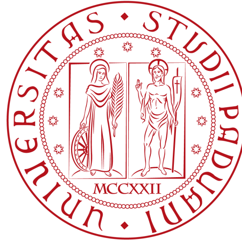


Università Degli Studi Di Padova

Facoltà di Ingegneria

Corso di Laurea: Ingegneria Elettronica

Investigation, Analysis and Design of a Sub-Bandgap Voltage Reference for Ultra-Low Voltage Applications

RELATORE
Prof. Bevilacqua Andrea

I CORRELATORE
Ing. Piselli Marco

II CORRELATORE
Ing. Flaibani Marco

TESI DI LAUREA DI
Zampieri Stefano
Matr. N. 586274

Anno Accademico [2011/2012]

To my parents, for all the immense love they have always been giving to me.

Contents

1 Introduction **1**

1.1 Voltage Reference Purpose . . . 1

1.2 Historical Overview . . . 3

1.3 Temperature Dependence . . . 6

1.3.1 $V_{BE}(T)$ Derivation . . . 6

1.3.2 Discussion on Approximations and Secondary Effects . 12

1.4 Compensation Orders . . . 14

1.5 Functional Blocks: Current References . . . 16

1.5.1 PTAT Current Generator . . . 16

1.5.2 $PTAT^2$ Current Reference, Bipolar Implementation . . 18

1.5.3 The Output Stage . . . 20

1.6 Impact On a Voltage Regulator Performance . . . 23

2 Design **29**

2.1 The Second Order Curvature-Corrected Sub-Bandgap . . . 29

2.1.1 Design Overview . . . 32

2.1.2 Performance Overview . . . 35

2.1.3 Performance Evaluation . . . 38

2.2 Choosing the Topology . . . 39

2.3 Preliminary PTAT Stage Analysis . . . 43

2.4 Errors Compensation Detailed Analysis . . . 45

2.4.1 Collector Current Mismatch . . . 45

2.4.2 Loop Analysis . . . 46

2.4.3 Base Currents 47
2.4.4 Mirror Channel Length Modulation and NPN Early Voltage 49
2.4.5 Leakage 52
2.5 Diode Loop Topology Design 58
2.5.1 The PTAT Stage 58
2.5.2 The CTAT and the Non-Linear Stage 66
2.5.3 The Output Stage 71
2.5.4 The Start-up 72
2.5.5 PSRR 74
2.5.6 Pseudo-Supply Design 77
2.5.7 AC Stability 79
2.5.8 Monte Carlo 96
2.5.9 Extracted Simulation 102
2.5.10 Performance 104
2.6 Proposed Topology Improvements 105

3 Measurements Vs Simulations 107
3.1 Analyzed Topologies and Procedures 107
3.2 Measurements VS Simulation 111
3.2.1 Voltage Reference and Current Consumption 111
3.2.2 Line Regulation 113
3.2.3 Leakage 113

4 Discussion And Comparison With Other Topologies 117
4.1 Main Objectives 117
4.2 Final Comparison and Evaluation 118
4.2.1 PSRR and Transient Response 120
4.2.2 Line Regulation and Start-Up 120
4.2.3 Nominal Precision and Statistical Analysis 121

5 Conclusions 125

A Appendix **127**

A.1 Base Emitter Voltage Calculation 127

A.1.1 General $V_{BE}(T)$ expression 127

A.1.2 Precise $V_{BE}(T)$ calculation 128

A.2 Linear Regualtor Analysis 129

A.2.1 Block Diagram of LDO Linear Regulator 129

A.2.2 PTAT ”W” transfer function calculation 129

A.3 AC Stability 130

A.4 Monte Carlo 132

A.4.1 Resistors Mismatch 132

Ringraziamenti **135**

List of Figures

1.1 Hilbiber's first voltage reference: the ancestor. 3
1.2 Widlar's first bandgap reference. 4
1.3 Brokaw's bandgap voltage reference. 5
1.4 Bandgap voltage versus absolute temperature and its first degree approximation (not to scale). 9
1.5 N-channel MOSFET, Id over Vgs, temperature as parameter. . 14
1.6 PTAT current generator. 17
1.7 Bipolar $PTAT^2$ current generator. 19
1.8 $PTAT^2$ current generator with horizontal NPN placement. . . 19
1.9 Output Stages. 21
1.10 Mixed-mode output topology. 23
1.11 LDO linear regulator. 24
1.12 Linear voltage regulator block diagram. 25
1.13 (a)$PSRR_{system}$ in function of $PSRR_{BG}$. (b)$PSRR_{BG}$. 28

2.1 Leakage trimmable compensation and $PTAT$ stages. 30
2.2 Spt5 sub-Bandgap: $CTAT$, $PTAT^2$ and output stages. 31
2.3 Output stage. 32
2.4 Curvature correction scheme. 33
2.5 Sub-bandgap reference performance. 36
2.6 Monte Carlo $V_{ref}(T)$ spread for three typical temperatures. . . 37
2.7 TC 100% yield for $Vref$. 38
2.8 "Diode Loop" topology. 40
2.9 Diode Loop Bandgap Core V2 (full version). 42
2.10 The $PTAT$ Stage. 43

2.11 I_C over V_{BE} characteristic. 45
2.12 $PTAT$ Loop block diagram. 46
2.13 Early voltage and saturation region approaching effects. 50
2.14 Early voltage test topology. 51
2.15 Collector-currents ratio. 51
2.16 Leakage parasitic devices and model. 52
2.17 The $PTAT$ Stage. 59
2.18 I_B compensation results. 60
2.19 Leakage compensation NPN. 61
2.20 V_{PTAT} error caused by leakage currents. 63
2.21 PTAT stage, version 3. 65
2.22 CTAT stage. 67
2.23 Tuned VS default mirror ratio. 69
2.24 $R_{NL} \pm 30\%$ span. Balance over T_r effect. 70
2.25 Output Stage, Version 2. 71
2.26 Start-up stage. 73
2.27 Pre-regulated pseudo-supply. 75
2.28 Pre-regulator circuit. 77
2.29 "Cold" current boost. 79
2.30 The pseudo-supply PTAT stage, complete representation. . . . 80
2.31 Pseudo-supply AC-loop block diagram. 82
2.32 No compensation STB analysis. 83
2.33 MATLAB verification of poles and zeros placement in the pseudo-supply: simulator data are dotted. 85
2.34 The Bandgap core PTAT stage, complete representation. . . . 86
2.35 MATLAB verification of poles and zeros placement in the PTAT stage: simulator data are dotted. 91
2.36 The Bandgap core CTAT stage, complete representation. . . . 92
2.37 MATLAB verification of poles and zeros placement in the CTAT stage: simulation data are dotted. 95
2.38 Diode loop topology, explicits circuit. 96
2.39 Corners on resistors mismatch error. 97
2.40 PTAT stage, mirror mismatch. 98

2.41 Current mirror mismatch, PTAT stage contribute. 99
2.42 Current mirror mismatch, CTAT stage contribute. 100
2.43 NPN mismatch, PTAT stage contribute. 101
2.44 NPN mismatch, CTAT stage contribute. 101
2.45 Improving steps in layout debug. 103

3.1 Leakage compensation structure. 108
3.2 Connections setup for "Set 1". 108
3.3 Connections setup for "Set 2". 109
3.4 Connections setup for "Set 2", inside view. 110
3.5 FIB cut: microscope picture. 110
3.6 V_{BG} measurements VS V_{BG} corner values. 111
3.7 I_{in} measurements VS I_{in} corner values. 112
3.8 Output resistance measurement Vs nominal value. 113
3.9 FIB measurements. 115
3.10 Leakage current: measurements VS simulation. 116

4.1 Performances comparison. 119
4.2 Nominal output voltage. 122
4.3 Mixed-mode output stage of Bg3. 123

A.1 NPN mirror effect, small signal schematic. 129
A.2 Small signal circuit for cascoded PTAT stage AC analysis. . . 131

List of Tables

1.1 Blocks table. 24

2.1 Specifications table. 36
2.2 Leakage explanation to V_{ref} variation discussed in 2.4.4. 57
2.3 Leakage values for $V_A = -800m$. 62
2.4 Specifications table: Diode-Loop Sub-Bandgap. 104
2.5 Improvements table: mixed mode VS diode loop. 104
2.6 Improvements table: first VS second design. 105

3.1 Line regulation estimation. 113
3.2 V_{BG} variation between no compensation and FIB2 modification. 116

4.1 Analyzed structures. 118
4.2 Precision, area and current consumption performance. 122

Abstract

The purpose of this thesis is to study and fully comprehend how to realize a very high performance sub-bandgap (low-voltage) structure. In order to achieve this result, it was necessary to begin from (and often come back to) the physics of semiconductor devices before moving to an analog approach in order to design the voltage reference itself. New formulas, as practical as accurate, will be derived in order to be able to successfully handle the many problems one can face during the design of the proposed topology. Parallel to this design activity, it was possible to study an already developed sub-bandgap structure, comparing measurements to simulation results and extracting important evaluations, useful to further improve both the designs. Layout and extracted simulations have also been taken into account, in order to assure the best reliability and performance matching.

Sommario

Lo scopo di questa tesi è quello di studiare e di comprendere appieno come realizzare un riferimento di tensione sub-bandgap ad alte prestazioni. Per ottenere questo risultato, è stato necessario partire dalla fisica dei dispositivi (e spesso ritornarci su) prima di passare all'ambito della progettazione analogica del circuito stesso. Saranno introdotte nuove formule di particolare utilità pratica, pur restando con un ottimo grado di accuratezza, così da poter gestire con successo i diversi problemi che ci si trova ad affrontare durante la fase di design. Parallelamente a quest'attività, è stato possibile studiare una struttura sub-bandgap già sviluppata facendo confronti tra le misure fatte e le simulazioni, cosa che ha portato ad interessanti deduzioni, utili per un ulteriore sviluppo di entrambi i design. Sono stati analizzati anche il layout e le simulazioni con estrazione di parassiti così da assicurare la miglior affidabilità del circuito ed il miglior matching con il design svolto.

Chapter 1

Introduction

1.1 Voltage Reference Purpose

Today, as much in the past, many functional blocks in an integrated circuit need a reference in order to work properly, be it voltage, current or time. A reference establishes what is that value that will be scaled, compared, followed and generally determines the value that will set a stable point that all other sub-circuits will use in order to work properly and to generate a predictable result. The most common examples of circuits that need a stable reference are analog-to-digital and digital-to-analog converters, operational amplifiers, sense comparators, reset circuitries, line regulators and so on. Being the reference for the circuit, a drift of the reference voltage will, in most situation, strongly impact over the performance of the whole IC itself. A change in the reference voltage can cause a reset circuit to move its threshold, or an analog-to-digital converter to have bit errors. A noisy reference can inject noise in a operational amplifier and in a linear regulator. The need for a very stable, low noise and precise reference clearly plays a pivotal role in modern integrated circuit. The design of such a reference has to be done considering the system in which it will work and a lot of other specifications in order to fulfill every requirement successfully and grant a high accuracy. In order to get high accuracy, there are two very important aspect to take care of:

One of the most important factor to consider is temperature: a temperature compensation can be done exploiting the temperature-dependent behavior of the components in order to reduce the sensitivity of the overall voltage reference to the temperature itself. Usually a voltage that is proportional to absolute temperature (from now on PTAT) is summed with some complementary to absolute temperature voltage (from now on CTAT) in a way that the output voltage of the reference will have low-voltage variations over the operating temperature range. These voltage references are named "Bandgap reference", because the output voltage produced is related to the bandgap of silicon, as will be shown in subsec. 1.3.1. The wider the temperature range to compensate, the more challenging it will be to grant a low-voltage variation. In fact, it is usually defined a temperature coefficient (TC) so to quantify this concept.

$$TC_{ref} = \frac{1}{Reference} \cdot \frac{\triangle Reference}{\triangle Temperature} \tag{1.1}$$

and it is usually expressed in parts-per-million per degree Celsius (ppm/°C).

The other critical factor is the line regulation and the PSRR. These functions describe the impact of the input voltage variations on the output, that in this case is the reference voltage. The bandgap voltage reference must have a very high rejection to power supply variation (high PSR) in order to have the lowest sensitivity to voltage variations of the battery and noise. Usually this is quantified by the PSRR (power supply rejection ratio) transfer function.

$$PSRR_{AC} = \frac{v_{ref}(f)}{v_{in}(f)} \tag{1.2}$$

This transfer function usually is intended to express the rejection ratio over frequency so we marked it as $PSRR_{AC}$; even though for

$$\lim_{f \to 0} PSRR(f) \approx PSRR_{DC}$$

This value is calculated for a single bias point, so I find useful to define

$$PSRR_{DC}(V_{IN}) = \frac{\partial V_{OUT}(V_{IN})}{\partial V_{IN}} \tag{1.3}$$

that describes the line regulation. This way, both the frequency and DC domain (that describes the line regulation) are taken into account. These and many other aspects will be discussed in detail, but before venturing into the many details that have to be disclosed, considering the important role that a voltage reference holds in IC electronic, an historical overview is given to elucidate the evolution of the bandgap reference over the years.

1.2 Historical Overview

In 1964, Hilbiber published the first bandgap reference [11]. He proposed to compensate for the temperature behavior of a base-emitter voltage by adding and subtracting several base-emitter voltages with different first-order temperature behaviors. Zener diodes were still very poor and he was looking for something that drifted less over time. It was already known that transistors with base and collector connected together made almost ideal diodes. Hilbiber took two of the Fairchild's discrete transistors with greatly different forward voltages (which he attributed to different diffusion profiles) and made two strings with different numbers of transistors. As his method used several stacked base-emitter voltages, as shown in Fig. 1.1, so the required power supply was relatively large compared with the reference voltage. He found a current level at which, over a narrow temperature

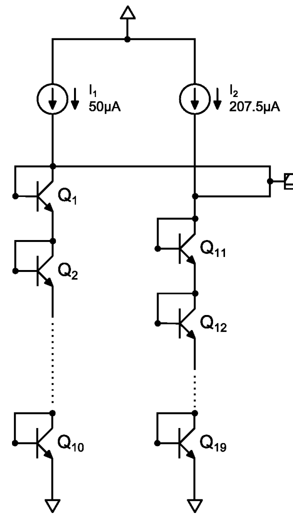


Figure 1.1: Hilbiber's first voltage reference: the ancestor.

range (±2°C), the voltage difference between the two strings change little and amounted to 1.2567V. He attempted to find a relationship between this voltage and the bandgap potential of silicon at zero Kelvin, but found that it was primarily a function of the semiconductor material used in the two different transistors. He got what he was after, a much better long-term stability, and he stopped at that.

Nothing happened for six years, when in 1971, Widlar put in the missing pieces proposing a new basic scheme of a bandgap reference requiring a lower supply voltage and this subsequently became commonly used. He recognized that the difference in diffusion profiles was only a secondary effect and the idea would work better if the two transistors where made by identical process. Plotting the V_{BE}, you will notice that it points at the bandgap potential at absolute zero. The bandgap voltage at zero K is strictly a theoretical concept: at that temperature the material is not a semiconductor anymore, being all the electrons absolutely still. His method was based on the compensation of the first-order temperature behavior of the base-emitter voltage with a voltage which is proportional to the absolute temperature. He had found this PTAT voltage in 1965 by using difference of two junction voltages. His idea was to create an opposite temperature coefficient that can be created by running transistors at different current densities:

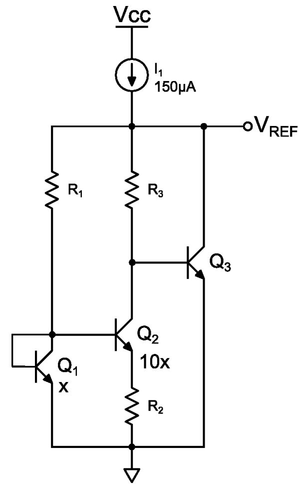


Figure 1.2: Widlar's first bandgap reference.

$$\Delta V_{BE} = \frac{kT}{q} \ln \left(\frac{A_{e1} I_2}{A_{e2} I_1} \right) \tag{1.4}$$

The problem was that this PTAT voltage raises slow over temperature, too slow to compensate the V_{BE} behavior. Widlar's solution was simple: multi-

ply the PTAT voltage for a resistor ratio. This circuit, shown in Fig. 1.2, can be briefly analyzed in the following way: R1 creates a current in Q1. Q2 has ten times the emitter area of Q1, so there is a ΔV_{BE} over the resistor between the two transistors of about 60mV at room temperature. This ΔV_{BE} shows up across R2. Neglecting the error due to base-current, emitter and collector currents of Q2 are equal. Thus the voltage drop across R3 is $\Delta V_{BE} \cdot \frac{R3}{R2}$. Adding to this voltage the V_{BE} of Q3, we get V_{REF}. The three transistors use a feedback loop, holding V_{REF} at a constant level. Where Hilbiber first made two appropriate stacks of base-emitter voltages, with a different fist-order temperature-compensated reference voltage, Widlar made a relatively small voltage with a linear temperature behavior, whereupon this voltage was amplified to cancel the first-order term of the base-emitter voltage. Widlar implemented the amplification of the voltages closer to the output of the reference. In 1973, Kuiji made an integrated bandgap reference using Hilbiber's idea. He used, however, an additional scaling factor in his reference such that output voltages different from the bandgap voltage could be realized.

Four years after the Widlar, Paul Brokaw published a paper entitled "A Simple Three-Terminal IC Bandgap Reference". This new two-transistor circuit uses a collector current sensing to eliminate errors due to base currents. He presented his voltage reference as a simpler and more flexible structure, especially for three-terminal applications. This cell offers separate control over output voltage and temperature coefficient in a circuit using only a single control loop. It also has low voltage capability, supplying a stable 2.5V output with operating supply bias down to 4V.

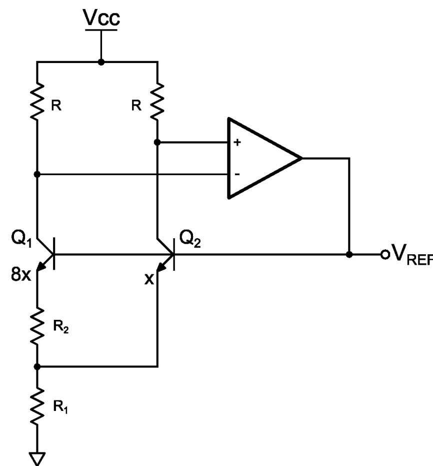


Figure 1.3: Brokaw's bandgap voltage reference.

1.3 Temperature Dependence

As mentioned in sec. 1.1, the components in the technology show different temperature dependence behaviors. However, the fundamental brick exploited for most temperature compensation is the forward-biased pn junction, being this in a diode or in a bipolar transistor. The well known equation

$$I_c(T) = I_s(T) \exp\left(\frac{qV_{BE}(T)}{kT}\right) \tag{1.5}$$

for the bipolar, where k is the Boltzmann constant and q is the electron charge, can be reversed to get one of the expressions for $V_{BE}(T)$. The I_C to V_{BE} characteristic needs to be carefully handled because many approximations in its derivation are commonly used that can bring to an excessive inaccuracy. It is very useful to remember also that

$$\frac{kT}{q} = V_T \tag{1.6}$$

For a certain reference temperature, we will also have

$$\frac{kT_r}{q} = V_{T_r} \tag{1.7}$$

1.3.1 $V_{BE}(T)$ Derivation

Solving (1.5) for $V_{BE}(T)$, we get

$$V_{BE}(T) = \frac{kT}{q} ln\left(\frac{I_C(T)}{I_S(T)}\right) \tag{1.8}$$

where I_C is the collector current, T is the absolute temperature, q the electron charge and k the Boltzmann constant. What about $I_S(T)$ then? As Barrie Gilbert said:

> ”If $V_{BE}(T)$ is the hearth of a bipolar transistor, then $I_S(T)$ must surely be its soul!”[1]

$I_S(T)$ is given by

$$I_S(T) = \frac{qA_E n_i^2(T)\bar{D}(T)}{N_B} \tag{1.9}$$

where A_E is the base-emitter junction area, $n_i(T)$ is the intrinsic carrier concentration, $\bar{D}(T)$ is the "effective" minority carrier diffusion constant in the base and N_B is the Gummel number (total number of impurities per unit area in the base).

The intrinsic carrier concentration

One of the pivotal point in this discussion is to correctly evaluate $n_i^2(T)$: as we know from the mass-action law

$$n_i^2(T) = n(T)p(T) \tag{1.10}$$

and

$$n(T) = N_C \exp\left(-\frac{E_C - E_F}{kT}\right) \tag{1.11}$$

where N_C is the effective density of states in the conduction band and is given by

$$N_C = 2\left(\frac{2\pi m_{de} kT}{h^2}\right)^{3/2} M_C \tag{1.12}$$

where M_C is the number of equivalent minima in the conduction band and m_{de} is the density-of-state effective mass for electrons and is given by

$$m_{de} = (m_1^* m_2^* m_3^*)^{1/3} \tag{1.13}$$

where m_i^* are the effective masses along the principal axes of the ellipsoidal energy surface. Similarly, we can obtain

$$p(T) = N_V \exp\left(-\frac{E_F - E_V}{kT}\right) \tag{1.14}$$

where N_V is the effective density of states in the valence band and is given

by

$$N_V = 2\left(\frac{2\pi m_{dh} kT}{h^2}\right)^{3/2} \tag{1.15}$$

where m_{dh} is the density-of-state effective mass of the valence band

$$m_{de} = \left(m_{lh}^{*3/2} m_{hv}^{*3/2}\right)^{2/3} \tag{1.16}$$

where m_{lh} and m_{hv} refer to light and heavy hole masses[1]. Substituting (1.11) and (1.14) in (1.10) gives, for the intrinsic carrier density:

$$np = n_i^2 = N_C N_V \exp\left(\frac{-E_g}{kT}\right) \tag{1.17}$$

$$n_i^2 = N_C N_V \exp\left(\frac{-E_g}{kT}\right) = \left(4.9 \cdot 10^{15}\left(\frac{m_{de} m_{dh}}{m_0^2}\right)\right)^2 M_c T^3 \exp\left(\frac{-E_g}{kT}\right) \tag{1.18}$$

where $E_g = (E_C - E_V)$ is the energy gap between the valence and conduction band. This equation leads to

$$n_i^2(T) = BT^3 \exp\left(-\frac{qV_G}{kT}\right) \tag{1.19}$$

where $V_G = {}^{E_g}/_{q}$ and B encloses all the non-temperature-dependence variables.

The Bandgap Voltage

The bandgap voltage V_G (and then the energy gap too of course) is in fact function of temperature and should be written as

$$n_i^2(T) = ET^3 \exp\left(-\frac{qV_G(T)}{kT}\right) \tag{1.20}$$

A plot of $V_G(T)$ is show in fig.1.4. A first approximation, denoted by $\hat{V}_G(T)$

[1] Please refer to [2] for more detailed physical explanation.

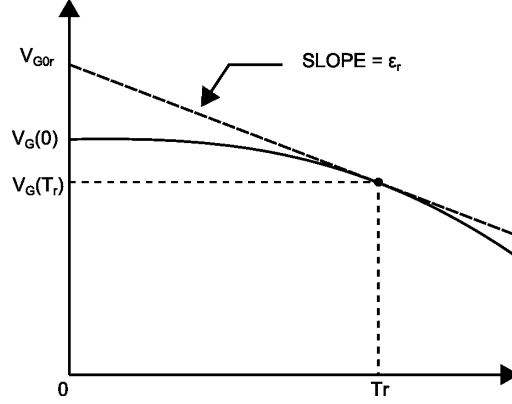


Figure 1.4: Bandgap voltage versus absolute temperature and its first degree approximation (not to scale).

is shown in fig.1.4. The straight line must be taken tangent to the exact curve for the certain T_R where we desire to have maximum accuracy. This will lead to

$$\hat{V}_G(T) = V_{G0r} + \epsilon_r T \tag{1.21}$$

where

$$\epsilon_r = \left(\frac{dV_g}{dT}\right)_{T=T_r} \tag{1.22}$$

The quantity V_{G0r} is then

$$V_{G0r} = V_G(T_r) - \epsilon_r T_r \tag{1.23}$$

It is important to keep in mind that what is commonly denoted as "the bandgap voltage" or "the bandgap voltage at 0K" is an extrapolated quantity and depends on the temperature T_r where the extrapolated line is chosen to be tangent to. From experimental data discussed in [3], the lower the temperature, the more severe the non-linearity in its temperature dependence gets as is shown in fig.1.4.

According to [4], an expression for $V_G(T)$ is

$$V_G(T) = V_G(0) - \frac{\alpha T^2}{T + \beta} \tag{1.24}$$

According to [4], these values are, for intrinsic silicon: $\alpha = 7.021 \cdot 10^{-4}$ $^{V}\!/_{K}$ and $\beta = 1108$ K, but other sources bring values that differ for more than 50% from those just mentioned. In addition, these values are chosen as to match (1.24) with measurements taken from near 0K to over 400K. Taking into account that the most non-linearity is at low temperature, these values are a compromise "best fit" for the whole temperature range, and are not accurate enough for IC circuits, that usually operate in a range from 200K to 450K. A more accurate measurement of $V_G(T)$is reported in [5], where the error with the following relation is within 0.2mV.

$$V_G(T) = 1.178 - 9.025 \cdot 10^{-5}T - 3.05 \cdot 10^{-7}T^2 \tag{1.25}$$

for $150K < T < 300K$. Sadly, this range is too little for our purpose, so we used an extrapolated first degree polynomial[2] for $T > 300K$:

$$V_G(T) = V_{G0npn} - \gamma T \tag{1.26}$$

where V_{G0npn} is the V_{G0r} extrapolated for the specific device and $\gamma = 2.7325 \cdot 10^{-4}$ $[^{V}\!/_{^\circ C}]$. Being the non-linearity occurring for temperatures much lower than 300K, this approximation still retains high accuracy for the range where it is defined. It is interesting that, if we substitute (1.26) into (1.19), we get

$$n_i^2(T) = ET^3 \exp\left(-\frac{q(V_{G0npn} - 2.7325 \cdot 10^{-4}T)}{kT}\right)$$

where the exponential part can be seen as the product of a constant term $\exp\left(\frac{q \cdot 2.7325 \cdot 10^4}{k}\right) = B$ and a temperature dependent term $\exp\left(-\frac{qV_{G0npn}}{kT}\right)$, so the final result is a corrected version of (1.19), $n_i^2(T) = BT^3 \exp\left(-\frac{qV_{G0npn}}{kT}\right)$

The effective minority carrier diffusion constant

Now that we know an exact formula for $V_G(T)$ and $n_i^2(T)$, we can now move to a brief evaluation of the last temperature dependent term: $\bar{D}(T)$. This

[2] This extrapolation is find in [3].

term is given by

$$\bar{D}(T) = V_T \bar{\mu_n}(T) \tag{1.27}$$

where $\bar{\mu_n}$ is the average mobility for minority carrier in the base

$$\bar{\mu_n}(T) = CT^{-n} \tag{1.28}$$

Accurate evaluation of Base-Emitter voltage temperature dependence

Now that all the temperature dependent parameters of our $I_S(T)$ have been explicitly related to temperature, we can express $V_{BE}(T)$ in its most general and accurate form. Let us consider two temperatures: an arbitrary temperature T and a reference temperature T_R. Applying (1.8) for these two temperature, we get

$$\begin{cases} V_{BE}(T) = \frac{kT}{q} ln\left(\frac{I_C(T)}{I_S(T)}\right) \\ V_{BE}(T_r) = \frac{kT_r}{q} ln\left(\frac{I_C(T_r)}{I_S(rr_r)}\right) \end{cases} \tag{1.29}$$

multiplying the first equation for T_r and the second one for T and subtracting them, we derive the following expression:

$$V_{BE}(T) = \left(\frac{T}{T_r}\right)\left\{V_{BE}(T_R) + \left(\frac{kT_r}{q}\right)\ln\left[\frac{I_S(T_r)}{I_S(T)} \cdot \frac{I_C(T)}{I_C(T_r)}\right]\right\} \tag{1.30}$$

Let us now use (1.9) in (1.30), with $n_i^2(T)$ given by (1.20) and $\bar{D}(T)$ given by (1.27). This brings to the following relation[3]:

$$V_{BE}(T) = V_G(T) - \left(\frac{T}{T_r}\right)V_G(T_r) + \left(\frac{T}{T_r}\right)V_{BE}(T_r) + \frac{kT}{q}\ln\left[\left(\frac{T}{T_r}\right)^4 \frac{\bar{\mu}(T_r)}{\bar{\mu}(T)}\frac{I_C(T)}{I_C(T_r)}\right] \tag{1.31}$$

[3]Calculation in appendix (A.1)

and using (1.28) in(1.31) we get:

$$\begin{aligned} V_{BE}(T) &= V_G(T) - \left(\frac{T}{T_r}\right)[V_G(T_r) - V_{BE}(T_r)] - \eta\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right) + \\ &+ \left(\frac{kT}{q}\right)\ln\left[\frac{I_C(T)}{I_C(T_r)}\right] \end{aligned} \quad (1.32)$$

where $\eta \equiv 4 - n$. A particular case for (1.32) happens when the collector current is proportional to some power of T:

$$I_C(T) = FT^x \quad (1.33)$$

Using (1.33) in (1.32), we obtain a simplified expression for $V_{BE}(T)$:

$$V_{BE}(T) = V_G(T) - \left(\frac{T}{T_r}\right)[V_G(T_r) - V_{BE}(T_r)] - (\eta - x)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right) \quad (1.34)$$

and using (1.26), our expression for $V_G(T)$ related to our NPN transistor, we get

$$V_{BE}(T) = V_{G0npn} - \left(\frac{T}{T_r}\right)[V_{G0npn} - V_{BE}(T_r) - \gamma T_r] - (\eta - x)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right) \quad (1.35)$$

1.3.2 Discussion on Approximations and Secondary Effects

The equations in subsec. 1.3.1 on page 6 have different degrees of precision, from quasi-exact to high precision. One of the most widely used expression for $V_{BE}(T)$ is (1.34) (very often used with $V_G(T) = \hat{V_{G0}}$). However, this case does not occur so easily: temperature dependence of resistors used to set $I_C(T)$ can lead to a current that does not vary as described in(1.33). In this case, (1.32) must be used and in order for this equation to be considered valid, the following comments are needed:

1. $\bar{\mu}$ is an "effective" mobility of the *minority* carrier in the base, but all the existing data on mobility is for *majority* carrier instead. So we

are assuming that the temperature dependence of the minority carrier mobility is the same as the mobility of majority carriers in a material of the opposite polarity and same doping concentration.

2. The constant in equation (1.28) depends on the impurity concentration. For a standard PNP device it is reasonable to consider this constant throughout the base. For a standard NPN device, the impurity concentration varies with depth, and therefore so do the constants in the mobility expression. $\bar{\mu}(T)$ is a single effective mobility, so it takes into account the global effect of all individual mobility. It follows, then, that what is true for a specific impurity concentration is true for the effective mobility.

3. Relation (1.28) is given at "around room temperature" and could not be accurate in the whole temperature range of interest.

For these reasons, from case to case, we have to wisely consider which equation to use in relation to what kind of data and accuracy we want to obtain. Let us now consider the intrinsic carrier concentration described in (1.20): the quantity "E" is proportional to the average density-of-states effective mass and this varies with temperature, but more complex expression for $n_i(T)$ does not lead to a significant improvement[6]. Fabrication process must also be considered, because they lead to a deviation of the parameter η in (1.35) and from the basic relation described in (1.5). This relation become

$$I_C(T) = I_S(T) \exp\left(\frac{V_{BE}(T)}{\alpha kT}\right) \tag{1.36}$$

where α is slightly grater then 1, and is current and temperature dependent and can change form device to device in the same wafer. Parasitic resistance can also impact on the I_C over V_{BE} characteristic. Let us consider R_C, a collector series parasitic resistance. The intrinsic V_{CE} will be

$$V'_{CE} = V_{CE} - R_C I_C \tag{1.37}$$

and this can become a problem, especially for low V_{CE}, especially for a PTAT I_C where, for higher temperature, the V_{BE} needed to support a given I_C may become higher than what would normally be required.

1.4 Compensation Orders

When designing a bandgap voltage reference, one of the main goals is to achieve a temperature independent output voltage. To reach this purpose in ICs domain, it is common to use devices or structures that have different temperature dependence in order to sum their contribute, or even trying to neglect every contribute exploiting some particular characteristic of the device itself. Just as an example, refer to fig.1.5 . This figure shows that if

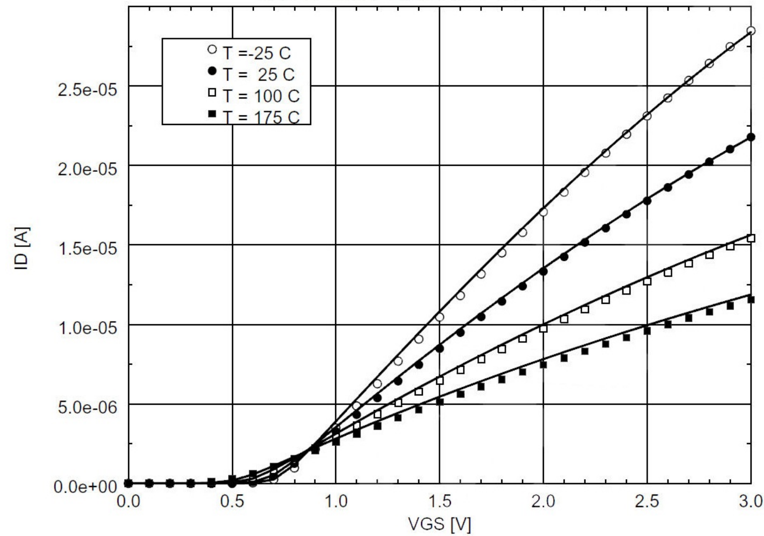


Figure 1.5: N-channel MOSFET, Id over Vgs, temperature as parameter.

this n-channel MOSFET is biased with a precise current, its V_{GS} shows very little temperature dependence. If the source of this n-channel MOSFET is grounded, its gate can be taken as a "stable" voltage reference. Needless to say, we would need a current that should be constant over temperature, so the problem just lays somewhere else. The bipolar transistor opens the possibility of exploiting its $V_{BE}(T)$ in order to get a temperature compensation with very high precision. In sec. (1.3), we have derived different descriptions of the

base-emitter voltage, from quasi-exact to very high precision approximation. Let us recall (1.35)

$$V_{BE}(T) = V_{G0npn} - \left(\frac{T}{T_r}\right)\left[V_{G0npn} - V_{BE}(T_r) - \gamma T_r\right] - (\eta - x)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right) \tag{1.38}$$

and let us take a look at its temperature dependence: theV_{BE} relationship can be rewritten as

$$V_{BE}(T) = A + BT + Cf(T) \tag{1.39}$$

where $f(T)$ represents all terms whose order is greater than 1. Ok, so our dream is to have $V_{ref} = A$, or at least $A + \sigma(T)$ in the temperature range of interest, but how can we get this result? One way is to use the Taylor series expansion of $V_{BE}(T)$ and try to compensate every order until the required precision has been reached. Another way can be exploiting the $(\eta - x)$ term, forcing a collector current proportional to T^{η} so that C=0 in (1.39) and easily compensate the B term. It is also possible to try to compensate B and $Cf(T)$ directly. All these solutions need to be implemented with a circuital topology that must also fulfill many other specifications other than extremely low temperature sensitivity and not all the parameters that are visible on the equation can be easily manipulated. The first and second order compensation of the Taylor series expansion of $V_{BE}(T)$ is a very effective technique and widely used for bandgap compensation, so it will be evaluated. The only term to expand in (1.35) is $(\eta - x)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right)$, being the other terms of zero or first degree.

$$(\eta - x)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right) = (\eta - x)\frac{V_{T_r}}{T_r}\left[T\ln\left(\frac{T}{T_r}\right)\right] \tag{1.40}$$

$$= K\left[T\ln\left(\frac{T}{T_r}\right)\right]$$

$$= K\sum_{n=0}^{\infty}\left[\frac{a_n(T - T_r)^n}{n!}\right] \tag{1.41}$$

with

$$a_k = \left. \frac{\partial^k \left[T \ln \left(\frac{T}{T_r} \right) \right]}{\partial^k T} \right|_{T=T_r} \tag{1.42}$$

With calculation, we get

$$\begin{aligned} V_{BE}(T) &= [V_{G0npn} + (\eta - x)V_{T_r}] \\ &\quad - \left(\frac{T}{T_r} \right) [V_{G0npn} - V_{BE}(T_r) + \gamma T_r + (\eta - x)V_{T_r}] \\ &\quad + \sigma(T) \end{aligned} \tag{1.43}$$

for the first order expansion, and

$$\begin{aligned} V_{BE}(T) &= \left[V_{G0npn} + \frac{(\eta - x)V_{T_r}}{2} \right] - \left(\frac{T}{T_r} \right) [V_{G0npn} - V_{BE}(T_r) + \gamma T_r] \\ &\quad - \frac{T^2}{T_r^2} V_{Tr} \left[\frac{(\eta - x)}{2} \right] + \sigma(T^2) \end{aligned} \tag{1.44}$$

for the second order.We need to implement some circuits that can generate a PTAT and a PTATT^2 signal. This blocks are essential not only for the straight order to order compensation, but also for most complex bandgap structures and pseudo-supply structures.

