin, H.-J. Liao, Y.-D. Chih, and H. Yamauchi, “An Offset-Tolerant Fast-Random-Read Current-Sampling-Based Sense Amplifier for Small-Cell-Current Nonvolatile Memory,” *IEEE Journal of Solid-State Circuits*, 2013.

[24] D. Schinkel, E. Mensink, E. Kiumperink, E. van Tuijl, and B. Nauta, "A Double-Tail Latch-Type Voltage Sense Amplifier with 18ps Setup+Hold Time," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers.* IEEE, 2007, pp. 314–605.

[25] V. B. Rao, "Delay Analysis of the Distributed RC Line," in *Proc. of 32nd Design Automation Conference*, 1995, pp. 370–375.

[26] E. Seevinck, P. J. van Beers, and H. Ontrop, "Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAM's," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 4, pp. 525–536, 1991.

[27] M.-F. Chang, C.-W. Wu, C.-C. Kuo, S.-J. Shen, K.-F. Lin, S.-M. Yang, Y.-C. King, C.-J. Lin, and Y.-D. Chih, "A 0.5 V 4Mb logic-process compatible embedded resistive RAM (ReRAM) in 65nm CMOS using low-voltage current-mode sensing scheme with 45ns random read time," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers.* IEEE, 2012, pp. 434–436.

[28] R.-A. Cernea, L. Pham, F. Moogat, S. Chan, B. Le, Y. Li, S. Tsao, T.-Y. Tseng, K. Nguyen, J. Li *et al.*, "A 34 MB/s MLC Write Throughput 16 Gb NAND With All Bit Line Architecture on 56 nm Technology," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 1, pp. 186–194, 2009.

[29] Y. Sofer, M. Edan, Y. Betser, M. Grossgold, E. Maayan, and B. Eitan, "A 55 mm^2 256 Mb NROM flash memory with embedded microcontroller using an NROM-based program file ROM," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers.* IEEE, 2004, pp. 48–512.

[30] Y. Polansky, A. Lavan, R. Sahar, O. Dadashev, Y. Betser, G. Cohen, E. Maayan, B. Eitan, F.-L. Ni, Y.-H. J. Ku *et al.*, "A 4b/cell NROM 1Gb Data-Storage Memory," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers.* IEEE, 2006, pp. 448–458.

[31] D. Miyashita, R. Yamaki, K. Hashiyoshi, H. Kobayashi, S. Kousai, Y. Oowaki, and Y. Unekawa, "A 10.4 pJ/b (32, 8) LDPC decoder with time-domain analog and digital mixed-signal processing," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers.* IEEE, 2013, pp. 420–421.

[32] P. E. Allen and D. R. Holberg, *CMOS Analog Integrated Circuit Design.* Oxford University Press,, 2002.

[33] W. M. Sansen, *Analog Design Essentials.* Springer, 2006, vol. 859.

[34] B. Razavi, *Design of Analog CMOS Integrated Circuits.* McGraw-Hill Companies, 2003.

[35] D.-S. Byeon, S.-S. Lee, Y.-H. Lim, J.-S. Park, W.-K. Han, P.-S. Kwak, D.-H. Kim, D.-H. Chae, S.-H. Moon, S.-J. Lee *et al.*, "An 8 Gb multi-level NAND flash memory

with 63 nm STI CMOS process technology," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2005, pp. 46–47.

[36] G. Marotta, A. Macerola, A. D'Alessandro, A. Torsi, C. Cerafogli, C. Lattaro, C. Musilli, D. Rivers, E. Sirizotti, F. Paolini *et al.*, "A 3bit/cell 32Gb NAND flash memory at 34nm with 6MB/s program throughput and with dynamic 2b/cell blocks configuration mode for a program throughput increase up to 13MB/s," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2010, pp. 444–445.

[37] S.-H. Chang, S.-k. Lee, S.-J. Park, M.-J. Jung, J.-C. Han, I.-S. Wang, J.-H. Lee, J.-H. Kim, W.-K. Kang, T.-K. Kang *et al.*, "A 48nm 32Gb 8-level NAND flash memory with 5.5 MB/s program throughput," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2009, pp. 240–241.

[38] K. Takeuchi, Y. Kameda, S. Fujimura, H. Otake, K. Hosono, H. Shiga, Y. Watanabe, T. Futatsuyama, Y. Shindo, M. Kojima *et al.*, "A 56-nm CMOS 99-mm^2 8-Gb Multi-Level NAND Flash Memory With 10-MB/s Program Throughput," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 1, pp. 219–232, 2007.

[39] D. Shum, J. Power, R. Ullmann, E. Suryaputra, K. Ho, J. Hsiao, C. Tan, W. Langheinrich, C. Bukethal, V. Pissors *et al.*, "Highly Reliable Flash Memory with Self-Aligned Split-Gate Cell Embedded into High Performance 65nm CMOS for Automotive & Smartcard Applications," in *IEEE 4th International Memory Workshop (IMW), 2012*. IEEE, 2012, pp. 1–4.

[40] *SH7256 Group User's Manual: Hardware Rev.2.00 Mar 2012 Renesas 32-Bit RISC Microcomputer SuperH^{TM} RISC engine Family*, Renesas, http://documentation.renesas.com/doc/products/mpumcu/doc/superh/r01uh0344ej0200_sh7256.pdf.

