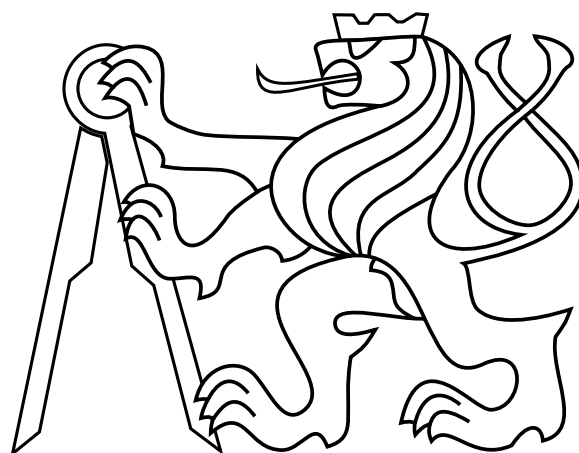CZECH TECHNICAL UNIVERSITY IN PRAGUE
FACULTY OF ELECTRICAL ENGINEERING
DEPARTMENT OF MICROELECTRONICS

DIPLOMA THESIS
Low voltage low power bandgap reference

Author: Jan Škoda
11.května 2015

České vysoké učení technické v Praze
Fakulta elektrotechnická

katedra mikroelektroniky

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Š K O D A Jan

Studijní program: Komunikace, multimédia a elektronika
Obor: Elektronika

Název tématu: **Návrh nízkopříkonové napěťové reference v technologii CMOS**

Pokyny pro vypracování:

1. Make a general research on voltage references structures and its functionality. Focus on low-power CMOS designs.

2. Design selected solution in CMOS 0.18 technology with given parameters:
- Supply voltage 0.9 V - 2 V
- Temperature range -50 °C - 100 °C
- Use inherent bipolar devices in CMOS technology
- Reference shall be started by enable signal
- Capacitive load 0.1 pF - 5 pF
- Focus on power minimization

Seznam odborné literatury:

[1] Gray, Hurst, Lewis, Meyer: Analysis and design of analog integrated circuits.
[2] Johns, Martin: Analog integrated circuit design.

Vedoucí: **doc.Ing. Jiří Jakovenko, Ph.D.**

Platnost zadání: 31. 8. 2015

prof. Ing. Miroslav Husák, CSc.   prof. Ing. Pavel Ripka, CSc.
vedoucí katedry   děkan

V Praze dne 30. 3. 2015

Declaration

I completed my diploma thesis on my own with the contribution of my supervisor and consultants. I used only materials (literature, projects, articles) specified in the attached list.

I agree with the utilisation of the information presented in my thesis, in compliance with copyright § 60 Zákona 121/2000 Sb., (autorský zákon).

Prague, ...
Signature

Acknowledgement

I would like to thank my supervisor Jiří Jakovenko and external consultants Oskar Křenek and Michal Münzberger. This thesis could not be written without their help and consultation.

Abstrakt

Tato práce se zabývá návrhem nízko napěťové, nízko příkonové napěťové reference. Úvodem jsou zde probraný MOS tranzistory a pasivní prvky dostupné v technologii CMOS. Tyto kapitoly tvoří s kapitolou popisující technologické odchylky výroby základ pro lepší pochopení návrhářské problematiky. Dále následuje souhrn několika jednoduchých i složitějších napěťových referencí, včetně reference na principu bandgap.

Hlavním úkolem práce je pak samotný návrh napěťové reference s nízkým napájecím napětím a nízkou spotřebou. V kapitole VI je čtenář postupně seznámen se základními prvky technologie CMOS a je provázen návrhem těchto hlavních částí: jádro napěťové reference, operační zesilovač, startovací obvod a trimovací obvod. Návrh a simulace je provedena v programu Cadence ver. 6.1.3. Během práce bylo třeba se alespoň okrajově seznámit s programovacím jazykem Skill a Verilog pro vytvoření simulačního skriptu a ideálního ADC převodníku.

Abstract

This diploma thesis describes design of a LWLP (low voltage and low power) voltage reference. At the beginning, transistors and passive elements available in CMOS technology are discussed. These two chapters and presented technology variations chapter form good base for a better understanding of IC design issues. This is followed by a summary of a few simple and complex types of voltage references, including the principle of bandgap references.

The main goal of this thesis is designing a low supply voltage and low power consumption voltage reference. In chapter VI, the reader is gradually familiarized with the CMOS basic elements and designing these main parts: voltage reference core, operational amplifier, start-up circuit and trimming circuit. Design and simulation is performed in the program Cadence ver. 6.1.3. During this work it was necessary to set familiar with the programming language Skill and Verilog for making an advanced corner simulation script and for creating the ideal ADC.

Contents

List of Figures

vii

List of Tables

Notations

A_V	Voltage gain
BGR	Bandgap reference
BJT	Bipolar transistor
Cc	Compensation capacitance
$CMOS$	Complementary metal oxide semiconductor
C_{OX}	Silicon oxide capacitance
$CTAT$	Complementary to absolute temperature
Ec	Energy of conduction band
Eg	Energy bandgap
Ev	Energy of valence band
I_D	Drain current
IC	Integrated circuit
Ic	Collector current
IF	Inversion factor
Is	Saturation current
k	Boltzmann constant
L	Channel length
MOS	Metal oxide semiconductor
$NMOS$	Negative metal oxide semiconductor
OTA	Operational transconductance amplifier
PM	Phase margin
$PMOS$	Positive metal oxide semiconductor

$PTAT$	Proportional to absolute temperature
T	Absolute temperature
TC	Temperature coefficient
U_T	Thermal voltage
V_{be}	Base-emitter voltage
V_{dd}	Positive supply voltage
V_{DS}	Drain-source voltage
V_{DSat}	Saturation voltage
V_{GS}	Gate-source voltage
V_{OS}	Offset voltage
V_{ss}	Negative supply voltage
V_{th}	Threshold voltage
V_{ref}	Reference voltage
W	Channel width

1 Introduction

Voltage references are important not only for ADC and DAC or memories. They are the essential part of every chip because circuit needs some biasing that sets proper working conditions.

Requirements of voltage reference are: accurate and constant voltage (independent on temperature and supply voltage) without noise transfer into the system (noiseless), low voltage and low power, preferably independent of the process and mismatch.

It is clear that it is impossible to design circuit with ideal parameters and in the most cases temperature and supply voltage independence are preferred.

Some simple circuits that provide poor voltage reference are later introduced. Despite poor voltage reference accuracy, these circuits can be useful for some applications due to their simplicity. Advanced topology usually exploits thermal voltage or base-emitter voltage etc. They usually contain parasitic diode on that account.

At the beginning the MOS transistor and passive device in CMOS technology are discussed for better insight of complex IC design.

2 MOS Transistor

2.1 Basic Parameters, Operating Region

MOS transistor is the four terminal (Gate, Source, Bulk, Drain) non-linear basic device in electronic design. The origin of short-cut MOS is from Metal-Oxide-Semiconductor structure in the transistor where metal creates a gate terminal, oxide forms insulator (SiO_2) under the gate terminal and semiconductor is P or N-type substrate. Fig.[1] shows simplified cross-section view of NMOS transistor without the Bulk terminal.
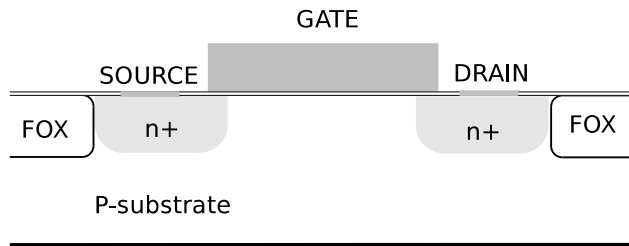
Figure 1: Simplified cross-sectional view of NMOS transistor

NMOS and PMOS transistors are two basic types of MOS transistor. NMOS transistor has an N-type source and drain area and P-type substrate, while PMOS has P-type source/drain and N-type substrate. Fig.[2] shows basic symbols of NMOS and PMOS transistors. The bulk terminal is connected to the negative supply voltage (often Gnd) in the case of the NMOS, while PMOS transistor has bulk terminal connected to the positive supply voltage (often Vdd).

In the next schematics the bulk terminals are not shown for better readability.

MOS transistor has two operating regions depending on V_{DS} and V_{GS} voltage. Linear region and saturation region. Eq.[1] shows condition for each saturation region and Eq.[2] for the linear region. Note that, these equations are valid only for strong inversion.

$$V_{DS} \geq V_{GS} - V_{th} \tag{1}$$

2

Figure 2: MOS transistor schematic symbols used in the thesis (a) NMOS (b) PMOS

$$V_{DS} \leq V_{GS} - V_{th} \tag{2}$$

The behaviour and equation for drain current are different for each region, see Fig.[3]. MOS transistor in the linear region behaves as a resistor, while in the saturation region its behaviour is rather near to a current source. Eq.[3] defines drain current in the saturation region and Eg.[4] in the linear region.

$$I_D = \frac{\beta}{2}(V_{GS} - V_{th})^2 = \frac{1}{2}K'\frac{W_{ef}}{L_{ef}}(V_{GS} - V_{th})^2 \tag{3}$$

$$I_D = \beta\left[(V_{GS} - V_{th}) - \frac{V_{DS}}{2}\right]V_{DS} = K'\frac{W_{ef}}{L_{ef}}\left[(V_{GS} - V_{th}) - \frac{V_{DS}}{2}\right]V_{DS} \tag{4}$$

where
I_D = drain current
$\beta = K'\frac{W_{ef}}{L_{ef}}$ - transconductance parameter
W_{ef}, L_{ef} = effective channel width and length
$K' = \mu_0 C_{ox}$ - technological constant
V_{th} = threshold voltage
V_{DS}, V_{GS} = drain-source and gate-source voltage

3
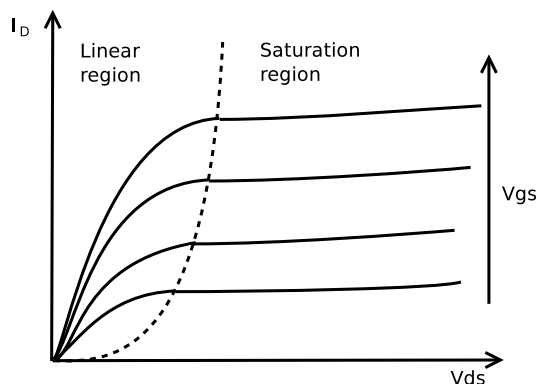
Figure 3: NMOS transistor, output characteristic

2.2 Large and Small Signal Model

The large signal model is important for determination of operation region. Drain current and saturation voltage can be calculated from it. This model is shown on Fig.[4]. It contains four terminals (Gate, Source, Drain and Bulk) and current source I_D. There are other devices in the large signal model. For instance, capacitances (Cgs, Cgd, Cgb, Cdb, Csb), substrate diodes (Dbd, Dbs) and serial resistances rs,rd. All these devices are parasitic. Ideally, only voltage control current source is present.

As mentioned before, MOS transistor is the non-linear device. Therefore, the large signal model is also non-linear. Small-signal model is needed for obtaining the linear characteristic/function. Simply, it is linearisation of the large-signal model in the specific operating point. Due to the linear function all parameters calculations are easier. Fig.[5] shows the principle of small-signal model. MOS transistor is biased by the DC gate-source voltage V_{GS} to the proper value of DC drain current I_D. In this operating point, an AC gate-source signal voltage v_{gs} is applied. The small-signal v_{gs} voltage causes change in the small-signal drain current i_d. As Fig.[5] shows, the small-signal variations are linear if the following conditions are valid (i.e $|v_{gs}| << V_{GS}$ and $|i_d| << I_D$).

Fig.[6(a)] shows the small-signal model. The most important parameter is transconductance g_m. This parameter relates to the small-signal gate voltage v_{gs} and to the small-signal drain current i_d as Eq.[5] shows. Other
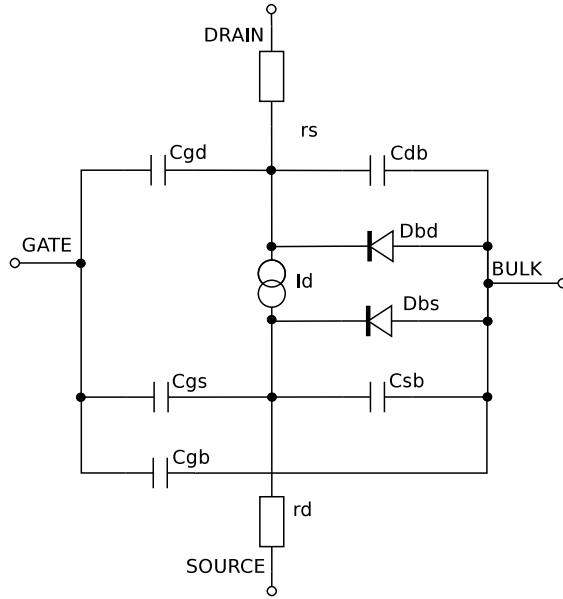
Figure 4: Large signal model, [All02]
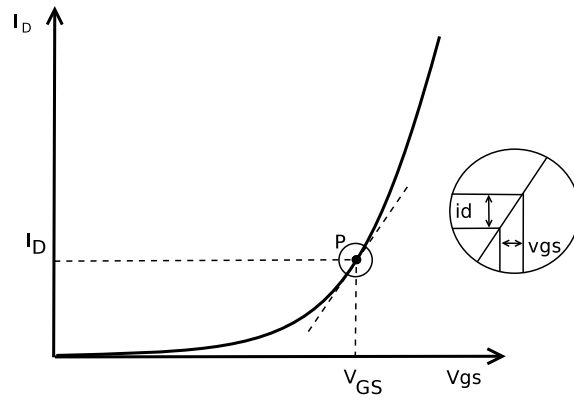
Figure 5: Small signal model explanation

5

transconductance is a g_{mbs}.

$$i_d = g_m \cdot v_{gs} \tag{5}$$

Model also contains capacitances from the large-signal model and three conductances g_{DS}, g_{BD}, g_{BS}. The conductance g_{DS} is a conduction of the channel and it is also an important small-signal parameter. The last two conductances represent conduction of parasitic substrate diodes. These diodes are usually off and therefore these conductivities (g_{BD}, g_{BS}) are often neglected. Fig.[6(b)] shows simplified small-signal model.
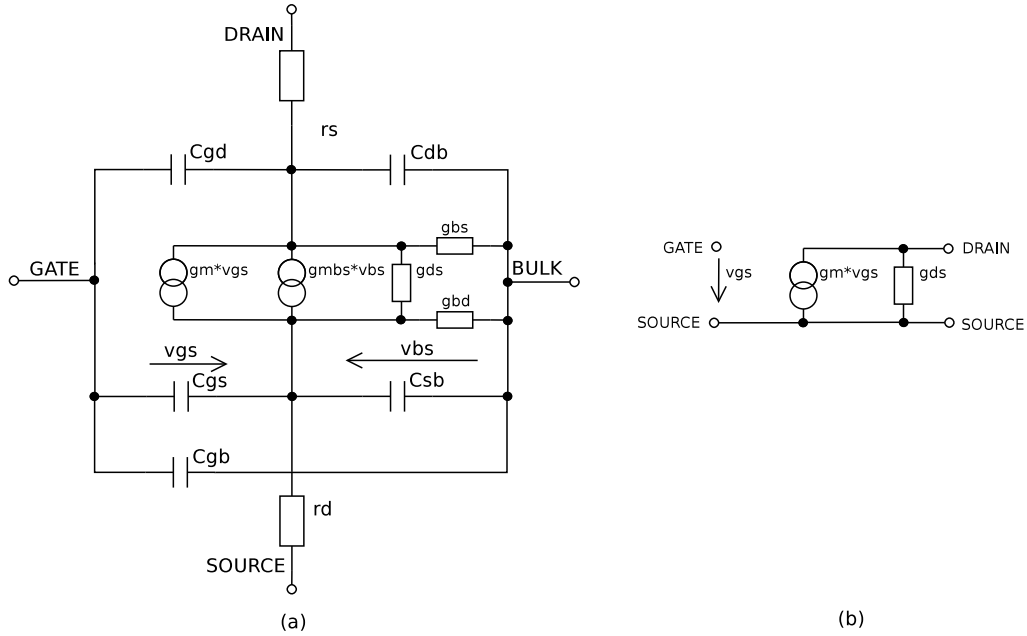
Figure 6: Small signal models (a) Full (b) Simplified, [All02]

The previous mentioned small-signal parameters can be calculated by Eq.[6,7,8]

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \tag{6}$$

$$g_{DS} = \frac{\partial I_D}{\partial V_{DS}} \tag{7}$$

$$g_{mbs} = \frac{\partial I_D}{\partial V_{BS}} \tag{8}$$

These parameters are valid only at given operating point and at its close vicinity. The following equations show the basic relations for the small-signal parameters in the linear (9,11) and in the saturation region (10,12). Note that strong inversion is assumed.

$$g_m = \beta V_{DS} \tag{9}$$

$$g_m = \frac{2I_D}{(V_{GS} - V_{th})} \tag{10}$$

$$g_{DS} = \beta((V_{GS} - V_{th}) - V_{DS}) \tag{11}$$

$$g_{DS} = \lambda I_D \tag{12}$$

2.3   MOS Capacitances Summary

Capacitances Csb and Cdb are the reversed biased junction capacitances. More about behaviour of junction capacitance can be found in the chapter 3.2.3.

The main three capacitances are Cgs, Cdb and Cgd. These capacitances are given by the overlap capacitance (C_{GDO}, C_{GBO} or C_{GSO}) and by the part of gate capacitance C'_{ox} (it is a capacitance between the gate and the channel). Note that a shape of the channel is not constant, it depends on V_{DS} and V_{GS}. Thus the value of capacitances Cgs, Cdb and Cgd are determined by the operating region. Table [1] lists the values of mentioned capacitances.

Name	Off	Triode	Saturation
Cgd	$C_{GDO} \cdot$ W	$\frac{1}{2}WLC'_{ox}$	$C_{GDO} \cdot$ W
Cgb	$C_{ox}WL + C_{GBO} \cdot$ L	$C_{GBO} \cdot$L	$C_{GBO} \cdot$ L
Cgs	$C_{GSO} \cdot$W	$\frac{1}{2}WLC'_{ox}$	$\frac{2}{3}WLC'_{ox} + C_{GSO} \cdot W$

Table 1: MOS transistor capacitances

2.4 Threshold Voltage and Body Effect

In the previous chapters it is assumed that MOS transistor is turned on. The threshold voltage V_{th} is applied to the gate terminal for keeping MOS transistor turned on.

The definition of threshold voltage is complex and the concrete value depends on many factors (thickness of gate oxide t_{ox}, substrate doping level, temperature etc.). Moreover the threshold voltage value depends on the bulk-source voltage V_{BS}.

The threshold voltage extraction can be described by several methods: Constant current method, Second derivative method, Extrapolation in the linear region method (ELR) etc...

According to the ELR method, the threshold voltage is found at the intercept of the tangent in the inflexion point with the V_G axis, [DPDR00]. The inflexion point P_{inflex} is the point where the transconductance g_m reaches its maximum. This method assumes NMOS with the source and bulk connected to the ground (therefore the I_D-V_G plot is used instead of the I_D-V_{GS}). Moreover this method neglects the parasitic series resistances and mobility degradation.
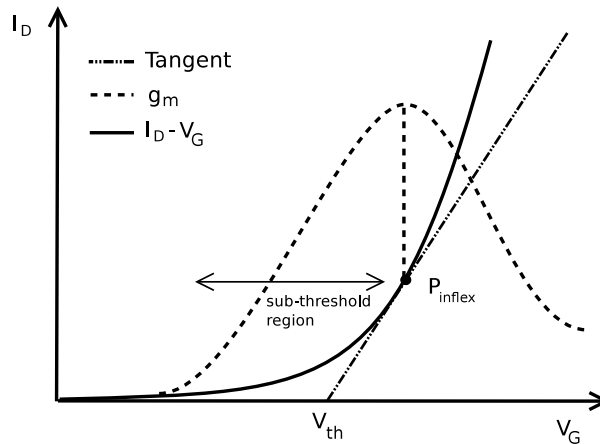
Figure 7: The V_{th} extraction, ELR method

From Fig.[7] is clear that some drain current flows under the threshold voltage. This operating area is called sub-threshold or weak-inversion region, see next sub-chapter for more information. Applying a gate-source voltage

$V_{GS} \leq V_{th}$ to the gate terminal, the sub-threshold region can be achieved. Thus only a small number of electrons is attracted under the gate oxide.

As mentioned before, the threshold voltage V_{th} depends on the voltage V_{BS}. If the source and bulk are not at the same potential then a Body effect occurs. The Body effect is usually unwanted phenomenon that simply steals some electrons from the channel. Thus the threshold voltage rises and the drain current I_D decreases (for constant V_{GS}), see Fig.[8].
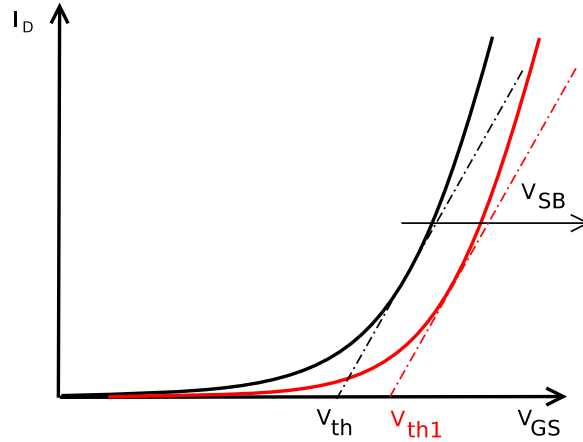
Figure 8: I_D-V_{GS} relation with marked threshold voltage

Eq.[13] describes influence of V_{BS} on the threshold V_{th}, [Bak10].

$$V_{th} = V_{th0} + \gamma \left(\sqrt{|2\phi_f| + V_{SB}} - \sqrt{|2\phi_f|} \right) \qquad (13)$$

where,
V_{th0} = zero-bias threshold voltage
$2\phi_f$ = flatband voltage - constant
V_{SB} = source-bulk voltage
γ = body factor

The body factor γ and the flatband voltage $2\phi_f$ are (for simplicity) considered as a constants.

2.5 MOS Transistor Inversion Level

As mentioned before MOS transistor has basically two operating regions (linear, saturation region) depending on voltages V_{GS} and V_{DS}. This approximation is sufficient for some design approaches. But for micro-power design it is not.

Therefore the the saturation region is divided into next three operating levels, [Ste08]: Weak inversion, Moderate inversion and Strong inversion. This three operating levels depend on the V_{GS} voltage and they are significant for modern low-power integrated circuit design.

The exact definitions of these region are based on: surface potential ψ_s, pinch-off voltage V_p, Inversion charge Q_{inv} and depletion charge Q_{dep}. It is clear that the definition is very complex and unfortunately the exact explanation of this matter is beyond the scope of this work.

The inversion level can be estimated by the $V_{GS} - V_{th}$ value, [SW94].

Weak inversion: $V_{GS} - V_{th} < -50mV$

Moderate inversion: $-50mV < V_{GS} - V_{th} < 250mV$

Strong inversion: $250mV < V_{GS} - V_{th}$

To find out how much the transistor is in the concrete inversion, the inversion factor IF is defined, [Ste08].

$$IF = \frac{I_{DSat}}{I_s} = ln^2 \left[1 + exp \left(\frac{V_p - V_s}{2U_T} \right) \right] \tag{14}$$

where
I_{DSat} = saturation drain current
I_s = Specific current
V_p = pinch-off voltage
V_s = local voltage at source end of channel

The following table defines the relation between the inversion level and inversion factor IF.

Weak Inversion	$IF < 0.1$
Moderate Inversion	$0.1 < IF < 10$
Strong Inversion	$10 < IF$

Table 2: Operating regions definition with respect to the IF, [Ste08]
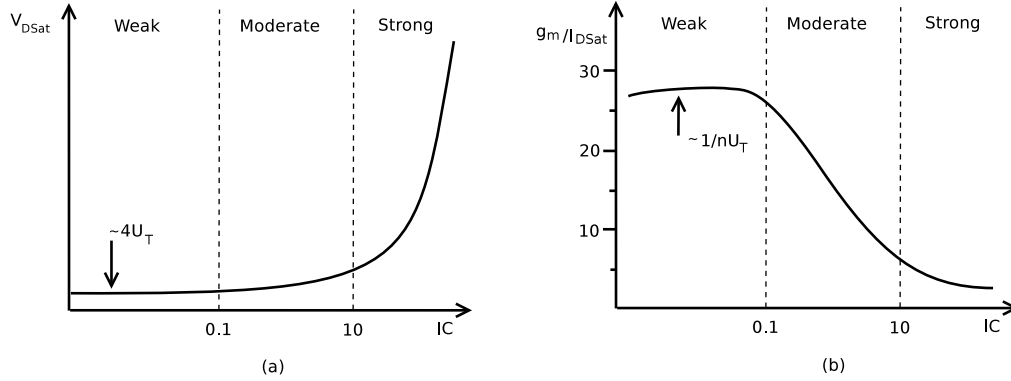
Figure 9: V_{DSat} and g_m/I_D parameters as a function of IF, [Ste08]

The following Fig.[9] shows two important MOS parameters (saturation voltage and g_m/I_D ratio) as a function of IF.