1.5 Functional Blocks: Current References

1.5.1 PTAT Current Generator

PTAT current reference can be implemented both with CMOS and bipolar transistors. For our purpose, we will exploit the NPN bipolar properties. Fig. 1.6 shows one possible implementation[4].

[4]For more BJT and CMOS implementations, see [7].

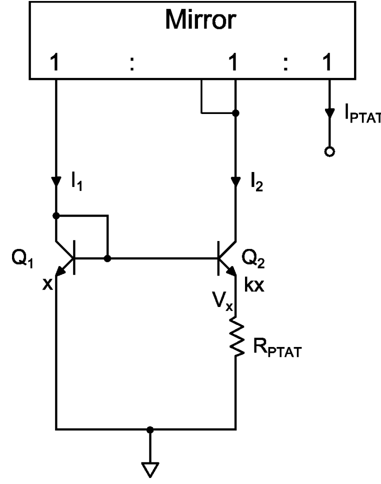


Figure 1.6: PTAT current generator.

This circuit exploit the $V_{BE}(T)$ logarithmic relation of eq. (1.8). Applying KVL to the loop with Q_1, Q_2 and R_{PTAT}, we get:

$$V_{BE1} - V_{BE2} - R_{PTAT}I_R = 0 \tag{1.45}$$

Solving (1.45) for I_R we get

$$I_R(T) = \frac{V_T \ln\left(\frac{I_{C1}(T)}{Ae_1 J_S(T)}\right) - V_T \ln\left(\frac{I_{C2}(T)}{Ae_2 J_S(T)}\right)}{R_{PTAT}} \tag{1.46}$$

$$= \frac{V_T \ln\left(\frac{I_{C1}(T)}{Ae_1 J_S(T)} \cdot \frac{Ae_2 J_S(T)}{I_{C2}(T)}\right)}{R_{PTAT}} \tag{1.47}$$

Because of the current mirror and neglecting the base-currents[5], we can write, for (1.47):

$$I_R(T) = \frac{V_T \ln(K)}{R_{PTAT}} = T \cdot \frac{k \ln(K)}{q} \tag{1.48}$$

where $K = \frac{I_{C1}}{I_{C2}}$. If R_{PTAT} had "zero order" temperature dependence, this current would be PTAT. We will see later in chapter 2 how this non-ideality will impact on the overall errors and performances. The drawback of this implementation is that it has a zero-current state, so it needs a start-up

[5]Errors will be discussed in subsec. 2.4.

in order to prevent the circuit from approaching this state. These PTAT generators can be implemented also with lateral PNP structure, but these topologies are less robust than the NPN counterpart because of two factors: a low collector-current efficiency (especially if a highly doped buried layer is not available) and a lightly doped base region. The first factor is due to the existence of a parasitic vertical PNP transistor whose collector is connected to substrate. As a result, some of the emitter current flows to substrate, thereby reducing the collector-current efficiency of lateral PNP device. The second drawback is due to the decrease of the current gain (forward-β) with increasing the collector-current because of high-level injection. This phenomenon occurs when minority carrier density in the base region becomes comparable with the majority carrier density. This happens because neutrality must be maintained. Higher majority carrier density shorten the minority carrier life (there are more majority carriers to recombine with) and increases the effective doping density of the base. For proper operation, the minority carriers should be well below the majority carrier level. This effect is still present but less effective on NPN because of the highly doping density of the base.

1.5.2 $PTAT^2$ Current Reference, Bipolar Implementation

A $PTAT^2$ current generator is also necessary in order to implement a precise, second order compensated, voltage reference. The common circuital implementation is shown in Fig. 1.7. Considering KVL composed by transistor Q_1, Q_2, Q_3 and Q_4 and using for each base-emitter voltage eq. (1.8), it yields

$$V_T \ln \left[\frac{I_{PTAT^2} \left(\frac{I_{PTAT}}{A} + I_{CTAT} \right)}{I_{PTAT} \cdot I_{PTAT}} \right] = 0 \tag{1.49}$$

so

$$I_{PTAT^2} = \frac{I_{PTAT}^2}{\left(\frac{I_{PTAT}}{A} + I_{CTAT} \right)} \approx \frac{I_{PTAT}^2}{K} \tag{1.50}$$

where

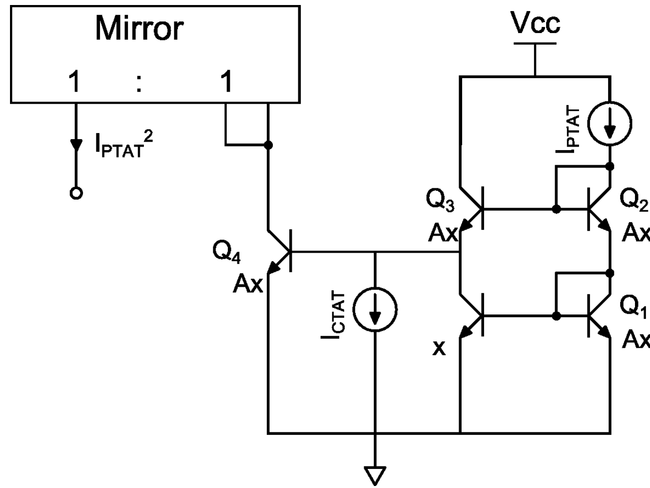


Figure 1.7: Bipolar $PTAT^2$ current generator.

$$\frac{I_{PTAT}}{A} + I_{CTAT} \approx constant \tag{1.51}$$

This circuit however stacks two NPN transistor and so it is not suitable for very-low voltage environment. Another way to get a $PTAT^2$ current generator without vertical NPN stacking is shown in Fig. 1.8.

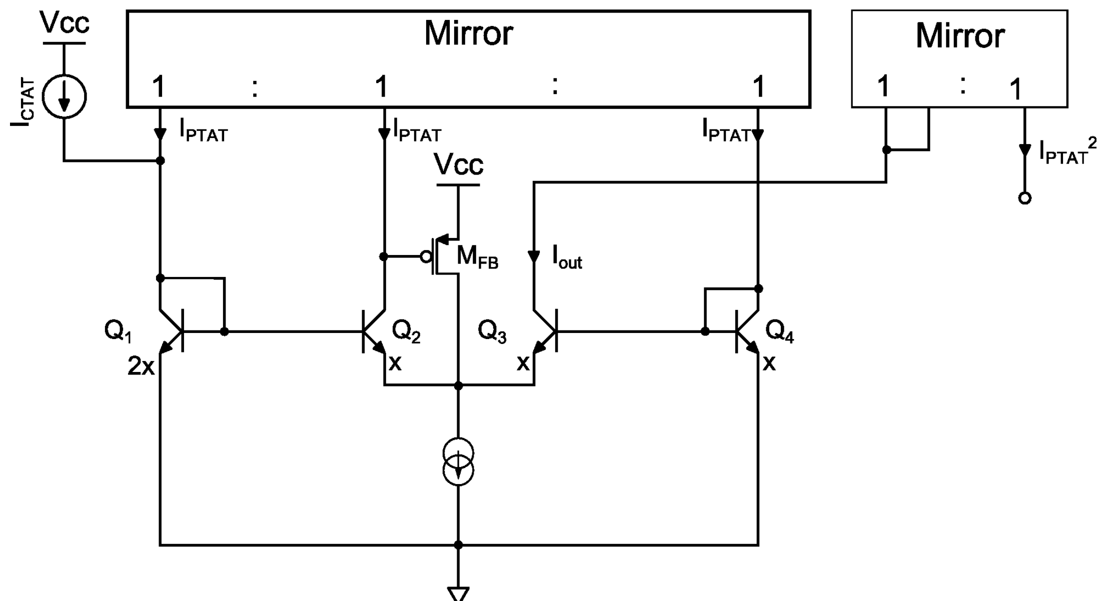


Figure 1.8: $PTAT^2$ current generator with horizontal NPN placement.

KVL for this circuit yields

$$V_T \ln \left[\frac{I_{out} \left(\frac{I_{PTAT}+I_{CTAT}}{2} \right)}{I_{PTAT} \cdot I_{PTAT}} \right] = 0 \tag{1.52}$$

so again

$$I_{out} = I_{PTAT^2} = \frac{I_{PTAT}^2}{\left(\frac{I_{PTAT}+I_{CTAT}}{2} \right)} \approx \frac{I_{PTAT}^2}{K} \tag{1.53}$$

but this circuit needs less headroom then the one shown in Fig. 1.7. Both the circuits suffer of inherent errors and approximations that can invalidate their precision. The most remarkable is due to the temperature coefficient of resistors used for generating the PTAT current that are CTAT in nature. Their behavior is not linear with temperature but with quadratic law so its effect will happen for high temperature, just where the second order compensation should kick in. When this effect begins, the $PTAT^2$ current begins to get linear, and the precision decreases. This interferes with (1.51) too because, the higher the temperature, the more the I_{CTAT} gets parabolic and the more I_{PTAT} gets flat, thereby leading to a CTAT current.

1.5.3 The Output Stage

The output stage of the voltage reference must be designed according to the load and the headroom limits for the given application. The main forms for an output stage are three: voltage-mode, current-mode and mixed-mode. Mixed-mode refers to circuits employing both voltage-mode and current-mode techniques. Voltage-mode easily implements low-impedance output stage but lacks the ability to choose the output level. On the other hand, current-mode is not suitable for implementing low-impedance output, but the output level can be easily regulated. If, for instance, the load is a large capacitor, there may not be the need for a low-impedance output.

Voltage-Mode

The voltage-mode output is the most common technique and has been used for long time because of its simplicity: in case of bandgap references, the

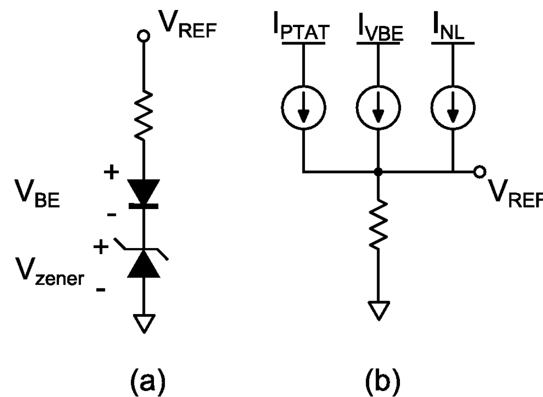


Figure 1.9: Output Stages.

output is trimmed merely changing the size of one resistor, realizing a PTAT trim voltage. Moreover, the basic temperature dependent components are used directly in the output stage, thereby not introducing further errors. Fig. 1.9(a) illustrates the typical output structure of first order Zener and bandgap references. The positive temperature coefficient (TC) of the Zener diode is summed to the negative TC of the forward-biased diode voltage. This kind of output stage provides a voltage reference of roughly 5-7 volts, and is not suitable for low voltage application. The market is driving power supply down, thereby creating a limit for this output stage: battery powered devices cannot be implemented with such an output stage. Even simple bandgap reference, like a resistance in series with a forward-biased diode, are appropriate only for devices with an output voltage lower limit of 1.5 V (roughly the 1.2V bandgap voltage plus the 300mV for p-channel MOSFET bias). This is not a limitation for battery powered circuits. However, single battery-cell operation and lower breakdown voltages will eventually require to operate at supply voltage of about 1V. Nickel-cadmium (NiCd) or nickel-hydride (NiMH) batteries have 1.5 V at their output but will decay to 0.9 V before collapsing: a suitable reference voltage for systems working in this environment is about 0.7 V, impossible to reach with voltage-mode output stage.

Current-Mode

A current-mode output stage is obtained summing temperature-dependent currents into a resistor, as show in Fig. 1.9(b). This way, only the value of the currents and of the resistor determine the value of the output voltage. This configuration offers a wide spectrum of choices for the output voltage, accommodating a range from millivolts to several volts. For most curvature-corrected bandgap references, the currents involved are PTAT, CTAT and non-linear. This makes possible to trim the output voltage simply by trimming one of the components, usually the PTAT current. Moreover, even though current-mode outputs suffer the fact that, in most cases, PTAT and CTAT current must be generated from a certain voltage over a resistance, then mirrored, and then summed into a load, they still retain a low sensitivity to resistors TC. This happens because the transfer function between the voltage that generates that current and its effect on the output voltage is a resistor ratio, like:

$$V_{out} = V_{in} \frac{R_{load}}{R_{in}} \tag{1.54}$$

so the TCs of the resistors are deleted. However, the effectiveness and simplicity of the voltage-mode is almost lost in the many errors occurring in transferring the voltage input reference (PTAT, CTAT, etc..) to the output load. This aspects will be further discussed in subsec. 2.4 of chapter 2.

Mixed-Mode

Mixed-mode, as the name states, combines the output stages mode above in order to get the benefits of both (but also inheriting some errors...). This structure is basically a current-mode topology for its current transfer function nature, but instead of summing all the contributes into a single load, it uses a resistor ladder, as snow in Fig. 1.10. This approach offers the possibility of lowering the output voltage while the voltage-mode ladder provides enhanced flexibility for temperature compensation. In fact, in the current-mode stage, the currents must be TC compensated before flowing into the load. In mixed-mode, it is the ladder what weights the single current contribution. Referring

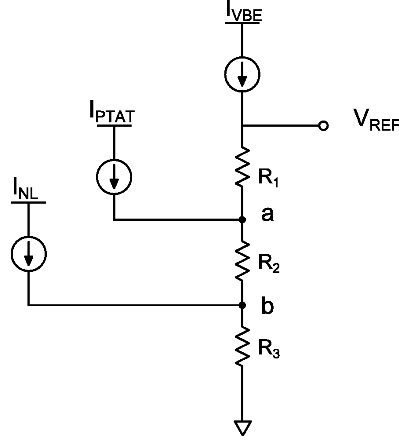


Figure 1.10: Mixed-mode output topology.

to Fig. 1.10:

$$V_{ref} = I_{V_{BE}}(R_1 + R_2 + R_3) + I_{PTAT}(R_2 + R_3) + I_{NL}R_3 \qquad (1.55)$$

where $I_{V_{BE}}$, I_{PTAT} and I_{NL} correspond to the base-emitter, the PTAT and the nonlinear temperature-dependent currents respectively.

1.6 Impact On a Voltage Regulator Performance

We now comprehend how important it is to consider the system and the environment in which our voltage reference will work in in order to choose the right implementation design. The environment we will consider is automotive for this project from now on. One of the most likely system that will see our designed bandgap as voltage reference is a LDO (low drop-out) linear regulator in Fig. 1.11 . One of the main task for this regulator is a low noise, high PSRR, high precision, low voltage operation.

Lets us begin noticing that the load of our bandgap reference is the input stage of an OTA: this, combined with the low voltage operation specific, makes us choose for a current or mixed-mode output stage. The high precision demand over temperature and supply battery variations in automotive set the operating range for our device to [-40 150] °C for temperature and [4

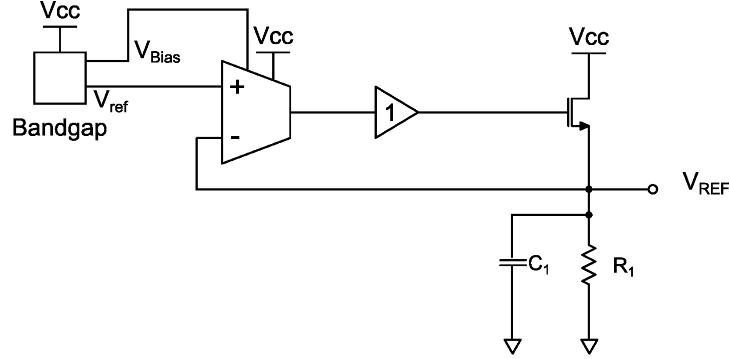


Figure 1.11: LDO linear regulator.

45] V for battery supply. Moreover, we want to guarantee a stable behavior also in the [150 180] °C range, where leakage effects dominate and can make over-temperature safe circuitry fail. The frequency range is set to be 10MHz so to cover also DC-DC converter environment.

PSRR

In order to study the PSRR impact over the linear regulator system, it is useful to describe it as a block diagram and, for doing so, we must also explicit all the transfer functions that are going to describe it. Let us start identifying the blocks.

Block Name	Block Function	Block Description
A	$\frac{V_{OTA}}{(V_+ - V_-)}$	The open loop gain of the OTA
B	$\frac{V_{0_{OTA}}}{V_{BIAS}}$	The open loop transfer function from V_{BIAS} to $V_{0_{OTA}}$
C	$\frac{V_{0_{OTA}}}{V_{CC}}$	The open loop transfer function from V_{CC} to $V_{0_{OTA}}$
H	1	The feedback gain
L	$\frac{V_{ref}}{V_{CC}}$	The "bandgap" PSRR to V_{ref} with no feedback from V_{BIAS}
M	$\frac{V_{BIAS}}{V_{CC}}$	The "bandgap" PSRR to V_{BIAS} with no feedback from V_{ref}
P	$\frac{V_{REF}}{V_G}$	Common source gain of the power stage
Q	$\frac{V_{REF}}{V_S}$	Common gate gain of the power stage

Table 1.1: Blocks table.

Now, suppose we have our signal $V_{in} = V_{CC}$ and $V_O = V_{REF}$, then we can write the block diagram of Fig. 1.12. We can now write the equation that

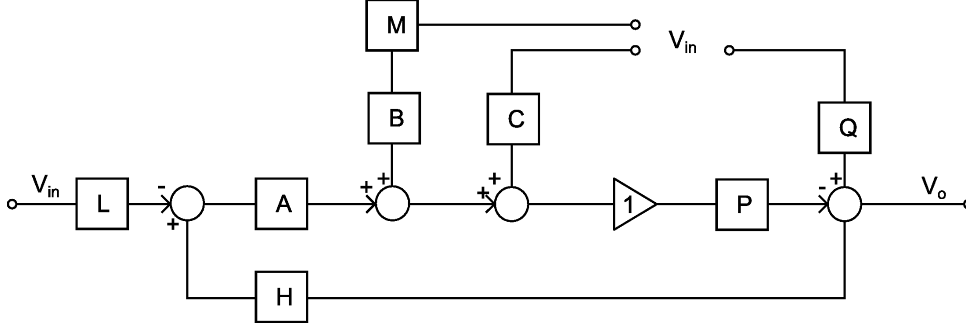


Figure 1.12: Linear voltage regulator block diagram.

describe this system[6]:

$$[(V_O H - LV_{in})\, A + V_{in} MB + V_{in} C]\,(-P) + V_{in} Q \;=\; V_O \tag{1.56}$$

$$\frac{V_O}{V_{in}} = \frac{P(LA - MB - C) + Q}{1 + HPA} \tag{1.57}$$

We can write this equation in a more intuitive way: doing so and substituting the diagram block letters with their meaning, we get

$$\frac{V_O}{V_{in}} = \frac{A_{OTA} L - MB - C + \frac{A_{CG}}{A_{CS}}}{\frac{1}{A_{CS}} + A_{OTA}} \tag{1.58}$$

Being $\frac{1}{A_{CS}} \ll A_{OTA}$, with $A_{OTA} \approx 370$ and $\frac{1}{A_{CS}} \approx 0.2$, we can write

$$\frac{V_O}{V_{in}} \approx \frac{A_{OTA} L - MB - C + \frac{A_{CG}}{A_{CS}}}{A_{OTA}} = L + \frac{\frac{A_{CG}}{A_{CS}} - C - MB}{A_{OTA}} \tag{1.59}$$

so we can write

$$PSRR = 20 \log \left(L + \frac{\frac{A_{CG}}{A_{CS}} - C - MB}{A_{OTA}} \right) \tag{1.60}$$

The OTA have $C \approx 1$, value that is confirmed by the simulator in the whole bandwidth of interest. The common-gate common-source ratio can be easily

[6]For full calculations see appendix A.2.1

calculated[7] and it follows that

$$\frac{A_{CG}}{A_{CS}} \approx 1 \tag{1.61}$$

and, like for the "C" term, this can considered a constant for the frequency range we are considering. A qualitative overview of eq. (1.60) tell us that the PSRR is dominated by the worst (higher) term between L and $\frac{\frac{A_{CG}}{A_{CS}}-C-MB}{A_{OTA}} = PSRR_{ideal}$[8]. Practically, we can consider dominant the term that is about 5 times greater than the other term. This way, we get a shift of

$$20 \cdot [\log(5x) - \log(5x + x)] = 20 \cdot log\left(\frac{5}{6}\right) = 1.58dB \tag{1.62}$$

It can come in handy to express the error shift in function of L and

$$\frac{\frac{A_{CG}}{A_{CS}} - C - MB}{A_{OTA}} = \Gamma$$

writing

$$\epsilon_{shift} = -20 \cdot \log\left(1 + \frac{L}{\Gamma}\right) \tag{1.63}$$

From simulations of this LDO regulator, we get that the $PSRR_{ideal} = -103dB$ so our target will be to design a bandgap reference with about a $\frac{1}{5}$ factor from linear $PSRR_{ideal}$, that in dB is $20log\left(\frac{1}{5}\right) = -14dB$, and what we expect is too get a $PSRR_{system} \approx -100dB$. All these calculations are confirmed by simulation as it is shown in Fig. 1.13. Looking at Fig. 1.13 (a), there is a $\Delta PSRR_{system} = 2.1dB$ for $PSRR_{BG}$ of -115dB (12dB less than $PSRR_{ideal}$ that is a x4 linear factor) where $PSRR_{BG} = 20\log(L)$. Equation (1.63) gives $\epsilon_{shift} = -1.93dB$, with good accuracy. Fig. 1.13 shows that, if the bandgap reference has $PSRR \gg PSRR_{ideal}$, then $PSRR_{system} = PSRR_{BG}$. We now have a target for our PSRR, that, in order not to impact on the performance of the regulator, must achieve a $PSRR_{BG} \lesssim -110dB$. Although all this dissertation regarded the DC part of the PSRR, AC speci-

[7] See [8], sec. 3.3.

[8] Note that the MB term depends on how we choose to bias the OTA and must be handled accordingly to the whole system.

fication will be treated in chapter 2.

Precision

Precision is one of the main matter with a voltage reference. Both design, layout and production (components tolerance, packaging stress, etc..) are involved. What we can do in order to achieve a good precision is to take into account all the errors, random and systematic, that may (and will) occur, both in the design and layout stage. Our goal for this voltage reference is to reach high precision, $\pm 1.5\%$ for $3\left(\frac{\sigma}{\eta}\right)$, being σ the standard deviation and η the average of the process.

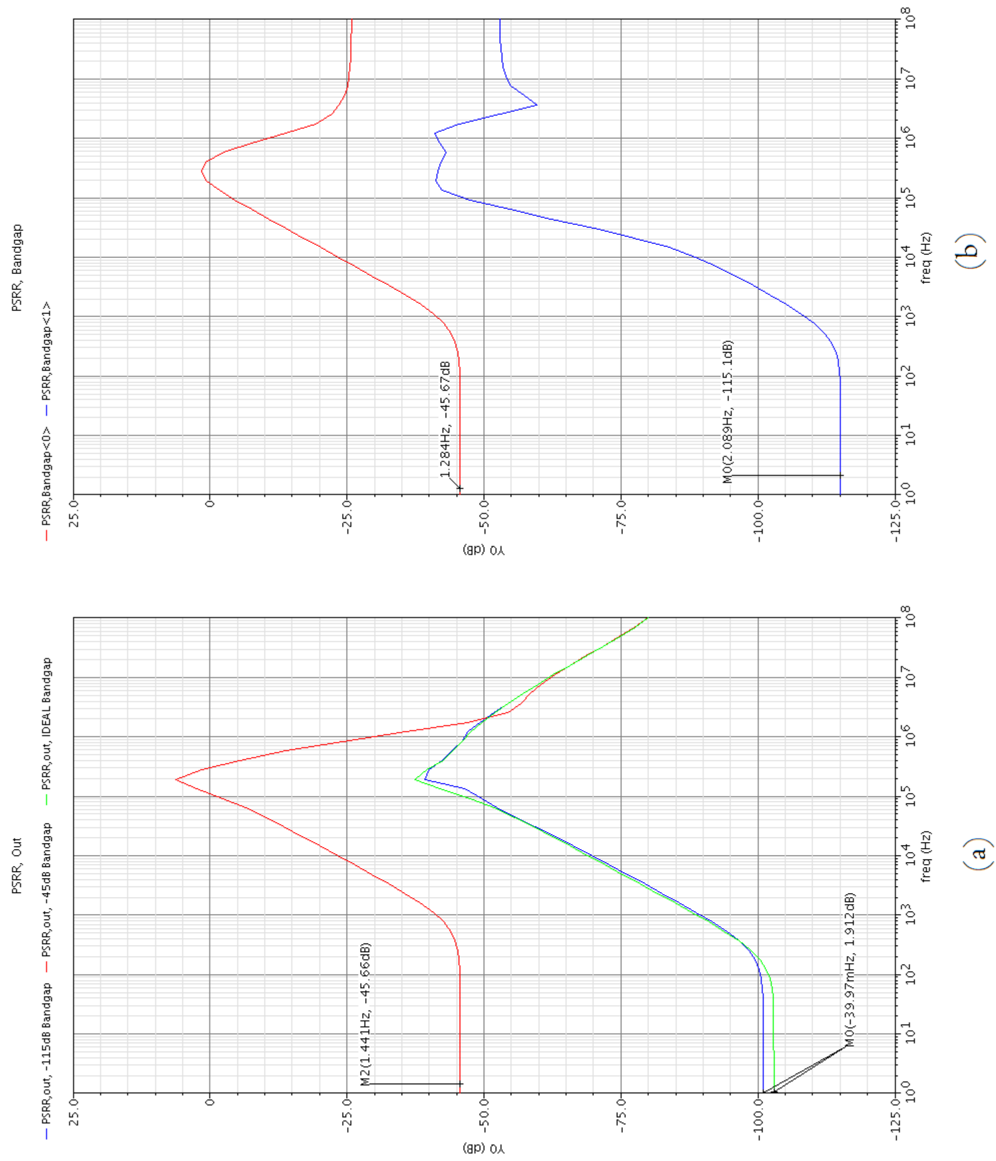


Figure 1.13: (a)$PSRR_{system}$ in function of $PSRR_{BG}$. (b)$PSRR_{BG}$.

Chapter 2

Design

In chapter 1 we saw which specification we have to fulfill. Our starting point in this thesis was however the analysis of an already develop sub-Bandgap structure. This analysis brought the main problems to light and showed what are the structural and physical limits for a voltage reference. Trying to overcome them will be our challenge.

2.1 The Second Order Curvature-Corrected Sub-Bandgap

The proposed sub-bandgap voltage reference is a second order curvature corrected, with mixed-mode output stage. Its schematic is in Fig. 2.1 and 2.2. It presents the $PTAT$ and low-voltage $PTAT^2$ stages seen in sec. 1.5 and uses a NPN transistor with a constant $I_C(T)$ to generate the $V_{BE}(T)$ reference. The mirrors are p-channel low-voltage MOSFETs "mp00p" with minimum channel length and all the circuit but the $PTAT^2$ stage use nst146p NPN standard transistor; $PTAT^2$ stage uses high-frequency oriented nhf112p. In Fig. 2.1 we see also a leakage-compensation trimmable stage, whose implementation is done such to simplify a FIB ("Focused Ion Beam") modification to cut one the desired net that connects one ore more of the NPN transistors connected to the drain of M1.

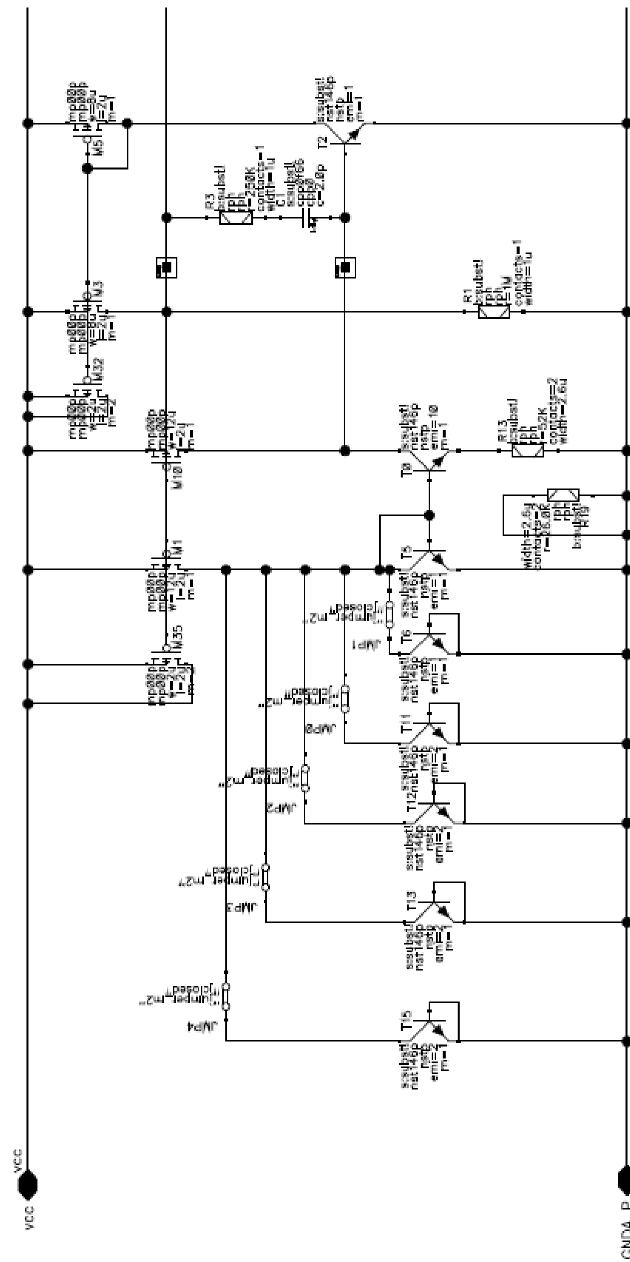


Figure 2.1: Leakage trimmable compensation and $PTAT$ stages.

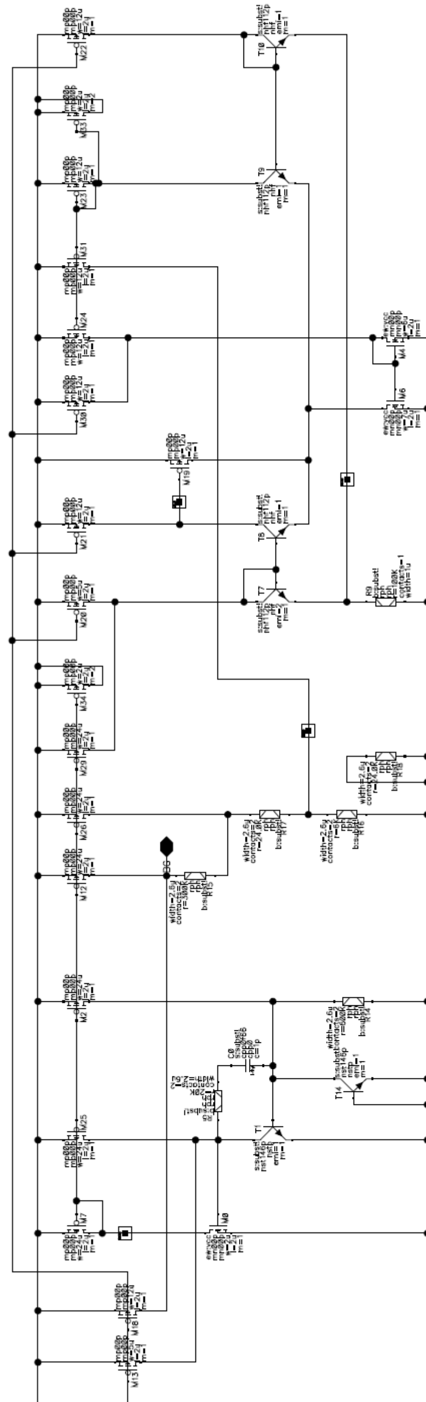


Figure 2.2: Spt5 sub-Bandgap: $CTAT$, $PTAT^2$ and output stages.

2.1.1 Design Overview

The mixed-mode resistor-ladder has been chosen as in Fig. 2.3, where I_{INDP} is the sum of a $PTAT$ and $CTAT$ current so that, once flowing into a load, the voltage they generate is "constant" in temperature. What really happens is that $I_{INDP} = I_{PTAT} + I_{CTAT}$ so it gives a first order compensated voltage over the load. It is very important to take into account that, having all resistors a certain TC (negative and quadratic for rph-poly resistors), a constant current does not generate a constant voltage over the load. So I_{INDP} will be $PTAT$ in nature, but with the TC that will compensate the resistors TC of the load. This is why the resistors involved in translating a voltage signal into current and those who will be the load for those currents must be of the same type.

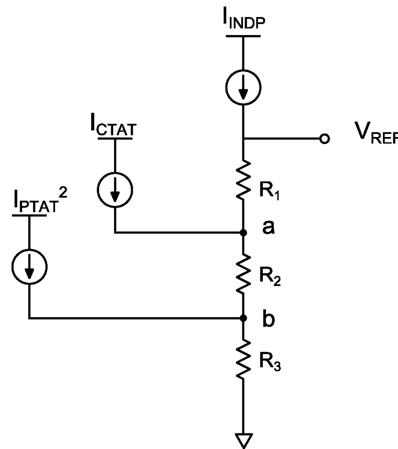


Figure 2.3: Output stage.

The output stage is configured in order to get the curvature compensation in Fig. 2.4: the classic first-order concave-down curvature is summed to a concave-up curvature so to yield a flat, second order compensated output. As shown in Fig. 2.4(a), the concave-up curvature is obtained by the sum of the $PTAT^2$ and the $CTAT$ voltage.

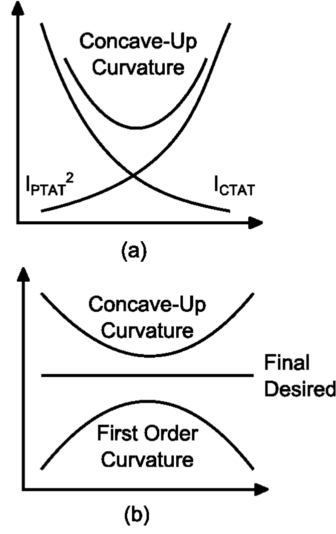


Figure 2.4: Curvature correction scheme.

The equation used for $V_{BE}(T)$ for this circuit is the Taylor series expansion stopped at first order of eq. (1.34), where $V_G(T)$ has been considered constant and equals to 1.2V. The $V_{BE}(T)$ formula used in this design is then

$$V_{BE_{first}}(T) = [V_{G0} + (\eta - x)V_{T_r}] - \left(\frac{T}{T_r}\right)[V_{G0} - V_{BE}(T_r) + (\eta - x)V_{T_r}] \tag{2.1}$$

The x term has been considered 0 since $I_{C_{T1}} \approx constant$. Recalling eq. (1.48), we have

$$I_{PTAT}(T) = \frac{V_t \ln(10)}{R_{PTAT}} \tag{2.2}$$

R_{PTAT} is dimensioned to set the magnitude of the $I_{PTAT}(T)$; this will set the whole circuit current consumption because all the other stage will be supplied by this current or have to compensate it, so they are strictly related to it. Being $V_{T_r}\ln(10) \approx 58mV$, choosing $R_{PTAT} = 52k\Omega$ sets the current to $1.1\mu A$ at ambient temperature and the span will be $[890nA - 1.73\mu A]$ considering a [-40°C +180°C] temperature range. The first order compensation will give

$$\left.\frac{\partial I_{CTAT_{first}}(T) + I_{PTAT}(T)}{\partial T}\right|_{T=T_r} = 0 \tag{2.3}$$

so it follows that

$$\frac{V_{T_r}\ln(10)}{T_r R_{PTAT}} = \frac{V_{G0} - V_{BE}(T_r) + \eta V_{T_r}}{T_r R_{CTAT}} \tag{2.4}$$

This gives straight forward a value for R_{CTAT}

$$R_{CTAT} = R_{PTAT} \cdot \frac{V_{G0} - V_{BE}(T_r) + \eta V_{T_r}}{V_{T_r}\ln(10)} \tag{2.5}$$

Using the simulated value for $V_{BE}(T_r)$ and $\eta = 4$, $R_{CTAT} = 598k\Omega \approx 600k\Omega$. I_{INDP} is also set and it becomes

$$I_{INDP}(T) = \frac{A_1 T}{R_{PTAT}} + \left[\frac{B_1}{R_{CTAT}} - \frac{C_1 T}{R_{CTAT}}\right]$$

with $A_1 = \frac{V_{T_r}\ln(10)}{T_r}$, $B_1 = V_{G0} + \eta V_{T_r}$ and $C_1 = \frac{V_{G0} - V_{BE}(T_r) + \eta V_{T_r}}{T_r}$. Writing the total output voltage, we can determine how to set load resistors seen in Fig. 2.2 on page 31

$$V_{ref} = I_{CONST} \cdot (R15 + R16 + R17) + I_{CTAT} \cdot (R16 + R17) + I_{PTAT^2} \cdot R16 \tag{2.6}$$

$V_{BE}(T)$ must be written with its Taylor series expansion for the second order so to calculate its second order coefficient, the $PTAT^2$ compensation and sizing R16 and R17. Doing so for eq. (1.34), again with the same approximations of eq. (2.1), we get

$$V_{BE}(T) = \left[V_{G0} + \frac{(\eta - x)V_{T_r}}{2}\right] - \left(\frac{T}{T_r}\right)[V_G(T_r) - V_{BE}(T_r)] - \frac{T^2}{T_r^2}\left[V_{Tr}\frac{(\eta - x)}{2}\right] \tag{2.7}$$

The $PTAT^2$ current has a not known coefficient due to the its nature: it is not the transduction of a voltage through a resistor, but the function of another current, so the TC of the resistor generating the source current (I_{PTAT} in this case) does not cancel itself. What is possible to do, is to estimate via simulation the law for $R16 \cdot I_{PTAT^2}$ and design R17 accordingly. R16 is so chosen just as to maintain the effect of $R16 \cdot I_{PTAT^2}$ in the right scale. Altering the circuit in order to evaluate $V_{PTAT^2} = I_{PTAT^2} \cdot (R16)$, we

are able to fit its curvature with

$$V_{PTAT^2} = 0.3\mu T^2 - 75\mu T + 10.7m \tag{2.8}$$

where the known term is just for fitting purpose, giving precision to the high range of temperature. Now we need that

$$\left.\frac{\left[\partial V_{PTAT^2} + \frac{V_{BE}}{R_{CTAT}}(R16 + R17)\right](T)}{\partial T}\right|_{T=T_r} = 0 \tag{2.9}$$

that yields

$$0.3\mu T_r - 75\mu - \frac{V_G(T_r) - V_{BE}(T_r)}{T_r R_{CTAT}} \cdot (R16 + R17) - \frac{V_{T_r}\eta}{T_r R_{CTAT}} \cdot (R16 + R17) = 0 \tag{2.10}$$

Easily inverting eq. (2.10), we get $R16 + R17 \approx 30k\Omega$, so $R16 = 24k\Omega$.