[41] A. Driskill-Smith, "Latest advances and future prospects of STT-RAM," in *Non-Volatile Memories Workshop*, 2010.

[42] A. Bette, J. DeBrosse, D. Gogl, H. Hoenigschmid, R. Robertazzi, C. Arndt, D. Braun, D. Casarotto, R. Havreluk, S. Lammers *et al.*, "A high-speed 128 Kbit MRAM core for future universal memory applications," in *2003 Symp. on VLSI Circuits Digest of Technical Papers*. IEEE, 2003, pp. 217–220.

[43] D. Gogl, C. Arndt, J. C. Barwin, A. Bette, J. DeBrosse, E. Gow, H. Hoenigschmid, S. Lammers, M. Lamorey, Y. Lu *et al.*, "A 16-Mb MRAM featuring bootstrapped write drivers," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 4, pp. 902–908, 2005.

[44] Y. Iwata, K. Tsuchida, T. Inaba, Y. Shimizu, R. Takizawa, Y. Ueda, T. Sugibayashi, Y. Asao, T. Kajiyama, K. Hosotani, S. Ikegawa, T. Kai, M. Nakayama, S. Tahara, and H. Yoda, "A 16Mb MRAM with FORK Wiring Scheme and Burst Modes," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*, 2006, pp. 477–486.

[45] J. J. Nahas, T. W. Andre, B. Garni, C. Subramanian, H. Lin, S. M. Alam, K. Papworth, and W. L. Martino, "A 180 Kbit Embeddable MRAM Memory Module," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 8, pp. 1826–1834, 2008.

[46] R. Takemura, T. Kawahara, K. Miura, H. Yamamoto, J. Hayakawa, N. Matsuzaki, K. Ono, M. Yamanouchi, K. Ito, H. Takahashi, S. Ikeda, H. Hasegawa, H. Matsuoka, and H. Ohno, "A 32-Mb SPRAM With 2T1R Memory Cell, Localized Bi-Directional Write Driver and '1'/'0' Dual-Array Equalized Reference Scheme," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 869–879, 2010.

[47] S. Chung, K.-M. Rho, S.-D. Kim, H.-J. Suh, D.-J. Kim, H.-J. Kim, S.-H. Lee, J.-H. Park, H.-M. Hwang, S.-M. Hwang *et al.*, "Fully integrated 54nm STT-RAM with the smallest bit cell dimension for high density memory application," in *IEEE International Electron Devices Meeting*. IEEE, 2010, pp. 12–7.

[48] H. Honigschmid, P. Beer, A. Bette, R. Dittrich, R. Gardic, D. Gogl, S. Lammers, J. Schmid, L. Altimime, S. Bournat *et al.*, "Signal-Margin-Screening for Multi-Mb MRAM," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2006, pp. 467–476.

[49] J. P. Kim, T. Kim, W. Hao, H. M. Rao, K. Lee, X. Zhu, X. Li, W. Hsu, S. H. Kang, N. Matt, and N. Yu, "A 45nm 1Mb embedded STT-MRAM with design techniques to minimize read-disturbance," in *2011 Symp. on VLSI Circuits Digest of Technical Papers*, 2011, pp. 296–297.

[50] K. Tsuchida, T. Inaba, K. Fujita, Y. Ueda, T. Shimizu, Y. Asao, T. Kajiyama, M. Iwayama, K. Sugiura, S. Ikegawa, T. Kishi, T. Kai, M. Amano, N. Shimomura, H. Yoda, and Y. Watanabe, "A 64Mb MRAM with clamped-reference and adequate-reference schemes," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2010, pp. 258–259.

[51] J. Kim, K. Ryu, S. H. Kang, and S.-O. Jung, "A Novel Sensing Circuit for Deep Submicron Spin Transfer Torque MRAM (STT-MRAM)," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 1, pp. 181–186, 2012.

[52] D. Halupka, S. Huda, W. Song, A. Sheikholeslami, K. Tsunoda, C. Yoshida, and M. Aoki, "Negative-resistance read and write schemes for STT-MRAM in 0.13μm CMOS," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2010, pp. 256–257.

[53] C. C. Enz and G. C. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization," *Proceedings of the IEEE*, vol. 84, no. 11, pp. 1584–1614, 1996.

[54] A. Kawahara, K. Kawai, Y. Ikeda, Y. Katoh, R. Azuma, Y. Yoshimoto, K. Tanabe, Z. Wei, T. Ninomiya, K. Katayama *et al.*, "Filament scaling forming technique and level-verify-write scheme with endurance over 107 cycles in ReRAM," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2013, pp. 220–221.

[55] M.-F. Chang, S.-S. Sheu, K.-F. Lin, C.-W. Wu, C.-C. Kuo, P.-F. Chiu, Y.-S. Yang, Y.-S. Chen, H.-Y. Lee, C.-H. Lien *et al.*, "A High-Speed 7.2-ns Read-Write Random Access 4-Mb Embedded Resistive RAM (ReRAM) Macro Using Process-Variation-Tolerant Current-Mode Read Schemes," *IEEE Journal of Solid-State Circuits*, 2013.

[56] T.-Y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap, J. Ouyang *et al.*, "A 130.7 mm^2 2-Layer 32Gb ReRAM Memory Device in 24nm Technology," in *IEEE Intern. Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 2013, pp. 210–211.