The saturation voltage V_{DSat} is the required voltage across drain-source terminal to keep it saturated. This parameter is often required to be low for better headroom. From Fig.[9(a)] is clear that minimum V_{DSat} is obtained in weak inversion and it is roughly equal to $V_{DSat,min} \simeq 4U_T$.

The headroom is the difference between V_{DS} and V_{DSat}. For instance, assume that NMOS transistor has a $V_{DSat} = 140mV$ and $V_{DS} = 180mV$. The V_{DS} can be reduced by roughly $40mV$ till the MOS transistor is still in saturation region. This voltage reserve is called Headroom.

The g_m/I_D is a very important parameter because it shows the transconductance efficiency, see Fig.[9(b)]. The maximum is reached at weak inversion $g_m/I_D \simeq 25$ (the maximum value depends on n which is technology parameter). The minimal value is reached in strong inversion.

11

On the other hand strong inversion provides the small device size and high bandwidth. This design tradeoffs are summarized by Fig.[10].

The IF factor is unique because all important parameters : saturation voltage V_{DSat}, g_m/I_D ratio, output conductance g_{DS}, capacitances (in saturation region), intrinsic gain, transition frequency and noise can be directly calculated as a function of IF, [Ste08].

Thus the new design methodology called g_m/I_D is used. Possible design approach provides three degrees of freedom in design: drain current I_D, inversion factor IF and transistor length L.

The maximum current condition and designer choice set the drain current of each transistor. The IF factor is than chosen accordingly to required MOS performance. The following Eq.[15] gives than the W/L ratio [1].

$$\frac{W}{L} = \frac{1}{IF}\frac{I_D}{I_s} \tag{15}$$

The last design step is the MOS length L choice. This choice is trade-off between high DC gain on one side and area, stability and speed on the other side. After this step the MOS width W is easily calculated.

---

[1]The W/L ratio is often called an aspect ratio AR

Rel. W, WL	100x	10x	1x	0.1x	0.01x
Rel. g_m/I_D, A_{Vi}	97%	89%	63%	28%	9.7%
Rel. f_{Ti}	0.03x	0.23x	1x	3.4x	10.9x
V_{DSAT}	104 mV	108 mV	135 mV	243 mV	595 mV
V_{EFF}	-163 mV	-72 mV	40 mV	225 mV	724 mV

Weak inversion   Moderate inversion   Strong inversion

High Side Weak Inversion Side Center Strong Inversion Side Onset, Low Side

← Deep   Heavy or Deep →

0.01   0.1   1   10   100

Inversion Coefficient

- Large device size
- Low bandwidth
- No improvement in g_m/I_D, V_{DSAT}
- High leakage current

- Moderate device size
- Moderate bandwidth
- g_m/I_D moderately high
- V_{EFF}, V_{DSAT} moderately low
- Little velocity saturation
- Important for power-efficient, low-voltage designs

- Small device size
- High bandwidth
- g_m/I_D low
- V_{EFP}, V_{DSAT} high
- Velocity saturation at short L
- Difficult for power-efficient, low-voltage designs

Figure 10: Design tradeoffs as a function of IF, [BFJ+07]

3 Passive Devices

Fundamental passive devices are presented in this chapter (resistors and capacitors). Quality of passive devices is very important for analog IC design. Passive devices are used for determination of gain, time constant, compensation, current to voltage conversion etc. Low power applications require resistors with a large resistance value. In this case, resistors consume a significant area.

Unfortunately resistors and capacitors in CMOS process have temperature and voltage dependence. Moreover process and mismatch variations cause significant inaccuracy. These effects are presented in the next chapter.

3.1 Resistor

3.1.1 Sheet Resistance

Generally, resistors are made of resistance body and heads, see Fig[11]. That is why the total resistance can be expressed by Eq.[16] and square resistance by Eq.[17]

$$R = 2R_{head} + R_{sqr}\frac{L}{W} \tag{16}$$

$$R_{sqr} = \rho/t \tag{17}$$

where
R_{head} = head resistance
W = body width
L = body length
ρ = material resistance
t = thickness of square layer
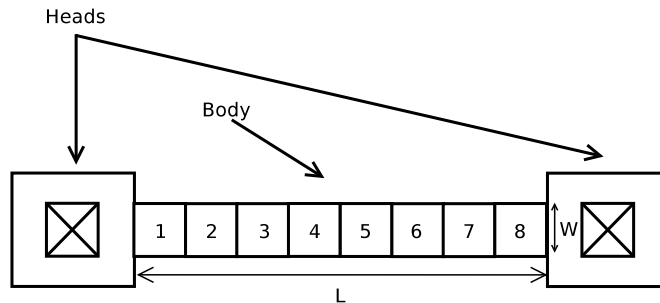
Figure 11: A simple resistor layout

3.1.2 Poly-Silicon Resistor

Poly-si (ie. poly crystal) is the most used type of resistor. The poly-si resistor can be manufactured with or without silicide layer. This layer is a mix of metal and silicon and it has better conduction ability. Silicide poly-si resistor has therefore generally smaller sheet resistance than non-silicide.

For instance, silicide poly-si in our process design kit (PDK-180 nm technology) has resistance about tens of ohm/square while non-silicide has about 200 ohm/square. Poly-si resistor is made of strip of resistive poly crystal material. Width and length of the strip can be calculated from final device resistance and sheet resistance.
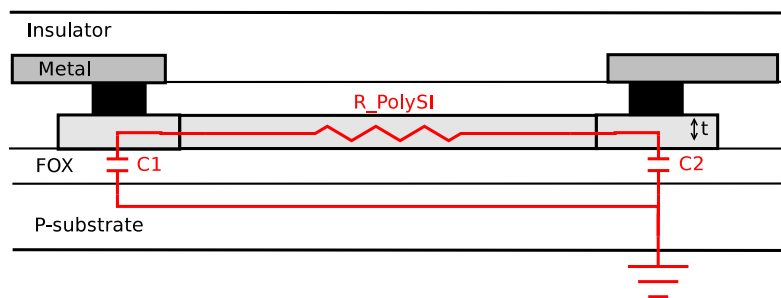
Figure 12: A first order Poly-Si resistor model (Cross section)

From the picture above is clear that this strip structure has parasitic capacitances to the silicon substrate. These parasitic capacitances are distributed along the strip but for simplicity assume the first-order model that

15

splits the total parasitic capacitance into two capacitors C1 and C2 see Fig.[12]

There are many types of resistor layout (meander, dog bone etc...). These layout technique should minimize impact of fabrication process.

3.1.3   N-Well Resistor

This type of resistor is made by P or N lightly doped diffusion area and two contacts. This n-well resistor has sheet resistance 10-100 ohm/square depending on the dopant concentration. Unfortunately, a parasitic diode D1 (pn junction) is presented in this structure. Attention to correctly polarized pn junction must be paid.
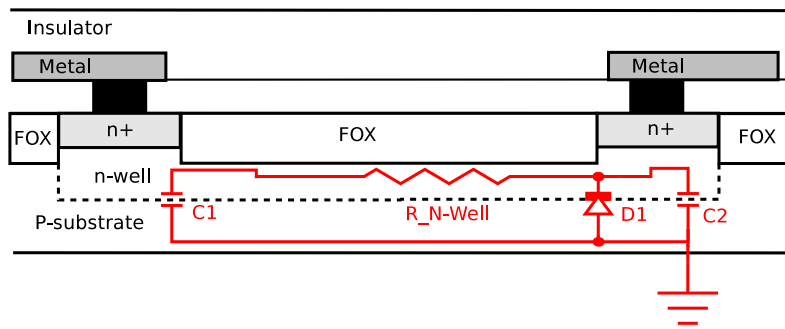
Figure 13: N-well resistor without depletion layer (Cross section)

For simplicity, the first-order model can be also used, see Fig.[13]. However, the parasitic capacitors C1,C2 are junction capacitances and therefore their values depend on the voltage across the resistor. So the total capacitance can change due to the voltage. Moreover width of the depletion region is voltage dependent, so the total resistance of diffusion resistor is nonlinear. These negatives can be problematic for analog design, where constant resistance and capacitance values are generally required.

3.1.4 MOS Resistor

The third method to realize resistor on chip is due to the MOS transistor operating in triode region. Assuming NMOS transistor the following condition must be set in order to achieve triode region, see Eg.[18]

$$V_{DS} =< V_{GS} - V_{th} \tag{18}$$
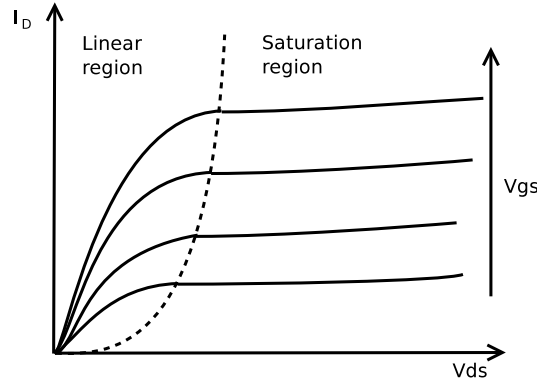
Figure 14: NMOS transistor, output characteristic

The resistance of mosfet resistor is calculated by Eq.[19], [All02].

$$rds = \frac{1}{(\mu_0 C_{ox}(W/L)(V_{GS} - V_{th}))} \tag{19}$$

where
μ_0 = carrier mobility
C_{ox} = oxide capacitance
W/L = geometric size of transistor channel (width and length)

Eq.[19] shows that resistance of triode NMOS transistor can be tuned by the gate-source voltage V_{GS} which is indeed an important advantage. Second advantage is a possibility of a large resistance value in the compact area.

On the other hand, non-linearity is a disadvantage of MOS resistor.

3.2 Capacitors

In this section three basic capacitor types are presented: metal-metal capacitor, pn junction capacitor and MOS capacitor. Capacitors have important role in the analog circuit design. They are often used for the signal filtering and for the stability compensation. Capacitors are essential part of charge pump and dynamic random access memory (DRAM) circuits.

3.2.1 Metal-Metal Capacitor

Assuming a simple plate capacitor that contains two parallel plates and dielectric material between them, Eq.[20] gives the total capacitance.

$$C = \varepsilon_{ox} \left(\frac{S}{t_{ox}} \right) \tag{20}$$

where
ε_{ox} = permittivity of the insulator between plates
S = plate area
t_{ox} = spacing between plates

It is clear from the Eq.[20] that the total capacitance is linear and a big S and ε_{ox} or small t_{ox} are required for the large value of capacitance.

Fortunately metal-metal capacitor can be realized by two plates with dielectric material between them (this structure is called parallel-plate). If the edge and parasitic capacitors are neglected, Eq.[20] is valid. Spacing between plates is in order of $0.1 - 10 \mu m$ (depending on technology), [Car12]. Silicon dioxide (SiO_2) is often used as dielectric material with $\varepsilon_{ox} \cong 3.9\varepsilon_o$, [Car12].

Fig.[15] show poly-poly capacitor Cpp with parasitic capacitances Cp. The most important parasitic capacitance Cbp is formed between the bottom plate of poly layer and the substrate. Therefore it is called bottom plate parasitic capacitance. It is clear that value of bottom plate parasitic capacitance can be large due to the large plate area.

The parallel-plate structure can be made of different elements, for instance by metal-metal, metal-substrate, metal-diffusion area, poly-poly, poly-metal etc...

18
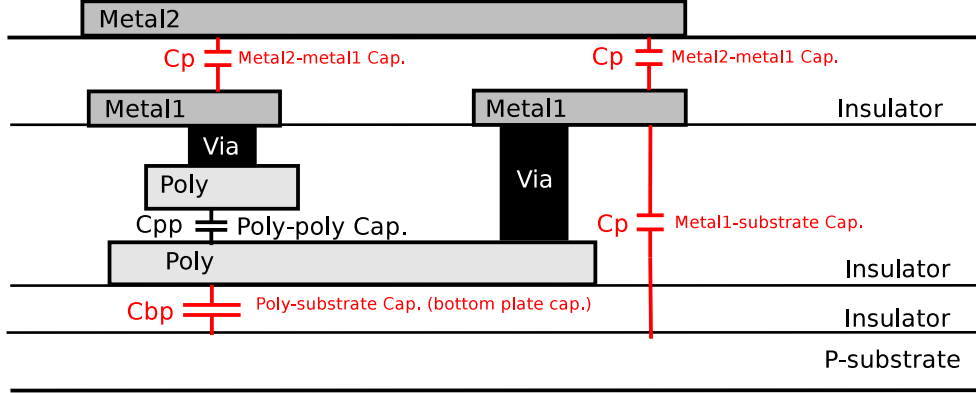
Figure 15: Poly-poly capacitor with parasitic capacitances (Cross section)

3.2.2 Pn Junction Capacitor

Reverse-biased pn junction has a charge-storage effect that is represented by the junction capacitor C_j. This junction capacitor C_j is formed by a bottom capacitor and a sidewall capacitor. Behaviour of the junction capacitor is given by Eq.[21], [All02].

$$C_j = \frac{C_{jb}}{(1 + \frac{V_R}{\Phi_0})^{m_j}} + \frac{C_{jsw}}{(1 + \frac{V_R}{\Phi_{0S}})^{m_{jsw}}} \tag{21}$$

where
C_{jb},C_{jsw}= zero-bias bottom pn capacitance for bottom, sidewall
Φ_O,Φ_{OS}= pn build potential for bottom, sidewall
m_j,m_{jsw}= pn grading coefficient for bottom, sidewall
V_R= applied reverse voltage

Pn junction capacitor provides large value of capacitance because depletion region can be thin and its relative permittivity is high ($\varepsilon_r = 11.8$), [Car12]. Furthermore, it is obvious from the Eq.[21] that the junction capacitance can be tuned by reverse voltage V_R across it. This feature is often used for oscillators and radio-frequency applications.

Voltage dependence is not required in some applications. The next disadvantage is capacitance dependence on dopant concentration, which is difficult controlled. Furthermore, leakage current through junction capacitor can be

problematic.

3.2.3   MOS Capacitor

As mentioned in Eq.[20], larger capacitance value can be achieved by thin dielectric material t_{ox}. Gate oxide of the MOS transistor is the thinnest layer in CMOS process, [Car12]. Therefore MOS capacitor provides a good ratio of capacitance/area.

Assume NMOS transistor now, if the positive voltage Vdd is applied to the gate electrode, a conductive channel is formed by attracted electrons. This channel creates with gate electrode plates of gate capacitor and gate oxide is a dielectric.

If source, drain and bulk are connected together, MOS transistor works as the capacitor, see Fig.[16]. Eq.[22] shows gate capacitance (sometimes called oxide capacitance).
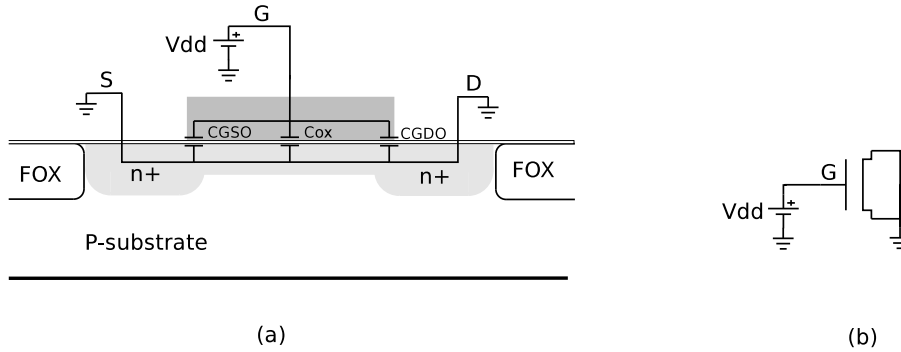
Figure 16: (a) NMOS capacitor (Cross section), (b) NMOS cap. symbol

$$C_{ox} = W_{ef} L_{ef} \frac{\varepsilon_{ox}}{t_{ox}} = W_{ef} L_{ef} C'_{ox} \tag{22}$$

where
$W_{ef} L_{ef} =$ effective geometrical size of channel
$C'_{ox} =$ gate capacitance per area unit

Fig.[16] also shows next two capacitors C_{GDO} and C_{GSO}. These two capacitors are called overlap capacitances and they are formed by inaccuracy of diffusion process (expansion of source and drain area under gate).

20

$$C_{GDO} = C_{GSO} = C'_{ox} L_{diff} \tag{23}$$

where L_{diff} is a length of overlapped area.

The total capacitance of MOS capacitor is given by Eq.[24]

$$C_{MOS} = W_{ef} L_{ef} C'_{ox} + 2 W_{ef} L_{diff} C'_{ox} \tag{24}$$

Unfortunately, gate capacitance C'_{ox} is dependent on voltage V_{GS} and it is not linear. Of course some minimal $V_{GS,min}$ voltage needs to be ensured for proper work of MOS capacitor, see Fig.[17].
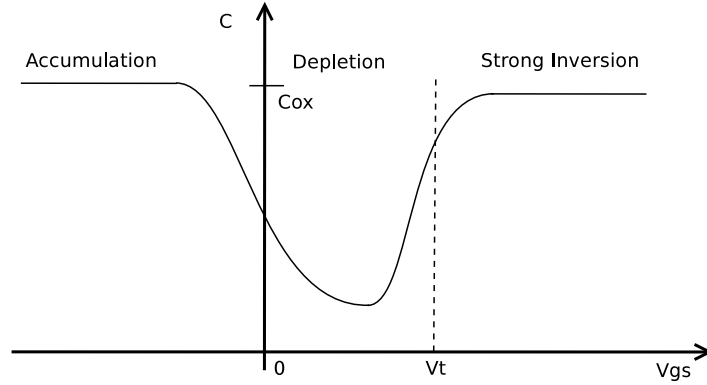
Figure 17: The variation of the gate capacitance with V_{GS} voltage

Note that MOS capacitor can operate in three regions: accumulation, depletion and strong inversion. These three regions are now discussed for NMOS transistor.

In accumulation region a negative voltage is applied on the gate. Thus mobile holes are attracted (accumulated) under the gate oxide. Capacitor is formed between gate electrode and substrate electrode. Unfortunately serial resistance is present and that is why accumulation region is not convenient choice for MOS capacitor.

If the gate voltage V_{GS} (smaller than threshold voltage V_{th}) is applied, only a small number of electrons are attracted. Now, the area under the gate is said to be nearly depleted. It means depleted of free electrons and holes.

21

From Fig.[17] is clear that total capacitance of MOS capacitor in depletion region (weak inversion) is small.

The last region is the suitable choice for MOS capacitor. Due to the large positive gate voltage a large number of electrons are attracted. Thus the area under gate oxide is no longer p-type but n-type. Moreover the parasitic serial resistance is smaller than in the accumulation region.

4   Voltage References

Voltage reference is a circuit that generates a specific output voltage. This output voltage should not depend on the load current, supply voltage, temperature, time or process corner. As mentioned before the voltage reference is essential component for ADC/DAC, memory and for circuit biasing. Thus it can be said that every complex integrated circuit (digital meters, smart sensors, threshold detectors, battery management systems etc..) needs some kind of voltage reference.

Before some voltage reference are presented, the temperature coefficient is introduced.

Temperature coefficient (TC) is an important parameter of a component that should be considered during the electronic design. TC describes the relationship between a change of parameter as a function of temperature.

With usage of Taylor polynom most of these functions can be expressed as follows:

$$\rho(T) = \rho(T_0) \left(1 + \alpha(\Delta T) + \beta(\Delta T)^2 + \gamma(\Delta T)^3 + ...\right) \quad (25)$$

where
ΔT = difference between temperature T and nominal temperature T_0
$\rho(T_0)$ = parameter ρ at temperature T_0
α, β, γ = are temperature coefficients

If TC coefficients of higher order$(\beta, \gamma,...)$ are negligible then we can consider the dependency to be linear. The Eq.[25] can then be written as:

$$\rho(T) = \rho(T_0) \left(1 + \alpha(\Delta T)\right) \quad (26)$$

It's clear that the parameter α describes the general tendency of the relationship.

For example, if the parameter ρ is the reference voltage V_{ref} and α is greater than zero then the value of V_{ref} increases with temperature. This phenomenon is called PTAT (Proportional To Absolute Temperature), see Fig.[18(a)]. If the α is smaller than zero then the V_{ref} decreases with temperature and it is called CTAT (Complementary To Absolute Temperature), see Fig.[18(b)].
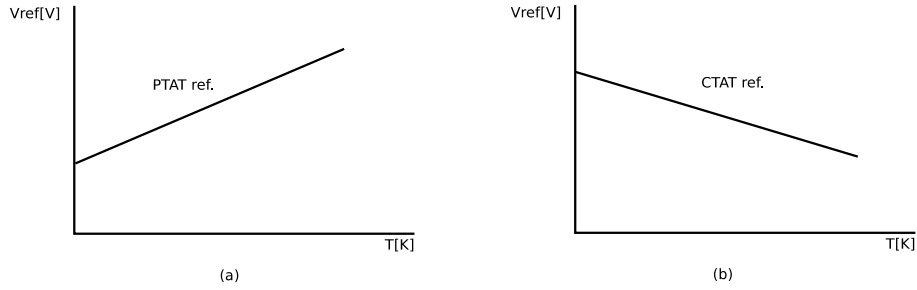
Figure 18: (a) PTAT and (b) CTAT voltage references

In next section PTAT and CTAT system are shown.

4.0.4 CTAT

Simple diode is used in order to generate CTAT TC. Diode or bipolar transistor can be used to obtain pn junction/diode. NPN transistor in Fig.[19] has base and collector connected.

In this configuration the base-emitter junction is exploited to obtain a diode. If positive voltage is applied to pn junction, forward biased region is entered. Relationship between I-V is approximated by Eq.[27] called Shockley's equation.

$$I = I_s \cdot exp\left(\frac{V}{nU_T} - 1\right) \tag{27}$$

where
n = material coefficient and its value is about 1.05 - 1.1. For simplicity, this coefficient is often considered to be equal to 1.
U_T = thermal voltage, given by Eq.[28]
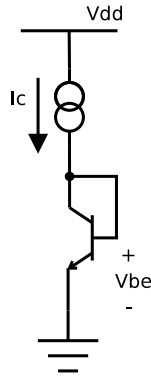
Figure 19: CTAT voltage realization

I_s = saturation current, its value is in range $10^{-12} - 10^{-18}$ [Sed13].

$$U_T = kT/q \approx 26mV/K \qquad T = 300K \qquad (28)$$

where
k = Boltzmann's constant = 1.38e-23 J/K
T = the absolute temperature in Kelvins = 273 + temperature in °C
q = the magnitude of elementary electron charge = 1.60e-19 C

It can be derived, [Sed13]:

$$\partial Vbe/\partial T \cong -2mV/K \qquad (29)$$

The temperature behaviour of diode connected bipolar transistor (i.e transistor with shorted base and collector, see Fig.[19]) was simulated. The $\partial Vbe/\partial T$ of NPN (Area=0.36μm^2, 180nm PDK) is $-2.27mV/K$ at room temperature, see simulation result Fig.[20].
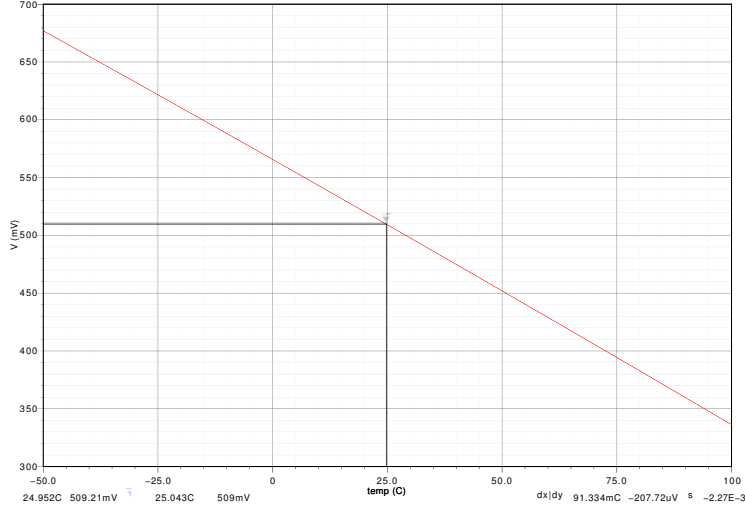
Figure 20: Temperature behaviour of the diode

4.0.5 PTAT

Fig.[21] shows topology that provides PTAT TC. Assuming two identical NPN bipolar transistors Q1 and Q2, each one has a different DC bias current. Their base and collector are connected and thus they behave like a diode. Eq.[30] is simplified Shockley's equations and it describes current through diode and from the same equation voltage V_{be} is expressed by Eq.[31].

$$I_c = I_s \cdot exp(V_{be}/U_T) \tag{30}$$

$$V_{be} = \frac{kT}{q} \cdot ln(I_c/I_s) \tag{31}$$

Eq.[32,33] show voltage across these two BJTs

$$V_{be1} = U_T \cdot ln(I_{c1}/I_s) \tag{32}$$

$$V_{be2} = U_T \cdot ln(I_{c2}/I_s) \tag{33}$$

$$\Delta V_{be} = V_{be2} - V_{be1} = U_T \cdot ln(I_{c2}/I_{c1}) \tag{34}$$

25
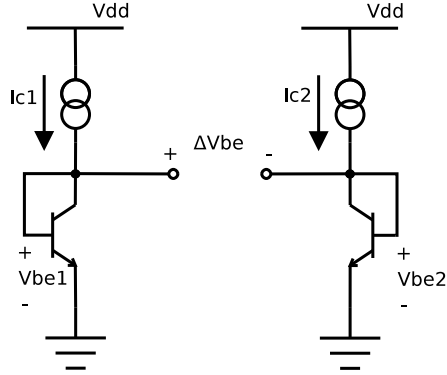
Figure 21: PTAT voltage realization

Eq.[34] shows that ΔV_{be} includes thermal voltage U_T and from Eq.[31] it is evident that this expression is rising with temperature ($U_T = kT/q$). Therefore ΔV_{be} is rising with temperature and ΔV_{be} is PTAT. Next term $ln(I_{c2}/I_{c1})$ is a design choice of different currents. The same circuit can be build with the same currents $I_{c1} = I_{c2}$ and use different area of pn junction. Or both principles together can be used, i.e. different current and area. Eq.[35] shows ΔV_{be} relation expressed by the current density J_c .

$$\Delta V_{be} = V_{be2} - V_{be1} = U_T \cdot ln(J_{c2}/J_{c1}) \tag{35}$$

Advanced topologies of voltage references are usually based on one of the previous mentioned phenomenons. Therefore they are PTAT or CTAT systems. Unfortunately TC is not sufficiently small for many applications and therefore bandgap reference (BGR) is used. The BGR combines PTAT and CTAT to achieve almost zero TC.