2.1.2 Performance Overview

A second order curvature-compensated bandgap reference should achieve a temperature drift performance from 1 to 20 ppm/°C [7], and we saw from sec. 1.6 that we need a $PSRR_{DC} < -100dB$ in order not to invalidate an hypothetical use in a voltage regulator. Another specification for more demanding voltage regulators is a $PSRR|_{f=10kHz} < -70dB$. We will however evaluate the $PSRR$ as function of V_{CC}. Moreover, we are going to evaluate the current consumption at nominal $V_{CC} = 3V$ in the temperature range of interest and the minimum supply to turn on the bandgap reference and the transient response to a$10\mu s$ start-up with all node discharged and $10\mu s$ shut down. Finally we will do a Monte Carlo analysis in order to see how process and mismatch will affect the precision of the bandgap reference output. In table 2.1, a quick view for this specifications.

	Specification	Value	Passed
Precision	1 to 20 ppm/°C	17ppm/°C	V
$PSRR_{DC}$	-100dB	-45.7dB	X
$PSRR_{10kHz}$	-70dB	-23.62dB	X
Low Voltage ($\leq 1.4V$)	$V_{on} < 1.4V$	$1.05V$	V
Monte Carlo	±1.5% for ($3\sigma/media$)	±13.1% for ($3\sigma/media$)	X
Current Consumption @$V_{CC} = 3V$	lowest	[16, 25, 40]μA for T [-40, 50, 180]°C	≈

Table 2.1: Specifications table.

Fig. 2.5 shows the simulation results for this schematic. Using eq. 1.1, we get

$$TC_{ref} = \frac{3m}{795m/220°C} = 17ppm/°\text{C} \tag{2.11}$$

so the TC performance are in specification.

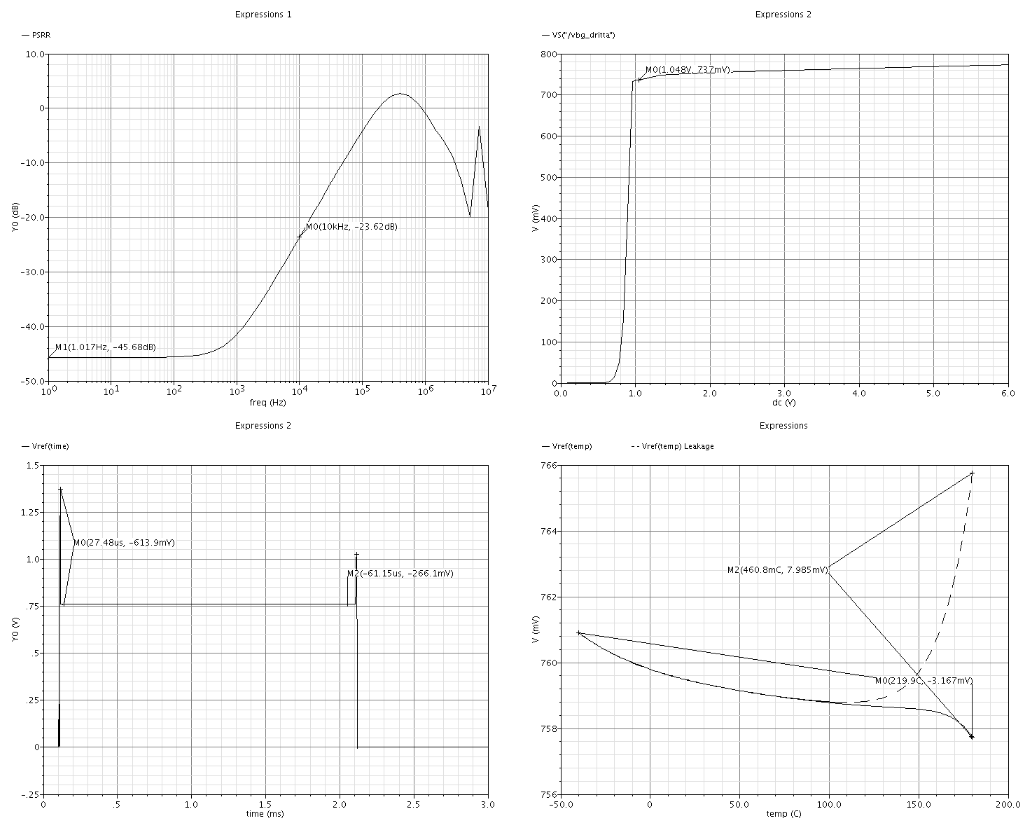


Figure 2.5: Sub-bandgap reference performance.

A statistical analysis has also been done for this bandgap. Even if the accuracy required sets a range of values the reference must respect in order to satisfy the specification (a normalized 3σ percentage for three different temperatures is considered), we find also important to keep track of the bandgap TC. A bandgap reference that statistically suffers of some offset but has stable temperature behavior is different from a bandgap reference with good bias at some reference temperature but with high TC. This way we can evaluate how much imprecision is due to an offset effect and how much to an error in TC. For this bandgap, we can see its process plus mismatch statistical analysis in Fig. 2.6.

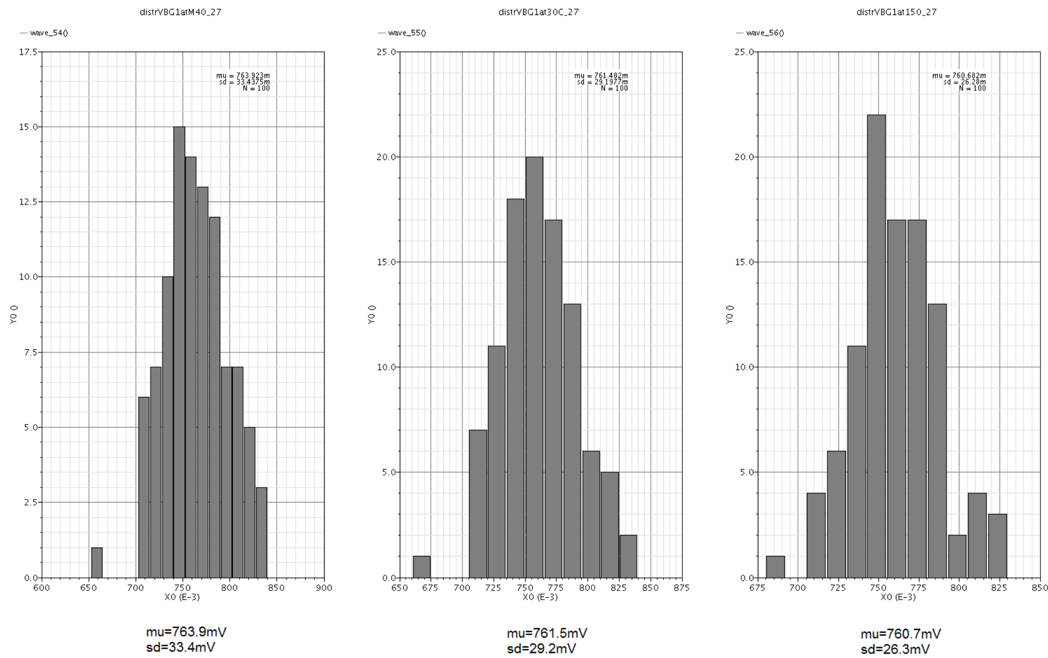


Figure 2.6: Monte Carlo $V_{ref}(T)$ spread for three typical temperatures.

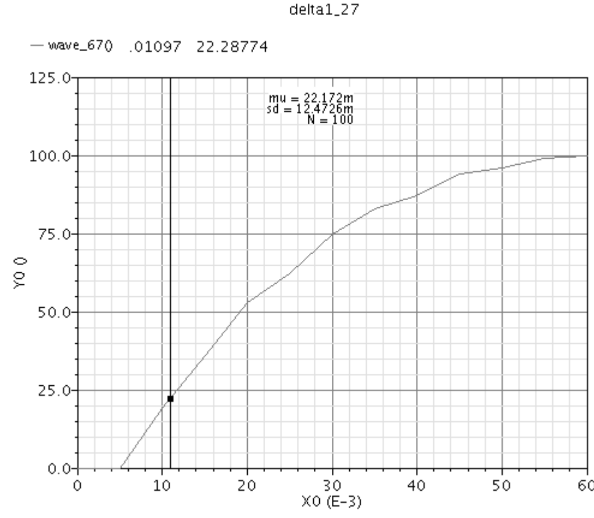


Figure 2.7: TC 100% yield for $Vref$.

We can see in Fig. 2.6, we can see we have a 3σ spread , normalized over the average, of

$$\epsilon_{\%} = \pm 3 \cdot \frac{33.4m}{763m} \cdot 100 = \pm 13.1\%$$

These values have to be combined with those of Fig. 2.7: this figure shows the percentage of samples that reaches a certain $\frac{\Delta V_{ref}}{V_{ref}}$ calculated in the whole 220°C of temperature range. As stated in [7], a second-order compensated circuit should have less than 50ppm/°C when realized. This means that we should have a 100% yield for a value, in the x-axis graph of Fig. 2.7, of

$$x_{ref} = 50ppm/°C \cdot T = 0.011$$

For that value, only the 22.3% of our samples are compliant.

2.1.3 Performance Evaluation

As we see from the previous subsec., nominal values for V_{ref} and current consumption are in specification, but Monte Carlo analysis reveals a weakness to process and mismatch (most is due to mismatch, see subsec. 2.5.8). Moreover, the second-order compensation requires a lot of current to be drained

from supply compared to the $PTAT$ and $CTAT$ stages because of the many branches involved in the $PTAT^2$ current generation. The $PSRR$ specification is just too high for a simple voltage reference[1]. We will have to find another solution for accomplish this task. We than see a strange behavior of V_{ref} for temperatures above 150°C. This is due to pn junction leakage currents. This phenomenon is very hard to compensate because it involves all the transistors in the circuit and all contributions have different weights depending on each transistor area and bias. This circuit has a leakage-compensation trimmable stage that tries to compensate this effect, but measurements will show that this strategy is effective only for a precise supply voltage. Leakage is also not precisely modeled in the simulator, so its compensation will require careful evaluations. Last but not least at all, the transient response shows a nervous behavior of the reference voltage when the supply is turned on and off; this shows an overshoot tendency that may not be acceptable, and must be taken into account while designing the loops frequency compensation.

Even though it is possible to redesign this circuit to better match all the performance required (and this will be done in sec. 2.6), some stages and design choices prevent further improvements. In order to create a "state of the art" voltage reference, we will now start a new design, beginning from the evaluation of the best-suited topology for our purpose and designing every stage aiming to prevent or reduce the errors before they occur and improving the synergy within the functional blocks of the circuit by avoiding errors propagation.

2.2 Choosing the Topology

In [7] we can find lots of ideas for voltage reference topology, but few envisage doing low-voltage low-current high precision voltage reference. Among the few, the one that best seems to suit our needs is the "Diode Loop" topology. The circuit in Fig. 2.8 describes the idea over which this topology is based. The trick consists in generating a non-linear voltage that tries to match ex-

[1]See subsec. 2.5.5

actly the $T\ln(T)$ behavior of $V_{BE}(T)$. This is obtained by the loop comprised from Q2, Q3 and R2:

$$V_{R3} = V_{BE2}(T) - V_{BE3}(T) = V_T \ln\left(\frac{I_{C1}A_2}{I_{C2}A_1}\right) \tag{2.12}$$

$$= V_T \ln\left(\frac{2I_{PTAT}}{I_{NL} + I_{constant}}\right) = V_{NL}(T) \tag{2.13}$$

where $I_{contant} = I_{PTAT} + I_{CTAT} + I_{NL}$. This formula shows recursive nature, but we will see how to handle this soon. The reference voltage will then be

$$V_{ref} = \left[\frac{V_{BE2}(T)}{R1} + I_{PTAT} + \frac{V_{NL}(T)}{R2}\right] \cdot R_{load} \tag{2.14}$$

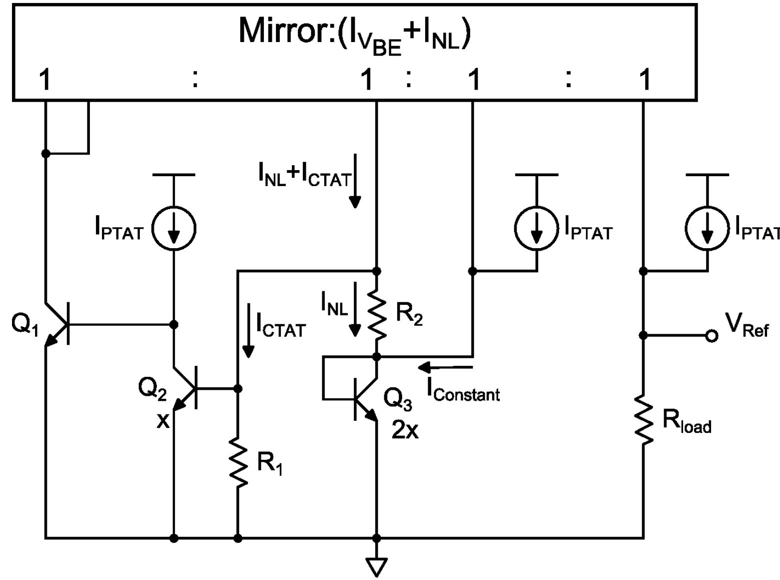


Figure 2.8: "Diode Loop" topology.

so it is necessary, in order to get resistors TC cancellation, to generate I_{PTAT} as $\frac{V_T}{R_{PTAT}}$, where R_{PTAT} must have the same TC of all other resistors. The $PTAT$ stage seen in sec. 1.5 suits our needs and then it will be used to generate the $PTAT$ current. Using the most accurate formula from subsec. 1.3.1, eq. (2.69), and the $PTAT$ stage standard equation, eq. (1.48), we can

write V_{ref} as

$$\begin{aligned} V_{ref} &= \left\{ V_{G0npn} - \left(\frac{T}{T_r}\right)\Theta_{T_r} - (\eta - x)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right)\right\} \cdot \frac{R_{load}}{R1} \\ &\quad + \frac{V_T \ln(K)}{R_{PTAT}} \cdot Rload + \frac{V_T}{R2}\ln\left(\frac{2I_{PTAT}}{I_{NL} + I_{constant}}\right) \cdot Rload \quad (2.15) \end{aligned}$$

with $\Theta_{T_r} = V_{G0npn} - V_{BE}(T_r) - \gamma T_r$ is constant. It is immediately visible that we should be able to get a perfect cancellation of both the linear and the $T \ln\left(\frac{T}{T_r}\right)$ to get $V_{ref} \approx V_{G0npn}$. The headroom limitation is ultimately defined by a diode-connected transistor, a drain-source voltage and a relatively small resistive voltage drop, which results in a voltage headroom limit of about 1V. We will now go through the implementation of this topology. In Fig. 2.9 we can see the final circuit in one of its three versions, divided into functional blocks. The main block is for sure the $PTAT$ stage, which sets the main variables for the whole circuit, so it will be the start of the design process. It is also the main source of errors. The $CTAT$ and non-linear stage will be sized after the $PTAT$ stage. The output stage is fully current-mode approach, to suite the low voltage requirement and because it is possible to sum currents that do not need a resistor ladder. The chosen $PTAT$ stage has a zero-current steady state, so it needs a start-up circuitry in order to work in the right state. This version includes a full leakage compensation and many output pins for testing purpose.

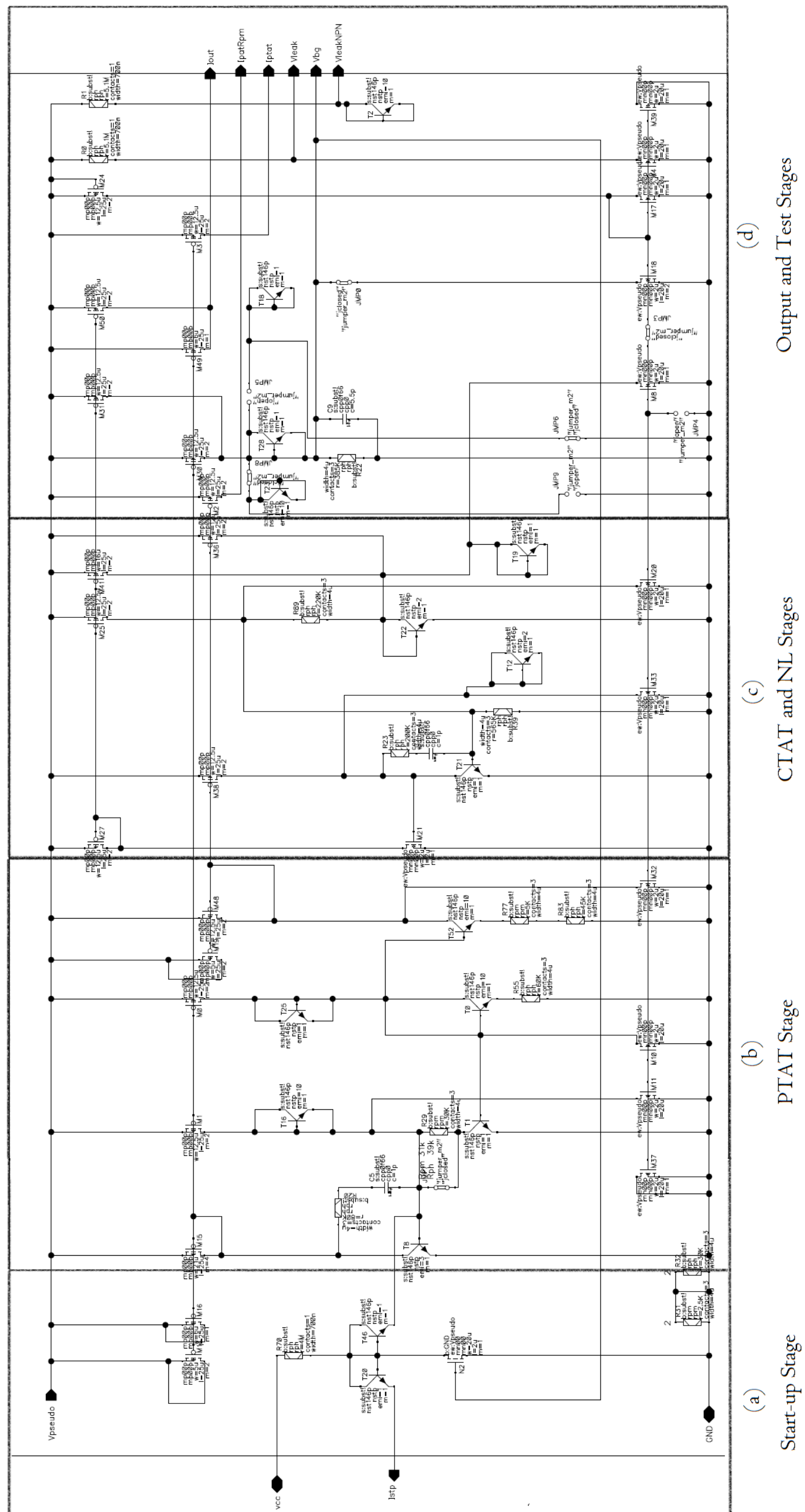


Figure 2.9: Diode Loop Bandgap Core V2 (full version).

2.3 Preliminary PTAT Stage Analysis

T_r is the reference temperature, and it is where the bandgap is more precisely compensated. The more is the distance from T_r, the more the errors are amplified. We decided to set $T_r = 333\text{K}$, so that V_{ref} is compensated at its mean value: setting $T_r = 300\text{K}$ would have left a very larger temperature range on the hotter side than it would on the colder side: this way, the error is distributed equally over the temperature range.

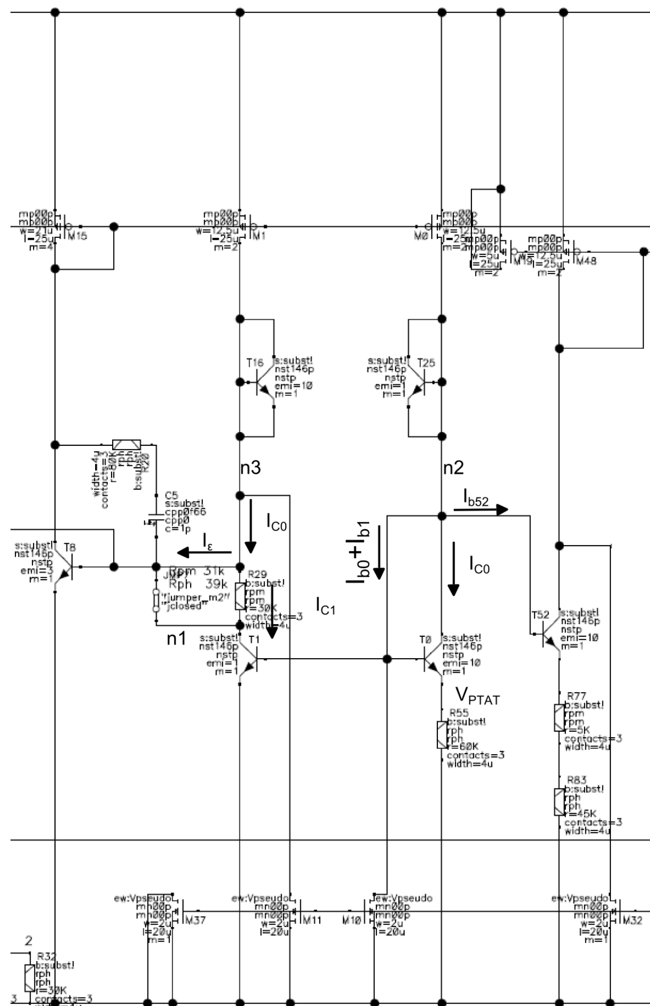


Figure 2.10: The *PTAT* Stage.

It will now be shown how the $PTAT$ stage of Fig. 2.10 has been designed. The first component that we are going to determine is the area ratio between T1 and T0. Area ratio for NPN device is well modeled and precise: using the maximum ratio will grant higher accuracy over the generated current, much higher than a resistor ratio. So the maximum value for A_{e0} has been chosen:

$$\frac{A_{e0}}{A_{e1}} = 10 \tag{2.16}$$

We can now choose the value of R55. This is set so to have a quiescent current of about $1\mu A$ at $T = T_r$. Using eq. (1.48)

$$\frac{V_{T_r} \ln(K)}{R55} = 1\mu A \tag{2.17}$$

$$R55 \approx 60k\Omega \tag{2.18}$$

For values above 50kΩ, rph poly silicon resistors are used. These resistors have a $CTAT$ temperature coefficient and this must be taken into account when using them to transduce a voltage into a current. In order for eq. (1.48) to be valid, let us write eq. (1.47) for this specific case:

$$I_{E0} = \frac{V_T \ln\left(\frac{I_{C1}(T)}{Ae_1} \cdot \frac{Ae_0}{I_{C0}(T)}\right)}{R_{PTAT}} \tag{2.19}$$

Eq. (2.19) shows that $\frac{I_{C1}}{I_{C0}}(T)$ must be constant over temperature. This task is performed by a loop that senses the difference $I_\epsilon(T) = [I_{C0} - I_{C1}](T)$ and regulates the mirror in order to minimize this error. This is necessary because of the different V_{BE} over I_C characteristics of T0 and T1. T0 is of an emitter-degenerated transistor while T1 is not, so they will behave as shown in Fig. 2.11: T0 has a greater I_C for low V_{BE} because of the low I_C itself that cause little voltage drop on the degenerating resistor and it has bigger area, but for higher V_{BE}, the I_C suffer the degeneration due to R55, the $PTAT$ resistor; T1 has lower area, so for low V_{BE}, the I_C is smaller, but for higher V_{BE} and I_C, it is free from degeneration, so the current can rise higher than the collector-current of T0.

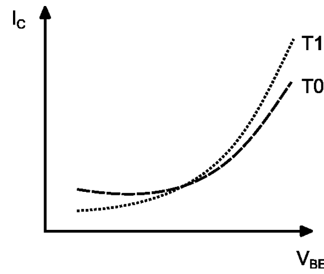


Figure 2.11: I_C over V_{BE} characteristic.

In order to fully comprehend and successfully design the $PTAT$ stage, now that we placed all the components, it is necessary to study the many errors affecting this stage before beginning the sizing of the components. This is necessary because of the high precision required. In fact, if we want to achieve a precision of few ppm over the [-40 150]°C range, we have to stay under about $150\mu V$ of ΔV_{ref}. We are biasing a load of about $400k\Omega$, so it takes only $0.5nA$ on the output bias current to have a error of $200\mu V$!

2.4 Errors Compensation Detailed Analysis

There are many sources of errors in the$PTAT$ stage. We are now going to evaluate them one by one in order to comprehend how they impact on the output voltage, in relation to temperature.

2.4.1 Collector Current Mismatch

We can see in $V_{PTAT}(T) = V_T \ln\left(\frac{I_{C1}(T)}{Ae_1} \cdot \frac{Ae_0}{I_{C0}(T)}\right)$ that if $\frac{I_{C1}(T)}{I_{C0}(T)} \neq 1$, $V_{PTAT}(T)$ is not linear with temperature anymore but it will become function of the collector-current ratio. There are two main sources of currents that can impact on the collector-current ratio: the base currents and the leakage currents. Moreover, we must not forget that we are inside a loop, so not only these currents will affect the V_{PTAT} generation, but will also be mirrored on the load and, beware, both these processes depend on the loop. This loop consists of two nested loops and must be studied before more considerations could be made.

2.4.2 Loop Analysis

The loop block diagram is shown in Fig. 2.12, where n1, n2 and n3 are the nodes marked in Fig. 2.10. These are the nodes where the leakage and base currents are injected or sunk, so it is useful to calculate each transfer function between them and the output current, that is taken as the collector current of M0 (that is also equal to the PTAT load current). The block diagram has been written accordingly to the circuit, where β is the current gain $\frac{I_{C,T8}}{I_{B,T8}}$, M is the mirror ratio between M15 and M0=M1=M30 that is set as $M = 3.36$ and W is the open loop transfer function $W = \frac{I_{C,T0}}{I_{C,T1}}$.

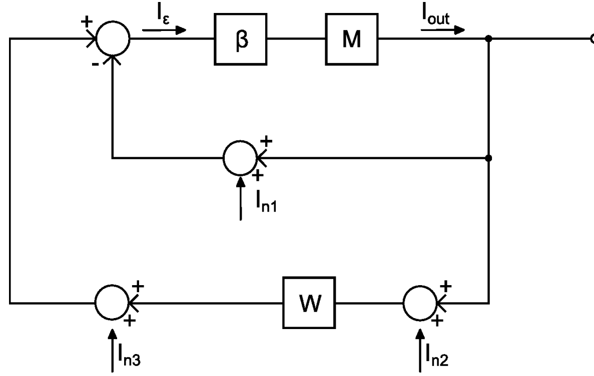


Figure 2.12: *PTAT* Loop block diagram.

W can be calculated from the small signal model[2] and we get

$$W \approx 3 \tag{2.20}$$

This value is stable in the whole temperature range. We are now able to evaluate all the transfer functions of interest. We set $A = \beta \cdot M$

I_{n1} transfer function

We can write, accordingly to the block diagram

$$[-(I_{out} + I_{n1}) + W \cdot I_{out}] \cdot A = I_{out} \tag{2.21}$$

[2]see appendix A.2.2

from which we get

$$\frac{I_{out}}{I_{n1}} = 0.497 \approx 0.5 \tag{2.22}$$

I_{n2} transfer function

For this current, we have that

$$[(I_{out} + I_{n2}) \cdot W - I_{out}] \cdot A = I_{out} \tag{2.23}$$

from which we get

$$\frac{I_{out}}{I_{n1}} = 1.497 \approx 1.5 \tag{2.24}$$

I_{n3} transfer function

For the conventions used for this loop analysis, the current I_{n3} has the opposite sign of the verse indicated by the arrow in Fig. 2.12. So we get

$$[-I_{out} + I_{n3} + W \cdot I_{out}] \cdot A = I_{out} \tag{2.25}$$

$$\frac{I_{out}}{I_{n3}} = \frac{1}{1 - W} \approx -0.5 \tag{2.26}$$

where $K = \frac{A}{A-1} \approx 1$. This result was predictable due to the fact that, in the circuit, this current is the same as I_{n1} even though in the block diagram it is taken from another branch. This is due to the fact that in the circuit these two currents are being taken from the same metal, so are practically the same, while in the block diagram they are placed under or above the node where the error current is generated, hence the different sign convention. Now that we know how the loop reacts, we can go move further for more specific error analysis.

2.4.3 Base Currents

We aimed to prevent the error due to the base-currents of T0, T1 and T52. We choose an NPN transistor over a n-channel MOSFET so to sink a base-current from the branch where it is connected. We dimensioned its area and

collector-current so that it would sink the same current sunk from the bases of T0, T1 and T52. This way, in first approximation, being T0, T1 and T52 driving all the same I_C and having very similar bias, they have the same base-current, so we can write

$$\begin{cases} I_{M0} = I_{C0} + 3I_B & (a) \\ I_{M1} = I_{C1} + 3\tilde{I_B} & (b) \end{cases} \tag{2.27}$$

Subtracting (2.27)(b) to (2.27)(a) we get

$$I_{C0} - I_{C1} = I_\epsilon = I_{M0} - I_{M1} \tag{2.28}$$

This way, we can reduce the error in the V_{PTAT} generation that would be caused by a difference in the two collector currents I_{C0} and I_{C1}. The drawback is that the output current carries also those $3I_B$ to the output, requiring an additional correction for their compensation

$$V_{ref}(T) = V_{ref}^{'} + I_{B,tot}(T)R_{load}(T) \tag{2.29}$$

where $I_{B,tot}$ is the total amount of current that the PTAT and NL stage cause to flow on the output load. This compensation involves also the resistor TC of the load, so we will model the

$$V_{ref,B} = I_{B,tot}(T)R_{load}(T) \tag{2.30}$$

term for a correct compensation. Exporting the data for (2.30), and fitting it with a first and a second order polynomial, we get two expression

$$V_{ref,B(a)} = -30\mu T + 0,0318 \tag{2.31}$$

$$V_{ref,B(b)} = 80nT^2 - 800\mu T + 0,0319 \tag{2.32}$$

The first order expansion has been calculated so to minimize the error on the whole temperature range instead of having a precise fitting for T_r and will be our first try into the compensation process.

2.4.4 Mirror Channel Length Modulation and NPN Early Voltage

The current mirrored from M0 and M1 can suffer of a slight channel length modulation. They share the same overdrive, but $V_{DS,M1} = V_{CC} - V_{CE,T1}$ while $V_{DS,M0} = V_{CC} - V_{BE,T0} - V_{PTAT}$. This difference is less than $6mV$ and, for the transistor length chosen of $25\mu m$ and the bias, the error current is negligible (magnitude of pico-ampere).

NPN Early voltage has more severe consequences on the PTAT stage. The different V_{CE} for T0 and T1 can cause an error on the PTAT voltage over R55. We can write eq. (1.8) taking into account the Early voltage; we get

$$V_{BE}(T) = V_T \ln \left[\frac{I_C}{A_e J_s \left(1 + \frac{V_{CE}}{V_A}\right)} \right] \tag{2.33}$$

where V_A is the Early voltage. We can now write the I_{PTAT} relation using (2.33) and it yields

$$I_{PTAT} = \frac{V_T}{R_{PTAT}} \ln \left[K \cdot \frac{I_{C1}\left(1 + \frac{V_{CE0}}{V_A}\right)}{I_{C0}\left(1 + \frac{V_{CE1}}{V_A}\right)} \right] \tag{2.34}$$

where K is the area ratio. It is easy to write I_{PTAT} as a nominal current plus an error current using the product property of logarithmic:

$$\begin{aligned} I_{PTAT} &= \frac{V_T}{R_{PTAT}} \ln \left(K \cdot \frac{I_{C1}}{I_{C0}} \right) + \frac{V_T}{R_{PTAT}} \ln \left[\frac{\left(1 + \frac{V_{CE0}}{V_A}\right)}{\left(1 + \frac{V_{CE1}}{V_A}\right)} \right] && (2.35) \\ &= I'_{PTAT} + I_{PTAT,error} && (2.36) \end{aligned}$$

Using the Early voltage for our technology we get a total error on Vref

$$\begin{aligned} V_{ref-error} &= I_{PTAT,error} \cdot R_{load} && (2.37) \\ &= \frac{R_{load}}{R_{PTAT}} \cdot \frac{kT}{q} \cdot \ln(0.9998) && (2.38) \end{aligned}$$

This error is about $-140\frac{nV}{°C}$, so we would have a total error over the whole temperature range of just $-30\mu V$ and normally does not need any compensation. However, simulation results in Fig. 2.13 show an unexpected behavior for V_{BG}, the output voltage of our circuit, when hot temperatures, above 100°C, are reached. This figure shows also the error on the V_{PTAT} voltage. Fig. 2.14 shows the three topologies which realize the three simulation results of Fig. 2.13.

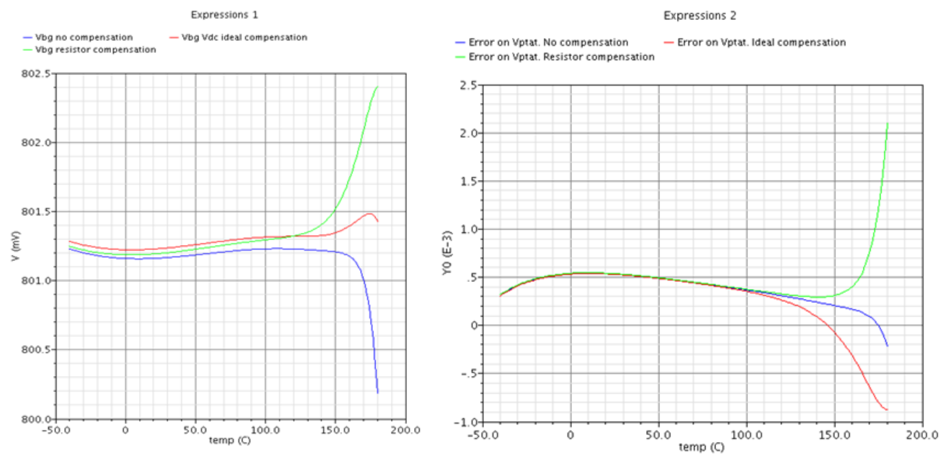


Figure 2.13: Early voltage and saturation region approaching effects.

The ideal compensation consists in adding a V_{DC} voltage generator which supplies the difference of the two collector-emitter voltages. The resistor compensation consists in placing a resistor which causes a voltage drop to reduce the collector-emitter voltage difference. These, beyond the fact that only one is physically feasible, differ because the first one rises the voltage collector of T0, the other one lowers the voltage collector of T1. We can see that the error from a non-compensated situation, for temperature below 100°C, is negligible both with the resistor and ideal compensation. Above 100°C, we can see the effect shown in Fig. 2.13. In order to make proper evaluations, we also keep track of the ratio $\frac{I_{C1}}{I_{C0}}$ shown in Fig. 2.15, the other source of error in the PTAT voltage generation. Let us examine the case "no compensation VS ideal compensation": we have a shift over all

the temperature range, but we see no changing until 100°C. Moreover, with the "ideal compensation", I_{C0} becomes greater than I_{C1} for $T \approx 145$°C and should cause the error $V_{PTAT,error}$ to became negative; the non compensated topology has a much more linear error behavior. However, the error on V_{BG} has the opposite behavior, but why? The "resistor compensation" case shows more coherency between the $V_{PTAT,error}$ and V_{BG}. The explanation is that there is the superposition of two effects: the $V_{PTAT,error}$ and the leakage-currents error. The "operative point" hypothesis does not give a full

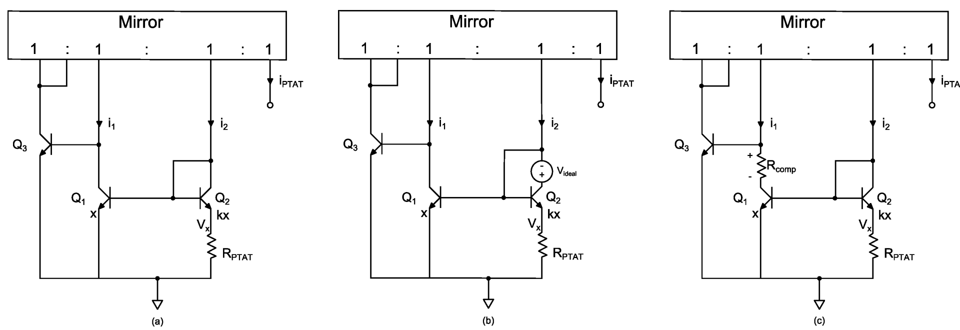


Figure 2.14: Early voltage test topology.

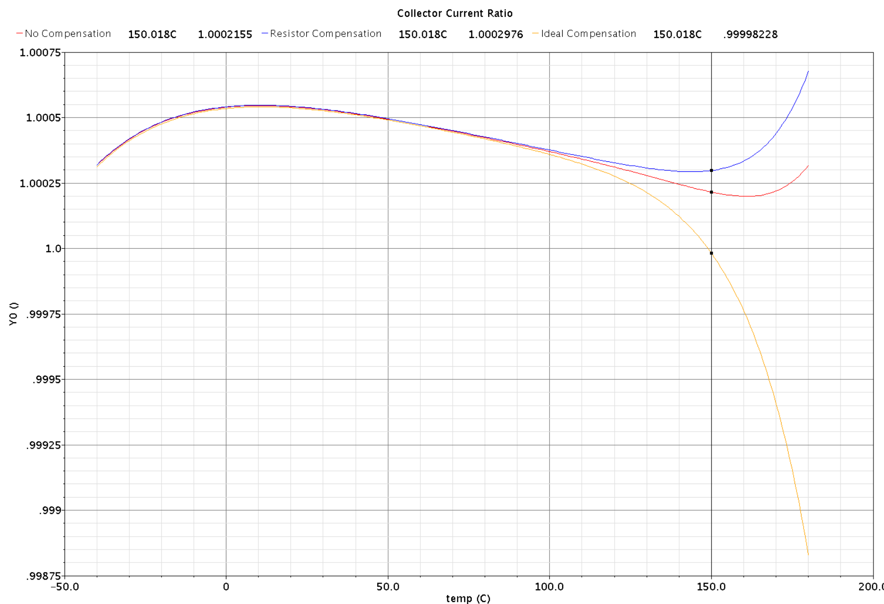
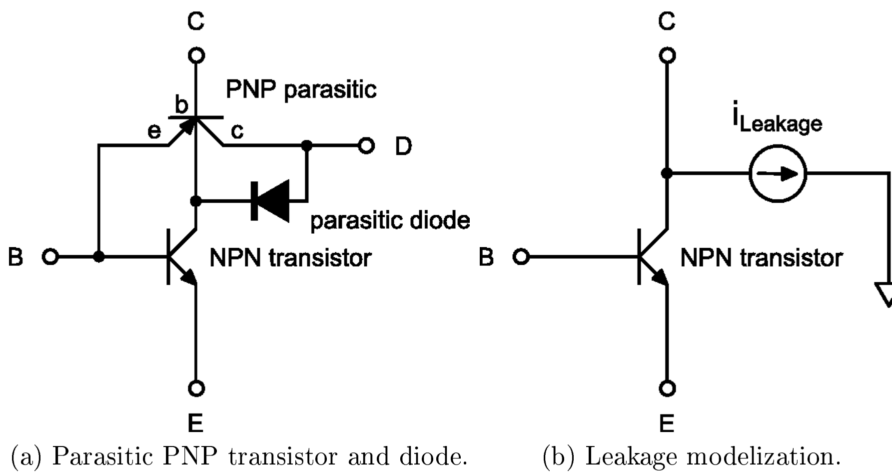


Figure 2.15: Collector-currents ratio.

explanation to these phenomena and it is not explainable through precise formulas. We will see that the other source of error is the leakage current of the parasitic PNP transistor and the parasitic diode. A deep insight in the leakage current phenomenon is then required. Then, in Sec. 2.4.5, a solution is formulated.

2.4.5 Leakage

With leakage current we mean the current in a reverse-biased pn junction. This current is usually negligible for our bias range and for temperature under 100°C, but for $T > 100°C$, it increases exponentially from less than 1pA to tens of nano-Ampere. This current is usually taken not as function of the reverse bias, and even though it has little sensitivity to reverse bias changes, measurements on the proposed sub-bandgap circuit showed that the leakage compensation performed worked fine for just one precise voltage[3]. Fig. 2.16a shows the main parasitic junction that cause current leakage. The D terminal is connected to a substrate ring that is connected to ground.



(a) Parasitic PNP transistor and diode. (b) Leakage modelization.

Figure 2.16: Leakage parasitic devices and model.

Those junctions, once leaking, drain more current from the real collector of the device, altering the total amount of current available for the main

[3]see chapter 3.

device. In the PTAT stage, leakage currents not only cause an exponential offset current (with temperature) that will be mirrored at the output load with a transfer function determined by where they are injected in, but they also cause an error in the $I_C - V_{BE}$ characteristic, as can be seen in eq. (2.35) when the I_C ratio becomes temperature dependent. Let us examine the physics behind this reverse current.

Leakage Physics[4]

In a pn junction, if a field is present, we will have both diffusion and drift current. Electron current density will be

$$J_{nx} = qn\mu_n E_x + qD_n \frac{\partial n}{\partial x} \tag{2.39}$$

where $D_n = \frac{kT}{q}\mu_n$ (as in eq. (1.9), but here the electrons mobility μ_n has not been taken as "average") and n are the free carrier (electrons). These relation is valid also for the hole diffusion current

$$J_{px} = qp\mu_p E_x - qD_p \frac{\partial p}{\partial x} \tag{2.40}$$

where the negative sign arises because of the positive charge of a hole. At thermal equilibrium, the total current (so also the current density) for both electrons and holes is zero, so we can write

$$qn\mu_n E_x + qD_n \frac{\partial n}{\partial x} = 0 \tag{2.41}$$

$$qp\mu_p E_x - qD_p \frac{\partial p}{\partial x} = 0 \tag{2.42}$$

from which we get (for the electrons)

$$n(x)\frac{\partial V(x)}{\partial x} = \frac{D_n}{\mu_n}\frac{\partial n(x)}{\partial x} \tag{2.43}$$

[4] For a complete physical analysis, please refer to [9]

this equation yield to the integer relation

$$\int_{V1}^{V2} \partial V(x) = \int_{n1}^{n2} \frac{1}{n(x)} \partial n(x) \tag{2.44}$$

so we can finally write that

$$V2 - V1 = V_T \ln \frac{n_2}{n_1} \tag{2.45}$$

and, dually for holes

$$V2 - V1 = V_T \ln \frac{p_1}{p_2} \tag{2.46}$$

These relation remain valid also for biased pn junction if the low-injection hypothesis is respected. In this case, noticing that

- holes concentration are equal to N_A in the p-doped region and $p_2 = p_n(x_n)$ in the n-doped region
- electrons concentration are equal to N_D in the n-doped region and $n_1 = n_p(-x_p)$ in the p-doped region

and supposing to have a reverse bias $V_A < 0$ we can write (i.e for holes) that

$$V_0 - V_A = V_T \ln\left(\frac{N_A}{p_2}\right) \tag{2.47}$$

from which we can find the holes concentration as

$$p_2 = p_n(x_n) = N_A e^{\left(\frac{V_A - V_0}{V_T}\right)} = p_{n0} e^{\left(\frac{V_A}{V_T}\right)} \tag{2.48}$$

where $p_{n0} = N_A e^{\frac{-V_0}{V_T}}$, that is when no bias is applied to the junction. For electrons, we get

$$n_1 = n_p(-x_p) = N_D e^{\left(\frac{V_A - V_0}{V_T}\right)} = n_{p0} e^{\left(\frac{V_A}{V_T}\right)} \tag{2.49}$$

Defining

$$p_n^{'}(x) = p_n(x) - p_{n0} \tag{2.50}$$

For $x > x_n$, $p_n^{'}(x)$ can be obtained from continuity equation and we get

$$p_n^{'}(x) = p_{n0}\left(e^{\frac{V_A}{V_T}} - 1\right)e^{-\frac{x-x_n}{L_P}} \tag{2.51}$$

where $L_P = \sqrt{D_P \tau_P}$ is the diffusion length and τ_p is the hole-lifetime. In our case, our NPN transistor has a highly doped collector diffusion, and can be demonstrated that in a reverse bias situation, if $N_D \gg N_A$, the drift current is mainly given by electrons minority carriers, so we get that

$$J_{r,drift} = q\sqrt{\frac{D_n}{\tau_n}}n_{p0}\left(1 - e^{\frac{V_A}{V_T}}\right) = q\sqrt{\frac{D_n}{\tau_n}}\frac{n_i^2}{N_A}\left(1 - e^{\frac{V_A}{V_T}}\right) \tag{2.52}$$

This current density is given by the minority carriers who are injected from a quasi-neutral region into the depletion region, but we must take into account also the current due to electron-hole pair generation in the depletion region.

This current is given by

$$J_{r,gen} = \frac{qn_iW}{\tau_e} \tag{2.53}$$

where τ_e is the effective life time of the carriers a W is the depletion-layer width. τ_e varies slowly with temperature and will be considered constant in the final evaluation. W is function of the applied field and it varies as

$$W\left(V_a\right) = W_{d0}\cdot\sqrt{1 - \frac{V_A}{V_0}} \tag{2.54}$$

where W_{d0} is the depletion width when no field is applied and V_0 is the built-it potential $\left|\frac{\phi_i}{q}\right|$. We can now write a proper formula for J_r:

$$J_r = q\sqrt{\frac{D_n}{\tau_n}}\frac{n_i^2}{N_A}\left(1 - e^{\frac{V_A}{V_T}}\right) + \frac{qn_i}{\tau_e}W_{d0}\cdot\sqrt{1 - \frac{V_A}{V_0}} \tag{2.55}$$

This formula shows that the reverse current

$$I_r = A_{eff}\cdot J_r \tag{2.56}$$

is function both of the temperature and voltage. Despite it has not been possible to evaluate $J_{r,drift}$ and $J_{r,gen}$ contribution for our technology, some important observation can still be done. Supposing a typical value for V_0 of 0.7V, if the drift and generation contribution are of the same magnitude we can see that:

- Defining $V_R = |V_A|$, if $V_r > 3V_T$, $J_{r,drift}$ does not change significantly with V_A: in our case we have at least about 300mV at T=180°C for the lower biased NPN transistor, so $V_{reverse} \approx 10V_T$. The change in this bias range is about 50ppm from 150 to 180°C. This error is negligible.
- Generation current shows that the reverse bias can affect the total leakage current, especially if the applied voltage become severe. This will explain also some behavior that will be shown and analyzed in chapter 3. Considering $\frac{qn_i}{\tau_e}W_{d0}$ as not being function of the reverse voltage applied, the term $\sqrt{1-\frac{V_A}{V_0}}$ shows great influence by the applied bias.

This last consideration impacts over two different situations:

- On the NPN transistors: with the decreasing of collector voltage from 150 to 180°C from about 360mV to 300mV for all the NPN transistor in the PTAT stage (neglecting the feedback transistor), $J_{r,gen}$ decrease. However, the changes can be easily calculated using eq. (2.54) and the total change is about -4%. This error is negligible, and is also reduce by the leakage compensation design. However, the leakage compensation done at the output[5] will suffer this effects, because the output voltage that is reverse-biasing the leakage compensation stage is fixed at about 800mV while the leaking junctions are biased from 360mV to 300mV, causing a $J_{r,gen}$ overcompensation of about 15%.
- On the p-channel MOSFETs: increasing the supply voltage in its range of operation (i.e. 1.5V to 5V), using eq.(2.54) we get a $J_{r,gen}$ leakage increase of 70% for the p-channel MOSFETs in the PTAT stage and

[5]Explained in sec. 2.5

90% increase for the p-channel MOSFETs in the output stage! This error could be symmetrical in the PTAT stage and so arises negligible errors, but it adds a voltage dependent current offset at the output. If the bandgap reference is designed to work with different supply voltage, then a compensation must be designed. However, if the bandgap reference is supplied by a pre-regulator or pseudo-supply, the need for a leakage-mirror correction is not necessary: leakage currents injected by p-channel MOSFETs to the load can be taken into account when choosing what counter-measures to take for the whole leakage phenomenon.

In Sec. 2.5 we will analyze some techniques in order to compensate these effects. Anyway, it is now possible to give an exhaustive explanation to the final issue find in Sub-sec. 2.4.4 on page 49. Using the loop transfer functions and applying them to the leakage currents, we get (for T=180°C)

Compensation	$I_{n1} = \|-I_{n3}\|$	I_{n2}	$I_{out} = 1.5I_{n2} - 0.5I_{n1}$
None	23nA	22.5nA	22.25nA
Ideal (-60mV)	23nA	23.3nA	23.45nA
Resistor	21.5nA	22.5nA	23nA

Table 2.2: Leakage explanation to V_{ref} variation discussed in 2.4.4.

Table 2.2 shows that we have a variation on the leakage that is transferred to the output load. This effect must be added to the $V_{PTAT,error}$ of each single case, shown in Fig. 2.13 on page 50. For 180°C, the error $V_{PTAT,error}$, calculated as differential from the non compensation version, is:

$$\begin{cases} \frac{kT}{q}\frac{R_{load}}{R_{PTAT}} \cdot \Delta V_{PTAT,res}\Big|_{T=(273+180)K} \approx 600\mu V & \text{for res. compensation} \\ \frac{kT}{q}\frac{R_{load}}{R_{PTAT}} \cdot \Delta V_{PTAT,ideal}\Big|_{T=(273+180)K} \approx -180\mu V & \text{for ideal compensation} \end{cases} \quad (2.57)$$

So the total error on the output stage, calculated as differential from the non compensated case, is

$$\begin{cases} 600\mu V + (23 - 22.25)n \cdot R_{load}|_{T=180°\mathrm{C}} = 850\mu V & \text{for res. compensation} \\ -180\mu V + (23.45 - 22.25)n \cdot R_{load}|_{T=180°\mathrm{C}} = 400\mu V & \text{for ideal compensation} \end{cases} \tag{2.58}$$

These values does still not match with those reported in Fig. 2.13 on page 50. This is because the other stages of the circuit depend on the PTAT reference generated in this early stage, and they generate other errors correlated to the errors of PTAT stage. However, it is already intuitive that if $\frac{V_{PTAT}}{R_{PTAT}}$ will be used for biasing other components (like the NPN who generates the CTAT voltage reference), the error on the total output current of the PTAT stage will cause an error of the same sign on the $V_{BE}(T)$ generated and this will add its contribute to the total error seen on V_{BG}, hence supplying the difference between what we calculated and the simulation.

2.5 Diode Loop Topology Design

Now that we are aware of all the snares hidden in this topology, we can begin sizing the components and implementing solutions for the main problems seen in Sec. 2.4. Three version of this topology have been developed. Version 1 is the default version, version 2 has a more accurate leakage compensation and different test structures and version 3 is leaned towards area optimization.

2.5.1 The PTAT Stage

In Sec. 2.3 we began to build our PTAT stage. We sized the PTAT resistor, $R55 = 60k\Omega$ and the area ration between T0 and T1 so to be $\frac{A_{e0}}{A_{e1}} = 10$. We recopy again the PTAT stage for convenience in Fig. 2.17 on the next page.

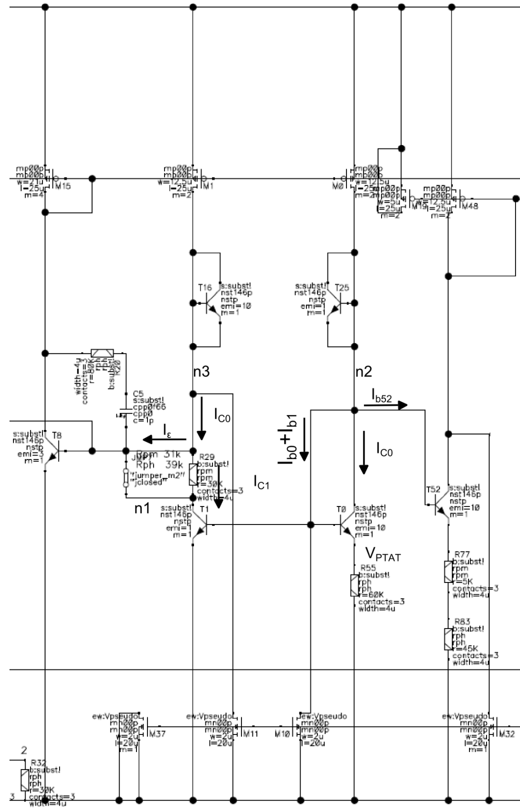


Figure 2.17: The *PTAT* Stage.

We have chosen an NPN BJT as feedback transistor so to sink from n3 approximately the same base currents that are sunk from n2. To make $I_{B,T8} \approx I_{B,T0} + I_{B,T1} + I_{B,T52}$ we set $I_{C,M15} \approx 3I_{C,M1}$ (being $I_C = \beta I_B$ and being $\beta_{T8} \approx \beta_{T1,0,52}$) by setting $\left(\frac{W}{L}\right)_{M15} = 3\left(\frac{W}{L}\right)_{M1}$ and $A_{e,T8} = 3$. A first simulation run showed significantly base-current error reduction, but the perfect tuning is obtained for $\left(\frac{W}{L}\right)_{M15} = 3.36\left(\frac{W}{L}\right)_{M1}$. Fig. 2.18 shows three cases, numbered from zero to two: case zero is with just one base-current compensation, case one is for $\left(\frac{W}{L}\right)_{M15} = 3\left(\frac{W}{L}\right)_{M1}$ (first approximation base current) and case two is for $\left(\frac{W}{L}\right)_{M15} = 3.36\left(\frac{W}{L}\right)_{M1}$ (tuned base-current compensation).

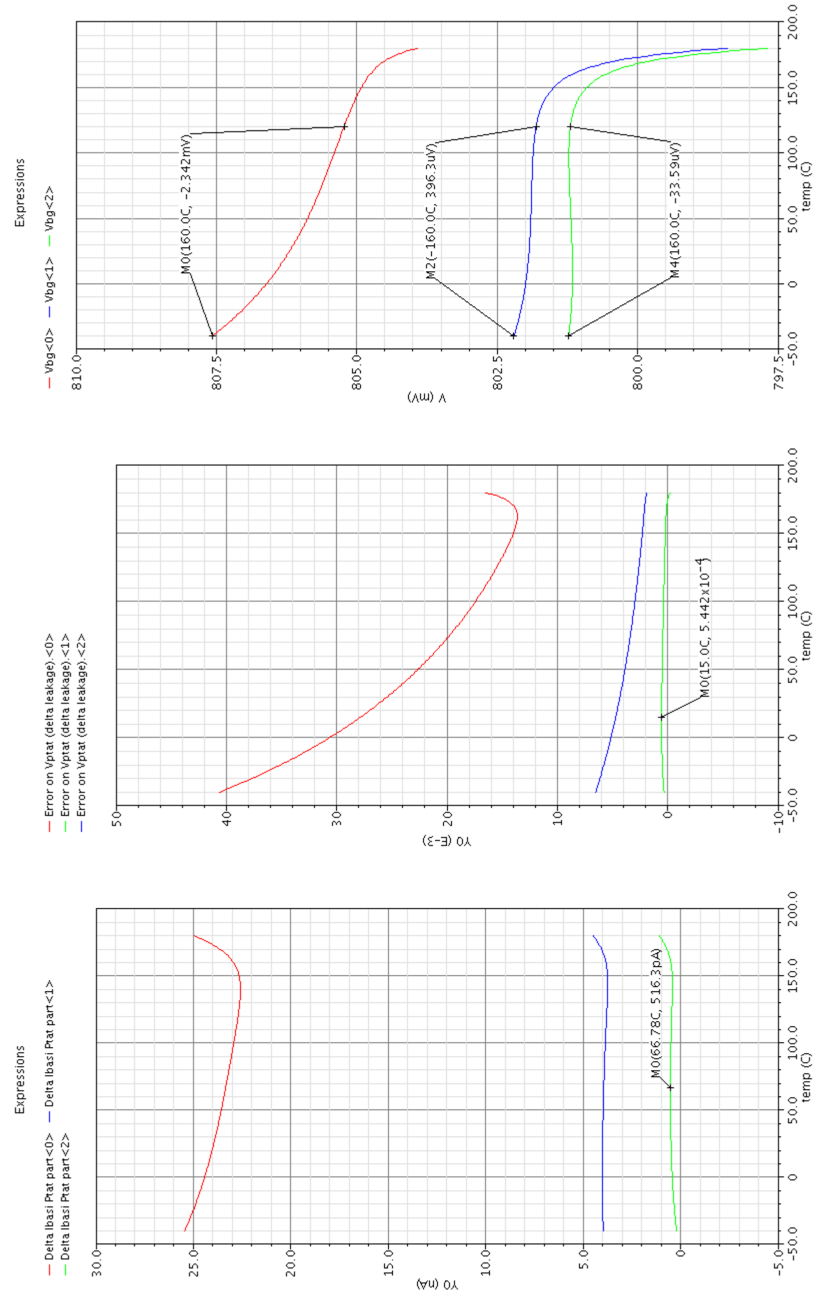


Figure 2.18: I_B compensation results.

For each case, we can see how the error propagates on the V_{PTAT} generation, changing the collector-current ratio of eq. (2.35), and finally afflicts the

output voltage. As we see from Fig. 2.18, this error would afflict the output voltage very hard if not compensated.

The collector current ratio is sensitive also to the leakage current mismatch of T0 and T1. As we have seen in eq. (2.56), leakage current is proportional to the effective area where J_r flows. T0 and T1 have very different area ratio so their leakage currents are different. This causes an error on the collector-current ratio in eq. (2.35). As we saw in subsec. 2.4.5, Fig. 2.16b, the effective collector current available for the transistor is reduced by the leakage current, so the bigger the transistor is, the more its collector current will drop. If we look at eq. (2.35), I_{C0} is at denominator, so, suffering more leaking, it will cause a positive V_{PTAT} error. This error will sum to the error due to leakage currents being mirrored to the output so it will not compensate itself, having both effects leaning towards the same direction. This error can be compensated in many ways, more or less accurate, but all of them will take advantages of the structure shown in Fig. 2.19

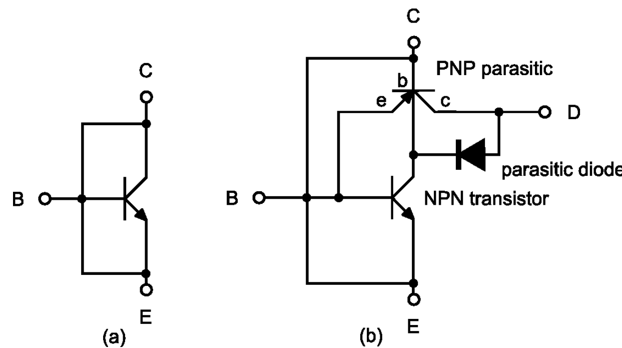


Figure 2.19: Leakage compensation NPN.

We will refer to this structure as "anti-leakage configuration". Short-circuiting all three terminals as shown in Fig. 2.19(a), we get a turned off transistor which has only the parasitic diode and the pn junction of the parasitic PNP enabled as can be seen in Fig. 2.19(b). We have got a structure that only produce leakage current and is turned-off for T<120°C. We can use this structure to sink leakage current from the net of interest. It is necessary to make an important observations before proceeding: the effective area for (2.56) is not strictly proportional to the emitter area used for the transistor:

table 2.3 will summarize the value obtained from the simulator for a reverse voltage $V_A = -800mV$ normalized for the leakage of an NPN with $A_e = 1$, which value is about $8.5\mu A$. This is due to the fact that the effective leaking

A_e	1	2	3	4	5	6	7	8	9	10
$\frac{I_{leak}(1)}{I_{leak}(A_e)}$	1	1.088	1.177	1.266	1.354	1.443	1.429	1.5	1.572	1.644

Table 2.3: Leakage values for $V_A = -800m$.

area depends also on the side area of the NPN. Multiplying the area does not increase the perimeter area as placing two space transistor having the same amount of area because, if these two transistor are put into contact, the leakage from the touching sides is deleted. We have developed three solutions.

1. The first and more precise solution is the one used in this version of the Diode Loop Sub-Bandgap[6]. It consists in adding two short-circuited NPN (T16 and T25 of Fig. 2.17) for each branch of the PTAT circuits that has the area of the transistor in the other branch. This way, each collector current sees the same leakage and the error is precisely compensated, as it come that

$$\begin{cases} I_{C0} = I_{M0} - I_{leak,1} - I_{leak,10} \\ I_{C1} = I_{M1} - I_{leak,1} - I_{leak,10} \end{cases} \tag{2.59}$$

where $I_{leak,x}$ is the leakage area for a single NPN transistor with $A_e = x$ and $I_{M0} = I_{M1}$. The downturn is that we add more leakage current that must then be sunk by the same structure at the output. The amount of leakage current can be calculated by exploiting the block diagram used for the loop study in Sub-sec. 2.4.2. We have that the total leakage current mirrored to the output from the PTAT stage is

$$I_{leak,OUT} = (I_{L,T0} + I_{L,T25})\,1.5 - (I_{L,T1} + I_{L,T16})\,0.5 \tag{2.60}$$

$$= I_{leak10} + I_{leak1} \tag{2.61}$$

[6] Version #2

Again, notice that $I_{leak10} + I_{leak1} \neq I_{leak11}$. This leakage current will than be sunk by exactly two NPN of A_e equals to 1 and 10, in anti-leakage configuration connected to the output net as shown in Fig. 2.25 on page 71. This way, the only error that occurs is due to the different reverse voltage biasing the leaking pn junction. This error is however very difficult to estimate and measurements will be necessary to see if result is acceptable or needs further corrections. These will apply to all the compensation that involves the anti-leakage NPN connected to the load.

2. The second solution is to place no compensation on the PTAT stage and just compensate the effect for T=180°C by the anti-leakage transistor at the load. This way, if the temperature is compensated for T=120°C, where leakage does not occur, and at T=180°C, where leakage is maximum, being $V_{BG}(T)$ continuous and leakage monotonic over temperature, we expect to see a quasi flat behavior from leakage contribution if this compensation should succeed. The V_{PTAT} error is shown in Fig. 2.20.

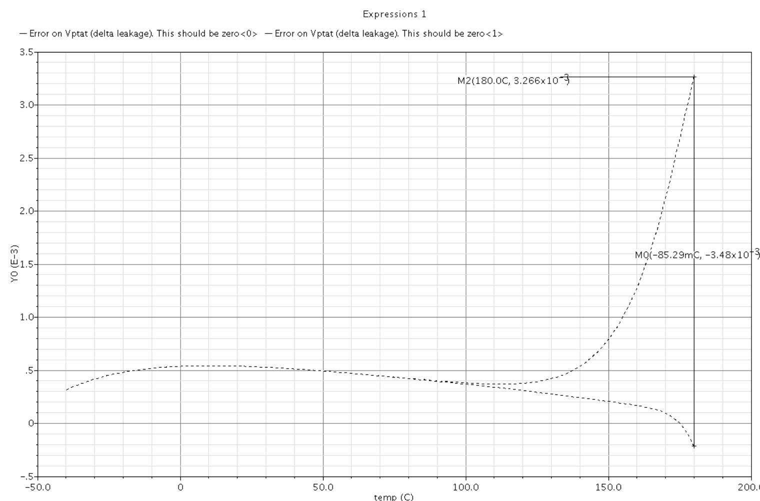


Figure 2.20: V_{PTAT} error caused by leakage currents.

For T=180°C, this error will cause an error ΔV_{BG}

$$\Delta V_{BG} = \frac{k}{q}(180 + 273.15) \cdot \epsilon_{180^{\circ}C} \cdot \frac{R_{load}}{R_{PTAT}} \approx 750\mu V \tag{2.62}$$

Being $R_{load} = 320k\Omega$ for T=180°C, we need a $\Delta I_{leak} = {}^{750\mu}/_{320k} \approx 2.35nA$. The contribution by leakage mirroring to output is

$$I_{lead,M} = 1.5 I_{L,T0} - 0.5 I_{L,T1} \tag{2.63}$$

$$= 1.5 L_{leak,10} - 0.5 I_{leak,1} \tag{2.64}$$

Using the values of table 2.3, we get

$$I_{lead,M} \approx 2 I_{leak,1} \tag{2.65}$$

The final configuration features two NPN in anti-leakage configuration with $A_e = 1$ and $A_e = 4$; the latter corresponds to the emitter area needed to obtain also the current contribution for the V_{PTAT} compensation[7]. The conceptual error behind this compensation is that we compensate also the V_{PTAT} error by sinking more leakage current. There is no connection between the error source and the compensation; this uncorrelation can bring to over or under-compensation if the model is not correct and if variations process occur. However, it saves a lot of area and with some trial on silicon, it may be tuned for a good performance.

3. The last version exploits the loop to reduce the leakage current mirrored to the output load without sacrificing the precision over the V_{PTAT} voltage. An anti-leakage NPN with $A_e = 1$ has been used only in the T1 branch, as shown in Fig. 2.21.

[7] It is the area that corresponds to a current ratio of $\frac{8.5+2.35}{8.5}$.

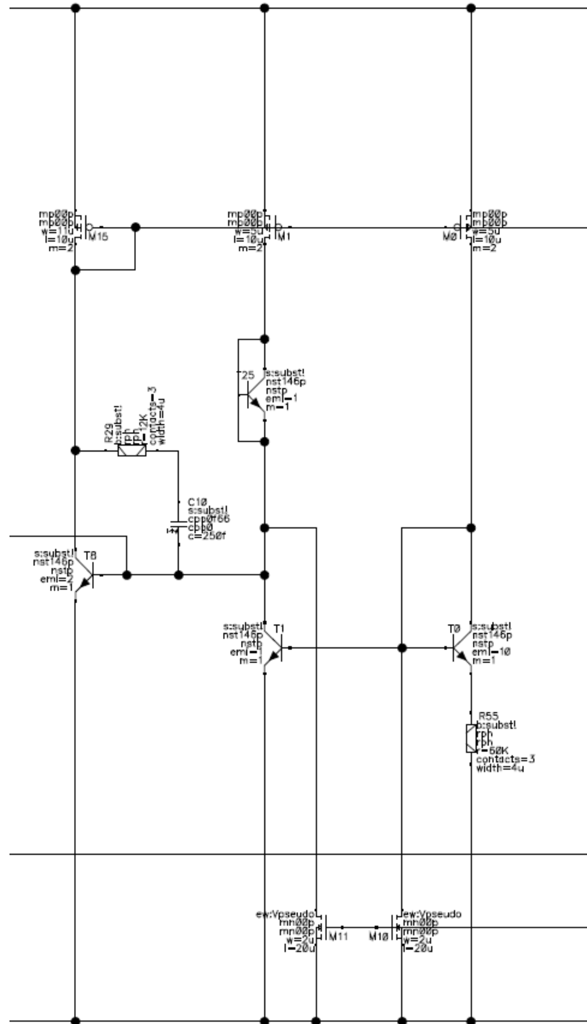


Figure 2.21: PTAT stage, version 3.

Using this configuration, we get

$$\begin{cases} I_{C0} = I_{M0} - I_{leak,10} \\ I_{C1} = I_{M1} - I_{leak,1} - I_{leak,1} = I_{M1} - \quad 1.2 I_{leak,10} \end{cases} \tag{2.66}$$

where leakage table 2.3 on page 62 has been used. This preserves a good precision for the V_{PTAT} generation, avoiding the need for a NPN transistor with $A_e = 10$ on T0 branch. The leakage current mirrored

at the output now becomes

$$I_{leak,M} = 1.5\,(I_{L,T0}) - 0.5\,(I_{L,T25} + I_{L,T1}) \quad (2.67)$$

$$= 1.5 I_{leak,10} - I_{leak,1} \approx I_{leak,6} \quad (2.68)$$

We decided to place an output anti-leakage NPN with $A_e = 3$ instead of 6, because we have tried to take into account also the difference in reverse bias. From measurement data, we estimated a correction factor for a reverse voltage $\Delta V_A = 500m$ of about 0.8. So we decided to apply this factor to $I_{leak,6} = 1.44 \cdot 0.8 \cdot I_{leak,1} = 1.152 I_{leak,1}$. This value correspond to an emitter area of 3. We have applied this leakage compensation on "version 3" because this version is made to obtain a better compromise between area and precision, so it is better suited.

The PTAT stage for version 1 and 2 presents also a second V_{PTAT} source, whose load is a series of two resistors with different TC. This is made by transistor T52 and R77 and R83. This transistor shares the same ΔV_{BE} configuration of the main PTAT stage to generate a V_{PTAT} reference, but being the load a mixed TC resistor series, the collector current will not follow the one imposed by the loop. This will cause two effects: the generation of a different I_{PTAT} and that, being $(\frac{\beta+1}{\beta})I_C$, an error on the V_{PTAT} itself. This behavior has been exploited to generate a current with the desired $\frac{\partial i(T)}{\partial T}$ with whom we feed the transistor responsible for the CTAT and non-linear voltage generation.

2.5.2 The CTAT and the Non-Linear Stage

If we were Asian-philosophy enthusiasts, we would say that PTAT and CTAT stages are the yin and the yang of this circuit. They are going to compensate themselves, being opposite in nature. So, once designed the stage that generates the simple, linear-behavior V_T reference, even with some tricky error compensation, now we are going to design the evil, non-linear part of the sub-bandgap reference, but that luckily has little need for error compensation. Leaving Asian philosophy for some more concrete math, let us examine

the CTAT stage in Fig. 2.22.

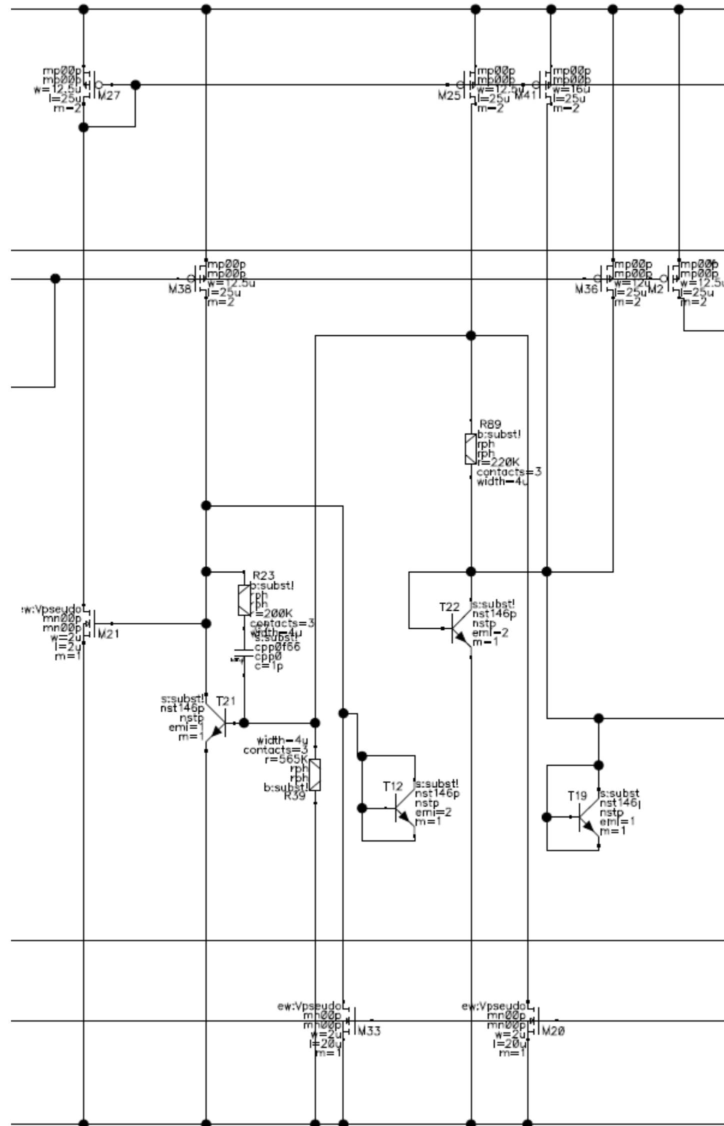


Figure 2.22: CTAT stage.