4.0.6 Simple Type Reference

Voltage divider.   A well know discrete form of voltage divider uses resistors, Fig.[22(a)], or capacitors. The voltage divider formed with resistors has the advantage of simplicity and temperature insensitivity but supply dependence and the power dissipation are a problem. Large resistors are required to minimize current through the resistor divider.

Voltage divider formed with MOS transistor is more space efficient and its behaviour is better described (BSIM3V3 or EKV model describe transistor).
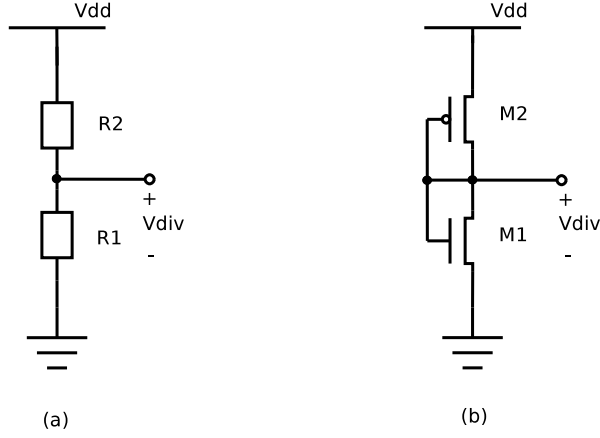
Figure 22: (a) resistor only divider, (b) MOS only divider

The MOS-only divider generates a fraction of the supply voltage. If the PMOS and NMOS transistor have the same drain current $I_{D1} = I_{D2}$ then Eq.[36] is valid.

$$\frac{\beta_1}{2}(V_{div} - V_{thn})^2 = \frac{\beta_2}{2}(V_{dd} - V_{div} - V_{thp})^2 \qquad (36)$$

Output voltage v_{div} can be then expressed by Eq.[37]

$$V_{div} = \frac{V_{dd} - V_{thp} + \sqrt{\frac{\beta_1}{\beta_2}}V_{thn}}{\sqrt{\frac{\beta_1}{\beta_2}} + 1} \qquad (37)$$

Voltage reference MOS-Resistor. Fig.[23] shows the simple voltage reference made by resistor and diode connected MOS transistor (diode connection guarantee the saturation region). The reference voltage V_{ref} is equal to the voltage V_{GS} of the MOS transistor M1.

$$I_D = \frac{V_{dd} - V_{ref}}{R_1} = \frac{\beta}{2}(V_{ref} - V_{thn})^2 \qquad (38)$$

$$V_{ref} = V_{thn} + \sqrt{\frac{2I_D}{\beta}} \qquad (39)$$

27
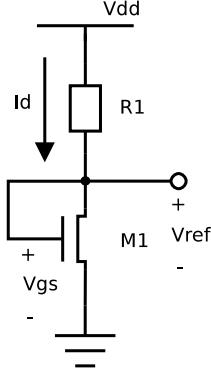
Figure 23: MOS-resistor voltage reference

Eq.[38,39] describes mentioned circuit.

4.0.7 Advanced Reference Topology

Voltage reference is used to set proper bias point in many cases. Therefore, the more precise reference the better choice. These advanced references are based on one of the following effects.

Voltage reference based on Vbe, [DB11]. Voltage on forward polarized diode is about 0.7 V (it depends on temperature and current density). Due to MOS transistors M4, M5 diode voltage is repeated on resistor R1, see Fig.[24]. The current $I1$ through resistor R1 is defined by Eg.[40]. This current is mirrored by PMOS mirror M2-M3 into the output branch. On the output resistor $kR1$ reference voltage V_{ref} is generated and it is described by Eq.[41]

$$I_1 = \frac{V_{be1}}{R_1} \tag{40}$$

$$V_{ref} = kR_1 I_{out} = kR_1 \frac{(W/L)_3}{(W/L)_2} I_1 = k\frac{(W/L)_3}{(W/L)_2} V_{be} \tag{41}$$

28
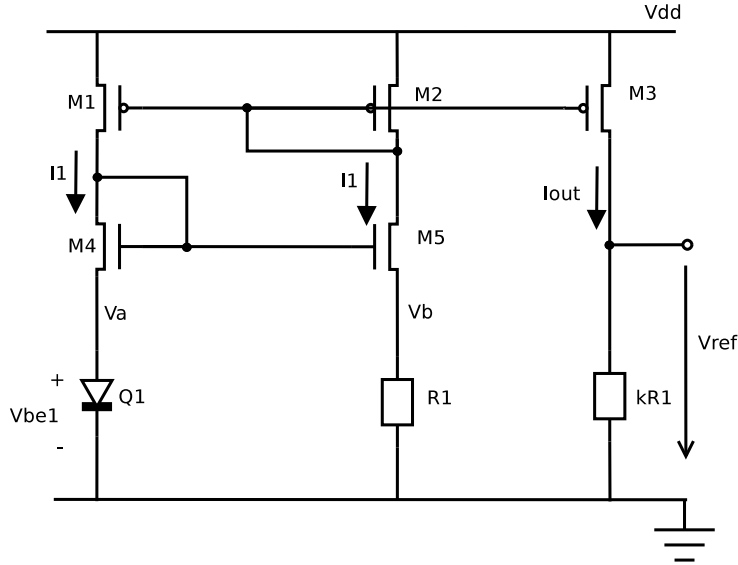
Figure 24: Voltage reference exploiting V_{be}

The reference voltage V_{ref} can be tuned by multiplying factors k and $\frac{(W/L)_3}{(W/L)_2}$.

Temperature behaviour of diode forward voltage can be estimated as -2mV/K and it is CTAT. Temperature on chip can rapidly change due to a chip self-heating. Of course, an ambient contribution is another important factor. Temperature working range might be wide (it depends on application), it is -50 - 100 °C in our case. Voltage reference based on base-emitter voltage V_{be} has variation about 300mV only due to the diode behaviour. Another temperature dependent devices are not taken into account.

Voltage reference based on Ut, [DB11]. Fig.[25] presents the voltage reference based on thermal voltage U_T. Note that two diodes Q1 and Q2 have a different area ratio $1 : N$. Thus voltage across diodes Q1, Q2 is not equal. Transistors M1, M2, M4 and M5 formed advanced mirror that ensure equality of voltages V_a and V_b and current through branches. Therefore voltage drop on resistor R1 obtain the thermal voltage U_T, as Eq.[42] shows.

$$I_1 = \frac{V_{R1}}{R_1} = \frac{\Delta V_{be}}{R_1} = \frac{U_T}{R_1}ln(N) \tag{42}$$

Current I_1 is mirrored by MOS transistors M2 and M3 into output branch. The reference voltage V_{ref} can be expressed by Eq.[43]

$$V_{ref} = kR_1 I_{out} = kR_1 \frac{(W/L)_3}{(W/L)_2} I_1 = kU_T ln(N) \frac{(W/L)_3}{(W/L)_2} \tag{43}$$

where k and $\frac{(W/L)_3}{(W/L)_2}$ terms can tune value of output voltage V_{ref}. Due to the PTAT coefficient in Eq.[43] temperature behaviour of circuit is PTAT.
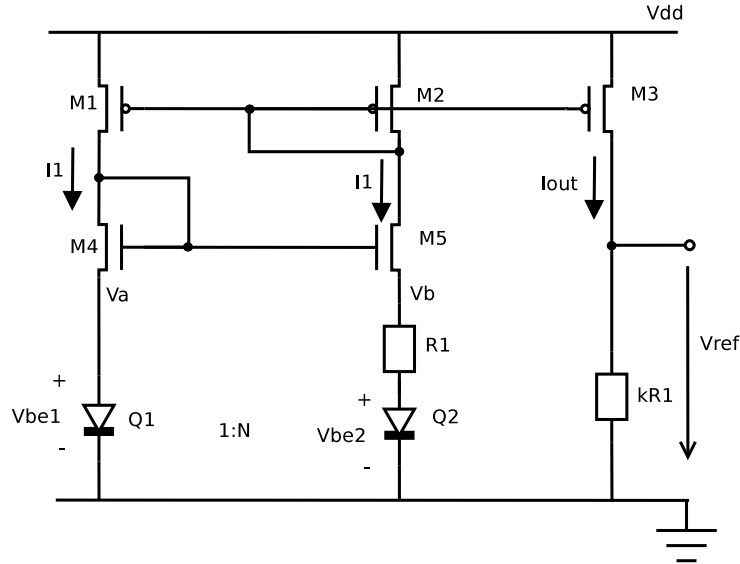
Figure 25: Voltage reference exploiting U_T

4.1 Bandgap Reference (BGR)

The name Bandgap has originates from the physics nature of the silicon. Fig.[26] shows a simplified energy band diagram of semiconductor, [Bak10].
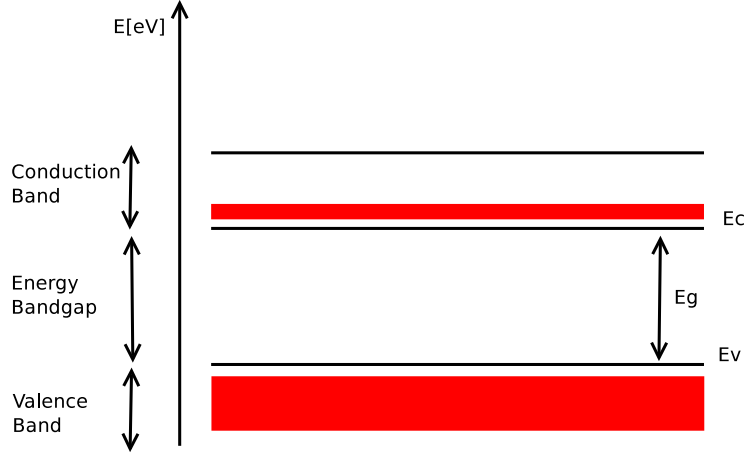
Figure 26: A simplified energy band diagram

The Ec line indicates bottom edge of the conduction band and Ev is top edge for valence band. The distance between these two bands is the bandgap energy Eg. This bandgap energy decreases as the temperature increases, see Eq.[44],[Bak10].

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T + \beta} \tag{44}$$

where $E_g(0)$, α, β are fitting parameters for different material (Ge,Si,GaAs). For the silicon and 300K temperature, Eq.[44] can be rewritten as

$$E_g(300) = 1.166 - \frac{0.473 \cdot 10^{-3} \cdot 300^2}{300 + 636} \cong 1.12eV \tag{45}$$

The result of Eq[45], 1.12eV is an amount of energy that electron needs to overcome gap between the valence and conduction band.

The theoretical output voltage of BGR is also equal to 1.12V, therefore it is called bandgap. BGR is one of the most popular voltage references due to minimized TC. Moreover some current references are based on BGR circuits.

After the explanation of BGR function a few implementations of BGR are presented.

4.1.1 Bandgap Basic Principle

As mentioned before, BGR is a circuit exploiting CTAT and PTAT to compensate temperature dependence.
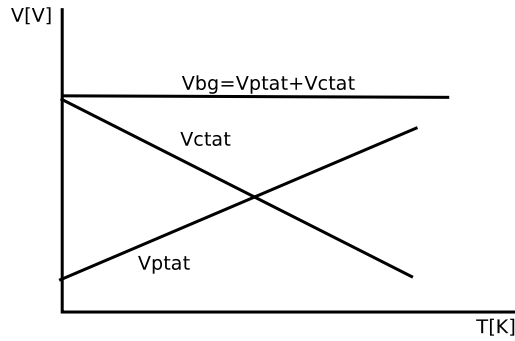
Figure 27: Zero TC due to PTAT and CTAT

As chapter [4.0.5] shows, the thermal voltage $U_T = kT/q$ has a positive temperature coefficient (PTAT) and it rises by $+85uV/K$. Voltage base-emitter V_{be} has a negative temperature coefficient (CTAT) and its drop is about -2mV/K. Ideally a zero TC is achieved by adding these two voltages, see Fig.[27]. It is evident that PTAT and CTAT are different scales, therefore PTAT is multiplied by constant K to ensure their mutual compensation. Then Eq.[46] is valid and it is a basic relation for BGR.

$$V_{bg} = V_{be} + KU_T \qquad (46)$$

Fig.[28] shows a principle of ideal BGR reference. Ideal means that output voltage V_{bg} has zero TC. Unfortunately there are reasons why the zero TC is never obtained. For instance voltage variations of V_{be} are not linear due to the higher order of TC. Moreover supply voltage V_{dd} changes, mismatch and process variation have also undesirable effect. Mismatch and process variation are discussed in chapter [5].
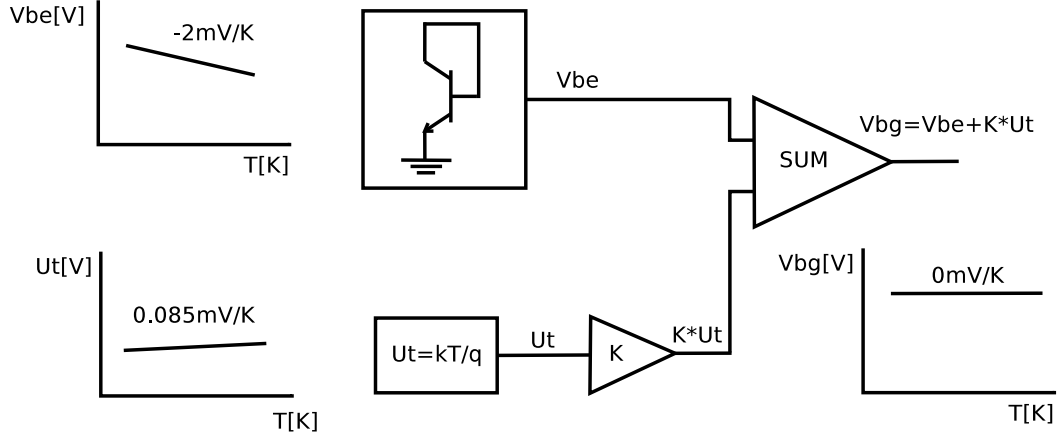
Figure 28: Principle of bandgap

4.2 Bandgap Basic Topology

4.2.1 Voltage Adding Topology

Fig.[29] shows voltage adding BGR topology [All02], that contains a Opamp, diodes, mosfets and resistors.

Assuming ideal Opamp we can write Eq.[47]

$$V_a = V_b = V_{be1} \tag{47}$$

The current I_{Rb} can be defined by the Eq.[48] below.

$$I_{Rb} = V_{Rb}/R_b = \Delta V_{be}/R_b = U_T \frac{ln(N)}{R_b} \tag{48}$$

Output voltage V_{ref} can be expressed by adding voltage across resistor R_{bg} and voltage V_b, Eq.[49].

$$V_{ref} = V_{rbg} + V_b \tag{49}$$

Voltage V_{rbg} can be expressed by the following Eq.[50].

$$V_{rbg} = R_{bg}I_{Rb} = R_{bg}\left(U_T \frac{ln(N)}{R_b}\right) \tag{50}$$

a final relation for V_{ref} is written Eg.[51].

33
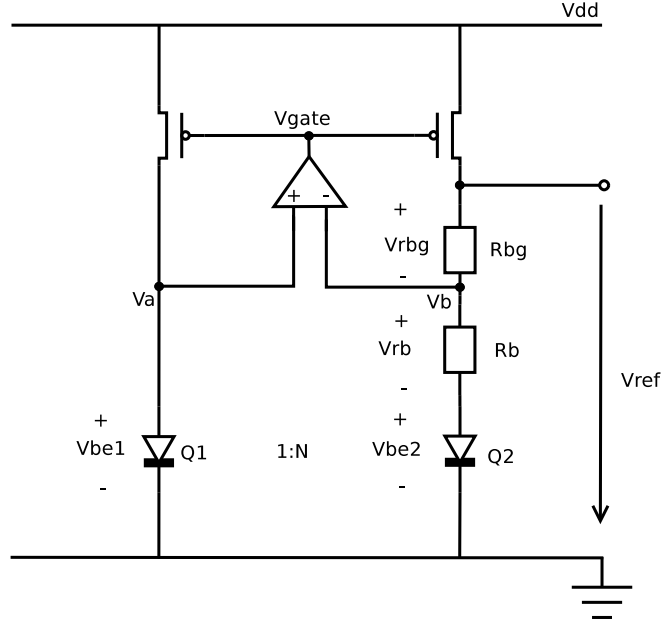
Figure 29: Voltage summing topology

$$V_{ref} = V_{rbg} + V_b = \frac{R_{bg}}{R_b}(U_T ln(N) + V_{be1}) \tag{51}$$

From Eg.[51] a typical PTAT(U_T) and CTAT(V_{be1}) terms are expressed. This circuit provides, from its nature, V_{bg} around 1.12 V and therefore supply voltage cannot be lower than 1.12 V. Due to this limitation, this topology is not suitable for low voltage reference.

4.2.2 Current Summing Topology

This concept was invented by H. Banba, H. Shiga and collective [BSU$^+$04]. The main advantage is sub-1-V operation with selectable output voltage V_{bg}. Fig.[30] presents this topology.

Voltages V_p and V_n are the same due to the Opamp and feedback loop. Feedback loop is made by the PMOS transistor P-BGRN, P-BGRP, which controls currents through branches. Because these PMOS devices have the same V_{GS} voltage, the currents through branches are the same.
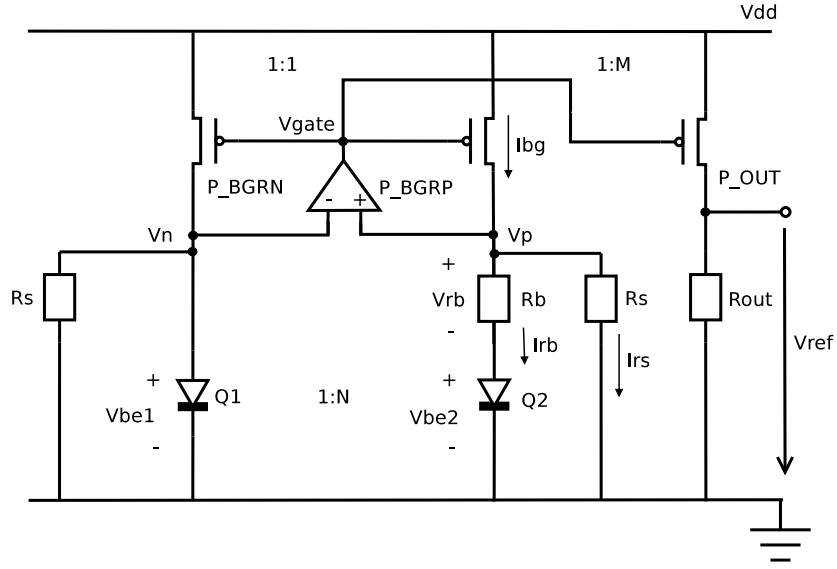
Figure 30: Principle of current Bandgap

From these considerations, the following equations are valid.

$$I_{bg} = I_{rb} + I_{rs} \qquad (52)$$

$$V_p = V_n \qquad (53)$$

$$V_{ref} = M(R_{out}(I_{rb} + I_{rs})) \qquad (54)$$

Currents I_{rb} and I_{rs} are expressed by Eq.[55,56]

$$I_{rs} = \frac{V_{be1}}{Rs} \qquad (55)$$

$$I_{rb} = \frac{V_{rb}}{R_b} \qquad (56)$$

where V_{rb} is ΔV_{be}.

$$V_{rb} = \Delta V_{be} = U_T ln(J_{c2}/J_{c1}) \qquad (57)$$

With Eq.[52,54,55,56,57] a final Eq.[58] for output voltage V_{bg} is written.

$$V_{ref} = M \frac{R_{out}}{R_s} \left(V_{be1} + \frac{R_s}{R_b} U_T ln(N) \right) \tag{58}$$

In Eq.[58] the term $(V_{be1} + \frac{R_s}{R_b} U_T ln(N))$ is bandgap voltage itself. It obtains both PTAT term (U_T) and CTAT term (V_{be1}) and its value is about 1.12 V as we can see from the voltage adding topology.

Notice that by the resistor ratio R_{out}/R_s and by the multiple factor M, the output voltage V_{ref} can be scaled.

The PMOS current mirror is set to saturation region with V_{ds} voltage about 120-140 mV, the V_p or V_n has maximum value about 0.7 V (this voltage is set by diode Q1 in low temperature case). So the current BGR has potential to work at the low supply voltage around 0.85 V. Unfortunately, building a high-gain low-voltage low-offset rail-to-rail Opamp can be quite difficult.

5 Offsets and Technology Variations

5.1 Offset Definition

The zero offset is one of the most wanted attributes of the circuit. Because it can cause large error in high gain Opamp application and it generally limits the precision of the analog design. Let's have a single output Opamp with input terminals tied together. The differential input voltage $V_{in,d}$ (voltage between the input terminals) is zero and the output voltage should be also zero (according to Eq.[59]).

$$V_{out} = A_V \cdot V_{in,d} \tag{59}$$

Unfortunately, this is not true and therefore $V_{out} \neq 0$. The offset voltage can be defined as the differential input voltage V_{OS} that is required for the zero output voltage, see Fig.[31].

Offset can be divided into systematic offset and random offset. The systematic one is caused by designer.

The random offset is caused by technological variations. It cannot be usually predicted because it is random. Fortunately, statistical result can
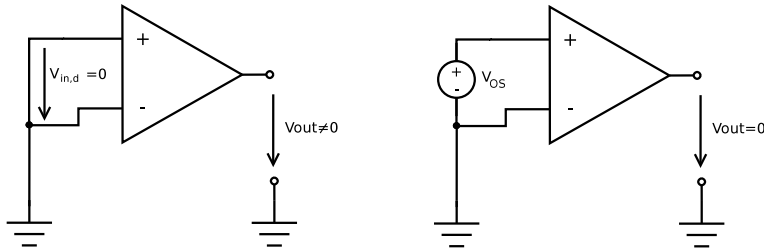
Figure 31: Definition of offset

be obtained by Monte Carlo simulation. The technology variations are now introduced.

5.2 Technology Variations

Technology variations have significant effect on analog precision and influence device performance. Clearly, geometrical size of manufactured MOS transistor differs from geometrical size of designed transistor (due to many difficult fabrication steps).

These effects can be categorized to process variations and mismatch variations. The process variation is applied on whole wafer, while mismatch variations are applied on each individual device.

5.2.1 Process Variations

There are many sources of process variations, for instance different temperature in fabrication steps, different dopant concentration, different thickness of growing oxide etc.

Several device parameters are used to estimate the impact of process variations. These models are called process corners. Generally slow and fast corner is defined for PMOS and NMOS devices. Junction capacitances and threshold voltage V_{th} in slow corner case are larger than expected in TT corner (TT corner is a case without any process variations). Therefore digital circuits are slower in slow corner.
On the other hand, fast corner reduces threshold voltage V_{th} and junction capacitances and that is why the digital devices are faster.

Five corners: SS, SF, FS, FF and TT are presented.

SS (NMOS slow, PMOS slow)

SF (NMOS slow, PMOS fast)

FS (NMOS fast, PMOS slow)

FF (NMOS fast, PMOS fast)

TT (NMOS typical, PMOS typical)

Every design should be tested in different corners, temperatures and supply voltages.

5.2.2 Mismatch Variation

As mentioned earlier, process variations are applied on the whole wafer, but mismatch variations are applied on each device. For instance, two identical designed MOS transistors have random differences in their parameters (threshold voltage V_{th}, body factor γ, technological factor κ, geometrical ratio W/L). These random variations can be established by the Gaussian distribution with an average μ and a spreading σ^2. According to Pelgrom's model [PTV98], following Eq.[60,61,62] describes MOS mismatch variance.

$$\sigma^2(V_{th}) = \frac{A_{vth}}{\sqrt{WL}} \tag{60}$$

$$\sigma^2(\kappa) = \frac{A_{\kappa}}{\sqrt{WL}} \tag{61}$$

$$\sigma^2(\gamma) = \frac{A_{\gamma}}{\sqrt{WL}} \tag{62}$$

where A_{vt}, A_{κ}, A_{γ} are experimental constants depending on technology. It is clear that WL size must be increased for better matching (i.e. for mismatch variations reduction). Unfortunately, mismatch variations could rise with an improper chip layout. Therefore the following layout techniques should be used. For instance, matching devices should be placed on the same isotherm, thus the temperature difference should be minimized. Matching

devices should have the same orientation and optionally the same area ratio. Centroid layout topology or dummy devices should be used for better matching, [SW94].

The main goal of mismatch and process optimization is to make a robust design. If the designed circuit is robust the technology variations does not affected its performance. Moreover the wafer yield is higher. The wafer yield is a number of acceptable chips produced per wafer. It is clear that the higher the wafer yields lead to a higher profit.

6   Circuit Design

Bandagap core, trimming circuit, operational transconductance amplifier (OTA) and start-up circuit are the main parts of chosen topology. Design of these four circuit blocks are presented in this chapter.

6.1   Design of the Bandgap Core

According to specification, the current summing BGR was chosen, [BSU$^+$04]. This topology is introduced in chapter [4.2.2] and it is shown again in FIG.[32]

Figure 32: Principle of the current BGR

Values of Rb, Rs, Rout, N and multiplying constant M are derived for proper BGR design. These values can be calculated from Eg.[63] derived in chapter [4.2.2].

$$V_{ref,WOF} = M\frac{R_{out}}{R_s}\left(V_{be1} + \frac{R_s}{R_b}U_T ln(N)\right) \tag{63}$$

Note that Eq.[63] shows the $V_{ref,WOF}$ voltage without the offset. Values M and ratio $\frac{R_{out}}{R_s}$ are used only for the absolute $V_{ref,WOF}$ value scale. Eg.[64]

shows the partial derivative of the bandgap voltage term $\left(V_{be1} + \frac{R_s}{R_b}U_T ln(N)\right)$ over the temperature.

$$\frac{\partial V_{ref}}{\partial T} = \frac{\partial \left(V_{be1} + \frac{R_s}{R_b}U_T ln(N)\right)}{\partial T} = \frac{\partial V_{be1}}{\partial T} + \frac{R_s}{R_b}ln(N)\frac{\partial U_T}{\partial T} \qquad (64)$$

To obtain ideally zero TC the Eg.[64] is equal to zero.

$$0 = \frac{\partial V_{ref}}{\partial T} = \frac{\partial V_{be1}}{\partial T} + \frac{R_s}{R_b}ln(N)\frac{\partial U_T}{\partial T} \qquad (65)$$

The exact value of partial derivation $\frac{\partial V_{be1}}{\partial T}$ was simulated at temperature $T = 300K$, see Fig.[20].

$$\frac{\partial V_{be1}}{\partial T} = -2.27mV/K \qquad (66)$$

The $\frac{\partial U_T}{\partial T}$ can be estimated by Eq.[67].

$$\frac{\partial U_T}{\partial T} = \frac{\partial \left(\frac{kT}{q}\right)}{\partial T} = \frac{k}{q} = \frac{1.38 \cdot 10^{-23}}{1.60 \cdot 10^{-19}} = 86.2\mu V/K \qquad (67)$$

Eq.[68] is obtained by substitution of Eq.[66],[67] into the Eq.[65]

$$0 \simeq -2.27 \cdot 10^{-3} + \frac{R_s}{R_b}ln(N)(86.2 \cdot 10^{-6}) \qquad (68)$$

Let's express the $\frac{R_s}{R_b}$ ratio.

$$\frac{R_s}{R_b} \simeq \frac{2.27 \cdot 10^{-3}}{86.2 \cdot 10^{-6} \cdot ln(N)} \simeq \frac{27.5}{ln(N)} \qquad (69)$$

41

From Eq.[69] it is clear that current density ratio N must be chosen. This parameter said, how many times current density in diode Q1 is larger then in diode Q2. This parameter has a significant effect on the reference generator performance.

Assume that the Opamp has an input referred offset voltage V_{OS}. Therefore Eq.[70] is valid.

$$V_n = V_p + V_{OS} \tag{70}$$

Currents I_{bg}, I_{rb} and I_{rs} are again derived.

$$I_{bg} = I_{rb} + I_{rs} \qquad I_{rb} = \frac{V_{rb} \cdot V_{OS}}{R_b} \qquad I_{rs} = \frac{V_{be} \cdot V_{OS}}{R_s} \tag{71}$$

Thus the reference voltage V_{ref} is expressed by the Eq.[72]

$$V_{ref} = R_{out} \cdot I_{bg} = R_{out} \left(\frac{V_{rb} \cdot V_{OS}}{R_b} + \frac{V_{be} \cdot V_{OS}}{R_s} \right) \tag{72}$$

After few calculation steps, Eg.[73] is derived. This relation shows unwanted dependence of the V_{ref} on the offset voltage V_{OS}

$$V_{ref} = V_{ref,OF} + V_{OS} = \frac{R_{out}}{R_s} \left(V_{be1} + \frac{R_s}{R_b} U_T ln(N) + V_{OS}(1 + \frac{R_s}{R_b}) \right) \tag{73}$$

Mismatch variations of the BGR core and the Opamp are the main contributors of the input offset voltage V_{OS}. Unfortunately, mismatch reduction is not easy task and every Opamp has the input referred offset.
From Eq.[73] is clear, that a small $\frac{R_s}{R_b}$ ratio can reduce the effect of V_{OS} on the V_{ref}. The large value of the N parameter provides small $\frac{R_s}{R_b}$ ratio, Eq.[69].

Moreover for good matching, the common centroid geometry is used. Thus N equal to $n^2 - 1$ is required for the square topology, see Fig.[33]. The

N parameter is chosen to be equal 24.
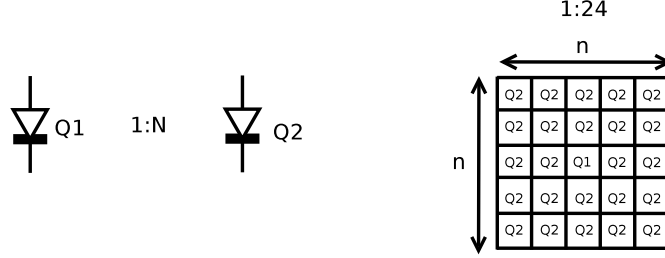
Figure 33: The common centroid geometry of diodes Q1,Q2 (N=24)

Now $\frac{R_s}{R_b}$ ratio can be calculated.

$$\frac{R_s}{R_b} \simeq \frac{27.5}{ln(24)} = 8.64 \tag{74}$$

Note that, this calculation is for the $T = 300K$. Different temperature results in the different ratio. It is therefore tentative calculation. The $\frac{R_s}{R_b}$ ratio is chosen to be equal to 8.

Now let's focus on the scaling factors $M \cdot \frac{R_{out}}{R_s}$. The current mirror multiplying factor M is chosen to $M = 1$ to reduce total current consumption. The $\frac{R_{out}}{R_s}$ ratio is chosen according to the required V_{ref} value. The $V_{ref} \simeq 0.7V$ is chosen because this voltage level creates a sufficient common mode voltage to the N-input stage of the subsequent Opamps.

$$V_{ref} \simeq M \frac{R_{out}}{R_s}(1.12) \simeq \frac{R_{out}}{8R_b}(1.12) \simeq \frac{5R_b}{8R_b}(1.12) \simeq 0.7V \tag{75}$$

As Eq.[75] shows, the $R_{out} = 5R_b$ is chosen for $V_{ref} \simeq 0.7V$.

The absolute value of R_b is still unknown. To satisfy the low power requirement $I_{rb} = 100nA$ is chosen.

$$R_b = \frac{V_r b}{I_r b} = \frac{U_T ln(24)}{100 \cdot 10^{-9}} = 826.3k\Omega \tag{76}$$

43

Table [3] summarizes the previous design choice.

Parameter	Designed Value
R_b	826.3kΩ
R_s	$8R_b$
R_{out}	$5R_b$
M	1
N	24

Table 3: BGR core design choice

PMOS current sources design. To achieve proper function of the PMOS current sources (P-BGRN, P-BGRP, P-OUT), the saturation operating region is required. The current through the PMOS transistors is set to 200nA (according to low power requirement). According to g_m/I_D design approach, the inversion factor is now chosen. The moderate or strong inversion is recommended for current mirror reduction (see chapter [6.3] for current error explanation). The IF equal to 9 is chosen. Using the previous mentioned Eq.[15] the $W/L = 1/6$ ratio is obtained. The length is set to 12um for mismatch reduction. For the selected W/L ratio and MOS length L the MOS width can be calculated $W = 2um$.

Note that too high inversion factor should not be chosen in this case, because the V_{DSat} would be to large. Low supply voltage $Vdd = 0.9V$ and quite high output voltage $V_{ref} = 0.7V$ left only 0.2V for V_{DS} of P-OUT transistor. It is clear, that for saturation region the V_{DSat} must be less then $V_{DS} = 0.2V$.

6.2 Design of the Trimming Circuit

It can be said that every design device has some variation from the designed value due to the technology variations and local mismatch, see Ch.[5]. In the previous chapter, the particular resistor value is designed according to our task. Due to the technology variation the absolute value of resistor can be

changed by the $\pm 20\%$, [Car12]. On the other hand, the relative variations (variation of ratio) are approximately $\pm 1 - 4\%$. Even this small variation can cause major inaccuracies.

It is clear that for achieving an accurate voltage reference some calibration device must be added. This device is called the "trimming" circuit and it simply provides post-fabrication adjustment. The typical trimming circuit can tune only one element (i.e resistor value, capacitor value, MOS transistor width etc.). It is important to note that this technique cannot ensure an ideal circuit performance but only an acceptable performance per some cost. It is tradeoff between trimming complexity and precision.

In this thesis the resistor trimming network is chosen. There are a few basic resistor trimming techniques: the laser trimming, metal fuse and software programmable trimming.

The metal fuse method is based on the resistor with parallel metal fuse. This fuse has almost zero resistance. Under normal conditions, the fuse shorts the resistor. Thus value of this shorted resistor is not counted to the final value of the resistor network. On the other hand, the fuse can be opened/burned and then the resistor value is counted. The disadvantage is need of large trimming pads (on this pads a current is applied).

The laser trimming is more accurate and area efficient. This technique is based on the thin metal resistor film which is cut by the laser beam. Unfortunately this procedure is expensive. Therefore it is mainly used for high performance circuits.

The software trimming is based on the electronic switches that change the final value of the resistor network in time. The trimming process can be repeated or changed. This is the main advantage indeed. Moreover this trimming is the cheapest variant. On the other hand some digital circuit is needed. In some complex integrated circuit project the digital circuits are present anyway. So this is not necessarily such big disadvantage. This trimming technique is chosen for this project.
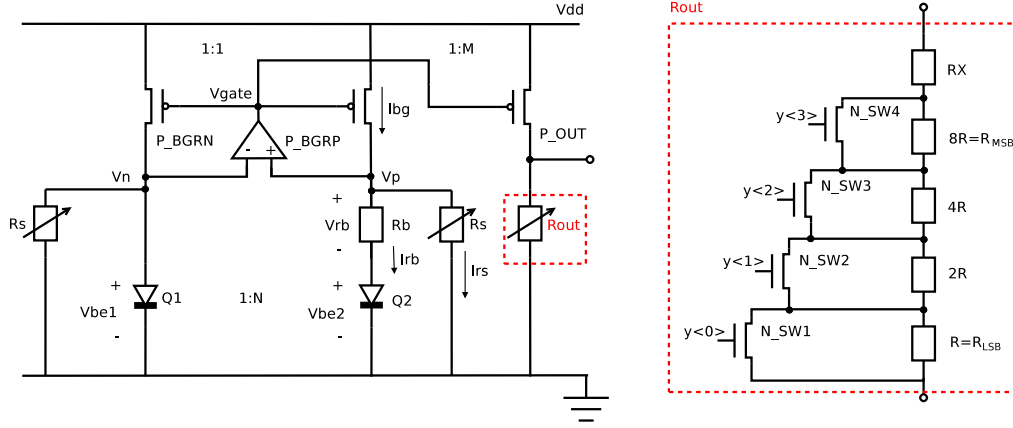
Figure 34: Trimming of the R_{out}, trimming resistor network

6.2.1 Trimming of the R_{out}

Fig.[34] shows designed trimming resistor network. This network is used to trim the value of the R_{out}. Thus the reference voltage V_{ref} can be tuned. For proper design is necessary to know the number of bits and maximum deviation of V_{ref}. Four bits are chosen for trimming. This choice is a compromise between trimming range/accuracy and cost. The maximum deviation due to the technology variation is determined by the Monte Carlo analysis. The deviation range roughly $\pm 125mV$ is detected, which is about $\pm 18\%$ of the nominal V_{ref}. In this situation trimming is linear. Thus the value of the R_{out} must be trimmed in the range of $\pm 18\%$, see Eq.[77]

$$R_{max} = R_{typ}(1.18) = 4.87M\Omega \qquad R_{min} = R_{typ}(0.82) = 3.38M\Omega \qquad (77)$$

where R_{typ} is nominal value of $R_{out} = 5R_b = 4.13M\Omega$ designed in previous chapter. From the R_{max}, R_{min} the trimmed range is obtained, Eq.[78]

$$\triangle R = R_{max} - R_{min} = 1.49M\Omega \qquad (78)$$

46

Note from Fig.[34] that binary weighted resistors are used. Resistors RX and $8R = R_{MSB}$ create the typical resistor value $R_{typ} = RX + 8R$. This R_{typ} can be trimmed to the maximum value $R_{typ} + 7R$ and to the minimal value $R_{typ} - 8R$ by the proposed structure. The values of R_{LSB} and R_{MSB} are defined by the Eq.[79].

$$R_{LSB} = \frac{\triangle R}{15} = 99.2k\Omega \qquad R_{MSB} = 8R_{LSB} = 793k\Omega \tag{79}$$

In the end the maximal resolution can be calculated by

$$Max.Resolution = \frac{R_{LSB} \cdot 100}{R_{typ}} = 2,03\% \tag{80}$$

The other resistor value can be calculated form the R_{LSB} value.

It is worth to mention that the smaller value of R_{LSB} the more accuracy but the narrower trimming range. To obtain high resolution and wide range network a large number of bits is required. This also relates with silicon area, cost and test-times. Therefore some compromise must be chosen.

6.2.2 Trimming of the R_s

TC can be trimmed by change of the Rb or by Rs resistor value. Trimming of the Rb is improper choice. Because voltage drop across Rb is given by $V_{rb} = U_T \cdot ln(N) \simeq 82.6mV$, (for N=24). This voltage drop is not sufficient for trimming switches. Therefore Rs resistors trimming is chosen (voltage drop across resistors Rs are equal to voltage across diode Q1).

Trimming structure is the same as structure of the R_{out} trimming, see Fig.[34]. The required trimming range is $\pm 8.5\%$ (it is obtained from Monte Carlo simulation). According to four bits possibilities and required trimming range $\pm 8.5\%$, the $R_{max}, R_{min}, R_{LSB}$ and maximal resolution are calculated.

$$R_{max} = 6.46M\Omega \qquad R_{min} = 7.67M\Omega \tag{81}$$

$$R_{LSB} = 80.5k\Omega \qquad Max.Resolution = 1,13\% \tag{82}$$

The whole schematics of the trimming circuits can be found in appendix [A.5, A.6].

6.3   Design of the Opamp

From the proposed circuit it is clear that the Opamp is required. The OTA is operational transconductance amplifier. This device is type of the Opamp but the output is not a voltage v_{out} but it is a current i_{out}, see Fig.[35]. The output node has therefore high impedance. The input is the differential voltage $V_{in,d}$ and thus input nodes has a high impedance. The whole device can be approximated/idealized as the one voltage controlled current source (VCCS). In the following text the Opamp and OTA words are considered to be the same.
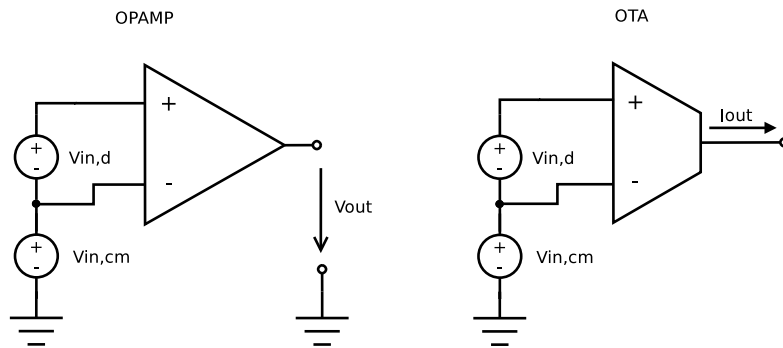
Figure 35: Opamp and OTA symbols

The BGR performance strongly depends on the Opamp performance. Therefore a robust Opamp is a crucial point of the whole BGR.

Let's summarize the Opamp requirements. The Opamp must work with the minimal supply voltage 0.9V. Due to the low voltage requirements the circuit topology range is limited. The next requirement is the wide input swing. The common mode voltage range is roughly 0.3-0.7 V (this range is given by the diodes temperature behaviour). Also low gain and high offset have impact on the BGR precision.

From the basic structure the telescope OTA cannot be used due to the low supply voltage. Also any two stage OTA is not preferred. The BGR

48

core is basically the next stage and therefore it creates the three stage system (with any two stage OTA). Generally the two stage system is easier to compensate[2]. Thus only one stage OTA is preferred.

The remaining structures are the single stage OTA, symmetrical OTA and folded cascode OTA. Previously mentioned OTA structures are outside the scope of this thesis. Explanation can be found in the [SW94].

6.3.1 Folded Cascode OTA with NMOS Differential Pair

The folded cascode OTA is chosen because it can work from the lowest supply voltage. It provides wide output swing, high gain and it can be modified for the rail to rail input stage. Fig.[36] shows the N-folded cascode OTA topology.
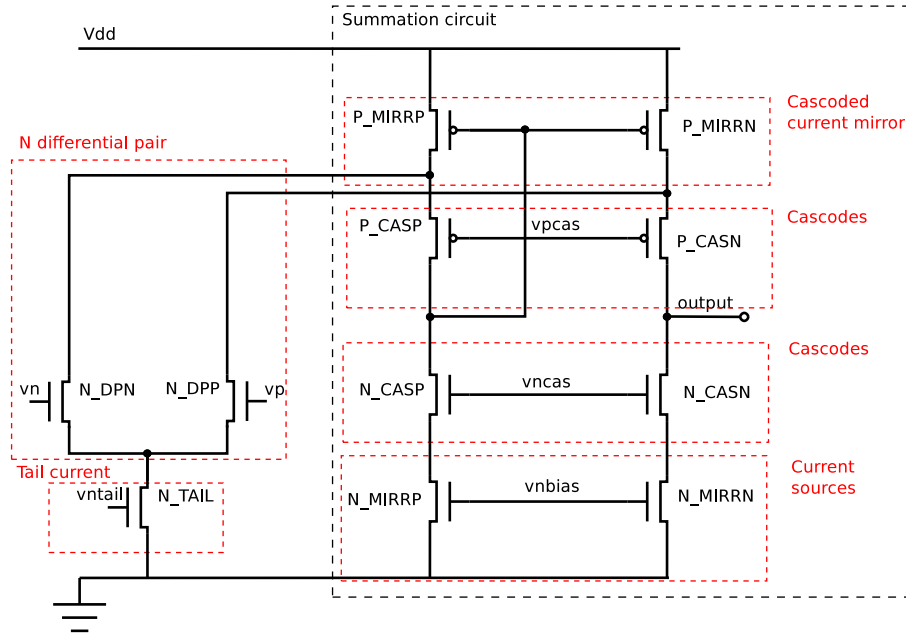
Figure 36: Folded cascode OTA with NMOS differential pair

This topology contains two parts: input differential stage (NMOS input differential pair, NMOS tail current) and the summation circuit (PMOS cur-

---

[2]compensate means ensure sufficient stability

rent mirrors, NMOS current sources, cascodes). These fundamental parts are now presented.

Differential Pair. Differential pair is made of two NMOS transistors N-DPP and N-DPN. The main function is convert the input differential voltage to the differential current. The diff.pair is biased[3] by the current source N-TAIL. For proper work of the N type diff.pair, Eq.[83] must be valid. This equation limits the minimal value of the $V_{in,cm}$.

$$V_{in,cm} \geq V_{GS} + V_{DS} \tag{83}$$

where V_{DS} is the voltage across drain-source of the N-TAIL transistor and the V_{GS} is the gate-source voltage of the N-DPP,N.

Current mirror and Cascodes. The current mirror is the most basic analog structure. The function is simply sense and reproduce the current. There are many types of current mirrors. Simple current mirror, Wilson, Cascode, [SW94]. Fig.[37] shows the N type basic, cascoded and low voltage current mirror.
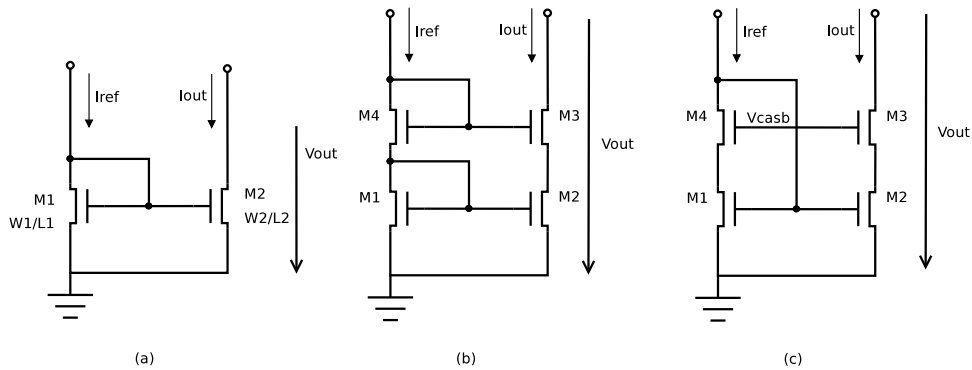
Figure 37: Current mirrors (a) basic (b) stacked (c) low voltage cascode

---

[3]biasing = setting the proper operational point

The important attribute is the current mirror error. It can be caused by different V_{DS} voltages and by mismatch. Increasing the output resistance (by adding the cascode or by longer channel) reduces the V_{DS} sensitivity but it also reduces the output swing. Eq.[84] shows the relation between I_{out} and I_{ref} of the basic current mirror.

$$\frac{I_{out}}{I_{ref}} = \left(\frac{K_2}{K_1}\right) \left(\frac{W_2/L_2}{W_1/L_1}\right) \left(\frac{V_{gs} - V_{T2}}{V_{gs} - V_{T1}}\right) \left(\frac{1 + \lambda V_{DS2}}{1 + \lambda V_{DS1}}\right) \tag{84}$$

Note, that the $\left(\frac{W_2/L_2}{W_1/L_1}\right)$ term is used to set the current ratio. The rest terms only describe the unwanted dependency. If the devices are matched and if the V_{DS} voltages are same ($V_{DS1} = V_{DS2}$) then Eq.[84] can be simplified as follows:

$$\frac{I_{out}}{I_{ref}} = \left(\frac{W_2/L_2}{W_1/L_1}\right) \tag{85}$$

Fig.[38] shows test of the current mirrors (basic, stacked, cascode). The current mirror error is caused by the different V_{DS} voltage. All these mirrors have the same value of testing current $I_{test} = 200nA$. It is clear that stacked current mirror provides smaller mirror error than the basic and cascode current mirror.