The first brick to design is the resistor translating the $V_{BE}(T)$ into current, R39. This resistor sets the slope for $I_{CTAT}(T)$. Let us recall eq. (2.69), that

we calculated at the end of Sec. 1.3.1:

$$V_{BE}(T) = V_{G0npn} - \left(\frac{T}{T_r}\right)[V_{G0npn} - V_{BE}(T_r) - \gamma T_r] - (\eta - x)\left(\frac{kT}{q}\right)\ln\left(\frac{T}{T_r}\right) \tag{2.69}$$

As we saw in subsec. 2.4.3 on page 47, there are also the base-currents from the PTAT stage that need to be compensated, and the same applies to the base current of T21. Supposing the base-currents linear in first approximation, we can write that the linear part of the output voltage is

$$V_{out,lin} = \frac{k}{q}\frac{R_{load}}{R_{PTAT}}T - \frac{T}{T_r}\frac{R_{load}}{R_{CTAT}}[V_{G0npn} - V_{BE}(T_r) - \gamma T_r] - \alpha T \tag{2.70}$$

where α is the slope for $\frac{\partial V_{out}}{\partial T}$ due to base currents only (taken from eq. (2.31)), $\gamma = 2.7325 \cdot 10^{-4}\,[V/°C]$ (from (1.26)) and $V_{G0npn} = 1.205$V. In order to have first order cancellation, we must have

$$R_{CTAT} = \frac{\left(\frac{k}{q}\frac{\ln(10)}{R_{PTAT}} - \frac{\alpha}{R_{load}}\right)}{T_r\left(V_{G0npn} - V_{BE}(T_r) - \gamma T_r\right)} \tag{2.71}$$

Using all these values, it yields that $R_{CTAT} = 545k\Omega$. However, once the design was ultimated, it resulted in a slightly overcompensated bandgap reference on the non-linear component. This meant that this value of R_{CTAT} was too small, causing a positive error on the linear CTAT slope[8] , that has been compensated by the non linear stage. A tuning process brought a $+3\%$ adjustment on the resistor value, that has been set to $R_{CTAT} = 565k\Omega$. The non-linear stage, while it may seems hard to design because of the unknown $(\eta - x)$ term and the recursive nature of I_{NL}, can be handled easily by a little trick. Let us recall eq. (2.13):

$$V_{NL}(T) = V_T \ln\left(\frac{2I_{PTAT}}{I_{NL} + I^*_{constant}}\right) \tag{2.72}$$

[8]Here it is meant in absolute value, that is, it was more negative.

In order to obtain a perfect cancellation we need that

$$\frac{(\eta - x)}{R_{CTAT}} \left(\frac{kT}{q}\right) \ln\left(\frac{T}{T_r}\right) = \frac{V_T}{R_{NL}} \ln\left(\frac{2I_{PTAT}}{I_{NL} + I^*_{PTAT} + I^*_{CTAT}}\right) \tag{2.73}$$

so all we need is to make the arguments of the two logarithms the same and than exploit the property of logarithm "ln(1) = 0" to check if the compensation has been done correctly. The first step then is to design the current ratio in eq. (2.72). Being a straight line with no intercept, they are equal if their derivative is the same, so we get

$$\frac{\partial\left(\frac{2I_{PTAT}}{I_{NL}+I^*_{PTAT}+I^*_{CTAT}}\right)}{\partial T} = \frac{1}{T_r} = 0.3^{m}/{}^{\circ}\mathrm{C} \tag{2.74}$$

In order to get a perfect flat slope, some simulation tuning on the mirrors supplying $I_{constant}$ was necessary. We wrote I^*_{PTAT} and I^*_{CTAT} to denote this tuning. In fig. 2.23 we can see the error that would be caused by keeping the default mirror ratio.

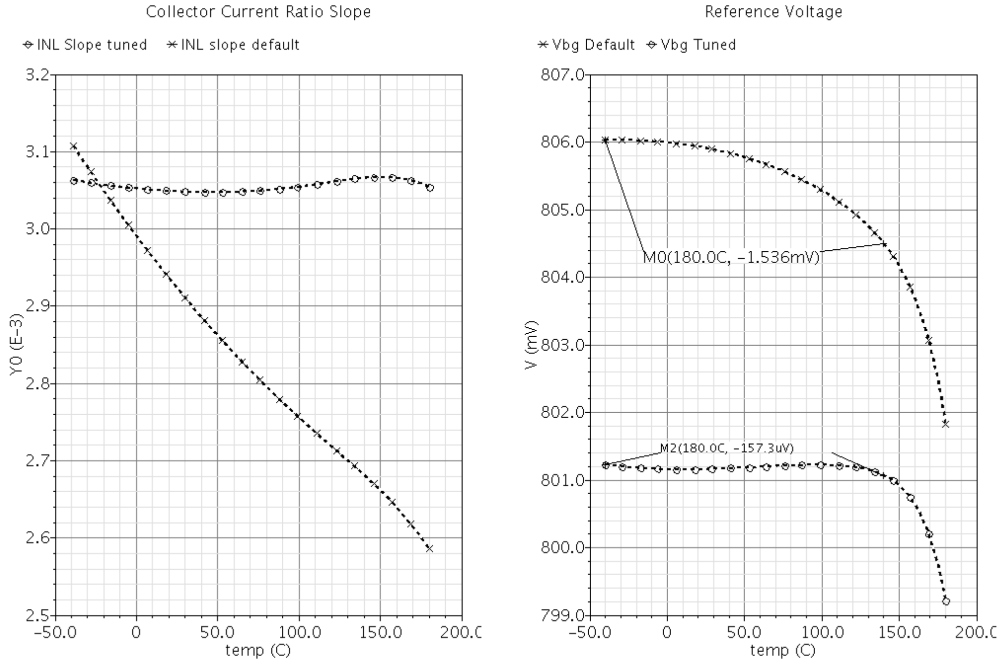


Figure 2.23: Tuned VS default mirror ratio.

The other problem is that not only the current ratio derivative must be constant, but also the ratio must be 1 for $T = T_r$. This can be checked simply monitoring the voltage across R_{NL}! In fact, when $2I_{PTAT} = I_{NL} + I^*_{PTAT} + I^*_{CTAT}$, $V_{NL} = 0$. Precisely tuning this configuration is a recursive task, because changing one of the currents also changes the total slope of the ratio. Once the ratio is designed, we can calculate R_{NL}

$$R_{NL} = \frac{R_{CTAT}}{(\eta - x)} = 226k\Omega \tag{2.75}$$

where we used $\eta = 3.5$ and $x = 1$. Being η and x uncertain, we can do a R_{NL} sweep in a $\pm 30\%$ range. This is used not only to find the correct value for the resistor, but also to check if the current ratio brings the right $\frac{1}{T_r}$ factor. In fact, we will see a rotation over the reference temperature set by the current ratio $\frac{2I_{PTAT}}{I_{NL}+I^*_{PTAT}+I^*_{CTAT}}$. Fig. 2.24 shows what has just been described.

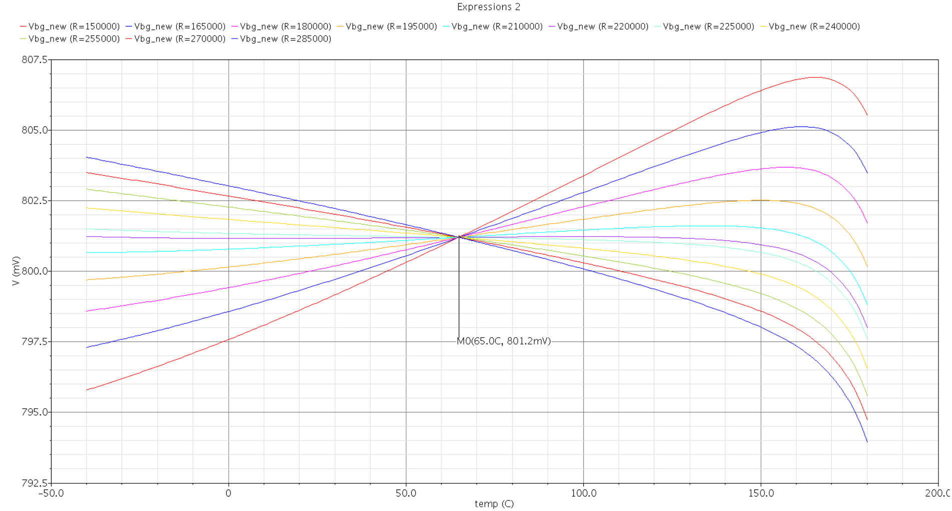


Figure 2.24: $R_{NL} \pm 30\%$ span. Balance over T_r effect.

It is easy to spot that we are centering the output voltage around $T_r = 65°C$, as we wanted, and the correct value for R_{NL}. The result shows no over compensations leaning towards first order or second order and the bandgap voltage has just an error of about a dozen of micro volt!

2.5.3 The Output Stage

The output stage is in current mode configuration and can been seen in Fig. 2.25. Over to the load mirrors and resistor, this stage presents many control structures useful to investigate as many variables of the circuit.

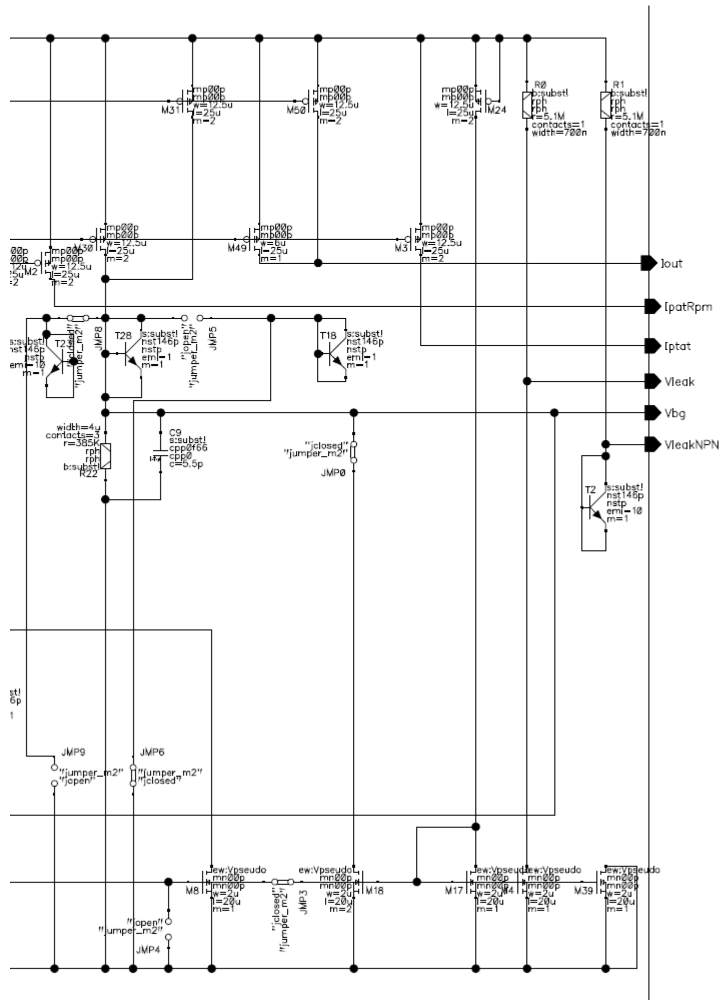


Figure 2.25: Output Stage, Version 2.

Transistors T23 and T28 are anti-leakage diode whose emitter-areas are designed accordingly to compensate the total leakage current mirrored to the output. However, measurements on the same compensation method suggest that the simulator has not precise models for temperatures over 150°C, so other transistors have been arranged so that they can be swapped or added to

those already present, for proper analysis and tuning. P-channel MOSFETs M49 and M58 have been sized to supply a constant output current. This is an optional part that as been added due to the simplicity of adding just another mirror structure; the result however is function of the "trans-conductance over temperature" characteristic, so it will be as precise as the model is. We also added a structure that aims at compensating the p-channel MOSFETs leakage. This structure consists in M24, with its $V_{SG} = 0$, so that the only current flowing to the drain of M17 is its leakage current. This current, that should be the same for all p-channel MOSFETs (there is the error due to the slightly different reverse bias of each node) is then sunk from all the vulnerable nodes of the circuit, or just from the output itself if jumper JMP3 is open (and JMP4 is closed to avoid floating gates). The problems of mirroring a leakage current are that the mirror structures will introduce other leakage and that the current involved is very small. If we recall the formula for a mos drain current, we have

$$I_D = \frac{1}{2}\mu C_{ox}\frac{W}{L}\left(V_{GS} - V_{th}\right)^2 \tag{2.76}$$

If I_D is very small, V_{GS} is very close to the threshold voltage, that is to say we have a very small overdrive. This makes the mirror very sensitive to the V_{th} mismatch. In order to reduce this sensitivity, we make a $\frac{W}{L} = \frac{2\mu}{20\mu}$ to considerably raise the overdrive. However, the risk that this structure fails is considered and JMP0 and JMP3 can completely separate this structure from the output stage. A copy of this current is sunk by resistor R0, so that we can read the voltage drop it causes and evaluate its magnitude and variation with reverse bias and temperature. The same thing is done with the NPN leakage current by T2 and R1. There are also two control structures for the two PTAT current generated in the PTAT stage.

2.5.4 The Start-up

The PTAT stage, as it has been designed, has also a zero-current operating point that must be avoided. A start-up circuit that will ensure the correct

bias point is then required. The start-up implementation is shown in Fig. 2.26. The idea is that the reference voltage V_{BG} is sensed: if it is low or zero, the bandgap reference is not working, if it is 800mV, it is on and fully working.

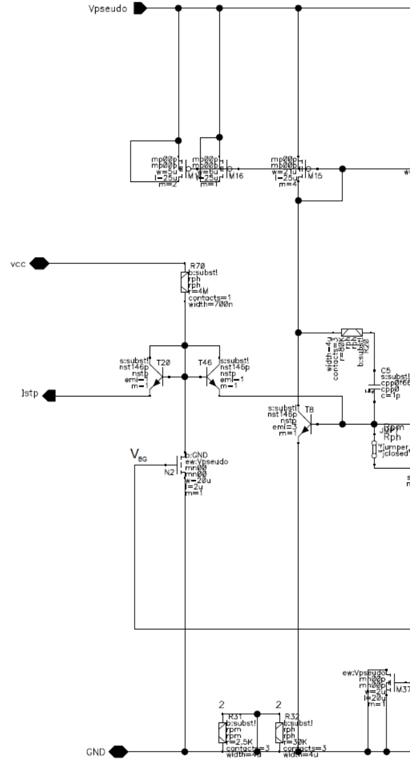


Figure 2.26: Start-up stage.

The device performing the voltage-sense must be a mos, because any current sunk from the output will cause an error. Moreover, the output voltage is set to be 800mV, that is close to $V_{th} \approx 0.7V$ of a standard low voltage n-channel MOSFET. We than choose to use a digital n-channel MOSFET, who differs from the low-voltage version also for a lower $V_{th} \approx 0.5V$. Opposite to the choice taken for the n-channel MOSFET used for the p-channel-MOSFET-leakage-current compensation, we want this transistor to need little overdrive to turn it on, so we dimensioned it for high g_m with a $\frac{W}{L} = \frac{20\mu}{2\mu}$ ratio. It follows that, when the circuit is off, N2 is sinking no current , there is then no drop across R70 so $V_{BE,T46} = vcc - v_x$, (where v_x is an unknown

value for $V_{BE,T8}$, although a transient simulation shows $V_{BE,T8} \approx 250mV$), so T46 is turned on, injecting current in the error sense node of the loop, that reacts to compensate the error by lowering $V_{G,mirror}$ so to increase M1 drain-current. This process starts up the circuit. Once the circuit is turned on, $V_{BG} \approx 800mV$ causes N2 to sink current from R70. This will turn off T46 reducing its base voltage while its emitter has risen to the proper bias point. We aim to have the lower possible $V_{B,T46}$ voltage (milli-volts) so that T46 has a negative V_{BE} and is securely turned off. This means that N2 is working in deep triode region. Calculating the "pull-up" resistor to have $I_{C,N2} < 1\mu A$, we set R70=4M. We have duplicated this start-up signal also for the pseudo-supply we have designed to meet the PSRR specification; this topic will be developed in subsec.s 2.5.5 and 2.5.6.

2.5.5 PSRR

The overall objective of designing a precision reference is to achieve high accuracy over all working conditions. The term *line regulation* indicates the deviation of the reference voltage caused by a variation of the supply. To be more precise, it indicates the steady state voltage changes in the reference resulting from DC changes in the supply voltage. The supply voltage can be also source of transient noise, depending on the environment where the circuit lies. This noise can be time variant, or can have precise frequency range where the interference is more severe. This is what is really meant by PSRR. However, *line regulation* and PSRR are connected because we can say that

$$\lim_{f \to 0} PSRR(f)_{V_{DC}=K} = \text{line regulation}\big|_{V_{DC}=K} \tag{2.77}$$

This formula means that PSRR tends to the line regulation value for $f \to 0$, but it also highlights that this is valid only for the DC bias that we are considering. The pivoting point for improving the PSRR and line regulation is to decrease the sensibility of sensitive nodes (the reference voltage in particular) to the input power supply voltage. This concept may be described assuming a voltage divider from the power supply to the reference voltage node. Let us take the current-mode output stage of Fig. 2.25. This stage

can be schematized by the "reference" part of Fig. 2.27.

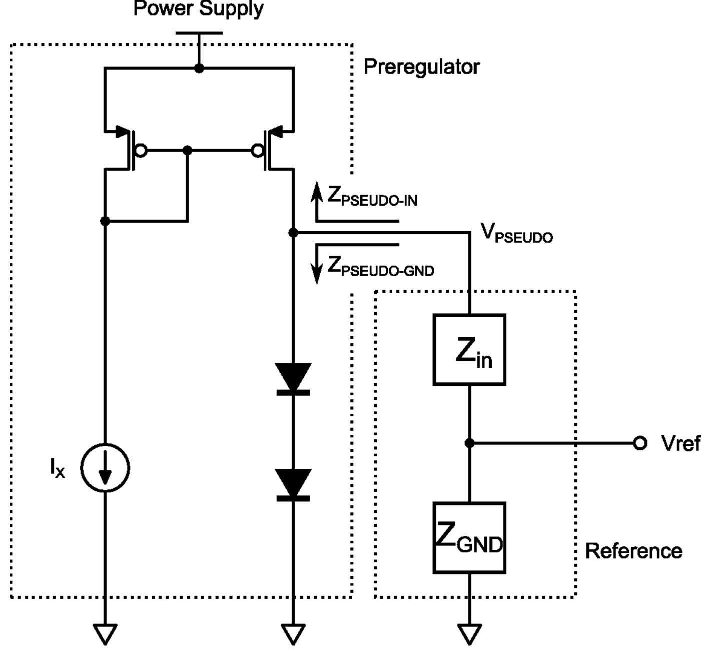


Figure 2.27: Pre-regulated pseudo-supply.

The transfer function between V_{ref} and V_{pseudo} is just the voltage divider

$$PSRR = \frac{\Delta V_{ref}}{\Delta V_{in}} = \frac{Z_{GND}}{Z_{GND} + Z_{IN}} \tag{2.78}$$

In order to get a high[9] PSRR, we need to design little Z_{GND} and high Z_{IN}. In the current mode output stage, Z_{GND} is the load and is set to be $385k\Omega$ in our case. Z_{IN} is $r_{0,P}$, the parallel of the output resistances for all the p-channel MOSFETs whose drains are connected to the load. This resistance is

$$r_0 = \frac{\eta L}{I_D} \tag{2.79}$$

where η is a technological parameter. So, being the current imposed to get the flat behavior of the reference voltage over temperature, the only variable we can play with is the p-channel MOSFET length. This is the limit to the

[9] PSRR is usually expressed in dB and, being an attenuation is negative: With "high PSRR", we are considering the absolute value.

PSRR for a current mode bandgap reference. In our case, with $L = 25\mu m$, we have $r_0 = 225M\Omega$ and we have the parallel of two p-channel MOSFETs, so $r_{o,P} \approx 140M\Omega$ and $R_{load} = 385k\Omega$. For these values, we get

$$PSRR_{DC} = 20\log\left(\frac{385k}{(140M + 385k)}\right) = -51dB \tag{2.80}$$

This is a limit that cannot be overtaken without a pre-regulator as shown in Fig. 2.27. The noisy and variable supply voltage is pre-regulated and essentially isolated from the reference. Using a pre-regulator, we can write

$$PSRR = \frac{Z_{PSEUDO-GND}}{Z_{PSEUDO-GND} + Z_{PSEUDO-IN}} \cdot \frac{Z_{GND}}{Z_{GND} + Z_{IN}} \tag{2.81}$$

This way, we can get another $-50dB$[10] from the real supply to V_{PSEUDO}. A particular attention must however be reserved for the bandgap drive. Lectures as [7] or articles like [10] suggest using diode connected transistor due to their low impedance in this configuration. This arise two problems:

1. Even supposing I_X of Fig. 2.27 being PTAT, the voltage generated from two diodes supposing their collector current being less than $5\mu A$ will be less than $1V$ for $T \approx 80$°C and it will fall down to 300mV for $T = 180$°C, so even stacking three diodes we will leave not enough headroom for $T > 120$°C. Diodes and diode-connected BJT are not suited for our automotive environment[11].

2. If using diodes or diode-connected BJT to create the voltage power supply for the bandgap, the total V_{PSEUDO} variation can go above 1V. A voltage reference with a normal PSRR of -40dB (100 times attenuation) will show 10mV of total temperature curvature due only to dependence of V_{PSEUDO} from temperature. A temperature compensated pre-regulator would be better than just a normal one.

[10] Speaking of DC analysis

[11] Temperature range for automotive products is [-40, 150]°C.

2.5.6 Pseudo-Supply Design

The basic ideas behind this pre-regulator are

- Driving the sub-bandgap reference core circuit by current, supplying the current needed and reserving a surplus for biasing the $Z_{PSEUDO-GND}$ impedance and for spare: components mismatch can lower the total current that the pre-regulator could supply and rise the sub-bandgap core circuit current consumption.

- Exploit the interesting property seen in chapter 1, on Fig. 1.5 on page 14. This figure shows the existence of V_{GS} for a precise I_C that has low sensitivity to temperature.

We can see the designed circuit in Fig.2.28.

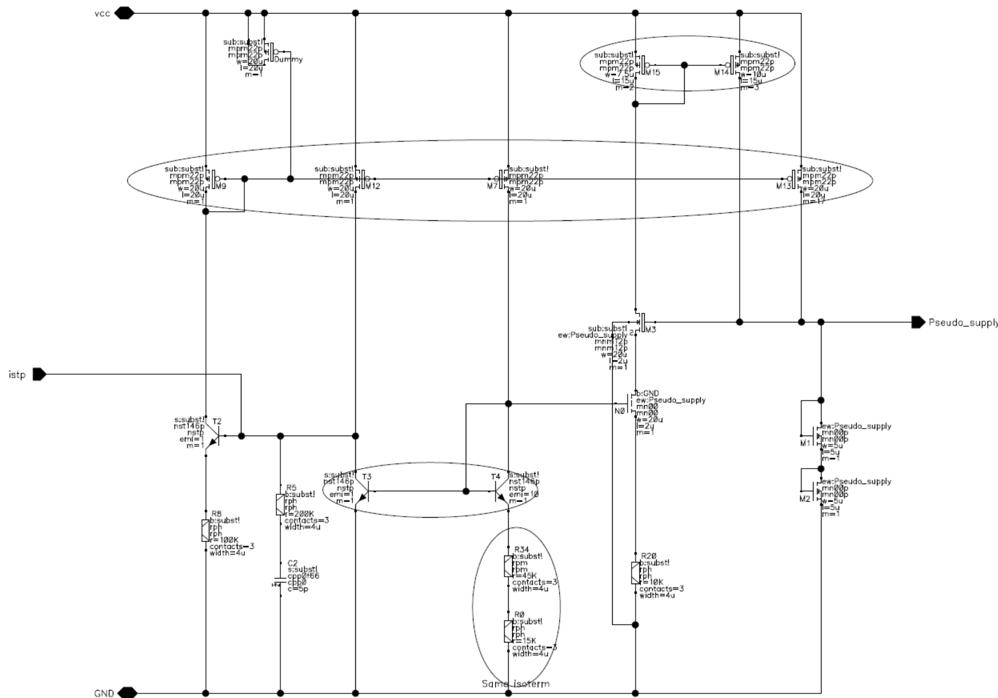


Figure 2.28: Pre-regulator circuit.

Being the current needed from the sub-bandgap reference core PTAT in nature, the first brick to realize this pre-regulator is the PTAT stage already seen in this chapter. The R_{PTAT} consists in the series of rph and rpm resistor: their temperature coefficient is slightly PTAT for the rpm and CTAT for the rph. This way, we can "modulate" the slope of the generated current so that, once multiplied for the mirror ratio, it had the same slope of the current required by the sub-bandgap reference core. However, Monte Carlo simulation showed that, for low temperatures, a significant percentage of samples suffered of low supplied current from the pre-regulator. As suspected, components mismatch was moving the bias and, at cold, the current generated is much less then at hot, so it takes only little mismatch quantities to reduce the pre-regulator current to the point that the sub-bandgap core circuit bias is compromised, creating unacceptable behavior. This issue would be solved increasing the PTAT current generated, but this would have brought a lot of current to be wasted during from-ambient-to-hot operating situation. We then thought to implement a circuit that generates a current boost only at low temperature. This was made by connecting the gate of a digital n-channel MOSFET to the bases of T3 and T4: this way, the current sunk by the n-channel MOSFET drain is driven by the base-emitter voltage, that is (as we know very well at this point) CTAT in nature and decreases of about -2mV/°C. V_{th} for the digital n-channel MOSFET is about 500mV, so when $V_{BE}(T_x) \approx 500mV$, the n-channel MOSFET will sink no more current, granting no current waste. Last little "pearl", being I_C proportional to $V_{OV}^2 \approx V_{BE}^2(T)$[12], I_C is kind of "$CTAT^2$". This and the resistor degeneration give even more precise boost where needed and make it wear off fast for $T > 30$°C. This behavior is shown in Fig. 2.29. The start-up for this stage has been realized exploiting the one already present in the PTAT stage on the sub-bandgap reference core. If the voltage reference is low, the same error current injected in the error-sensing node of the PTAT loop is injected in the pre-regulator loop, starting up the system.

The load of this pre-regulator are two diode-connected n-channel MOSFETs. We set the total output current to mach the sub-bandgap reference

[12] When in saturation region.

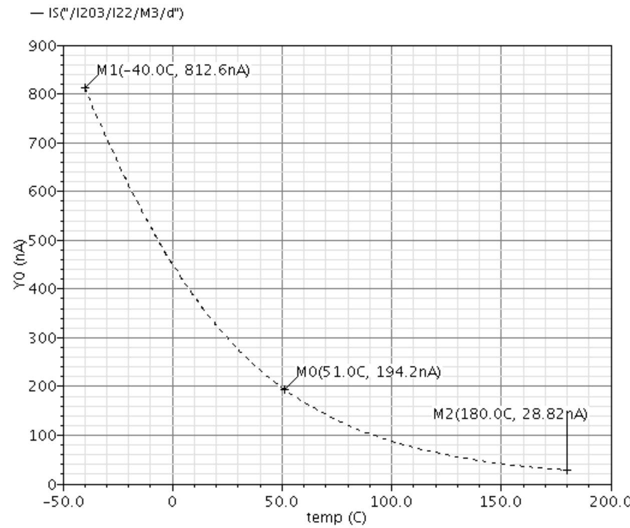


Figure 2.29: "Cold" current boost.

core current consumption plus about $5\mu A$ to bias the two diode-connected n-channel MOSFETs so that their $V_{GS}(T)$ matched the point of Fig. 1.5 on page 14. This way, not only we realized low impedance path to ground, for improved PSRR, but we also granted a temperature-stabilized voltage for the sub-bandgap reference core circuit!

2.5.7 AC Stability

Both in PTAT and CTAT stages, collector currents are controlled by loops, so an AC analysis is required to grant stability and smooth transient response. However, the loop bandwidth impact on the ability to reject noise from supply, hence the PSRR. An overcompensation that will affect the loops bandwidth will result in a poor $PSRR_{ac}$ performance, while low phase margins compromise in favor of greater bandwidth will cause oscillating transient response to fast stimuli. The pseudo-supply and the bandgap core shares the same PTAT stage; however, the compensation for these two stages are different. Let us take a look at the more precise sketch of the PTAT stage for the pseudo-supply in Fig. 2.30.

Pseudo Supply

This stage focuses on generating a voltage supply as stable as possible, so we have to keep this in mind while choosing how to compensate this loop.

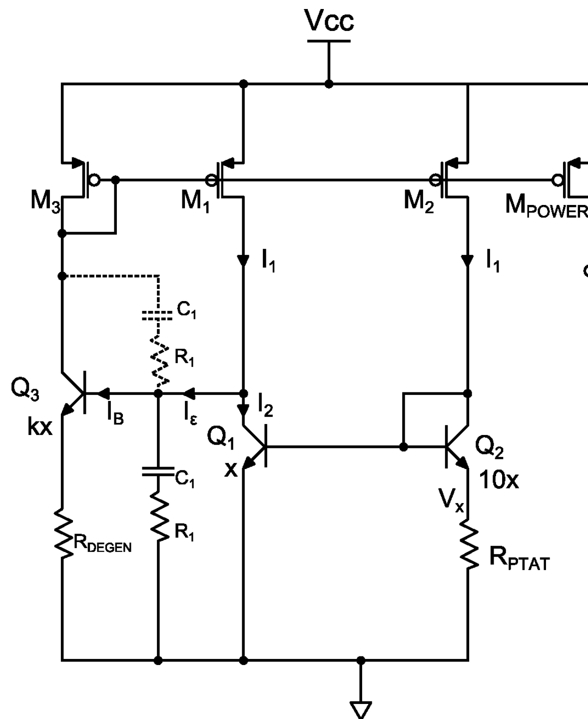


Figure 2.30: The pseudo-supply PTAT stage, complete representation.

The Miller compensation, dashed in Fig. 2.30, has to be avoided in this stage. The first reason is because it would couple the amplifying point of the loop with the gate of M3. This terminal follows the supply voltage in order to regulate its current (being a V_{GS} fixed for a certain I_C at a certain temperature): coupling this point to a sensitive point like Q3 base has to be avoided. Moreover, we would have basically no Miller effect placing an impedance between base and collector of Q3 because there is no voltage gain in that stage. This circuit (and the one for the PTAT current generation in the bandgap core) presents two loops, one positive and one negative, so we have to pay particular attentions about opening the loop in the right place for a correct evaluation of the loop gain. We can write the DC voltage gain

for an emitter-degenerated NPN

$$A_{V,npn} = \frac{\beta R_{load}}{r_\pi + (\beta + 1) R_{DEGEN}} \tag{2.82}$$

Being the load a diode connected p-channel MOSFET, $R_{load} = \frac{1}{g_{m,M3}}$. Even with no degeneration ($R_{DEGEN} = 0$), $A_{V,npn} < 5$, so we will have very little gain for Miller compensation. Note that Miller's effect is function of frequency, because $A_V(f)$ is. So it must be carefully applied. Let us write the block diagram for this loop. We can see it in Fig. 2.31 where

$$W_1(s) = \frac{Z_{comp}}{Z_{comp} + R_{in,NPN}} \tag{2.83}$$

is the current divider between the compensation impedance and the input impedance of Q3 (without loss of generality even if the compensation is not present, because in that case $Z_{comp} = \frac{1}{sC_P}$). We have

$$Z_{comp}(s) = \frac{sC_1R_1 + 1}{sC_1} \backslash\backslash \frac{1}{sC_P} = \frac{1}{s\left(C_P + C_1\right)} \cdot \frac{sC_1R_1 + 1}{1 + s\left(\frac{C_1C_P}{C_1 + C_P}\right)R_1} \tag{2.84}$$

where $C_P = C_\pi + A_{v,Q3}(f) C_{mu}$ and $A_{v,Q3}(f)$ is the voltage gain from the base to the collector of Q3. Be careful to correctly evaluate C_P according to the frequency.

$$R_{in,NPN} = r_\pi + (\beta + 1) R_{DEGEN} \tag{2.85}$$

where again, if no degeneration is applied, R_{DEGEN} can be set equal to zero, without loss of generality.

$$W_2(s) = \frac{r_{0,M3}}{1 + g_{m,M3} r_{0,M3}} \cdot \frac{1}{1 + sC_{TOT}\left(\frac{r_{0,M3}}{1 + g_{m,M3} r_{0,M3}}\right)} \tag{2.86}$$

$W_2(s)$ is the impedance seen from the collector of Q3, that will set the V_{GS} for the mirrors. C_{TOT} is the total capacitance seen from Q3 collector to ground. M1 and M2 are identical and share the same bias, so they have the same g_m. However, as already calculated in subsec. 2.4.2 on page 46, there is also the

two NPN transistors that work like a current-mirror so we have $H_1 = 1$ and $H_2(s) \approx 3.15\frac{1}{1+s\tau}$.

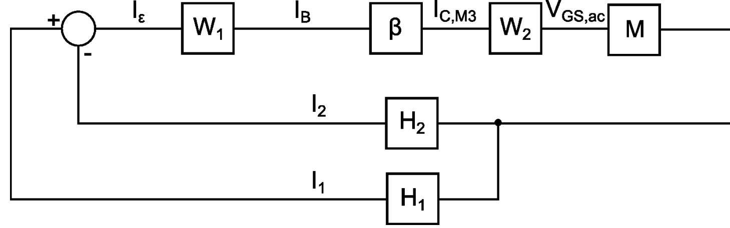


Figure 2.31: Pseudo-supply AC-loop block diagram.

We can write the loop gain

$$\begin{aligned} T(s) &= \beta \cdot W_1 \cdot W_2 \cdot g_{m,M1}(-3+1) &(2.87)\\ &= \beta \cdot \frac{1+sC_1R_1}{1+sX+s^2\left(C_PC_1R_1R_{in,NPN}\right)} \cdot \\ &\quad \cdot \frac{r_{0,M3}}{1+g_{m,M3}r_{0,M3}} \cdot \frac{1}{1+sC_{TOT}\left(\frac{r_{0,M3}}{1+g_{m,M3}r_{0,M3}}\right)} \cdot \\ &\quad \cdot g_m \cdot \left(-2 \cdot \frac{1-s\frac{\tau}{2}}{1+s\tau}\right) &(2.88) \end{aligned}$$

where

$$X = C_1R_1 + \left(C_P + C_{mu}\right)R_{in,NPN} + C_1R_{in,NPN} \tag{2.89}$$

If no compensation is applied (both degeneration and $C_1 - R_1$ series), we get a DC gain of

$$\begin{aligned} |T(0)| &\approx 20\log\left(\beta \cdot \frac{1}{g_{m,M3}} \cdot g_m \cdot 2\right) \\ &= 46.4\text{dB} &(2.90) \end{aligned}$$

and the first pole is at

$$\begin{aligned} f_{p1} &= \left[2\pi \cdot C_{TOT}\left(\frac{r_{0,M3}}{1+g_{m,M3}r_{0,M3}}\right)\right]^{-1} \\ &= 27.5kHz &(2.91) \end{aligned}$$

and the second pole, using W_1 with $C_1 = 0$

$$f_{p2} = (2\pi C_P + r_\pi)^{-1} \approx 580kHz \tag{2.92}$$

this is confirmed by STB simulation, as can been seen in Fig. 2.32. This figure also shows that there are others zeros and poles that have been neglected in the previous analysis. In fact, we have a negative phase margin, meaning that we are crossing f_c, the frequency where $|T(f_C)| = 0$, with more than -40dB/dec.