On the other hand the stacked mirror needs the largest voltage V_{out} for its proper function ($V_{out,min} = 2V_{ov} + V_{th}$). While the basic current mirror requires only ($V_{out,min} = V_{ov}$). Cascode current mirror provides compromise between the current error and minimal $V_{out} = 2V_{ov}$ voltage, [All02].

The P-MIRRP and P-MIRRN create the P current mirror. This mirror uses cascodes P-CASP, P-CASN to increase the output impedance. This topology provides the high output impedance but it does not reduce the output swing as the stacked current mirror.
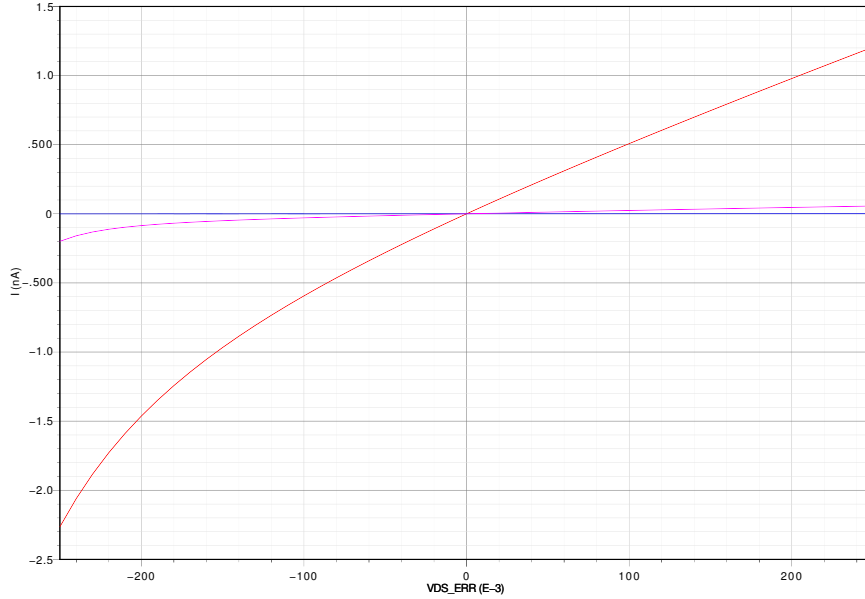
Figure 38: Current mirror error vs. V_{DS} for basic (red), stacked (blue) and cascoded (violet)

The current sources are used to set the current through the structure. Basically it is also the extended current mirror as can be seen in the whole schematic. N-MIRRP and N-MIRRN should have a long channel length, to minimize effect of V_{DS} mismatch.

Bias circuits. Any designed transistor needs to be set to the correct operating point. Thus a bias is required. The bias circuit can provide bias voltage (i.e gate voltage of cascodes) or bias current (i.e bias current of N-TAIL). The whole bias circuit basically works as the current mirrors network. Therefore ideal current mirrors and one current source are needed to replicate bias current I_{ref} without additional mirror error, see Fig.[39(a)].

The Fig.[39(b)] shows the simple bias realization by the diode-connected MOS transistor and resistor. This easy circuit heavily depends on supply voltage, temperature and resistance value. Therefore it is not a convenient choice indeed.

The interesting solution is to use the I_{bg} current from the BGR. As men-

Figure 39: Bias network with ideal source (a) and with R_{REF} resistor (b)

tioned before, this current should be temperature and supply voltage independent. It seems to be great choice for bias circuit, but there is one issue to note. The I_{bg} current is the reference current only if the BGR works in proper operating point (and if the Opamp is turned on). But the Opamp cannot works without I_{bg}. These statements form a condition loop.

The specific biased solutions are now presented.

Figure 40: Current sources biasing

Current sources biasing, Fig.[40]. N-MIRRP,N bias is basically only the current mirror with current source made by transistor P-OUT1. Note

that transistor N-BIAS1 is a part of the bias structure not a part of summing structure. The ratio of N-BIAS1/N-MIRRN,P is set to 1/2 for current saving. The current density of each transistor in current mirror should be the same. Therefore the multiplying factor M is used. Note from the Eq.[83] that for better current matching[4] it is good to achieve the same voltages $V_{DS1} = V_{DS2}$.
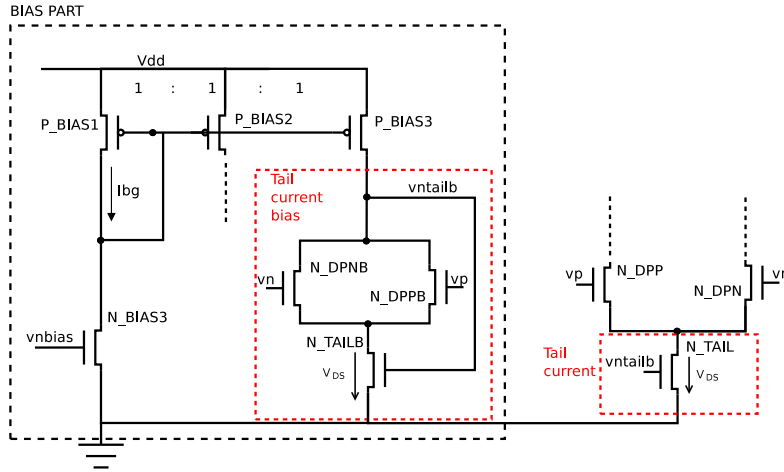
Figure 41: Tail current biasing

N-TAIL biasing, see Fig.[41]. Note that it is the basic current mirror with input diff.pair replica N-DPNB, N-DPPB. These two transistors help to keep V_{DS} voltages of N-TAILB and N-TAIL same. In this case it is more important than in case of current mirror. Moreover this feature does not need any extra current. The main disadvantage is that if the input common mode voltage $V_{in,cm}$ is low than the diff.pair replica is shut off. Thus the current cannot flow through this branch. This issue makes the Opamp startup more difficult.

The cascode biasing. Fig.[42] shows the biasing for N cascode. The principle of P cascode bias is the same. Thus only N cascode biasing is now introduced. The voltage applied on the gate of cascodes v_{ncas} must be $v_{ncas} \geq V_{Dsat,NMIRR} + V_{GS,NCAS}$. The $V_{Dsat,NMIRR}$ voltage is the necessary voltage to keep the N-MIRRP,N in the saturation region and $V_{GS,NCAS}$ is the

---

[4]the matching term means that two or more parameters is the same
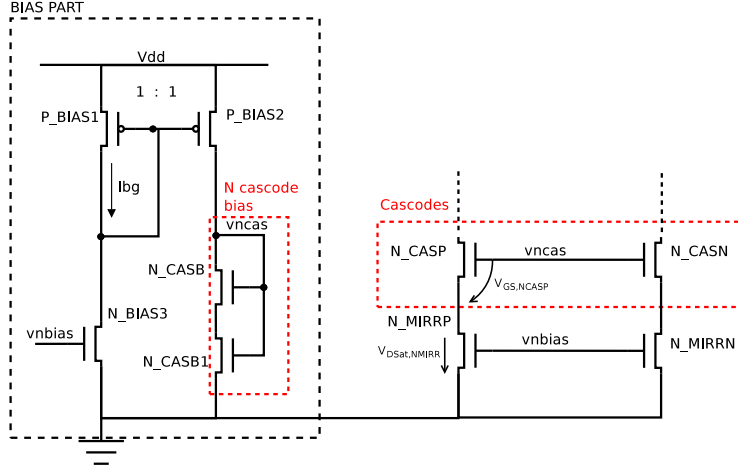
Figure 42: Cascode biasing

gate-source voltage of the cascode. As mentioned before, to obtain smaller current error of the current mirror/source, the strong inversion is required. Thus the $V_{Dsat,NMIRR}$ rises and the V_{ncas} should rise too (if the $V_{GS,NCAS}$ is constant). Unfortunately, this leads to headrooms reduction of cascodes N-CASP,N. It is obvious that some trade-off must be chosen. Transistor N-CASB1 is in resistive region. The resistance of the N-CASB1 is chosen by the W/L of the N-CASB1. The chosen bias current and the resistance value give the appropriate V_{DS} voltage of the N-CASB1. Thus v_{ncas} is shifted up by the voltage across the N-CASB1. Thus Eq.[86] is valid.

$$v_{ncas} = V_{DS,CASB1} + V_{GS,CASB} \geq V_{Dsat,NMIRR} + V_{GS,NCAS} \qquad (86)$$

The whole biasing circuit is shown on the Fig.[43]. Fortunately the circuit can be optimized to lower current consumption, as the Fig.[44] shows. Note that branch made by transistors N-BIAS3 and P-BIAS1 is skipped. The N cascode bias block and the N tail bias block are now supplied directly from BGR core by transistors P-OUT1,2,3.
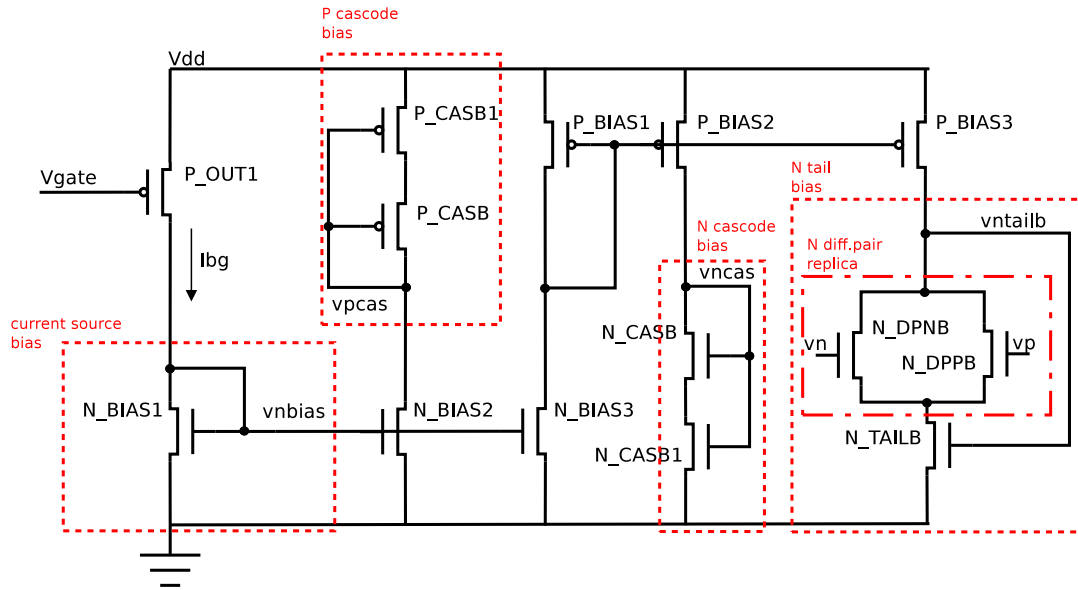
Figure 43: Self-biased network
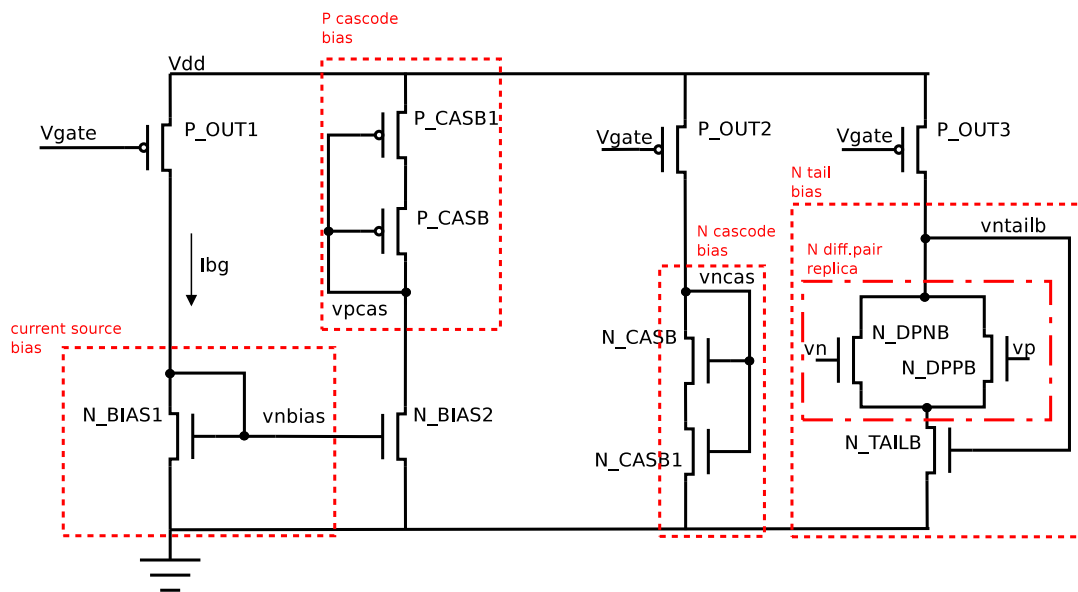
Figure 44: Upgraded self-biased network

6.3.2 Rail-to-Rail Folded Cascode OTA

Due to the wide $V_{in,cm}$ range and low supply voltage, the N diff.pair is not able to work at the whole temperature range. Thus the rail to rail (R2R) input stage is presented. The rail to rail input is able to work with the $V_{in,cm}$ equal to zero or supply voltage.

Fig.[45] shows the R2R input stage. It contains the N diff.pair and also the P diff.pair. The P diff.pair has also a condition for required $V_{in,cm}$ level, see Eq.[87]. Eq.[88] shows the $V_{in,cm}$ level condition for N diff.pair.
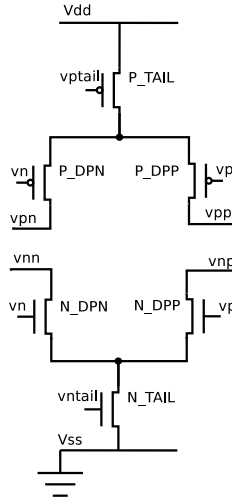
Figure 45: Rail to rail input stage

$$V_{in,cm} \leq V_{dd} - V_{DSat} - V_{GS} \tag{87}$$

$$V_{in,cm} \geq V_{GS} + V_{DSat} + V_{SS} \tag{88}$$

Input operating range can be estimated from Eq.[87,88]. The well known problem is with the unequal value of the total transconductance $g_{mtot} = g_{m,p} + g_{m,n}$, see Fig.[46]. The g_{mtot} reaches its maximum in the middle of the $V_{in,cm}$ because both N diff.pair and P diff.pair work and their transconductance are summed together. On the other hand at the each side of $V_{in,cm}$,
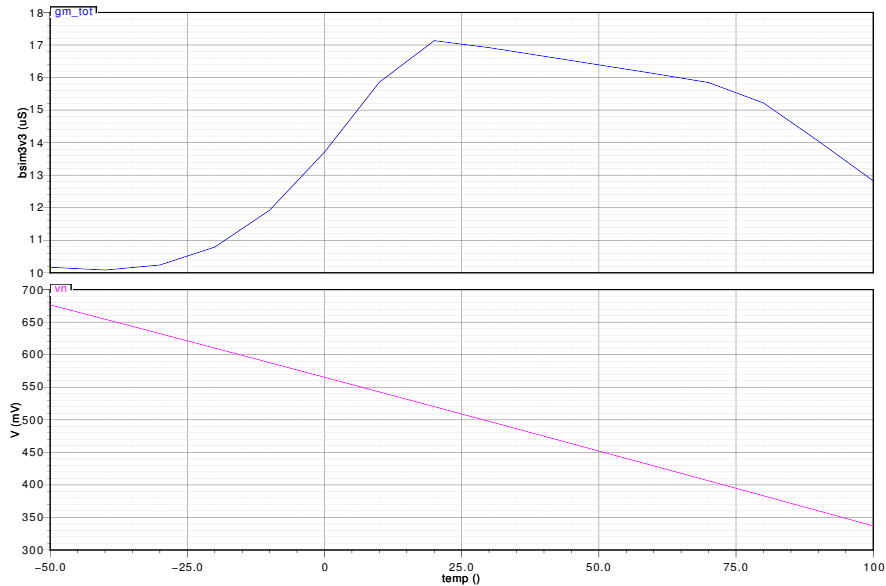
Figure 46: g_{mtot} and $V_{in,cm}$ temperature dependence

only one diff.pair works. Thus the g_{mtot} is smaller there.

The Fig.[47] shows designed R2R folded cascode OTA. As mentioned before the offset of the Opamp has major influence on the BGR precision. Therefore the offset reduction is one of the main goals.

Unfortunately, this structure provides unwanted input referred offsets on the 1V supply voltage, see Fig.[48 - left window]. These simulation results are obtained from temperature sweep of reference voltage and input referred offset of the Opamp, these simulations were repeated by Monte Carlo for different process corner and mismatch (20 runs). As it can be seen, the Opamp input offset is transferred to the reference voltage V_{ref}, Fig.[48 - right window]. Moreover the offsets are not constant over temperature range. This is serious problem. It follows to additional bend of the V_{ref} curve, see Fig.[48 - right window]. With the trimming circuit the constant offset can be compensated to the nominal V_{ref} value. But the extra bend character simply devastates the BGR stability over temperature.

The same simulation with $V_{dd} = 1.2V$ is made and the problem vanishes, see the Fig.[49].

58
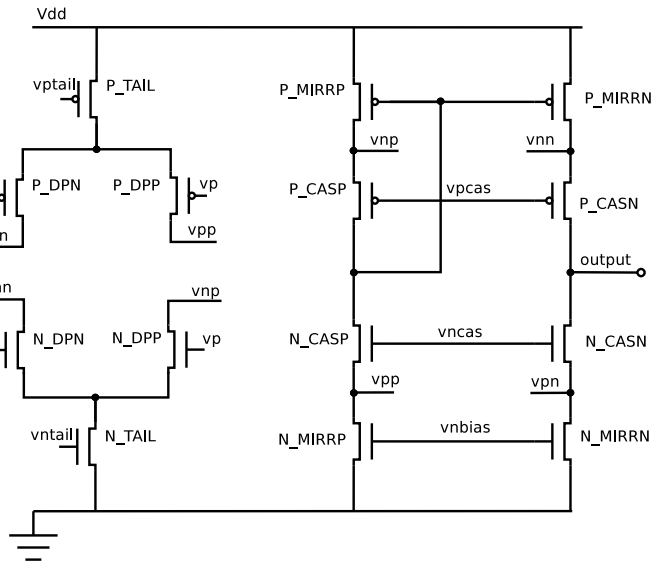
Figure 47: Rail to rail OTA

Figure 48: Offset issue Vdd=1V, (Input referred offset and reference voltage as a function of temperature, Monte Carlo 20 runs)
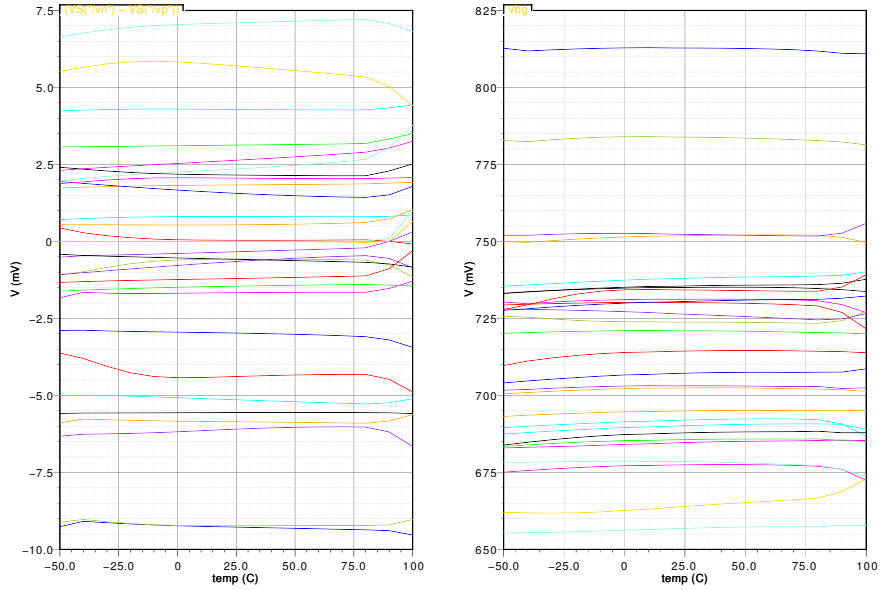
Figure 49: Offset issue Vdd=1.2V, (Input referred offset and reference voltage as a function of temperature, Monte Carlo 20 runs)

The reason of these problems is in the P diff.pair function. When low V_{dd} is applied the P diff.pair has not enough headroom and it is turned off. While in the $V_{dd} = 1.2V$ case, the P diff.pair is always on. A significant current changes occur in the summing circuit due to this difference.

Unfortunately the supply voltage condition is 0.9-2V, therefore this topology cannot be used. Several constant-gm methods exist [LHV11, DHYW10, WRZYWD09]. In the end the basic N-folded cascode OTA topology is chosen and upgraded, see next chapter.

6.3.3 Final OTA Design

The N-folded cascode OTA topology has already been presented and explained in the previous chapter. The following two equations describe the DC gain and offsets of this structure. These two factors are the main design goals.

$$A_0 = g_{m,1} \cdot R_{OUT} \tag{89}$$

Eq.[89] said that for high DC gain both $g_{m,1}$ and R_{OUT} need to be large. Thus the NMOS diff.pair should work in weak inversion where the g_m/I_D is maximum. And the output resistance of the summing circuit should be large, this can be achieved by long channel transistor or by adding the cascode.

$$V_{OS} = \frac{V_{OS,SYS}}{A_0} + \Delta V_{th,1} + \frac{g_{m6}}{g_{m1}}\Delta V_{th6} + \frac{g_{m11}}{g_{m1}}\Delta V_{th11} + \frac{V_{GS1} - V_{th,1}}{2}S \quad (90)$$

From Eq.[89] is clear, that the resulting offset is caused by many factors. First the systematic offset $V_{OS,SYS}$ should be minimized by design, its contribution is fortunately divided by the DC open loop gain. The ΔV_{th} is the threshold voltage mismatch, it can be reduced by increasing area, see chapter [5]. The $\frac{g_{m6}}{g_{m1}}$ term is a ratio of P-MIRR transconductance to N-DP transconductance, and $\frac{g_{m11}}{g_{m1}}$ is ratio of N-MIRR/N-DP transconductance. It is clear that for less offset contribution the N-MIRRP,N and P-MIRRP,N should have smaller transconductance than the N-DP. The last term express that the $V_{GS1} - V_{th,1}$ of N-DPP,N transistor should be small for offset reduction, this is also achieved by the choice of weak inversion, (the S factor obtains the sizes and κ mismatch), [SW94].

The NMOS differential pair requires weak inversion for maximum g_m/i_d. That is a proper choice for the high DC gain and small offsets. The weak inversion is achieved by chosen inversion factor IF=0.1 at current I_D equal to 400nA. Thus the W/L ratio is 6/1. The length is set to 2um to get reasonable mismatch. The MOS transistor area is important due to the random offset reduction, see Ch.[5].

The N-TAIL is a biased current source. It has twice the current of the diff.pair MOS transistor. The moderate inversion is chosen because it provides low V_{Dsat} voltage. The inversion factor equal to 5 is chosen, it is again compromise between low V_{Dsat} against the higher current mirror error. The drain current is 800nA. The W/L ration is then 1/4. Length is set to L=4um.

Unfortunately this choice makes characteristic curvature of the input offsets, see Fig.[48,49] at temperature range 75-100 °C. This is due to low $V_{in,cm}$.

61

NMOS diff.pair try to keep its V_{GS} and therefore V_{DS} of N-TAIL transistor and tail current are reduced.

Thus the N-TAIL size is set closer to weak inversion (IF=0.8).The aspect ratio is then W/L=3/2 and length is the same as before L=2um. By this choice N-TAIL operates in near to weak inversion. This region is not the most suitable for tail current device (N-TAIL) but it provides lower V_{Dsat}. Thus some current for NMOS diff.pair is still available at higher temperature and offset bending is improved, see Fig.[50] .
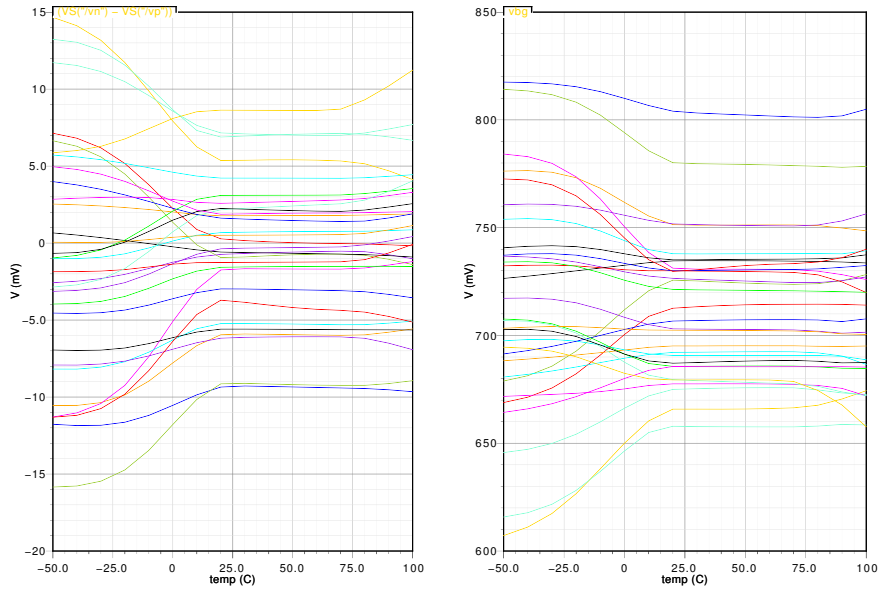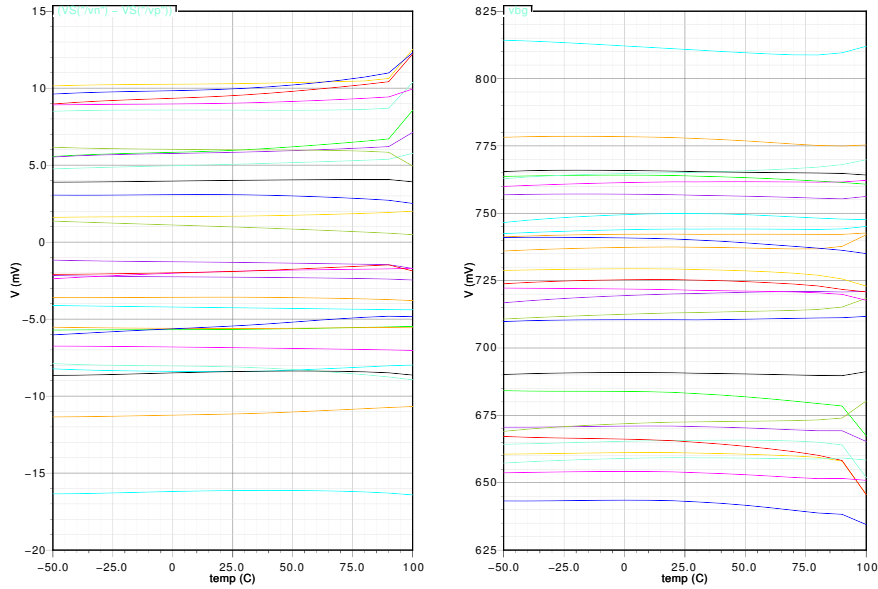

Figure 50: Offset Vdd=1V, (Input referred offset and reference voltage as a function of temperature, Monte Carlo 20 runs), upgraded N-TAIL

The P-MIRRP,N are set into the strong inversion, inversion factor is IF=15 and drain current is 1000nA. This region is ideal to reduce current mirror errors [Ste08]. Thus the W/L is 1/2 and length is set to L=8um, for random offset reduction and higher output resistance.

The P-CASP,N cascodes operate in closed to weak inversion, IF=0.5, I_D=800nA. They are set to W/L=12/1, L=1um. The g_m/i_d is high and therefore $g_m r_{ds}$ term is also high (the $g_m r_{ds}$ term improves the output impedance). Moreover the weak inversion provide low V_{Dsat} value and therefore the more

headroom is obtained.

The current sources N-MIRRP and N-MIRRN set the current through branches. These two NMOS transistor should provide equal current 800nA. The current mirror precision is achieved by strong inversion region. Unfortunately the inversion factor IF cannot be chosen to high (i.e IF=10-15), because the V_{Dsat} is too large and N-MIRRP,N transistors would be in linear region. Thus the DC gain and current mirror precision would drop.

Therefore the inversion factor is chosen to be near to the strong inversion IF=7.2. Thus the ratio is set to W/L= 1/6. The L=12um to increase the output resistance.

Opamp Current scaling down. Designed OTA is now resized to achieve lower current consumption. Two scale factors (SF_A, SF_B) are chosen. SF_B is the scale factor for the bias part and SF_A for the amplifier. These factors indicate how the current through transistor needs to be scaled. The SF_A is larger than SF_B factor (for instance $SF_A : SF_B = \frac{1}{32} : \frac{1}{40}$) because it is undesirable if bias part consumes more current then amplifier part. The example of one transistor resizing is shown next.

$$AR_0 = \frac{W}{L} = \frac{2}{1} \qquad SF_B = \frac{1}{40} \tag{91}$$

The $\frac{W}{L}$ ratio is called aspect ratio AR and it is given by the concrete inversion factor IF and by the drain current I_D. AR_0 is the original aspect ratio. The scaled aspect ratio AR_1 is obtained by multiplying the AR_0 with the proper scale factor.

$$AR_1 = AR_0 \cdot SF_B = \frac{2}{1} \cdot \frac{1}{40} = \frac{1}{20} = \frac{W_1}{L_1} \tag{92}$$

Now the new length L_1 must be chosen. The length choice is often a trade-off depending on specific basic block. The L_1=4um is chosen. Then the scaled MOS transistor width $W1$ is directly calculated.

$$W_1 = AR_1 \cdot L_1 = \frac{4}{20} = 0.2um \tag{93}$$

Unfortunately, the new width W_1=0.2um is not allowed by the PDK. The minimal width is W_{min}=0.4um. Moreover the MOS transistor width is chosen to be larger than W_{min} to obtain at least two contacts. This choice provides higher device reliability.

Thus the L_1=20um is chosen and width W_1=1um is calculated. The new transistor sizes (W_1, L_1) are obtained. In case of strong inversion (small W/L) the new length L_1 exceeds the 20um PDK limit. Therefore the new device composed from serial connected MOS transistors must be created.

For instance, two serial connected NMOS transistors W/L=2/20, L=20um create one NMOS transistor with ratio W/L=2/40 and length L=40um. On the other hand, a parallel connection increases the total width.

Each transistor in OTA needs to be scaled by this procedure.
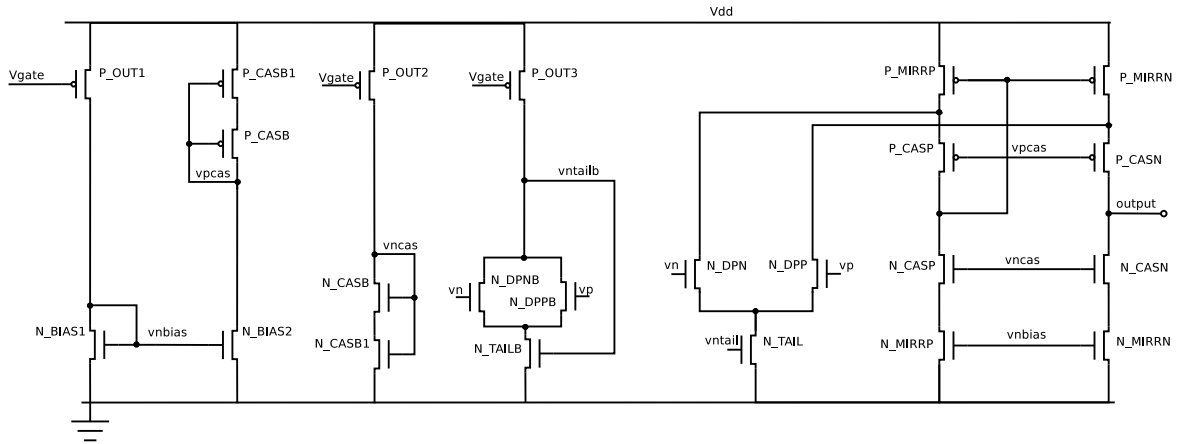
Figure 51: Complete OTA schematic with the bias part

The complete list of device sizes of the designed OTA is in table [4]. The presented sizes L_0 and W_0 are directly calculated according to inversion factor

Device Name	IF	I_D[uA]	W/L	L_0[um]	W_0[um]	Num.	W_1[um]	L_1[um]
N-DPN,P	0.1	0.4	25/4	2	12.5	1	5	10
N-TAIL	0.8	0.8	25/16	2	3.13	2	5	12
P-CASN,P	0.5	0.8	320/27	1	11.85	1	3	1
P-MIRRN,P	12	1.0	50/81	8	4.94	1	1	18
N-MIRRN,P	7.2	0.8	17/98	12	2.08	32	2	12.5
N-CASN,P	0.5	0.8	5/2	1	2.5	1	1	1
N-BIAS1	7.2	0.8	17/98	12	2.08	32	1	12.5
N-BIAS2	7.2	0.4	2/23	12	1.04	32	1	12.5
N-CASB	0.5	0.4	5/4	1	1.25	2	1	20
N-CASB1	20	0.4	1/32	14	0.44	32	1	14
N-TAILB	0.8	0.4	25/32	2	1.56	2	1	12
N-DPNB,P	0.1	0.2	25/8	2	6.25	1	1	10
P-CASB	0.5	0.4	25/8	2	6.25	1	1	5
P-CASB1	20	0.4	4/27	10	1.48	32	1	17.5
P-OUT1,2,3	6	0.2	20/81	20	4.94	3	1	20

Table 4: OTA Device sizing, scaling factors are 32 (for amplifier) and 40 (for bias part)

IF and drain current I_D. The new MOS device sizes L_1, W_1 are the final design choices after the offset optimization and current scaling procedure. Scaling factors are chosen 32 for amplifier and 40 for the bias part. Parameter $Num.$ is a number of serial connected MOS transistor.

	Before	After
$I_{Consum.}[nA]$	3938	163
DC GAIN[dB]	66	70

Table 5: OTA Scaling results

Table[5] shows the OTA performance before and after the current scaling process. It can be seen, that current consumption is reduced to 4.13% of the initial value. Moreover the DC gain and the Phase margin are even better than non-scale version. The DC gain is given by $Av = g_{m1} \cdot R_{out}$. The g_{m1} is a transconductance of the N diff.pair. Due to the scaling procedure the g_{m1} drops (N diff.pair is in weak inversion where g_m is proportional to current). On the other hand, transistor channel lengths are longer. The R_{out} rise more rapidly then the g_{m1} drops. Therefore the resulting DC gain is bigger after scaling procedure.

The GBW can be expressed by $GBW = g_{m1}/(2\pi C_L)$. Due to the g_{m1} drops the GBW is moved closer to lower frequencies. Therefore the Phase margin is improved (the distance between second pole f_2 and unity gain point is increased).

6.3.4   Stability Compensation

The Opamp is used with a negative feedback in the previous mentioned BGR topology. Thus a stability problem may occur (if the phase shift turns from 180 degrees to 0 degrees then a positive feedback occur and oscillation might appear).

Phase margin (PM) is an important parameter that describes stability. The sufficient value of PM depends on situation, sometimes even PM=45 degrees is acceptable. In most cases PM should be at least 60 degrees to obtain stable system.

If PM is not sufficient then some stability compensation is required. In this work compensation is made by MOS capacitor, see Fig.[52]. This PMOS transistor is same as the P-BGRN,P and its area is 120 μm^2.

66

Figure 52: BGR schematic with a stability compensation (MOS capacitor Cc)

6.4 Design of the Start-up Circuit

6.4.1 Inroduction, Start-up Examples

BGR has an attribute of multiple possible stable operating points. Only one of those is the proper operating point and the rest are zero points or some unwanted points. Therefore a special start-up circuit is required. The specification is that the BGR shall be started by enable signal. From this follows that if the supply voltage is applied but enable signal is low, the BGR needs to be switched off. The BGR can be turned on only if it is necessary or it can be periodically turned on and off. Both options should reduce power consumption.

Start-up design is one of the key tasks. In many cases this circuit is not strong enough for BGR starting. If start-up obtains feedback it can be unstable. Moreover this circuit sometimes has a problem with its own switching off.

So the start-up requirements are:

• stable performance

• reliable start

67

- small current consumption

- negligible BGR affection

- start by enable signal

Start-up circuits can be divided according to their function to the static and dynamic. The static start-up works with some DC voltage or current, while the dynamic needs some rising/falling edge or some pulse.

The main question is what to sense and what to drive. It means that some specific voltage or current indicating the state of BGR (on/off) must be found. Also some way to reliable BGR start must be found.

Figure 53: Start-up sensing options

Following options were found. Fig.[53] shows proposed sensing options.

- Sensing
  - V_{gate}
  - V_{rb}
  - V_{ref}
  - Branch replica

- Interact

  - Current injection into nodes V_n/V_p
  - Voltage V_{gate} connect to ground
  - Voltage V_{gate} push to start-point

Fig.[54] shows the start-up example based on V_{gate} sensing.

Figure 54: Start-up example based on V_{gate} sensing

Assume that the BGR is off. Thus I_{bg} is lower than supposed current or more usually case it is zero. The voltage V_s is due to N-DIODE transistor pushed down (nearly to zero). Therefore P-START transistor is turned on and some current is injected into V_n node.

If the BGR is already in correct operating point, then the $I_{bg} \simeq 200nA$ (an accurate value depends on temperature). This specified current creates voltage drop across the N-DIODE transistor (the V_s voltage is rising). This voltage drop needs to be large enough to close safely P-START transistor and the start-up does not affect BGR core any more.

Disadvantage of this circuit is that it does not sense current through diodes. In some cases the current I_{bg} is flowing only through resistors Rs. This case is unwanted because start-up circuit is turned off and BGR core does not work correctly.

69

The following circuit solves this problem, Fig.[55]. It is the start-up based on V_{rb} sensing.
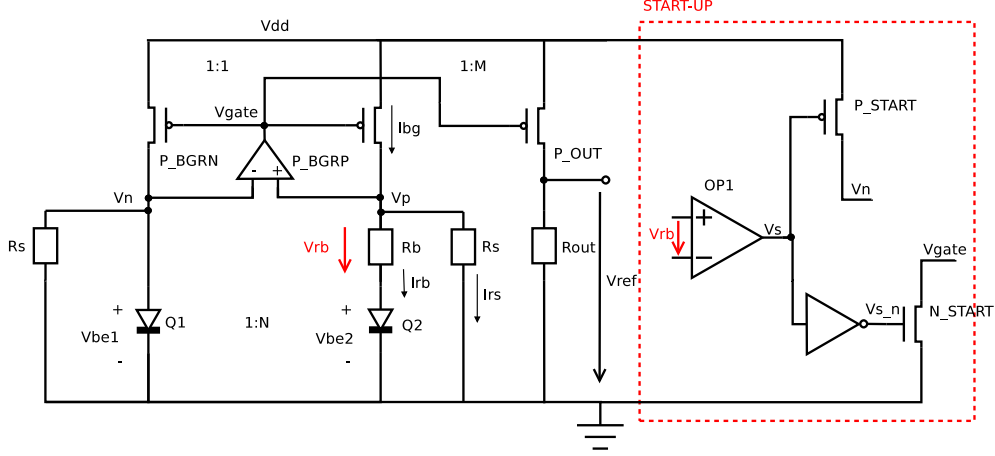

Figure 55: Start-up example based on V_{rb} sensing

Assume that BGR is turned off, or as mentioned before, the whole I_{bg} current flow into resistors Rs. In both cases the voltage drop across Rb is zero. The start-up Opamp $OP1$ is connected without feedback, and it serves as comparator. The $OP1$ is designed with intend offset $V_{OS,OP1} \simeq V_{rb}$. Therefore voltage V_s is zero and V_{s-n} is due to logic NOT almost V_{dd}. The P-START transistor injects current into V_n node. N-START transistor push the V_{gate} voltage near to zero. Thus currents in all BGR core branches rise.

In the required operating point the current about 100nA flow through each diode Q1 and Q2. Thus the voltage drop V_{rb} across the resistor Rb occur. This voltage drop can be estimated at $T = 300K, I_{rb} = 100nA$ by the following calculation.

$$V_{rb} = U_T \cdot ln(N) \simeq 26 \cdot 10^{-3} ln(24) \simeq 82.6mV \qquad (94)$$

If the V_{rb} reaches this value, then Opamp $OP1$ should switch the output voltage V_s near to V_{dd}. Thus P-START is switched off and due to the logic gate the V_{s-n} is zero and N-START is also turned off.

70

It is clear that the total consumption rises by adding extra Opamp $OP1$. According to the low power task, this topology is not selected.

6.4.2   Final start-up design

The following Fig.[56] shows the final start-up topology. The left branch is a part of BGR core. As it is shown, the sensing voltage is directly the reference voltage V_{ref}. The output is connected to the V_{gate} voltage in BGR core. The V_{gate} voltage should be pushed into the right operating point by this structure.

Figure 56: Final Start-up Circuit

The V_{ref} voltage target is 0.7V. If it is less something is wrong. Now assume that BGR operates in wrong operating point so the V_{ref} voltage is for instance 0.2V. This voltage is not enough for opening N-SWITCH transistor ($V_{ref} < V_{th,N-SWITCH}$). Thus the voltage V_S goes up to V_{dd}. This is because N-SWITCH acts as the open switch and P-RES is resistor. Following logic gates (Smith trigger, logic NOT) set the voltages $V_{int-n} = 0$, $V_{int} = V_{dd}$ according to the V_S voltage. Therefore P-SW, N-START are switched on. The N-START transistor is a current source which drains some current through the diode connected transistor P-DIODE. Due to the diode and particular current through it, the V_{gate} voltage is pulled down. This causes that all currents in the BGR core rise. Therefore the V_{ref} rises and the Opamp starts

to operate.

On the other hand, if $V_{ref} > V_{th,N-SWITCH}$ then the N-SWITCH is switched on and V_S voltage is nearly zero. The indicating voltages are set to $V_{int-n} = Vdd$, $V_{int} = 0$. Therefore N-START is switched off and it does not drain current any more. P-SW transistor is also turned off thus no current can flow through the diode. The BGR core is than no more affected by the start-up circuit.

Transistors P-PWD1,P-PWD2 and N-PWD1 are used for power down switching. When enable signal is low, the whole BGR is turned off. This is achieved by these power down switches, which are placed in every block of BGR. If enable signal is low then non-enable signal e_{n-n} is Vdd. Thus P-PWD1,2 disconnect the logic gates and N-PWD1 shorts the gate of N-START to the ground. So no current is able to flow through N-START transistor. To avoid possible oscillation around decision level the Schmitt trigger is added.
Compared to the previous example, this start-up consumes less statical current. Unfortunately the logic gates consume some current while they are switching. This logic gates are build from long channel MOS transistors. Thus these logic gates are slower but the dynamic current consumption is reduced.

Final design sizes are now presented.

Transistor P-OUT is connected as a resistor. Its W/L ratio should be small for achieving small current I_{res}. Thus the W/L ratio is set to 1/640 and length is set to L=640um. Unfortunately the PDK does not allow the larger length than L_{max} =20um. Therefore the new device(PMOS-32S) with 32 serial connected PMOS transistors is designed. The current I_{res} is not sufficiently small, so one more PMOS-32S is added to serial connection. The maximum value of I_{res} is equal to 50nA. This maximum is reached at T=-50°C and V_{dd}=2V. On the other hand the minimum of I_{res} is 5nA (T=-50°C, V_{dd}=0.9V).

The N-INIT transistor is set to W/L=1/40 and length is L=20um. This transistor must be long enough to provide sufficient voltage drop across it, while the I_{ref} is only 5nA.

The N-START transistor acts as the current limiter. The provided starting current should be at least 3 times larger than a current in the OTA summing branch for reliable start. Thus this transistor is the same as the current source in OTA N-MIRRN (W/L=1/400, L=400um) but it is multiply by 3.

Transistor P-DIODE is diode connected PMOS. Thus the voltage V_{gate} is influenced by the voltage drop across the P-DIODE. If the starting current flows through the P-DIODE then the voltage V_{gate} should be set near to the 0.3-0.4V. The sizes was experimentally obtained W/l=4.8/20, where length is equal to L=20um.

Device Name	Num.	W_1[um]	L_1[um]
P-OUT	1	4.8	20
P-RES	64	1	20
P-PWD1,2	1	0.4	0.4
P-SW	1	0.4	0.4
P-DIODE	1	4.8	20
N-START	32	4	12.5
N-PWD1	1	0.4	0.4
N-SWITCH	1	0.4	20

Table 6: Start-up device sizing

7 Simulation and Results

Voltage reference circuit designed in the previous chapter was tested. The simulation results are now presented. The whole circuit schematic can be seen in appendix A.

Fig.[57] shows the reference voltage V_{ref} over the whole temperature range (-50-100°C). The maximal deviation is 2.99mV. This simulation was performed under these conditions: supply voltage 1V and NN corner.

The previous simulation was repeated for the same supply voltage but for all other corners, see Fig.[58]. It can be seen that some voltage deviation at temperature range 50-100°C existed. The maximal peak to peak value is now roughly 5mV. The causes of these voltage deviation are rising leakage current[5] and lower N-TAIL current. The N-TAIL current issue was already presented. This problem was improved by proper design choice of N-TAIL transistor, but it was not eliminated.

Fig.[64] shows a 100 runs of Monte Carlo (process and mismatch). This simulation is a real test of the circuit robustness. From this results, the trimming requirement can be obtained.

Figures [61, 62, 63] show the current consumptions as a functions of temperature. On the Fig.[61] the OTA consumption is the blue curve. Its current consumption is less than 200nA. It can be seen, that the total circuit consumption (the red curve) is not higher than a $1\mu A$. Fig.[62,63] shows the current consumption of the whole circuit and the OTA as function of temperature and supply voltage.