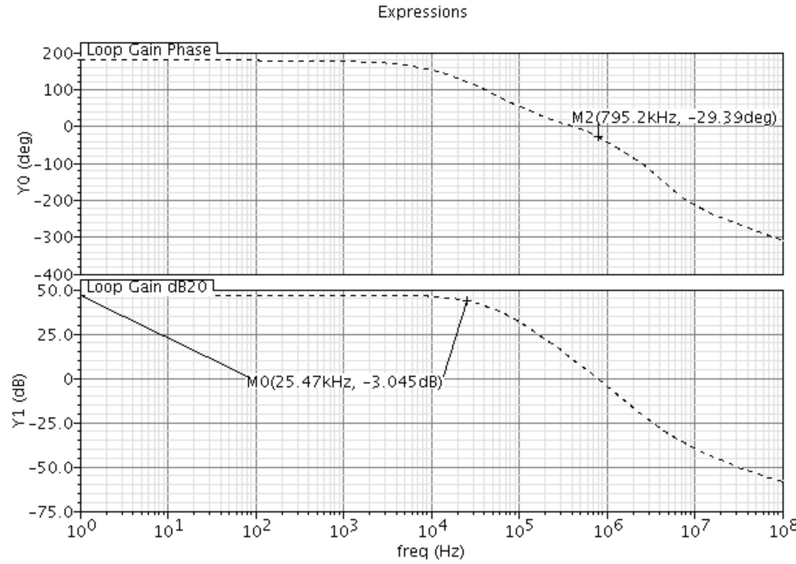


Figure 2.32: No compensation STB analysis.

One of the neglected effect is C_{mu} of Q3. C_{mu} introduces an high frequency zero at $f_{Z2} = \left(2\pi \frac{C_{mu}}{g_{m,Q3}}\right)^{-1} = 150MHz$ that does not influence our analysis but also add a third pole at $f_{P3} = (2\pi C_{mu} R_L)^{-1} = 39MHz$ that is more than a decade away from our crossing frequency, so it should have little effect. The pole that has great impact on the phase margin is due the current mirror created by Q1 and Q2. The last contribute to open loop gain and has been approximated by the simulator

$$[H_1 - H_2](s) = \left(-2 \cdot \frac{1 - s\frac{\tau}{2}}{1 + s\tau}\right) \tag{2.93}$$

with $f_{PH} = \frac{1}{2\pi\tau} \approx 5MHz$ and so $f_{ZH} = 10MHz$ with real part greater then zero.

Compensation

We will now proceed with the compensation.

$$A_{v,Q3} = \frac{\beta}{r_\pi} \cdot W_2(s) \tag{2.94}$$

$A_{v,Q3}$ has a DC value of 14dB and drops at -20dB/dec rate from about 25kHz, being $W_2(s)$ the input impedance of the load seen by Q3: it follows that one decade after, at $250kHz$, we have no gain, but on the contrary we are already attenuating, hence no Miller effect. In order to compensate this stage, let us summarize what poles and zero we have or we can get using degeneration resistor and $C_1 - R_1$ series. In this case we will have

$$W_1(s) = K_1 \cdot \frac{1 + sC_1R_1}{1 + s\left(C_1R_1 + C_PR_{in,NPN} + C_1R_{in,NPN}\right) + s^2\left(C_PC_1R_1R_{in,NPN}\right)} \tag{2.95}$$

$$W_2(s) = K_2 \cdot \frac{1}{1 + sC_{TOT}\left(\frac{r_{0,M3}}{1+g_{m,M3}r_{0,M3}}\right)} \tag{2.96}$$

$$W_3(s) = K_3 \cdot \frac{1 - s\frac{\tau}{2}}{1 + s\tau} \tag{2.97}$$

There are many ways to exploit R_{DEGEN}, C_1 and R_1 to obtain a compensation. We aimed for a stable 45° phase margin compensation. With "stable" we mean that, if possible, we tried to keep a certain gain margin over a constant phase. Process and mismatch will cause variations over circuits parameters, so a phase that is not flat at $\omega = \omega_c$ will be more sensitive over circuits variations. On the contrary, a phase that remains constant over a certain range of frequency, will be less sensitive to circuit variations. In Fig. 2.33, gain and phase for this compensation are visible. We used $R_{DEGEN} = 100k\Omega$ and $C_1 = 5pF$ to get a dominant pole at $f_{PD} = 2,4kHz$ and the first -90° phase rotation that ends a decade after. At $f_{P1} = 27.5kHz$, the second pole introduced by W_2 brings another -20dB/dec and another -90° rotation. This

way we would approach $f_c \approx 150kHz$ at -40dB/dec, hence PM=0°. We have still less than one decade before the third pole at $f_{PH} = 800kHz$ and the first zero at $f_{ZH} = 1.6MHz$, so these will slightly affect the Bode diagram at f_c and this will be evaluated by simulation. We then decide to play our negative real part zero just over f_c. Being C_1 already set to $5pF$, we need

$$R_1 = \frac{1}{2\pi C f_c} = 212k\Omega \tag{2.98}$$

R_1has been rounded up to $200k\Omega$ for layout purpose. We verified that our modeling were congruent with the circuit superimposing the data from cadence to the Bode diagram traced with MATLAB, as shown in Fig. 2.33.

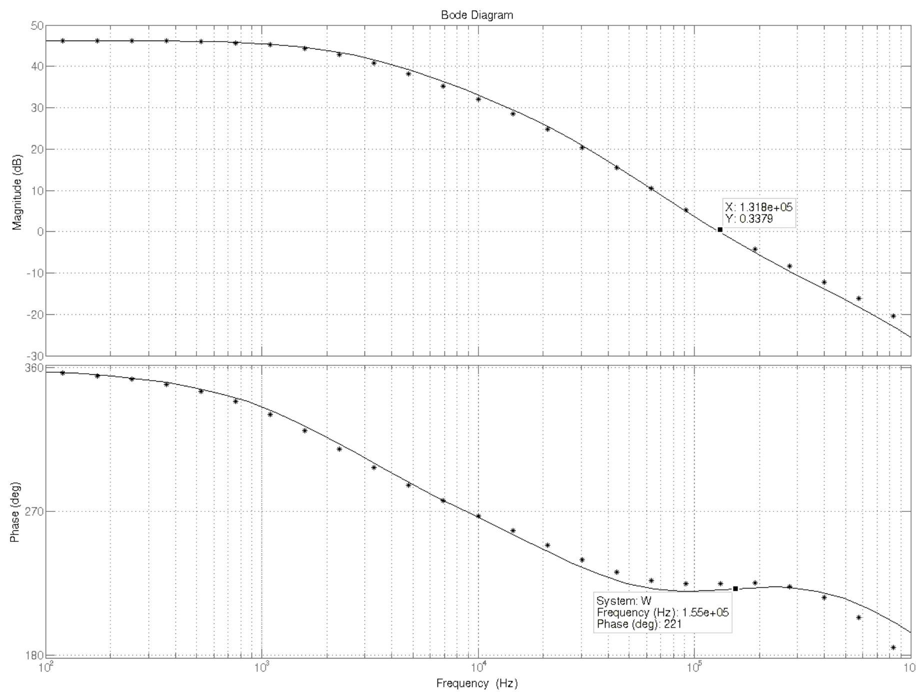


Figure 2.33: MATLAB verification of poles and zeros placement in the pseudo-supply: simulator data are dotted.

This choice for f_C gives a phase boost so to cross that frequency with exactly 45° of phase margin.

Bandgap Core, PTAT Stage

The compensation of this part is similar. However, as can be seen in Fig. 2.34, this stage results cascoded by M_C and the low impedance DC path consists in two diode connected n-channel MOSFETs, N_1 and N_2, in series. The peculiarity of this stage is that all the AC current sunk from all the p-channel MOSFETs in the circuit does not come from virtual ground (like in the pseudo-supply stage, where it was sunk from supply) but from N_1 and N_2 series, so the low impedance given by those two diode-connected n-channel MOSFETs works like a source-degeneration. Let us see what we mean.

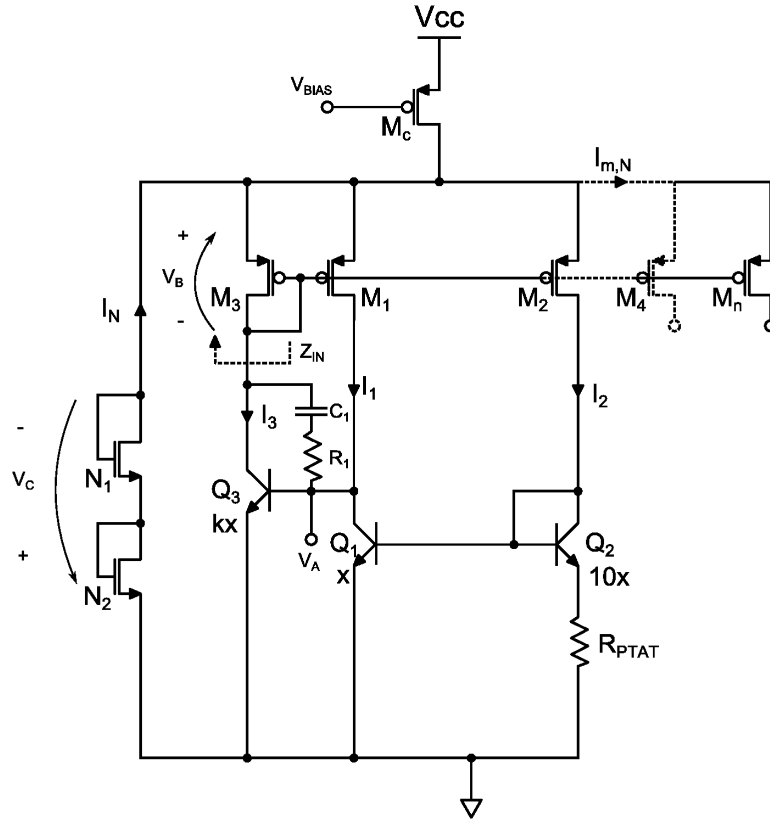


Figure 2.34: The Bandgap core PTAT stage, complete representation.

Using the notations of Fig. 2.34, we can write that the open loop A_v gain for Q3 is

$$A_v = \frac{V_C + V_B}{V_A} \tag{2.99}$$

We can write V_B and V_C:

$$V_B = \frac{V_A}{r_\pi}\frac{\beta}{g_{m,M3}} \tag{2.100}$$

$$V_C = I_N \cdot \left(\frac{1}{g_{m,N1}} + \frac{1}{g_{m,N2}}\right) \tag{2.101}$$

but now I_N is

$$I_N = I_{C,M3} + I_{C,M1} + I_{C,M2} + I_{M,n} \tag{2.102}$$

that is, all the current mirrored by the all the p-channel MOSFETs have to come from N_1 and N_2 series. This complicates things a lot since it would be necessary to evaluate the transfer functions $\frac{I_{M,n}}{V_B}$. We will do a theoretical analysis considering the PTAT stage only and we will use some simplifications and the simulator to extend this analysis to the whole circuit; this is equivalent to say that we are going to analyze $I_{M,n}$ contribution in a second time. With calculation made from the small signal circuit[13] we calculated the input impedance drawn in Fig. 2.34. It results:

$$\begin{aligned} Z_{in}(s) &= \frac{(1 + kg_{m,M1}R_L) + R_L g_{m,M3} + \frac{R_L}{r_0}}{g_{m,M3}} \cdot \\ &\cdot \frac{1 + sC_T\left[\frac{r_0}{\frac{r_0}{R_L}\left(1+kg_{m,M1}R_L\right)+g_{m,M3}R_L+1}\right]}{1 + sC_T\left[\frac{r_0}{1+g_{m,M3}r_0}\right]} \end{aligned} \tag{2.103}$$

where k is a multiplying factor that takes into account the number of p-channel MOSFETs connected so to share the same V_{GS} of M_1 and C_T is the sum of all the C_{GS} connected to the gate of M_3. Looking at its DC value, we get

$$Z_{in,DC} \approx R_L + \frac{(1 + kg_{m,M1}R_L)}{g_{m,M3}} \tag{2.104}$$

Using the values of the circuit, we get $Z_{in,DC} \approx 170k\Omega$, $f_p \approx 590kHz$ and $f_z = 1.2MHz$. These value are very close to simulation results, but the zero is at higher frequency. However we only need the bandwidth where Miller effect can be applied, that is roughly $500kHz$. We are now able to evaluate

[13] See Appendix A.3 on page 130

the voltage gain for transistor Q3, assuming that we can use the DC value for Z_{in} for $f < 500kHz$.

$$A_v = \frac{\beta}{r_\pi} Z_{in,DC} \tag{2.105}$$

Substituting the values β and r_π, we get $A_v = 21.95$, that fits with the $A_v = 20.7$ given by the simulator. Correcting the k factor so to include all the other p-channel MOSFETs we get $A_v = 32 + A_X$ where A_X is the gain due to the p-channel MOSFET in the CTAT stage. Using the simulator to get this result by manual verification of $Z_{in}(f)$, we get $A_v = 34.5$. We will exploit this gain for implementing a Miller compensation. What will follow however is the compensation of the PTAT stage excluding all the contributes from other stages of the Bandgap. So our calculation will be done for $A_V = 20.7$. The open loop gain is given by

$$T(s) = [W_1 \cdot W_2 \cdot W_3](s) \tag{2.106}$$

$$W_1(s) = \beta \cdot \frac{1}{1 + s\left[(A_v + 1) C_{mu} + C_\pi\right] r_\pi} \tag{2.107}$$

that is the transfer function $\frac{I_c}{I_B}(f)$ for Q3. Then the AC voltage V_B (see Fig. 2.34) pilots the gates of M_1 and M_2, that transduces this voltage to collector current by their transconductance g_m. So the transfer function $W_2 = \frac{I_{D,M1}}{I_C} = \frac{I_{D,M2}}{I_C}$ is given by

$$W_2(s) = \frac{g_{m,M1} r_{0,M3}}{1 + g_{m,M3} r_{0,M3}} \cdot \frac{1}{1 + s\frac{C_T r_{0,M3}}{1 + g_{m,M3} r_{0,M3}}} \tag{2.108}$$

which introduces a pole at $f_{p3} = 580kHz$ and then, like the pseudo-supply, we have a transfer function due to the mirror made by transistors Q1 and Q2.

$$W_3(s) = -2 \cdot \frac{1 - s\frac{\tau}{2}}{1 + s\tau} \tag{2.109}$$

that introduces a pole at $f_{P,W3} = 5MHz$ and a zero with real part greater than zero at $f_{Z,W3} = 2 \cdot f_{P,W3} = 10MHz$.

Compensation

The compensation is made by introducing an impedance made by resistor-capacitor in series R_1 and C_1, so that $Z_C = \frac{1+sC_1R_1}{sC_1R_1}$, as shown in 2.34. This changes the transfer function W_1 because now in the current is divided between the total impedance between base and collector of Q3 and the impedance between Q3 base and ground. The first impedance is $Z_C//C_{mu}$ referred to ground and divided for A_v, for $f < 500kHz$. The impedance seen from Q3 base is $r_\pi//C_\pi$. W_{1C} describes the current that flows through r_π only and so that will be amplified by β.

$$W_{1C} = \beta \frac{1 + sC_1R_1}{1 + s\left[C_1R_1 + A(C_1 + C_{mu})r_\pi\right] + s^2\left[A\left(C_1 + C_{mu}\right)r_\pi\left(\frac{C_1C_{mu}}{C_1+C_{mu}}\right)R_1\right]} \tag{2.110}$$

with $A = (A_v + 1)$ so, writing

$$\begin{cases} a = & (A_v + 1)\left(C_1 + C_{mu}\right)r_\pi\left(\frac{C_1C_{mu}}{C_1+C_{mu}}\right)R_1 \\ b = & C_1R_1 + (A_v + 1)(C_1 + C_{mu})r_\pi \end{cases} \tag{2.111}$$

we can place a dominant pole with $R_1 = 85k\Omega$ and $C_1 = 1pF$ at

$$f_{p1} = \left[-\frac{1}{2\pi}\cdot\frac{-b+\sqrt{b^2-4ac}}{2a}\right]^{-1} = 7.8kHz \tag{2.112}$$

a second pole at

$$f_{p2} = \left[-\frac{1}{2\pi}\cdot\frac{-b-\sqrt{b^2-4ac}}{2a}\right]^{-1} \approx 2MHz \tag{2.113}$$

and the zero at

$$f_{z1} = 1.9MHz \tag{2.114}$$

that almost cancels the second pole. We can say that while the pole at f_{p1} is correct, the pole at f_{p2} probably is placed at higher frequency because the Miller's hypothesis on A_v are not valid in that frequency range. The

zero,instead, is not influenced by Miller's effect and so f_{z1} is reliable. Now we have to evaluate the transfer function $\frac{I_{C,Q3}}{I_{Z_{in}}}$. The impedance between base and collector of Q3 should also be placed between collector and ground multiplied by $\frac{A_v}{1+A_v}$. For $f < 500kHz$ there is no effect on this impedance, but for $f < 500kHz$, it tends to zero. For $f < 500kHz$, we can write:

$$W_{2C}(s) = \frac{1}{1+\frac{Z_{in}}{Z_{out,Q3}}} = \frac{\left(1+s\frac{C_T}{g_{m,M3}}\right)\cdot(1+sC_1R_1)}{1+sX_1+s^2X_2} \tag{2.115}$$

where

$$\begin{cases} X_1 = C_1R_1 + \frac{C_T}{g_{m,M3}} + C_1Z_{in,DC} \\ X_2 = \frac{C_T}{g_{m,M3}}C_1R_1 + C_TR_ZC_1Z_{in,DC} \end{cases} \tag{2.116}$$

and

$$R_Z = \frac{r_0}{\frac{r_0}{R_L}\left(1+kg_{m,M1}R_L\right)+g_{m,M3}R_L+1} \tag{2.117}$$

This gives two poles at $f_{p3} = 546kHz$ and $f_{p4} = 678kHz$ and two zeroes at $f_{z2} = 587kHz$ and $f_{z3} = 1.87MHz$. While the zero located at f_{z2} is not sensitive on Miller's effect, the other two poles and the zero are in a frequency range where $5 < A_v < 20$, so $0.83 \cdot Z_{out,Q3} < \tilde{Z}_{out,Q3} < 0.95 \cdot Z_{out,Q3}$, and so a slight inaccuracy may afflict them. $W_{3C} = W_2$, that remains unaltered and has a single pole at $f_{p5} = 586kHz$ and finally we have $W_{4C} = W_3$ with a pole at $f_{P,W3} = 5MHz$ and a zero with real part greater than zero at $f_{Z,W3} = 2 \cdot f_{P,W3} = 10MHz$. The DC gain is then

$$20Log\left(|T(0)|\right) = 20Log\left(2\beta\frac{g_{m1}}{g_{m3}}\right) = 35.6dB \tag{2.118}$$

The results from simulator have been superimposed to the Bode diagram drawn with MATLAB and can be seen in Fig. 2.35.

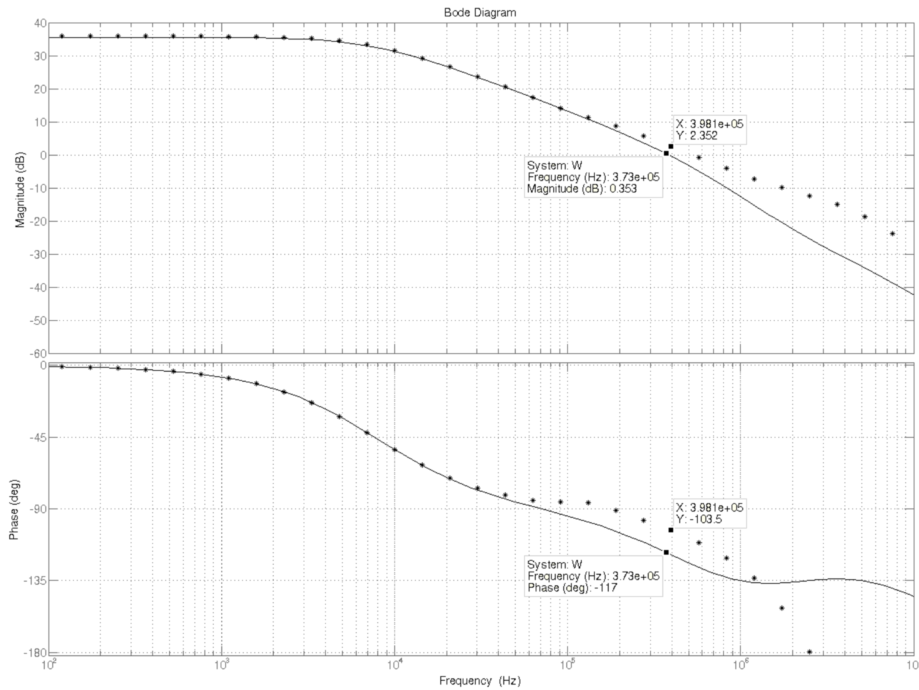


Figure 2.35: MATLAB verification of poles and zeros placement in the PTAT stage: simulator data are dotted.

We have a calculated phase margin of 63° against a simulated value of 77°. However, this analysis excluded part of the circuit. We verified with a stability analysis that the global circuit does not compromise the stability, and it results a phase margin of 60°.

Bandgap Core, CTAT Stage

The same approach and similar calculation have been used for the compensation of the CTAT stage, shown in Fig. 2.36.

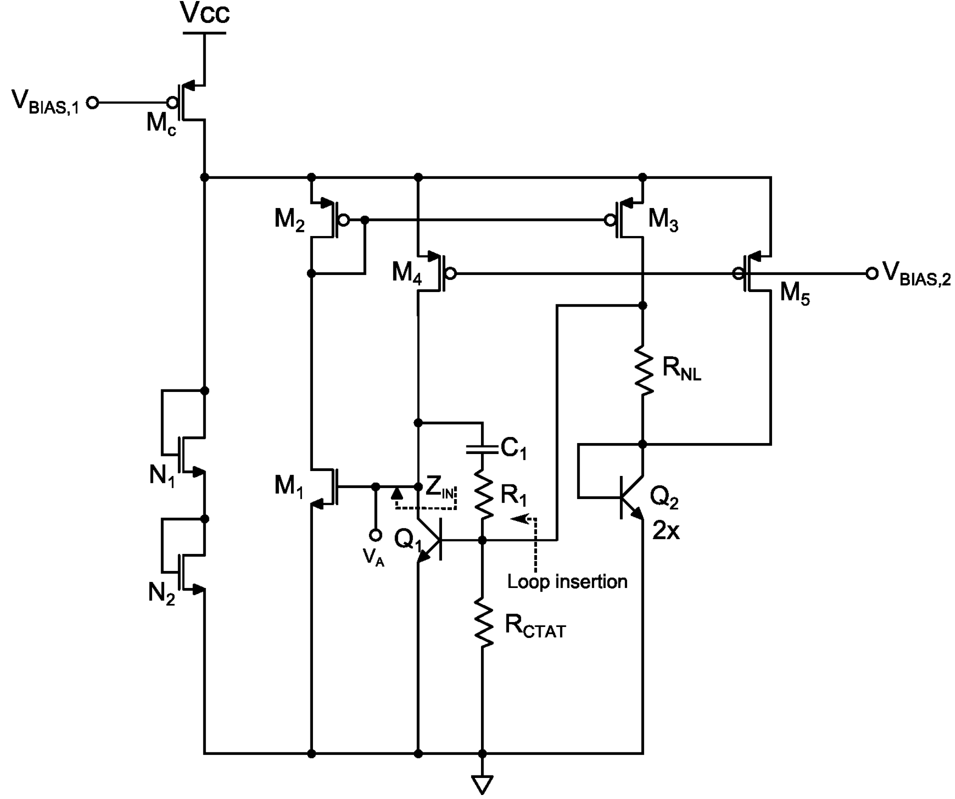


Figure 2.36: The Bandgap core CTAT stage, complete representation.

Let us consider the circuit without the compensation impedance. Considering a current insertion where marked in Fig.2.36 we can follow the signal and evaluate its return ratio. The first transfer function we find is due to the current divider at Q_1 base node. We have

$$Z_{in,Q1}(s) = \frac{r_\pi}{1 + s\left[(A_v + 1)\, C_{mu} + C_\pi\right] r_\pi} \tag{2.119}$$

so the current flowing on $r_\pi$will be

$$W_{1,in}(s) = \frac{R_{CTAT}}{r_\pi + R_{CTAT}} \cdot \frac{1}{1 + s\left[(A_v + 1)\, C_{mu} + C_\pi\right] \frac{R_{CTAT} \cdot r_\pi}{R_{CTAT} + r_\pi}} \tag{2.120}$$

that yields a pole at $f_{p1} = 3.5kHz$ and a Bode gain of 0.2. In order to calculate this transfer function with had to evaluate the A_v gain by the simulator. This current is then transferred to the output with the transfer

function

$$W_{1,out}(s) = \beta \cdot \left[1 + sC_{mu}\left(\frac{r_\pi + \beta R_L}{\beta}\right)\right] \tag{2.121}$$

introducing a zero at $f_{z1} = 25kHz$. This current now flow through Z_{in}, shown in Fig. 2.36. This impedance yields the transfer function

$$Z_{in}(s) = 125M\Omega \cdot \frac{1}{1 + sC_{gs,M1}\left(r_{o,M2} + R_L\right)} \tag{2.122}$$

that transduces the collector current of Q_1 to a voltage that drives the gate of M_1, so that its drain current will sink a current proportional to its transconductance $g_{m,M1}$. We can write

$$W_2 = Z_{in} \cdot g_{m,M1} \tag{2.123}$$

A pole at $f_{p2} = 6kHz$ is then introduced. This current will be mirrored to M_3 drain with the transfer function

$$\begin{aligned} W_3(s) &= \frac{g_{m,M3} r_0}{1 + g_{m,M2} r_0} \cdot \frac{1}{1 + sC_T \frac{r_0}{1 + g_{m,M2} r_0}} \\ &\approx \frac{g_{m,M3}}{g_{m,M2}} \cdot \frac{1}{1 + s\frac{C_T}{g_{m,M2}}} \end{aligned} \tag{2.124}$$

where C_T is the total capacitance seen from the gate of M_2. W_3 introduces a pole at $f_{p3} = 205kHz$. The current flowing through the drain of M_3 will now see another impedance divider that can be approximated as

$$W_4 = \frac{R_{NL}}{R_{NL} + R_{CTAT}} \approx 0.3 \tag{2.125}$$

The total DC gain is the product of all the bode constant of the transfer function:

$$K_{bode} = \frac{R_{CTAT}}{R_{CTAT} + r_\pi} \cdot \beta \cdot 125M \cdot g_{m,M1} \cdot \frac{g_{m,M3}}{g_{m,M2}} \cdot 0.3 \tag{2.126}$$

that expressed in dB, gives

$$20Log\left(|K_{bode}|\right) = 74dB \tag{2.127}$$

We can notice that we will cross f_c with at less than -40dB/dec, hence no phase margin. Moreover, we neglected many high frequency poles so the phase margin will result negative.

Compensation

Let us introduce the compensation impedance consisting in the series of $R_1$and C_1. The transfer function $W_{1,in}$ become

$$\begin{aligned} W_{1,in,C}(s) &= \frac{R_{CTAT}}{r_\pi + R_{CTAT}} \cdot \\ &\cdot \frac{1 + sC_1R_1}{1 + s\left[C_1R_1 + A_v\left(C_1 + C_{mu}\right)R_P\right] + s^2\left[A_vC_1C_{mu}R_PR_1\right]} \end{aligned} \tag{2.128}$$

Being $A_v \approx 60dB$, A_v heavily boost the compensation impedance (and we simplified $A_v + 1 \approx A_v$), giving a pole at $f_{p1} = 230Hz$ and a pole-zero cancellation at $f_{p2} \approx f_{z1} = 790kHz$. We note by the simulator, that even though $Z_{in,C}$ (Z_{in} *compensated*) has the expected shape given by a new transfer function can be calculate as the parallel of Z_{in} the $R_1 - C_1$ series, the transfer function $g_{m,M1} \cdot Z_{in}$ is flat, with probably some zero-pole cancellations. W_3 remains unaltered with a pole at $f_{p3} = 205kHz$ and finally the impedance divider changes in

$$W_4(s) \approx \frac{R_{NL}}{R_{NL} + R_{CTAT}} \cdot \frac{1 + sA_vC_1R_1}{1 + sC_1\left(\frac{AR_1R_{NL}}{R_{CTAT} + R_{NL}}\right)} \tag{2.129}$$

which has a zero in $f_{z2} = 530kHz$ and a pole at $f_{p4} = 2MHz$. This zero-pole placement grants a quasi-constant slope of -20dB/dec from $74dB$ beginning at $f_{p1} = 280Hz$. That yields, in first approximation, an $f_c < f_{p4}$, so the phase margin should be greater than 45°. We calculated with MATLAB what exact phase margin our zero pole placement yields and verified it with the simulator. The result is shown in Fig. 2.37.

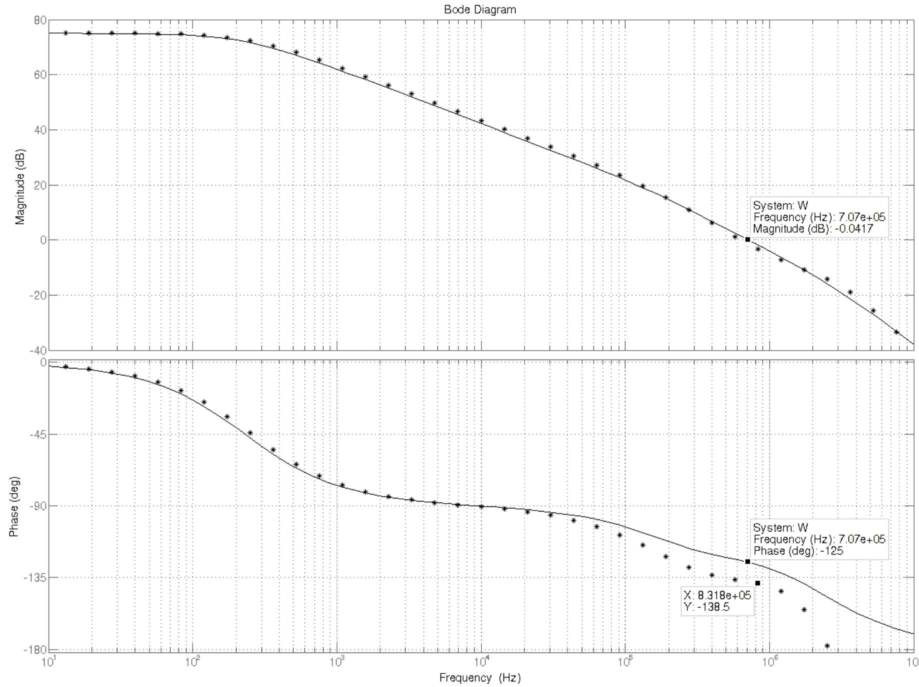


Figure 2.37: MATLAB verification of poles and zeros placement in the CTAT stage: simulation data are dotted.

The simulator yields a phase margin of 43° versus the 55° of the calculated value. This imprecision is due also to the high frequency range where f_c has been placed, where simplification hypothesis begins to have a big influence on the Bode diagram.

These compensations impact also on transient response and PSRR. An high phase margin would require a compensation that reduces gain or bandwidth greatly, but this would make the loop weaker in rejecting the noise from the supply, even if guaranteeing a smoother transient response. On the other hand, a low phase margin would mean higher bandwidth and better PSRR but overshoot in the output-voltage transient response. This is why these compensations aim at a compromise between bandwidth and phase margin of 45°, that should grant a smooth transient response, but still fast enough to grant a good PSRR.

2.5.8 Monte Carlo

The production of wafers involves two typical kind of errors: process and mismatch. Process errors are related to a phenomenon that afflicts all the same devices in the same way: a shift of resistors nominal values is an example. However, the ratio between the process-afflicted parameter of the devices remains constant. Mismatch errors afflict the ratio between device parameters and can cause more severe consequences in this topology. We will now analyze the main effects that can afflict the reference precision. Let us refer to Fig. 2.38 for the analysis of this subsec..

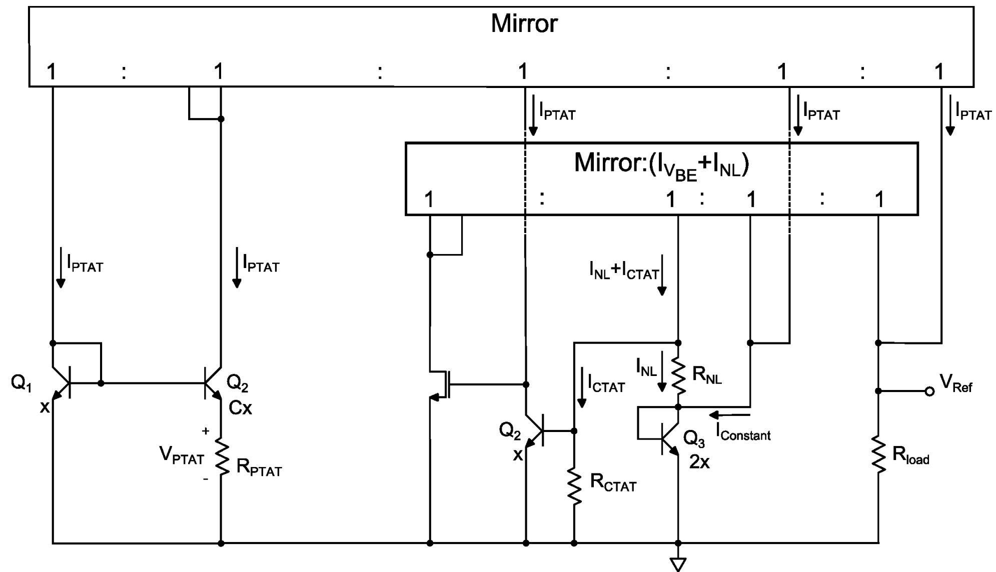


Figure 2.38: Diode loop topology, explicits circuit.

Resistor mismatch refers to relative differences in value among resistors that should nominally have a fixed proportion. We can refer the mismatch of R_{NL}, R_{CTAT}, and R_{PTAT} to R_{load} for simplicity (R_{load} constant and all the other resistors varies all with different coefficients). We can write new ratios for these mismatching resistor where α, β and γ are the mismatching coefficient of each resistor (they are all different, otherwise we would fall into a process-kind case). With calculations shown in A.4.1, we determine that

the error on the reference voltage is

$$\Delta V_{ref} = 0.1 \cdot 1.273m \cdot T + 0.1 \cdot 0.15m \cdot T \ln\left(\frac{T}{T_r}\right) \quad (2.130)$$

The error is graphically displayed in Fig. 2.39. We can see that this error causes a maximum $\pm 28mV$ offset and a maximum $|\Delta V_{ref}| = 26mV$. This error is severe and will require a care full layout design to minimize this effect.

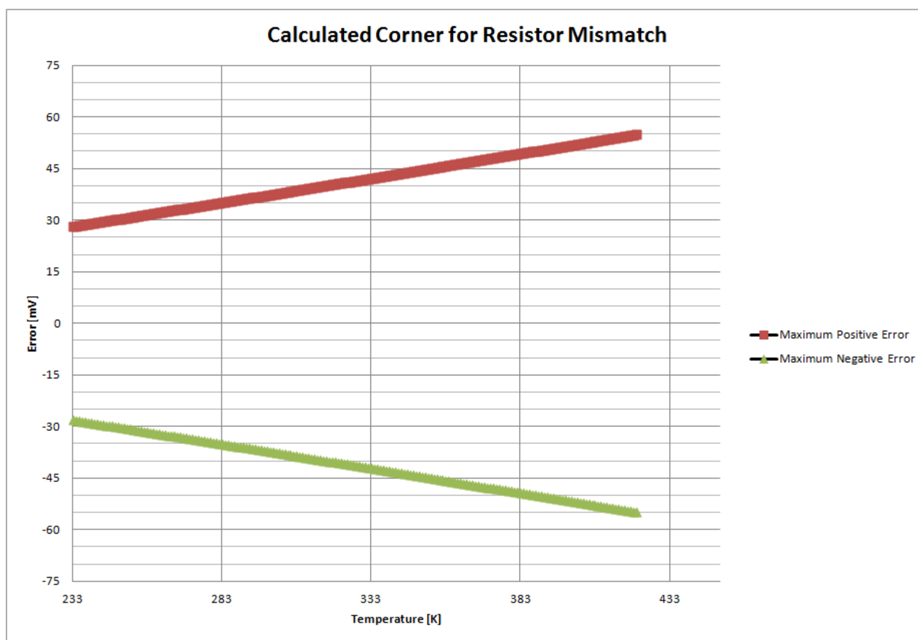


Figure 2.39: Corners on resistors mismatch error.

The other main source of error is due to mirror mismatch. This happens when the $\frac{W}{L}$ ratio of the p-channel MOSFETs changes from one transistor to another. However, Fig. 2.38 shows that the mirror are divided into two blocks. The mismatch between p-channel MOSFETs of different blocks is not connected, so it is not necessary to layout the mirror stages so to match all the p-channel MOSFETs. What is important, is to design a layout scheme that ensures a good matching for the p-channel MOSFETs of the each mirror block. *Divide and conquer!* The mismatch on the PTAT mirror block triggers also many of the errors described in Sec.2.4. The most serious is the error

introduced on the collector-current ratio. Referring to Fig. 2.40, we can analyze this error using eq. (1.47).