---

[5]leakage current is a current flowing through the transistor, which is turned off
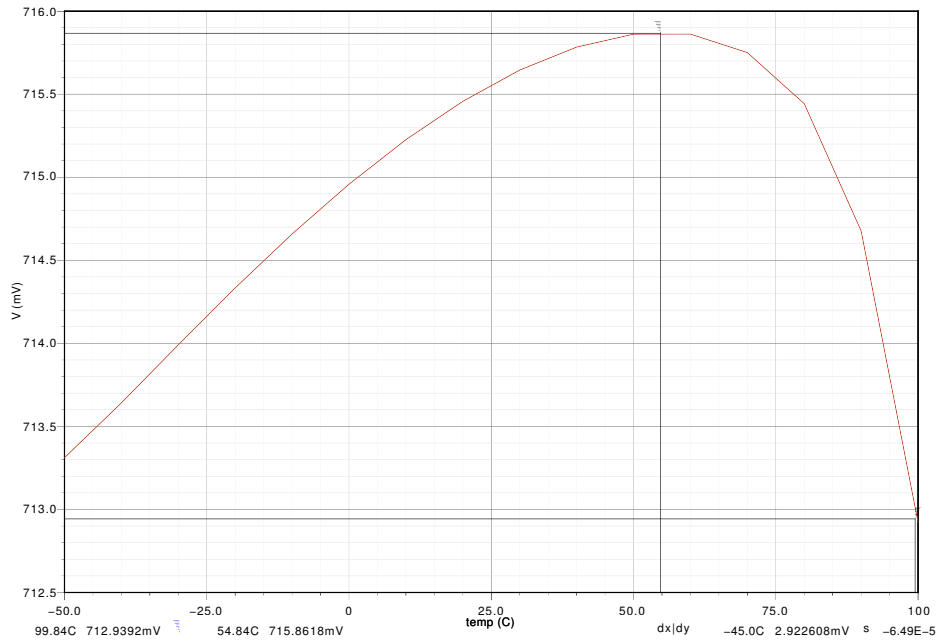
Figure 57: Reference voltage V_{ref} vs. temperature (-50-100℃), NN corner
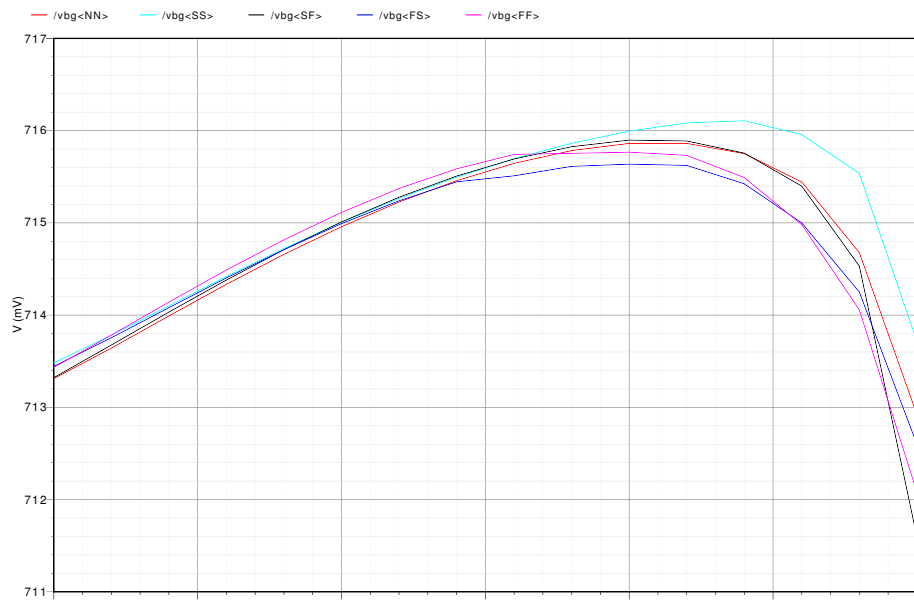
Figure 58: Reference voltage V_{ref} vs. temperature (-50-100℃), for various technological corners (NN,SS,SF,FS,FF)
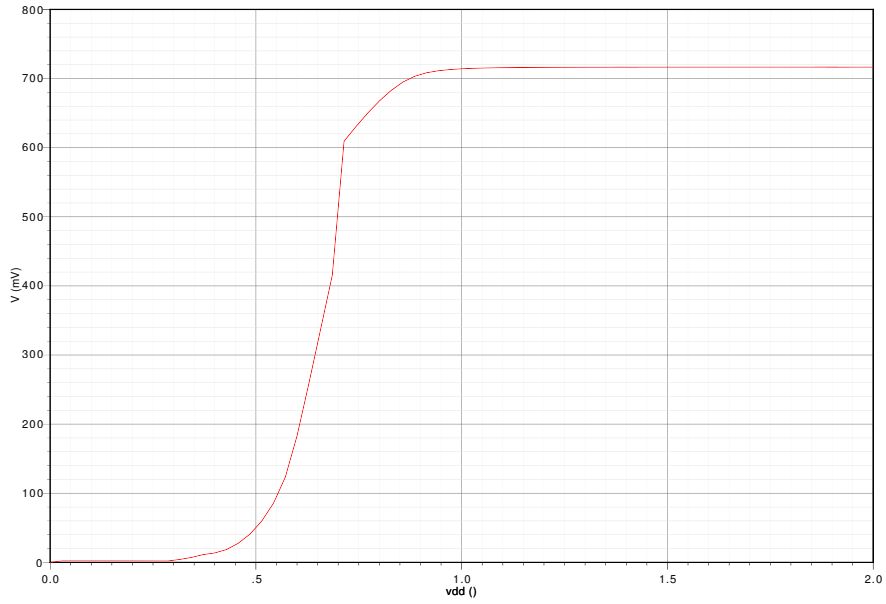
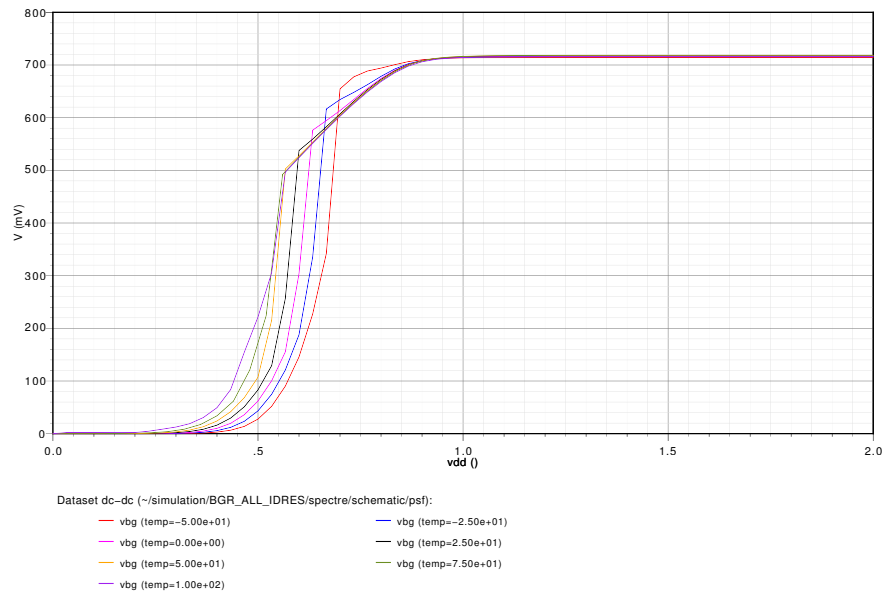Figure 59: Reference voltage V_{ref} vs. supply voltage (0-2V), NN corner, -50℃

Dataset dc–dc (~/simulation/BGR_ALL_IDRES/spectre/schematic/psf):

— vbg (temp=–5.00e+01)   — vbg (temp=–2.50e+01)
— vbg (temp=0.00e+00)   — vbg (temp=2.50e+01)
— vbg (temp=5.00e+01)   — vbg (temp=7.50e+01)
— vbg (temp=1.00e+02)

Figure 60: Reference voltage V_{ref} vs. supply voltage (0-2V), for various temperatures (-50, -25, 0, 25, 50, 75, 100 ℃)
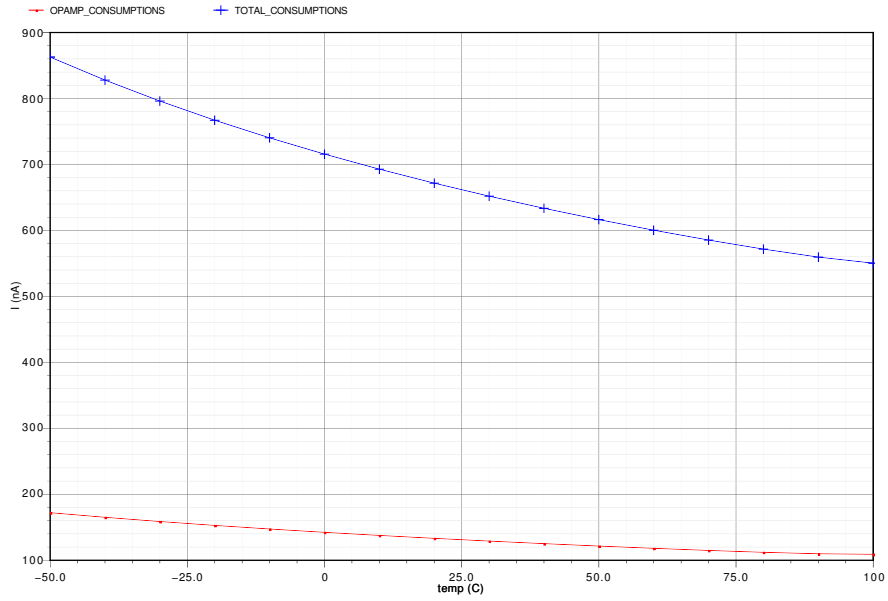
Figure 61: BGR current consumption and OTA consumption vs. temperature (-50-100℃)
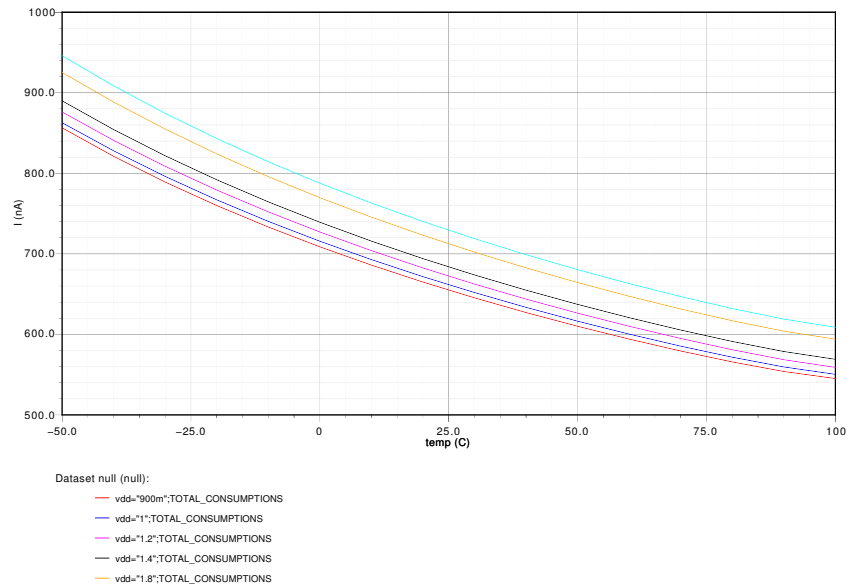
Figure 62: BGR current consumption vs. temperature (-50-100℃), for various power supply (0.9, 1.0, 1.2, 1.4, 1.8, 2.0V)

Figure 63: OTA current consumption vs. temperature (-50-100°C), for various power supply (0.9, 1.0, 1.2, 1.4, 1.8, 2.0V)

The following figures [64, 65, 66, 67, 68, 69] show the trimming possibilities. Fig.[64] shows the reference voltage V_{ref} vs. temperature (-50-100°C), Monte Carlo 100 runs. This simulation result specify required trimming range. Two worst cases are marked (iteration 62 and 15). Figures [65, 66] show successfully trimmed worst cases (62,15). These simulation confirm proper design of the absolute (R_{out}) trimming.

Fig.[67] shows normalized V_{ref} results from Fig.[64]. These results show the curvatures of the V_{ref} voltages. Ideally all runs would be equal to one for the whole temperature range. For the better curvature the TC trimming is used (resistors Rs). Two worst cases are again marked (iteration 41 and 1). Figures [68, 69] show trimmed worst cases (41,1).
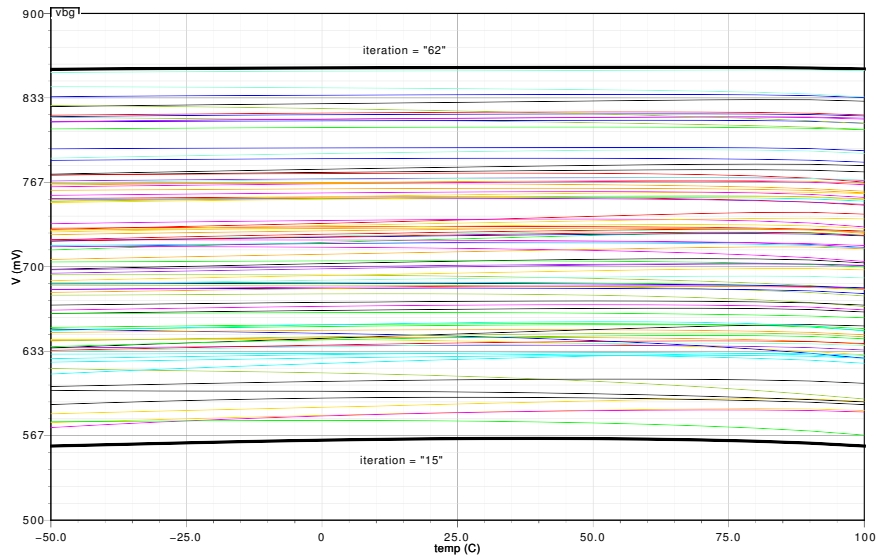
Figure 64: Reference voltage V_{ref} vs. temperature (-50-100°C), Monte Carlo 100 runs process and mismatch, worst cases = 62 and 15 iteration
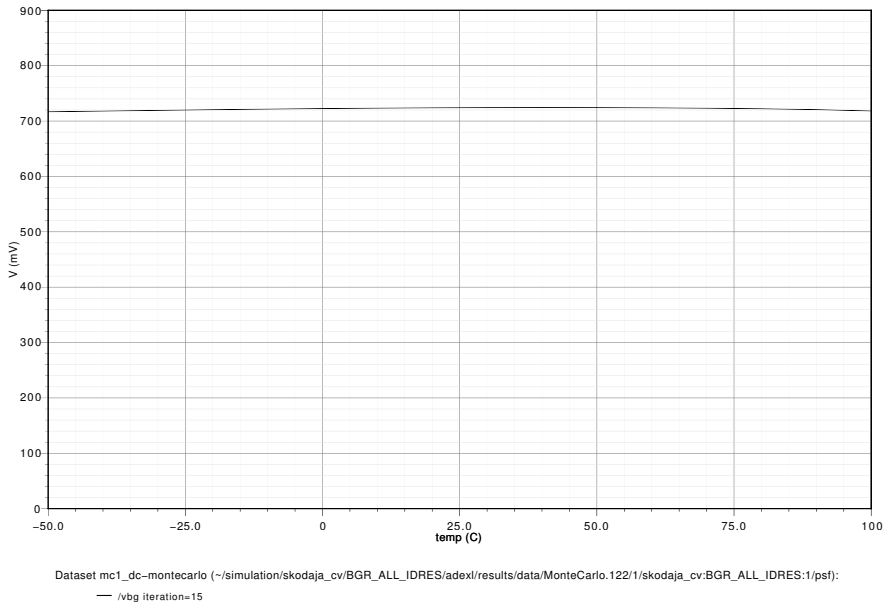

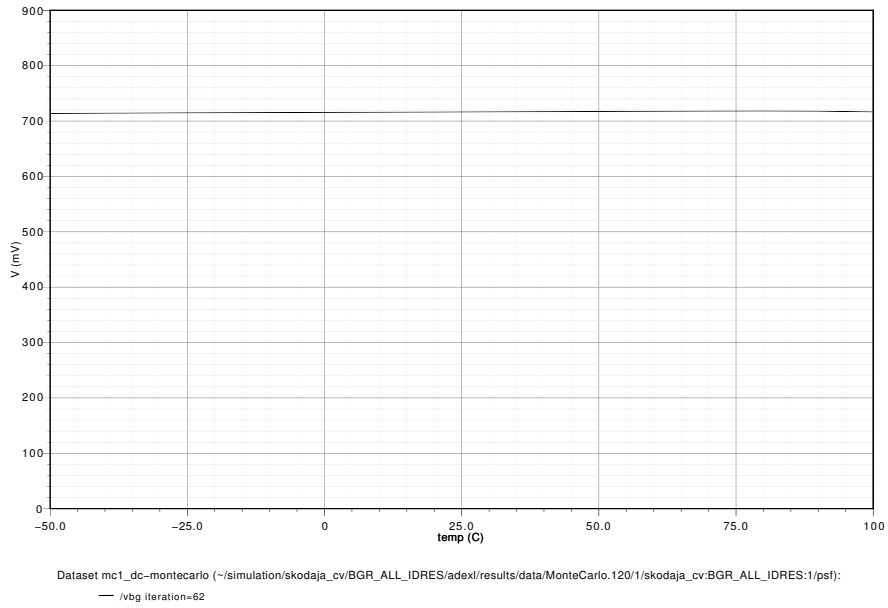Figure 65: Reference voltage V_{ref} vs. temperature (-50-100°C) - trimmed worst case 15 iteration

Figure 66: Reference voltage V_{ref} vs. temperature (-50-100°C) - trimmed worst case 62 iteration
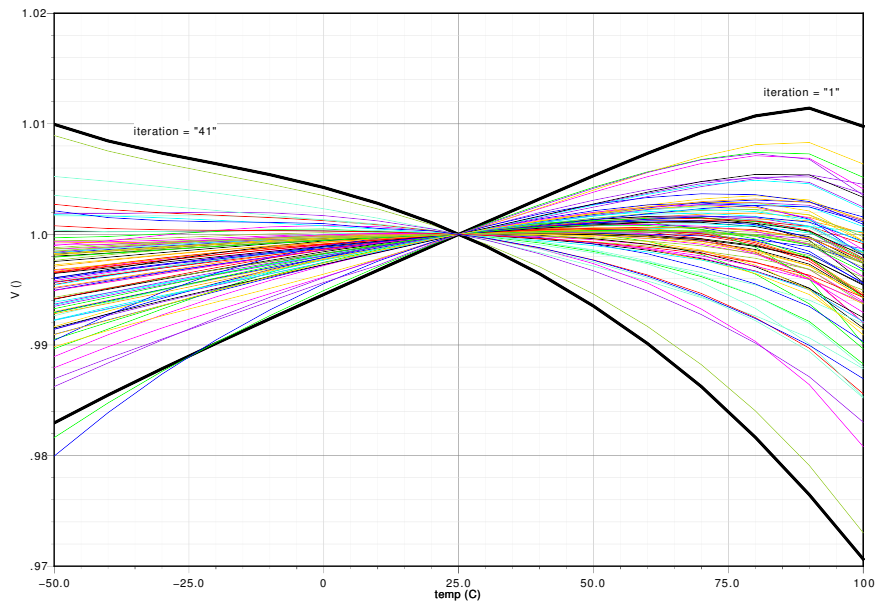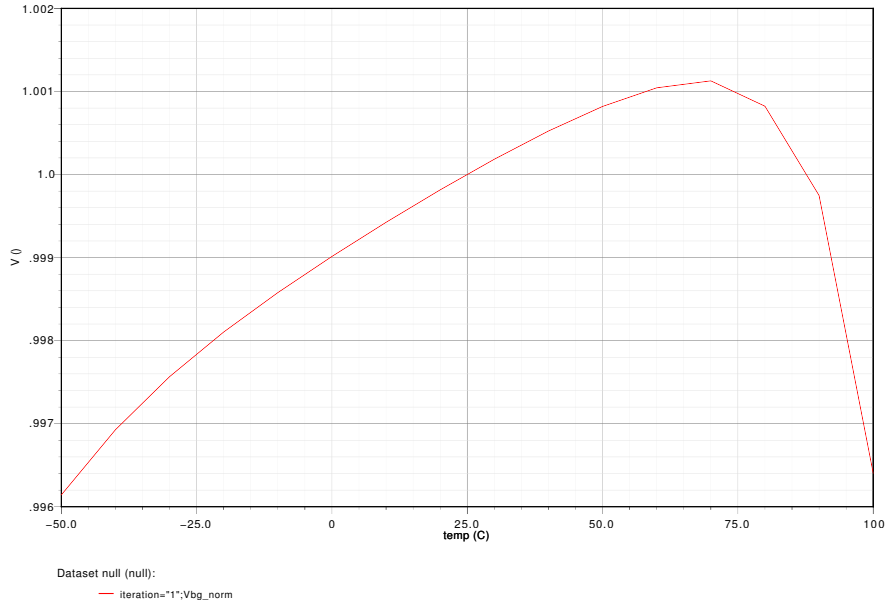

Figure 67: Normalized reference voltage V_{ref} vs. temperature (-50-100°C), Monte Carlo 100 runs process and mismatch, worst cases = 41 and 1 iteration

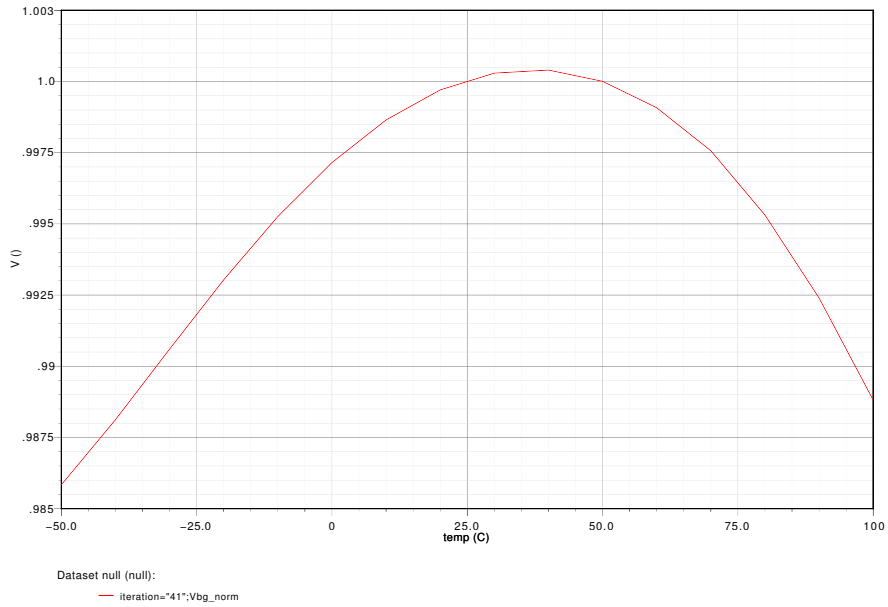Figure 68: Normalized reference voltage V_{ref} vs. temperature (-50-100°C) - trimmed worst case 1 iteration

Figure 69: Normalized reference voltage V_{ref} vs. temperature (-50-100°C) - trimmed worst case 41 iteration
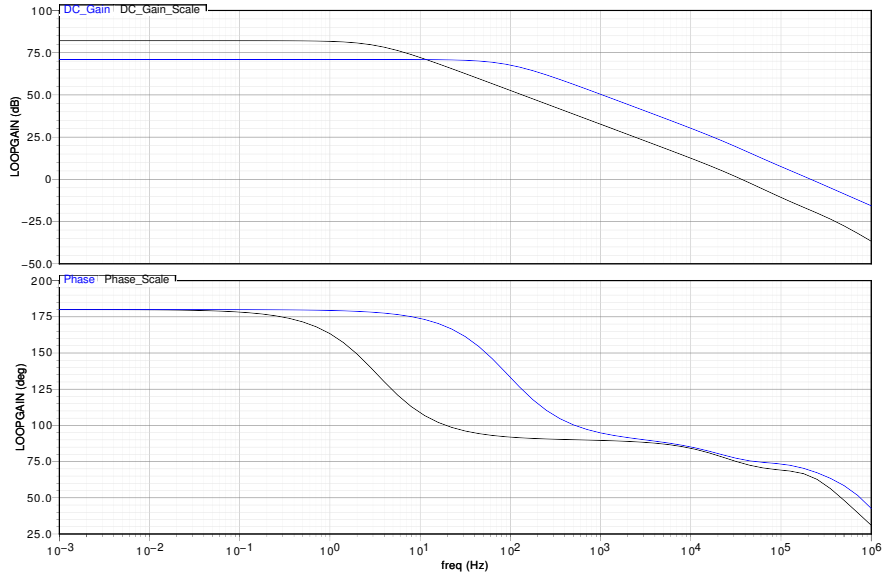
Figure 70: BGR AC. simulation, DC gain and phase

Fig.[70] shows the AC simulation of the designed BGR (scaled and non-scaled version). The scaled version provides higher DC gain but lower BW. From the phase characteristic the pole-zero issue can be seen. Fortunately this is far beyond GBW, thus it does not corrupt the system stability. Table[7] summarizes the AC simulation results (NN corner, temperature $-50°C$, supply voltage 1V).

The AC parameters of scaled BGR are simulated for various supply voltages (0.9, 1.0, 1.2, 1.4, 1.8, 2.0), temperatures (-50, -25, 0, 25 ,50 ,75, 100) and technological corners (NN, SS, SF, FS, FF), see table [8]. These tests confirm the proper stability compensation design.

	Unscaled	Scaled
$I_{TOTAL}[nA]$	6005	862.5
DC GAIN[dB]	70	81
PM[°]	68	73
BW[Hz]	112	3
GBW[kHz]	214	32.2
Cc[pF]	4.0	0.4

Table 7: AC results of the scaled and unscaled BGR, (NN corner, temperature -50℃, supply voltage 1V)

	Min	Nom	Max
DC GAIN[dB]	64	81	86
PM[°]	69	73	79
GBW[kHz]	17.1	32.2	36.9

Table 8: Scaled BGR AC parameters, tested for various supply voltages (0.9, 1.0, 1.2, 1.4, 1.8, 2.0), temperatures (-50, -25, 0, 25 ,50 ,75, 100), technological corners (NN, SS, SF, FS, FF)

Following three figures [71,72,73] are the start-up simulation results. The start-up testing is achieved by transient simulation. It is clear that despite some ringing the circuit started in all simulations.

The worst case (0.9V supply voltage) was chosen and then temperature sweep was simulated (-50, -25, 0, 25, 50, 75, 100 °C), Fig.[72]. The last simulation is again the worst case (0.9V supply voltage, -50 °C) for various technological corners (NN, SS, SF, FS, FF), see Fig.[73].