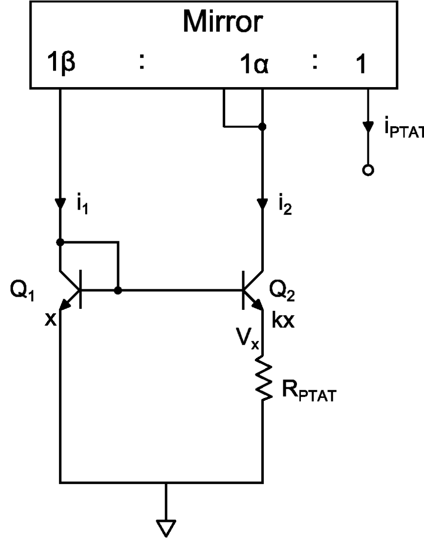


Figure 2.40: PTAT stage, mirror mismatch.

We can rewrite this equation to see the error effect by

$$I_{PTAT} = \frac{V_T \ln\left(\frac{\alpha I_{C1}(T)}{Ae_1 J_S(T)} \cdot \frac{Ae_2 J_S(T)}{\beta I_{C2}(T)}\right)}{R_{PTAT}} \tag{2.131}$$

$$= \frac{V_T \ln(C) + V_T \ln\left(\frac{\alpha}{\beta}\right)}{R_{PTAT}} \tag{2.132}$$

so the error introduced is

$$\Delta I_{PTAT} = \frac{V_T}{R_{PTAT}} \ln\left(\frac{\alpha}{\beta}\right) \tag{2.133}$$

This effect must be superimposed to the effective mirror error. The error is just the consequence of the modified mirror ratio, so it yields

$$I_{PTAT,out} = \frac{I_{PTAT}}{\alpha} \tag{2.134}$$

Using (2.132) into eq.(2.134), we get

$$I_{PTAT} = \frac{V_T \ln(C) + V_T \ln\left(\frac{\alpha}{\beta}\right)}{\alpha R_{PTAT}} \tag{2.135}$$

so the error is

$$\Delta V_{ref,P} = \Delta I_{PTAT} \cdot R_{load} = \frac{V_T}{R_{PTAT}}\left[\frac{(1-\alpha)\ln(C)}{\alpha} + \frac{\ln\left(\frac{\alpha}{\beta}\right)}{\alpha}\right] \tag{2.136}$$

The error for the CTAT stage is instead dominated by the error ratio between the $I_{CTAT} + I_{NL}$ generated and its mirrored version sent to the output load. This is simply given by

$$\Delta V_{ref,C} = \Delta I_{CTAT} \cdot R_{load} = \gamma I_{CTAT} \cdot R_{load} \tag{2.137}$$

The errors of these two stages have been plotted in Fig. 2.41 and 2.42 for the PTAT and CTAT stage respectively.

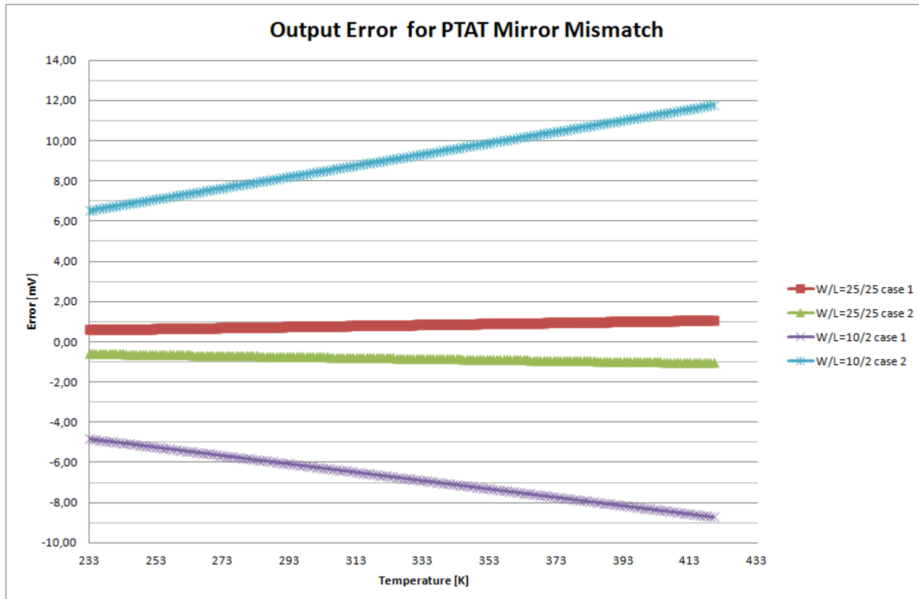


Figure 2.41: Current mirror mismatch, PTAT stage contribute.

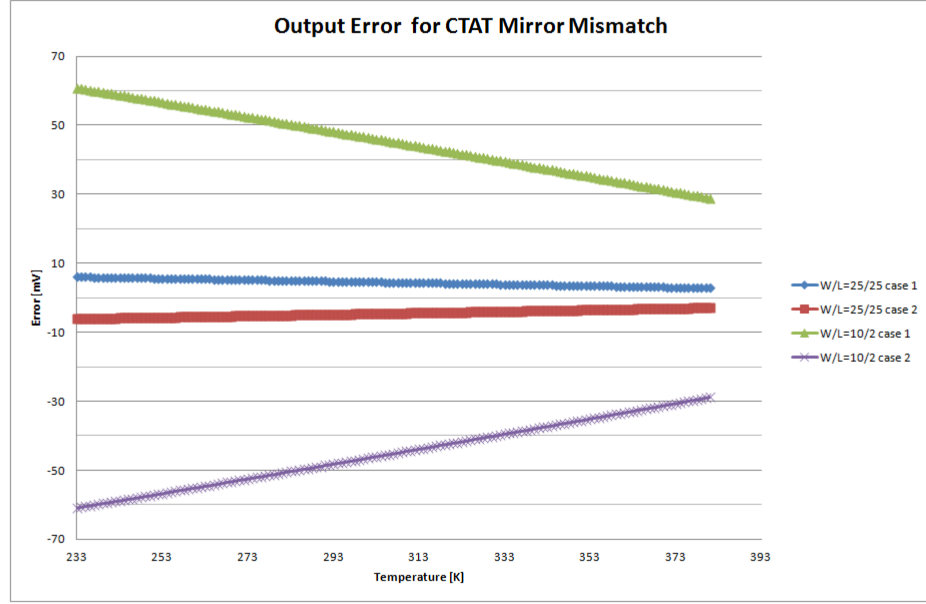


Figure 2.42: Current mirror mismatch, CTAT stage contribute.

The PTAT stage presents the interesting characteristic of cancellation if $\beta > \alpha$, this reduces the sensitivity to this stage to mirror mismatching. Figures 2.41 and 2.42 show also the worst case for two $\frac{W}{L}$ ratio. These cases give an idea on how big the p-channel MOSFETs should be sized in order to obtain certain "statistical-precision" result and how the error is influenced by the area of the transistor.

NPN transistor area can also mismatch, arising an error in the PTAT generation. This error is however caused by a phenomenon with low variance and usually lays around $\delta_{NPN} = \pm 2\%$. This error cause a $\Delta V_{ref,NPN}$

$$\Delta V_{ref,NPN} = V_T \frac{R_{load}}{R_{PTAT}} \cdot \ln(1 - \delta_{NPN}) \tag{2.138}$$

The same calculation can be done for the CTAT stage, where the area ratio is used to control the non-linear voltage. So, in our case, these errors effect on V_{ref} in a worst case scenario are shown in Fig. 2.43 and 2.44 respectively.

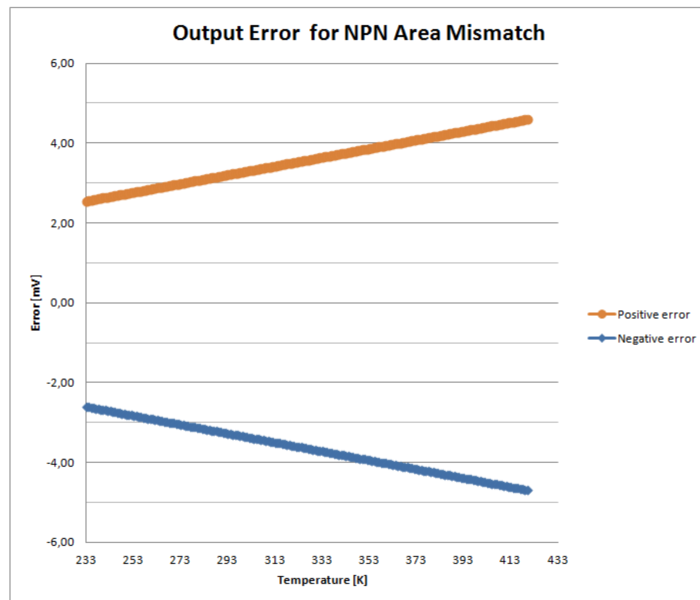


Figure 2.43: NPN mismatch, PTAT stage contribute.

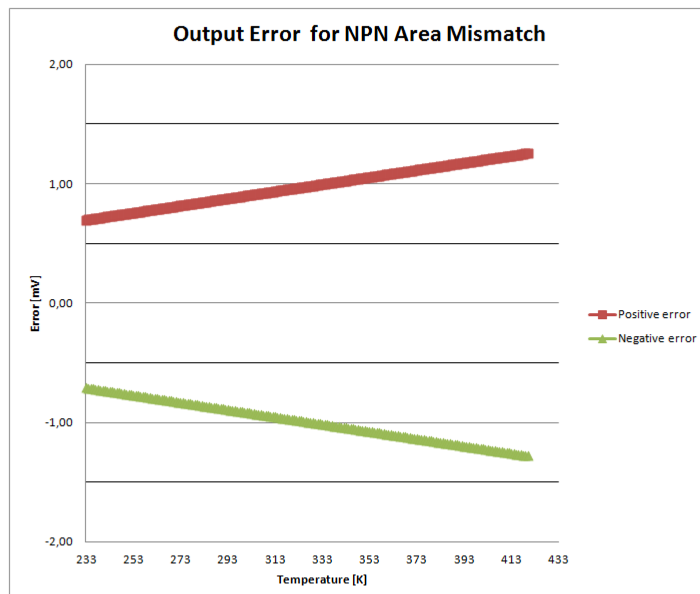


Figure 2.44: NPN mismatch, CTAT stage contribute.

2.5.9 Extracted Simulation

Once the circuit design has been done and the schematic has been settled, it is time to perform it layout. Once the layout view is ready, it is possible to simulate the circuit behavior taking into account parasitic resistances and capacitors. Thanks to these simulation, we have spotted an anomaly in the PSRR behavior of version 2. At a glance, the layouts of the three versions were very similar, but the PSRR at 10kHz for the extracted view for version 2 was 20dB lower[14] than the ideal circuit. This loss was not being encountered in the other bandgap, so the problem was associated to that particular layout. This behavior interested only the medium frequency range and not the DC region, so we simulated the extracted view with only some well defined range of parasitic capacitors together with all the parasitic resistors. This way, we approximately associated to every parasitic capacitors set their contribution to the PSRR decrease. Once identified the main source of degradation, we identified every parasitic capacitors of that set and tried to see which could contribute to the effect. Following this method, we identified the main problems: namely the capacitance coupling between sensitive metal traces and PAD placing. Rerouting some metal paths and changing the PAD placement (supply and pseudo-supply PAD placed far apart and far from the output PAD and if possible with the GND PAD between them is a good solution) brought a big improvement over the PSRR for both medium and high frequency, of approximately -20dB for $300Hz < f < 10kHz$ and -15dB for $f > 2MHz$, as shown in Fig. 2.45.

[14] In absolute value.

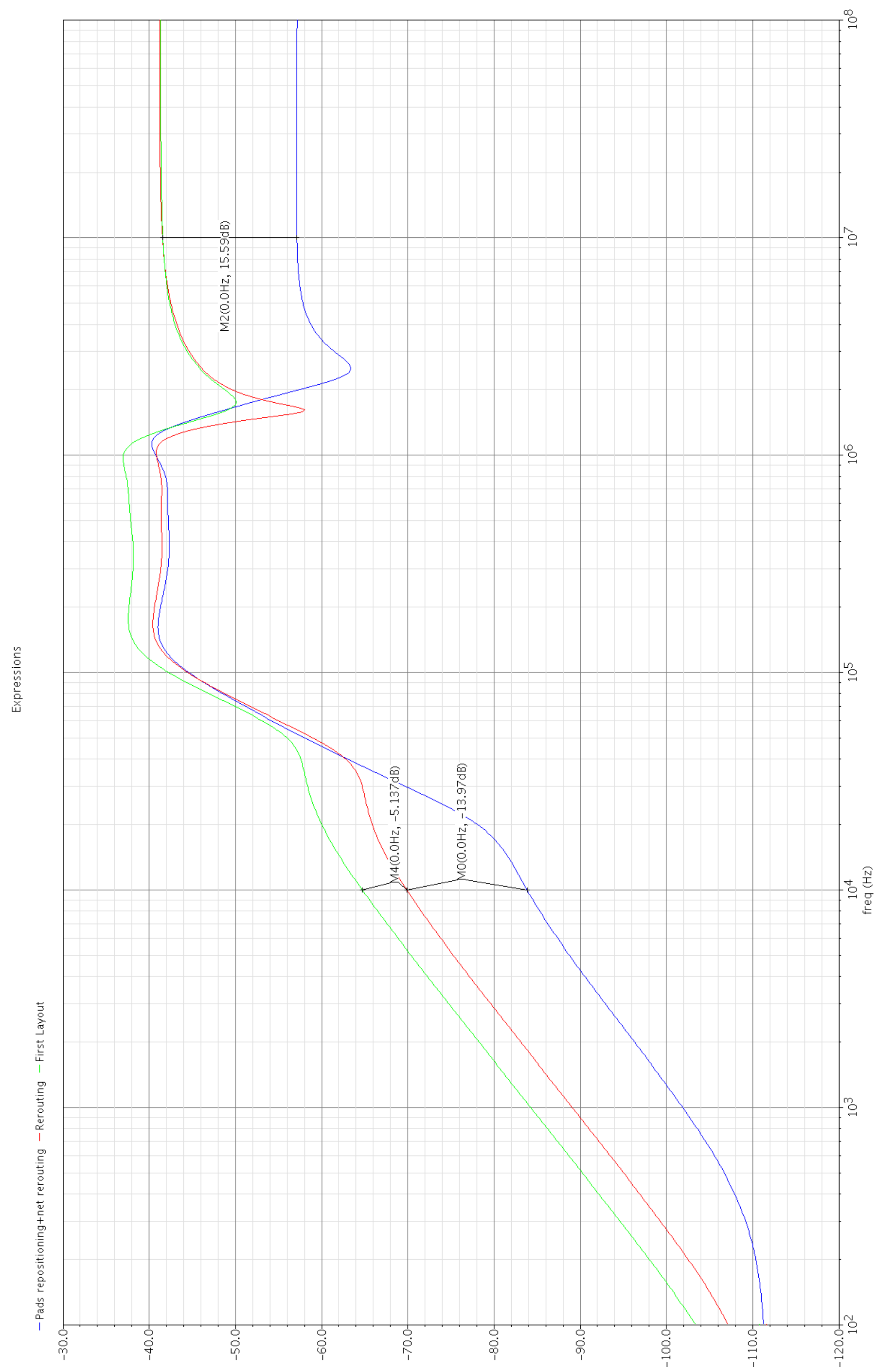


Figure 2.45: Improving steps in layout debug.

2.5.10 Performance

It is now time to summarize the achieved performance. Let us use the same table shown in subsec. 2.1.2 on page 35.

	Specification	Value	Passed
Precision	1 to 20 ppm/°C	0.5ppm/°C	V
$PSRR_{DC}$	-100dB	-121dB	V
$PSRR_{10kHz}$	-70dB	-95dB	V
Low Voltage ($\leq 1.4V$)	$V_{on} < 1.4V$	$1.25 \div 1.4V$	V
Monte Carlo	±1.5% for ($3\sigma/mean$)	±1.27% for ($3\sigma/mean$)	V
Core Current Consumption	lowest	[18, 22.5, 36]μA for T [-40, 50, 180]°C	V
Current Consumption	lowest	[11, 15, 21.5]μA for T [-40, 50, 180]°C	V

Table 2.4: Specifications table: Diode-Loop Sub-Bandgap[15].

Comparing the diode loop sub-bandgap core with the first sub-bandgap reference analyzed, we can notice the improvements reported in Table 2.5

	Mixed Mode sub-bandgap	Diode Loop (core only)
Nominal Precision	17ppm/°C	0.5ppm/°C
$PSRR_{DC}$	-45.7dB	-54dB
$PSRR_{10kHz}$	-23.62dB	-47dB
Low Voltage ($\leq 1.4V$)	$1.05V$	$1.25 \div 1.4V$
Monte Carlo	±13.1% for ($3\sigma/mean$)	±1.27% for ($3\sigma/mean$)
Current Consumption @$V_{CC} = 3V$ @ $T = 50$°C	$25\mu A$	$22.5\mu A$

Table 2.5: Improvements table: mixed mode VS diode loop.

We achieved very good performance for all the required specifications while still retaining the low-voltage and low-quiescent current specification.

Low quiescent current for the sub-bandgap core circuit is mostly due to the simpler topology: it does not require a $PTAT^2$ stage, saving all the current needed for biasing all the branches required to implement that stage. Moreover, the mixed mode is based on summing waveform-like signal, like a second order curvature-up with a first order curvature-down compensation: this however is redundant and wastes more current that would actually be needed to get the compensation because the same current (the CTAT current

in our second order compensated bandgap) is used twice over different loads to generate the desired curvature compensations. Current mode, on the contrary, is already compensated *before* the currents get to the load, and the load just transduces the current into voltage, so there is no need for redundant current superposition on different loads to get the compensation.

2.6 Proposed Topology Improvements

After the design of the Diode Loop topology, we moved to the re-design of the proposed second-order compensated sub-bandgap reference. The main focus of the redesign was to improve the yield of the Monte Carlo statistical analysis. It is useless to design a second order compensation if the mismatch errors invalidate the precision that the circuit is capable of. As we saw in Sec. 2.5.8, resistors and mirrors mismatch are our worst enemy. Resistor mismatch can be counteracted by layout solutions, that are more effective if the resistors value are easy to recreate with series or parallel of the same resistor-module. This allows an inter-digitation layout of resistors and their placement in the same area of the circuit, improving matching. Big p-channel MOSFET mirrors also improve matching, and reduce Early effect at the same time. The AC compensation must now be tuned on the new capacitance introduced by the bigger p-channel MOSFETs and the whole circuit must be tuned to compensate the changes in errors like different base currents and Early effects. The result of this redesign is shown in Table 2.6.

	First Design sub-bandgap	Second design sub-bandgap
Nominal Precision	17ppm/°C	4ppm/°C
$PSRR_{DC}$	-45.7dB	-45.7dB
$PSRR_{10kHz}$	-23.62dB	-23.6dB
Low Voltage ($\leq 1.4V$)	$1.05V$	$1.05V$
Monte Carlo	±13.1% for ($3\sigma/media$)	±1.7% for ($3\sigma/media$)
Current Consumption @$V_{CC} = 3V$ @ $T = 50$°C	$25\mu A$	$20\mu A$

Table 2.6: Improvements table: first VS second design.

The most important improvement can be seen in the statistical analysis,

where the error significantly reduced from 13.1% to 1.7%. The error correction brought also a significant improvements over the nominal precision of the temperature coefficient of the reference voltage and a -20% current consumption at ambient. The drawback is a larger area for the whole bandgap reference circuit, that increased of 35%.

Chapter 3

Measurements Vs Simulations

This chapter will discuss the measurements we collected for three test-chip samples, each one holding two of the second-order curvature-corrected bandgap reference, pre-redesign version seen in chapter 2. These measurements will then be compared to the nominal and statistical values expected in order to investigate if the statistical model predictions are respected and to check if the implementation of the solution to the leakage problem worked correctly.

3.1 Analyzed Topologies and Procedures

The two circuits tested in these three test-chip samples are substantially the same circuit showed in 2.1 on page 29. What makes them different is the presence ("LGK" version) or less ("SBG") of a leakage compensation in the PTAT stage, highlighted in Fig. 3.1. The measurement sets and procedures are the following:

Set 1

This set of measurements aims at evaluating the precision of the voltage reference, the current consumption and indirectly the line regulation. It consists in two measurements:

- Connecting the circuit as shown in Fig. 3.2.

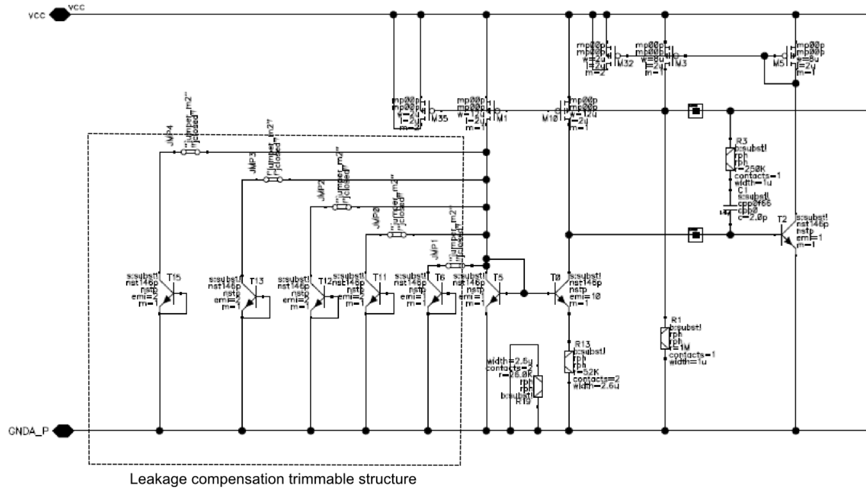


Figure 3.1: Leakage compensation structure.

- Output voltage (that we will call V_{BG}) at four different input voltages [1.5, 2, 3, 4]V from -40°C to 180°C, with steps of 10°C.

- Input current (that we will call I_{in}) at four different input voltages: [1.5, 2, 3, 4]V from -40°C to 180°C, with steps of 10°C.

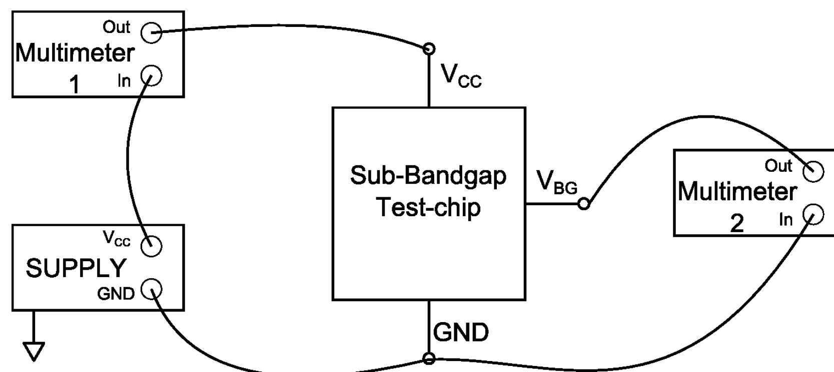


Figure 3.2: Connections setup for “Set 1”.

Set 2

This set exploits the availability of the output pin and the low impedance output stage to measure the value of the resistor ladder (that is the series of the three resistors used for the mixed mode output) of this stage. This will bring a direct comparison between measured resistances and their nominal value. The measure consists in:

- Connecting the circuit as shown in Fig. 3.3
- Measurement of V_{BG} while sinking $I_{test} = 100\mu A$ from V_{BG} pin as shown in Fig. 3.4 to avoid turning on the anti-parallel diodes of the p-channel MOSFET mirrors.
- Measurement of V_{BG} while sinking $I_{test} = 300\mu A$ from V_{BG} pin to check that this measurement is not bias-depending.

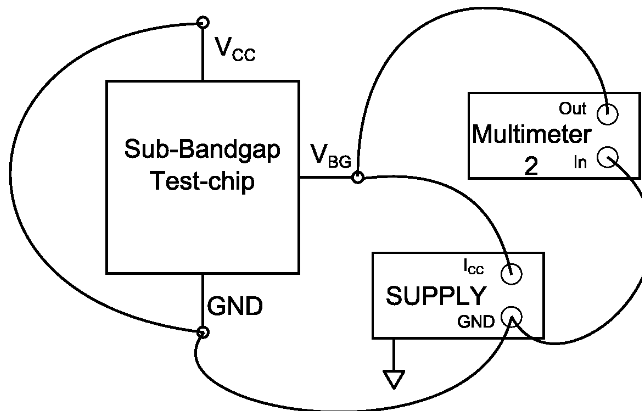


Figure 3.3: Connections setup for “Set 2”.

Set 1 will be used for measurements of the output voltage also after two FIB[1] modification. The modifications will be called FIB1 and FIB2.

FIB1

FIB1 is a cut FIB that disengage the whole compensation stage from the rest of the circuit. Fig. 3.5 shows a picture taken from the microscope.

[1]Focused Ion Beam.

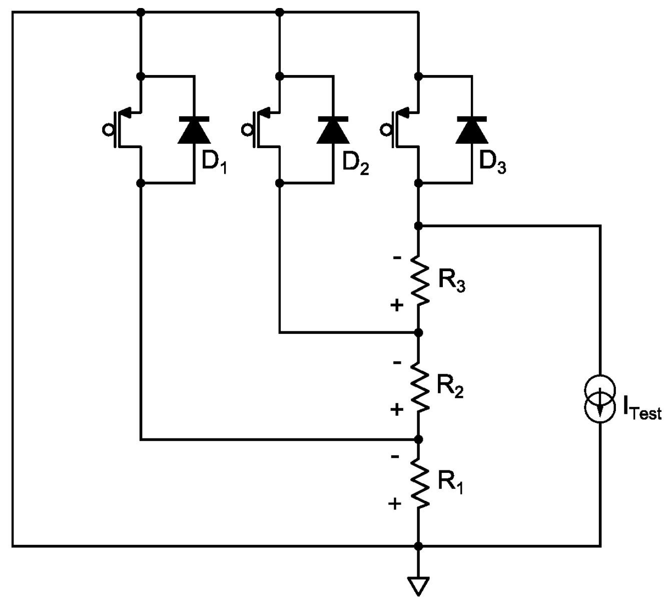


Figure 3.4: Connections setup for “Set 2”, inside view.

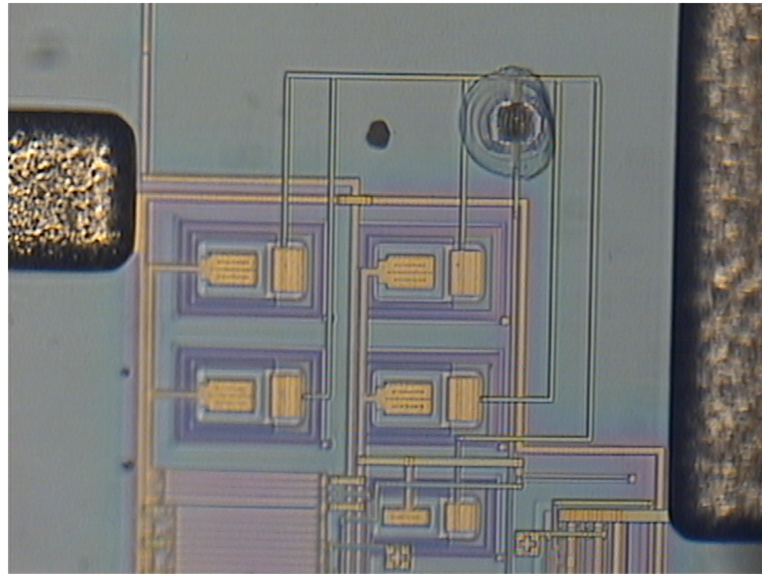


Figure 3.5: FIB cut: microscope picture.

FIB2

FIB2 is a FIB that reconnects 2 NPN transistor of $A_e = 2$ in anti-leakage configuration to the output net of the circuit. The idea, as explained in

chapter 2, is to sink the excess of current mirrored to the load due to leakage errors all over the circuits before it reaches the load itself. This solution is not precise, because there is no connection between the error current and this sunk current, other than they are consequences of the same phenomenon. However, for a certain input voltage and transistor in anti-leakage configuration connected to the load, we can have a good reduction of the V_{BG} error. We will now compare the measurements taken and see if they are inside the statistical range foreseen by the simulator.

3.2 Measurements VS Simulation

3.2.1 Voltage Reference and Current Consumption

Because we are eager to know if our sub-bandgap reference is precise or not, we will evaluate the results that comes from "Set 1". Results are shown in Fig. 3.6, where SBG and LGK version are compared to the same corner, being practically the same circuit. We can see that one of the six voltages

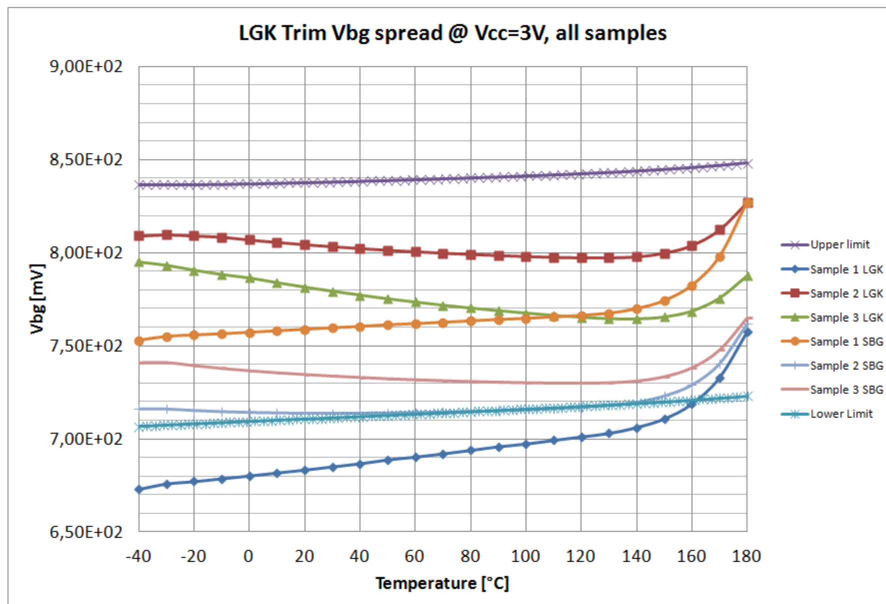


Figure 3.6: V_{BG} measurements VS V_{BG} corner values.

measured is border-line with the lower corner and one is even out of the cor-

ner range. Also the input current sees a little shift towards the upper region of the region of the statistical foreseen values. This is shown in Fig. 3.7. This

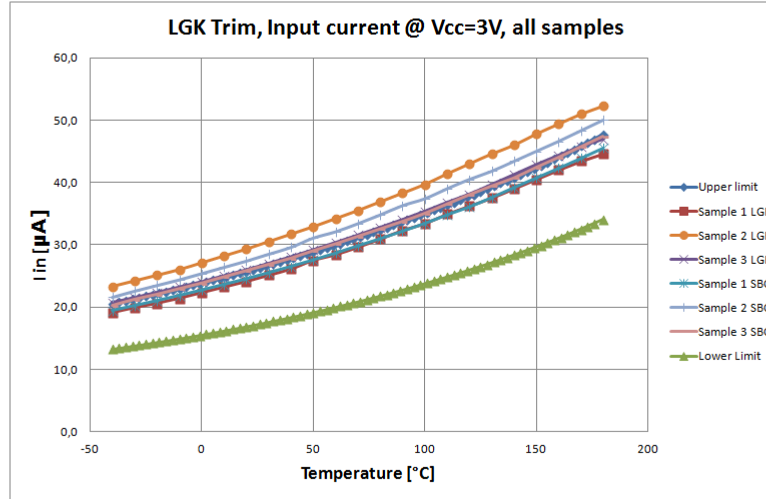


Figure 3.7: I_{in} measurements VS I_{in} corner values.

can find an explanation by measurements done with "Set 2", that has shown a very severe process variation on the output resistance. Fig. 3.8 shows a -20% process variation for this resistance (series). These measurements moves accordingly to the analysis shown in Sec. 2.5.8 on page 96: short channel length and resistor mismatches actually have a big impact on the precision of this voltage reference to the point that the efforts for developing a second order compensation are wasted. So we have a bad compromise between area and precision and power consumption and precision. For instance, a first order sub-bandgap reference with longer channel length and less components would probably have had better performance, needing less area and less power. It is important to remember that a $25\mu m$x$25\mu m$ p-channel MOSFET is still 2.5 times smaller than a NPN with $A_e = 1$ in this technology, so saving a total of 5 A_e (that is the total area occupied only by the NPN necessary for implementing the $PTAT^2$ stage) we would be able to use just about half of that area for all the $25\mu m$x$25\mu m$ p-channel MOSFETs needed to implement a first order current mode sub-bandgap reference, that would show higher precision.

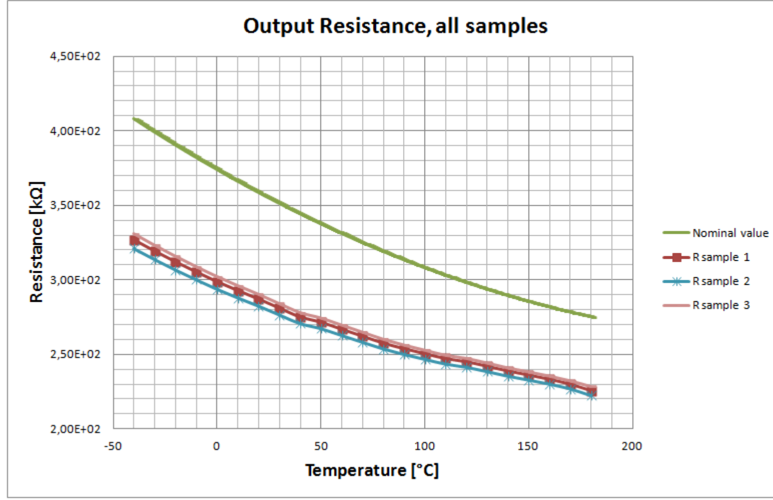


Figure 3.8: Output resistance measurement Vs nominal value.

3.2.2 Line Regulation

It is also possible to evaluate an approximate estimation for the line regulation: adapting eq. (1.3) for a discrete case, we have

$$PSRR_{DC}(V_{IN}) = \frac{\Delta V_{OUT}(V_{IN})}{\Delta V_{IN}} \tag{3.1}$$

Applying this formula to the measurements taken we get the values reported in Tab. 3.1: These values are in line with those reported by the simulator.

@V_{CC}	Sample 1	Sample 2	Sample 3
1.75V	-45.5dB	-46dB	-44.5dB
2.5V	-43.1dB	-44.4dB	-43.1dB
3.5V	-43dB	-44.4dB	-43.1dB

Table 3.1: Line regulation estimation.

3.2.3 Leakage

Some very interesting observation can be extrapolated from the FIB measurement about the leakage currents. Fig. 3.9 resumes all the measurements done and will come in handy to get the analysis easier. FIB1 is indicated as

“Post FIB” and FIB2 as “Post Connection FIB”. At first glance, the FIB2 shows the most effective compensation, and is the less sensitive to supply variations, with a $\Delta V_{BG,FIB2} = 0.74 - 0.7 - DC_{shift} \approx 32mV$ versus $\Delta V_{BG,FIB2} = 0.780 - 0.715 - DC_{shift} \approx 58mV$, where DC_{shift} considers the effect of line regulation. The leakage compensation shown in 3.1 is ineffective for $V_{CC} > 2V$. Despite the fact that the current I_{LC} sunk by the compensation stage generate a negative ΔV_{PTAT}

$$\Delta V_{PTAT} = \ln\left(1 - \frac{I_{LC}}{I_{C2}}\right) \tag{3.2}$$

and, due to the loop in the PTAT stage, a proportional current is also subtracted to the total output load, this strategy does not work. There is no circuital reason for this effect to take place because, even if this temperature range could make transistor work in a “week” operative point (such as low V_{CE}). One of our guess is that a PAD lying very close to one of the NPN anti-leakage transistor, being connected to an high voltage net, turns on a parasitic structure. Acting like a gate for a parasitic MOSFET, it could slightly turn on this parasitic device in a weak-inversion operating mode. This would explain the current injected that counteracts the compensation. FIB2 directly sinks current from the total output current, re-absorbing part of the current introducing by leaking junctions. From this measurement, we could do an interesting evaluation. Let us consider the difference between the value of V_{BG} for FIB2 and for V_{BG} with no compensation applied and let us call it $\Delta V_{BG,F}$: considering V_{BG} constant when biasing the anti-leakage NPN in FIB2, they will sink always the same current $I_{L,FIB2}$, no matter at what voltage the power supply is. What we expect is that, for a fixed temperature, the contribute $\Delta V_{BG,F}$ remains constant, but this is not what it is happening. In fact, we can see a variation over $\Delta V_{BG,F}$. Let us take a look at Tab. 3.2. Knowing the output resistance and its temperature dependence, we can plot the current I_{LC} for each supply voltage, and comparing it with the current calculated by the simulator. The results are shown in Fig. 3.10. We can see a big discrepancy between the simulated model and the measured one, and this will cause the effects of leaking current all over the circuit to be

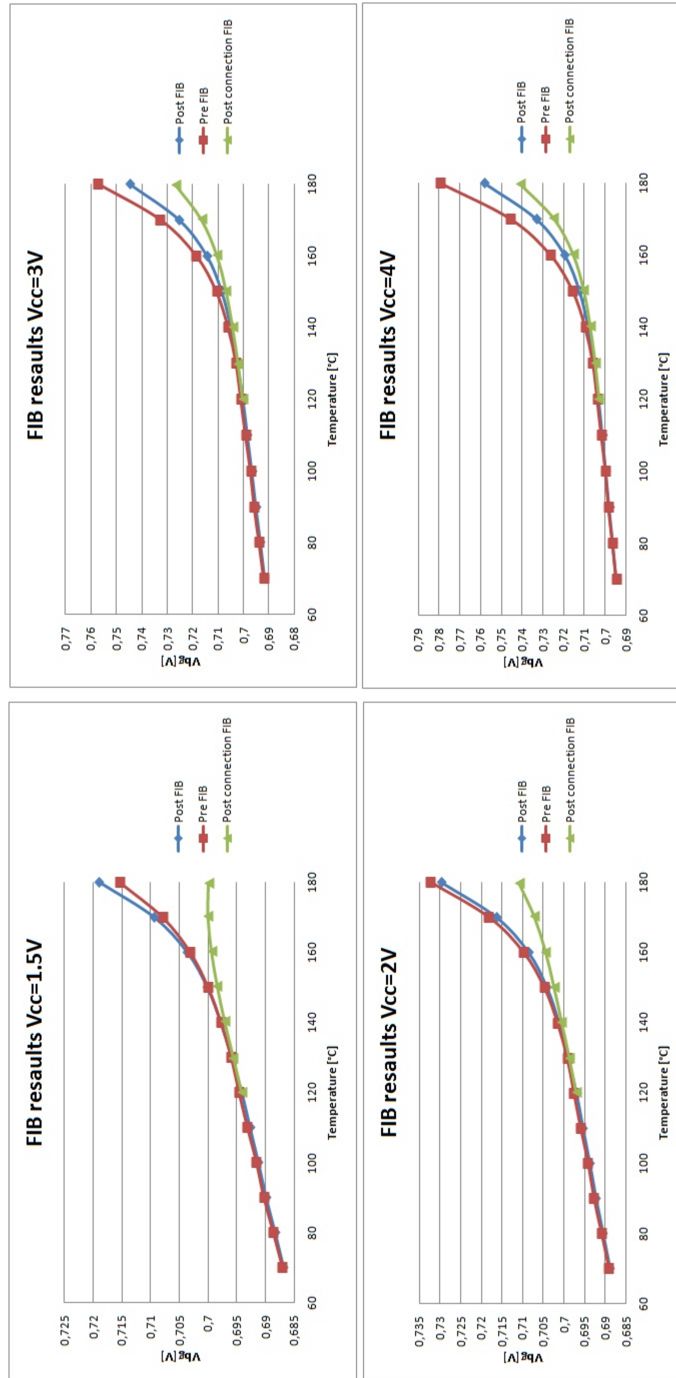


Figure 3.9: FIB measurements.

Temp [°C]	$\Delta V_{BG,F}$ @ $V_{CC} = 4V$ $[mV]$	$\Delta V_{BG,F}$ @ $V_{CC} = 3V$ $[mV]$	$\Delta V_{BG,F}$ @ $V_{CC} = 2V$ $[mV]$
180	17,37	18.34	18.96
170	8,42	8.92	9.28
160	3,99	4.12	4.23
150	1,91	1.87	1.796
140	0,90	0.82	0.72
130	0,59	0.51	0.451
120	0,07	0.05	0

Table 3.2: V_{BG} variation between no compensation and FIB2 modification.

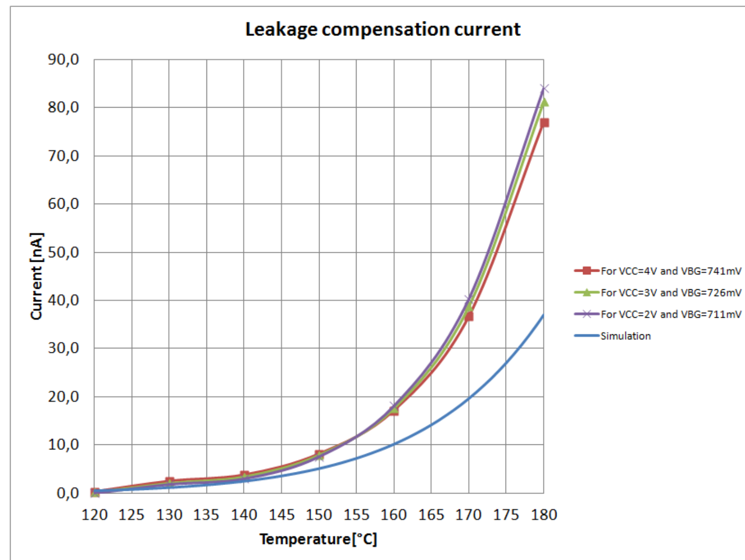


Figure 3.10: Leakage current: measurements VS simulation.

difficult to predict. The small changes between the different values for I_{LC} at different supply voltages can be due to the slightly different bias that both the NPN in anti-leakage configuration and the p-channel MOSFETs of the mirror are subjected to. Down to this values, Early voltage and instrument limits bring non-negligible errors.

Chapter 4

Discussion And Comparison With Other Topologies

In this chapter, we will summarize the results we achieved with this design, and we will compare them with other topologies developed at Infineon Technologies.

4.1 Main Objectives

Our purpose was to study the implementation of a sub-bandgap voltage reference that fulfills very strict specifications. The first important choice is of course the topology, that must be chosen to be the one that suits the specification best. Before starting any implementation (or re-design) activity, silicon physic has been evaluated and all the main sources of errors in the sub-circuits that will constitute the whole project have been analyzed. This pre-evaluation allowed to be aware of the causes of certain effects and to prevent or mitigate them effectively. The result seems very promising, with an astonishing nominal precision, a very robust behavior to process and mismatch deviations, ultra-low voltage and excellent PSRR and Line Regulation performances. However, an electronic designer knows that all comes at a price, and if no trade off are accepted, some parameters will then take on their shoulders all the weight of our choices, most of the time area and

	Compensation	Output Stage	Pre-regulated Supply	Sub-bandgap
Bg1	Exact	Current-Mode	Yes	Yes
Bg2	First Order	Voltage-Mode	Yes	No
Bg3	Second Order	Mixed-Mode	No	Yes
Bg4	First Order	Current-Mode	No	Yes

Table 4.1: Analyzed structures.

current consumption. So, assumed that excellent design has been performed, some topologies will just perform better in some fields than others. This is why we think it is useful to compare different bandgap and sub-bandgap structures, so to understand what their capabilities are.

4.2 Final Comparison and Evaluation

We have compared our developed projects to other voltage references. The analyzed topologies are reported in Tab. 4.1. These topologies have been chosen because they represent the widest combination of different possibilities available.

- Bg1 refers to the pre-regulated current mode sub-bandgap voltage reference designed in this thesis.
- Bg2 is a pre-regulated voltage-mode bandgap reference already developed by Infineon DC ATV STD VREG team.
- Bg3 is the proposed mixed-mode circuit before the re-design process.
- Bg4 is another sub-bandgap we designed using the same concept used for Bg1, with the difference that is it is first-order compensated and that it is designed for very low quiescent current.

Fig. 4.1 shows where each one performs best and is commented in subsec. 4.2.1 and 4.2.2.

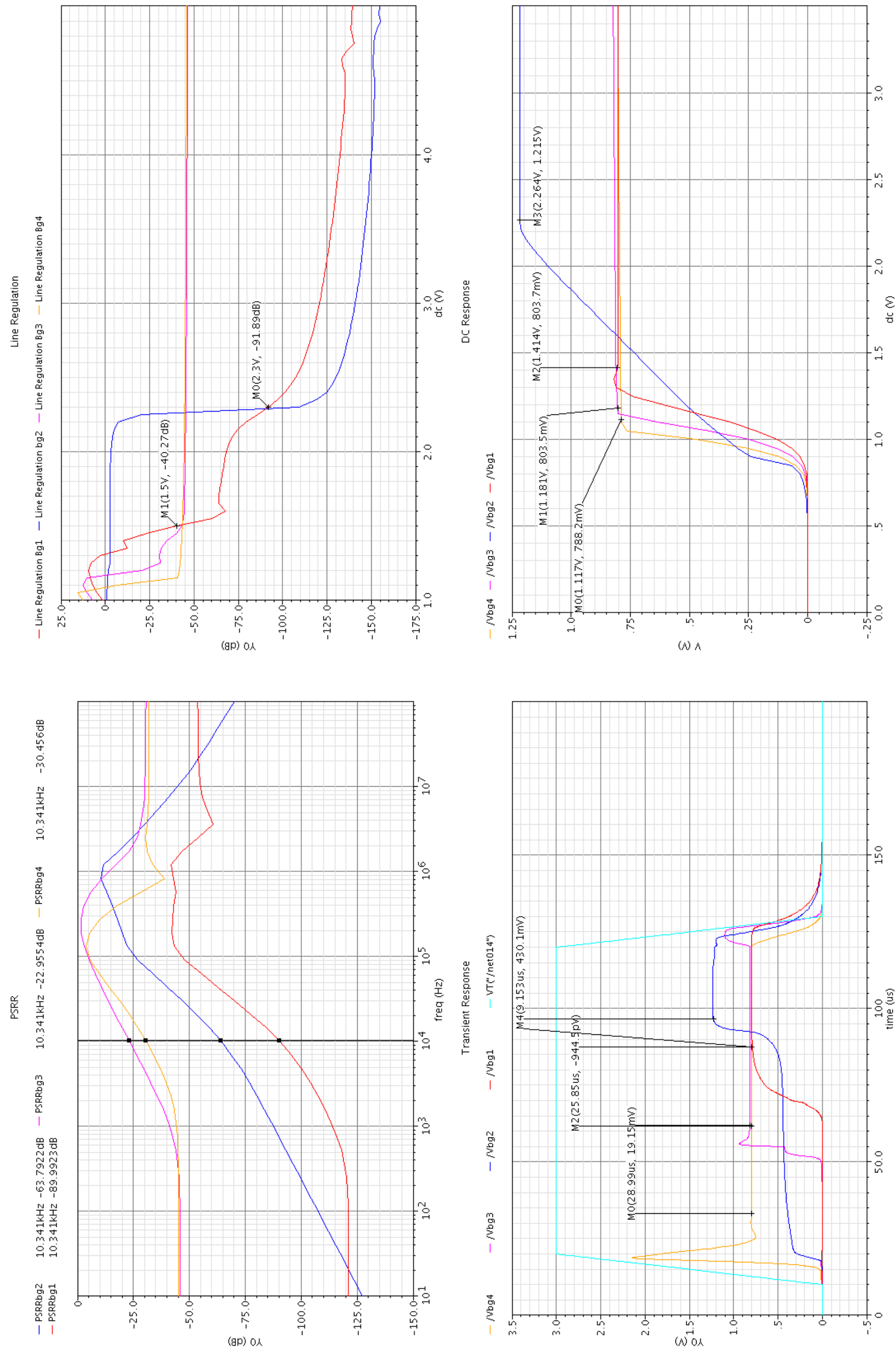


Figure 4.1: Performances comparison.

4.2.1 PSRR and Transient Response

Transient response shows the turn on of the circuits with a $10\mu s$ slope from 0V to 3V of the supply voltage. PSRR[1] values have been taken for a supply voltage of 3V, so all the circuits are in their optimal working bias condition. In this condition, the pre-regulated voltage references have a -75dB boost in the quasi-DC frequency range and then the frequency behavior is dependent on how the compensation is done and what structures implements the voltage reference. We can see that Bg1 is scoring about -27dB against the other pre-regulated circuit at $f = 10kHz$, while the non pre-regulated bandgap references differ because of different compensation choices: Bg3 has a more aggressive AC compensation, meaning more phase margin and less bandwidth where Bg4 has less phase margin but more bandwidth. This reflects on the transient response, that shows much more overshoot for Bg4 and a slower but smoother response for Bg3. Pre-regulated supply shows slower response to power-supply turn on, due mostly to the turn on of the power stage (because of the big gate capacitance associated). Bg2 has a similar structure as Bg1, with smaller power stage, so the reference voltage raise faster and in two different times, but the total amount of time taken is still little greater then Bg1.

4.2.2 Line Regulation and Start-Up

Line regulation shows three distinct zones where each voltage reference works best. The pre-regulated references have better line regulation when also the power supply is turned completely on. If the cascoding stages or the p-channel MOSFETs constituting the mirrors of the bandgap core circuit are not in saturation region, their output resistance drops abruptly and, as shown in subsec. 2.5.5 on page 74, the PSRR drops too. For a supply voltage lower than 1.5V, non pre-regulated references have better PSRR and line regulation, and this threshold must but moved about 300mV higher if considering operation at cold (-40°C), because of the higher $V_{BE}(T)$ offset.

[1]The measurements where performed with 1pF of external capacitance connected between V_{BG} and ground.

After that threshold, pre-regulated voltage regulator have up to -90dB better performance, depending on the pseudo-supply and the bandgap core circuit implementation. Bg1, having a current-mode low-voltage output stage, even if pre-regulated, has minimum voltage of about 1.4V against the 2.3V of Bg2, which features a voltage-mode output stage. Bg1 has practically a minimum voltage that is one overdrive higher than that required for a current-mode or mixed-mode reference.

4.2.3 Nominal Precision and Statistical Analysis

One of the most important tasks for a voltage reference is precision. This is why we struggle for second order compensation, fancy curvature compensation and exact compensation, and every compensation has its impact in terms of area and power consumption. The physical implementation of the chip has the most crucial impact over the performance of the circuit and could impair all the benefit of the most complex compensation. This is also why bandgap references are usually trimmable, but this operation can be risky and will add additional costs. We will analyze different implementation methods to see what are their benefits and their drawback. Fig. 4.2 shows the nominal output voltages for the four examined voltage references. The temperature range stops at 120°C to avoid dragging leakage effects into this evaluation. Bg1 is designed to hit the best compensation possible: in fact, the topology that has been chosen exploits an exact compensation method and an extensive errors prevention and correction has been applied. However, in the layout phase, not all resistors where proportional and easy to be layouted with combinations of resistor modules; however, the modularity has been reached, but at the cost of more area usage. Bg2 has been designed to accept the compromise of a slightly sloped output, but retains easy matching for resistors, that have been layouted easily; moreover, the pre-regulator and the bandgap core circuit have been designed to optimize the area, without the need (like for Bg1), of a real "power stage". Bg2 is implemented in voltage mode, and its precision is strongly depending on resistor matching, while current and mixed-mode needs also an accurate mirror matching. Bg3 is

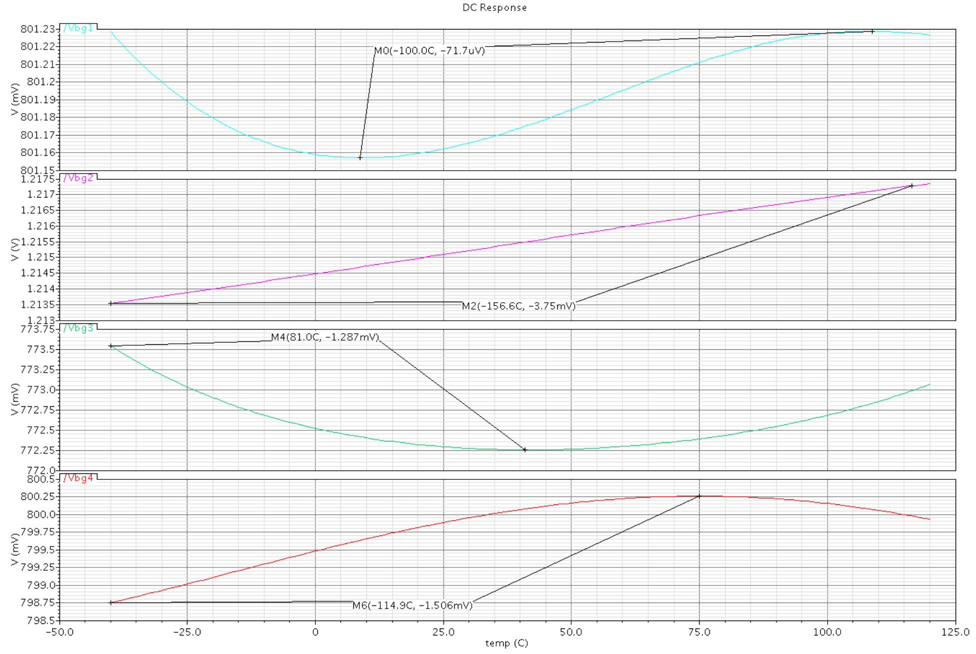


Figure 4.2: Nominal output voltage.

also set for high accuracy, but has been designed for low area usage (minimal length p-channel MOSFETs!). Bg4 is set for maximum precision (for a first order bandgap reference) and minimum current consumption; due to the low area required for its implementation and to mitigate the errors introduced by such a low current usage ($I_{cc}@3V < 6\mu A$, T=25°C), mirrors are sized for high matching precision. We will now cross in Tab. 4.2[2] the simulated results for nominal output voltage, current consumption, area and statistical analysis to make some interesting considerations...

	TC_{ref} [ppm]	Statistical Precision $\left[\frac{3\sigma}{\mu}\%\right]$	Area $[mm^2]$	Current @ $V_{CC} = 3V$
Bg1	0.55	±1.5%	0.14 (a) - 0.24 (b)	13 (a) - 20.5 (b)
Bg2	19	±1.73%	0.13	12
Bg3	10.4	±10%	0.07	21.5
Bg4	12	±1.72%	0.06	5.5

Table 4.2: Precision, area and current consumption performance.

[2](a) and (b) refers respectively to pseudo-supply disengaged or engaged. Current consumption is evaluated at T=25°C

Tab. 4.2 brings very important consideration to light. Let us compare Bg3 and Bg4. They had same PSRR, and now we can see that they have quite the same nominal precision, but BG3 has a very high statistical precision and 75% less current usage, requiring the same area for its implementation. This is caused by the fact that all the room used by the second-order compensation brings some nominal precision, but sink lots of current (due to the its horizontal implementation) and brings no significant statistical precision improvement. On the other hand, a first order compensation will be as flat as the the second order term in the $V_{BE}(T)$ Taylor expansion series is. That is to say that, if we write

$$V_{ref} = K \cdot \left\{ \left[V_{G0npn} + \frac{(\eta - x)V_{T_r}}{2} \right] - \frac{T^2}{T_r^2} V_{Tr} \left[\frac{(\eta - x)}{2} \right] \right\} \qquad (4.1)$$

where K is a constant that takes into account the resistor ratio for the transduced $V_{BE}(T)$, the concavity (hence the precision) of the reference voltage is a non-reachable variable, so should it be changes in $\frac{V_{Tr}}{T_r^2}\left[\frac{(\eta-x)}{2}\right]$, we would have no control over it. The second-order correction nature, on the contrary, allows to control this term, even if the mixed-mode output stage has also a mixed way to weight the currents: let us take a look at Fig. 4.3.

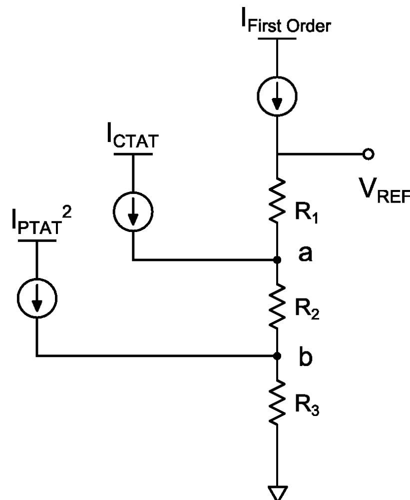


Figure 4.3: Mixed-mode output stage of Bg3.

We can see that changing R_3 (that is usually much smaller than R_1 and R_2) we will alter V_b, but also slightly[3] modify V_a and of course V_{ref}, unbalancing how the three currents are translated into voltages and how this voltages are summed, not just the V_b component.

[3]Slightly because $\frac{R_3}{R_2} < 1$ and $\frac{R_3}{R_1} \ll 1$ so, being the total impedance where $I_{first\ order}$ and I_{CTAT} flow is respectively $(R_1 + R_2 + R_3)$ and $(R_2 + R_3)$, R_3 changes have different impacts on the other two voltage components depending on the weight of R_3 on the total impedance.

Chapter 5

Conclusions

This work focused on studying the different implementations of sub-bandgap structures and all the related source of errors, both systematic and random. Our investigation began from silicon physical properties, went through the solution of many analog problems and analyzed how to reduce the impact of statistical events (like components mismatch) over the specification of reliability and precision we aimed to achieve. With all these knowledge, it was possible to design a new structure that aims at representing a state-of-the-art device between the ultra-low voltage references. However, no structure will be definitive: as we showed in chapter 4, every structure has its strength and weakness, and it should be chosen accordingly to the system into which it will work. Our hope however is that this work could have shown how to deal with the many problematics related to the implementation of a high performance ultra-low voltage sub-bandgap circuit and be itself a ”reference” and an instrument for any engineer who needs to design one.

Appendix A

Appendix

A.1 Base Emitter Voltage Calculation

A.1.1 General $V_{BE}(T)$ expression

$$\begin{cases} V_{BE}(T) = \frac{kT}{q}\ln\left(\frac{Ic(T)}{I_S(T)}\right) & (1) \\ V_{BE}(T_r) = \frac{kT_r}{q}\ln\left(\frac{Ic(T_r)}{I_S(T_r)}\right) & (2) \end{cases} \tag{A.1}$$

$$\begin{cases} T_r \cdot V_{BE}(T) = \frac{kT}{q}\ln\left(\frac{Ic(T)}{I_S(T)}\right)\cdot T_r & (1) \\ T \cdot V_{BE}(T_r) = \frac{kT_r}{q}\ln\left(\frac{Ic(T_r)}{I_S(T_r)}\right)\cdot T & (2) \end{cases} \tag{A.2}$$

Now, subtracting eq(1) and eq(2) leads to

$$V_{BE}(T) = \frac{T}{T_r}V_{BE}(T_r) + \frac{kTT_r}{qT_r}\ln\left(\frac{Ic(T)}{I_S(T)}\right) - \frac{kTT_r}{qT_r}\ln\left(\frac{Ic(T_r)}{I_S(T_r)}\right) \tag{A.3}$$

$$= \left(\frac{T}{T_r}\right)\left\{V_{BE}(T_r) + \frac{kT_r}{q}\left[\ln\left(\frac{I_S(T_r)}{I_S(T)}\frac{I_C(T)}{I_C(T_r)}\right)\right]\right\} \tag{A.4}$$

A.1.2 Precise $V_{BE}(T)$ calculation

Expliciting all temperature-related parameters in 1.9 on page 7 lead to

$$\begin{aligned}\frac{I_S(T_r)}{I_S(T)} &= \frac{BT_r^3 \exp\left(\frac{-qV_G(T_r)}{kT_r}\right)\cdot \bar{D}(T_r)}{BT^3 \exp\left(\frac{-qV_G(T)}{kT}\right)\cdot \bar{D}(T)} \\ &= \frac{T_r^3 \exp\left(\frac{-qV_G(T_r)}{kT_r}\right)\cdot V_{T_r}\cdot CT_r^{-n}}{T^3 \exp\left(\frac{-qV_G(T)}{kT}\right)\cdot V_T\cdot CT^{-n}} \\ &= \frac{T_r^{(4-n)} \exp\left(\frac{-qV_G(T_r)}{kT_r}\right)}{T^{(4-n)} \exp\left(\frac{-qV_G(T)}{kT}\right)}\end{aligned} \tag{A.5}$$

then, for $\eta = 4 - n$

$$\begin{aligned}\ln\left(\frac{I_S(T_r)}{I_S(T)}\right) &= \ln\left[\left(\frac{T}{T_r}\right)^{\eta}\right] + \ln\left[\frac{\exp\left(\frac{-qV_{G0}(T_r)}{kT_r}\right)}{\exp\left(\frac{-qV_{G0}(T)}{kT}\right)}\right] \\ &= \eta\ln\left(\frac{T_r}{T}\right) + \ln\left[\exp\left(\frac{V_{G0}(T)}{V_T} - \frac{V_{G0}(T_r)}{V_{T_r}}\right)\right] \\ &= -\eta\ln\left(\frac{T}{T_r}\right) + \frac{V_{G0}(T)}{V_T} - \frac{V_{G0}(T_r)}{V_{T_r}}\end{aligned} \tag{A.6}$$

Using (A.6) in (A.4) we get

$$\begin{aligned}V_{BE}(T) &= \frac{T}{T_r}\cdot V_{BE}(T_r) + \frac{T}{T_r}V_{T_r}\left(-\eta\ln\left(\frac{T}{T_r}\right) + \frac{V_G(T)}{V_T} - \frac{V_G(T_r)}{V_{T_r}}\right) \\ &\quad + \frac{T}{T_r}V_{T_r}\ln\left(\frac{I_C(T)}{I_C(T_r)}\right) \\ &= \frac{T}{T_r}\cdot V_{BE}(T_r) - V_T\eta\ln\left(\frac{T}{T_r}\right) + V_G(T) - \frac{T}{T_r}V_G(T_r) \\ &\quad + V_T\ln\left(\frac{I_C(T)}{I_C(T_r)}\right)\end{aligned} \tag{A.7}$$

A.2 Linear Regualtor Analysis

A.2.1 Block Diagram of LDO Linear Regulator

This is the explicit calculation for the Block diagram seen in Fig. 1.12 on page 25:

$$\begin{aligned} \left[(V_O H - L V_{in}) A + V_{in} M B + V_{in} C\right](-P) + V_{in} Q &= V_O \\ -V_O H P A + L P A V_{in} - V_{in} M B P - V_{in} C P + V_{in} Q &= V_O \end{aligned} \tag{A.8}$$

$$V_{in}\left[LPA - MBP - CP + Q\right] = V_O(1 + HPA) \tag{A.9}$$

so we get

$$\frac{V_O}{V_{in}} = \frac{P(LA - MB - C) + Q}{1 + HPA} \tag{A.10}$$

A.2.2 PTAT "W" transfer function calculation

In Fig. A.1 is reported the small signal equivalent model for the NPN couple of the PTAT stage.

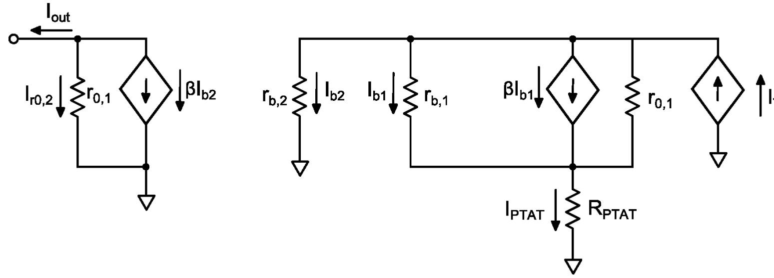


Figure A.1: NPN mirror effect, small signal schematic.

We can write

$$I_T = \beta_1 I_{B1} + I_{B1} + I_{B2} = (\beta_1 + 1) I_{B1} + I_{B2} \tag{A.11}$$

$$I_{PTAT} = (\beta_1 + 1) I_{B1} \tag{A.12}$$

$$I_{B2} = \frac{((\beta_1 + 1) I_{B1} R_{PTAT} + I_{B1} r_{b,1})}{r_{b,2}} \tag{A.13}$$

Using (A.13) in (A.12), we get

$$I_{B2} = (I_T - I_B) \cdot \frac{R_{PTAT}}{r_{b,2}} + \frac{(I_T - I_B)}{(\beta_1 + 1)} \cdot \frac{r_{b,1}}{r_{b,2}} \tag{A.14}$$

$$I_{B2}\left(1 + \frac{R_{PTAT}}{r_{b,2}} + \frac{r_{b,1}}{r_{b,2}(\beta_1 + 1)}\right) = I_T\left(\frac{R_{PTAT}}{r_{b,2}} + \frac{r_{b,1}}{r_{b,2}(\beta_1 + 1)}\right) \tag{A.15}$$

Substituing the values from spectre operative point, we get

$$\frac{\beta_2 I_{B2}}{I_T} \approx \frac{I_{out}}{I_T} = 3.15 \tag{A.16}$$

A.3 AC Stability

The use of a pseudo-supply introduces some complications in the AC analysis of the bandgap core stages. We are then going to do an explicit evaluation of the impedance seen from Q3 collector of Fig. 2.34 on page 86 as an example of how to handle the analysis when a pre-regulator like our pseudo-supply is used. In fact, not all the transfer functions will be explictly evaluated, as this is not the purpose of this thesis. However, we think that it should be usefull to see just this calculation, being slightly different from what we usually run into.

This analysis will consider only the p-channel MOSFETs responsible for the PTAT-current generation-loop, and is extandable to any structure with other pmos that share the same V_{GS}. In this case, the small signal modellization for the circuit is shown in Fig. A.2. The "k" indicates an eventual multiplier equivalent to the number of identical pmos added for mirroring purpouse. If they are not identical, $g_{m,TOT}$ will just be the sum of the different g_m. The same for the output resistor: if they are different, manual calculation for the total impdence must be done.

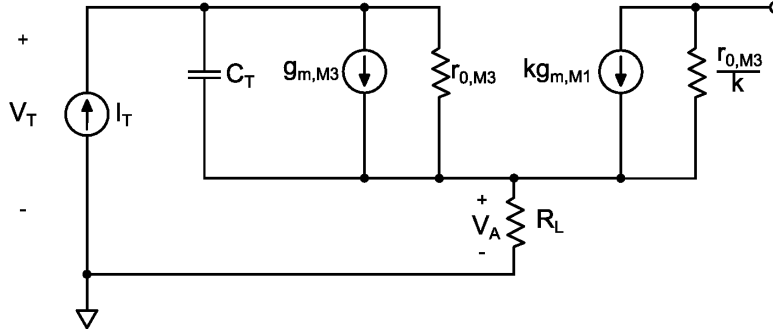


Figure A.2: Small signal circuit for cascoded PTAT stage AC analysis.

Using the notations of Fig. A.2, we can write the following couple of equations

$$\begin{cases} V_A = [I_T + (V_T - V_A)\, kg_{m,M1}]\, R_L & (a) \\ (V_T - V_A)\left(\frac{1}{Z_0} + \frac{1}{g_{m,M3}}\right) = I_T & (b) \end{cases} \tag{A.17}$$

where $Z_0 = \frac{r_{0,M3}}{1+sC_T r_{0,M3}}$. Solving eq. (A.17)(a) for V_A, we get

$$V_A = \frac{I_T R_L + V_T \cdot kg_{m,M1} R_L}{(1 + kg_{m,M1} R_L)} \tag{A.18}$$

Now we can solve eq. (A.17)(b). After some calculation, we get

$$\frac{V_T}{I_T} = Z_{in}(f) = \frac{Z_0\,(1 + kg_{m,M1} R_L)}{1 + g_{m,M3} Z_0} + R_L \tag{A.19}$$

We can immediatly see its DC impendance, being $Z_0 = r_0$, we get

$$Z_{in,DC} = R_L + \frac{(1 + kg_{m,M1} R_L)}{g_{m,M3}} \tag{A.20}$$

Explicitng Z_0 we get

$$Z_{in}(f) = \frac{(1 + kg_{m,M1} R_L) + R_L g_{m,M3} + \frac{R_L}{r_0}}{g_{m,M3}} \cdot \frac{1 + sC_T \left[\frac{r_0}{\frac{r_0}{R_L}\left(1 + kg_{m,M1} R_L\right) + g_{m,M3} r_0 + 1}\right]}{1 + sC_T \left[\frac{r_0}{1 + g_{m,M3} r_0}\right]} \tag{A.21}$$

A.4 Monte Carlo

A.4.1 Resistors Mismatch

Refering the mismatch of R_{NL}, R_{CTAT}, and R_{PTAT} to R_{load}, we can write new ratios for these mismatchig resistors

$$\begin{cases} \frac{\alpha R_{load}}{R_{PTAT}} \\ \frac{\beta R_{load}}{R_{CTAT}} \\ \frac{\gamma R_{load}}{R_{NL}} \end{cases} \tag{A.22}$$

where α, β and γ are the mismatching coefficients of each resistor (they are all different, otherwise we would fall into a process-kind case). The resistors are dimensioned so that

$$V_{ref} = V_{GO,npn} + \left(\frac{R_{load}}{R_{PTAT}}k_1 - \frac{R_{load}}{R_{CTAT}}k_2\right)T + \tag{A.23}$$

$$+ \left(\frac{R_{load}}{R_{NL}}k_3 - \frac{R_{load}}{R_{CTAT}}k_4\right)T\ln\left(\frac{T}{T_r}\right) \tag{A.24}$$

$$\approx V_{GO,npn} \tag{A.25}$$

where k_x are process and temperature indipendent constant. This means that

$$\begin{cases} \frac{R_{load}}{R_{PTAT}}k_1 = \frac{R_{load}}{R_{CTAT}}k_2 = x \\ \frac{R_{load}}{R_{NL}}k_3 = \frac{R_{load}}{R_{CTAT}}k_4 = y \end{cases} \tag{A.26}$$

Combining eq. (A.22) and (A.26) in (A.23), we get

$$V_{ref} = V_{GO,npn} + x\left(\alpha - \beta\right)T + y\left(\gamma - \beta\right)T\ln\left(\frac{T}{T_r}\right) \tag{A.27}$$

No mismatch would imply $\alpha = \beta = \gamma$. Mismatching variations percent are usually beetwen 0.1 and 5%. Assuming the worst case where this percentage we would get and error of

$$\Delta V_{ref} = \pm 0.1x \pm 0.1yV_T\ln\left(\frac{T}{T_r}\right) \tag{A.28}$$

Evaluating x and y for our circuit, their values are

$$\begin{cases} x = \frac{k}{q}\ln(10)\frac{R_{load}}{R_{PTAT}} = 1.273\frac{mV}{°C} \\ y = \frac{k}{q}\frac{R_{load}}{R_{NL}} = 0.15\frac{mV}{°C} \end{cases} \tag{A.29}$$

Using eq. (A.29) in (A.28) we get

$$\Delta V_{ref} = 0.1 \cdot 1.273m \cdot T + 0.1 \cdot 0.15m \cdot T \ln\left(\frac{T}{T_r}\right) \tag{A.30}$$

Ringraziamenti

I would like to thanks, with all my heart:

- Prof. Andrea Bevilacqua and Infineon Technologies, especially Pietro Piersimoni, for letting me live this wonderful experience.
- Marco Piselli, for being such an example to follow, as person and as engineer, and a brilliant man to learn from.
- Marco Flaibani, for his constant presence and for the many things he taught me with so much patience and kindness.
- Infineon STD team, and in particular Giovanni Bisson, Letizia Fragomeni, Fabio Gini, Giovanni Vadalà and Fabio Ballarin for the precious help and happy moments they gave me during these wonderful months.
- My parents Giuseppe and Rosetta, who encouraged, supported and helped me greatly, so that I could focus totally in this work.

Bibliography

[1] Advanced Engineering Course On Practical Aspects in Mixed-Signal ICs, EPFL Premises, Lousanne, Switzerland, August 22-26 2011

[2] S.M.Sze, Physics of Semiconductor Devices, 2nd Edition

[3] Yannis P. Tsividis, "Accurate Analysis of Temperature Effects in $I_C - V_{BE}$ Characterstics with Application to Bandgap Reference Sources", IEEE journal of solid state circuits, Vol SC-15, No. 6, December 1980.

[4] Y. P. Varshni, "Temperature Dependance of The Energy Gap in Semiconductors", Physica, vol.34, 1967.

[5] W. Bludau, A. Onton, W. Heinke, "Temperature Dependance of Banda Gap in Silicon", J. Appl. Phys., vol. 45, no.4, 1974

[6] C. D. Thurmond, "The Standard Thermodynamic Functions for the formation of electrons and holes in Ge, Si, GaAs, and GaP", J. Electrochem. Soc.: Solid State Sci. and Technol., vol.122, no. 8, 1975

[7] Gabriel A. Rincòn-Mora, Voltage References,Wiley-Interscience, 2000.

[8] Gray Mayer, Et., Al., Analysis And Design Of Analog Integrated Circuits, 4Th Ed, 2001

[9] R. Muller, T. Kamins, "Device Electronics for Integrated Circuits", Third Edition, 2003.

[10] P. Rajarshi, A. Patra, S. Baranwal, K. Dash, "Design of Second-Order Bandgap mixed-mode Voltage reference Circuit for Low Voltage Applications, 2005.

[11] D. F. Hilbiber. "A New Semiconductor Voltage Standard". In ISSCC Digest Technical Papers, Vol. 7, 1967.