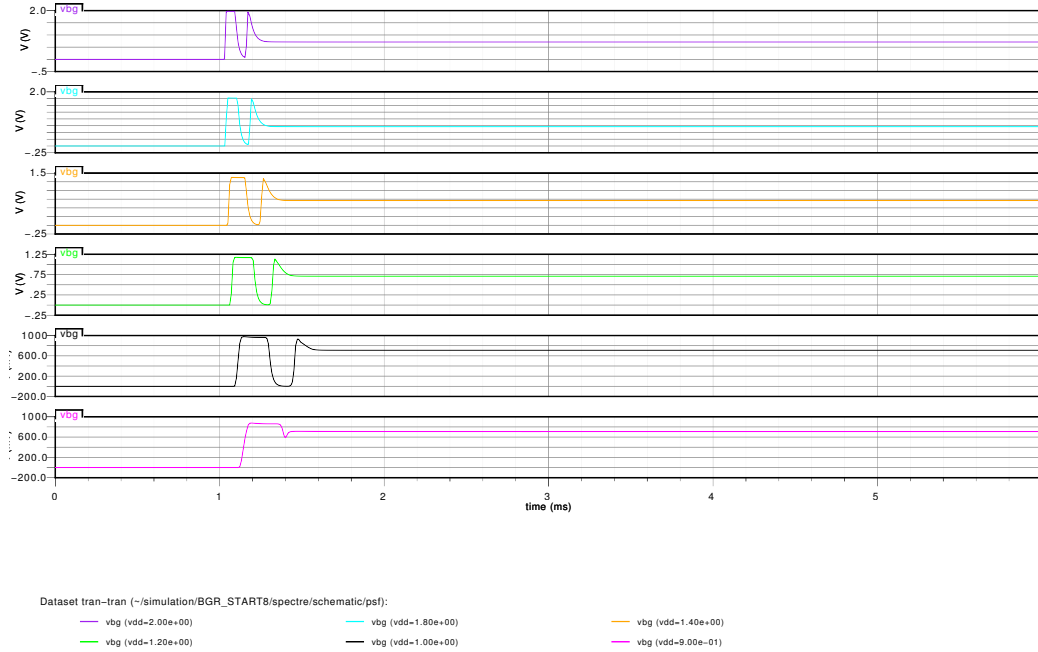The circuit always successfully started in time interval 1ms.

Figure 71: Start-up test results, different supply voltage (0.9V, 1V, 1.2V , 1.4V, 1.8V, 2V), temperature -50°C
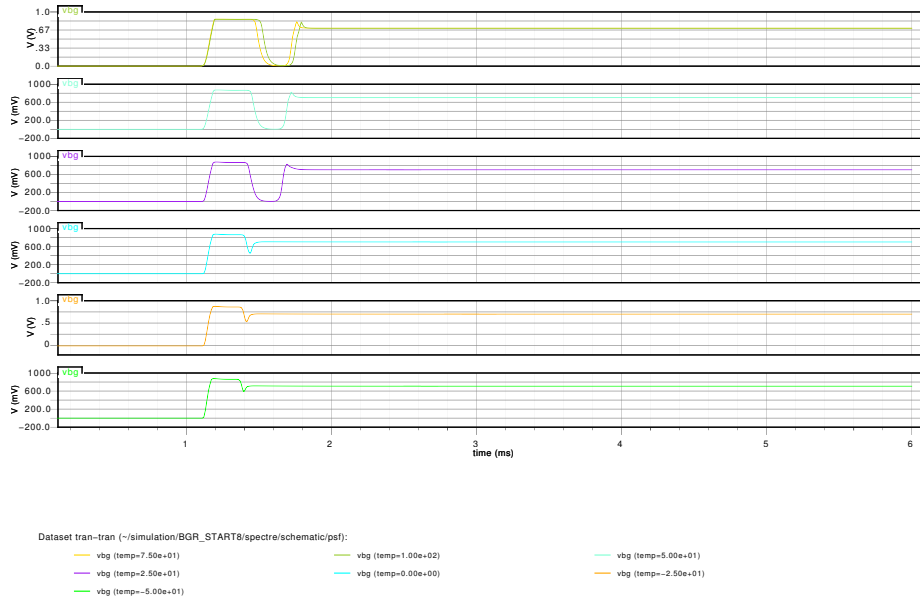
Figure 72: Start-up test, worst case (0.9V supply voltage) for various temperatures (-50, -25, 0, 25, 50, 75, 100 °C)
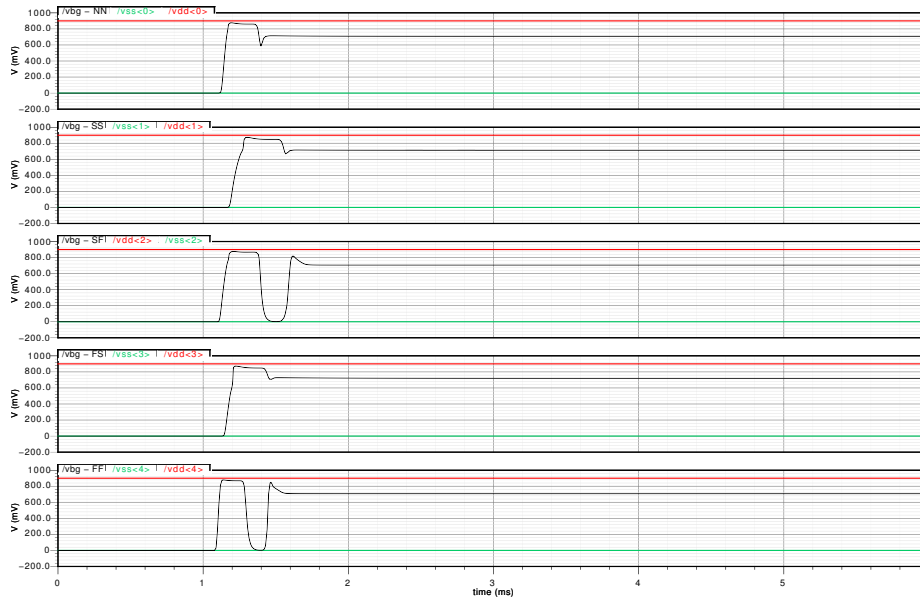

Figure 73: Start-up test, worst case (0.9V supply voltage, -50 °C) for various technological corners (NN, SS, SF, FS, FF)

85

8 Conclusion

In this thesis a CMOS bandgap reference has been presented. The robust and low power cell is designed with Cadence Virtuoso PDK 180nm process. BGR core, Opamp, Start-up and trimming circuit were designed and simulated.

Circuit is able to work at 0.9 V and its consumption is below 1uA. The nominal output voltage is 715mV at T=25℃. The start-up circuit provides reliable start in 1ms. The voltage deviation due to the temperature change is 2.99mV at nominal conditions ($V_{dd} = 1V$, NN corner).

This BGR has following advantages: It is low voltage and low power circuit, thus it is suitable for modern low power applications. The core can be also used as the current reference. The wide temperature range meets the requirements of most applications.

The reference voltage can be tuned due to the trimming circuit. Moreover circuit works with the enable signal. Thus additional power optimization can be achieved.

Table [9] shows the performance of other BGR.

	Leung et al.[1]	Annema [2]	Malcovat et al.[3]	Banba et al.[4]	This Work
Technology	0.6μm CMOS	0.35μm CMOS	0.8μm BiCMOS	0.4μm CMOS	0.18μm CMOS
Min. Power Supply [V]	0.98	0.85	0.95	0.84	0.90
Current Consumption [μA]	180	<1.2	<92	<2.2	<1.0
Output Voltage [mV]	603	650	536	515	715
Temperature Coefficient [ppm/℃]	15	57	19	59	27
Temperature Range [℃]	0-100	-20-100	0-80	27-125	-50-150

Table 9: Performance comparison of low-power BGR

References

[All02] Holberg D. R Allen, P. E. *CMOS Analog Circuit Design.* Oxford University Press, January 2002.

[Bak10] R. J. Baker. *CMOS Circuit Design, Layout, and Simulation.* IEEE Press, 2010.

[BBK02] A. Boni, Sergio Bampi, and Hamilton Klimach. Op-amps and startup circuits for cmos bandgap references with near 1-v supply. *IEEE Journal of Solid-State Circuits*, vol. 37(issue 10):1339–1343, 2002.

[BBK07] Juan Pablo Martinez Brito, Sergio Bampi, and Hamilton Klimach. A 4-bits trimmed cmos bandgap reference with an improved matching modeling design. *2007 IEEE International Symposium on Circuits and Systems*, pages 1911–1914, 2007.

[BFJ+07] D. M. Binkley, D. Flandre, P.G.A. Jespers, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui. Tradeoffs and optimization in analog cmos design. *2007 14th International Conference on Mixed Design of Integrated Circuits and Systems*, vol. 31(issue 9):47–60, 2007.

[BSU+04] H. Banba, H. Shiga, A. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui. A cmos bandgap reference circuit with sub-1-v operation. *IEEE Journal of Solid-State Circuits*, vol. 34(issue 5):670–674, 2004.

[Car12] Johns D. A. Carusone, T. CH. *Analog Integrated Circuit Design.* John Wiley and Sons, Inc., 2012.

[CSG+05] A. Cabrini, G. De Sandre, L. Gobbi, P. Malcovati, M. Pasotti, M. Poles, F. Rigoni, and G. Torelli. A 1 v, 26 µw extended temperature range band-gap reference in 130-nm cmos technology. *Proceedings of the 31st European Solid-State Circuits Conference, 2005. ESSCIRC 2005*, pages 503–506, 2005.

[DB11] J. Stehlík D. Bečvář. *Návrh analogových integrovaných obvodů.* VUTBR, 2011.

[DHYW10] Guo-Ding Dai, Peng Huang, Ling Yang, and Bo Wang. A constant gm cmos op-amp with rail-to-rail input/output stage. *2010 10th IEEE International Conference on Solid-State and Integrated Circuit Technology*, pages 123–125, 2010.

[DPDR00] L. Dobrescu, M. Petrov, D. Dobrescu, and C. Ravariu. Threshold voltage extraction methods for mos transistors. *2000 International Semiconductor Conference. 23rd Edition. CAS 2000 Proceedings (Cat. No.00TH8486)*, pages 371–374, 2000.

[KWM03] Q.A. Khan, S.K. Wadhwa, and K. Misri. Low power startup circuits for voltage and current reference with zero steady state current. *Proceedings of the 2003 International Symposium on Low Power Electronics and Design, 2003. ISLPED '03*, pages 184–188, 2003.

[LHV11] Boram Lee, Ted Higman, and M. Vertregt. Extremely simple constant-gm technique for low voltage rail-to-rail amplifier input stage. *2011 18th IEEE International Conference on Electronics, Circuits, and Systems*, pages 314–317, 2011.

[LTL$^+$09] Timothée Levi, Jean Tomas, Noëlle Lewis, Pascal Fouillat, T. Tanzawa, S. Atsumi, and K. Sakui. A cmos resizing methodology for analog circuits. *IEEE Design*, vol. 26(issue 1):78–87, 2009.

[MLK04] P.K.T. Mok, Sergio Ka Nang Leung, and Hamilton Klimach. Design considerations of recent advanced low-voltage low-temperature-coefficient cmos bandgap voltage reference. *Proceedings of the IEEE 2004 Custom Integrated Circuits Conference (IEEE Cat. No.04CH37571)*, vol. 37(issue 10):635–642, 2004.

[PTV98] M.J.M. Pelgrom, H.P. Tuinhout, and M. Vertregt. Transistor matching in analog cmos applications. *International Electron Devices Meeting 1998. Technical Digest (Cat. No.98CH36217)*, pages 915–918, 1998.

[Raz03] B. Razavi. *Design of Analog CMOS Integrated Circuits.* McGraw-Hill Companies, 2003.

[Sed13] Smith C. K. Sedra, A. S. *Microelectronic Circuits Theory and Applications.* Oxford University Press, 2013.

[SFJ+04] F. Silveira, D. Flandre, P.G.A. Jespers, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui. A g/sub m//i/sub d/ based methodology for the design of cmos analog circuits and its application to the synthesis of a silicon-on-insulator micropower ota. *IEEE Journal of Solid-State Circuits*, vol. 31(issue 9):1314–1319, 2004.

[Ste08] Kayal M. Stefanovic, D. *Structured Analog CMOS Design.* Springer Science+Business Media B.V, 2008.

[SW94] K. Sansen W., Laker. *Design of Analog CMOS Integrated Circuits and Systems.* McGraw-Hill Companies, 1994.

[WRZYWD09] Zhao Wu, Fu Rui, Zhang Zhi-Yong, and Cheng Wei-Dong. Design of a rail-to-rail constant-gm cmos operational amplifier. *2009 WRI World Congress on Computer Science and Information Engineering*, pages 198–201, 2009.
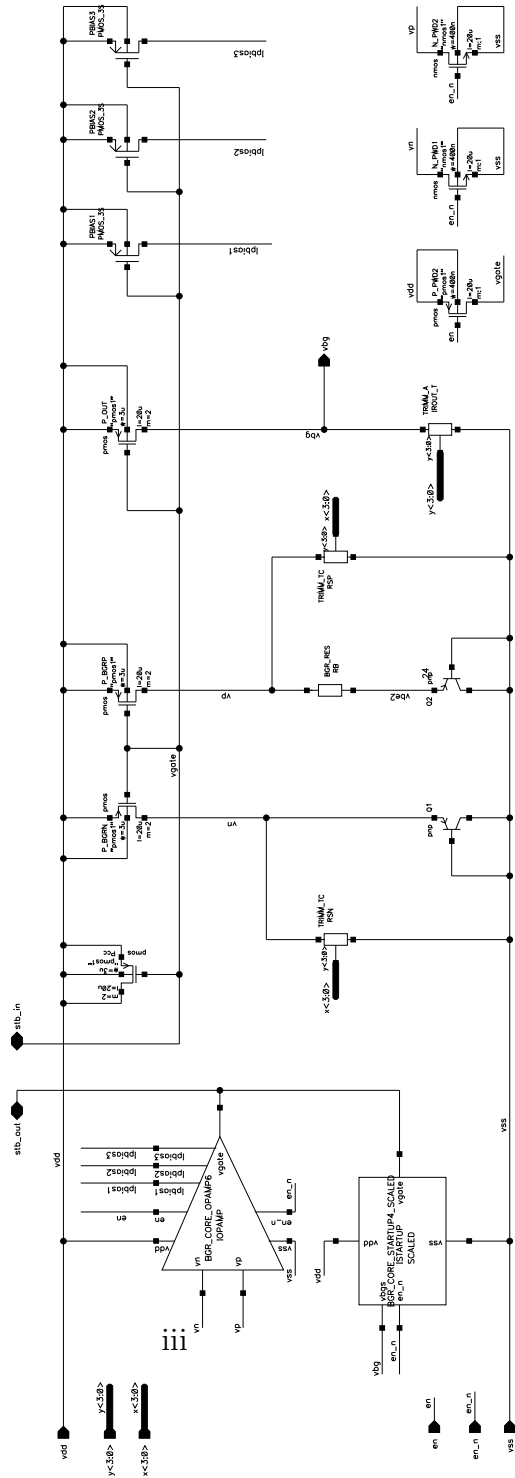
Appendices

Appendix A.1: BGR testbench
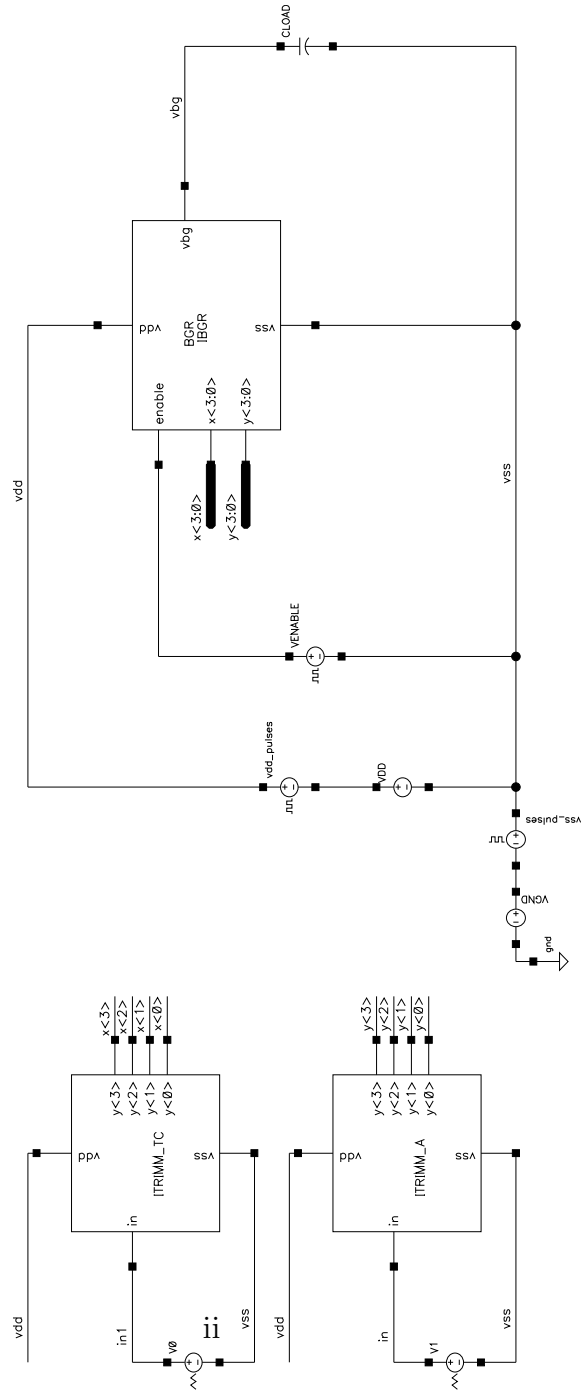
Appendix A.2: BGR core schematic

Appendix A.3: Opamp schematic

Appendix A.4: Start-up circuit schematic

vgate

Appendix A.5: Trimming Rout schematic
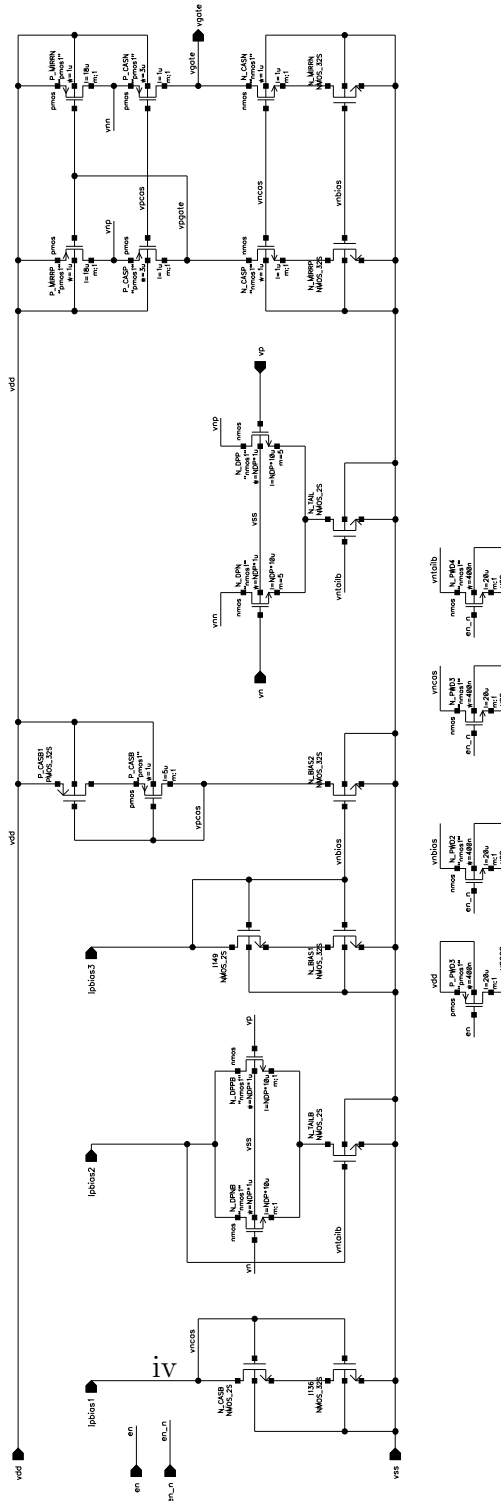
Appendix A.6: Trimming Rs schematic

XRT — 40x160k
8RT — 4x160k
4RT — 2x160k
2RT — 1x160k
1RT — 0.5x160k (2x160k parallel)
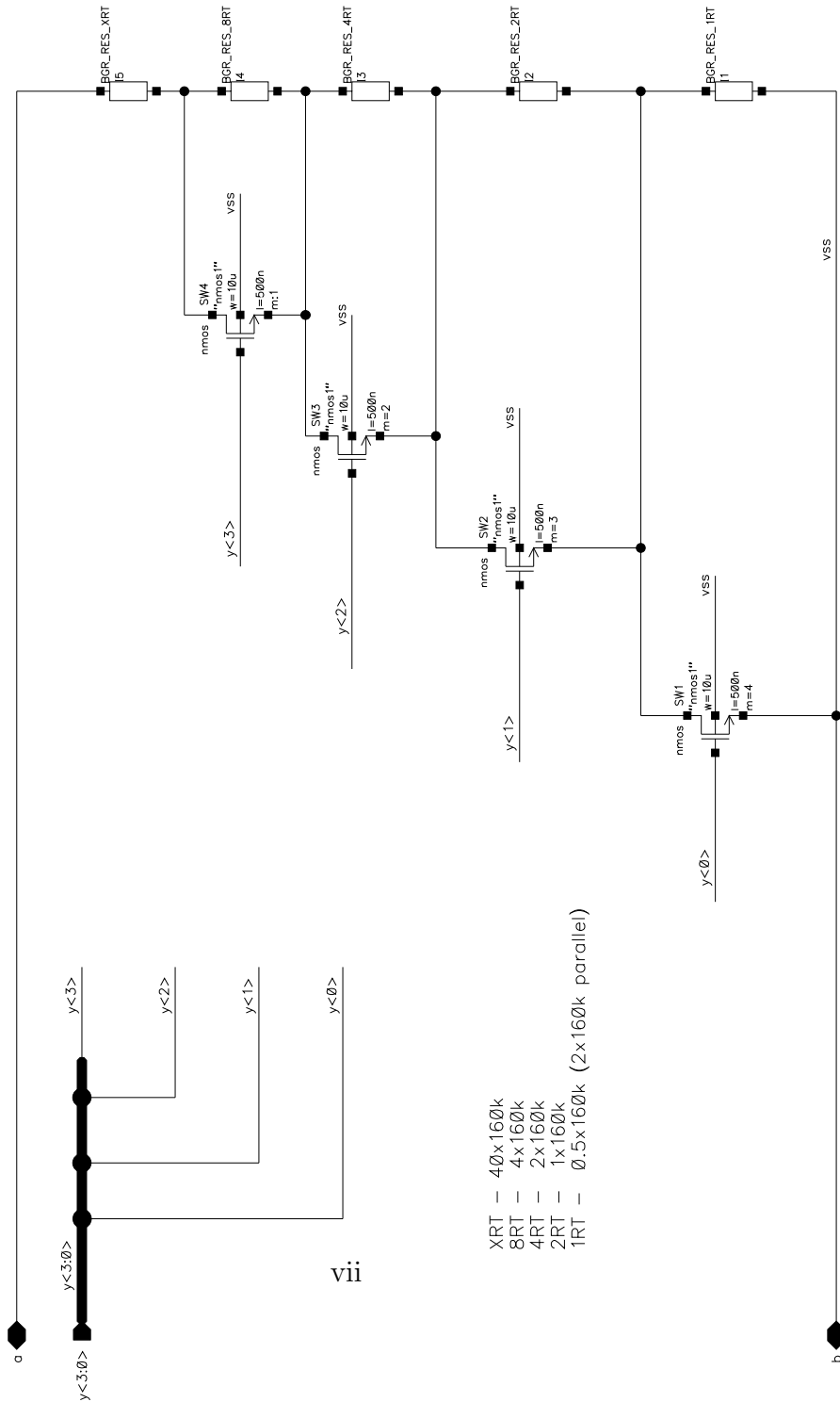
Appendix B.1: Verilog-A code for logic NOT gate

```
// VerilogA for skodaja_cv, BGR_V_NOT, veriloga

`include "constants.vams"
`include "disciplines.vams"

module BGR_V_NOT(a,vdd,vss,y);

input a, vdd, vss;
electrical a, vdd, vss;

output y;
electrical y;

parameter delay=0, ttime=1p;
real result;

analog begin
if(V(a)>=(V(vdd)/2))
result = V(vss);
else
result = V(vdd);

V(y) <+ transition(result, delay, ttime);
end

endmodule
```

Appendix B.2: Verilog-A code for 4-bits ADC

```
// VerilogA for skodaja_cv, BGR_V_ADC4, veriloga

'include "constants.vams"
'include "disciplines.vams"
'define bits 4
module BGR_V_ADC4(in, out);

parameter full_scale=16.0, delay=0, ttime=10n;
parameter real  vth = 2.5;

input in;
electrical in;

output ['bits-1:0] out;
electrical ['bits-1:0] out;

real sample, thresh;
real result[0:'bits-1];
integer ii;

analog begin
@(initial_step or initial_step("dc", "ac", "tran", "xf")) begin
generate i ( 'bits-1, 0) begin
result[i] = 0;
end
end

begin  //@(cross(V(clk)-vth, +1)) begin
sample = V(in);
thresh = full_scale/2.0;
 for (ii='bits-1; ii>=0; ii=ii-1) begin
if (sample > thresh) begin
result[ii] = 1;
sample = sample - thresh;
```

```
end
else result[ii] = 0;
sample= 2.0*sample;
end
 end
generate i ( 'bits-1, 0) begin
V(out[i]) <+ transition( result[i], delay, ttime);
end
end

endmodule

'undef bits
```